УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА

Аутоматска анализа студентских повратних информација на српском језику ради унапређења образовног софтвера и разумевања емоција у учењу

ДОКТОРСКА ДИСЕРТАЦИЈА

Ментор:
проф. др Милан Сегединац

Кандидат:
Драган Видаковић

Нови Сад, 2026. године

УНИВЕРЗИТЕТ У НОВОМ САДУ **ОБРАЗАЦ – 5а**
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА[1]

Врста рада:	Докторска дисертација
Име и презиме аутора:	Драган Видаковић
Ментор (титула, име, презиме, звање, институција):	др Милан Сегединац, редовни професор, Универзитет у Новом Саду, Факултет техничких наука
Наслов рада:	Аутоматска анализа студентских повратних информација на српском језику ради унапређења образовног софтвера и разумевања емоција у учењу
Језик и писмо рада:	Српски (ћирилица)
Физички опис рада:	Унети број: Страница: 194 Поглавља: 7 Референци: 144 Табела: 51 Слика: 27 Графикона: 0 Прилога: 2
Научна област:	Електротехничко и рачунарско инжењерство
Ужа научна област (научна дисциплина):	Примењене рачунарске науке и информатика
Кључне речи / предметна одредница:	обрада природног језика, емоције усмерене на учење, инжењерство захтева засновано на групи корисника, интелигентни системи подучавања, корпус златног стандарда, велики језички модели
Апстракт на језику рада:	Развој образовних технологија и платформи за учење на даљину повећао је доступност образовања, али је нарушио повратну спрегу између наставника и студената. У окружењима са великим бројем корисника, попут интелигентних система подучавања, ручна обрада

[1] Аутор докторске дисертације потписао је и приложио следеће Обрасце:
5б – Изјава о ауторству;
5в – Изјава о истоветности штампане и електронске верзије докторског рада и дозвола за објављивање личних података;
5г – Изјава о коришћењу.
Ове Изјаве се чувају у институцији у штампаном и електронском облику и не кориче се са радом.

	неструктурираних текстуалних повратних информација постаје непрактична, што намеће потребу за аутоматизованим приступом. У дисертацији је обрађена аутоматска анализа студентских повратних информација на српском језику, кроз два домена: инжењерство захтева засновано на групи корисника и детекцију емоција усмерених на учење. Главни циљ је развој и евалуација модела машинског учења чије су перформансе упоредиве са људским анотаторима. Прикупљени коментари су анонимизовани, сегментирани на ниво реченице и анотирани, чиме је формиран корпус златног стандарда од 1850 реченица. Дефинисане су таксономије намере и теме у првом, односно аспекта и емоције у другом домену, што чини четири задатка вишекласне класификације на којима је упоређено више приступа: развој модела од нуле, фино подешавање претренираних *Transformer* модела, промпт инжењерство са великим језичким моделима и ансамбл модела. Предложени тежински ансамбл достиже људски ниво у класификацији намера и аспеката, без статистички значајне разлике, док код сложенијих задатака тема и емоција заостаје за људским учинком, иако са мањом варијансом у одлучивању. Балансирање класа није значајно побољшало макро-просек, али је стабилизовало моделе и побољшало препознавање мањинских класа. Научни допринос чине јавно доступан корпус студентских коментара на српском језику и пратеће анотационе шеме, као референтна тачка за будућа истраживања. Развијени модел омогућава праћење афективних стања студената и ефикасније одржавање образовног софтвера.
Датум прихватања теме од стране надлежног већа:	30.04.2026.
Датум одбране: (Попуњава накнадно институција)	
Чланови комисије: (титула, име, презиме, звање, институција)	Председник: др Милан Видаковић, редовни професор, Универзитет у Новом Саду, Факултет техничких наука Члан: др Јована Видаковић, ванредни професор, Универзитет у Новом Саду, Природно-математички факултет Члан: др Маја Босанац, доцент, Универзитет у Новом Саду, Природно-математички факултет Члан: др Милан Рапаић, редовни професор, Универзитет у Новом Саду, Факултет техничких наука Ментор: др Милан Сегединац, редовни професор, Универзитет у Новом Саду, Факултет техничких наука
Напомена:	

UNIVERSITY OF NOVI SAD
FACULTY OF TECHNICAL SCIENCES

KEY WORD DOCUMENTATION[2]

Document type:	Doctoral dissertation
Author:	Dragan Vidaković
Supervisor (title, first name, last name, position, institution)	dr Milan Segedinac, Full Professor, University of Novi Sad, Faculty of Technical Sciences
Thesis title in English:	Automatic analysis of student feedback in Serbian to improve educational software and understand emotions in learning
Language and script:	Serbian (Cyrillic script)
Physical description:	Number of: Pages: 194 Chapters: 7 References: 144 Tables: 51 Illustrations: 27 Graphs: 0 Appendices: 2
Scientific field:	Electrical and computer engineering
Scientific subfield (scientific discipline):	Applied computer science and informatics
Subject, Key words:	natural language processing, learning-centered emotions, crowd-based requirements engineering, intelligent tutoring systems, gold-standard corpus, large language models
Abstract in English:	The development of educational technologies and distance learning platforms has increased the accessibility of education, but has disrupted the feedback loop between teachers and students. In environments with a large number of users, such as intelligent tutoring systems, manual processing of unstructured textual feedback is impractical, necessitating an automated approach.

[2] The author of the doctoral dissertation has signed the following Statements:
5б – Statement on the authorship,
5в – Statement that the printed and e-version of the doctoral dissertation are identical and authorization to use personal data,
5г – Copyright statement.
The paper and e-versions of Statements are held at the institution and are not included into the printed thesis.

	This dissertation addresses the automatic analysis of student feedback in the Serbian language, across two domains: crowd-based requirements engineering and the detection of learning-centered emotions. The main goal is to develop and evaluate machine learning models whose performance is comparable to that of human annotators. The collected comments were anonymized, segmented at the sentence level, and annotated, resulting in a gold-standard corpus of 1850 sentences. Taxonomies of intention and topic were defined for the first domain, and of aspect and emotion for the second, yielding four multiclass classification tasks on which several approaches were compared: developing models from scratch, fine-tuning pretrained Transformer models, prompt engineering with large language models, and an ensemble model. The proposed weighted ensemble achieves human-level performance in classifying intention and aspect, with no statistically significant difference from human performance. In contrast, for the more complex tasks of topic and emotion, it falls short of human performance, though with lower decision variance. Class balancing did not significantly improve the macro-average, but it stabilized the models and improved the recognition of minority classes. The scientific contribution comprises a publicly available corpus of student comments in Serbian and accompanying annotation schemes, which serve as a reference point for future research. The developed model enables the monitoring of students' affective states and more efficient maintenance of educational software.
Date of endorsement by the scientific board:	30.04.2026.
Date of defence: (Filled in by the institution)	
Thesis defence board: (title, first name, last name, position, institution)	Chair: dr Milan Vidaković, Full Professor, University of Novi Sad, Faculty of Technical Sciences Member: dr Jovana Vidaković, Associate Professor, University of Novi Sad, Faculty of Sciences Member: dr Maja Bosanac, Assistant Professor, University of Novi Sad, Faculty of Sciences Member: dr Milan Rapaić, Full Professor, University of Novi Sad, Faculty of Technical Sciences Supervisor: dr Milan Segedinac, Full Professor, University of Novi Sad, Faculty of Technical Sciences
Note:	

Захвалница

Искрену захвалност дугујем свом ментору, проф. др Милану Сегединцу, на стручном вођењу, подстицајним разговорима и несебичној подршци током израде ове дисертације.

Захваљујем се члановима Комисије на драгоценим саветима и корисним сугестијама које су значајно унапредиле квалитет дисертације.

Посебну захвалност дугујем својој породици на разумевању и подршци током свих ових година.

Резиме

Развој образовних технологија и платформи за учење на даљину знатно је повећао доступност образовања, али је истовремено нарушио непосредну повратну спрегу између наставника и студената. У окружењима са великим бројем корисника, попут масивних отворених онлајн курсева и интелигентних система подучавања, појединачне реакције студената постају статистички невидљиве, а ручна обрада неструктурираних текстуалних повратних информација постаје непрактична и подложна грешкама. Из тог разлога јавља се потреба за аутоматизованим приступом који омогућава систематско прикупљање, обраду и интерпретацију повратних информација студената, чиме се истовремено очувава увид у потребе корисника и приступ афективним стањима студената која традиционални инструменти евалуације тешко обухватају.

У овој докторској дисертацији обрађена је област аутоматске анализе студентских повратних информација на српском језику у интелигентним системима подучавања, кроз два домена. Први домен припада инжењерству захтева заснованом на групи корисника и усмерен је на идентификацију захтева за одржавање и развој образовног софтвера, док се други односи на детекцију емоција усмерених на учење из текстуалних коментара студената. Главни циљ истраживања је развој и евалуација модела машинског учења чије су перформансе упоредиве са перформансама људских анотатора. У складу са тим циљем, испитано је да ли се перформансе модела статистички значајно разликују од људских у оба домена, као и да ли примена техника балансирања класа доводи до статистички значајног побољшања резултата на небалансираним вишекласним класификационим задацима.

Истраживање је спроведено у реалном образовном окружењу, у оквиру платформе за адаптивно учење која се користи у универзитетској настави. Прикупљени студентски коментари су анонимизовани, нормализовани, сегментирани на ниво реченице и анотирани, чиме је формиран корпус златног стандарда од 1850 реченица, од којих 591 припада домену инжењерства захтева, а 1259 домену емоција усмерених на учење. Процес анотације заснован је на *MATTER* методологији, уз учешће четири анотатора по домену, а поузданост анотације проверена је мерењем међуанотаторске сагласности. Дефинисане су две анотационе шеме, са

таксономијама намере и теме за домен инжењерства захтева, односно аспекта и емоције за домен емоција усмерених на учење.

Развој модела обухватио је евалуацију више приступа кроз четири задатка вишекласне класификације: развој модела од нуле, фино подешавање претренираних *Transformer* модела, промпт инжењерство са великим језичким моделима и креирање ансамбла модела који интегрише најуспешнија појединачна решења. Тиме је омогућено поређење различитих приступа обраде природног језика, уз испитивање утицаја техника балансирања и аугментације података на перформансе модела у условима неравнотеже класа.

Резултати евалуације показују да предложени модел тежинског ансамбла остварује најбоље перформансе у свим задацима. Модел достиже људски ниво у задацима класификације намера и аспеката, без статистички значајне разлике у односу на људске анотаторе. Код сложенијих задатака класификације тема и емоција, забележена је статистички значајна разлика у корист људи, али је модел показао мању варијансу у одлучивању. Иако примена техника балансирања класа није довела до статистички значајног побољшања на нивоу макро-просека, детаљна анализа показала је њихову практичну вредност у стабилизацији модела и бољем препознавању мањинских класа. Анализа грешака показала је да највећи део погрешних класификација произилази из преклапања семантички блиских категорија, губитка ширег контекста при издвајању појединачних реченица и специфичних одлика српског језика, као што су висок степен флексије, слободан редослед речи и изостављање дијакритика.

Научни допринос дисертације огледа се у дефинисању јавно доступних анотационих шема и смерница на основу којих је формиран први јавно доступан корпус студентских коментара на српском језику у оба домена. Ови ресурси представљају референтну тачку за будућа истраживања, олакшавајући покретање сличних студија и омогућавајући директно поређење резултата. Развијени модел омогућава трансформацију неструктурираних повратних информација у употребљиве податке за праћење афективних стања студената и ефикасније управљање одржавањем и еволуцијом образовног софтвера. Добијени резултати потврђују да је аутоматизована анализа студентских повратних

информација на српском језику методолошки изводљива и довољно поуздана за интеграцију у реалне образовне системе, а истовремено представљају конкретан допринос трећој мисији универзитета кроз трансфер знања ка локалној академској заједници и индустрији образовних технологија.

Abstract

The development of educational technologies and distance learning platforms has considerably increased access to education. However, it has disrupted the direct feedback loop between teachers and students. In environments with large numbers of users, such as massive open online courses and intelligent tutoring systems, individual student reactions become statistically invisible. In contrast, manual processing of unstructured textual feedback is impractical and error-prone. For this reason, there is a need for an automated approach that enables the systematic collection, processing, and interpretation of student feedback, thereby preserving both insight into user needs and access to students' affective states, which traditional evaluation instruments capture only with difficulty.

This doctoral dissertation addresses the automatic analysis of student feedback in Serbian within intelligent tutoring systems across two domains. The first domain pertains to crowd-based requirements engineering. It focuses on identifying requirements for the maintenance and development of educational software, while the second concerns detecting learning-centered emotions in students' textual comments. The main goal of the research is to develop and evaluate machine learning models whose performance is comparable to that of human annotators. In line with this goal, the study examines whether model performance differs significantly from human performance in both domains, and whether applying class-balancing techniques leads to a statistically significant improvement in results on imbalanced multiclass classification tasks.

The research was conducted in a real educational environment, within an adaptive learning platform used in university teaching. The collected student comments were anonymized, normalized, segmented at the sentence level, and annotated, resulting in a gold-standard corpus of 1850 sentences, of which 591 belong to the requirements engineering domain and 1259 to the learning-centered emotions domain. The annotation process was based on the MATTER methodology, with four annotators per domain, and the reliability of the annotation was verified by measuring inter-annotator agreement. Two annotation schemes were defined, with taxonomies of intention and topic for the requirements engineering domain, and aspect and emotion for the domain of learning-centered emotions.

The model development included evaluating several approaches across four multiclass classification tasks: developing models from scratch, fine-tuning pretrained Transformer models, applying prompt engineering with large language models, and creating an ensemble model that integrates the most successful individual solutions. This made it possible to compare different natural language processing approaches and examine the impact of class balancing and data augmentation techniques on model performance under class-imbalanced conditions.

The evaluation results show that the proposed weighted ensemble model achieves the best performance across all tasks. The model achieves human-level performance on the intention and aspect classification tasks, with no statistically significant difference compared with human annotators. In the more complex tasks of topic and emotion classification, a statistically significant difference in favor of humans was observed, although the model exhibited lower decision variance. Although the application of class balancing techniques did not yield a statistically significant improvement at the macro-average level, a detailed analysis demonstrated their practical value in stabilizing models and improving recognition of minority classes. Error analysis showed that the majority of misclassifications arise from the overlap of semantically close categories, the loss of broader context when extracting individual sentences, and specific features of the Serbian language, such as a high degree of inflection, free word order, and the omission of diacritics.

The scientific contribution of the dissertation lies in the definition of publicly available annotation schemes and guidelines, on which the first publicly available corpus of student comments in Serbian was created for both domains. These resources serve as a reference point for future research, facilitating the initiation of similar studies and enabling direct comparison of results. The developed model enables the transformation of unstructured feedback into usable data for monitoring students' affective states and for more efficient management of the maintenance and evolution of educational software. The obtained results confirm that the automatic analysis of student feedback in the Serbian language is methodologically feasible and sufficiently reliable for integration into real educational systems, and at the same time represent a concrete contribution to the third mission of the university through the transfer

of knowledge to the local academic community and the educational technology industry.

Садржај

Списак слика v

Списак табела vi

Списак скраћеница ix

1. Увод 1

1.1. Садржинска и афективна димензија студентских повратних информација 5

1.2. Предмет истраживања 7

1.3. Циљ истраживања и хипотезе 9

1.4. Структура дисертације 11

2. Теоријске основе истраживања 13

2.1. Квалитативна анализа садржаја 13

2.2. Модели емоција у образовању 15

2.3. Анотација природног језика 19

2.3.1. MATTER методологија 19

2.3.2. Међуанотаторска сагласност 21

2.3.3. Најбоље праксе анотације природног језика 22

2.4. Обрада природног језика 23

2.4.1. Методолошки оквир обраде текста 23

2.4.2. Претпроцесирање текста 24

2.4.3. Статистичка репрезентација текста 26

2.5. Аугментација података у обради природног језика 28

2.6. Машинско учење за класификацију текста 31

2.6.1. Основе класификације текста 32

2.6.2. Евалуација модела 32

2.6.3. Методе оптимизације хиперпараметара 34

2.7. Статистички модели машинског учења 35

2.8. Ансамбли 38

2.9. Вештачке неуронске мреже 43
2.9.1. Рекурентне неуронске мреже 46
2.10. Дистрибуирана репрезентација текста 47
2.11. Transformer архитектура 50
2.11.1. Оригинална архитектура и механизам пажње 51
2.11.2. Претренирани енкодерски модели и трансфер учење 52
2.12. Велики језички модели 55
2.12.1. Архитектура GPT модела 56
2.12.2. Архитектура Gemma модела 58
2.12.3. Промпт инжењерство 60
3. Преглед актуелног стања у области 62
3.1. Инжењерство захтева засновано на групи корисника 62
3.2. Емоције усмерене на учење 66
3.3. Обрада српског језика 68
4. Корпус златног стандарда 71
4.1. Прикупљање података 71
4.2. Методологија анотације 73
4.3. Анотационе шеме 75
4.3.1. Анотациона шема за инжењерство захтева засновано на групи корисника 76
4.3.2. Анотациона шема за емоције усмерене на учење 79
4.4. Карактеристике корпуса 83
4.5. Ограничења корпуса 86
5. Развој модела за анализу студентских повратних информација 88
5.1. Методологија 88
5.2. Развој модела од нуле 89
5.3. Фино подешавање претренираних Transformer модела 91

5.4. Примена промпт инжењерства са великим језичким моделима 91
5.5. Креирање ансамбл модела 93
5.6. Експериментални поступак 93
6. Експериментални резултати и дискусија 96
6.1. Резултати у домену инжењерства захтева заснованог на групи корисника..... 96
6.2. Резултати у домену емоција усмерених на учење 100
6.3. Анализа статистичке значајности резултата 103
6.3.1. Тестирање хипотеза $\mathbf{H_1}$ и $\mathbf{H_2}$ 103
6.3.2. Тестирање хипотезе $\mathbf{H_3}$ 104
6.4. Анализа грешака најуспешнијег модела 106
6.4.1. Анализа грешака по задацима класификације..... 106
6.4.2. Систематска категоризација типова грешака.....110
6.4.3. Утицај структуре коментара на прецизност класификације 113
6.4.4. Лингвистички изазови српског језика 114
6.5. Анализа рачунарске ефикасности 115
6.6. Дискусија доприноса и могућности примене.....117
6.7. Ограничења истраживања и претње валидности 118
7. Закључак..... 121
Литература..... 125
Биографија..... 142
Прилози 144
А. Анотационе инструкције за домен инжењерства захтева заснованог на групи корисника..... 144
1. Кораци анотације и општи савети..... 145
2. Намера 145
3. Тема..... 147

4. Контекст 150
Б. Анотационе инструкције за домен емоција усмерених на учење . 151
1. Кораци анотације и општи савети 152
2. Аспект 152
3. Емоција 155
4. Имплицитност 158
5. Контекст 160

Списак слика

Слика 1 Фазе квалитативне анализе садржаја [34] 15
Слика 2 Динамика афективних стања током учења [27] 17
Слика 3 *MATTER* методологија [44] 20
Слика 4 Методолошки оквир обраде текста у задацима класификације 24
Слика 5 Пример *SVM* модела у дводимензионалном простору за две класе [57] 37
Слика 6 Пример *SVM* модела са меким појасом 37
Слика 7 Линеарна, полиномна и радијално-базна функција језгра 38
Слика 8 Основна структура просте агрегације 39
Слика 9 Илустрација рада метода случајних шума 40
Слика 10 Основна структура *XGBoost* модела 42
Слика 11 Структура вештачког неурона 44
Слика 12 Основне архитектуре *Word2Vec* модела 48
Слика 13 Визуелизација линеарних односа у векторском простору речи 49
Слика 14 Архитектура *Transformer* модела [80] 51
Слика 15 Структура *BERT* модела 53
Слика 16 Структура *Transformer* блокова *GPT* модела [86] 57
Слика 17 Основне стратегије промпт инжењерства 60
Слика 18 Анотациона процедура 74
Слика 19 Хијерархијска организација таксономије *Тема* 79
Слика 20 Хијерархијска организација таксономије *Аспект* 81
Слика 21 Преглед коришћених модела машинског учења 89
Слика А.1 Процес анотације 144
Слика А.2 Вредности атрибута Намера 145
Слика А.3 Вредности атрибута Тема 147
Слика Б.1 Процес анотације 151
Слика Б.2 Вредности атрибута Аспект 153
Слика Б.3 Вредности атрибута Емоција 155

Списак табела

Табела 1 Типичне смернице за интерпретацију вредности *IAA* 22
Табела 2 Матрица конфузије 33
Табела 3 Еволуција *GPT* серије модела 58
Табела 4 *Fleiss*-ова κ након *PoC* итерација 75
Табела 5 Примери из анотационе шеме и смерница за *CrowdRE* 79
Табела 6 Примери из анотационе шеме и смерница за емоције усмерене на учење 83
Табела 7 Дескриптивна статистика корпуса 83
Табела 8 Дистрибуција категорија *CrowdRE* корпуса 84
Табела 9 Дистрибуција категорија корпуса емоција усмерених на учење 86
Табела 10 Структура промпта који се користи за класификацију реченица 92
Табела 11 Просечне макро *F1*-мере за задатке класификације намера и тема кроз 10 независних и стратификованих подела података. Подебљане вредности означавају најбољи учинак унутар приступа, док је свеукупно најбољи резултат додатно подвучен 97
Табела 12 Упоредни приказ прецизности (*P*), одзива (*R*) и *F1*-мере по појединачним класама за људски учинак и најбољи модел у *CrowdRE* домену 99
Табела 13 Просечне макро *F1*-мере за задатке класификације аспеката и емоција кроз 10 независних и стратификованих подела података. Подебљане вредности означавају најбољи учинак унутар приступа, док је свеукупно најбољи резултат додатно подвучен 101
Табела 14 Упоредни приказ прецизности (*P*), одзива (*R*) и *F1*-мере по појединачним класама за људски учинак и најбољи модел у домену емоција усмерених на учење 102
Табела 15 Статистичко поређење перформанси најбољег модела и људског учинка 104
Табела 16 Статистичка анализа утицаја балансирања класа на перформансе модела 104
Табела 17 Утицај аугментације података на перформансе мањинских класа 105
Табела 18 Репрезентативни примери погрешних класификација у задатку класификације тема 108

Табела 19 Репрезентативни примери погрешних класификација у задатку класификације емоција....110
Табела 20 Дистрибуција типова грешака по задацима и класама....112
Табела 21 Утицај коментара са више реченица на учесталост погрешних класификација по задацима....113
Табела 22 Примери лингвистичких изазова који утичу на перформансе модела....115
Табела 23 Време обуке модела по задатку на једној репрезентативној подели података....116
Табела 24 Утрошак ресурса *GPT-3.5 Turbo* модела по задатку....116
Табела А.1 Примери реченица чија је намера Пријава грешке 146
Табела А.2 Примери реченица чија је намера Захтев за функционалношћу 146
Табела А.3 Примери реченица чија је намера Оцена 146
Табела А.4 Примери реченица чија је тема Цела апликација.... 147
Табела А.5 Примери реченица чија је тема Наставни материјал 148
Табела А.6 Примери реченица чија је тема Задаци 148
Табела А.7 Примери реченица чија је тема Садржај 148
Табела А.8 Примери реченица чија је тема Употребљивост 149
Табела А.9 Примери реченица чија је тема Поузданост 149
Табела А.10 Примери реченица чија је тема Компатибилност 149
Табела А.11 Примери реченица чија је тема Кориснички интерфејс .. 150
Табела А.12 Примери реченица чију намеру или тему није могуће прецизно утврдити без разматрања контекста 150
Табела Б.1 Примери реченица чији је аспект Интерфејс 153
Табела Б.2 Примери реченица чији је аспект Искуство.... 153
Табела Б.3 Примери реченица чији је аспект Софтвер 154
Табела Б.4 Примери реченица чији је аспект Наставни материјал.... 154
Табела Б.5 Примери реченица чији је аспект Задаци.... 154
Табела Б.6 Примери реченица чији је аспект Лекција 155
Табела Б.7 Примери реченица чија је емоција Досада 156
Табела Б.8 Примери реченица чија је емоција Збуњеност 156
Табела Б.9 Примери реченица чија је емоција Фрустрација.... 157
Табела Б.10 Примери реченица чија је емоција Задовољство 157
Табела Б.11 Примери реченица чија је емоција Изненађеност.... 158
Табела Б.12 Примери реченица чија је емоција Ангажованост 158

Табела Б.13 Примери реченица у којима је Аспект исказан на имплицитан начин..... 159
Табела Б.14 Примери реченица у којима је Емоција исказана на имплицитан начин..... 159
Табела Б.15 Примери реченица чији аспект или емоцију није могуће прецизно утврдити без разматрања контекста 160

Списак скраћеница

Скраћеница	Пун назив
ABEA	Aspect-Based Emotion Analysis
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
BoW	Bag-of-Words
Clean CaDET	Clean Code and Design Education Tool
cMOOC	connective Massive Open Online Courses
CrowdRE	Crowd-based Requirements Engineering
DDDM	Data-driven Decision Making
DL	Deep Learning
ENN	Edited Nearest Neighbors
FN	False Negative
FP	False Positive
GloVe	Global Vectors
GPT	Generative Pre-trained Transformer
GRU	Gated Recurrent Unit
IAA	Inter-Annotator Agreement
IDF	Inverse Document Frequency
ITS	Intelligent Tutoring Systems
LLM	Large Language Models
LSTM	Long Short-Term Memory
mBERT	multilingual BERT
ML	Machine Learning
MLM	Masked Language Modeling
MOOC	Massive Open Online Courses
NCR	Neighborhood Cleaning Rule
NER	Named Entity Recognition
NLP	Natural Language Processing
NSP	Next Sentence Prediction
PoC	Proof-of-Concept
POS	Part-of-Speech
PTM	Pre-trained Transformer Models
QCA	Qualitative Content Analysis
RBF	Radial Basis Function

RE	Requirements Engineering
ReLU	Rectified Linear Unit
RMSNorm	Root Mean Square Layer Normalization
RNN	Recurrent Neural Networks
RoBERTa	Robustly Optimized BERT Pretraining Approach
RoPE	Rotary Position Embeddings
SMD	Serbian Morphological Dictionary
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
tanh	Hyperbolic Tangent
TBED	Text-Based Emotion Detection
TF	Term Frequency
TF-IDF	Term Frequency - Inverse Document Frequency
TN	True Negative
TP	True Positive
TPACK	Technological Pedagogical Content Knowledge
UI	User Interface
XGBoost	eXtreme Gradient Boosting
XLM-R	Cross-lingual Language Model
xMOOC	extended Massive Open Online Courses

1. Увод

Образовни системи у савременом дигиталном добу пролазе кроз дубоку трансформацију коју покрећу образовне технологије и развој дигиталних платформи за учење на даљину. Појава масивних отворених онлајн курсева (енгл. *Massive Open Online Courses - MOOC*) и њихова масовна институционализација кроз платформе попут *Coursera*-е, *edX*-а и *Udacity*-ја отворила је могућност да наставне садржаје истовремено прати велики број студената, при чему географске и институционалне границе више не представљају препреку приступу образовању [1]. Овај формат, у литератури познат као проширени масовни отворени онлајн курсеви (енгл. *extended Massive Open Online Courses - xMOOC*), претежно је заснован на традиционалним педагошким принципима преноса знања и разликује се од конективистичких *MOOC*-ова (енгл. *connective Massive Open Online Courses - cMOOC*), чији ће теоријски оквир бити описан у наставку [2]. Њихов утицај није ограничен само на формате наставе, већ обухвата и шире промене у организационим и сервисним моделима образовних институција [3]. Пандемија *COVID-19* додатно је убрзала овај процес тиме што је принудила образовне институције на свим нивоима да у краткорочном периоду пређу на хибридне или потпуно онлајн моделе наставе, чиме су образовне технологије постале саставни део традиционалне наставе, не само њена допуна [4]. Ова трансформација донела је до сада незабележени ниво доступности образовања, али је истовремено увела значајан изазов који се тиче структуре комуникације између наставника и студената.

У традиционалној учионици, наставник непосредно прати реакције студената и њихов напредак у разумевању градива, на основу чега прилагођава приступ, темпо излагања или избор примера. Ова непосредна повратна спрега представља основу за индивидуализован приступ настави и омогућава наставнику да благовремено идентификује студенте који имају потешкоће или су у фази пада ангажованости. У окружењу *MOOC*-ова, где један курс може истовремено похађати десетине хиљада студената [5], оваква повратна спрега се прекида: појединачне реакције студената постају статистички невидљиве у

мноштву података, а наставници губе могућност да прате афективна и когнитивна стања сваког појединца.

Иако је овај проблем израженији у *MOOC*-овима, он се у ублаженом облику јавља и у универзитетској настави, где се коментари акумулирају кроз семестре и кроз изворе попут анкета и интерактивних модула. У оба случаја, ручна обрада неструктурираних повратних информација постаје непрактична, а изостанак систематске процедуре отвара питање репликабилности увида до којих се на тај начин долази. Овакав контекст захтева методолошки приступ који омогућава систематско прикупљање, обраду и интерпретацију индивидуалних повратних информација свих студената.

Теоријски оквир *cMOOC* формата представља конективизам, концепт који су његови аутори предложили као нову теорију учења за дигитално доба [6] [7]. Иако неки аутори доводе у питање његов статус формалне теорије учења [8] [9], конективизам знање схвата као мрежу међусобно повезаних чворова и веза, а учење у дигиталном добу описује као процес повезивања кроз ту мрежу, у којој студент, наставник и образовни систем чине комплементарне чворове, а знање настаје и одржава се кроз континуирану размену информација између њих. У оквиру педагошке литературе, окружење учења у дигиталном добу не сматра се неутралним подлогом за пренос знања, већ активним фактором који обликује сам процес учења [3]. Кључни предуслов исправног функционисања ове мреже је слободан проток информација између њених делова, што у контексту образовних технологија значи да повратне информације студената морају бити доступне и обрадиве у разумном временском периоду. Када број студената у мрежи постане превелик за људску обраду, неструктурирани текстуални коментари претварају се у заостале информације које се не могу искористити, чиме се мрежа функционално деградира. Аутоматизована анализа студентских повратних информација стога представља једно од методолошких решења које може очувати функционалност овакве мреже.

Поред разумевања динамике учења у мрежним системима, ефикасан образовни систем захтева и сагледавање структурне сложености наставног процеса. Један од најутицајнијих концептуалних оквира за

анализу ове сложености је оквир технолошког, педагошког и садржинског знања (енгл. *Technological Pedagogical Content Knowledge - TPACK*), који полази од претпоставке да квалитетна настава настаје у пресеку три комплементарне димензије знања [10]. Технолошка димензија обухвата познавање алата и платформи којима се настава одвија, педагошка димензија односи се на методе и стратегије преношења знања, а садржинска димензија подразумева стручно познавање материје која се предаје. У контексту савремених образовних система, у којима технолошки слој представља неодвојиви део наставног процеса, *TPACK* оквир пружа структурирани начин за разумевање међусобних интеракција и идентификацију проблема који могу настати у било којој од три димензије.

Студентске повратне информације у образовним системима представљају извор података, који, у зависности од садржаја коментара, може осветлити сваку од три димензије *TPACK* оквира. Технички проблеми у функционисању платформе, попут грешака у интерфејсу или нестабилних функционалности, сигнализирају слабости у технолошкој димензији. Изражена фрустрација или досада студената током рада са одређеним наставним модулом могу указати на недостатке у педагошком приступу, попут неприлагођене методологије. Потешкоће у разумевању конкретних концепата, са друге стране, откривају проблеме у садржинској димензији, било због неадекватног излагања градива или због погрешних претпоставки о предзнању студената. Аутоматска анализа студентских повратних информација стога не служи само као алат за инжењерство захтева у развоју софтвера, већ као дијагностички инструмент за идентификацију и одржавање равнотеже између три димензије *TPACK* оквира. Поред дијагностичке улоге, систематска анализа повратних информација може довести и до развојних промена у техничким унапређењима, педагошким реформама или ревизији наставног садржаја, чиме се отвара и динамичка перспектива овог оквира.

Поред структурне перспективе *TPACK* оквира, савремена теорија курикулума указује да образовни процес обухвата и димензије које превазилазе техничку, педагошку и садржинску. Реконцептуалистички правац у теорији курикулума заговара схватање курикулума као сложене комуникације која обухвата социјализацију, формирање идентитета и

афективни развој студента, не само пренос когнитивног знања [11] [12]. Из ове перспективе, образовни систем који своди наставу на проверу когнитивних исхода путем стандардизованих тестова губи приступ ширим димензијама учења, попут социјалне укључености, мотивације и емотивног односа према градиву. Студентске повратне информације у интелигентним образовним системима представљају један од ретких сигнала кроз које се ове димензије могу систематски опазити. Аутоматска анализа оваквих повратних информација доприноси обнављању шире димензије образовног процеса коју реконцептуалистички приступ препознаје као суштинску за квалитет учења.

Док се теорија курикулума бави широким оквиром образовног процеса, конкретна анализа учења на нивоу појединца захтева систематичну класификацију образовних исхода. Један од најутицајнијих концептуалних оквира за ову сврху је Блумова таксономија образовних циљева, која учење сагледава кроз три комплементарна домена: когнитивни, афективни и психомоторни. Когнитивни домен обухвата интелектуалне способности попут памћења, разумевања, анализе и синтезе знања, и представља димензију која је најчешће примењена у савременим образовним системима. Афективни домен односи се на емоције, ставове, вредности и мотивацију студента у односу на градиво и процес учења. Психомоторни домен обухвата физичке вештине и координацију и претежно је релевантан за специфичне типове наставе попут лабораторијских вежби или техничког тренинга. Касније ревизија ове таксономије задржала је исту структуру три домена, али је унапредила опис когнитивних процеса и њихових међусобних односа [13] [14].

Иако когнитивни домен доминира у савременим методама евалуације образовних исхода, оригинална Блумова таксономија експлицитно наглашава да дубинско учење захтева ангажовање сва три домена истовремено. Афективни домен у том смислу није допунска димензија учења, већ предуслов за остваривање виших нивоа когнитивне таксономије. Студент у стању афективног дисбаланса, попут продужене фрустрације или досаде, не може ефикасно ангажовати когнитивне процесе попут анализе и синтезе градива, без обзира на предзнање или способности. Савремена истраживања потврђују ову повезаност и

показују да емоције нису одвојене од когнитивних процеса, већ су њима неуролошки и функционално интегрисане [15]. Из ове перспективе, праћење афективних стања студената у образовним системима представља суштински елемент за разумевање квалитета процеса учења, не само његових формалних исхода.

Сагледано из целокупне педагошке перспективе, аутоматска анализа студентских повратних информација добија вишеструку улогу. Она омогућава одржавање комуникационог тока знања у мрежи онлајн учења, представља дијагностички инструмент за оцену равнотеже између технолошке, педагошке и садржинске димензије образовног система, и отвара приступ афективном домену учења који традиционалним инструментима евалуације остаје недоступан. У оквиру постојећих образовних технологија, интелигентни системи подучавања (енгл. *Intelligent Tutoring Systems - ITS*) пружају погодан оквир у којем се аутоматска обрада студентских повратних информација може интегрисати у наставни процес.

Интелигентни системи подучавања представљају интерактивне софтверске платформе засноване на методама вештачке интелигенције, чија је основна улога пружање персонализованог образовног искуства прилагођеног индивидуалним потребама студената. Кључна карактеристика ових система јесте њихова способност да континуирано прате и процењују активности студената, прилагођавају наставни садржај у реалном времену и пружају правовремене информације, у циљу унапређења процеса учења [16].

1.1. Садржинска и афективна димензија студентских повратних информација

У оквиру *ITS*-а, прикупљање и анализа повратних информација студената заузимају централну улогу у континуираном унапређењу система. Ове информације су по својој природи двојаке: садржинска димензија односи се на оцене и утиске студената о методама наставе, квалитету лекција, као и о интерфејсу и функционалности платформе, док се афективна димензија односи на емоционалне реакције студената током интеракције са образовним садржајем [17] [18]. Повратне информације пружају

директан увид у корисничко искуство, помажући наставном особљу и развојним тимовима да боље разумеју потребе студената и ефикасније прилагоде софтвер и едукативне материјале њиховим очекивањима [19] [20]. Свака од ових димензија захтева сопствени методолошки приступ, што чини окосницу овог истраживања.

У литератури уобичајено се прави разлика између квантитативних и квалитативних повратних информација. Квантитативне повратне информације обухватају оцењивање путем нумеричке вредности или скала и због једноставности обраде чешће су заступљене у савременим истраживањима [21]. Насупрот томе, квалитативне повратне информације, иако захтевају сложеније технике анализе, пружају дубљи увид у ставове, емоције, мишљења и потребе студената [22]. Управо због њихове сложености и вредности које пружају, у овом истраживању фокус је на аутоматској анализи квалитативних повратних информација студената. Ова сложеност је додатно изражена код језика са ограниченим ресурсима, попут српског, где стандардни алати за обраду природног језика често не пружају задовољавајуће резултате. Истраживање је спроведено у контексту платформе *Clean Code and Design Education Tool* (*Clean CaDET*) [23], која се користи у настави на универзитетском нивоу.

У оквиру садржинске димензије, благовремене и конструктивне повратне информације корисника од кључног су значаја за програмере јер помажу у идентификацији грешака, препознавању потреба за новим функционалностима и унапређењу корисничког искуства [24]. Међутим, с порастом броја корисника, ручна анализа повратних информација постаје неодржива. Зато је неопходно применити аутоматизоване методе за анализу повратних информација. Инжењерство захтева засновано на групи корисника (енгл. *Crowd-based Requirements Engineering - CrowdRE*) представља приступ унутар инжењерства захтева (енгл. *Requirements Engineering - RE*) који користи аутоматизоване или полуаутоматизоване методе за прикупљање и анализу података од великог броја корисника, у циљу идентификације кључних потреба корисника [25]. У овом истраживању, *CrowdRE* се примењује за аутоматску анализу студентских повратних информација у склопу *ITS*-а, са циљем да се развојном тиму и наставном особљу омогући ефикасна идентификација потребних

функционалности, брзо препознавање грешака и континуирано унапређење система.

У оквиру афективне димензије, емоције које се јављају током интензивних активности учења, познате као емоције усмерене на учење (енгл. *learning-centered emotions*) [26] [27], утичу на различите аспекте процеса учења. На пример, могу утицати на когнитивне механизме и способности задржавања информација [28] [29]. Иако се емоције могу аутоматски детектовати анализом израза лица, карактеристика гласа или можданих сигнала, овакви приступи често изазивају забринутост због нарушавања приватности и нелагодност код студената [30]. Због тога се ово истраживање фокусира на детекцију емоција из текста (енгл. *Text-Based Emotion Detection - TBED*) која је прихватљива алтернатива која и даље омогућава значајне повратне информације. Резултати аутоматске анализе детекције емоција из текста могу бити посебно корисни наставном особљу, јер му омогућава да идентификује наставне садржаје или задатке који изазивају негативне емоције код студената (попут фрустрације или досаде), те да благовремено прилагоде наставне методе, курикулум или материјале ради унапређења квалитета наставе.

За реализацију аутоматизоване анализе квалитативних повратних информација у виду текста кључну улогу има обрада природног језика (енгл. *Natural Language Processing - NLP*). *NLP* технике омогућавају аутоматско издвајање и класификацију значајних информација из великих количина текстуалних података и тиме постављају технолошки темељ за методологију ове дисертације. Методолошки, ово истраживање ослања се на *CrowdRE* за обраду садржинских коментара и *TBED* за анализу афективних реакција, чиме се ове димензије покривају на истом корпусу студентских повратних информација.

1.2. Предмет истраживања

Главни предмет овог истраживања је аутоматска анализа студентских повратних информација на српском језику, са фокусом на емоције усмерене на процес учења и на идентификацију захтева релевантних за унапређење образовног софтвера. Истраживање се спроводи у специфичном домену *ITS*-а, где се текстуални коментари корисника

користе као примарни извор података за разумевање корисничких потреба и афективних стања током учења.

Методолошки аспект предмета истраживања усмерен је на превазилажење уоченог недостатка стандардизованих и транспарентних процедура анотације података. С обзиром на то да је у литератури идентификовано одсуство јасних инструкција за анотаторе и изостанак евалуације међуанотаторске сагласности (енгл. *Inter-Annotator Agreement - IAA*), овај рад ставља акценат на развој златног стандарда за српски језик. Ово обухвата дефинисање прецизних анотационих шема усклађених са образовним контекстом и проверу поузданости анотација применом *IAA* метрика, чиме се тежи превазилажењу проблема ниске репликабилности резултата у областима *CrowdRE* и *TBED*.

Техничка димензија истраживања обухвата евалуацију и поређење различитих архитектура за обраду природног језика, у распону од статистичких модела машинског учења до савремених великих језичких модела. Посебан фокус је на испитивању адекватности евалуационих метрика за вишекласне класификационе задатке, с обзиром на то да проста тачност често маскира стварне перформансе на мањинским класама. У том смислу, предмет истраживања укључује и анализу утицаја техника за балансирање података на ефикасност модела у условима изражене неравнотеже класа, што представља значајан истраживачки јаз у постојећим студијама.

На крају, предмет истраживања обухвата технолошку и друштвену димензију која се односи на превазилажење недостатака висококвалитетних језичких ресурса за српски језик. Као ресурсно сиромашан језик, српски се суочава са значајним ограничењима при примени савремених *NLP* метода. Иако савремени велики језички модели доносе значајан напредак у обради свих језика, њихова непосредна примена у образовном контексту има значајна ограничења. Општи језички модели нису специјализовани за специфичне доменске вокабуларе и нијансе, што ограничава њихову примену без додатне адаптације на доменски специфичним подацима [31]. Овај рад зато тежи да, кроз креирање доменски специфичног корпуса и модела, истовремено успостави поуздану основу за научно поређење различитих приступа и

створи практичне ресурсе доступне локалној академској заједници и индустрији образовних технологија. Креирањем отворених ресурса доступних истраживачкој заједници, овакав допринос концептуално је близак схватању треће мисије универзитета у којем се академски рад препознаје као допринос ширем друштвеном добру [32] [33].

1.3. Циљ истраживања и хипотезе

Основни циљ овог истраживања је развој и евалуација модела за аутоматску анализу студентских повратних информација на српском језику, са перформансама које су упоредиве са људским анотаторима. Фокус је на дубљем разумевању емоционалног искуства учења и побољшању функционалности постојећег *ITS*-а *Clean CaDET*. Модел треба да успешно обрађује неструктуриран текст специфичан за студентску комуникацију на српском језику, који се одликује варијабилном дужином и стилом, са мешавином писама (ћирилица и латиница), честим изостављањем дијакритичких знакова, повременим убацивањем енглеских израза, употребом жаргона и скраћеница, неуједначеном интерпункцијом и правописом, као и елементима попут емотикона.

Да би се остварио овај главни циљ, истраживање је подељено на следеће задатке:

1. Дефинисање теоријског оквира и шема анотације: избор модела емоција усмерених на учење и *CrowdRE* таксономије, након чега следи дефинисање шема и смерница за анотацију које обухватају карактеристичне примере, граничне случајеве и утврђивање нивоа анализе (фраза, реченица или целокупан коментар).
2. Дефинисање методологије анотације: дефинисање поступка анотације који укључује избор алата за сам процес анотације, проналазак и обуку анотатора, избор метрике и дефинисање прага за рачунање *IAA* као и поступак усаглашавања анотација и ревизије смерница.
3. Прикупљање, претпроцесирање и анотирање података: прикупљање и анонимизација студентских коментара, нормализација формата која подразумева конверзију писма на

латиницу и припрему података за изабрани ниво анализе, након чега следи процес анотирања до постизања задатог *IAA* прага.
4. Селекција модела: анализа постојећих приступа за класификацију текста, одабир референтних приступа, уз прецизно дефинисање експерименталне поставке и стратегија за балансирање класа.
5. Развој и оптимизација модела: развој и оптимизација модела за аутоматску класификацију текста у циљу ефикаснијег унапређења образовног софтвера и разумевања емоција корисника током процеса учења.
6. Евалуација перформанси модела и анализа грешака: евалуација развијених модела метрикама примереним небалансираним вишекласним задацима, анализа грешака и утицаја балансирања класа, као и завршно поређење перформанси модела са перформансама људских анотатора.

На основу постављеног циља истраживања дефинисане су следеће хипотезе:

- ***H*$_1$***:* Перформансе модела за аутоматску анализу коментара студената на српском језику, прилагођеног домену образовних технологија у високом образовању, за издвајање захтева за одржавање и развој интелигентних система подучавања, не разликују се статистички значајно од људских перформанси.
- ***H*$_2$***:* Перформансе модела за аутоматску анализу коментара студената на српском језику, прилагођеног домену образовних технологија у високом образовању, за детекцију емоција усмерених на учење, не разликују се статистички значајно од људских перформанси.
- ***H*$_3$***:* Примена техника балансирања класа доводи до статистички значајног пораста метрика примерених небалансираним вишекласним задацима у односу на исту архитектуру модела, и за *CrowdRE* и за *TBED*.

Очекивани резултати истраживања обухватају:

- формирање јасно дефинисаних анотационих шема и смерница за анотацију коментара у оквиру *CrowdRE* и емоција усмерених на учење,

- креирање јавно доступног висококвалитетног анотираног корпуса студентских коментара на српском језику,
- развој поузданих *NLP* модела за аутоматску класификацију студентских коментара, са циљем ефикаснијег унапређења *ITS*-а и бољег разумевања емоција усмерених на учење.

Очекивани резултати имају потенцијал примене како у научном, тако и у индустријском контексту. У научном контексту, резултати у виду развијених анотационих шема, јасних смерница за анотацију и јавно доступног корпуса коментара на српском језику представљају значајан допринос истраживањима у области *NLP*-а, нарочито за језике са ограниченим ресурсима. Демонстрација ефикасности *NLP* техника у класификацији повратних информација и детекцији емоција усмерених на учење пружа солидну основу за будућа истраживања у областима *CrowdRE*-а и текстуалне детекције емоција у образовном контексту.

У индустријском контексту, аутоматизована анализа студентских коментара има велики потенцијал за сектор образовних технологија. *NLP* модели развијени у оквиру овог истраживања имају висок потенцијал за интеграцију у постојеће и будуће образовне технологије, чиме се развојним тимовима омогућава аутоматизована идентификација недостатака и ефикасније планирање будућих унапређења образовног софтвера. Поред тога, наставно особље може користити резултате аутоматске анализе емоција како би правовремено идентификовало проблеме у учењу и прилагодило наставни приступ. Све наведено доприноси укупном унапређењу квалитета образовних услуга и корисничког искуства студената, у складу са трећом мисијом универзитета која подразумева трансфер знања и технолошких решења ка локалној академској заједници и индустрији образовних технологија.

1.4. Структура дисертације

Структура ове дисертације је организована у седам поглавља. Ово, прво поглавље представља увод у област истраживања, дефинише значај анализе студентских повратних информација, предмет, главни циљ истраживања, истраживачке хипотезе, као и очекиване резултате.

Друго поглавље обухвата теоријске основе неопходне за разумевање примењених метода. Оно укључује моделе емоција у образовању, кључне концепте анотације и обраде природног језика, технике аугментације података и евалуације, као и архитектуре класификационих модела.

Треће поглавље даје преглед актуелног стања у области инжењерства захтева заснованог на групи корисника, детекције емоција усмерених на учење и обраде српског језика.

Четврто поглавље описује методологију креирања корпуса студентских повратних информација на српском језику. Оно обухвата прикупљање података, дефинисање анотационих шема, процес анотације, као и карактеристике и ограничења корпуса.

Пето поглавље посвећено је експерименталном раду и развоју модела. Овде су описани методолошки приступи, као и експериментална поставка укључујући технике за балансирање класа и поступке евалуације.

Шесто поглавље приказује резултате евалуације модела и поређење са људским анотаторима ради тестирања хипотеза. Поглавље обухвата и анализу грешака, дискусију о доприносима и могућностима примене, као и критички осврт на ограничења и претње валидности истраживања.

Седмо поглавље представља закључна разматрања, сумира остварене доприносе и указује на могуће правце будућих истраживања.

На крају рада наведен је попис коришћене литературе.

2. Теоријске основе истраживања

Развој модела за аутоматску анализу студентских повратних информација захтева интеграцију више комплементарних научних дисциплина које заједно чине теоријски и методолошки оквир овог истраживања. Ово поглавље пружа свеобухватан преглед концепата неопходних за разумевање процеса трансформације сирових текстуалних података у структурирано знање путем алгоритама вештачке интелигенције.

Поглавље почиње квалитативном анализом садржаја и моделима емоција у образовању, који служе као основа за конструисање валидних анотационих шема и анотацију података. Након тога, фокус се помера на обраду природног језика, укључујући технике претпроцесирања, аугментације и репрезентације текста. Други део поглавља детаљно обрађује алгоритме машинског учења, почевши од статистичких модела, преко вештачких неуронских мрежа, до *Transformer* и великих језичких модела. На тај начин поставља се темељ за разумевање модела који се развијају и тестирају у наставку дисертације.

2.1. Квалитативна анализа садржаја

Квалитативна анализа садржаја (енгл. *Qualitative Content Analysis* - *QCA*) представља систематичну, правилима вођену методологију за анализу текстуалних података, која омогућава идентификацију централних тема, концепата и образаца у тексту. *QCA* почива на јасно дефинисаном процесу кодирања који обезбеђује транспарентност и поновљивост анализе. Основна идеја овог приступа јесте да се текст редукује на информативне сегменте који се систематски сврставају у категорије, чиме се сложени текстуални материјал претвара у структуриран аналитички корпус [34].

Категорије представљају централни елемент *QCA* јер чине аналитичке јединице које истраживачу омогућавају да групише садржај према заједничким значењима. Оне се могу разумети као основни концепти сазнања који указују на оно што је заједничко одређеним стварима, односно термини, ознаке или појмови којима се обухватају међусобно слични елементи посматрани у одређеним аспектима. У контексту анализе садржаја категорије представљају супстанцу истраживања и

темељне градивне елементе будуће теорије. Управо због тога рад са категоријама и развој кодног система (енгл. *coding frame*) представља темељ сваке *QCA* [34]. Квалитет и ефикасност истраживања директно су одређени квалитетом тих категорија, јер истраживање не може бити боље од самих категорија које садрже суштину истраживачког процеса [35].

Поменути кодни систем представља структурирану целину категорија која служи као основни аналитички инструмент за класификацију квалитативног материјала. Он се састоји од главних категорија које заједно покривају значења релевантна за опис и интерпретацију корпуса, а сложеније категорије могу се по потреби додатно разложити на поткатегорије, чиме се гради хијерархијски организован систем. На тај начин се сложени текстуални садржај преводи у сажетији и аналитички употребљив категоријални приказ [36].

У изградњи кодног система, постоје три основна приступа за формирање категорија. Први је концептуално (дедуктивно) развијање категорија, где се категорије изводе из теорије, релевантне литературе или истраживачких питања. Други је подацима вођено (индуктивно) развијање категорија, које започиње отвореним кодирањем и постепеном организацијом категорија све до засићења, након чега се оне систематизују у општије категорије и поткатегорије у складу са садржајем самог текста. Трећи приступ је комбиновано развијање категорија, у којем се иницијални концептуални оквир допуњава индуктивним категоријама како би се прецизније обухватиле карактеристике података [34]. Комбинација дедуктивних и индуктивних стратегија омогућава да кодни систем истовремено буде теоријски утемељен и емпиријски прилагођен конкретном корпусу [36].

Слика 1 илуструје фазе у којима се одвија процес *QCA*, укључујући иницијално читање корпуса, развој прелиминарних категорија, кодирање, ревизију категорија и презентацију резултата. У иницијалној фази истраживач стиче општи увид и одређује прве концептуалне целине. Накнадне фазе омогућавају постепено усаглашавање категорија са текстом, проверу њихове доследности и изградњу стабилне структуре категорија која може бити предмет даље анализе података [34].

Слика 1 Фазе квалитативне анализе садржаја [34]

Пажљиво дефинисане категорије представљају основу за квалитетну анализу текста, јер осигуравају поузданост процеса и омогућавају валидно тумачење садржаја. Управо због тога *QCA* често служи као полазиште за развој кодних система који ће касније бити коришћени у аутоматским системима класификације текста, без обзира на домен или врсту текстуалног корпуса [34] [36]. Водећи се овим принципима, комбиновани приступ *QCA* представља концептуалну основу за развој анотационих шема у овој дисертацији.

Поред методолошког оквира, развој анотационих шема захтева и теоријско разумевање доменских концепата који се анализирају. У контексту емоција усмерених на учење, неопходан је преглед савремених модела емоција у образовању, што је тема наредног поглавља.

2.2. Модели емоција у образовању

Појам емоције представља један од централних концепата савремене психологије, при чему у литератури не постоји јединствена дефиниција већ више теоријских приступа који различито тумаче његову природу и структуру. У основи савремене класификације разликују се три главна

приступа. Теорије основних емоција претпостављају постојање ограниченог скупа основних емоционалних категорија, попут радости, туге, страха, беса, гађења и изненађења, које се сматрају биолошки утемељеним и стабилним кроз различите културе [37]. Димензионални модели емоције представљају као тачке у континуалном простору дефинисаном димензијама пријатности и активације [38]. Трећи приступ полази од претпоставке да емоције настају као одговор на когнитивну процену ситуације у односу на циљеве и вредности појединца [39].

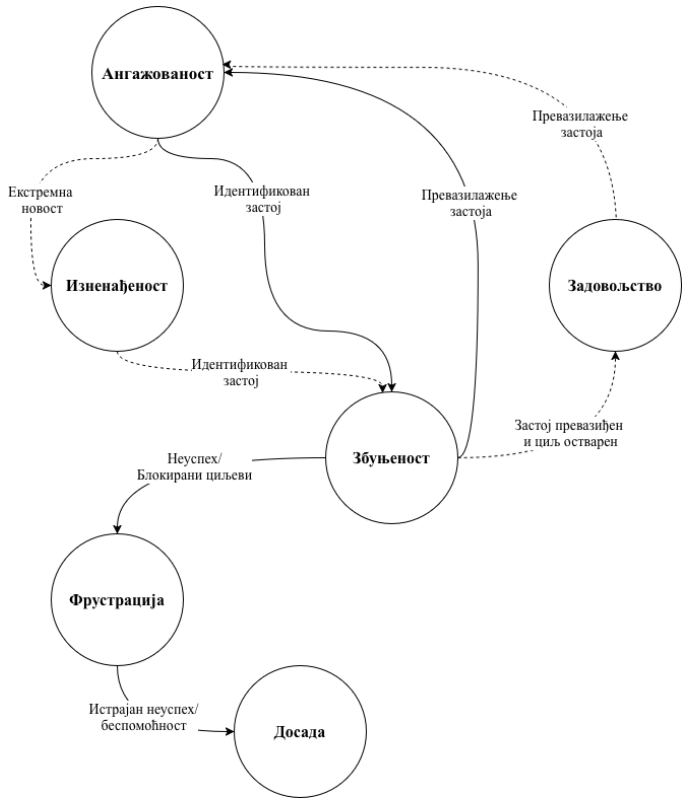
У оквиру овог истраживања, појам емоције тумачи се у складу са трећим приступом, према којем емоције настају као одговор на процену ситуације током учења. Под емоцијом се подразумева афективно стање које се јавља у интеракцији студента са образовним садржајем, а зависи од доживљене тежине задатка, успешности у његовом решавању и сопственог осећаја контроле над процесом учења. Овакво схватање одговара моделу емоција усмерених на учење, на којем је заснована анотациона шема развијена у овом истраживању.

Разумевање улоге емоција у процесу учења представља кључан аспект савремене образовне психологије и интелигентних система подучавања. За разлику од теорија основних емоција, које претпостављају ограничен скуп биолошки укорењених категорија, афективни одговори ученика у контексту учења имају специфичне когнитивне и мотивационе компоненте, те се не могу једноставно подвести под опште психолошке класификације [28]. У ситуацијама сложеног когнитивног напора, решавања проблема или интеракције са наставним материјалом, јављају се емоционална стања која нису предвиђена класичним теоријама, већ су непосредно повезана са самим процесом учења [26] [27] [28].

Емпиријска истраживања у домену образовних технологија показују да се током учења јављају карактеристична когнитивно-афективна стања као што су збуњеност (енгл. *confusion*), фрустрација (енгл. *frustration*), ангажованост (енгл. *engagement*), задовољство (енгл. *delight*) и досада (енгл. *boredom*), која имају специфичне улоге у регулацији когнитивних активности и нису обухваћена традиционалним моделима емоција [26] [27]. Ова стања се јављају у динамичним секвенцама током решавања задатака, а њихови прелази често одражавају кораке у процесу мисаоног

закључивања. Слика 2 илуструје такве афективне трајекторије, при чему се види да емоције попут збуњености природно претходе ангажованости или задовољству када се успешно превазиђе когнитивна баријера. Ова динамика потврђује да емоције у учењу нису статичне, већ су чврсто повезане са интеракцијама ученика и напредовањем кроз задатак [27].

Додатна истраживања на тему доменске специфичности афективних стања показују да у контексту учења није могуће правити једноставну разлику између позитивних и негативних емоција. На пример, фрустрација, иако традиционално негативна емоција, може бити индикатор когнитивног напретка и дубљег ангажовања у задатку. Насупрот томе, досада, такође негативна по валенци, повезана је са одустајањем од активности и падом перформанси, те има јасно деструктивну улогу у процесу учења [40]. Ово наводи на закључак да традиционални, валентно засновани модели емоција не могу поуздано описати комплексан однос између афекта и когнитивних процеса.



Слика 2 Динамика афективних стања током учења [27]

Динамика приказана на слици 2 осликава кључне постулате модела емоција усмерених на учење (енгл. *learning-centered emotions*). Иако је овај модел извршио значајан утицај на област афективног рачунарства у образовању, новија истраживања указују и на његова ограничења. Систематска анализа десет ранијих студија показала је да се прелази између афективних стања, утврђени у различитим културним и образовним контекстима, не подударају увек са обрасцима које предвиђа изворни модел [41]. Међу факторима који доприносе оваквим разликама истичу се различите узрасне групе, типови образовних окружења, домени учења, као и културолошке карактеристике популације у којој су подаци прикупљени. Поред тога, методологија детекције емоција у изворним студијама претежно се ослањала на ретроспективне самопроцене и експертске анотације видео записа, што уноси одређени степен субјективности у саме податке [42].

Међу алтернативним теоријским приступима, посебно место заузима модел контроле и вредности (енгл. *Control-Value Theory of Achievement Emotions*). Овај модел обухвата шири скуп афективних стања повезаних са постигнућем, укључујући задовољство, наду, понос, олакшање, бес, анксиозност, безнађе, стид и досаду [28]. Ипак, његова практична примена захтева сложеније инструменте за детекцију и валидацију емоција [43], што га чини мање погодним за аутоматску анализу студентских текстуалних повратних информација. За потребе овог истраживања изабран је модел емоција усмерених на учење због експлицитне оријентације на осликавање стања кроз језичке сигнале.

Синтеза наведених налаза показује да су емоције релевантне за образовање посебна група афективних стања која се разликују од основних емоција, како по природи тако и по улози у когнитивним процесима. У савременој литератури управо због тога преовладава модел емоција усмерених на учење, односно афективних стања која настају као последица когнитивних изазова, успеха или неуспеха током учења. Овакав модел пружа теоријски адекватну основу за опис и класификацију афективних реакција у образовном контексту, а његова ограничења узета су у обзир приликом дизајна анотационе шеме у овом истраживању.

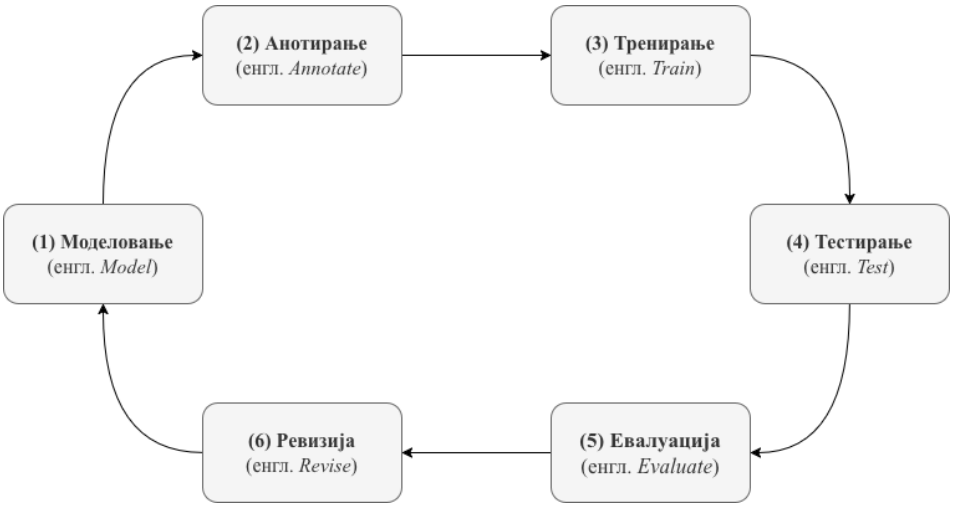
2.3. Анотација природног језика

Након представљених теоријских основа за развој анотационих шема, потребно је представити и формалну методологију њихове примене на текстуалним подацима. Анотација природног језика представља процес систематског додељивања структурних, семантичких или прагматичких ознака над текстуалним подацима са циљем да се необрађени текст претвори у формални ресурс погодан за анализу, тренирање и евалуацију модела [44]. Анотације могу бити дескриптивне, када се једноставно описују језичке или семантичке појаве које се налазе у подацима, или прескриптивне, када анотатори применом формалних правила доносе одлуке у складу са унапред дефинисаним категоријама и критеријумима. У контексту машинског учења и изградње модела, прескриптивне анотације су чешћи приступ јер омогућавају формирање доследног златног стандарда (енгл. *gold standard*) који служи као референтна истина за тренирање и тестирање модела [45].

Поступак анотације је посебно критичан у задацима који укључују апстрактне концепте, као што су емоције, где је неопходно обезбедити међусобно усаглашавање и стабилност анотирања. Савремена литература показује да се неслагања између анотатора најчешће не јављају због погрешних интерпретација, већ због недовољно прецизно дефинисаних задатака или концепата [46]. Управо због тога процес мора бити вођен јасним методолошким принципима и добро структурираном процедуром.

2.3.1. MATTER методологија

Развој поузданог анотационог корпуса представља кључни корак у задацима машинског учења за обраду природног језика, будући да квалитет анотација непосредно одређује квалитет модела који се на том корпусу обучавају. *MATTER* (*Model - Annotate - Train - Test - Evaluate - Revise*) методологија представља један од најприхваћенијих приступа у области анотације природног језика, јер омогућава систематичан развој анотационих шема и смерница. Овај приступ наглашава да се анотација мора третирати као итеративан процес, при чему се шема, смернице и сам корпус постепено унапређују кроз више циклуса [44].



Слика 3 *MATTER* методологија [44]

Слика 3 илуструје овај приступ. У почетној фази (Моделовање) дефинишу се анотациона шема и смернице, односно концептуални оквир, категорије и њихове формалне дефиниције. Ова фаза обухвата разграничење категорија, формулисање позитивних и негативних примера и утврђивање критеријума за граничне случајеве. У фази Анотирања, више анотатора примењује шему на одабрани део корпуса, што служи као практична провера јасноће и применљивости смерница. Фазе Тренирања и Тестирања користе анотирани материјал за иницијално тренирање и проверу модела, чиме се установљава да ли је анотација употребљива у реалним задацима машинског учења. У фази Евалуације, анализирају се неслагања међу анотаторима и перформансе модела, а добијени увиди представљају основу за ревизију шема и смерница. Завршна фаза (Ревизија) затвара циклус кроз унапређење шема и смерница, након чега процес може започети поново [44].

Оваква итеративна структура осигурава стабилност, теоријску заснованост и емпиријску поузданост анотационог корпуса, што га чини погодним као златни стандард за тренирање и евалуацију модела машинског учења. Управо због ових својстава, *MATTER* методологија је одабрана као оквир за развој анотационог процеса у овом истраживању.

2.3.2. Међуанотаторска сагласност

Квалитет анотације се не може проценити само на основу интерних смерница или искустава анотатора, већ је потребно формално измерити колико су независни анотатори доследни у примени истих шема и смерница. Овај показатељ, познат као међуанотаторска сагласност (енгл. *Inter-Annotator Agreement - IAA*), представља темељну меру поузданости анотационог процеса у анотацији природног језика. Висока *IAA* вредност указује да су категорије добро дефинисане и стабилно применљиве, док ниска вредност упућује на потребу за ревизијом шема и смерница [44].

Избор метрике зависи од броја анотатора, природе података и типа категорија. Када на истом корпусу раде два анотатора, најчешће се користи *Cohen*-ова κ, која мери колико је посматрана сагласност већа од оне која би се очекивала случајно [47]:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{1}$$

где је p_o посматрана сагласност, а p_e очекивана сагласност заснована на расподели категорија.

У случајевима када анотацију обавља више од два анотатора, примењује се *Fleiss*-ова κ, која представља генерализацију *Cohen*-ове κ:

$$\kappa = \frac{\bar{P} - \bar{P_e}}{1 - \bar{P_e}} \tag{2}$$

где $\bar{P}$ означава просечну посматрану сагласност, а $\bar{P_e}$ просечну очекивану сагласност.

Вредности ових метрика се крећу од -1 (савршено неслагање) до 1 (савршено слагање). Табела 1 приказује типичне смернице за интерпретацију ових вредности кроз шест нивоа сагласности. Ове смернице могу послужити као оквирна референца, али коначни прагови зависе од сложености и субјективности задатка [48].

Табела 1 Типичне смернице за интерпретацију вредности *IAA*

Вредност	**Интерпретација**
(-1, 0)	лоше слагање
(0.01 - 0.2)	благо слагање
(0.21 - 0.4)	прихватљиво слагање
(0.41 - 0.6)	адекватно слагање
(0.61 - 0.8)	значајно слагање
(0.81 - 1)	готово савршено слагање

У овом истраживању, имајући у виду да у анотацији учествују четири анотатора, за процену поузданости коришћена је *Fleiss*-ова κ, а постигнуте вредности интерпретиране су у складу са горе наведеним смерницама.

2.3.3. Најбоље праксе анотације природног језика

Поступак анотације мора бити методолошки чист, репликативан и транспарентан како би се обезбедила употребљивост корпуса у истраживачке и практичне сврхе. Најзначајније препоруке формулисане у литератури могу се систематизовати у следеће принципе [44] [46]:

1. Прецизно дефинисана анотациона шема: категорије морају бити јасно формулисане, међусобно разграничене и теоријски утемељене. Неодређене дефиниције представљају најчешћи узрок ниске *IAA*.
2. Свеобухватне и доследне смернице: упутства треба да садрже дефиниције, позитивне и негативне примере, граничне ситуације и поступке за неодређене случајеве. Ово осигурава међусобну сагласност у примени категорија.
3. Тренинг и пилот-анотација: пре главне анотације неопходно је спровести пилот фазу, са дискусијом спорних случајева и ревизијом смерница. Ово је у складу са *MATTER* методологијом и значајно подиже квалитет корпуса.

4. Систематско решавање неслагања: неслагања између анотатора не треба третирати као грешку у анотирању, већ као показатељ проблема у шеми или смерницама.
5. Детаљна документација: коначан корпус мора садржати методологију, шему, смернице, *IAA* вредности и ограничења. Ово је предуслов за научну репликабилност.

Оваква организација рада омогућава да анотација буде стабилна, методолошки заснована и погодна као ресурс за развој модела машинског учења у наставку истраживања. Сви наведени принципи примењени су у развоју анотационог процеса у овом истраживању.

2.4. Обрада природног језика

Анотирани корпус, развијен према претходно описаној методологији, представља полазиште за процес аутоматске обраде. Обрада природног језика (енгл. *Natural Language Processing - NLP*) је интердисциплинарна област рачунарства, вештачке интелигенције и лингвистике која се бави развојем метода и алгоритама за аутоматску анализу, интерпретацију и генерисање људског језика. Ова област омогућава рачунарима да раде са текстуалним и говорним подацима који су по својој природи неструктурирани, вишезначни и подложни великим варијацијама у значењу и форми [49] [50].

Ово поглавље представља методолошки оквир за обраду текста, укључујући кораке претпроцесирања, избор техника представљања текста и интеграцију ових поступака у системе за аутоматску анализу. Посебан нагласак ставља се на процедуре кључне за задатак класификације текста, које чине окосницу модела развијених у наставку дисертације.

2.4.1. Методолошки оквир обраде текста

Обрада природног језика представља прецизно дефинисан процес у ком се неструктурирани текст трансформише у нумеричке (векторске) репрезентације погодне за алгоритме машинског учења. Савремена литература истиче да се већина *NLP* система, без обзира на домен или

задатак, може описати кроз тростепени методолошки оквир који обухвата [49] [50]:

1. претпроцесирање текста,
2. репрезентацију текстуалних података,
3. моделовање.

У првој фази, систем врши нормализацију текста и сегментацију на минималне језичке јединице, чиме се подаци припремају за даље кораке. У другој фази примењују се приступи статистичке или дистрибуиране (неуронске) репрезентације текста, који редукују лингвистички комплексне садржаје на векторе фиксне димензије. Трећа фаза обухвата примену модела машинског учења, што доводи до доделе коначних класа. Слика 4 илуструје целокупан процес, од сировог текстуалног уноса до финалне предикције.

Слика 4 Методолошки оквир обраде текста у задацима класификације

2.4.2. Претпроцесирање текста

Претпроцесирање текста представља темељни корак сваког система за обраду природног језика и има за циљ да сирови текст преведе у структурираније и уједначеније облике погодне за даље алгоритамско моделовање. Текст добијен из реалних извора обично садржи различите облике шума, као што су интерпункцијски артефакти, правописне грешке, нестандардне форме или вишак празнина, па је због тога неопходно применити поступке чишћења и нормализације. Први корак у овом процесу обично обухвата уклањање непотребних симбола, исправљање неправилности у форматирању и, у зависности од задатка, нормализацију великих и малих слова ради смањења варијабилности у текстуалним репрезентацијама [49] [50].

Након почетног чишћења спроводи се токенизација, процес разбијања текста на основне јединице - токене, при чему карактери, делови речи, речи, изрази, или реченице постају појединачне аналитичке јединице. Она представља основу за већину каснијих алгоритама, јер се анализа обично одвија над токенима, а не над необрађеним низовима карактера [49].

У наставку процеса, често се примењује уклањање неважних речи (енгл. *stopwords removal*), односно фреквентних функционалних речи које углавном не носе информативан садржај релевантан за већину задатака. Спискови неважних речи могу се преузети из стандардних ресурса или конструисати доменски, при чему је важно водити рачуна да елиминација таквих токена не наруши интегритет значења у специфичним анализама. Поред уклањања неважних речи, уобичајена је и елиминација врло ретких токена, насумичних низова карактера или техничких обележја, јер они најчешће представљају шум без семантичке вредности [49] [50].

Финални корак претпроцесирања обично подразумева морфолошку нормализацију. Кореновање (енгл. *stemming*), које се базира на уклањању префикса и суфикса према једноставним правилима, омогућава брзо и ефикасно редуковање речи на основу, али ствара облике који нису нужно лингвистички исправни. Лематизација (енгл. *lemmatization*), насупрот томе, користи морфолошке речнике и граматичка правила како би добила граматички исправан основни облик речи (енгл. *lemma*), што омогућава већу прецизност у разумевању значења. Избор између кореновања и лематизације зависи од језика, природе корпуса и захтева конкретног модела [49]. У овом истраживању, за обраду текста на српском језику примењено је кореновање, због доступности алата за овај поступак.

Целокупан процес претпроцесирања мора бити пажљиво уравнотежен између смањења шума и очувања информација кључних за даљу анализу. Ова фаза представља основну фазу трансформације текста пре његове трансформације у векторску репрезентацију, а њен квалитет директно утиче на стабилност и ефикасност каснијих модела машинског учења [49] [50].

2.4.3. Статистичка репрезентација текста

Статистичке репрезентације текста заснивају се на идеји да се документ опише као вектор нумеричких карактеристика изведених из појављивања речи или термина у тексту. Овај приступ занемарује редослед речи и граматичку структуру, али обезбеђује једноставан, ефикасан и у пракси веома употребљив модел за различите задатке природног језика. Два најзначајнија модела ове групе су модел скупа речи (енгл. *Bag-of-Words - BoW*) и *TF-IDF* (енгл. *Term Frequency - Inverse Document Frequency*).

BoW посматра документ као неуређени скуп речи, а његова нумеричка репрезентација добија се формирањем речника корпуса и бележењем фреквенције којом се свака реч појављује у документу. Редослед речи се при томе занемарује, што доводи до високодимензионалних, разређених вектора [50].

Предности *BoW* модела огледају се у једноставности, транспарентности и ефикасности, јер омогућава брзу израду репрезентација и лако комбиновање са класичним класификаторима. Са друге стране, недостатак модела је занемаривање контекста и значењских односа међу речима, као и осетљивост на величину речника и учесталост неинформативних речи. Упркос овим ограничењима, *BoW* се успешно примењује у задацима класификације текста, нарочито када је реч о краћим текстовима или када редослед речи није од пресудног значаја за остваривање циља класификације [50].

TF-IDF представља надоградњу *BoW* модела и уводи концепт пондерисања, тако што речи које су специфичне за документ добијају већу тежину, док речи које су честе у целом корпусу добијају мању вредност [51].

Поступак *TF-IDF* трансформације обухвата два корака. Прво, рачуна се учесталост речи у документу (енгл. *Term Frequency - TF*), чиме се добија локална мера важности. Формално, мера учесталости термина t у документу d:

$$\mathrm{tf}(t,d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}}, \tag{3}$$

где је $f_{t,d}$ број појављивања термина t у документу d, а именилац представља укупан број термина у документу. Потом се израчунава инверзна фреквенција појављивања речи у осталим документима (енгл. *Inverse Document Frequency - IDF*), која умањује значај речи које се јављају у великом броју докумената:

$$\mathrm{idf}(t) = log\,\frac{N}{\mathrm{df}(t)}, \tag{4}$$

где је N укупан број докумената у корпусу, а $\mathrm{df}(t)$ број докумената у којима се термин t јавља бар једном. Множењем ова два фактора добија се коначна тежина коју *TF-IDF* додељује свакој речи у документу [51]:

$$\mathrm{tfidf}(t,d) = \mathrm{tf}(t,d) \cdot \mathrm{idf}(t). \tag{5}$$

Овај модел је посебно користан за препознавање термина специфичних за тему текста, чиме се омогућава бољи начин разликовања докумената у задацима претраге и класификације [51]. Иако не узима у обзир контекст или редослед речи, *TF-IDF* је и даље широко примењиван због своје једноставности, интерпретабилности и конкурентних перформанси у бројним практичним *NLP* задацима [50]. Управо због ових својстава, *TF-IDF* је у овом истраживању примењен као једна од векторских репрезентација текста за обуку статистичких модела, како би се успоставила референтна тачка за поређење са напреднијим контекстуалним репрезентацијама.

Иако статистичке репрезентације као што су *BoW* и *TF-IDF* обезбеђују ефикасан начин трансформисања текста у нумерички облик, њихова употребна вредност директно зависи од квалитета и квантитета доступних података. У практичним примерима често се јавља изазов рада са ограниченим скуповима података или израженом неравномерношћу класа, где мањи број примера у одређеним категоријама отежава процес учења и генерализације. Директна примена алгоритама на такве неуравнотежене дистрибуције неминовно доводи до пристрасности модела и фаворизовања већинских класа.

Превазилажење ових структурних недостатака у подацима захтева примену метода које вештачким путем проширују и балансирају тренинг скуп пре самог процеса моделовања. Технике аугментације података

омогућавају генерисање нових и разноликих инстанци на основу постојећих узорака, чиме се повећава робусност система и смањује ризик од преприлагођавања. Ове методе представљају неопходан међукорак који осигурава да одабране репрезентације пруже оптималне информације алгоритмима машинског учења.

2.5. Аугментација података у обради природног језика

Аугментација података (енгл. *data augmentation*) представља скуп техника за вештачко генерисање нових инстанци за обуку на основу постојећих података, са циљем повећања обима и разноликости скупа, као и смањења преприлагођавања модела. Док је у рачунарској визији аугментација релативно интуитивна, у обради природног језика ово представља значајан изазов. Због дискретне природе језика, чак и мале измене, попут замене једне речи, могу променити граматичку структуру или потпуно инвертовати значење реченице (на пример: додавање негације). У зависности од тога у којој фази обраде се примењују, технике аугментације у *NLP*-у се могу поделити на две широке категорије: аугментацију у простору атрибута (енгл. *feature space augmentation*) и аугментацију у простору података (енгл. *data space augmentation*) [52].

Аугментација у простору атрибута се примењује након што је текст већ трансформисан у нумеричке векторе. Циљ је генерисање нових синтетичких вектора или селективно уклањање постојећих, како би се смањила неравнотежа између класа. Технике у овој групи се могу поделити у три основна приступа: преузорковање мањинске класе (енгл. *oversampling*), подузорковање већинске класе (енгл. *undersampling*), и хибридне методе које комбинују оба приступа.

Најпознатији алгоритам за преузорковање је *SMOTE* (енгл. *Synthetic Minority Over-sampling Technique*). Уместо простог дуплирања постојећих примера, које води ка преприлагођавању јер модел памти идентичне тачке, *SMOTE* креира нове инстанце линеарном интерполацијом између постојећих примера мањинске класе и њихових најближих суседа. За сваки вектор мањинске класе x_i алгоритам бира једног од његових најближих суседа и генерише нову тачку x_{new} према формули:

$$x_{new} = x_i + \lambda \times (x_{zi} - x_i), \qquad (6)$$

где је λ насумични број у интервалу $[0, 1]$. Овај поступак ефективно попуњава простор одлучивања између мањинских примера [53].

SMOTE не уноси нове информације у строгом статистичком смислу, јер све синтетичке тачке леже унутар конвексног омотача постојећих мањинских примера. Корист алгоритма произилази из његовог регуларизационог ефекта, јер спречава модел да научи превише уске границе одлучивања око појединачних примера, и из имплицитне претпоставке локалне глаткости према којој простор између два блиска мањинска примера такође припада истој класи [53]. Ова претпоставка представља и главно ограничење *SMOTE*-а у високодимензионалним просторима какви настају статистичком репрезентацијом текста. Због велике димензионалности, појам најближих суседа губи на поузданости, а интерполирани вектори могу представљати семантички неконзистентне тачке. Из тог разлога се *SMOTE* неретко примењује у комбинацији са техникама чишћења података [52] [54].

Алтернативни приступ проблему неравнотеже представља подузорковање, које уместо генерисања нових примера уклања одређене инстанце већинске класе. Једна од најчешће коришћених техника из ове групе је правило чишћења околине (енгл. *Neighborhood Cleaning Rule - NCR*), које за сваку инстанцу мањинске класе проналази њене најближе суседе и уклања оне међу њима који припадају већинској класи, док саму мањинску инстанцу задржава нетакнуту. Додатно, у другом кораку алгоритма, из већинске класе се уклањају и преостале инстанце чија се класа разликује од већине њихових суседа. Овим поступком се побољшава сепарабилност класа у граничном региону, без увођења синтетичких података и без значајног губитка информација из већинске класе [55].

Преузорковање и подузорковање се могу комбиновати у хибридне методе, од којих је најпознатија *SMOTEENN*. Овај поступак прво примењује *SMOTE* за генерисање нових примера, а затим користи правило измењених најближих суседа (енгл. *Edited Nearest Neighbors - ENN*) за уклањање шума. *ENN* брише инстанце чија се класа разликује од класе већине њихових суседа. Резултат су чистије групе података и јасније границе одлучивања [54]. У овом истраживању, ове три технике

аугментације у простору атрибута примењене су за решавање проблема небалансираних класа у обуци статистичких модела.

За разлику од манипулације нумеричким векторима, аугментација у простору података подразумева трансформације директно над сировим текстом. Циљ је креирање нових, лингвистички валидних реченица које задржавају оригинално значење и ознаку класе, али уносе лексичку и синтаксичку варијабилност.

Традиционалне методе замене речи синонимима често нарушавају контекст реченице јер не узимају у обзир полисемију. Савремени приступ превазилази овај проблем коришћењем контекстуалне аугментације, која се ослања на претриниране језичке моделе, о којима ће детаљније бити речи у поглављу 2.11. Ови модели поседују способност разумевања контекста целе реченице, што се користи за две основне врсте трансформације [52]:

1. Замена (енгл. *substitution*): одређене речи у реченици се маскирају, а модел предвиђа највероватнију замену која се семантички и граматички уклапа у окружење (нпр. замена придева сродним придевом који одговара тону реченице).
2. Уметање (енгл. *insertion*): модел предвиђа нову реч која се може природно уметнути у структуру реченице, чиме се повећава варијабилност дужине и структуре узорака.

Употреба дубоких језичких модела за ове трансформације осигурава знатно већу семантичку конзистентност у поређењу са једноставним техникама заснованим на речницима. Примена наменских модела за специфичне језике осигурава да су генерисане реченице морфолошки исправне, чиме се значајно повећава робусност класификатора на варијације у природном језику [52]. У овом истраживању, контекстуална аугментација путем замене и уметања речи примењена је приликом финог подешавања претрениранх *Transformer* модела.

Велики језички модели, о којима ће детаљније бити речи у поглављу 2.12, се у овом раду примењују као један од приступа моделовања, али се не користе за аугментацију података. Иако су они способни да генеришу нове текстуалне примере, њихови излази захтевају додатну контролу

квалитета, јер могу одступити од оригиналне лабеле класе, увести фактичке нетачности или произвести реченице које значајно мењају стил и дужину у односу на постојеће примере. Модели који се користе за контекстуалну аугментацију, насупрот томе, задржавају структуру полазне реченице и модификују је искључиво на нивоу појединачних токена, чиме се обезбеђује лексичка варијабилност уз минималан ризик од промене лабеле класе.

2.6. Машинско учење за класификацију текста

Након припреме података кроз претпроцесирање, репрезентацију и аугментацију, следећи корак у методолошком оквиру обраде текста је примена модела машинског учења. Машинско учење (енгл. *Machine Learning - ML*) представља област вештачке интелигенције која се бави развојем модела који на основу података могу да уочавају обрасце и доносе одлуке. Суштина машинског учења огледа се у приближном моделовању функције која повезује улазне податке са излазима, при чему је циљ да се постигне добра генерализација, односно да модел оствари високе перформансе и на подацима који нису били обухваћени током обуке [56].

У литератури се уобичајено разликују три главна приступа машинском учењу [57]:

- Надгледано учење (енгл. *supervised learning*) подразумева да су улазни подаци праћени познатим излазима, а модел се обучава да научи мапирање између улаза и излаза.
- Ненадгледано учење (енгл. *unsupervised learning*) има за циљ да открије скривене структуре у неозначеним подацима (нпр. кластере).
- Учење поткрепљивањем (енгл. *reinforcement learning*) заснива се на интеракцији агента са окружењем и оптимизацији награде.

За потребе аутоматске анализе студентских повратних информација, где су категорије унапред дефинисане, користи се парадигма надгледаног учења.

2.6.1. Основе класификације текста

Класификација текста се формално дефинише као задатак додељивања једне или више категорија из унапред дефинисаног скупа $C = \{c_1, c_2, \ldots, c_k\}$ датом документу d. Циљ алгоритма је да на основу скупа обележених примера за обуку D апроксимира функцију мапирања $\gamma : D \rightarrow C$ која ће са највећом вероватноћом доделити исправну класу новом документу [50]. У случају бинарне класификације, скуп C садржи две категорије, док у вишекласној класификацији скуп обухвата три или више категорија. У овом истраживању све класификације имају вишекласну природу.

Овај процес обухвата четири кључне компоненте: репрезентацију података која трансформише текст у нумеричке векторе, модел који мапира нумеричке векторе у класе, функцију грешке (енгл. *Loss function*) која квантификује одступање предвиђања од стварних вредности, и алгоритам оптимизације који прилагођава параметре модела ради минимизације грешке [50].

Критични изазов у процесу обуке представља балансирање између способности модела да научи обрасце из тренинг скупа и његове способности генерализације. Уколико је модел превише једноставан да би ухватио структуру података, долази до потприлагођавања (енгл. *underfitting*). Са друге стране, уколико модел постане превише комплексан и почне да учи детаље и шум специфичан за тренинг скуп, настаје проблем преприлагођавања (енгл. *overfitting*), што резултује лошим перформансама модела на новим подацима. Оптималан модел постиже равнотежу између тачности на обучавајућем и тестирајућем скупу података [56].

2.6.2. Евалуација модела

Објективна процена успешности класификационих модела заснива се на двема комплементарним компонентама: квантитативним метрикама које мере квалитет предикција и стратегијама валидације које обезбеђују поузданост ове процене. Овај одељак најпре излаже основне метрике евалуације, а затим стратегије валидације које омогућавају поуздану процену генерализације модела.

Основу за израчунавање метрика чини матрица конфузије (енгл. *Confusion Matrix*), коју илуструје табела 2. Матрица конфузије приказује однос између стварних и предвиђених вредности кроз четири исхода: стварно позитивно (енгл. *True Positive - TP*), стварно негативно (енгл. *True Negative - TN*), лажно позитивно (енгл. *False Positive - FP*) и лажно негативно (енгл. *False Negative - FN*).

Табела 2 Матрица конфузије

Стварно / Предвиђено	Позитивно	Негативно
Позитивно	стварно позитивно (TP)	лажно негативно (FN)
Негативно	лажно позитивно (FP)	стварно негативно (TN)

Иако се често користи тачност (енгл. *Accuracy*), ова мера може бити заваравајућа код неуравнотежених скупова података. Због тога се у *NLP*-у примарно користе следеће метрике [50]:

- Прецизност (енгл. *Precision*) мери удео релевантних примера међу онима које је модел означио као позитивне:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

- Одзив (енгл. *Recall*) мери способност модела да идентификује све релевантне примере у скупу:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

- *F1*-мера (енгл. *F1-score*) представља хармонијску средину прецизности и одзива, пружајући јединствену меру квалитета:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{9}$$

У случају вишекласне класификације, агрегација перформанси се најчешће врши на два начина. Макро усредњавање (енгл. *macro-average*) подразумева израчунавање метрике за сваку класу независно, након чега се налази њихов просек, чиме се све класе третирају равноправно, што је од кључног значаја за валидацију модела на мањинским класама. Насупрот томе, микро усредњавање (енгл. *micro-average*) функционише

тако што агрегира укупне вредности *TP*, *FP* и *FN* исхода свих класа пре самог израчунавања метрике, због чега је овај приступ доминантно одређен перформансама на већинским класама [58].

Применљивост наведених метрика директно зависи од стратегије валидације која одређује начин поделе података на тренинг и тест скупове и тиме поузданост саме процене перформанси. Класична подела у фиксном односу пружа само једну тачкасту процену, која може бити осетљива на специфичну поделу. Унакрсна валидација (енгл. *cross-validation*) превазилази ово ограничење поделом скупа на k једнаких делова и понављањем процеса обуке и тестирања, при чему се сваки део наизменично користи као тест скуп. Иако пружа стабилнију процену, овај приступ не омогућава квантификацију варијабилности перформанси кроз стандардну девијацију или интервале поузданости [56].

Bootstrap валидација пружа алтернативно решење засновано на статистичком поновном узорковању, које омогућава експлицитну квантификацију варијабилности перформанси. У *out-of-sample* варијанти, у свакој од више независних итерација врши се стратификована подела података у фиксном односу, чиме се очувава дистрибуција класа, а финалне перформансе се извештавају као средња вредност и стандардна девијација кроз све итерације. Овакав приступ омогућава поуздану процену генерализације и квантификацију несигурности у мерењу перформанси модела [59]. У овом истраживању примењена је управо *out-of-sample bootstrap* валидација као примарна стратегија за процену генерализације модела, уз употребу макро усредњене *F1*-мере као примарне метрике за оптимизацију.

2.6.3. Методе оптимизације хиперпараметара

Перформансе модела машинског учења зависе од хиперпараметара, односно конфигурационих вредности које се постављају пре почетка процеса учења. Процес проналажења оптималне комбинације ових вредности назива се оптимизација хиперпараметара и најчешће се спроводи уз помоћ унакрсне валидације (енгл. *cross-validation*) [60].

У литератури се издвајају три доминантне стратегије претраге простора хиперпараметара. Најосновнији приступ, мрежна претрага (енгл. *Grid*

Search), подразумева систематско испитивање свих могућих комбинација унапред дефинисаних вредности параметара. Иако овај метод гарантује проналазак најбољег решења унутар задатог опсега, он је рачунарски захтеван и често непрактичан за високодимензионалне проблеме. С друге стране, мрежна претрага представља логичан избор у сценаријима у којима је простор хиперпараметара унапред сужен на мали број вредности заснованих на препорукама из литературе, што је често случај код финог подешавања претренираних модела. Као ефикаснија алтернатива често се примењује насумична претрага (енгл. *Random Search*), која бира комбинације из задатих дистрибуција вероватноће. Емпиријска истраживања показују да овај приступ омогућава темељније истраживање простора стања уз знатно мањи утрошак ресурса у поређењу са мрежном претрагом [60].

За разлику од претходна два метода који не користе информације из прошлих итерација, Бајесова оптимизација (енгл. *Bayesian Optimization*) представља софистицираније приступ који гради пробабилистички модел функције циља, најчешће користећи Гаусове процесе. На основу претходних евалуација, овај метод интелигентно бира наредну комбинацију хиперпараметара, балансирајући између истраживања непознатих делова простора и искоришћавања региона за које је већ утврђено да дају добре резултате [61]. У овом истраживању примењене су две стратегије претраге простора хиперпараметара: Бајесова оптимизација за развој статистичких модела и мрежна претрага за фино подешавање претренираних *Transformer* модела, у складу са препорукама из литературе.

2.7. Статистички модели машинског учења

Као што је описано у претходним поглављима, избор одговарајућег алгоритма за класификацију директно зависи од природе проблема и карактеристика задатка. Иако су неуронске мреже донеле револуцију у обради природног језика, статистички модели машинског учења и даље представљају незаменљиву основу у задацима класификације текста. Њихова важност огледа се у томе што служе као референтни модели који често нуде оптималан баланс између рачунарске ефикасности и тачности, посебно на скуповима података умерене величине. Осим тога, за разлику

од комплексних архитектура дубоког учења, ови модели омогућавају већи степен интерпретабилности резултата [62].

Ови модели се заснивају на чврстом математичком оквиру који омогућава прецизну процену параметара и контролу сложености кроз принципе оптимизације [56]. У контексту ове дисертације, статистички модели се користе над векторским репрезентацијама текста (*TF-IDF*) како би се успоставила референтна тачка за евалуацију напреднијих приступа. Међу најзначајнијим алгоритмима у овој групи издваја се метод потпорних вектора (енгл. *Support Vector Machine - SVM*).

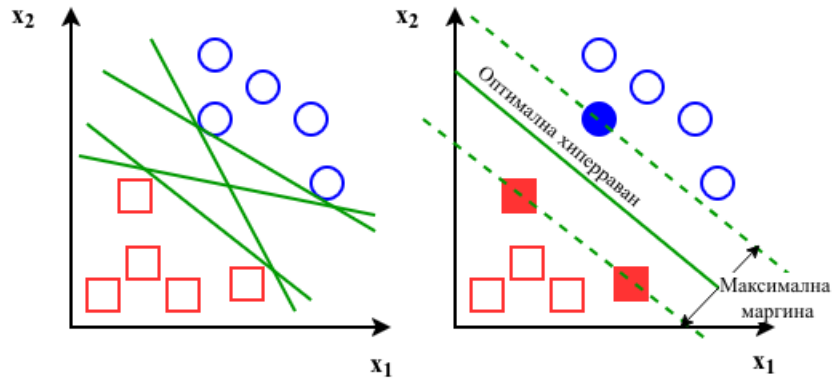
Метод потпорних вектора представља један од најробуснијих алгоритама надгледаног учења за класификацију високодимензионалних података, какав је текст. Основна идеја *SVM*-а је проналажење оптималне хиперравни која раздваја податке различитих класа са максималном могућом маргином. Теоријски гледано, већа маргина имплицира мању грешку генерализације [63].

Формално, за дати скуп парова $(\boldsymbol{x_i}, y_i)$, где је x_i вектор атрибута, а $y_i \in \{-1, 1\}$ ознака класе, хипер раван је дефинисана једначином:

$$w^T x + b = 0, \tag{10}$$

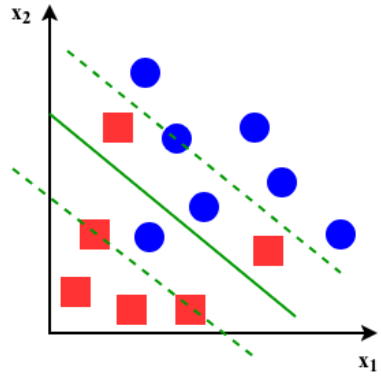
где w представља вектор тежина (нормалу на раван), а b помак. Оптимална хипер раван се добија решавањем оптимизационог проблема који максимизује маргину $\frac{2}{\|\boldsymbol{w}\|}$, уз услов да су сви примери исправно класификовани. Примери који се налазе на самим границама маргине називају се потпорни вектори и једино они утичу на дефинисање модела. Слика 5 илуструје могуће хиперравни у дводимензионалном простору и избор оптималне хиперравни за раздвајање две класе, при чему се потпорни вектори налазе на испрекиданим линијама.

У реалним применама, подаци често нису линеарно раздвојиви. Због тога се користи метод потпорних вектора са меким појасом (енгл. *soft margin*), који дозвољава извесну толеранцију грешке у циљу боље генерализације [57]. Слика 6 илуструје овај метод у дводимензионалном простору на проблему са две класе.



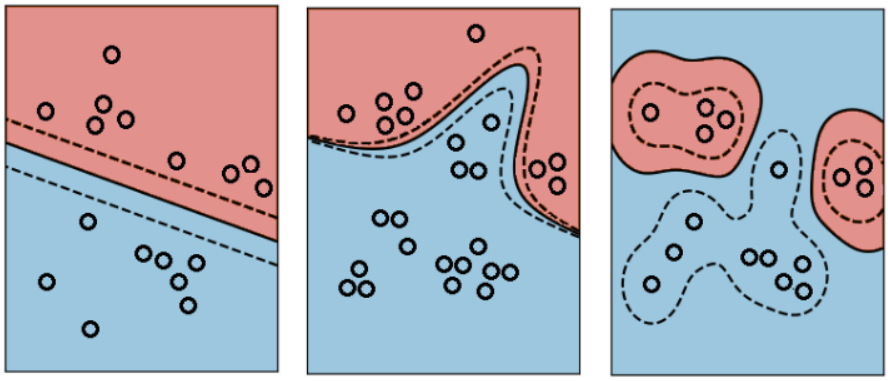
Слика 5 Пример *SVM* модела у дводимензионалном простору за две класе [57]

У овај модел уводе се променљиве толеранције $\xi_i \geq 0$, које мере степен прекорачења маргине за сваки појединачни пример, и параметар контроле грешке $C > 0$, који одређује однос између ширине маргине и дозвољеног нивоа грешке. Веће вредности C теже строжем раздвајању (ризик од преприлагођавања), док мање вредности дозвољавају већу толеранцију грешке ради боље генерализације [56].



Слика 6 Пример *SVM* модела са меким појасом

За моделовање нелинеарних граница користи се трик језгра (енгл. *kernel trick*), који користи функције језгра (енгл. *kernel functions*) које имплицитно мапирају податке у виши димензионални простор у коме су класе линеарно раздвојиве. Основни принцип се заснива на израчунавању скаларног производа у трансформисаном простору. Најчешће коришћене функције језгра су линеарна, полиномна и радијално-базна (енгл. *Radial Basis Function - RBF*) [56] [63]. Слика 7 илуструје ове функције језгра.



Слика 7 Линеарна, полиномна и радијално-базна функција језгра

У задацима класификације текста, *SVM* се најчешће примењује са линеарном функцијом језгра, јер су текстуални вектори већ довољно високодимензионални да омогућавају линеарну сепарацију [50].

Главне предности *SVM*-а огледају се у његовој чврстој теоријској утемељености и способности добре генерализације, чак и у високодимензионалним просторима карактеристичним за текстуалне податке. Чињеница да коначну границу одлучивања дефинише само мали подскуп примера (потпорни вектори) чини модел меморијским ефикасним у фази предикције. Међутим, значајна ограничења представљају висока рачунска захтевност током обуке, која квадратно расте са бројем примера, као и осетљивост на избор хиперпараметара и шум у подацима [56] [63]. И поред наведених ограничења, *SVM* је одабран као један од референтних модела у овом истраживању због своје доказане ефикасности у задацима класификације текста, дајући поуздану основу за поређење са напреднијим архитектурама.

Поред *SVM*-а као појединачног класификатора, у овом истраживању примењени су и ансамбл приступи који комбинују више базних модела ради остваривања веће тачности и стабилности предвиђања, што је представљено у наредном поглављу.

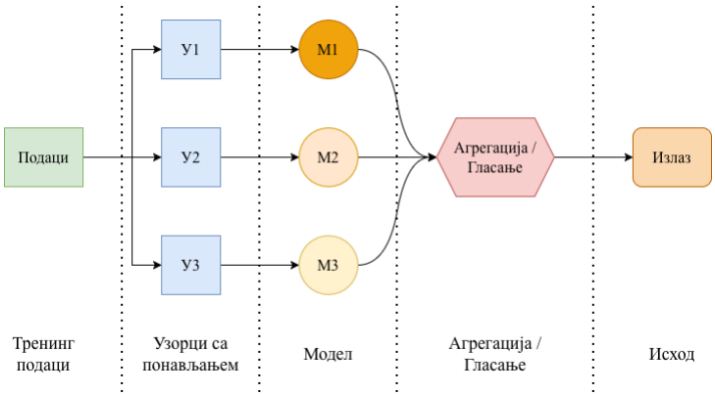
2.8. Ансамбли

Ансамбли (енгл. *ensemble*) представљају класу модела машинског учења у којима се више појединачних модела комбинује у јединствен модел са

циљем постизања боље тачности и стабилности предвиђања. Основна мотивација за овај приступ произилази из анализе грешака модела, која се у теорији машинског учења често описује кроз компоненте пристрасности и варијансе. Пристрасност (енгл. *bias*) означава систематско одступање предвиђања модела од стварних вредности услед претерано једноставне хипотезе, док варијанса (енгл. *variance*) представља осетљивост модела на варијације у тренинг подацима, односно степен до ког се модел преприлагођава шуму или случајним одступањима. Ансамбли имају за циљ да уравнотеже ове компоненте, смањујући ризик од нестабилности (високе варијансе) или недовољног моделовања структуре података (високе пристрасности) [56].

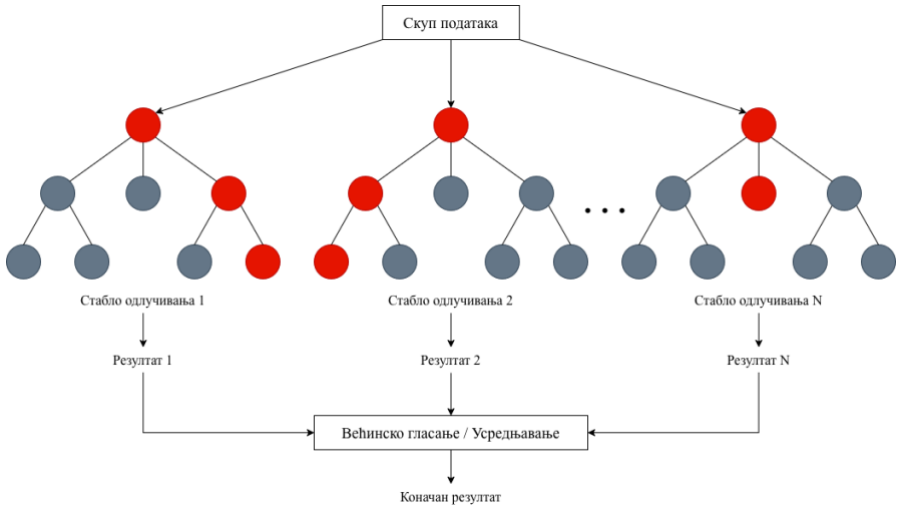
Најпознатији приступи конструкције ансамбла обухватају три концептуално различите стратегије: просту агрегацију (енгл. *bootstrap aggregation*, *bagging*), појачавање (енгл. *boosting*) и гласање (енгл. *voting*). Прве две стратегије подразумевају дизајн посебног процеса обуке базних модела, док трећа стратегија комбинује предвиђања више независно обучених модела.

Проста агрегација се заснива на идеји да се више модела обучава на различитим варијацијама истог скупа података. Подскупови примера бирају се случајним избором са понављањем, што доводи до диверзитета у базним моделима и смањења корелисаности њихових грешака. Током предвиђања, појединачни модели гласају или дају просечну вредност, чиме се добија стабилнија коначна предикција [64]. Слика 8 илуструје основну структуру овог приступа.



Слика 8 Основна структура просте агрегације

Метод случајних шума (енгл. *Random Forests*) представља проширење просте агрегације увођењем додатне случајности током изградње стабла. Поред избора различитих подскупова примера, у сваком чвору стабла разматра се само насумични подскуп карактеристика. Овај поступак додатно смањује корелисаност између појединачних стабала и повећава диверзитет ансамбла [65]. Слика 9 илуструје рад метода случајних шума, где се коначни резултат добија комбиновањем одлука свих стабала путем већинског гласања или усредњавања.



Слика 9 Илустрација рада метода случајних шума

Иако су појединачна стабла одлучивања склона високој варијанси, односно преприлагођавању тренинг подацима, њиховим усредњавањем у оквиру шуме та варијанса се драстично смањује без значајног повећања пристрасности. Ова особина чини случајне шуме изузетно робусним и отпорним на шум у подацима [65]. Метод случајних шума примењен је у овом истраживању као један од референтних статистичких модела.

Појачавање представља класу ансамбл техника у којима се више слабих модела гради секвенцијално, при чему сваки наредни модел настоји да исправи грешке претходних. За разлику од просте агрегације, која углавном умањује варијансу комбиновањем независних модела, појачавање поступно унапређује модел усмеравањем процеса учења ка тешким примерима (онима који су претходно погрешно класификовани). На овај начин се истовремено смањују и пристрасност и варијанса, што

доводи до изузетно ниске грешке генерализације [56] [57]. Историјски гледано, развој ове области започео је алгоритмом *AdaBoost* [66], који користи механизам повећања тежина за погрешно класификоване примере.

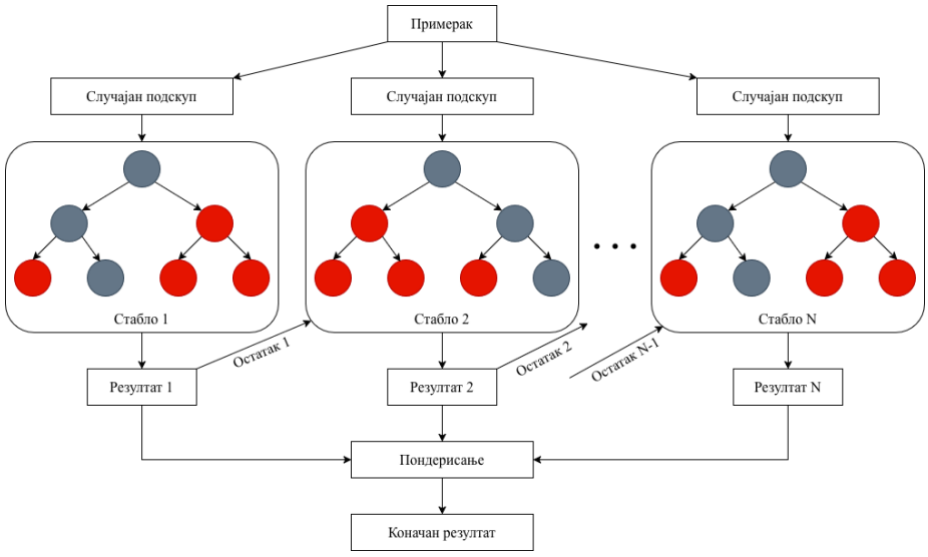
Модеран приступ овој идеји представља градијентно појачавање (енгл. *Gradient boosting*), које процес учења посматра као функционалну оптимизацију. Уместо једноставног мењања тежина примера, сваки модел у низу се обучава да апроксимира негативан градијент функције грешке у односу на тренутну предикцију ансамбла [67].

Као тренутно најнапреднија имплементација овог концепта издваја се екстремно градијентно појачавање (енгл. *eXtreme Gradient Boosting - XGBoost*). Овај алгоритам се заснива на изградњи плитких стабала одлучивања, али уз кључно унапређење у виду регуларизације. Функција циља коју *XGBoost* оптимизује експлицитно укључује члан за контролу сложености:

$$\mathrm{Obj}(\Theta) = L(\Theta) + \Omega(\Theta), \tag{11}$$

где L представља функцију грешке (одступање предвиђања од стварности), док Ω представља регуларизациони термин који кажњава прекомерну сложеност стабла (нпр. велики број листова или велике вредности тежина). Овај механизам спречава преприлагођавање, што је од критичног значаја при раду са разређеним подацима као што су *TF-IDF* вектори [68]. Слика 10 илуструје секвенцијални процес где се финална одлука добија сумирањем доприноса свих стабала.

Технике појачавања, а нарочито *XGBoost*, показале су се као веома ефикасне у задацима класификације текста, јер ефикасно моделују сложене, нелинеарне односе у високодимензионалним просторима. Механизми контроле сложености и регуларизације доприносе стабилности и отпорности на преприлагођавање, што их чини погодним избором у анализи текстуалних корпуса [68]. У овом истраживању, *XGBoost* је примењен као један од статистичких модела за класификацију текста, управо због наведене робусности у раду са *TF-IDF* репрезентацијама.



Слика 10 Основна структура *XGBoost* модела

За разлику од просте агрегације и појачавања, који подразумевају изградњу читавог ансамбла кроз посебно дизајниран процес обуке, гласање се ослања на комбиновање предвиђања више независно обучених модела. Ова стратегија посебно је корисна када се располаже моделима различитих архитектура чија су предвиђања делимично некорелисана. Концептуална основа лежи у претпоставци да модели различитих структура праве различите типове грешака, па њихова комбинација може довести до тачнијих и стабилнијих коначних предвиђања него било који појединачни модел [69].

У основном облику, познатом као већинско гласање (енгл. *hard majority voting*), коначна класа се одређује већинским гласањем међу предвиђањима свих модела у ансамблу. У случајевима изједначеног броја гласова, потребно је дефинисати стратегију за разрешење неодлучности, која се најчешће своди на коришћење предвиђања најуспешнијег појединачног модела. Унапређени облик представља тежинско гласање (енгл. *weighted voting*), у којем се сваком моделу додељује коефицијент значаја сразмеран његовим перформансама. Коначна одлука се тада доноси као тежински збир предвиђања, чиме се омогућава да модели са доказаном већом поузданошћу имају пресудан утицај на крајњи резултат [70]. Обе стратегије гласања примењене су у овом истраживању.

За разлику од просте агрегације и појачавања, гласање не намеће ограничења на тип базних модела. Базни модели могу бити статистички класификатори, неуронске мреже или велики језички модели, што гласање чини флексибилном стратегијом за интеграцију хетерогених приступа моделовању. Управо ову флексибилност овог приступа искоришћава последњи приступ моделовања у овом истраживању, у којем се комбинују статистички класификатори, фино подешени *Transformer* модели и велики језички модели.

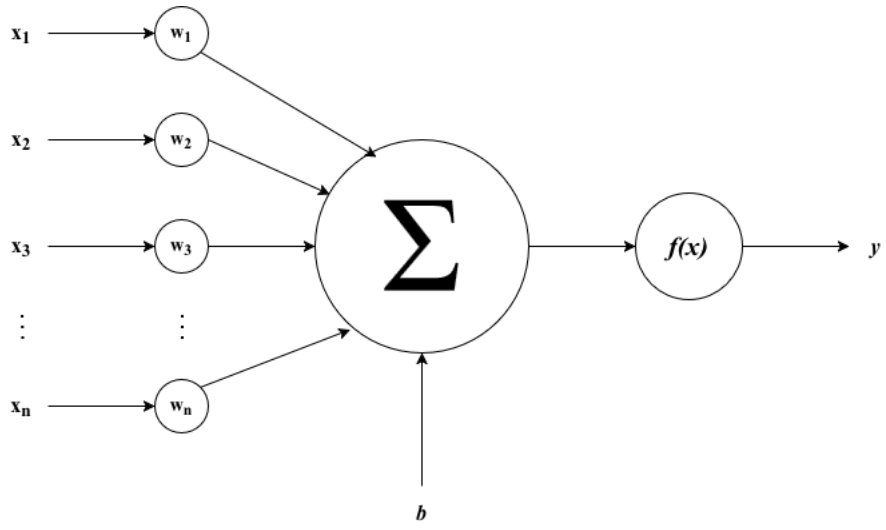
И поред доказане ефикасности статистичких модела и ансамбала у многим применама, перформансе појединачних статистичких класификатора суштински зависе од квалитета улазних репрезентација и ручно дизајнираних атрибута. Они често не успевају да у потпуности ухвате дубоке семантичке зависности и сложену хијерархијску структуру природног језика без сложених трансформација улазних података. Превазилажење ових ограничења захтева другачију парадигму учења, која је способна да аутоматски издваја битне обрасце директно из сирових података и моделује високо нелинеарне односе. Такав приступ нуде вештачке неуронске мреже, које ће бити предмет наредног поглавља.

2.9. Вештачке неуронске мреже

Вештачке неуронске мреже представљају математички оквир за машинско учење инспирисан биолошком структуром неурона у мозгу сисара. Њихова основна снага лежи у способности универзалне апроксимације, односно могућности да апроксимирају било коју нелинеарну непрекидну функцију са произвољном тачношћу, под условом да мрежа поседује довољан број неурона у скривеним слојевима [71]. У контексту обраде природног језика, вештачке неуронске мреже су омогућиле прелазак са статистичких модела репрезентације текста на моделе који аутоматски уче репрезентације језичких јединица. У овом истраживању, неуронске мреже представљају теоријску основу за разумевање претренираних *Transformer* модела (поглавље 2.11) и великих језичких модела (поглавље 2.12), који чине основу два методолошка приступа примењена у дисертацији.

Како би се описала математичка структура неуронске мреже, потребно је најпре дефинисати њен основни градивни елемент. Основни градивни елемент мреже је вештачки неурон, чију структуру илуструје слика 11. Он прима улазни вектор x, врши линеарну трансформацију користећи вектор тежина w и параметар помака (енгл. *bias*) b, а затим на резултат примењује нелинеарну активациону функцију f. Излаз једног неурона y дефинише се на следећи начин [56]:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) = f(w^T x + b). \tag{12}$$



Слика 11 Структура вештачког неурона

Појединачни неурони се организују у слојеве, при чему излаз једног слоја служи као улаз наредног. Мреже са више скривених слојева између улаза и излаза називају се дубоким неуронским мрежама и представљају основу савремених приступа машинском учењу [56].

Активациона функција f је кључна јер уводи нелинеарност, омогућавајући мрежи да учи комплексне границе одлучивања које нису линеарно сепарабилне. Уколико нелинеарна активација изостане, неуронска мрежа би се, без обзира на број слојева, математички свела на обичну линеарну регресију. У литератури и пракси се најчешће издвајају

три функције, свака са својим специфичним применама и историјским контекстом [56] [72].

Историјски гледано, сигмоидна функција (енгл. *Sigmoid*) је била најчешћи избор због своје сличности са стопом окидања биолошког неурона. Она пресликава улазне вредности у опсег $(0, 1)$, што је чини погодном за интерпретацију излаза као вероватноће:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{13}$$

Иако корисна у излазним слојевима за бинарну класификацију, у дубоким мрежама пати од проблема засићења (енгл. *saturation*), где градијенти постају занемарљиво мали за веома високе или ниске вредности улаза, што успорава или зауставља учење. Овај феномен у литератури је познат као проблем ишчезавања градијента (енгл. *vanishing gradient*) и чини обуку дубоких архитектура изузетно тешком или немогућом [72].

Као алтернатива сигмоидној функцији, хиперболични тангенс (енгл. *Hyperbolic Tangent - tanh*) пресликава улаз у опсег $(-1, 1)$. Ова функција је центрирана око нуле (енгл. *zero-centered*), што олакшава оптимизацију у скривеним слојевима мреже:

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{14}$$

И даље се често користи у скривеним слојевима неуронских мрежа за обраду секвенцијалних података [50].

У савременим дубоким архитектурама, стандард представља *ReLU* (енгл. *Rectified Linear Unit*). Она је рачунарски ефикаснија (не захтева експоненцијалне операције) и значајно ублажава проблем ишчезавања градијента јер не засићује за позитивне вредности:

$$f(x) = max(0, x). \tag{15}$$

Због ових својстава, *ReLU* је подразумевани избор у савременим дубоким архитектурама, омогућавајући тренирање знатно дубљих модела који раније нису били реализовани [72].

Учење мреже подразумева оптимизацију параметара θ (тежина и помака) минимизацијом функције грешке L (енгл. *loss function*), која мери одступање предвиђања модела од стварних вредности. За проблеме класификације у *NLP*-у стандард је употреба унакрсне ентропије (енгл. *Cross-Entropy Loss*):

$$L(\theta) = -\sum_{c=1}^{C} y_c \, log(\hat{y_c}) \,, \tag{16}$$

где је y_c стварна класа, а $\hat{y_c}$ предвиђена вероватноћа. Ажурирање тежина врши се алгоритмом пропагације уназад (енгл. *Backpropagation*), који израчунава парцијалне изводе функције грешке по сваком параметру мреже и користи их за прилагођавање тежина ради минимизације грешке [73]. У пракси се ажурирање најчешће изводи варијантама стохастичког градијентног спуста.

2.9.1. Рекурентне неуронске мреже

За разлику од стандардних потпуно повезаних неуронских мрежа које процесирају улазе независно један од другог, рекурентне неуронске мреже (енгл. *Recurrent Neural Networks - RNN*) су архитектура специјализована за обраду секвенцијалних података. Оне уводе концепт темпоралне динамике путем повратних веза, што им омогућава креирање интерног стања које функционише као меморија система. Ова карактеристика их чини природним избором за моделовање језика, где семантичко значење зависи од редоследа речи и ширег контекста [74].

У сваком временском кораку t, мрежа прима тренутни улаз x_t и ажурира своје скривено стање h_t на основу тренутног улаза и претходног скривеног стања h_{t-1} [50]:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b), \tag{17}$$

где W_h представља матрицу тежина рекурентне везе која преноси информацију кроз време, W_x матрицу тежина која повезује улаз са скривеним слојем, b вектор помака, а σ нелинеарну активациону функцију (најчешће хиперболични тангенс или *ReLU*).

Упркос теоријској способности да моделују зависности произвољне дужине, класичне *RNN* архитектуре у пракси пате од проблема нестабилних градијената, посебно ишчезавајућег градијента, чиме мрежа губи способност учења дугорочних зависности. Као одговор на овај проблем развијене су архитектуре са механизмима капија (енгл. *gates*), међу којима су најзначајније *LSTM* (енгл. *Long Short-Term Memory*) [75] и *GRU* (енгл. *Gated Recurrent Unit*) [76]. Ови модели уводе експлицитну контролу над протоком информација, чиме омогућавају ефикасније моделовање дугорочних зависности.

У области обраде природног језика, *LSTM* и *GRU* мреже представљале су доминантан приступ пре појаве *Transformer* архитектуре, омогућавајући значајан напредак у задацима као што су језичко моделовање, машинско превођење и препознавање именованих ентитета [50]. Међутим, њихова секвенцијална природа онемогућава паралелизацију тренинга, што значајно ограничава ефикасност обуке на великим корпусима. Ово ограничење, заједно са изазовима у моделовању веома дугачких зависности, касније је мотивисало развој *Transformer* архитектуре.

Поред развоја специфичних архитектура за обраду секвенци, један од најважнијих доприноса неуронских мрежа у области обраде природног језика јесте могућност учења дистрибуираних векторских репрезентација речи и реченица. Ове репрезентације, које представљају математичку основу за модерне приступе моделовању текста, биће представљене у наредном поглављу.

2.10. Дистрибуирана репрезентација текста

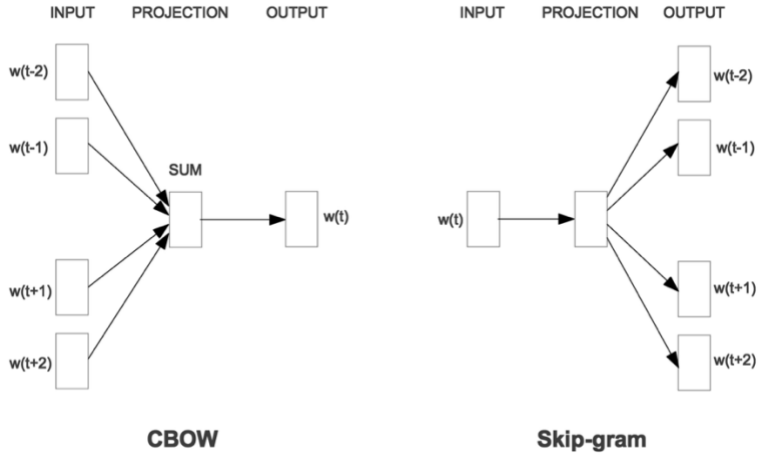
Као што је дискутовано у претходном поглављу, суштина дубоког учења лежи у формирању ефикасних и сажетих репрезентација података [72]. Овај принцип је посебно значајан у обради природног језика, где се дискретни језички симболи могу пројектовати у континуални векторски простор тако да се очувају семантичке сличности. У обради природног језика, статистички приступи репрезентације текста (одељак 2.4.3) резултују векторима изузетно високе димензионалности и разређености, будући да је дужина вектора једнака величини вокабулара. Оваква репрезентација онемогућава генерализацију, јер модел третира сваку реч

као изоловану јединицу без дељених карактеристика [50]. Превазилажење овог ограничења постиже се применом дистрибуираних векторских репрезентација (енгл. *word embeddings*), где се речи мапирају на густе векторе реалних бројева знатно ниже димензије.

Теоријско утемељење овог приступа даје дистрибуирана хипотеза према којој се значење речи дефинише контекстом у којем се она појављује [77]. Дистрибуирани вектори су дизајнирани тако да њихова геометријска близина (најчешће мерена косинусном сличношћу) одржава семантичку и синтаксну сличност [50]. Овај приступ омогућио је развој ефикасних алгоритама за учење репрезентација на великим корпусима, међу којима су најзначајнији *Word2Vec* и *GloVe*.

Word2Vec представља породицу плитких неуронских мрежа које уче векторске репрезентације речи на основу њиховог контекста појављивања. Слика 12 приказује две основне архитектуре овог модела:

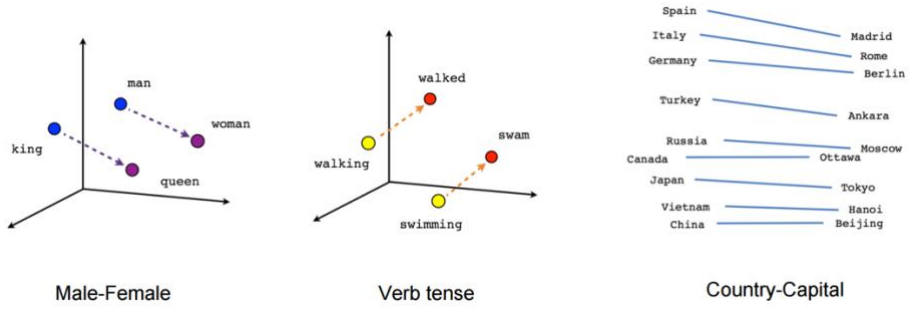
1. *CBOW* (*Continuous Bag-of-Words*) у којем модел предвиђа циљну реч на основу усредњеног контекста околних речи, при чему редослед речи у контексту не утиче на резултат.
2. *Skip-gram* где модел функционише инверзно, односно на основу циљне речи покушава да предвиди речи из њеног окружења унутар дефинисаног прозора.



Слика 12 Основне архитектуре *Word2Vec* модела

Због рачунарске захтевности великих вокабулара, у пракси се користи техника негативног узорковања (енгл. *negative sampling*), која проблем вишекласне класификације апроксимира низом бинарних класификација (разликовање праве речи од насумично одабраног шума) [78].

Слика 13 илуструје једну од најзначајнијих особина ових вектора, односно очување линеарних семантичких релација, што директно омогућава векторску аритметику.



Слика 13 Визуелизација линеарних односа у векторском простору речи

Док *Word2Vec* учи репрезентације ослањајући се на локалне прозоре контекста, алгоритам *GloVe* (енгл. *Global Vectors*) комбинује локални контекст са глобалном статистиком заједничког појављивања речи у целом корпусу. Основа модела је матрица заједничког појављивања чији елементи показују колико често се две речи јављају у заједничком контексту. Кључна теза приступа је да односи вероватноћа заједничког појављивања носе више семантичких информација него појединачне вероватноће, па се вектори речи уче тако да њихов скаларни производ апроксимира логаритам ових вероватноћа [79].

Упркос успеху, модели попут *Word2Vec* и *GloVe* поседују фундаментално ограничење: статичност репрезентације. Свакој речи се додељује фиксан вектор, независно од контекста у ком се користи. То значи да реч са више значења (полисемија) има само једну репрезентацију која представља просек свих њених употреба, чиме се губе битне семантичке нијансе.

Потреба за моделовањем речи у складу са динамичким контекстом реченице довела је до развоја механизма пажње и *Transformer* архитектуре, која представља тему наредног поглавља. За разлику од *Word2Vec* и *GloVe* модела, *Transformer* не генерише статичке векторе већ

контекстуализоване репрезентације које се мењају у зависности од окружења речи. Управо овакве контекстуалне репрезентације коришћене су у овом истраживању. Архитектура која омогућава њихово стварање представља се у наредном поглављу.

2.11. Transformer архитектура

Традиционални приступи обраде природног језика, превасходно засновани на *RNN*, дуго су представљали стандард. Међутим, ови модели процесирају податке секвенцијално, што онемогућава паралелизацију процеса обучавања и чини моделовање дугорочних зависности у дугим текстовима изазовним. Прекретницу у овој области означила је појава *Transformer* архитектуре [80]. У овом истраживању, *Transformer* архитектура представља теоријску основу за два методолошка приступа моделовања.

За разлику од претходника, *Transformer* у потпуности одбацује рекурентност. Уместо тога, ослања се искључиво на механизам пажње (енгл. *attention mechanism*) како би успоставио глобалне зависности између улаза и излаза. Ова промена парадигме омогућила је значајно већи степен паралелизације, што је довело до могућности обучавања модела на изузетно великим количинама текста у знатно краћем временском року.

Кључна иновација *Transformer* архитектуре је механизам пажње заснован на скаларном производу (енгл. *scaled dot-product attention*). Овај механизам омогућава моделу да приликом обраде сваке речи динамички обрати пажњу на друге релевантне речи у секвенци, без обзира на њихову удаљеност. На тај начин се ефикасно решава проблем полисемије, јер значење речи постаје директна функција њеног контекста [50].

Ово поглавље најпре представља оригиналну *Transformer* архитектуру и механизам пажње као њену кључну иновацију, а затим излаже породицу претрениранх енкодерских модела који су на овој архитектури засновани, заједно са парадигмом трансфера учења која омогућава њихову практичну примену.

2.11.1. Оригинална архитектура и механизам пажње

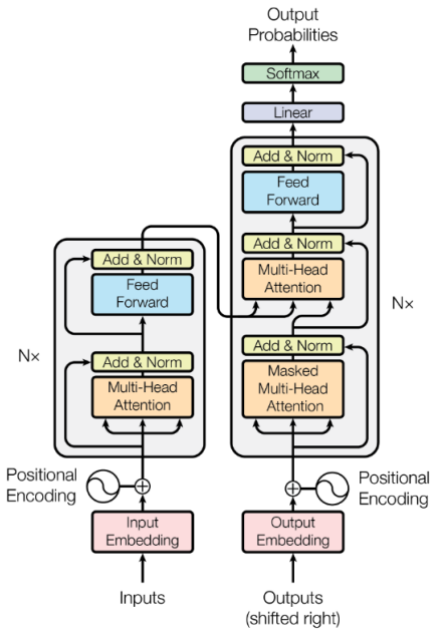
Функција пажње оперише над три вектора: упитом (Q), кључем и вредношћу (K, V). Излаз представља пондерисану суму вредности, где се тежине добијају мерењем сличности између упита и кључева [80]:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{18}$$

где фактор скалирања $\frac{1}{\sqrt{d_k}}$ служи за стабилизацију градијента током тренинга.

Како би модел могао да истовремено прати различите аспекте језичких структура (на пример: морфолошка слагања и семантичке везе), користи се вишеглава пажња (енгл. *multi-head attention*). Овај процес подразумева паралелно извршавање функције пажње кроз више независних пројекција (глава), чији се резултати на крају спајају. То омогућава моделу да креира богатију репрезентацију текста [50].

Слика 14 илуструје архитектуру *Transformer* модела.



Слика 14 Архитектура *Transformer* модела [80]

Transformer модел је оригинално конципиран као енкодер-декодер архитектура:

1. Енкодер: чини га низ слојева задужених за разумевање улазног текста. Сваки слој садржи механизам вишеглаве пажње и потпуно повезану мрежу, уз примену резидуалних веза и нормализације.
2. Декодер: генерише излазни текст. Поред стандардних слојева, садржи и додатни слој пажње који повезује излаз декодера са репрезентацијама из енкодера.

Пошто модел не обрађује речи редом, информација о редоследу речи се експлицитно уводи путем позиционог кодирања (енгл. *positional encoding*), које се додаје улазним векторима [80].

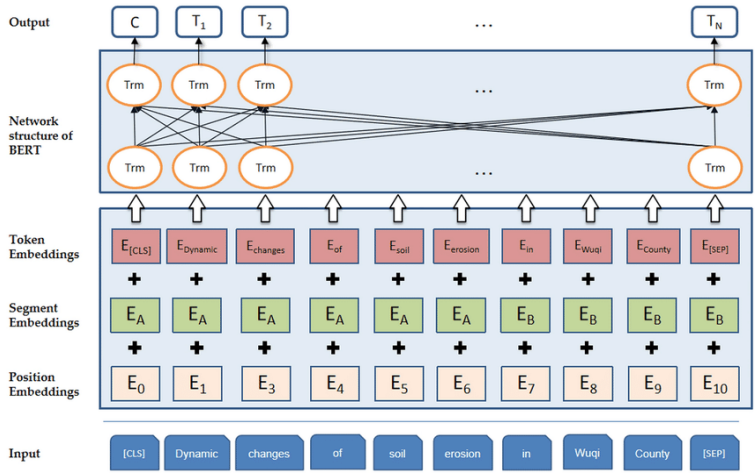
2.11.2. Претренирани енкодерски модели и трансфер учење

На основу *Transformer* архитектуре представљене у претходном одељку, развијена је читава фамилија претренираних модела за обраду природног језика. У задацима класификације текста, где је циљ изградња дубоких контекстуалних репрезентација улазне секвенце, доминирају модели засновани искључиво на енкодерској страни *Transformer* архитектуре. Ови модели се користе у комбинацији са парадигмом трансфера учења, која омогућава њихову ефикасну примену на специфичне задатке уз минималан додатни тренинг. Овај приступ је популаризовао модел *BERT* (енгл. *Bidirectional Encoder Representations from Transformers*).

BERT је вишеслојни двосмерни *Transformer* енкодер. Његова иновативност лежи у начину претренирања кроз два задатка [81]:

1. Маскирано језичко моделовање (енгл. *Masked Language Modeling - MLM*): модел не предвиђа само наредну реч, већ учи контекст из оба смера. Током тренинга, 15% токена се насумично маскира, а циљ модела је да их реконструише. Ово приморава модел да научи дубоке семантичке везе унутар реченице.
2. Предвиђање наредне реченице (енгл. *Next Sentence Prediction - NSP*): да би разумео односе између реченица, модел се тренира да за дати пар реченица *(A, B)* предвиди да ли реченица *B* логички следи након *A*.

Слика 15 илуструје структуру *BERT* модела.



Слика 15 Структура *BERT* модела

Поред модела за енглески језик, значајан допринос представља и вишејезични *BERT* (енгл. *multilingual BERT - mBERT*), трениран на корпусу Википедије на 104 језика. Овај модел дели исти векторски простор за све језике, што омогућава међујезички трансфер знања (енгл. *cross-lingual transfer*), односно способност модела да решава задатак на једном језику иако је фино подешаван на другом [82].

RoBERTa (енгл. *Robustly Optimized BERT Pretraining Approach*) је унапређена варијанта *BERT*-а. Аутори су утврдили да је оригинални *BERT* био недовољно трениран, те су увели неколико кључних побољшања. Прво, уклоњен је *NSP* задатак, јер су истраживања показала да он не доприноси значајно перформансама у задацима примене модела (енгл. *downstream tasks*), те да је тренирање искључиво на непрекидним блоковима текста ефикасније. Друго, уведено је динамичко маскирање уместо статичког, што омогућава моделу да у свакој епохи учи из другачијих образаца маскираних токена. На крају, модел је трениран на знатно већој количини података и са значајно већим бројем итерација. Резултат је модел који доследно надмашује *BERT* у широком спектру задатака [83].

Иако *RoBERTa* у својој оригиналној форми покрива искључиво енглески језик, њена методологија претренирања показала се изузетно успешном и у вишејезичном контексту. Као одговор на ограничења *mBERT*-а, развијен

је *XLM-RoBERTa* (енгл. *Cross-lingual Language Model - XLM-R*), вишејезична варијанта *RoBERTa* модела претренирана на корпусу текстова са Интернета на 100 језика. За разлику од *mBERT*-а, који је трениран на знатно мањем корпусу Википедије, *XLM-R* користи већи и разноврснији корпус, што посебно доприноси перформансама на језицима са ограниченим ресурсима. Поред тога, *XLM-R* наслеђује унапређене стратегије претренирања из *RoBERTa* приступа, чиме доследно надмашује *mBERT* у задацима класификације на језицима са ограниченим ресурсима [84].

Иако су вишејезични модели попут *mBERT*-а и *XLM-RoBERTa* широко применљиви, они често не успевају да ухвате фине нијансе специфичних језика због дељења капацитета мреже на велики број језика. Као одговор на то, развијен је *BERTić*, *Transformer* модел претрениран на корпусу од 8 милијарди токена прикупљених са домена босанског, хрватског, црногорског и српског језика (такозвани *BCMS* макројезик). За разлику од генеричких модела, *BERTić* је оптимизован искључиво за ову језичку групу. Емпиријска евалуација показала је да овај модел доследно надмашује *state-of-the-art* вишејезичке моделе у широком спектру задатака, чиме се потврђује неопходност коришћења наменски развијених модела за постизање врхунских резултата у локалном језичком домену [85].

Примена ових модела у пракси ослања се на парадигму трансфера учења (енгл. *transfer learning*). Овај приступ омогућава да се знање стечено на једном (општем) задатку пренесе и искористи за решавање другог (специфичног) задатка [50]. Процес се састоји из две фазе:

1. Претренирање (енгл. *pre-training*): у овој фази модел се обучава на огромним корпусима необележеног текста. Циљ је да модел научи општа својства језика: структуру реченице, значење речи, морфологију и синтаксу. Резултат су параметри модела који садрже богато лингвистичко предзнање.
2. Фино подешавање (енгл. *fine-tuning*): претренирани модел се затим прилагођава конкретном задатку. На излаз модела се додаје један нови слој специфичан за задатак, а затим се цела мрежа додатно тренира на мањем скупу означених података.

Кључна предност овог приступа је ефикасност. Док претренирање захтева огромне ресурсе, фино подешавање се може извршити брзо, постижући притом врхунске резултате чак и са релативно малим скупом података за обуку [50] [81].

Увођење механизма пажње и *Transformer* архитектуре решило је фундаменталне проблеме паралелизације и моделовања дугорочних зависности, који су ограничавали претходне *RNN* моделе. Међутим, прави потенцијал ове архитектуре постао је очигледан тек њеним скалирањем, односно значајним повећањем броја параметара, слојева и количине података за обуку. Емпиријска истраживања су показала да једноставно увећање *Transformer* модела доводи до тога да модел почиње да испољава способности решавања задатака за које није био наменски обучаван. Овај феномен означио је прелазак са специјализованих модела на моделе опште намене, постављајући темеље за развој великих језичких модела који ће бити описани у наредном поглављу.

2.12. Велики језички модели

Велики језички модели (енгл. *Large Language Models - LLM*) представљају пробабилистичке моделе засноване на неуронским мрежама који се обучавају на масивним текстуалним корпусима са циљем предвиђања наредног токена у секвенци. За разлику од ранијих приступа који су захтевали специфично тренирање за сваки појединачни *NLP* задатак, савремени *LLM*-ови функционишу као основни модели (енгл. *Foundation Models*). То значи да се један претренирани модел може адаптирати за широк спектар задатака, од машинског превођења и сумирања текста до писања кода и логичког закључивања, често без потребе за ажурирањем његових параметара. Суштина њиховог рада заснива се на моделовању условне вероватноће појављивања секвенце речи коришћењем ланчаног правила вероватноће [50]:

$$P(w_{1:N}) = \prod_{i=1}^{N} P(w_i | w_1, \dots, w_{i-1}) . \tag{19}$$

Иако се *LLM*-ови могу реализовати кроз различите варијанте *Transformer* модела, модели из *GPT* (енгл. *Generative Pre-trained Transformer*) серије

постали су синоним за напредак у овој области захваљујући својој способности да скалирањем архитектуре постигну изузетне генеративне способности. У овом истраживању, велики језички модели представљају један методолошки приступ за класификацију студентских повратних информација, при чему су коришћена два модела различитих породица: *GPT* и *Gemma*. Архитектуре ових модела представљене су у наредним одељцима.

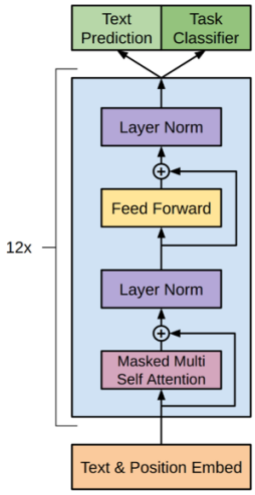
2.12.1. Архитектура GPT модела

Доминантна архитектура за изградњу савремених генеративних модела, коју је популаризовала *GPT* серија, јесте архитектура која користи само декодер (енгл. *Decoder-only*) [86]. За разлику од оригиналног *Transformer* модела који је користио енкодер-декодер структуру примарно за машинско превођење, *GPT* модели користе низ декодерских блокова за ауторегресивно генерисање текста [50]. Овакви модели се називају каузални језички модели јер предвиђање наредног токена зависи искључиво од претходних токена у секвенци, чиме се поштује временска каузалност информација [86].

Централна компонента ове архитектуре је механизам маскиране самопажње (енгл. *masked self-attention*). У стандардној самопажњи, каква се користи код енкодера попут *BERT* модела, сваки токен има приступ информацијама из целе реченице. Међутим, у ауторегресивном моделовању, такав механизам би омогућио моделу увид у будуће токене које тек треба да предвиди. Маскирање решава овај проблем тако што онемогућава приступ свим наредним позицијама у секвенци, чиме се осигурава да сваки токен може обрадити информације искључиво из свог претходног контекста и самог себе [80].

GPT модели се састоје од низа наслаганих *Transformer* блокова. Сваки блок садржи слој маскиране самопажње са више глава, након чега следи потпуно повезана неуронска мрежа. Унутар сваког блока примењују се резидуалне везе (енгл. *residual connections*), које функционишу тако што се улаз у слој сабира са његовим излазом, чиме се формира континуирани проток информација који олакшава пренос градијента кроз мрежу. Поред тога, користи се нормализација слојева (енгл. *layer normalization*), која

стандардизује вредности вектора унутар скривених слојева (центрирањем око нуле и скалирањем), чиме се стабилизује процес учења и омогућава бржа конвергенција модела. Ова структура, коју илуструје слика 16, омогућава стабилно тренирање веома дубоких мрежа [50] [86].



Слика 16 Структура *Transformer* блокова *GPT* модела [86]

Излаз последњег блока за дати токен пролази кроз линеарни слој који га пресликава у димензију вокабулара, након чега се примењује *softmax* функција како би се добила дистрибуција вероватноћа за следећу реч:

$$P(w_i|w_{1:i-1}) = \text{Softmax}(h_i W_{vocab}), \tag{20}$$

где h_i представља скривено стање последњег слоја за корак i, а W_{vocab} матрицу тежина [50] [87].

Ова архитектура је показала изузетну скалабилност, што је омогућило развој све моћнијих модела једноставним увећањем броја слојева и ширине вектора, уз адекватно повећање количине података за обуку. Еволуција *GPT* серије модела, приказана у табели 3, илуструје како експоненцијални раст броја параметара доводи до квалитативних скокова у способностима модела, омогућавајући им да генеришу кохерентан текст дужине више хиљада речи и решавају сложене задатке [86] [87].

Табела 3 обухвата генерације модела чија је архитектура званично документована. Касније генерације попут *GPT-3.5* и *GPT-4*, иако су задржале основне архитектонске принципе декодерске *Transformer*

структуре, своде се на интерне моделе чији технички детаљи (укупан број параметара, број слојева и димензија вектора) нису јавно објављени од стране *OpenAI*-ја. Због тога се ови модели у литератури обично описују кроз њихово понашање и могућности, а не кроз архитектонске спецификације.

Табела 3 Еволуција *GPT* серије модела

Модел	Година	Број параметара	Број слојева	Димензија модела
GPT-1	2018.	117 милиона	12	768
GPT-2	2019.	1.5 милијарди	48	1600
GPT-3	2020.	175 милијарди	96	12288

Експоненцијални раст капацитета ових модела није само побољшао њихове перформансе у моделовању језика, већ је довео до фундаменталне промене у начину на који се они примењују у пракси. Док су претходне генерације *NLP* модела захтевале опсежно ажурирање параметара (фино подешавање) за сваки специфичан задатак, архитектура *GPT* модела омогућава решавање нових проблема искључиво манипулацијом улазним контекстом. Ова способност елиминише потребу за додатним тренингом и поставља темељ за нову методолошку дисциплину познату као промпт инжењерство (енгл. *prompt engineering*) [50].

2.12.2. Архитектура Gemma модела

Породица модела *Gemma* представља серију модела отворених тежина (енгл. *open-weight*) коју је развио *Google DeepMind*, засновану на истраживањима и технологији коришћеној за развој *Gemini* модела. Иако се ослања на исту каузалну архитектуру која користи само декодер као и *GPT* архитектура, *Gemma* архитектура уводи специфичне техничке модификације усмерене на побољшање стабилности тренинга и ефикасности закључивања на хардверу са ограниченим ресурсима [88].

За разлику од стандардних имплементација *Transformer* архитектуре, *Gemma* архитектура уводи три кључне модификације усмерене на побољшање стабилности и ефикасности мреже. Прва иновација је примена *RMSNorm* (енгл. *Root Mean Square Layer Normalization*) механизма [89] на улазу сваког слоја, који нормализује активације без

потребе за рачунарски захтевним поновним центрирањем података. Тиме се осигурава нумеричка стабилност дубоких мрежа и омогућава бржа конвергенција током обуке. Даље, традиционална *ReLU* функција замењена је *GeGLU* активационом функцијом [90], која кроз механизам капије (енгл. *gate*) омогућава моделу да ефикасније моделује сложене нелинеарне интеракције између карактеристика, што директно побољшава капацитет учења и финалне перформансе на језичким задацима. Коначно, за инкорпорирање информација о поретку токена користе се ротациони позициони вектори (енгл. *Rotary Position Embeddings - RoPE*) [91] који уместо апсолутних позиција користе ротацију вектора у комплексном простору. Овај приступ омогућава механизму пажње да природно препознаје релативне удаљености између токена, што је кључни предуслов за ефикасну генерализацију модела на дуге секвенце и рад са контекстуалним прозорима великог капацитета.

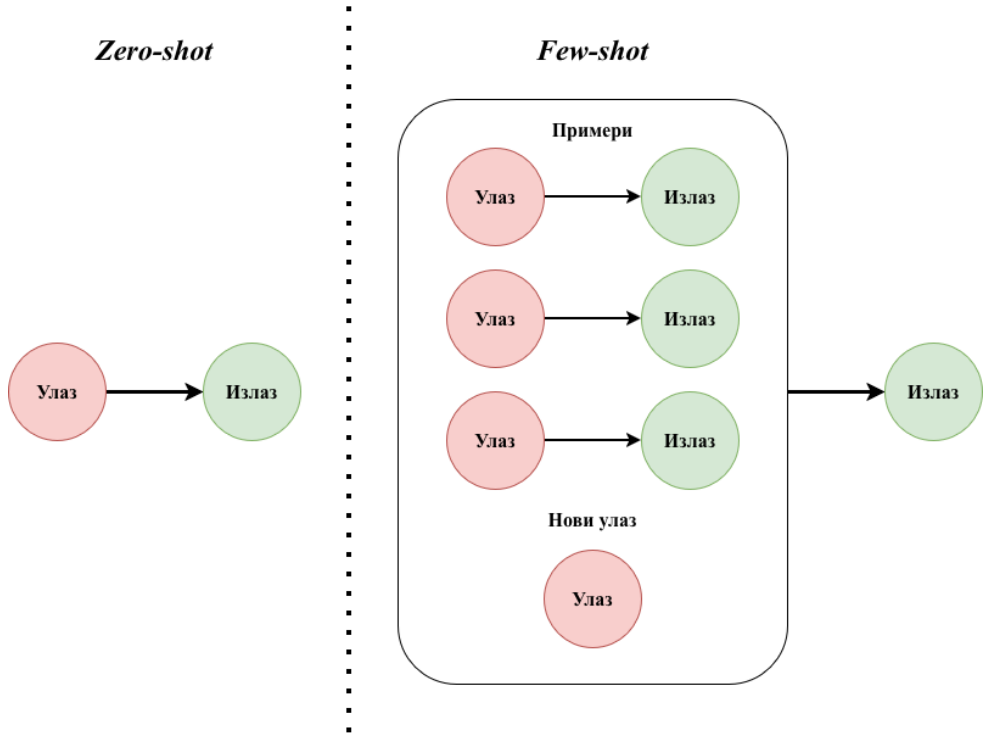
Најновији модел, *Gemma 3*, доноси фундаменталне промене у начину на који се процесира контекст, прелазећи са чисто текстуалног на мултимодално моделовање уз увођење напредних оптимизација за дуге секвенце. Кључна архитектонска новина је хибридни механизам пажње који уводи смену слојева локалне пажње са клизним прозором и слојева глобалне пажње у односу 5:1. У овој конфигурацији, локални слојеви оперишу на кратким распонима од 1024 токена, док само глобални слојеви обрађују комплетан контекст, чиме се ефикасно редукује рачунарска сложеност и потрошња меморијских ресурса при обради веома дугих секвенци. Додатно побољшање обухвата оптимизацију рачунања пажње кроз увођење *QK*-нормализације, која осигурава стабилније вредности унутар самог механизма и доприноси већој нумеричкој поузданости модела са великим бројем параметара [92].

За потребе анализе обимних скупова података од великог је значаја и проширени контекстуални прозор од 128000 токена. *Gemma 3* модел је доступан у варијантама од 1, 4, 12 и 27 милијарди параметара, при чему највећи модел постиже перформансе које су упоредиве са најмоћнијим комерцијалним моделима затвореног типа, али уз задржавање локалне имплементације и пуне контроле над приватношћу података [92].

2.12.3. Промпт инжењерство

Промпт инжењерство представља емпиријску методологију дизајнирања и оптимизације улазних текстова са циљем да велики језички модел усмери ка извршавању специфичног задатка и то без потребе за ажурирањем његових параметара. Овај приступ се ослања на способност модела да препознаје обрасце у тексту и на основу њих генерише одговарајући наставак што се у литератури назива учење у контексту (енгл. *in-context learning*) [87]. За разлику од традиционалног финог подешавања где се знање уграђује у тежине модела кроз градијентни спуст, овде се инструкције и примери пружају као део улазног контекста у тренутку закључивања [50].

Ефикасност промпта зависи од његове формулације и структуре, као и количине пружених информација. У зависности од тога да ли се моделу пружају примери жељеног излаза разликују се две основне *zero-shot* и *few-shot*. Слика 17 илуструје ове стратегије.



Слика 17 Основне стратегије промпт инжењерства

Zero-shot подразумева да се моделу прослеђује само опис задатка или инструкција без икаквих демонстрација или примера како тај задатак треба решити. Модел се у овом случају ослања искључиво на знање о језику и свету које је стекао током фазе претренирања како би разумео инструкцију и генерисао одговор [50]. Иако је овај метод рачунарски најефикаснији јер захтева најмањи контекст, истраживања показују да може бити мање поуздан код сложених задатака где формат излаза није очигледан или где је потребно специфично резоновање [93].

Са друге стране, *few-shot* стратегија подразумева да се у промпт поред инструкције укључи и мали број примера (парова улаз и излаз). Ови примери служе као образац који модел треба да прати, чиме му се омогућава да научи специфичности задатка и жељени формат одговора само на основу контекста. Истраживања су показала да повећање броја примера у промпту, до одређене границе, значајно повећава тачност модела и често достиже перформансе које су упоредиве са моделима који су фино подешени на мањим скуповима података [87].

Коришћење *few-shot* приступа је посебно значајно у сценаријима са ограниченим ресурсима, јер омогућава брзу адаптацију модела на нове домене без потребе за скупим процесом поновног тренирања, што је био основни мотив за његову примену у овом истраживању. Међутим, треба напоменути да квалитет резултата у великој мери зависи од избора репрезентативних примера и њиховог редоследа у промпту [50].

Представљени теоријски концепти, од статистичких алгоритама преко неуронских мрежа до великих језичких модела, чине технолошку основу за аутоматску обраду текста. Међутим, примена ових метода на комплексан домен анализе студентских повратних информација носи јединствене изазове. Како би се разумело на који начин се ови алгоритми користе у пракси за екстракцију знања из неструктурираних коментара, потребно је сагледати досадашња научна достигнућа у овој области. Наредно поглавље пружа детаљан преглед сродних радова, са фокусом на инжењерство захтева засновано на групи корисника и аутоматску детекцију емоција у образовном контексту.

3. Преглед актуелног стања у области

Развој система за аутоматску анализу студентских повратних информација захтева дубоко разумевање постојећих теоријских и методолошких приступа у три области: инжењерству захтева заснованом на групи корисника, детекцији емоција усмерених на учење и обради српског језика. Иако свака од ових истраживачких области има сопствену историју, доменске специфичности и изазове, њихова интеграција постаје неопходна за развој поузданих модела способних да интерпретирају сложене аспекте студентских коментара у окружењу интелигентних система за подучавање.

Ово поглавље пружа преглед релевантне литературе у наведеним доменима. Најпре се разматрају основе инжењерства захтева заснованог на групи корисника, затим се анализирају модели и налази о емоцијама усмереним на учење, а на крају се даје преглед стања у области обраде српског језика као језика са ограниченим ресурсима. Оваква структура омогућава идентификацију кључних изазова, ограничења постојећих приступа и истраживачких празнина које мотивишу развој ресурса и модела у оквиру ове дисертације.

3.1. Инжењерство захтева засновано на групи корисника

Инжењерство захтева засновано на групи корисника (енгл. *Crowd-based Requirements Engineering* - *CrowdRE*) представља приступ који користи аутоматизоване или полуатоматизоване методе за прикупљање и анализу података од великог броја учесника у циљу идентификације и валидације корисничких захтева [25]. Овај приступ омогућава развојним тимовима приступ широкој и разноликој бази стварних корисника, чиме се обезбеђује да дефинисани захтеви верније одражавају реалне потребе корисника и доприносе већој одрживости софтверског решења [94]. С обзиром на то да је ручна екстракција информација из великог броја корисничких коментара скупа, временски захтевна и подложна грешкама услед субјективних интерпретација и људске пристрасности, примена напредних техника обраде природног језика постала је неопходан предуслов за ефикасну анализу корисничких повратних информација [95].

Да би се велики обим неструктурираних корисничких података могао систематично анализирати, неопходно је успоставити одговарајуће таксономије које дефинишу јасне категорије за класификацију. Ранија истраживања у овој области уложила су значајан напор у дефинисање таквих таксономија, примарно се фокусирајући на разликовање пријава грешака (енгл. *bug report*) и захтева за новим функционалностима (енгл. *feature request*) у оквиру рецензија мобилних апликација [24] [96] [97] [98] [99] [100] [101]. Иако већина предложених таксономија дели ове основне категорије, уочене су значајне варијације у погледу нивоа детаљности, броја поткласа и коришћене терминологије, што је отежавало поређење резултата различитих студија. У циљу превазилажења овог проблема и хармонизације различитих приступа, спроведена су свеобухватна истраживања којима су постојеће категорије консолидоване у унифковану таксономију, груписану према намери корисника и теми коментара, чиме је постављен стандардизован оквир за анализу повратних информација [102].

Међутим, доминантне таксономије у литератури претежно су изведене из домена мобилних апликација доступних на глобалним дистрибутивним платформама, што намеће ограничења приликом њихове примене у специфичним доменима као што су интелигентни системи за подучавање. Док су мобилне апликације често дизајниране са фокусом на кратке и фреквентне интеракције, образовни софтвер наглашава персонализовано и адаптивно искуство учења које подразумева дуготрајнији и когнитивно захтевнији ангажман корисника. Услед ових суштинских разлика у намени и начину употребе софтвера, таксономије развијене за општи домен мобилних апликација не могу се директно пресликати на образовни контекст без одговарајућих модификација и проширења која би обухватала специфичне педагошке и дидактичке аспекте интеракције.

Рана истраживања у домену аутоматске класификације корисничких захтева претежно су се ослањала на технике тематског моделовања (енгл. *Topic Modeling*), са циљем идентификације скривених тематских структура у великим количинама неструктурираних текстуалних података без потребе за претходним анотирањем. Иако су ови приступи омогућавали иницијални увид у садржај коментара, њихова ограничена прецизност и интерпретабилност су отежавале практичну примену у

инжењерству захтева [24] [96] [97]. Као одговор на ова ограничења, фокус истраживања померен је ка примени алгоритама надгледаног машинског учења, попут метода потпорних вектора или ансамбала, који су у комбинацији са статистичким репрезентацијама текста постизали значајно боље перформансе [98] [99] [100] [101] [103] [104] [105]. Међутим, већина ових истраживања није примењивала технике за балансирање података, упркос чињеници да су скупови података у *CrowdRE* изразито небалансирани, где већина података припада општим категоријама пријављивања грешака и захтева за новим функционалностима [102].

Новија истраживања усмерена су на превазилажење ограничења традиционалног машинског учења применом метода дубоког учења, а посебно претренираних модела базираних на *Transformer* архитектури (енгл. *Pre-trained Transformer Models - PTM*) [106] [107]. Показано је да модели као што су *BERT* и његови деривати надмашују претходне приступе. Међутим, ни ове студије нису примењивале технике за балансирање података.

Како би се смањила зависност од скупих и ретких анотираних података, појавили су се приступи засновани на ненадгледаном машинском учењу. Ови приступи комбинују технике за генерисање векторских репрезентација реченица са алгоритмима кластеровања и тематског моделовања, чиме се елиминише потреба за ручним анотирањем података. Иако су експерименти показали да овакви приступи могу обезбедити ефективну категоризацију захтева у одсуству анотираних података, значајан пад перформанси у ситуацијама када различите категорије деле велики део вокабулара указује на ограничену способност ненадгледаних алгоритама да раздвоје суптилне семантичке разлике [94].

Најновији правац истраживања испитује потенцијал великих језичких модела у области *CrowdRE*. Примена контекстуализованих векторских репрезентација добијених из ових модела омогућава делимично премошћавање семантичког јаза између језика корисника и техничке терминологије. Ипак, посебан изазов представља обрада вишејезичних података, јер је показано да машинско превођење рецензија пре обраде уводи шум и негативно утиче на перформансе система. Налази сугеришу

да приступи који генеришу векторске репрезентације директно на изворном језику обезбеђују поузданије резултате у поређењу са онима који се ослањају на превођење [108].

Поред методолошких изазова, значајно ограничење постојеће литературе представља недостатак разноврсних и висококвалитетних скупова података. Већина доступних корпуса обухвата корисничке рецензије мобилних апликација написане на енглеском језику, док су ресурси за језике са мањом заступљеношћу у дигиталном простору изузетно ретки. Ово представља озбиљну препреку за широку примену *CrowdRE* решења, будући да је експериментално потврђено да машинско превођење података на енглески језик уводи шум и негативно утиче на тачност класификације у поређењу са обрадом на изворном језику [108].

Други кључни недостатак односи се на процес анотације података. У контексту надгледаног машинског учења, доступност висококвалитетних анотираних скупова података представља један од главних предуслова за развој поузданих модела. У претходним истраживањима, приступи класификацији корисничких коментара најчешће су евалуирани на скуповима података креираним ручном анотацијом [96] [97] [98] [99] [100] [101], ручно дефинисаним правилима [24], или комбинацијом *CrowdRE* алата и ручне анотације [103]. Већина наведених студија користила је више независних анотатора, али без транспарентног извештавања о њиховој обуци и без доследне примене формалних мера сагласности између више анотатора, што представља значајно одступање од најбољих пракси анотације природног језика и отежава објективну процену поузданости и репликабилности доступних ресурса.

Синтеза постојеће литературе указује на више отворених истраживачких изазова у области *CrowdRE*. Пре свега, уочава се недостатак адекватне подршке за *CrowdRE* у домену *ITS*, где су корисничке интеракције, циљеви и повратне информације суштински различити у односу на домене у којима су постојеће таксономије и модели развијени. Додатно, евидентан је јаз у доступности језичких ресурса и методолошких решења за језике изван енглеског говорног подручја, што намеће изазове у раду са језицима са ограниченим ресурсима. Такође, постоји простор за унапређење перформанси модела применом техника за балансирање

података, с обзиром на изразиту неравнотежу класа у стварним скуповима података. Коначно, идентификован је ограничен број студија које у овом контексту систематски испитују и упоређују ефикасност савремених приступа заснованих на великим језичким моделима у односу на *PTM*-ове и традиционалне технике машинског учења.

3.2. Емоције усмерене на учење

Област аутоматске детекције емоција из текста (енгл. *Text-Based Emotion Detection - TBED*) прошла је кроз значајну еволуцију, пратећи шири развој обраде природног језика. Рана истраживања ослањала су се на приступе засноване на кључним речима, статистичком машинском учењу и хибридним методама за детекцију емоција, најчешће у оквиру општих модела као што су Екманов модел основних емоција или *OCC* модел (енгл. *Ortony, Clore, and Collins*) [109]. Иако једноставни и интерпретабилни, ови приступи су показали значајна ограничења у разумевању контекста и вишезначности речи. Како показују новији прегледни радови, напредак у овој области условио је прелазак ка техникама дубоког учења, векторским репрезентацијама речи и трансферу учења, превасходно за енглески језик [110] [111] [112]. Студије указују да приступи засновани на дубоком учењу превазилазе раније методе, будући да су способни да моделују секвенцијалне зависности и семантички контекст, али и да њихове перформансе критично зависе од квалитета претпроцесирања и величине доступних скупова података. Значајан помак представљају модели засновани на *Transformer* архитектури, који користе механизам пажње за дубинско разумевање контекста. Додатно, савремени велики језички модели, попут модела из *GPT* породице, омогућавају једноставнију примену и остварују резултате који су упоредиви са специјализованим приступима дубоког учења (енгл. *Deep Learning - DL*), чак и у сценаријима са ограниченом количином података [113] [114].

Имплементација ових технологија у образовном домену, а посебно у *ITS*, започела је применом приступа базираних на кључним речима за детекцију основних људских емоција [115] [116]. Примарни недостатак оваквих приступа огледа се у томе што су били усмерени на детекцију емоција које се ретко јављају у контексту решавања академских задатака.

Поред тога, ове ране студије нису обезбедиле транспарентне процедуре анотације података нити детаљне описе експерименталне методологије, што је отежало репликацију резултата. Савременије студије које су користиле вишејезичне приступе засноване на лексиконима за категоризацију студентских коментара према Плучиковом моделу (енгл. *Plutchik's Wheel of Emotions*) такође нису у потпуности превазишле наведене недостатке [117].

Увидевши ограничења општих модела емоција, истраживачка заједница се постепено окренула ка детекцији емоција усмерених на учење, које су детаљније представљене у поглављу 2.2. Већина истраживања у овом специфичном домену примењивала је статистичке *ML* и *DL* технике за анализу студентских коментара, најчешће у контексту учења програмирања у *ITS* окружењима на шпанском језику [118] [119] [120]. Иако су ове студије направиле значајан искорак креирањем сопствених скупова података, њихове процедуре анотације карактерише недостатак транспарентности. Недостајале су јасно дефинисане инструкције за анотаторе, док метрике *IAA* нису биле израчунате. Значајан методолошки недостатак ових радова представља ослањање на тачност као примарну метрику за евалуацију. С обзиром на то да су образовни подаци инхерентно небалансирани, са доминацијом позитивних и неутралних коментара, метрика тачности не пружа реалистичну слику о способности модела да детектује мањинске, али педагошки кључне класе попут фрустрације и збуњености.

Овај проблем валидности података додатно је истакнут у најновијем истраживању које је упоредило *ML*, *DL* (укључујући *BERT*) и еволуционе приступе на задатку класификације студентских коментара [121]. Аутори су развили два скупа података на шпанском језику, који су иницијално анотирани применом приступа базираном на лексикону, а затим додатно рафинирани од стране само једног анотатора. Ослањање на само једног анотатора уводи значајан ризик субјективности у процес анотације, што негативно утиче на поузданост података који служе као основа за тренирање модела. Као и претходне, и ова студија користила је тачност као метрику евалуације, занемарујући метрике погодније за вишекласне проблеме са небалансираним подацима.

Док су глобална истраживања усмерена ка детекцији емоција специфичних за процес учења, истраживања у контексту српског језика примарно су остала у домену анализе сентимента. Ранији радови бавили су се одређивањем поларности (позитивно/негативно) коментара о професорима и настави, користећи приступе засноване на правилима и техникама машинског учења [122]. Такође, спроведена је и аспектна анализа сентимента на подацима из студентских анкета [123]. За разлику од поменутих студија, ово истраживање помера фокус са општег сентимента на финију гранулацију емоција специфичних за когнитивне процесе учења. Ипак, поменуте студије истичу недостатак јавно доступних, домен-специфичних скупова података и ресурса за српски језик, што додатно оправдава значај овог истраживања.

Свеобухватном анализом стања у области уочене су јасне потребе за даљим истраживањем, које проистичу из ограничења тренутних приступа. Првенствено, неопходно је успостављање стандардизованих и транспарентних процедура анотације података које укључују више независних анотатора и рачунање *IAA* ради осигурања валидности података. Друго, потребно је унапређење евалуације на небалансираним скуповима применом техника балансирања података и коришћењем метрика попут *F1*-мере, уместо просте тачности која маскира перформансе на мањинским класама. Коначно, евидентан је недостатак ресурса за српски језик, с обзиром на то да су постојећа истраживања ограничена на општу анализу сентимента.

3.3. Обрада српског језика

Примена *NLP* техника на српски језик носи специфичне изазове који га издвајају од језика са богатим ресурсима. Српски језик, као део јужнословенске групе језика, одликује се богатом морфологијом, слободним редом речи у реченици и феноменом диграфије, односно активном употребом и ћириличног и латиничног писма [124] [125]. Висок степен флексије, односно промена речи кроз падеже, родове, бројеве и лица, доводи до значајног раста броја јединствених облика речи, што у статистичким моделима често узрокује проблем разређености података (енгл. *data sparsity*) [125]. Због тога је обрада српског језика

традиционално захтевала развој сложених лексичких ресурса и алата за нормализацију текста.

Рана истраживања у овој области била су усмерена на изградњу темељних ресурса неопходних за савладавање морфолошке комплексности језика. Кључни напредак постигнут је развојем Српског морфолошког речника (енгл. *Serbian Morphological Dictionary - SMD*) и пратећих алата за лематизацију и морфосинтаксичко означавање (енгл. *Part-of-Speech (POS) tagging*). Овај ресурс је био значајан за развој и евалуацију различитих алата, од оних заснованих на Марковљевим моделима до новијих приступа заснованих на дубоким неуронским мрежама [126]. Паралелно са тим, развијани су алгоритамски алати за кореновање [127]. У домену класификације текста, истраживања пре ере дубоког учења ослањала су се на статистичке алгоритме машинског учења, показујући да морфолошка нормализација значајно доприноси перформансама модела, али и да без великих анотираних корпуса статистички модели тешко хватају сложене синтаксичке и семантичке зависности [128].

Значајну прекретницу у обради српског језика означила је појава *Transformer* модела. Док су иницијално коришћени вишејезични модели попут *mBERT*-а, истраживања су показала да модели тренирани на специфичним језичким групама постижу врхунске резултате. Кључни напредак у овом домену представља развој модела *BERTić*, *Transformer* модела базираног на *BERT* архитектури, који је претрениран на више од 8 милијарди токена текста прикупљених са *Web* домена Србије, Хрватске, Босне и Херцеговине и Црне Горе [85]. Евалуација овог модела на задацима као што су препознавање именованих ентитета (енгл. *Named Entity Recognition - NER*), *POS tagging* и геолоцирање, показала је да *BERTić* доследно надмашује вишејезичне моделе попут *mBERT*-а, јер боље моделује специфичне морфосинтаксичке карактеристике локалног језичког простора [85].

Најновији талас технолошког развоја донео је генеративне *LLM*-ове и промпт инжењерство. Најактуелније истраживање систематски је испитало перформансе *LLM*-ова у поређењу са фино подешеним *BERT* моделима на задацима класификације текста за јужнословенске језике

[129]. Резултати указују на то да *LLM*-ови показују изузетне способности у *zero-shot* и *few-shot* сценаријима, често достижући или превазилазећи перформансе фино подешених *BERT* модела у задацима као што су анализа сентимента и класификација тема. Међутим, истраживање истиче да, иако *LLM*-ови нуде импресивне перформансе без претходног тренинга, фино подешени модели засновани на *BERT*-у и даље представљају практичније и поузданије решење у сценаријима који захтевају високу ефикасност и брзину, при чему се највећи потенцијал види у хибридним приступима где се *LLM* користи за аутоматску анотацију података на којима се потом тренирају мањи, ресурсно ефикаснији модели [129].

Синтезом прегледа литературе за српски језик уочава се јасан недостатак ресурса у кључним доменима овог истраживања. Првенствено, потребно је нагласити потпуно одсуство ресурса и алата за *CrowdRE* на српском језику. Паралелно са тим, у домену детекције емоција, иако је задатак анализе сентимента (поларности) релативно добро покривен постојећим скуповима података, не постоје јавно доступни корпуси нити модели развијени за детекцију финијих, педагошки релевантних емоција. Поред тога, поређење између финог подешавања и промпт инжењерства још увек није извршено у контексту *ITS*-а на српском језику. Ове идентификоване празнине мотивишу креирање наменских ресурса и систематску евалуацију обе парадигме како би се утврдио оптималан приступ за развој напредних образовних система на српском језику.

4. Корпус златног стандарда

Развој и евалуација система за аутоматску анализу студентских повратних информација захтевају постојање поузданих скупова података који служе као златни стандард за тренирање и евалуацију модела машинског учења. У контексту сложених задатака, као што су инжењерство захтева засновано на групи корисника и детекција емоција усмерених на учење, транспарентност анотационог процеса и квалитет података директно утичу на валидност добијених резултата.

Ово поглавље описује процес креирања корпуса развијеног за потребе овог истраживања. Представљени су поступци прикупљања и претпроцесирања текстуалних података у реалном образовном окружењу, примењена методологија анотације, као и дизајн анотационих шема за оба истраживачка задатка. На крају, приказане су карактеристике креираних скупова података и дискутована њихова ограничења, чиме су дефинисани предуслови за спровођење експерименталних анализа у наставку рада.

4.1. Прикупљање података

Подаци коришћени у овом истраживању прикупљени су у реалном образовном окружењу, у оквиру платформе за адаптивно учење *Clean CaDET* [23] која се користи у настави на универзитетском нивоу. Подаци су прикупљени током четворонедељног теренског истраживања спроведеног у циљу евалуације и унапређења дизајна платформе, као и њеног образовног садржаја. Прикупљање података током више наставних циклуса омогућило је обухватање разноврсних искустава студената са различитим нивоима предзнања, мотивације и ангажованости. Учесници истраживања били су студенти друге године на предмету *Увод у софтверско инжењерство* на Факултету техничких наука Универзитета у Новом Саду.

Студенти су учествовали у недељним активностима учења путем платформе које су претходиле настави, како би се припремили за сесије уживо. На крају активности, студенти су одговарали на следећа питања отвореног типа:

1. Како бисте унапредили *Clean CaDET* платформу као софтвер?
2. Како бисте описали ваше искуство учења са *Clean CaDET* платформом?

Одговори на прво питање чинили су *CrowdRE* корпус коришћен у овом истраживању, док су одговори на друго питање чинили корпус за емоције усмерене на учење. Одговори су прикупљани у анонимизованом облику, без личних идентификатора, у складу са етичким принципима истраживања и важећим прописима о заштити података. На овај начин обезбеђено је да анализирани текстуални садржаји одражавају аутентичне реакције студената, без утицаја социјално пожељног понашања или страха од потенцијалних последица. У даљем тексту, сви текстуални одговори биће означавани термином *коментари*.

Добијени коментари обухватали су текстове различите дужине, језичког стила и квалитета формулације, укључујући кратке исказе, непотпуне реченице, разговорне конструкције и правописне варијације. Ове карактеристике су типичне за спонтане корисничке повратне информације и представљају додатни изазов за каснију аутоматску анализу, али истовремено обезбеђују висок степен валидности прикупљених података.

Пре приступања анотацији, извршена је основна селекција података како би се уклонили дупликати и коментари који не садрже информативан текстуални садржај. Ради обезбеђивања конзистентности, сви коментари изворно написани ћириличним писмом аутоматски су конвертовани у латиницу, узимајући у обзир диграфску природу српског језика. Будући да коментари студената често садрже више реченица које се тичу различитих аспеката платформе, сваки одговор је сегментиран на појединачне реченице употребом програмског језика *Python* и *NLTK* библиотеке [49].

Пратећи раније истраживачке праксе у областима *CrowdRE* и *TBED*, усвојена је анотација на нивоу реченице како би се омогућила детаљнија анализа повратних информација корисника [106] [111]. Иако овај ниво грануларности повремено може довести до губитка ширег контекста, он олакшава прецизнију класификацију и боље одржава структуру прикупљених података.

4.2. Методологија анотације

У оквиру прегледа релевантне литературе у поглављу 3 идентификовани су бројни методолошки недостаци у постојећим истраживањима, пре свега изражена субјективност анотације, недовољна транспарентност поступка и изостанак формалне процене поузданости анотације. Ови проблеми су посебно изражени у задацима који укључују апстрактне и доменски специфичне категорије, као што су корисничке намере у *CrowdRE* и емоције усмерене на учење, где нејасно дефинисане шеме и недоследна примена категорија могу значајно угрозити валидност добијених резултата.

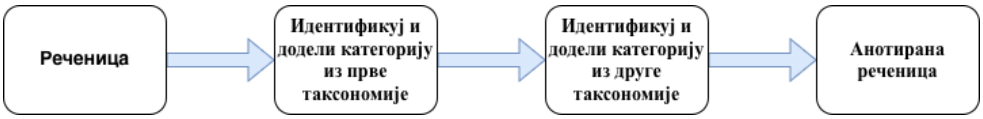
У циљу превазилажења наведених ограничења, у овом истраживању успостављена је процедура анотације заснована на *MATTER* методологији (поглавље 2.3.1), уз доследну примену најбољих пракси анотације природног језика описаних у теоријским основама (поглавље 0). С обзиром на карактер задатака који обухватају апстрактне и концептуално сложене категорије, примењен је прескриптивни приступ анотацији, при чему су анотатори доносили одлуке на основу унапред дефинисаних анотационих шема и детаљних смерница. Анотациони процес је осмишљен као итеративан и транспарентан, са пилот фазом и систематском анализом неслагања међу анотаторима, чиме је обезбеђена висока поузданост и методолошка утемељеност креираних скупова података.

Анотација оба корпуса спроведена је коришћењем алата *LightTag*, специјализованог *Web* окружења за анотацију текстуалних података који омогућава рад више анотатора, праћење верзија и ефикасно управљање анотационим процесом [130]. Избор овог алата омогућио је транспарентну и контролисану реализацију поступка анотације.

У процесу анотације оба корпуса учествовало је по четири анотатора са универзитетским образовањем, укључујући аутора ове дисертације. *CrowdRE* корпус су анотирали доктор наука и три докторанда из области електротехнике и рачунарства, док су корпус емоција анотирала два мастер психолога, докторанд српског језика и књижевности и докторанд електротехнике и рачунарства. Укључивање више независних анотатора омогућило је систематску процену поузданости анотације и смањење

ризика од индивидуалне пристрасности. Након дефинисања почетних анотационих шема и смерница, организована је двочасовна обука анотатора, током које су детаљно представљене анотационе шеме, смернице, као и начин коришћења алата.

Приликом анотације оба корпуса праћена је процедура коју илуструје слика 18, која се заснива на примени дефинисаних таксономија. За *CrowdRE* корпус коришћене су таксономије *Намере* и *Теме*, док су за корпус емоција категорије додељиване према таксономијама *Аспекта* и *Емоције*. Оваква декомпозиција проблема омогућила је прецизнију класификацију и дубљу анализу корисничких повратних информација. Детаљније објашњење структуре и дефиниција ових таксономија биће дато у наредном поглављу.



Слика 18 Анотациона процедура

Како би се емпиријски проверила јасноћа и применљивост дефинисаних категорија, спроведене су пилот фазе анотације (енгл. *proof-of-concept, PoC*). У свакој *PoC* итерацији, насумично је биран подскуп који је чинио приближно 10% корпуса, при чему су сви анотатори независно анотирали све јединице у изабраном подскупу. Након завршетка анотације, израчуната је међуанотаторска сагласност како би се проценила стабилност и недвосмисленост анотационог задатка. У случајевима када сагласност није достигла унапред дефинисани праг, анотатори су заједнички анализирали спорне случајеве, након чега су анотационе шеме и смернице ревидиране. Потом је поступак поновљен над новим насумичним подскупом података.

За процену поузданости анотације коришћена је *Fleiss*-ова κ, која представља стандардни избор у ситуацијама када више од два анотатора анотира исте јединице података. При интерпретацији добијених вредности коришћене су широко прихваћене смернице које класификују ниво сагласности у опсеге од слабог до готово савршеног слагања. И метрика и смернице интерпретације су детаљно описани у поглављу 2.3.2.

Табела 4 *Fleiss*-ова κ након *PoC* итерација

	CrowdRE (60 реченица по рунди)		**Емоције** (140 реченица по рунди)	
	Намера	**Тема**	**Аспект**	**Емоција**
Прва рунда	0.73	0.44	0.58	0.44
Друга рунда	0.82	0.64	0.82	0.82

Као што је приказано у табели 4, за оба корпуса је било потребно спровести две *PoC* итерације да би се постигла задовољавајућа сагласност за све категорије. За већину таксономија постигнут је готово савршен ниво сагласности ($\kappa \geq 0.81$), док је за таксономију *Тема* постигнут значајан ниво слагања ($\kappa \geq 0.61$). Овако високе вредности *IAA* указују да су анотационе шеме и смернице јасно дефинисане и доследно примењиве, те да креирани корпуси представљају поуздан златни стандард за тренирање и евалуацију модела машинског учења [131].

Након постигнуте сагласности, приступљено је завршној анотацији целокупног корпуса. У овој фази, сви анотатори су независно анотирали преостале јединице података, примењујући финализоване анотационе шеме и смернице. Коначна ознака за сваку јединицу података одређивана је на основу већинског гласања (енгл. *majority vote*), при чему је као финална класа усвајана она ознака коју је доделио највећи број анотатора. Оваквим приступом смањен је утицај индивидуалних интерпретација и додатно повећана стабилност златног стандарда.

У случајевима када је дошло до изједначеног броја додељених ознака, коначну одлуку доносио је аутор дисертације, ослањајући се на дефинисане смернице и резултате дискусија спорних случајева из претходних *PoC* итерација. Оваква улога адјудикатора у складу је са препорученим праксама у анотацији природног језика, посебно у задацима који укључују апстрактне и доменски специфичне категорије.

4.3. Анотационе шеме

У овом поглављу детаљно су представљене анотационе шеме. Анотационе шеме дефинишу скуп категорија, њихове формалне описе и међусобне односе, заједно са пратећим смерницама за анотацију.

Категорије унутар сваке шеме су организоване у таксономије, односно структуриране скупове класа, којима се одређени аспекти текста класификују. Концепт анотационе шеме је методолошки наследник кодних система из квалитативне анализе садржаја (поглавље 2.1).

У складу са истраживачким циљевима ове дисертације, дефинисане су две независне анотационе шеме. Прва шема намењена је анализи корисничких повратних информација у контексту инжењерства захтева заснованог на групи корисника и обухвата таксономије *Намере* и *Теме*. Друга шема усмерена је на детекцију емоција специфичних за процесе учења и обухвата таксономије *Аспекта* и *Емоције*. Обе шеме су развијене итеративно, у складу са *MATTER* методологијом, и финализоване након више циклуса пилот анотације и анализе међуанотаторске сагласности. Категорије су дефинисане тако да буду теоријски утемељене, међусобно разграничене и применљиве на реалне, неструктуриране студентске повратне информације.

4.3.1. Анотациона шема за инжењерство захтева засновано на групи корисника

Анотациона шема за *CrowdRE* осмишљена је тако да омогући систематску анализу студентских повратних информација у контексту унапређења образовног софтвера. У складу са налазима из прегледа литературе (поглавље 3.1), кориснички коментари често истовремено садрже информације о *намери* корисника и *теми* на коју се коментар односи. Из тог разлога, проблем је декомпонован на два задатка: класификацију намере и класификацију теме.

Како би се осигурала релевантност и практична употребљивост дефинисаних категорија за развојни тим, спроведена је квалитативна анализа садржаја у сарадњи са развојним тимом платформе *Clean CaDET*. Као полазна основа коришћена је стандардизована *CrowdRE* таксономија [102], која обухвата шири скуп категорија *Намере* и *Теме*. Индуктивним поступком селекције, из шире таксономије *Намере* преузете су оне категорије које су идентификоване као најрелевантније за домен образовног софтвера и циљеве истраживања. На тај начин задржана је

концептуална усклађеност са постојећим истраживањима, уз истовремено прилагођавање конкретном контексту примене.

За таксономију *Теме*, категорије које се односе на квалитет софтверског производа су прецизиране у складу са стандардом *ISO/IEC 25010* [132], док су категорије које описују контекст производа дефинисане на основу образовног садржаја и педагошке намене платформе *Clean CaDET*.

Оваква стратегија развоја анотационих категорија директно је заснована на комбинованом приступу *QCA* описаном у поглављу 2.1, који обједињује дедуктивно преузимање теоријски утемељених категорија и њихово индуктивно прилагођавање специфичностима анализираног корпуса.

У оквиру таксономије *Намере*, анотатор свакој реченици додељује тачно једну категорију, бирајући ону која најбоље осликава примарну комуникациону сврху реченице:

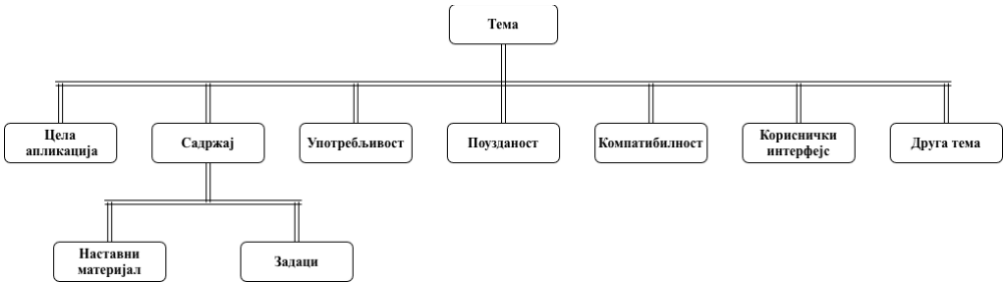
- **Пријава грешке** (енгл. *Bug report*) има за циљ да информише програмера о проблему или дефекту комплетног софтвера или неке његове функционалности.
- **Захтев за функционалношћу** (енгл. *Feature request*) представља захтеве корисника да се додају нове функционалности или садржај у софтвер, или да се уклоне, измене или побољшају постојеће.
- **Оцена** (енгл. *Rating*) има за циљ да корисник изрази општу захвалност или опште незадовољство апликацијом. Фокусира се на опште расуђивање и укључује похвале, омаловажавања и критике.
- **Друга намера** (енгл. *Other intention*) се користи за реченице чија се намера не може разазнати.

Као и у претходном случају, и у оквиру таксономије *Теме* свакој реченици додељује се једна категорија. Специфичност ове таксономије је њена хијерархијска организација, коју илуструје слика 19. Приликом анотације, анотатори су инструисани да, кад год је могуће, бирају најспецифичнију категорију. Категорије вишег нивоа користе се искључиво у случајевима када није могуће извести прецизнији закључак. На пример, категорија **Садржај** се додељује само ако анотатор не може

да разлучи да ли студент реферише на **Наставни материјал** или **Задатке**. Категорије ове таксономије су:

- **Цела апликација** (енгл. *Whole App*) се односи на реченице које посматрају софтвер у целини, без његових функционалности.
- **Наставни материјал** (енгл. *Instructional items*) се односи на наставне материјале доступне у софтверу - инструкционе видее, наговештаје (енгл. *hint*), текстове и илустрације.
- **Задаци** (енгл. *Assessment items*) се односи на елементе за проверу знања - питања са понуђеним одговорима, задатке са превлачењем и изазове за програмирање.
- **Садржај** (енгл. *Content*) се односи на сав садржај доступан кроз софтвер, а који се не може категорисати у једну од претходне две категорије (**Наставни материјал** и **Задаци**).
- **Употребљивост** (енгл. *Usability*) се односи на лакоћу са којом корисник може научити и користити софтвер. Ово укључује интуитивност, ефикасност, лакоћу разумевања, навигације и довршавања задатака.
- **Поузданост** (енгл. *Reliability*) се односи на способност софтвера да функционише исправно и доследно током времена. Ово такође укључује способност софтвера да се опорави од кварова и настави са радом без прекида.
- **Компатибилност** (енгл. *Compatibility*) се односи на могућност софтвера да ради са другим софтверима. Ово омогућава да се подаци и функционалности деле између различитих софтверских апликација и платформи, омогућавајући им да ефикасно комуницирају и размењују информације.
- **Кориснички интерфејс** (енгл. *User Interface - UI*) је тачка интеракције између корисника и софтвера. То је начин на који корисници комуницирају са софтвером и контролишу га, као и начин на који софтвер представља информације и повратне информације кориснику. Кориснички интерфејс укључује графичке елементе, као што су иконе, дугмад и менији, као и команде засноване на тексту које корисник уноси за интеракцију са софтвером.

- **Друга тема** (енгл. *Other topic*) се користи за реченице чију тему није могуће разазнати.



Слика 19 Хијерархијска организација таксономије *Тема*

Представљена анотациона шема праћена је детаљним смерницама са конкретним примерима за сваку категорију. Део ових примера приказан је у табели 5, док су комплетна шема и смернице јавно доступне и документоване у оквиру прилога А ове дисертације.

Табела 5 Примери из анотационе шеме и смерница за *CrowdRE*

Реченица	Намера	Тема
Ne mogu da pređem na sledeci zadatak.	Пријава грешке	Поузданост
Tamna tema je super.	Оцена	Кориснички интерфејс
Moguće automatsko preuzimanje materijala.	Захтев за функционалношћу	Употребљивост

4.3.2. Анотациона шема за емоције усмерене на учење

Дизајн анотационе шеме за овај корпус заснива се на приступу аспектне анализе емоција (енгл. *Aspect-Based Emotion Analysis - ABEA*). За разлику од традиционалне класификације која детектује само присуство и врсту емоције, овај приступ захтева да се прво идентификује аспект (објекат) ка којем је емоција усмерена, а тек потом и сама емоција.

Увођење нивоа аспекта мотивисано је потребом за већом прецизношћу и употребном вредношћу аутоматске анализе у контексту *ITS*-а. Сама детекција негативне емоције, попут фрустрације, није довољна за адекватан одговор система. На пример, фрустрација изазвана техничким неисправностима софтвера захтева инжењерску интервенцију (исправку кода), док фрустрација изазвана тежином градива захтева педагошку

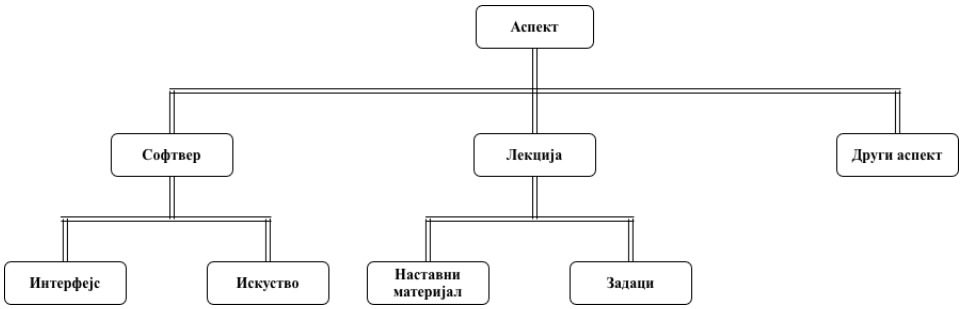
интервенцију (пружање помоћи или прилагођавање садржаја). Због тога су дефинисане две таксономије: *Аспект* и *Емоција*.

За разлику од *CrowdRE*, у домену детекције емоција усмерених на учење не постоји широко прихваћена и стандардизована таксономија аспеката на које се емоционалне реакције студената односе. Из тог разлога, аспекти у овом истраживању нису преузети из постојећих таксономија, већ су развијени посебно за анализирани домен индуктивним приступом *QCA*, описаним у поглављу 2.1.

Таксономија *Аспекта* организована је хијерархијски, као што илуструје слика 20. Приликом анотације, анотатори су инструисани да свакој реченици додељују тачно једну, најспецифичнију категорију. Категорије вишег нивоа користе се искључиво у случајевима када није могуће извести прецизнији закључак. На пример, категорија **Софтвер** се додељује само ако анотатор не може да разлучи да ли студент реферише на **Интерфејс** или **Искуство**. Категорије ове таксономије су:

- **Интерфејс** (енгл. *Interface*) се односи на начин на који корисник комуницира са софтверском апликацијом и како се креће кроз њу. Укључује изглед, дугмад, текст, слике и друге визуелне елементе који чине интерфејс. Кориснички интерфејс је дизајниран да олакша корисницима да остваре своје задатке и пронађу информације које су им потребне.
- **Искуство** (енгл. *Experience*) се односи на целокупно искуство које корисник има приликом интеракције са софтвером. Укључује факторе као што су једноставност коришћења, ефикасност и опште задовољство.
- **Софтвер** (енгл. *Software*) се додељује реченици која се односи на софтвер, али се не може категорисати у једну од претходне две категорије.
- **Наставни материјал** (енгл. *Instructional items*) се односи на доступност и квалитет наставних материјала - инструкционих видеа, текстова и илустрација.
- **Задаци** (енгл. *Assessment items*) се односи на доступност и квалитет елемената за проверу знања - питања са понуђеним одговорима, задатака са превлачењем и изазова за програмирање.

- **Лекција** (енгл. *Lesson*) се додељује реченици која се односи на лекцију, али се не може категорисати у једну од претходне две категорије.
- **Други аспект** (енгл. *Other aspect*) се односи на реченице чији аспект није могуће разазнати.



Слика 20 Хијерархијска организација таксономије *Аспект*

Као и у претходном случају, и у оквиру таксономије *Емоција* свакој реченици додељује се тачно једна категорија. Категорије ове таксономије представљају емоције усмерене на учење, представљене у поглављу 2.2. У односу на изворни модел, укључена је категорија **Неутрално**, како би се избегло принудно сврставање реченица које не одговарају ниједном од карактеристичних афективних стања:

- **Досада** (енгл. *Boredom*) је негативно емоционално стање које карактерише осећање апатије и незаинтересованости за задатак, и често је праћено недостатком пажње, ниском мотивацијом и општим осећајем неангажованости. Може се препознати употребом кључних речи или фраза: *досадно*, *монотоно*, *незанимљиво*. Одликује је недостатак ентузијазма у писању.
- **Збуњеност** (енгл. *Confusion*) је афективно стање које карактерише осећај неизвесности, двосмислености и тешкоћа да се разуме информација или ситуација. Може се препознати употребом кључних речи или фраза: *збуњен*, *несигуран*, *изгубљен*. Одликује се употребом језика који указује на недостатак јасноће, као што је изражавање сумње у разумевање или постављање питања која захтевају додатна појашњавања.
- **Фрустрација** (енгл. *Frustration*) је негативан емоционалан одговор на препреке или сметње у постизању циљева. То је стање

несигурности и незадовољства које произилази из нерешених проблема или неиспуњених потреба. Може се препознати употребом кључних речи или фраза: *фрустриран*, *изнервиран*, *обесхрабрен*. Одликује се употребом оштрог језика.

- **Задовољство** (енгл. *Delight*) или одушевљење је позитиван емоционалан одговор на аспекте окружења за учење, као што су занимљив и пријатан садржај или успешно извршење задатка. Често је резултат доживљавања или предвиђања пожељног или пријатног догађаја, или задовољења личне потребе или циља. Карактеришу га осећања радости и усхићења. Може се препознати употребом кључних речи или фраза: *одушевљен*, *задовољан*, *узбуђен*. Одликује се употребом позитивног језика, ентузијастичним тоном или енергичним стилом писања.
- **Изненађеност** (енгл. *Surprise*) је осећај неочекиваности или чуђења у погледу наставног материјала или задатка. То је изненадно откриће нечег неочекиваног и карактерисано је осећајем неочекиваности или шока. Може се препознати употребом кључних речи или фраза: *изненађен*, *зачуђен*, *шокиран*, *неочекиван*. Одликује се употребом узвичног језика, повећањем енергије или узбуђења у писању или изненадна промена тона.
- **Ангажованост** (енгл. *Engagement*) је стање укључености или интересовања за материјал или задатак за учење. Карактерише се осећањем усредсређености, посвећености и потпуног урањања и фокусирања на предмет учења. Бројни фактори, укључујући личну мотивацију, потешкоће задатка и присуство ометања, утичу на ангажовање. Може се препознати употребом кључних речи или фраза: *заинтересован*, *фокусиран*, *укључен*, *очаран*, *фасциниран*. Одликује се доследним стилом писања, употребом описног језика или позитивног тона.
- **Неутрално** (енгл. *Neutral*) се односи на реченице на основу чијег садржаја се није могла утврдити ниједна од горе наведених емоција, или се иста није могла са сигурношћу утврдити.

Представљена анотациона шема праћена је детаљним смерницама са конкретним примерима за сваку категорију. Део ових примера садржи

табела 6, док су комплетна шема и смернице јавно доступне и документоване у оквиру прилога Б ове дисертације.

Табела 6 Примери из анотационе шеме и смерница за емоције усмерене на учење

Реченица	Аспект	Емоција
Sjajno iskustvo!	Искуство	Задовољство
Zadaci su zanimljivi.	Задаци	Ангажованост
Tutor me uspavljuje.	Софтвер	Досада
Izbor boja me izluđuje	Интерфејс	Фрустрација
Izgubljen sam u ovoj lekciji	Лекција	Збуњеност
Nisam očekivao ovakvo pitanje!	Задаци	Изненађеност

4.4. Карактеристике корпуса

Из укупно 1414 информативних коментара изведено је 1850 реченица просечне дужине 9 речи. Лексичка структура корпуса је разнолика и обухвата комбинацију стандардног српског језика, стручне терминологије из домена софтверског инжењерства, као и неформалних израза типичних за дигиталну комуникацију. Присуство правописних грешака, изостављање дијакритика и употреба колоквијалних конструкција додатно потврђују аутентичност и валидност прикупљених података.

Дескриптивна статистика оба корпуса, приказана у табели 7, указује на изражене разлике у језичкој структури између два посматрана домена.

Табела 7 Дескриптивна статистика корпуса

Корпус	Број коментара	Број реченица	Број речи	Број речи у реченици	Просечан број речи у реченици
CrowdRE	396	591	9236	[1, 66]	16
Емоције	1018	1259	7727	[1, 47]	6

Реченице у *CrowdRE* корпусу су знатно комплексније, са просечном дужином од 16 речи. Оваква структура је очекивана, будући да пријаве софтверских проблема и предлози нових функционалности подразумевају детаљније описе, често уз навођење техничког контекста или конкретних корака. Насупрот томе, реченице у корпусу емоција су

краће и језгровитије, са просечном дужином од 6 речи, што је у складу са природом афективног изражавања које је често импулсивно и директно.

Ове разлике у дужини и структури реченица представљају важан фактор при избору и подешавању *NLP* модела. Дуже и технички богатије реченице захтевају моделе са већим контекстуалним капацитетом, док кратке емотивне реченице захтевају моделе осетљиве на суптилне лексичке сигнале.

Табела 8 садржи дистрибуцију категорија унутар *CrowdRE* корпуса. Анализа *Намера* указује на изразиту доминацију категорије **Захтев за функционалношћу**, која обухвата више од половине укупног корпуса (56.2%). Ово указује на то да су студенти, као будући инжењери, примарно фокусирани на конструктивно унапређење софтвера, предлажући нове функционалности и побољшања постојећих. С друге стране, **Пријаве грешке** и **Оцене** су готово равномерно заступљене (22.5% и 20.8%).

Када је реч о дистрибуцији *Тема*, запажа се знатно равномернија расподела међу водећим категоријама. Највећи број коментара усмерен је на **Кориснички интерфејс** (20.3%) и **Употребљивост** (19.3%), које у стопу прате педагошки аспекти попут **Задатака** (18.1%) и **Наставног материјала** (15.9%).

Табела 8 Дистрибуција категорија *CrowdRE* корпуса

Намера	**Број анотација**	**% анотација**
Захтев за функционалношћу	332	56.2
Пријава грешке	133	22.5
Оцена	123	20.8
Друга намера	3	0.5
Тема	**Број анотација**	**% анотација**
Кориснички интерфејс	120	20.3
Употребљивост	114	19.3
Задаци	107	18.1
Наставни материјал	94	15.9
Цела апликација	56	9.5
Поузданост	56	9.5
Компатибилност	28	4.7
Садржај	12	2.0
Друга тема	4	0.7

Насупрот томе, категорије **Друга намера** (0.5%) и **Друга тема** (0.7%) су готово занемарљиве. С обзиром на то да изузетно мали број узорака онемогућава стабилну обуку и евалуацију модела, а саме класе не носе информативну вредност за развојни тим, оне су искључене из даљег експерименталног рада.

Значајан увид пружа укрштена анализа између *Намера* и *Тема*. Како подаци показују, већина захтева за новим функционалностима односила се на категорије **Употребљивост** (28.6%) и **Кориснички интерфејс** (22.6%). Ово сугерише да студенти највећи простор за унапређење виде у интеракцији са платформом. Насупрот томе, када је реч о пријавама грешака, оне су примарно повезане са **Поузданошћу** (41.4%), што је и очекивано јер се багови директно одражавају на стабилност рада. Занимљиво је приметити да су оцене (било позитивне или негативне) најчешће биле усмерене ка **Целој апликацији** (39%), што указује на тенденцију студената да дају холистички суд о платформи, пре него о појединачним компонентама.

Дистрибуција у корпусу емоција усмерених на учење, коју приказује табела 9, открива другачије обрасце. У погледу *Аспеката*, апсолутну доминацију има категорија **Искуство** (51.6%), што показује да студенти емоције најчешће везују за свој субјективни доживљај коришћена софтвера, пре него за конкретне техничке елементе попут **Интерфејса** (6.4%) или **Софтвера** (4.9%).

Када је реч о самим *Емоцијама*, најзаступљенија је **Ангажованост** (35.1%), праћена **Задовољством** (29.8%). Овакав резултат је охрабрујући са педагошког аспекта, јер указује на то да је платформа *Clean CaDET* успела да одржи пажњу студената. Међутим, присуство **Фрустрације** у 16.8% случајева не сме се занемарити.

Дубља анализа односа између *Аспеката* и *Емоција* открива кључне зависности. **Ангажованост** је доминантна емоција у готово свим аспектима везаним за садржај учења, што указује на квалитет образовног материјала. **Задовољство** је најизраженије код аспекта **Искуство** (40.1%), што сугерише да је општи утисак о платформи позитиван. **Фрустрација** је, међутим, доминантна емоција везана за аспект **Интерфејс** (88.8%). Овај податак је од критичне важности за развојни тим, јер прецизно

лоцира извор незадовољства корисника у начину интеракције са платформом, а не у наставном садржају.

Табела 9 Дистрибуција категорија корпуса емоција усмерених на учење

Аспект	**Број анотација**	**% анотација**
Искуство	650	51.6
Задаци	257	20.4
Лекција	121	9.6
Наставни материјал	89	7.1
Интерфејс	80	6.4
Софтвер	62	4.9
Емоција	**Број анотација**	**% анотација**
Ангажованост	442	35.1
Задовољство	375	29.8
Фрустрација	212	16.8
Неутрално	117	9.3
Збуњеност	52	4.1
Досада	51	4.1
Изненађеност	10	0.8

4.5. Ограничења корпуса

Иако креирани корпуси представљају значајан допринос као први ресурс ове врсте за српски језик, важно је истаћи ограничења која проистичу из контекста њиховог настанка и природе прикупљених података. Идентификација ових ограничења је кључна за правилно тумачење добијених резултата и процену њихове опште применљивости.

Пре свега, корпуси су прикупљени у оквиру једне образовне платформе, на једном предмету и у оквиру једне високошколске институције. Оваква доменска и контекстуална ограниченост подразумева да добијени налази одражавају специфичности конкретног образовног окружења и циљне групе студената. Иако то може ограничити директну генерализацију резултата на друге домене или системе, истовремено обезбеђује висок степен валидности података, будући да они потичу из реалног наставног процеса.

Друго ограничење односи се на изражену неравнотежу класа у оба посматрана домена. Као што је приказано у претходном поглављу, **Захтев**

за функционалношћу у *CrowdRE* корпусу или **Ангажованост** у корпусу емоција доминира скупом података. Са друге стране, класе од изузетног педагошког значаја, као што су **Збуњеност** или **Досада**, релативно слабо су заступљене. Оваква дистрибуција представља значајан изазов за алгоритме машинског учења, јер стандардни модели имају тенденцију фаворизовања већинских класа, што може довести до слабијег препознавања мањинских. Ово ограничење директно мотивише примену техника балансирања и аугментације података у експерименталном делу рада.

Коначно, иако величина корпуса од 1850 реченица омогућава тренирање робусних модела уз примену трансфера учења, она је и даље релативно ограничена у поређењу са ресурсима доступним за језике са богатим ресурсима, попут енглеског. Ограничен обим података може утицати на способност модела да генерализује ретке језичке конструкције или да препозна суптилне варијације у изражавању емоција, што додатно оправдава избор напредних архитектура и стратегија за ефикасно искоришћење доступних података.

5. Развој модела за анализу студентских повратних информација

У овом поглављу описан је процес развоја и архитектура модела машинског учења намењених аутоматској анализи студентских повратних информација на српском језику. Главни циљ овог поглавља је да представи примењену методологију, архитектуру одабраних модела и експерименталну поставку за решавање идентификованих истраживачких задатака. Тиме се поставља основа за евалуацију и анализу резултата која следи у наредном делу дисертације.

5.1. Методологија

Методолошки оквир развоја модела заснован је на четири задатка вишекласне класификације текста. Сви модели су пројектовани за класификацију на нивоу појединачне реченице, што је у потпуности усклађено са процесом анотације и структуром креираног корпуса златног стандарда. Задаци класификације су груписани према два кључна домена истраживања.

CrowdRE домен, који се односи на процес прикупљања корисничких захтева кроз анализу повратних информација, обухвата задатке:

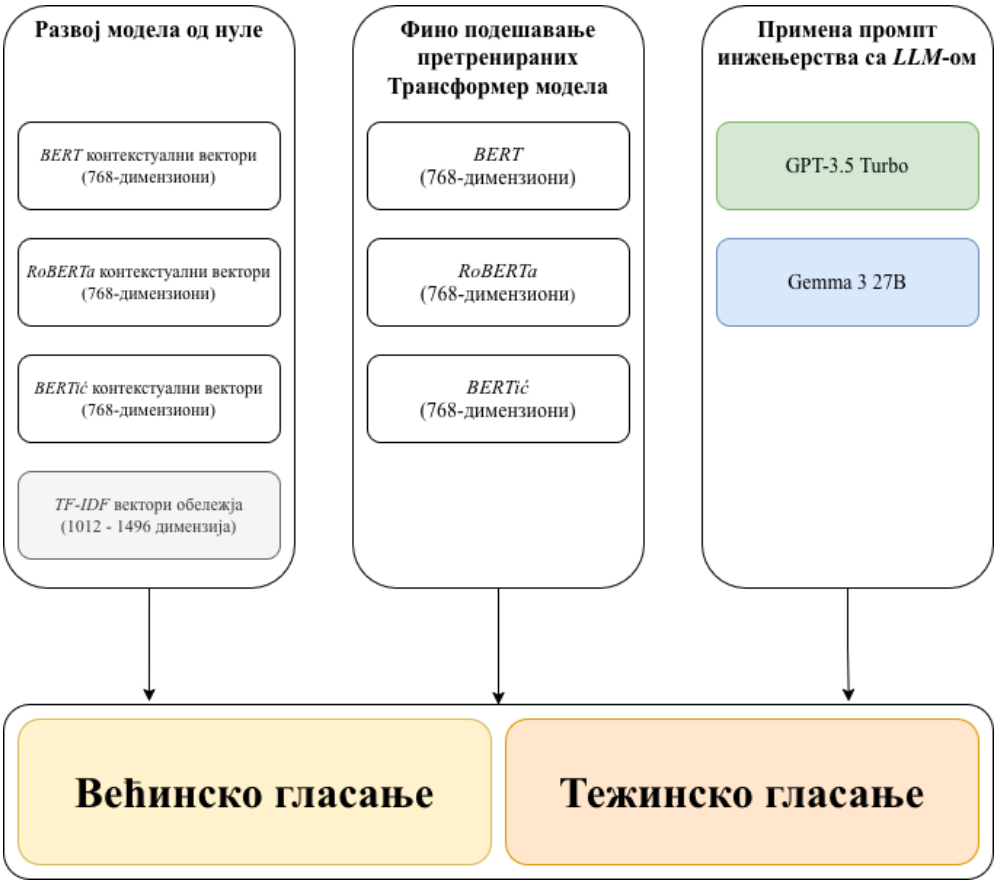
- Класификација **намере** (енгл. *Intention*) за идентификацију сврхе реченице, попут пријаве грешке у софтверу, предлога за унапређење или похвале.
- Класификација **теме** (енгл. *Topic*) за препознавање ширег контекста на који се намера односи.

Домен емоција, који је усмерен на разумевање афективних стања студената током интеракције са *ITS*-ом, обухвата задатке:

- Класификација **аспекта** (енгл. *Aspect*) за идентификацију специфичних елемената *ITS*-а или корисничког искуства.
- Класификација **емоције** (енгл. *Emotion*) за препознавање афективног стања које се односи на идентификовани аспект.

Како би се идентификовао најефикаснији приступ класификацији, процес моделовања је спроведен кроз четири различита приступа на сва четири

задатка: (1) развој модела од нуле, (2) фино подешавање претренираних *Transformer* модела, (3) примена промпт инжењерства са *LLM*-овима и (4) креирање ансамбл модела кроз интеграцију најбољих појединачних модела. Ове приступе илуструје слика 21.



Слика 21 Преглед коришћених модела машинског учења

Овакав избор четири приступа омогућава систематско поређење три различите парадигме обраде природног језика: статистичких метода, претренираних дубоких архитектура и приступа заснованих на великим језичким моделима. Као четврти, интегрисани слој користи се ансамбл који комбинује предности претходних метода.

5.2. Развој модела од нуле

Први приступ развоју модела подразумева изградњу класификатора од нуле коришћењем статистичких *ML* модела како би се успоставила

основна линија (енгл. *baseline*) за поређење са сложенијим архитектурама. За трансформацију реченица у векторе обележја примењена су два различита поступка ради систематског поређења ефеката статистичких и контекстуалних репрезентација текста.

Први поступак се ослања на методолошки оквир за обраду текста представљен у одељку 2.4.1. Претпроцесирање текста (представљено у поглављу 2.4.2) обухвата токенизацију реченица на појединачне речи коришћењем *NLTK* библиотеке [49]. Из речника су затим уклоњене неважне речи за српски језик [133], након чега је примењен алгоритам за кореновање специфичан за српски језик [127]. Коначна репрезентација сваке реченице добијена је применом *TF-IDF* модела (описан у одељку 2.4.3), чиме су формирани вектори обележја чије су димензије варирале од 1012 до 1496, зависно од специфичног класификационог задатка.

Други поступак обухвата екстракцију фиксираних контекстуалних вектора из претренираних *Transformer* модела, представљених у поглављу 2.11. За разлику од претходног поступка, овде се прослеђује текст без претпроцесирања, а коначна репрезентација реченице добија се усредњавањем излазних вектора свих токена из последњег скривеног слоја модела (енгл. *mean pooling*). На овај начин добијају се 768-димензиони вектори употребом *Transformers* библиотеке [134]. Овакав приступ омогућава статистичким моделима машинског учења да искористе богатије семантичке информације садржане у контекстуалним репрезентацијама.

На основу описаних репрезентација, извршена је обука неколико статистичких *ML* модела чије су теоријске основе детаљно изложене у поглављу 2.7. За потребе овог приступа коришћени су модел случајних шума, метод потпорних вектора, проста агрегација са *SVM*-ом као основним класификатором, као и екстремно градијентно појачавање. За *XGBoost* употребљена је његова примарна *Python* имплементација [68], док је за остале класификаторе и *TF-IDF* коришћена *scikit-learn* библиотека [135].

5.3. Фино подешавање претренираних Transformer модела

Други приступ развоју модела заснива се на парадигми трансфера учења кроз фино подешавање претренираних *Transformer* модела, чије су теоријске основе и архитектуре детаљно изложене у поглављу 2.11. За потребе овог истраживања одабрана су три различита приступа претренираним моделима, како би се испитао утицај различитих стратегија претренирања на прецизност класификације студентских повратних информација: *mBERT, RoBERTa* и *BERTić*. *mBERT* је коришћен у две варијанте: оној која разликује величину слова (енгл. *cased*) и оној која све карактере своди на мала слова (енгл. *uncased*), како би се испитао утицај морфолошке нормализације кроз величину слова на класификацију текста на српском језику. Када је реч о *RoBERTa* моделу, у експериментима је наменски коришћена њена вишејезична варијанта *XLM-RoBERTa*, како би се осигурала адекватна подршка за обраду српског језика. Док *mBERT* и *RoBERTa* обезбеђују широку лингвистичку покривеност кроз претренирање на великом броју језика, *BERTić* је одабран због своје специјализације за *BCMS* макројезик. Ова специјализација је од кључног значаја за препознавање морфолошких и синтаксичких нијанси у корпусу на српском језику.

Процес финог подешавања реализован је употребом *Transformers* библиотеке [134]. За разлику од претходно описаног приступа моделовања, ови модели примају сиров текст без претходног претпроцесирања, при чему је сваки модел применио сопствену токенизацију и специјалне токене у складу са својом архитектуром. На овај начин омогућава се потпуно коришћење семантичког контекста и лингвистичког предзнања модела за решавање четири дефинисана класификациона задатка.

5.4. Примена промпт инжењерства са великим језичким моделима

Трећи приступ моделовању обухвата примену *LLM*-ова кроз методе промпт инжењерства, чије су теоријске основе представљене у одељку 2.12.3. У оквиру овог приступа коришћени су претренирани *GPT-3.5 Turbo* модел [136] и *Gemma 3* модел са 27 милијарди параметара. Архитектура *GPT-3.5 Turbo* модела није званично документована, али се

ослања на исте архитектонске принципе представљене у одељку 2.12.1, док је архитектура *Gemma 3* модела представљена у одељку 2.12.2.

Овим моделима се приступало путем програмског интерфејса (енгл. *Application Programming Interface - API*). За *GPT-3.5 Turbo* коришћен је јавни *API* сервиса компаније *OpenAI*, док је *Gemma 3* модел покренут у локалном окружењу коришћењем *Ollama* платформе[3], која је омогућила приступ моделу путем локалног *API*-ја унутар затворене инфраструктуре. За разлику од претходна два приступа који се ослањају на потпуно или делимично тренирање модела, овде је коришћена *few-shot* стратегија, где се понашање модела усмерава кроз пажљиво структуриране инструкције унутар самог промпта.

За сваки од четири класификациона задатка креирани су специфични промптови на српском језику. Структура промпта, коју илуструје табела 10, дизајнирана је тако да, поред општег описа задатка и дефиниција могућих лабела, садржи и репрезентативне примере преузете директно из финалних анотационих инструкција које су користили и људски анотатори. На тај начин је моделу пружен контекст неопходан за разумевање специфичних нијанси у студентским повратним информацијама.

Табела 10 Структура промпта који се користи за класификацију реченица

Želim da klasifikuješ tekst za mene. Pogledaj ispod sve moguće labele i njihove opise. Prilikom klasifikacije mi vrati samo broj koji predstavlja labelu, bez pratećeg teksta. { Labele sa pripadajućim opisom } U nastavku pogledaj nekoliko primera: { Primeri sa pripadajućom labelom } Evo teksta koji treba klasifikovati: { Tekst }

За сваку реченицу из тест скупа података генерисан је посебан промпт и прослеђен моделу на обраду. Након одговора модела, прикупљене су

[3] https://ollama.com/

лабеле ради даље евалуације и поређења са осталим приступима. Овај приступ је омогућио испитивање капацитета савремених *LLM*-ова да реше комплексне задатке вишекласне класификације на српском језику без потребе за додатним финим подешавањем.

5.5. Креирање ансамбл модела

Четврти приступ моделовању заснива се на изградњи ансамбл модела са циљем експлоатације специфичних предности сваког од претходна три приступа. Кроз интеграцију модела различитих архитектура тежи се искоришћавању грешака које праве модели различите структуре, чиме се остварује смањење варијансе у предвиђањима и идентификација поузданијих граница између класа.

Креирање је реализовано кроз два концептуално различита модела гласања, чије су теоријске основе представљене у поглављу 2.8. У оба случаја увршћени су најбоље рангирани појединачни модели из сваког од три приступа. Први модел се ослања на већинско гласање, где се коначна лабела доноси на основу консензуса већине укључених модела, чиме се минимизује утицај случајних грешака специфичних за изоловане приступе. У случајевима изједначеног броја гласова, коначна одлука је донесена на основу предвиђања најуспешнијег појединачног модела. Други приступ обухвата тежинско гласање, у којем је сваком моделу додељен специфичан коефицијент значаја сразмеран његовој релативној перформанси на тренинг скупу, чиме се избегава ризик од цурења информација из тест скупа (енгл. *data leakage*). На овај начин је моделима са доказаном већом поузданошћу омогућен пресудан утицај на коначну одлуку.

5.6. Експериментални поступак

Експериментални поступак је дизајниран тако да обезбеди поуздану меру генерализације развијених модела на новим подацима. Целокупан процес је реализован коришћењем програмског језика *Python*, уз употребу библиотека за машинско учење (*scikit-learn*), дубоко учење (*Transformers*) и обраду текста (*NLTK*). За сваки од четири задатка класификације примењена је *out-of-sample bootstrap* валидација, чије су теоријске основе

изложене у одељку 2.6.2. Ово обухвата генерисање 10 независних и стратификованих подела података на тренинг и тест скуп у односу 80:20. Овакав приступ омогућава да се резултати модела, као и људске перформансе, прикажу кроз средњу вредност и стандардну девијацију перформанси на свих 10 тест скупова.

Приликом развоја модела од нуле, оптимизација хиперпараметара је извршена путем Бајесове оптимизације (представљене у одељку 2.6.3) у оквиру *KerasTuner* библиотеке [137] кроз 40 итерација на тренинг скупу. С обзиром на изражену неравнотежу класа унутар корпуса, простор за претрагу је проширен техникама аугментације података у простору атрибута: *SMOTE* за генерисање нових примера мањинских класа, *NCR* за уклањање инстанци већинских класа које су блиске инстанцама мањинских класа, и *SMOTEENN* као хибридни приступ за генерисање нових примера мањинских класа и уклањање инстанци већинских класа. Ови алгоритми су представљени у поглављу 2.5, а примењени су уз помоћ *imbalanced-learn* библиотеке [138]. Свака техника је имала по 10 итерација у оквиру процеса Бајесове оптимизације.

Са друге стране, фино подешавање *PTM* спроведено је кроз четири епохе уз коришћење *AdamW* оптимизатора [139]. За ове моделе је примењена оптимизација хиперпараметара величине групе података (енгл. *batch size*) и стопе учења (енгл. *learning rate*) мрежном претрагом, пратећи препоруке аутора *BERT* модела [140]. Проблем небалансираних класа адресиран је кроз аугментацију у простору података коришћењем замене и уметања речи (описано у поглављу 2.5) путем *nlpaug* библиотеке [141]. За контекстуалну аугментацију коришћен је базни модел који одговара моделу који се фино подешава, чиме се осигурава да аугментоване реченице деле исти токенизатор и вокабулар са моделом који се фино подешава. На пример, приликом финог подешавања *BERTić*-а, његов базни модел је коришћен у *nlpaug*-у за генерисање аугментованих реченица.

Сви експерименти спроведени су на преносивом рачунару *MacBook Pro* са *Apple M3 Pro* чипом, 36 *GB* радне меморије и оперативним системом *macOS Sequoia*, чиме је обезбеђено да све измерене перформансе одговарају широко доступном потрошачком хардверу. Статистички

модели су обучавани на процесору, док је фино подешавање *PTM* и локално покретање *Gemma 3* модела извршено коришћењем *Apple Metal Performance Shaders* окружења. Ради процене рачунарске ефикасности приступа, времена обуке и предикције измерена су на једној репрезентативној подели података, а добијени резултати анализирани су у поглављу 6.5.

Евалуација свих приступа извршена је применом макро-усредњене (енгл. *macro-averaged*) прецизности, одзива и *F1*-мере, при чему је *F1*-мера одабрана као примарна метрика за процесе оптимизације. Ове метрике су представљене у одељку 2.6.2, а за њихово израчунавање је коришћена *scikit-learn* библиотека. Како би се утврдила статистичка значајност разлика у перформансама између испитиваних модела, као и у односу на људски учинак, примењен је Велчов *t*-тест (енгл. *Welch's t-test*) [142], реализован помоћу *SciPy* библиотеке [143]. Овај тест представља модификацију стандардног *t*-теста за поређење средњих вредности два узорка, при чему не претпоставља једнакост варијанси између узорака, што га чини погодним за поређење различитих модела који испољавају различите нивое варијабилности у перформансама. Као референтна вредност за успех модела дефинисане су људске перформансе, израчунате као просек резултата појединачних анотатора у односу на коначну (адјудиковану) лабелу на истом тест скупу који је коришћен за моделе. Експериментални поступак је подељен у три фазе: (1) примарно поређење перформанси различитих приступа моделовања, (2) испитивање утицаја стратегија балансирања класа на резултате и (3) завршна анализа грешака најбољих модела ради идентификације лингвистичких и контекстуалних изазова специфичних за српски језик.

6. Експериментални резултати и дискусија

У овом поглављу су представљени резултати евалуације модела за четири задатка класификације у оквиру *CrowdRE* домена и домена емоција усмерених на учење. Анализа обухвата компаративни приказ четири методолошка приступа у односу на референтни људски учинак. Поред квантитативног приказа, поглавље доноси статистичку верификацију хипотеза, као и квалитативну анализу грешака. На крају, дискутују се доприноси истраживања, могућност примене модела, као и ограничења истраживања.

6.1. Резултати у домену инжењерства захтева заснованог на групи корисника

У оквиру овог поглавља представљени су резултати задатака класификације намера и тема, који се односе на процес прикупљања корисничких захтева у оквиру *CrowdRE* домена. Главни циљ спровођења ових експеримената је био да се испита у којој мери аутоматизовани приступи могу успешно идентификовати сврху студентских коментара и специфичне аспекте софтвера на које се ти коментари односе.

Табела 11 приказује компаративни преглед макро-просечних вредности *F1*-мере за све испитане методолошке приступе у односу на референтни људски учинак. Резултати указују на значајне разлике у перформансама испитиваних методолошких приступа. У задатку класификације намера, фино подешени *PTM* ојачани техникама аугментације података, значајно су надмашили статистичке моделе. Посебно се истиче модел *BERTić* који се приближио људском учинку. Оваква супериорност *PTM* је у складу са скорашњим истраживањима у доменима *CrowdRE* [106] [107]. Вреди напоменути да је *RoBERTa* модел у основној фази финог подешавања остварио значајно слабије резултате у поређењу са *mBERT* и *BERTić* моделима. Услед ових слабих перформанси у иницијалним експериментима, овај модел није био укључен у наредну фазу експеримената са аугментацијом података.

Интересантан увид пружа евалуација *LLM*-ова. Док је *GPT-3.5 Turbo* остварио скромније резултате, модел *Gemma 3* је остварио изванредан резултат у класификацији намера, чиме се изједначио са просечним

људским учинком. Међутим, у задатку класификације тема, сви појединачни модели су остварили резултате приметно испод људске референце. Ово се може објаснити специфичном комплексношћу овог задатка. Овај задатак карактерише највећи број класа и најнижа *IAA*, што указује на објективне потешкоће у семантичком разграничавању аспеката софтвера.

Табела 11 Просечне макро *F1*-мере за задатке класификације намера и тема кроз 10 независних и стратификованих подела података. Подебљане вредности означавају најбољи учинак унутар приступа, док је свеукупно најбољи резултат додатно подвучен

Методолошки приступ	Модел	Намера	Тема
Људски учинак		0.84 ± 0.09	0.75 ± 0.13
Развој од нуле	*TF-IDF* вектори	0.71 ± 0.03	**0.51 ± 0.06**
	mBERT cased вектори	0.64 ± 0.03	0.45 ± 0.06
	mBERT uncased вектори	**0.73 ± 0.04**	0.46 ± 0.07
	RoBERTa вектори	0.68 ± 0.06	0.34 ± 0.04
	BERTić вектори	0.69 ± 0.05	0.35 ± 0.07
Фино подешавање *PTM*	*mBERT cased*	0.76 ± 0.04	0.38 ± 0.08
	mBERT uncased	0.77 ± 0.04	0.47 ± 0.09
	RoBERTa	0.63 ± 0.21	0.22 ± 0.1
	BERTić	0.79 ± 0.19	0.29 ± 0.16
	mBERT cased аугментован заменом	0.77 ± 0.04	0.4 ± 0.13
	mBERT uncased аугментован заменом	0.78 ± 0.04	0.38 ± 0.14
	BERTić аугментован заменом	**0.83 ± 0.04**	0.35 ± 0.16
	mBERT cased аугментован уметањем	0.77 ± 0.03	0.38 ± 0.13
	mBERT uncased аугментован уметањем	0.81 ± 0.03	**0.52 ± 0.09**
	BERTić аугментован уметањем	0.83 ± 0.05	0.48 ± 0.05
Промпт инжењерство	*GPT-3.5 Turbo*	0.68 ± 0.03	0.33 ± 0.05
	Gemma 3 27B	**0.84 ± 0.04**	**0.56 ± 0.04**
Ансамбл	Већинско гласање	0.82 ± 0.03	0.56 ± 0.08
	Тежински ансамбл	**0.84 ± 0.03**	**0.59 ± 0.06**

Најбоље укупне перформансе остварио је предложени тежински ансамбл. Код класификације намера, овај модел је постигао макро *F1*-меру од 0.84, што је идентично људском учинку, али уз знатно мању стандардну девијацију ($\sigma_{\text{модел}} = 0.03$ у односу на $\sigma_{\text{људи}} = 0.09$). Ово сугерише да је ансамбл конзистентнији од људских анотатора. У изазовнијем задатку класификације тема, ансамбл је постигао $F1 = 0.59 \pm 0.06$, што је најбољи аутоматизовани резултат у овом истраживању и потврда да интеграција различитих архитектура успешно компензује појединачне недостатке модела.

Како би се стекао дубљи увид у понашање најуспешнијег модела у односу на специфичне класе унутар *CrowdRE* домена, извршена је декомпозиција резултата по категоријама. Табела 12 садржи детаљне податке о прецизности, одзиву и *F1*-мери за тежински ансамбл, упоредо са вредностима људског учинка за сваку класу.

Детаљна анализа перформанси по класама указује на неколико кључних карактеристика тежинског ансамбла у односу на референтни људски учинак. Први значајан увид односи се на изражену конзистентност модела, што се огледа у знатно нижим вредностима стандардне девијације кроз готово све категорије. Овај тренд је посебно видљив код класа са високим степеном субјективности, као што је **Употребљивост**, где је варијанса резултата модела знатно нижа у поређењу са људима ($\sigma_{\text{модел}} = 0.05$ наспрам $\sigma_{\text{људи}} = 0.13$). Ово указује на то да су људски анотатори склони различитим субјективним проценама истог садржаја, док редукована варијанса модела сугерише његову већу стабилност и поузданост при аутоматизованој обради великих количина података.

Други кључни увид односи се на способност модела да идентификује специфичне техничке аспекте софтвера. У оквиру класе **Поузданост**, тежински ансамбл остварује већи одзив ($R = 0.85$) у односу на људски учинак ($R = 0.83$), што је резултат од великог значаја за практичну примену у системима за прикупљање захтева где је идентификација критичних грешака приоритет. Посебно се издвајају резултати за класу **Садржај**, где је тежински ансамбл са $F1 = 0.64 \pm 0.11$ значајно надмашио људски учинак ($F1 = 0.43 \pm 0.33$). С обзиром на то да ову класу карактерише мали број узорака и врло висока варијанса код људи,

супериорност модела у овом сегменту може се приписати ефикасности примењене аугментације података и ансамбл методе. Ове технике су омогућиле моделу да успешно идентификује лингвистичке обрасце чак и код мањинских класа, чиме је потврђена робусност модела у условима изражене небалансираности података.

Табела 12 Упоредни приказ прецизности (*P*), одзива (*R*) и *F1*-мере по појединачним класама за људски учинак и најбољи модел у *CrowdRE* домену

Намера	**Метод**	***P***	***R***	***F1***
Захтев за функционалношћу	Људски учинак	0.95 ± 0.06	0.96 ± 0.04	0.95 ± 0.03
	Тежински ансамбл	0.9 ± 0.03	0.89 ± 0.03	0.89 ± 0.03
Пријава грешке	Људски учинак	0.98 ± 0.02	0.91 ± 0.07	0.94 ± 0.04
	Тежински ансамбл	0.88 ± 0.03	0.88 ± 0.02	0.88 ± 0.03
Оцена	Људски учинак	0.88 ± 0.11	0.88 ± 0.14	0.87 ± 0.09
	Тежински ансамбл	0.9 ± 0.03	0.86 ± 0.06	0.87 ± 0.04
Тема	**Метод**	***P***	***R***	***F1***
Кориснички интерфејс	Људски учинак	0.87 ± 0.08	0.83 ± 0.12	0.84 ± 0.09
	Тежински ансамбл	0.82 ± 0.02	0.76 ± 0.03	0.78 ± 0.03
Употребљивост	Људски учинак	0.8 ± 0.13	0.78 ± 0.15	0.78 ± 0.13
	Тежински ансамбл	0.78 ± 0.07	0.73 ± 0.04	0.75 ± 0.05
Задаци	Људски учинак	0.78 ± 0.12	0.88 ± 0.07	0.82 ± 0.09
	Тежински ансамбл	0.73 ± 0.03	0.79 ± 0.04	0.75 ± 0.03
Наставни материјал	Људски учинак	0.88 ± 0.06	0.87 ± 0.08	0.87 ± 0.07
	Тежински ансамбл	0.82 ± 0.06	0.81 ± 0.06	0.81 ± 0.06
Цела апликација	Људски учинак	0.95 ± 0.02	0.91 ± 0.06	0.92 ± 0.04
	Тежински ансамбл	0.85 ± 0.06	0.91 ± 0.05	0.87 ± 0.04
Поузданост	Људски учинак	0.84 ± 0.14	0.83 ± 0.16	0.82 ± 0.13
	Тежински ансамбл	0.78 ± 0.05	0.85 ± 0.05	0.8 ± 0.04
Компатибилност	Људски учинак	0.94 ± 0.08	0.73 ± 0.26	0.78 ± 0.17
	Тежински ансамбл	0.94 ± 0.07	0.69 ± 0.08	0.75 ± 0.08
Садржај	Људски учинак	0.47 ± 0.3	0.45 ± 0.38	0.43 ± 0.33
	Тежински ансамбл	0.62 ± 0.1	0.7 ± 0.18	0.64 ± 0.11

Сумирајући резултате у оквиру *CrowdRE* домена, може се закључити да модел тежинског ансамбла представља робусно решење које успешно досеже, а у појединим аспектима и превазилази, просечан људски учинак у задацима класификације намера и тема. У наредном поглављу анализирају се перформансе истих методолошких приступа у домену детекције емоција.

6.2. Резултати у домену емоција усмерених на учење

У овом поглављу анализирају се перформансе модела на задацима класификације аспеката и емоција унутар домена емоција усмерених на учење. За разлику од претходно анализираних техничких захтева, овај домен карактерише висок степен лингвистичке суптилности и субјективности, што детекцију афективних стања чини знатно изазовнијом.

Табела 13 приказује макро *F1*-мере за све методолошке приступе. Фино подешени *PTM* доминирају у оба задатка, што је у складу са налазима претходних студија [111] [113] у *TBED* домену. Остварени бенефити примене аугментације података додатно пружају емпиријску подршку налазима систематског прегледа литературе [111], који истиче ефикасност ових техника не само у рачунарској визији, већ и у *NLP* контексту.

Иако су резултати за класификацију емоција генерално нижи него код аспеката, што осликава инхерентну тежину детекције осећања која захтева нијансирано разумевање контекста, предложени ансамбл приступи поново остварују најбоље и најконзистентније резултате. Интересантно је приметити да су у домену емоција оба ансамбл модела остварила идентичне макро *F1* вредности ($F1 = 0.73 \pm 0.07$). Постигнути резултати указују на то да ансамбл методе пружају стабилно решење које је отпорно на специфичне недостатке појединачних модела. Таква стабилност је кључна за анализу емоција, с обзиром на лингвистичке суптилности и различите начине на које студенти изражавају своја афективна стања.

Иако су оба ансамбл приступа остварила висок ниво успешности, тежински ансамбл је одабран за даљу детаљну анализу по класама. Овај модел је остварио боље вредности макро *F1*-мере у задатку класификације аспеката, док је у задатку класификације емоција, упркос идентичним макро-просечним резултатима, показао већу статистичку стабилност са мањим одступањима у вредностима микро и тежинске *F1*-мере. Табела 14 садржи декомпозицију резултата по појединачним класама тежинског ансамбла упоредо са референтним људским учинком.

Табела 13 Просечне макро *F1*-мере за задатке класификације аспеката и емоција кроз 10 независних и стратификованих подела података. Подебљане вредности означавају најбољи учинак унутар приступа, док је свеукупно најбољи резултат додатно подвучен

Методолошки приступ	Модел	Аспект	Емоција
Људски учинак		0.84 ± 0.1	0.82 ± 0.1
Развој од нуле	*TF-IDF* вектори	**0.75 ± 0.03**	**0.62 ± 0.08**
	mBERT cased вектори	0.65 ± 0.02	0.59 ± 0.05
	mBERT uncased вектори	0.73 ± 0.04	0.53 ± 0.06
	RoBERTa вектори	0.52 ± 0.1	0.52 ± 0.06
	BERTić вектори	0.49 ± 0.08	0.51 ± 0.06
Фино подешавање *PTM*	*mBERT cased*	0.78 ± 0.02	0.67 ± 0.05
	mBERT uncased	0.79 ± 0.03	0.69 ± 0.04
	RoBERTa	0.48 ± 0.23	0.47 ± 0.04
	BERTić	0.78 ± 0.05	0.60 ± 0.19
	mBERT cased аугментован заменом	0.76 ± 0.05	0.65 ± 0.07
	mBERT uncased аугментован заменом	**0.8 ± 0.02**	0.67 ± 0.07
	BERTić аугментован заменом	0.61 ± 0.29	**0.71 ± 0.06**
	mBERT cased аугментован уметањем	0.76 ± 0.04	0.66 ± 0.06
	mBERT uncased аугментован уметањем	0.8 ± 0.03	0.68 ± 0.07
	BERTić аугментован уметањем	0.64 ± 0.24	0.57 ± 0.25
Промпт инжењерство	*GPT-3.5 Turbo*	0.51 ± 0.03	0.51 ± 0.04
	Gemma 3 27B	**0.66 ± 0.02**	**0.62 ± 0.03**
Ансамбл	Већинско гласање	0.79 ± 0.02	**0.73 ± 0.07**
	Тежински ансамбл	**0.81 ± 0.03**	**0.73 ± 0.07**

Упоредни приказ перформанси по класама указује на изузетну ефикасност тежинског ансамбла у оквиру домена емоција усмерених на учење. Посебно је значајно што модел у неколико кључних категорија остварује боље резултате у односу на људе. Код класификације аспекта, ансамбл остварује већу *F1*-меру на класама **Задаци**, **Лекција**, **Наставни материјал** и **Интерфејс**.

Сличан тренд приметан је и код задатка класификације емоција. За класе **Ангажованост**, **Задовољство**, **Неутрално**, **Досада** и **Изненађеност** модел остварује боље резултате у односу на просечан људски учинак. Посебно се истиче емоција **Изненађеност**, где је ансамбл остварио $F1 = 0.6 \pm 0.2$, што представља значајан помак у односу на $F1 = 0.41 \pm 0.21$, колико су постигли људски анотатори, иако ова класа има свега 10 примера у корпусу.

Табела 14 Упоредни приказ прецизности (*P*), одзива (*R*) и *F1*-мере по појединачним класама за људски учинак и најбољи модел у домену емоција усмерених на учење

Аспект	**Метод**	***P***	***R***	***F1***
Искуство	Људски учинак	0.97 ± 0.02	0.95 ± 0.02	0.96 ± 0.02
	Тежински ансамбл	0.96 ± 0.01	0.95 ± 0.01	0.95 ± 0.01
Задаци	Људски учинак	0.94 ± 0.06	0.93 ± 0.05	0.93 ± 0.05
	Тежински ансамбл	0.94 ± 0.02	0.94 ± 0.02	0.94 ± 0.02
Лекција	Људски учинак	0.92 ± 0.04	0.94 ± 0.02	0.93 ± 0.03
	Тежински ансамбл	0.96 ± 0.02	0.96 ± 0.02	0.96 ± 0.02
Наставни материјал	Људски учинак	0.78 ± 0.15	0.79 ± 0.15	0.78 ± 0.15
	Тежински ансамбл	0.89 ± 0.04	0.9 ± 0.04	0.89 ± 0.03
Интерфејс	Људски учинак	0.89 ± 0.07	0.93 ± 0.04	0.91 ± 0.05
	Тежински ансамбл	0.97 ± 0.03	0.93 ± 0.03	0.95 ± 0.02
Софтвер	Људски учинак	0.68 ± 0.17	0.68 ± 0.23	0.67 ± 0.2
	Тежински ансамбл	0.77 ± 0.11	0.62 ± 0.04	0.66 ± 0.05
Емоција	**Метод**	***P***	***R***	***F1***
Ангажованост	Људски учинак	0.89 ± 0.08	0.78 ± 0.28	0.81 ± 0.22
	Тежински ансамбл	0.92 ± 0.01	0.91 ± 0.01	0.91 ± 0.01
Задовољство	Људски учинак	0.85 ± 0.18	0.92 ± 0.04	0.87 ± 0.11
	Тежински ансамбл	0.92 ± 0.01	0.94 ± 0.01	0.93 ± 0.01
Фрустрација	Људски учинак	0.94 ± 0.04	0.9 ± 0.08	0.92 ± 0.05
	Тежински ансамбл	0.89 ± 0.02	0.92 ± 0.02	0.9 ± 0.01
Неутрално	Људски учинак	0.92 ± 0.04	0.93 ± 0.03	0.93 ± 0.03
	Тежински ансамбл	0.97 ± 0.02	0.93 ± 0.03	0.95 ± 0.02
Збуњеност	Људски учинак	0.93 ± 0.1	0.89 ± 0.07	0.9 ± 0.07
	Тежински ансамбл	0.83 ± 0.09	0.87 ± 0.06	0.85 ± 0.06
Досада	Људски учинак	0.75 ± 0.19	0.92 ± 0.11	0.82 ± 0.15
	Тежински ансамбл	0.96 ± 0.04	0.79 ± 0.05	0.85 ± 0.04
Изненађеност	Људски учинак	0.45 ± 0.18	0.42 ± 0.24	0.41 ± 0.21
	Тежински ансамбл	0.6 ± 0.2	0.6 ± 0.2	0.6 ± 0.2

Сумирајући резултате у домену емоција усмерених на учење, може се закључити да најбољи модел не само да успешно досеже људски ниво препознавања аспеката и емоција, већ га у већини категорија и превазилази. Знатно мања варијанса указује на висок степен употребљивости оваквог модела у реалним образовним условима.

У наредном поглављу приказана је анализа статистичке значајности разлика у перформансама између тежинског ансамбла и људског учинка, чиме се пружа коначан методолошки доказ о предностима овог приступа.

6.3. Анализа статистичке значајности резултата

У овом поглављу приказани су резултати статистичке анализе којом се потврђују или одбацују постављене истраживачке хипотезе. У складу са експерименталним поступком дефинисаним у поглављу 5.6, за проверу значајности разлика средњих вредности макро-просечне *F1*-мере примењен је Велчов *t*-тест. Овај тест омогућава поуздано поређење перформанси у ситуацијама када узорци могу имати неједнаке варијансе. Сва статистичка тестирања извршена су на нивоу значајности $\alpha = 0.05$. У контексту поређења модела са људским учинком (хипотезе $\boldsymbol{H_1}$ и $\boldsymbol{H_2}$), вредност $p > 0.05$ би указала на одсуство статистички значајне разлике, чиме би се потврдила упоредивост перформанси. Насупрот томе, код испитивања утицаја балансирања класа (хипотеза $\boldsymbol{H_3}$), вредност $p < 0.05$ би потврдила да примењене технике доводе до статистички значајног побољшања резултата.

6.3.1. Тестирање хипотеза $\boldsymbol{H_1}$ и $\boldsymbol{H_2}$

У табели 15 приказани су збирни резултати најбољег модела за сва четири класификациона задатка у поређењу са просечним људским учинком.

На основу резултата приказаних у табели 15, може се закључити да су хипотезе $\boldsymbol{H_1}$ и $\boldsymbol{H_2}$ делимично потврђене. Модел је успешно достигао људски ниво перформанси у задацима детекције **намере** и **аспекта** ($p > 0.05$), док је код семантички сложенијих задатака детекције **теме** и **емоција** забележена статистички значајна разлика у корист људи.

Табела 15 Статистичко поређење перформанси најбољег модела и људског учинка

Задатак	Човек *F1*	Модел *F1*	*p*-вредност	Одлука
Намера	0.84 ± 0.09	0.84 ± 0.03	1.0	***H₁* потврђена**
Тема	0.75 ± 0.13	0.59 ± 0.06	0.004	***H₁* није потврђена**
Аспект	0.84 ± 0.1	0.81 ± 0.03	0.3837	***H₂* потврђена**
Емоција	0.82 ± 0.1	0.73 ± 0.07	0.033	***H₂* није потврђена**

Парцијална потврда хипотеза H_1 и H_2 одражава објективну разлику у комплексности задатака. Задаци класификације **намера** и **аспеката** имају мањи број класа и јаснију семантичку разграниченост, што моделима олакшава учење. Код задатка класификације **тема**, изражена комплексност уочљива је и кроз нижу вредност *IAA* међу анотаторима у поглављу 4.2, док код задатка класификације **емоција**, упркос високој људској сагласности, модели још увек не достижу ниво суптилног афективног расуђивања.

6.3.2. Тестирање хипотезе H_3

Хипотеза H_3 претпоставља да примена техника балансирања класа доводи до статистички значајног пораста макро-просечне *F1*-мере. Тестирање је извршено поређењем најбољих фино подешених *PTM*-ова пре и након примене стратегија аугментације у простору података (замена и уметање речи). Вредности добијене Велчовим *t*-тестом за сва четири класификациона задатка приказани су у табели 16.

Табела 16 Статистичка анализа утицаја балансирања класа на перформансе модела

Задатак	Модел	Без балансирања	Са балансирањем	*p*-вредност	Одлука
Намера	*BERTić*	0.79 ± 0.19	0.83 ± 0.04	0.5297	***H₃* није потврђена**
Тема	*mBERT uncased*	0.47 ± 0.09	0.52 ± 0.09	0.2301	***H₃* није потврђена**
Аспект	*mBERT uncased*	0.79 ± 0.03	0.8 ± 0.02	0.3937	***H₃* није потврђена**
Емоција	*BERTić*	0.60 ± 0.19	0.71 ± 0.06	0.1092	***H₃* није потврђена**

Иако вредности у табели 16 показују конзистентан пораст средње *F1*-мере, Велчов *t*-тест не сугерише статистичку значајност на нивоу макро-

просека ($p > 0.05$). Главни разлог за одсуство статистичке значајности је изузетно висока варијанса модела пре балансирања. Међутим, ефекат ових техника је јасно видљив кроз драстичну стабилизацију модела. Код детекције **намере**, стандардна девијација је опала са 0.19 на 0.04, а код **емоција** са 0.19 на 0.06. Ово указује на то да балансирање суштински чини модел робуснијим и поузданијим за реалну примену.

Да би се дубље истражило зашто долази до оваквих помака, извршена је анализа на нивоу појединачних мањинских класа корпуса. Табела 17 приказује промене (Δ) у прецизности (*P*), одзиву (*R*) и *F1*-мери након примена техника аугментације.

Табела 17 Утицај аугментације података на перформансе мањинских класа

Задатак	Класа	Δ*P*	Δ*R*	Δ*F1*
Намера	Пријава грешке	+0.10	0	+0.05
Намера	Оцена	+0.10	+0.05	+0.06
Тема	Задаци	+0.06	+0.16	+0.10
Тема	Поузданост	+0.10	+0.02	+0.06
Тема	Компатибилност	+0.56	+0.46	+0.47
Тема	Садржај	-0.10	-0.09	-0.11
Аспект	Софтвер	-0.05	+0.10	-0.04
Емоција	Изненађеност	0	0	0

Анализа мањинских класа открива да је аугментација значајно унапредила перформансе у већини мањинских класа, са изузетним порастом код класе **Компатибилност** од 47%. Чак и у случајевима где је укупна *F1*-мера благо опала (класа **Софтвер**), забележен је значајан пораст одзива од 10%, што потврђује да је модел постао осетљивији на ретке примере. Код класе **Садржај** забележен је пад свих метрика, док код емоције **Изненађеност** није дошло до промене услед изузетно малог броја оригиналних узорака за учење. Ови налази указују на то да, иако је стратешко балансирање путем уметања и замене речи веома ефикасно, за екстремно ретке класе би у будућим истраживањима требало размотрити примену *LLM*-ова за генерисање већег обима синтетичких података.

Иако добијени резултати не омогућавају потврду хипотезе ***H₃*** у њеном најстрожем статистичком смислу на нивоу целог корпуса, детаљна анализа показује да технике балансирања значајно унапређују

класификацију мањинских класа и стабилизују понашање модела . Ово пружа снажну индикацију о оправданости примене ових техника, посебно у контексту поузданости инжењерског решења.

6.4. Анализа грешака најуспешнијег модела

Након анализе статистичке значајности резултата, у овом поглављу спроведена је детаљна анализа грешака тежинског ансамбла као најуспешнијег модела. Циљ ове анализе је идентификација специфичних лингвистичких и семантичких образаца који доводе до погрешне класификације, чиме се добија увид у границе примењених архитектура у домену образовног контекста. Анализа обухвата детаљан преглед погрешних класификација по задацима, њихову систематску категоризацију, као и процену утицаја структуре скупа података и специфичности српског језика на успешност модела.

6.4.1. Анализа грешака по задацима класификације

У задатку класификације **намера**, најуспешнији модел показује највећу успешност у идентификацији класе **Захтев за функционалношћу** у односу на мање заступљене категорије. Међутим, уочена је изражена тенденција модела ка лажно позитивним резултатима управо у овој класи. Овакав образац грешака директно рефлектује дистрибуцију података у корпусу, где су **Захтеви за функционалношћу** заступљенији него преостале две класе заједно, што условљава благу пристрасност модела ка овој доминантној класи приликом обраде семантички вишезначних примера.

Анализа указује на то да најчешћи тип грешке представља погрешно класификовање **Оцене** као **Захтева за функционалношћу**, док се као пар класа који је најтеже међусобно разграничити издвајају **Пријава грешке** и **Захтев за функционалношћу**. Ове грешке потичу из неспособности модела да јасно разграничи исказивање незадовољства или опис техничких потешкоћа од формалних предлога за унапређење система. Специфично, повратне информације које описују лоше корисничко искуство, попут неинтуитивног интерфејса или конфузне навигације, модел често интерпретира као имплицитни **Захтев за**

функционалношћу[4]. Ови резултати сугеришу да модел успешно препознаје општу потребу за променом, али услед специфичног начина студентског изражавања, које често меша критику и предлоге, не успева увек да прецизно одреди да ли је примарни фокус поруке субјективна оцена квалитета или конкретан захтев за новом функционалношћу.

У задатку класификације **тема**, најуспешнији модел показује комплексну дистрибуцију погрешних класификација, примарно вођену семантичким преклапањем између педагошких и функционалних категорија. Као најбројнији пар класа који модел најтеже међусобно разграничава издвајају се **Кориснички интерфејс** и **Употребљивост**, што указује на изазов у дефинисању границе између визуелног приказа и саме логике тока апликације. С друге стране, појединачно најчешћа грешка модела је погрешно препознавање теме **Употребљивост** као **Задаци**, што сугерише да се општи проблеми у коришћењу платформе често перципирају као потешкоће у изради конкретних задатака. Такође, у многим случајевима модел није у стању да препозна контекстуалну разлику између теоријских објашњења **Наставних материјала** и практичних захтева **Задатака**.

Додатни извор конфузије јесте и разграничавање између **Поузданости** и категорија као што су **Кориснички интерфејс** или **Компатибилност**. Било која реч која помиње грешке у раду, дефекте или неодзивност система значајно утиче да модел предвиђа **Поузданост**, чак и када је из контекста јасно да је узрок проблема везан за дизајн интерфејса или компатибилност са окружењем, попут проблема са екстензијама. Карактеристични примери изазовних случајева који илуструју ове обрасце приказани су у табели 18.

У задатку класификације **аспеката**, најуспешнији модел се суочава са значајним изазовима у разграничавању категорија **Искуство**, **Софтвер** и **Задаци**. Убедљиво најбројнији пар погрешних класификација представља мешање субјективног **Искуства** корисника са коментарима који се тичу **Софтвера** као алата. Модел показује снажну тенденцију ка прекомерном сврставању повратних информација у категорију **Искуство**,

[4] Пример: *Nije mi bilo intuitivno da elementi koriste jednu od kategorija kao svoju početnu lokaciju*.

кад год студенти користе личну заменицу у првом лицу или евалуативне придеве[5].

Табела 18 Репрезентативни примери погрешних класификација у задатку класификације тема

Реченица	Тачно	Предвиђено	Напомена
Primetio sam da na par rezolucija responsiv design nije bas najoptimalniji, tako da bih to promenio	Кориснички интерфејс	Задаци	Јасан проблем са корисничким интерфејсом погрешно интерпретиран као потешкоћа са задатком.
Mozda bih dodao jos kracih zadataka	Задаци	Наставни материјал	Предлог за још задатака погрешно класификован као захтев за наставним материјалом.
Prikazivanje svih predmeta koje sam do sada radio na tutoru (mislim na SIMS prevenstveno)	Употребљивост	Задаци	Захтев за функцију навигације/историје погрешно схваћен као опис задатка.
Trenutno imam problem gde mi se nista desava kada pokusam da submitujem challenge u visual studio code.	Компатибилност	Поузданост	Екстерни технички проблем погрешно приписан унутрашњој поузданости платформе.
I jedna sugestija za kraj: je li moze neka precica na tastaturi za "Save" u biljeskama?	Употребљивост	Кориснички интерфејс	Захтев за функцију која повећава ефикасност коришћења софтвера погрешно интерпретиран као захтева за визуелном променом интерфејса.
U poslednjem videu je samo receno sta je greska, ali ne i na koji nacin bi ona trebala da se ispravi :)	Наставни материјал	Задаци	Критика методологије наставног материјала протумачена као проблем са задатком.

Додатно, примећује се специфична пристрасност модела при којој се технички проблеми и системске грешке често погрешно класификују као **Задаци** уместо **Софтвер**. Ово сугерише да је модел превише осетљив на кључне речи као што су *задаци*, *одговори* или *питања*, те не успева да препозна да се студентски коментар односи на алат, а не на квалитет или тежину задатака[6].

У задатку класификације **емоција**, на перформансе модела снажно утиче лингвистичка сличност између парова позитивних и негативних сентимената. Кључно запажање је систематска конфузија између

[5] Пример: *Deluje kao zanimljiva ideja*.
[6] Пример: *Daje mi da radim zadatke koje sam već rešila, bez razloga.*

Ангажованости и **Задовољства**. Модел често не успева да направи разлику између активног учешћа студента (**Ангажованост**) и његовог исказивања задовољства, те фактичке изјаве о разумевању често погрешно класификује као **Задовољство**.

Други значајан изазов представља негативна пристрасност модела у вези са тежином и репетитивношћу. Модел тежи да свако помињање тежине, апстрактности или монотоније прекомерно класификује као **Фрустрацију**, чак и када је стварна емоција **Досада**, **Збуњеност** или чак неутрална **Ангажованост**. На пример, реченице које изражавају исцрпљеност или монотоност готово искључиво се предвиђају као **Фрустрација**. Ово сугерише да модел интерпретира стања високог напора или понављања као активну иритацију, а не као пасивну незаинтересованост.

Модел има потешкоћа са класом **Изненађеност**, коју је најтеже идентификовати због њене екстремне малобројности у скупу података. Већина примера **Изненађености** погрешно је класификована као **Ангажованост**. **Неутрална** класа се често меша са **Задовољством** када су повратне информације кратке и не садрже индикаторе негативног сентимента. Насупрот томе, неспособност модела да детектује **Досаду**, коју учестало меша са **Фрустрацијом**, указује на недостатак осетљивости на пасивну природу досаде у односу на активну природу фрустрације.

Свеукупно, иако је модел способан да детектује присуство емоције, он се бори да прецизно одреди њен емотивни поларитет и интензитет, посебно у присуству кључних речи везаних за задатке попут *задаци*, *понављање* или *објашњење*. Репрезентативни примери ових изазовних класификација, који илуструју границе препознавања поларитета и интензитета емоције, приказани су у табели 19.

Табела 19 Репрезентативни примери погрешних класификација у задатку класификације емоција

Реченица	Тачно	Предвиђено	Напомена
Dobro, razumljivo	Ангажованост	Задовољство	Чињенична изјава о разумевању погрешно протумачена као израз одушевљења.
Sve je bilo jasno	Задовољство	Ангажованост	Задовољство јаснoћом погрешно класификовано као стандардно ангажовање.
Polako postajem umorna :(	Досада	Фрустрација	Пасивно стање исцрпљености/досаде погрешно протумачено као фрустрација.
Kodovi u lekcijama postaju nepregledni.	Фрустрација	Ангажованост	Критичке повратне информације о јасноћи кода се погрешно тумаче као неутрално запажање везано за задатак.
Ova lekcija je bila malo teza.	Ангажованост	Фрустрација	Неутрално запажање о тежини лекције погрешно протумачено као фрустрација.
Malo me zbunjuje ova lekcija, ali se nekako snalazim.	Збуњеност	Фрустрација	Модел не препознаје стање збуњености корисника, подразумевано предвиђајући општу фрустрацију.
Zapravo veoma zanimljivo i korisno (iznad ocekivanja)	Изненађеност	Ангажованост	Нијанса позитивног изненађења се губи, при чему модел предвиђа стандардно ангажовање.
bez problema	Неутрално	Задовољство	Одсуство проблема погрешно протумачено као позитивно емоционално стање.

6.4.2. Систематска категоризација типова грешака

Како би се систематски анализирале грешке у класификацији кроз десет независних подела на тренинг и тест скуп, прво су идентификоване све јединствене погрешно класификоване реченице унутар сваког тест скупа за сваки задатак класификације. С обзиром на то да су се поједине реченице појављивале у више различитих тест скупова, за сваку од њих одређена је најчешћа погрешно предвиђена класа кроз сва појављивања. Свака јединствено погрешно класификована реченица затим је прегледана и ручно јој је додељена једна од четири унапред дефинисане категорије грешака:

- **Преклапање категорија** - грешке код којих класе деле веома сличне или преклапајуће семантичке карактеристике, што доводи до конфузије.
- **Субјективност** - грешке које произилазе из субјективне интерпретације вишезначних изјава од стране анотатора.
- **Недовољно информација** - грешке узроковане недостатком довољно информација о класи унутар саме реченице, што отежава прецизну класификацију.
- **Контекстуално неразумевање** - грешке које су последица неуспеха модела да интерпретира исправно значење или суптилни контекст реченице.

Затим су избројана појављивања сваке категорије грешке за сваку класу, чиме су јасно идентификовани најзаступљенији типови грешака. Резултати ове анализе сумирани су у табели 20.

Резултати приказани у табели 20 указују на то да је преклапање категорија доминантан извор грешака у готово свим задацима класификације, што сугерише да се највећи изазов налази у самој семантичкој блискости дефинисаних класа.

У задатку класификације **намера**, преклапање категорија чини највећи удео грешака (44%), што је посебно изражено код класа **Захтев за функционалношћу** и **Оцена**. Ово указује на то да модел често препознаје општу намеру за променом, али не успева да повуче јасну границу између субјективног вредновања и формалног захтева. Код задатка **тема**, преклапање је још израженије (61%), нарочито код функционалних категорија као што су **Употребљивост** и **Кориснички интерфејс**, где су границе често суптилне и испреплетене у студентском изражавању.

Контекстуално неразумевање је значајан фактор код задатака класификације **емоција** (28%) и **тема** (23%). Модел повремено греши у интерпретацији правог значења специфичног жаргона или сажетости коментара којима недостаје довољно околног текста за потпуно информисано одлучивање. Занимљиво је да се код емоција субјективност појављује као изразито висок фактор (27%), примарно због класе **Ангажованост**, где су анотатори вероватно користили додатно

разумевање ширег контекста комуникације које модел, ограничен на појединачне реченице, не поседује у истој мери.

Табела 20 Дистрибуција типова грешака по задацима и класама

Намера	**Преклапање категорија**	**Субјективност**	**Недовољно информација**	**Контекстуално неразумевање**
Захтев за функционалношћу	**14**	3	3	8
Пријава грешке	6	1	6	**14**
Оцена	**18**	5	4	4
Тема	**Преклапање категорија**	**Субјективност**	**Недовољно информација**	**Контекстуално неразумевање**
Кориснички интерфејс	**37**	1	10	7
Употребљивост	**45**	1	6	12
Задаци	**15**	3	4	7
Наставни материјал	**18**	3	6	5
Цела апликација	1	0	2	**8**
Поузданост	**14**	0	1	4
Компатибилност	**10**	0	0	9
Садржај	**5**	0	1	1
Аспект	**Преклапање категорија**	**Субјективност**	**Недовољно информација**	**Контекстуално неразумевање**
Искуство	**8**	5	3	**8**
Задаци	**8**	5	3	**8**
Лекција	**4**	0	0	3
Наставни материјал	**14**	0	0	3
Интерфејс	**8**	1	0	3
Софтвер	**19**	1	10	11
Емоција	**Преклапање категорија**	**Субјективност**	**Недовољно информација**	**Контекстуално неразумевање**
Ангажованост	13	**41**	0	12
Задовољство	**15**	3	1	8
Фрустрација	3	0	3	**21**
Неутрално	0	5	**8**	3
Збуњеност	**13**	0	0	0
Досада	**16**	0	2	4
Изненађеност	**4**	0	0	**4**

Коначно, одређени број грешака произилази из недостатка информација унутар појединачних реченица, што онемогућава недвосмислену класификацију без обзира на комплексност модела. На основу ових налаза, будући рад би требало усмерити ка увођењу вишеструке

анотације (енгл. *multilabel annotation*), што би омогућило моделу да додели више класа у случајевима када је преклапање семантички оправдано. Такође, унапређење смерница за анотацију могло би додатно смањити утицај субјективности, посебно у домену афективне анализе.

6.4.3. Утицај структуре коментара на прецизност класификације

Како би се додатно истражио утицај структуре скупа података на перформансе модела, анализиран је ефекат реченица проистеклих из коментара који се састоје од више реченица. За сваки задатак класификације измерена је пропорција тест реченица које потичу из оваквих коментара, као и удео погрешних класификација повезаних са њима, што је омогућило директно поређење заступљености ових реченица у скупу и њиховог стварног доприноса укупном броју грешака. Анализа је спроведена кроз свих десет независних подела скупа података, а резултати су усредњени како би се добиле средње вредности и стандардне девијације. Ови резултати су приказани у табели 21.

Табела 21 Утицај коментара са више реченица на учесталост погрешних класификација по задацима

Задатак	**Реченице из коментара са више реченица**	**Погрешне класификације из коментара са више реченица**
Намера	0.56 ± 0.04	0.70 ± 0.08
Тема	0.55 ± 0.03	0.60 ± 0.08
Аспект	0.25 ± 0.01	0.32 ± 0.13
Емоција	0.25 ± 0.01	0.32 ± 0.07

Резултати приказани у табели 21 показују да су реченице изведене из коментара са више реченица значајно допринеле стопама погрешне класификације у задацима **намера** и **тема**. У задатку класификације **намера**, иако је 56% тест реченица потицало из коментара са више реченица, оне су чиниле чак 70% погрешних класификација. Сличан тренд, иако мање изражен, примећен је и у задатку класификације **тема**. Насупрот томе, задаци класификације **аспеката** и **емоција** показали су мању пропорцију реченица из коментара са више реченица, што је резултирало сразмерно мањим утицајем на погрешну класификацију.

Ови налази сугеришу да губитак ширег контекста током екстракције реченица може непропорционално утицати на задатке који захтевају суптилније разумевање. Како би се ублажио овај проблем, будућа унапређења могла би да обухвате коришћење већих контекстуалних прозора или очување додатних околних реченица током претпроцесирања, како би се боље обухватило првобитно значење.

6.4.4. Лингвистички изазови српског језика

Како би се истражило на који начин специфичне лингвистичке карактеристике српског језика утичу на перформансе модела, анализирано је неколико понављајућих проблема уочених у погрешно класификованим реченицама. Као што је претходно описано у поглављу 3.3, висок степен флексије српског језика доводи до великог броја јединствених облика речи, што директно усложњава учење засновано на токенима и ствара изазове у семантичком мапирању различитих граматичких облика истог појма. Поред тога, релативно слободан редослед речи [124] уводи значајне синтаксичке варијације. Ове варијације могу представљати изазов за моделе засноване на *Transformer* архитектури, који су примарно тренирани на језицима са стриктнијом синтаксичком структуром, попут енглеског.

Додатно, у неформалном писању често долази до изостављања дијакритичких знакова (појава позната као *ћелава латиница*), што узрокује неусклађеност токена и смањује поузданост модела у препознавању тачног значења. Специфични и често сложени обрасци негације, својствени индиректном или учтивом говору, додатно замагљују намеру говорника, што је нарочито уочљиво у задатку детекције емоција. Ове лингвистичке одлике заједнички уводе двосмисленост у синтаксичку и семантичку интерпретацију коју *Transformer* модели тешко разрешавају без ширег контекста.

Репрезентативни примери оваквих грешака илустровани су у табели 22, а добијени налази указују на неопходност увођења специфичних корака претпроцесирања прилагођених српском језику.

Табела 22 Примери лингвистичких изазова који утичу на перформансе модела

Реченица	Задатак	Тачно	Предвиђено	Лингвистичка одлика	Напомена
Odlican sadrzaj	Тема	Садржај	Цела апликација	Изостављање дијакритичких знакова	Недостатак дијакритичких знакова смањује тачност подударања токена.
Tekst pre zadataka bih organizovao u manje pasuse koji su sazetiji	Тема	Наставни материјал	Садржај	Слободан редослед речи	Инверзија објекта и глагола уводи структурну двосмисленост за модел.
Nisam ga dovoljno koristio	Емоција	Неутрално	Збуњеност	Висок степен флексије	Облик негације мења емотивну нијансу и отежава интерпретацију.
Ne bih znao kasti	Емоција	Неутрално	Збуњеност	Сложена негација	Учтива негација изражава општу несигурност пре него експлицитну збуњеност.

6.5. Анализа рачунарске ефикасности

Поред класификационих перформанси и претходно анализираних извора грешака, важан аспект практичне применљивости развијених модела представља њихова рачунарска ефикасност. У оквиру окружења описаног у поглављу 5.6, измерена су времена обуке и предикције за све методолошке приступе на једној репрезентативној подели података. Ради упоредивости, мерења фино подешених *PTM* спроведена су на њиховим основним варијантама, будући да аугментација утиче на састав тренинг скупа, а не на архитектуру, те не мења битно рачунарски профил модела. Времена обуке приказана су у табели 23.

Табела 23 приказује времена обуке за приступе који то захтевају. Статистички модели, ослоњени на *TF-IDF* репрезентацију, обучавају се готово тренутно, за мање од једне секунде по задатку. Фино подешавање *PTM* представља рачунарски захтевнију операцију, реализовану у распону од приближно четири до осам минута по задатку. Иако ова

вредност није занемарљива, реч је о једнократном трошку који се у потпуности реализује на потрошачком хардверу, без потребе за наменском серверском инфраструктуром. Приступи засновани на *LLM*-овима не захтевају тренирање, те су из ове табеле изостављени.

Табела 23 Време обуке модела по задатку на једној репрезентативној подели података

Приступ	Модел	Задатак	Реченице	Време обуке
Развој од нуле	*TF-IDF* вектори	Намера	470	0.3 s
		Тема	469	0.4 s
		Аспект	1007	0.2 s
		Емоција	1007	0.4 s
Фино подешавање *PTM*	*BERTić*	Намера	470	230.7 s
	mBERT uncased	Тема	469	238.4 s
	mBERT uncased	Аспект	1007	504.8 s
	BERTić	Емоција	1007	484.4 s

Насупрот једнократном трошку обуке, време предикције представља трошак који се понавља при свакој обради нових података. Статистички модели обрађују појединачну реченицу за мање од једне милисекунде, а фино подешени *PTM* за приближно 10 до 55 милисекунди, при чему се обрада у потпуности извршава локално.

Приступи засновани на *LLM*-овима показали су вишеструко дужа времена предикције. *GPT-3.5 Turbo* је обрадио сва четири тест скупа за укупно приближно 12 минута, уз укупан трошак од $0.59. Детаљан утрошак ресурса по задатку приказан је у табели 24. Локални *Gemma 3* модел не изискује новчани трошак, али захтева приближно 19 *GB* меморије уз време обраде од приближно 1 секунде на краћим, до 20 секунди по реченици на задацима са дужим промптовима.

Табела 24 Утрошак ресурса *GPT-3.5 Turbo* модела по задатку

Задатак	Реченице	Улазни токени	Излазни токени	Време обраде	Трошак
Намера	118	108 884	119	136.8 s	$0.055
Тема	118	118 759	118	117.0 s	$0.060
Аспект	252	473 118	252	245.3 s	$0.237
Емоција	252	478 158	252	196.0 s	$0.239
Укупно	740	1 178 919	741	695.1 s	$0.591

Приказани резултати потврђују да предложени оквир може да функционише у потпуности локално, на стандардном потрошачком хардверу и без зависности од екстерних сервиса. При томе је важно разликовати две конфигурације. Најбоље рангирани тежински ансамбл постиже највише вредности *F1*-мере, али укључивањем локалног *Gemma 3* модела наслеђује његов меморијских захтев и дуже време предикције. Фино подешени модели мањег формата, са друге стране, задржавају перформансе блиске ансамблу из вишеструко краће време предикције, чиме представљају рачунарски ефикасну алтернативу погодну за примену у реалним образовним системима са ограниченим ресурсима.

6.6. Дискусија доприноса и могућности примене

Резултати представљени у овој дисертацији потврђују да је аутоматизована анализа студентских повратних информација на српском језику методолошки изводљива и довољно поуздана за интеграцију у реалне образовне системе. Примарни научни допринос овог рада огледа се у дефинисању анотационих шема и инструкција за *CrowdRE* и емоције усмерене на учење, на основу којих је формиран висококвалитетан корпус студентских коментара. Посебно је значајно што су целокупан корпус [144] и пратеће анотационе инструкције[7] јавно доступни, чиме је постављена прва референтна тачка за будућа истраживања у овој области на српском језику. Оваквим приступом се значајно смањује напор неопходан за покретање сличних студија и омогућава директно поређење резултата.

У практичном инжењерском смислу, развијени модел тежинског ансамбла омогућава трансформацију неструктурираних повратних информација у податке који помажу у доношењу одлука (енгл. *Data-driven Decision Making - DDDM*). У оквиру *ITS*-ова, ови модели могу служити као софтверски сензори за праћење афективних стања студената. Истраживање је демонстрирало високу робусност модела на неформалан студентски говор. Детекција збуњености или фрустрације у интеракцији са задацима и наставним материјалом омогућава наставном особљу да идентификује критичне тачке у курикулуму и правовремено га прилагоди,

[7] https://github.com/Clean-CaDET/serbian-student-feedback-corpora

што директно доприноси унапређењу корисничког искуства током процеса учења.

Поред образовног значаја, остварени резултати имају директну примену у процесу одржавања и еволуције софтвера кроз аутоматизацију *CrowdRE* задатака. Са инжењерског аспекта, развијени тежински ансамбл представља робусно решење које достиже људски ниво учинка, а притом се у потпуности извршава локално, без зависности од екстерних сервиса. Ова независност чини решење погодним за примену у индустрији образовних технологија и локалним академским оквирима, где су заштита података и локална имплементација приоритети. Овакав практичан исход истраживања представља конкретну манифестацију треће мисије универзитета кроз трансфер знања и технолошких решења ка локалној академској заједници и привреди. Стабилност коју су показали модели након примене техника за балансирање класа сугерише да је предложено решење спремно за интеграцију у сложене софтверске архитектуре, нудећи робустан алат за дубље разумевање потреба корисника у академском окружењу.

6.7. Ограничења истраживања и претње валидности

Приликом интерпретације остварених резултата потребно је размотрити одређена ограничења и потенцијалне претње валидности истраживања. Једна од примарних претњи конструктној валидности односи се на ослањање на људску процену приликом креирања скупа података. Како би се минимизовао ризик од погрешног тумачења категорија у оквиру таксономија, анотација је спроведена кроз две групе од по четири анотатора уз примену детаљних смерница. Ове смернице су итеративно ревидиране на основу анализа неслагања у пробним фазама, чиме је осигурана већа објективност корпуса. Додатно, одабир специфичних модела и архитектура заснован је на детаљном прегледу релевантне литературе и праксе, чиме је осигурана методолошка утемељеност изабраних техника.

Интерна валидност истраживања могла је бити угрожена потенцијалним преклапањем категорија унутар одабраних таксономија. Како би се овај ризик ублажио, извршена је квалитативна анализа садржаја, при чему је

свакој реченици додељена само једна доминантна категорија. Иако је овакво редуктивно мапирање могло довести до губитка одређених суптилних информација у тексту, оно је било неопходно за постизање високе стабилности класификације. Такође, ризик од преприлагођавања услед примене технике аугментације контролисан је кроз употребу десет независних стратификованих подела података. Ипак, важно је напоменути да код екстремно мало заступљених класа чак ни технике балансирања нису могле у потпуности да надокнаде недостатак изворних језичких образаца. Додатно ограничење представља и природа примењених *Transformer* модела, чија структура *црне кутије* (енгл. *black box*) ограничава интерпретабилност донетих одлука, што може бити релевантно у образовним сценаријима који захтевају експлицитна образложења. Слично ограничење важи и за коришћене велике језичке моделе, чија интерпретабилност донетих одлука кроз промпт инжењерство остаје ограничена.

Закључна валидност истраживања односи се на поузданост статистичких закључака изведених из експерименталне евалуације. За проверу значајности разлика средњих вредности макро-просечне *F1*-мере примењен је Велчов *t*-тест, који је одабран јер не претпоставља једнакост варијанси узорака, што је адекватно за поређење модела са различитим нивоима варијабилности у перформансама. Кроз 10 независних стратификованих подела података обезбеђена је процена варијабилности перформанси, али релативно мали број итерација може ограничити снагу теста при детекцији мањих ефеката. Ово ограничење је најуочљивије код задатка детекције емоција у оквиру хипотезе $\boldsymbol{H_3}$, где је забележено највеће одступање у средњим вредностима $(+0.11)$ и најнижа p-вредност (0.1092), али без достизања прага статистичке значајности.

У погледу екстерне валидности, очекује се да се закључци истраживања могу генерализовати на повратне информације студената унутар других *ITS*-ова који подучавају принципе чистог кода (енгл. *clean code*). Међутим, специфичне карактеристике различитих едукативних окружења и лингвистичке нијансе могу ограничити применљивост модела изван овог контекста без додатне адаптације. Ефикасност решења је такође инхерентно везана за квалитет базних модела тренираних на општим језичким корпусима, што представља зависност која се

рефлектује на перформансе у специфичним доменима. Коначно, остварени резултати су специфични за српско говорно подручје. Иако је архитектура тежинског ансамбла универзално применљива, њен крајњи успех директно зависи од доступности и квалитета језичких ресурса за конкретан језик, што остаје изазов за даљу примену овог приступа на друге језике са ограниченим ресурсима.

7. Закључак

У оквиру ове дисертације прикупљене су и анализиране повратне информације студената унутар интелигентног система подучавања (*ITS*). Истраживање је обухватило домене инжењерства захтева заснованог на групи корисника (*CrowdRE*) и детекцију емоција усмерених на учење, са циљем унапређења *ITS*-а и разумевања афективних стања студената током образовног процеса. С порастом броја корисника, ручна анализа постаје непрактична, што условљава потребу за развојем аутоматизованих метода.

Методологија истраживања заснивала се на примени *MATTER* методологије за анотацију повратних информација прикупљених унутар платформе *Clean CaDET*. У процесу анотације учествовала су четири анотатора по домену, а резултујући корпус обухвата 1850 реченица из два домена истраживања: 591 у *CrowdRE* домену и 1259 у домену емоција усмерених на учење. Развој модела обухватио је евалуацију широког спектра приступа кроз четири задатка класификације: намере и теме у оквиру *CrowdRE* домена, те аспекте и емоције у оквиру емоција усмерених на учење. Испитивани приступи укључују статистичке методе машинског учења, фино подешавање претренираних *Transformer* модела, као и промпт инжењерство са великим језичким моделима.

Резултати евалуације показују да је основни циљ ове дисертације успешно остварен. У оквиру тестирања постављених хипотеза, потврђено је да предложени модел тежинског ансамбла достиже људске перформансе у задацима класификације намера (хипотеза $\boldsymbol{H_1}$) и аспеката (хипотеза $\boldsymbol{H_2}$), без статистички значајне разлике у односу на људске анотаторе. Код сложенијих задатака класификације тема и емоција забележена је статистички значајна разлика у корист људи, али је модел показао већу конзистентност и мању варијансу у одлучивању. Иако примена техника балансирања података није довела до статистички значајног побољшања на нивоу макро-просека (хипотеза $\boldsymbol{H_3}$), детаљна анализа показала је њихову практичну вредност у стабилизацији модела и побољшању препознавања мањинских класа.

Научни допринос ове дисертације огледа се у дефинисању јавно доступних анотационих шема и смерница на основу којих је формиран

први јавно доступан корпус студентских коментара у овим доменима [144]. Ови ресурси представљају референтну тачку за будућа истраживања и олакшавају даљи рад на српском језику. Развијени модел тежинског ансамбла омогућава трансформацију неструктурираних повратних информација у употребљиве податке за праћење афективног стања студената и ефикасније управљање еволуцијом софтвера. Коришћење специјализованих модела мањег формата нуди рачунарски ефикасно решење са високим перформансама, без потребе за скупим хардверским ресурсима или зависности од екстерних сервиса. Овакав практичан исход истраживања представља конкретан допринос трећој мисији универзитета у виду трансфера знања и технолошких решења ка локалној академској заједници и индустрији образовних технологија.

Преклапање категорија и недостатак изворних језичких образаца за мањинске класе идентификовани су као кључна ограничења. Такође, губитак ширег контекста приликом екстракције појединачних реченица из дужих коментара доприноси паду перформанси модела у задацима унутар *CrowdRE* домена. Додатно, специфичности српског језика, као што су висок степен флексије, слободан редослед речи, изостављање дијакритичких знакова и сложени обрасци негације, стварају двосмисленост коју модел тешко разрешава. Ови налази наглашавају објективне границе модела у разумевању неформалног студентског говора и указују на неопходност даљег проширења корпуса и увођења специфичних корака претпроцесирања прилагођених српском језику.

Даљи правци истраживања обухватају више димензија. На образовном плану, планира се интеграција резултата детекције емоција ради транзиције платформе *Clean CaDET* из интелигентног у афективни систем подучавања. На методолошком плану, увођење вишеструке анотације (енгл. *multilabel annotation*) омогућило би прецизније моделовање случајева са семантичким преклапањем категорија, док би коришћење великих језичких модела за генерисање синтетичких података могло да прошири ефекат техника балансирања код екстремно ретких класа. На инжењерском плану, истраживаће се методе за груписање сродних пријава грешака и захтева за новим функционалностима, као и аутоматизовано генерисање задатака (енгл. *ticket*) унутар система управљања пројектима, чиме би се повећала

ефикасност одржавања и еволуције образовних софтверских система. Резултати ове дисертације показују да је аутоматизована анализа студентских повратних информација на српском језику методолошки изводљива и да предложени оквир може послужити као основа за даљи развој образовних технологија на српском језику.

Литература

[1] J. Daniel, "Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility," *Journal of Interactive Media in Education,* vol. 2012, no. 3, p. 18, 2012.

[2] O. C. Rodrigez, "MOOCs and the AI-Stanford Like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses," *European Journal of Open, Distance and E-Learning,* 2012.

[3] M. M. Bosanac, Masovni otvoreni onlajn kursevi u oblasti visokog obrazovanja, 2025.

[4] C. Hodges, S. Moore, B. Lockee, T. Trust and A. Bond, "The difference between emergency remote teaching and online learning," *Educause review,* vol. 27, no. 1, pp. 1-9, 2020.

[5] J. Reich and J. A. Ruiperez-Valiente, "The MOOC pivot," *Science,* vol. 363, no. 6423, pp. 130-131, 2019.

[6] G. Siemens, "Connectivism: A Learning Theory for the Digital Age," *International Journal of Instructional Technology and Distance Learning,* vol. 2, 2005.

[7] S. Downes, "Connectivism," *Asian Journal of Distance Education,* vol. 17, no. 1, 2022.

[8] R. Kop and A. Hill, "Connectivism: Learning theory of the future or vestige of the past?," *International Review of Research in Open and Distributed Learning,* vol. 9, no. 3, pp. 1-13, 2008.

[9] M. Clara and E. Barbera, "Learning Online: Massive Open Online Courses (MOOCs), Connectivism, and Cultural Psychology," *Distance Education,* vol. 34, no. 1, pp. 129-136, 2013.

[10] M. Koehler and P. Mishra, "What is technological pedagogical content knowledge (TPACK)?," *Contemporary issues in technology and teacher education,* vol. 9, no. 1, pp. 60-70, 2009.

[11] W. F. Pinar, "The reconceptualisation of curriculum studies," *Journal of curriculum studies,* vol. 10, no. 3, pp. 205-214, 1978.

[12] W. F. Pinar, Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses, vol. 17, Peter Lang, 1995.

[13] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill and D. R. Krathwohl, Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain, David McKay, 1956.

[14] L. W. Anderson and D. R. Krathwohl, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition, Addison Wesley Longman, Inc., 2001.

[15] M. H. Immordino-Yang and A. Damasio, "We feel, therefore we learn: The relevance of affective and social neuroscience to education," *Mind, brain, and education,* vol. 1, no. 1, pp. 3-10, 2007.

[16] M. A. Hasan, N. F. M. Noor, S. S. B. A. Rahman and M. M. Rahman, "The Transition From Intelligent to Affective Tutoring System: A Review and Open Issues," *IEEE Access,* vol. 8, pp. 204612-204638, 2020.

[17] A. D. Rowe, "Feelings about feedback: The role of emotions in assessment for learning," *Scaling up assessment for learning in higher education,* pp. 159-172, 2017.

[18] L. Balachandran and A. Kirupananda, "Online reviews evaluation system for higher education institution: An aspect based sentiment analysis tool," in *2017 11th international conference on software, knowledge, information management and applications (SKIMA)*, 2017.

[19] N. Altrabsheh, M. M. Gaber and M. Cocea, "SA-E: sentiment analysis for education," *Intelligent Decision Technologies,* pp. 353-362, 2013.

[20] M. H. Abdulla, "The use of an online student response system to support learning of Physiology during lectures to medical students," *Education and Information Technologies,* vol. 23, no. 6, pp. 2931-2946, 2018.

[21] F. F. Balahdia, C. G. Fernando and I. C. Juanatas, "Teacher's performance evaluation tool using opinion mining with sentiment analysis," in *2016 IEEE region 10 symposium (TENSYMP)*, 2016.

[22] K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using lexicon based approach," in *2017 IEEE/ACIS 16th international conference on computer and information science (ICIS)*, 2017.

[23] N. Luburić, L. Dorić, J. Slivka, D. Vidaković, K.-G. Grujić, A. Kovačević and S. Prokić, "An Intelligent Tutoring System to Support Code Maintainability Skill Development," *IEEE Transactions on Learning Technologies,* vol. 18, pp. 289-303, 2025.

[24] N. Chen, J. Lin, S. C. Hoi, X. Xiao and B. Zhang, "AR-miner: mining informative reviews for developers from mobile app marketplace," in *Proceedings of the 36th international conference on software engineering*, 2014.

[25] E. C. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol and A. Perini, "The crowd in requirements engineering: The landscape and challenges," *IEEE software,* vol. 34, no. 2, pp. 44-52, 2017.

[26] R. S. Baker, M. M. T. Rodrigo and U. E. Xolocotzin, "The dynamics of affective transitions in simulation problem-solving environments," in *International conference on affective computing and intelligent interaction*, Berlin, Heidelberg, 2007.

[27] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learning and Instruction,* vol. 22, no. 2, pp. 145-157, 2012.

[28] R. Pekrun, "The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice," *Educational psychology review,* vol. 18, no. 4, pp. 315-341, 2006.

[29] R. Pekrun, T. Goetz, W. Titz and R. P. Perry, "Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research," *Educational psychologist,* vol. 37, no. 2, pp. 91-105, 2002.

[30] H. L. Fwa, "An architectural design and evaluation of an affective tutoring system for novice programmers," *International Journal of Educational Technology in Higher Education,* vol. 15, no. 1, pp. 1-19, 2018.

[31] C. Ling, X. Zhao, J. Lu, C. Deng, C. Zheng, J. Wang, T. Chowdhury, Y. Li, H. Cui, X. Zhang, T. Zhao, A. Panalkar, D. Mehta, S. Pasquali, W. Cheng, H. Wang, Y. Liu and Z. Chen, "Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey," *ACM Computing Surveys,* vol. 58, no. 3, p. 39, 2025.

[32] R. B. Grandić and M. M. Bosanac, "Uticaj nove produkcije znanja na savremene univerzitete," *Inovacije u Nastavi,* vol. 32, no. 3, 2019.

[33] M. M. Bosanac and J. J. Milutinović, "Treća misija sveučilišta - osnovne razlike u pristupima," *Obrazovanje za poduzetništvo - E4E: znanstveno stručni časopis o obrazovanju za poduzetništvo,* vol. 12, no. 1, pp. 20-32, 2022.

[34] U. Kuckartz, "Qualitative text analysis: A systematic approach," *Compendium for early career researchers in mathematics education,* pp. 181-197, 2019.

[35] B. Berelson, "Content Analysis in Communication Research," *The ANNALS of the American Academy of Political and Social Science,* vol. 283, no. 1, pp. 197-198, 1952.

[36] M. Schreier, Qualitative content analysis in practice, Sage Publications, 2012.

[37] P. Ekman, "An argument for basic emotions," *Cognition & emotion,* vol. 6, no. 3-4, pp. 169-200, 1992.

[38] J. A. Russel, "A circumplex model of affect," *Journal of personality and social psychology,* vol. 39, no. 6, p. 1161, 1980.

[39] K. R. Scherer, "What are emotions? And how can they be measured?," *Social science information,* vol. 44, no. 4, pp. 695-729, 2005.

[40] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments," *International Journal of Human-Computer Studies,* vol. 68, no. 4, pp. 223-241, 2010.

[41] S. Karumbaiah, R. S. Baker, J. Ocumpaugh and J. M. A. L. Andres, "A re-analysis and synthesis of data on affect dynamics in learning," *IEEE Transactions on Affective Computing,* vol. 14, no. 2, pp. 1696-1710, 2021.

[42] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on affective computing,* vol. 1, no. 1, pp. 18-37, 2010.

[43] R. Pekrun, T. Goetz, A. C. Frenzel, P. Barchfeld and R. P. Perry, "Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ)," *Contemporary educational psychology,* vol. 36, no. 1, pp. 36-48, 2011.

[44] J. Pustejovsky and A. Stubbs, Natural Language Annotation for Machine Learning, Massachusetts: O'Reilly Media, 2012.

[45] P. Roettger, B. Vidgen, D. Hovy and J. B. Pierrehumbert, "Two Contrasting Data Annotation Paradigms for Subjective NLP Tasks," *arXiv preprint arXiv:2112.07475,* 29 April 2022.

[46] Y. Oortwijn, T. Ossenkoppele and A. Betti, "Interrater disagreement resolution: A systematic procedure to reach consensus in annotation tasks," in *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, 2021.

[47] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement,* vol. 20, no. 1, pp. 37-46, 1960.

[48] R. J. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics,* pp. 159-174, 1977.

[49] S. Bird, E. Loper and E. Klein, Natural Language Processing with Python, O'Reilly Media Inc., 2009.

[50] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models, 2025.

[51] G. Salton and M. J. McGill, Introduction to Modern Information Retrieval, New York, NY: McGraw-Hill, Inc., 1986.

[52] M. Bayer, M.-A. Kaufhold and C. Reuter, "A survey on data augmentation for text classification," *ACM Computing Surveys,* vol. 55, no. 7, pp. 1-39, 2022.

[53] N. V. Chawla, K. W. Bowyer, L. O. Hall and P. W. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.

[54] G. E. Batista, R. C. Prati and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter,* vol. 6, no. 1, pp. 20-29, 2004.

[55] J. Laurikkala, "Improving identification of difficult small classes by balancing class distribution," in *Conference on artificial intelligence in medicine in Europe*, 2001.

[56] C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2006.

[57] M. Nikolić i A. Zečević, Mašinsko učenje, Beograd: Matematički fakultet, 2019.

[58] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management,* vol. 45, no. 4, pp. 427-437, 2009.

[59] C. Tantithamthavorn, S. McIntosh, A. E. Hassan and K. Matsumoto, "An empirical comparison of model validation techniques for defect prediction models," *IEEE Transactions on Software Engineering,* vol. 43, no. 1, pp. 1-18, 2016.

[60] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The journal of machine learning research,* vol. 13, no. 1, pp. 281-305, 2012.

[61] J. Snoek, H. Larochelle and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012.

[62] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu and J. Gao, "Deep Learning-based Text Classification: A Comprehensive Review," *ACM Computer Surveys,* vol. 54, no. 3, pp. 1-40, 2021.

[63] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[64] L. Breiman, "Bagging predictors," *Machine learning,* vol. 24, no. 2, pp. 123-140, 1996.

[65] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.

[66] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences,* vol. 55, no. 1, pp. 119-139, 1997.

[67] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics,* pp. 1189-1232, 2001.

[68] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016.

[69] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, 2000.

[70] Z.-H. Zhou, Ensemble methods: foundations and algorithms, Chapman and Hall/CRC, 2025.

[71] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks,* vol. 2, no. 5, pp. 359-366, 1989.

[72] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[73] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors," *nature,* vol. 323, no. 6088, pp. 533-536, 1986.

[74] J. L. Elman, "Finding structure in time," *Cognitive science,* vol. 14, no. 2, pp. 179-211, 1990.

[75] S. Hochreiter и J. Schmidhuber, „Long short-term memory,“ *Neural computation,* т. 9, бр. 8, pp. 1735-1780, 1997.

[76] K. Cho, B. v. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings*

of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.

[77] J. Firth, "A synopsis of linguistic theory, 1930-1955," *Studies in linguistic analysis,* pp. 10-32, 1957.

[78] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of Workshop at ICLR*, 2013.

[79] J. Pennington, R. Socher and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014.

[80] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is all you need," 2017.

[81] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019.

[82] T. Pires, E. Schlinger и D. Garrette, „How Multilingual is Multilingual BERT?," у *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[83] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv preprint arXiv:1907.11692,* 26 July 2019.

[84] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzman, E. Grave, M. Ott, L. Zettlemoyer and V. Stoyanov, "Unsupervised Cross-lingual Representation Learning at Scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.

[85] N. Ljubešić and D. Lauc, "BERTić - The Transformer Language Model for Bosnian, Croatian, Montenegrin and Serbian," *arXiv preprint arXiv:2104.09243,* 19 April 2021.

[86] A. Radford, K. Narasimhan, T. Salimans and I. Sutskever, "Improving Language Understanding by Generative Pre-Training," 2018. [Online]. Available: https://www.mikecaptain.com/resources/pdf/GPT-1.pdf. [Accessed December 2025].

[87] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child and Ram, "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, 2020.

[88] Gemma Team, "Gemma: Open Models Based on Gemini Research and Technology," *arXiv preprint arXiv:2403.08295,* no. 4, 16 April 2024.

[89] B. Zhang and R. Sennrich, "Root mean square layer normalization," *Advances in neural information processing systems,* vol. 32, 2019.

[90] N. Shazeer, "GLU Variants Improve Transformer," *arXiv preprint arXiv:2002.05202,* 12 February 2020.

[91] J. Su, M. Ahmed, Y. Lu, S. Pan, W. Bo and Y. Liu, "RoFormer: Enhanced transformer with Rotary Position Embedding," *Neurocomputing,* vol. 568, p. 127063, 2024.

[92] Gemma Team, "Gemma 3 Technical Report," *arXiv preprint arXiv:2503.19786,* 25 March 2025.

[93] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Boregaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean and W. Fedus, "Emergent Abilities of Large Language Models," *arXiv preprint arXiv:2206.07682,* 26 October 2022.

[94] N. Sharma and A. Sharma, "Classification of Crowd-Based Software Requirements via Unsupervised Learning," in *Quality of Information and Communications Technology*, 2024.

[95] E. C. Groen, J. Schowalter, S. Kopczynska, S. Polst and S. Alvani, "Is there Really a Need for Using NLP to Elicit Requirements? A Benchmarking Study to Assess Scalability of Manual Analysis.," in *REFSQ Workshops*, 2018.

[96] D. Pagano and W. Maalej, "User feedback in the appstore: An empirical study," in *2013 21st IEEE international requirements engineering conference (RE)*, 2013.

[97] E. Guzman and W. Maalej, "How do users like this feature? a fine grained sentiment analysis of app reviews," in *2014 IEEE 22nd international requirements engineering conference (RE)*, 2014.

[98] W. Maalej and H. Nabil, "Bug report, feature request, or simply praise? on automatically classifying app reviews," in *2015 IEEE 23rd international requirements engineering conference (RE)*, 2015.

[99] X. Gu and S. Kim, "What parts of your apps are loved by users?," in *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2015.

[100] E. Guzman, M. El-Haliby and B. Bruegge, "Ensemble methods for app review classification: An approach for software evolution," in *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2015.

[101] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora and H. C. Gall, "What would users change in my app? summarizing app reviews for recommending software changes," in *Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering*, 2016.

[102] R. Santos, E. C. Groen and K. Villela, "A Taxonomy for User Feedback Classifications," in *REFSQ Workshops*, 2019.

[103] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora and H. C. Gall, "ow can I improve my app? Classifying user reviews for software maintenance and evolution," in *2015 IEEE international conference on software maintenance and evolution (ICSME)*, 2015.

[104] L. Villarroel, G. Bavota, B. Russo, R. Oliveto and M. Di Penta, "Release planning of mobile apps based on user reviews," in *In 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, 2016.

[105] K. Phetrungnapha and T. Senivongse, "Classification of mobile application user reviews for generating tickets on issue tracking system," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 2019.

[106] R. R. Mekala, A. Irfan, E. C. Groen, A. Porter and M. Lindvall, "Classifying user requirements from online feedback in small dataset environments using deep learning," in *2021 IEEE 29th International requirements engineering conference (RE)*, 2021.

[107] M. A. Hadi and F. H. Fard, "Evaluating pre-trained models for user feedback analysis in software engineering: A study on classification of app-reviews," *Empirical Software Engineering,* vol. 28, no. 4, p. 88, 2023.

[108] A. Pilone, P. Meirelles, F. Kon and W. Maalej, "Multilingual Crowd-Based Requirements Engineering Using Large Language Models," *arXiv preprint arXiv:2408.06505,* 12 August 2024.

[109] H. Binali, C. Wu and V. Potdar, "Computational approaches for emotion detection in text," in *4th IEEE international conference on digital ecosystems and technologies*, 2010.

[110] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *A review on sentiment analysis and emotion detection from text,* vol. 11, no. 1, p. 81, 2021.

[111] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora and I. Pappas, "A Review on Text-Based Emotion Detection - Techniques, Applications,

Datasets, and Future Directions," *arXiv preprint arXiv:2205.03235,* 26 April 2022.

[112] T. Chutia and N. Baruah, "A review on emotion detection by using deep learning techniques," *Artificial Intelligence Review,* vol. 57, no. 8, 2024.

[113] S. Rathje, D.-M. Mirea, I. Sucholutsky, R. Marjieh, C. Robertson and J. J. Van Bavel, "GPT is an effective tool for multilingual psychological text analysis," *Proceedings of the National Academy of Sciences,* vol. 121, no. 34, 2024.

[114] A. Al Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha and Z. Aung, "Challenges and Opportunities of Text-Based Emotion Detection: A Survey," *IEEE Access,* vol. 12, pp. 18416-18450, 2024.

[115] X. Mao and Z. Li, "Implementing emotion-based user-aware e-learning," *CHI'09 Extended Abstracts on Human Factors in Computing Systems,* pp. 3787-3792, 2009.

[116] P. Rodriguez, A. Ortigosa and R. M. Carro, "Extracting emotions from texts in e-learning environments," in *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, 2012.

[117] S. Rani and P. Kumar, "A sentiment analysis system to improve teaching and learning," *Computer,* vol. 50, no. 5, pp. 36-43, 2017.

[118] R. Z. Cabada, M. L. B. Estrada and R. O. Bustillos, "Mining of educational opinions with deep learning," *Journal of Universal Computer Science,* vol. 24, no. 11, pp. 1604-1626, 2018.

[119] R. Oramas-Bustillos, M. L. Barron-Estrada, R. Zatarain-Cabada and S. L. Ramirez-Avila, "A corpus for sentiment analysis and emotion recognition for a learning environment," in *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 2018.

[120] M. L. Barron-Estrada, R. Zatarain-Cabada and R. O. Bustillos, "Emotion Recognition for Education using Sentiment Analysis," *Res. Comput. Sci.,* vol. 148, no. 5, pp. 71-80, 2019.

[121] M. L. B. Estrada, R. Z. Cabada, R. O. Bustillos and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Systems with Applications,* vol. 150, p. 113265, 2020.

[122] O. Grljević, Z. Bošnjak and A. Kovačević, "Opinion mining in higher education: a corpus-based approach," *Enterprise Information Systems,* vol. 16, no. 5, 2020.

[123] N. Nikolić, O. Grljević and A. Kovačević, "Aspect-based sentiment analysis of reviews in the domain of higher education," *The Electronic Library,* vol. 38, no. 1, pp. 44-64, 2020.

[124] S. Kitić, "On function of word order in English and Serbian," *FACTA UNIVERSITATIS-Linguistics and Literature,* vol. 2, no. 09, pp. 303-312, 2002.

[125] U. Marovac, A. Avdić and N. Milošević, "A Survey of Resources and Methods for Natural Language Processing of Serbian Language," *arXiv preprint arXiv:2304.05468 ,* 11 April 2023.

[126] R. Stanković, B. Šandrih, C. Krstev, M. Utvić and M. Škorić, "Machine Learning and Deep Neural Network-Based Lemmatization and Morphosyntactic Tagging for Serbian," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020.

[127] N. Milošević, "Stemmer for Serbian language," *arXiv preprint arXiv:1209.4471 ,* 20 September 2012.

[128] V. Batanović, B. Nikolić and M. Milosavljević, "Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The Serbian Movie Review Dataset," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016.

[129] T. Kuzman Pungeršek, P. Rupnik, I. Porupski, V. Dinić and N. Ljubešić, "State of the Art in Text Classification for South Slavic Languages: Fine-Tuning or Prompting?," *arXiv preprint arXiv:2511.07989,* 11 November 2025.

[130] T. Perry, "LightTag: Text Annotation Platform," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2021.

[131] N. Ide and J. Pustejovsky, Handbook of Linguistic Annotation, vol. 1, Springer Dordrecht, 2017.

[132] International Organization for Standardization, ISO/IEC 25010 - Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Product quality model, ISO/IEC, 2010.

[133] Google, "Glossary stopwords," 2024. [Online]. Available: https://cloud.google.com/translate/docs/advanced/stopwords. [Accessed 11 January 2026].

[134] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu and C. Xu, "HuggingFace's Transformers: State-of-the-art Natural Language Processing," *arXiv preprint arXiv:1910.03771,* 14 July 2020.

[135] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, "Scikit-learn: Machine learning in Python," *Journal of machine learning research,* vol. 12, no. Oct, pp. 2825-2830, 2011.

[136] OpenAI, "Models," 2024. [Online]. Available: https://platform.openai.com/docs/models. [Accessed 11 January 2026].

[137] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin and L. Invernizzi, "KerasTuner," 2019. [Online]. Available: https://github.com/keras-team/keras-tuner. [Accessed 11 January 2026].

[138] G. Lemaitre, F. Nogueira and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *Journal of Machine Learning Research,* vol. 18, no. 17, pp. 1-5, 2017.

[139] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *arXiv preprint arXiv:1711.05101,* no. 3, 4 January 2019.

[140] I. Turc, M.-W. Chang, K. Lee and K. Toutanova, "Well-Read Students Learn Better: On the Importance of Pre-training Compact Models," *arXiv preprint arXiv:1908.08962,* no. 2, 25 September 2019.

[141] E. Ma, *NLP Augmentation,* https://github.com/makcedward/nlpaug, 2019.

[142] B. L. Welch, "The generalization of STUDENT'S problem when several different population varlances are involved," *Biometrika,* vol. 34, no. 1-2, pp. 28-35, 1947.

[143] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. Van der Walt, M. Brett, J. Wilson and J. Millman, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods,* vol. 17, no. 3, pp. 261-272, 2020.

[144] D. Vidaković, N. Luburić, A. Kovačević and J. Slivka, "Datasets for Crowd-based Requirements Engineering and aspect-based detection of learning-centered emotion from the text in Serbian language," 28 July 2024. [Online]. Available: www.doi.org/10.5281/zenodo.13119214. [Accessed March 2026].

Биографија

Драган Видаковић рођен је 1993. године у Бијељини, где је 2012. године завршио Техничку школу ”Михајло Пупин”. Исте године уписује Факултет техничких наука Универзитета у Новом Саду, на студијском програму Рачунарство и аутоматика. Основне академске студије завршава 2016. године одбраном дипломског рада на тему ”Систем за препоруку са дистрибуираним заједничким филтрирањем”.

Образовање наставља на мастер академским студијама у оквиру истог студијског програма, које успешно приводи крају 2017. године одбранивши мастер рад под називом ”Пробабилистички простори знања”.

Докторске академске студије на смеру Рачунарство и аутоматика уписује 2017. године. Од исте године је запослен на матичном факултету, најпре у звању сарадника у настави, а од 2018. године у звању асистента. Аутор је и коаутор више научних радова објављених у међународним часописима и презентованих на међународним конференцијама.

Прилози

А. Анотационе инструкције за домен инжењерства захтева заснованог на групи корисника

У циљу унапређења интелигентног система за учење *Clean CaDET* вршимо анализу повратних информација студената на српском језику. Повратне информације представљају текстуални одговор студената на питање ”Како бисте унапредили *Clean CaDET* платформу као софтвер?”. Наш задатак је да развијемо решење, засновано на вештачкој интелигенцији, које ће вршити аутоматско генерисање валидних корисничких развојних захтева (енгл. *Crowd Requirements Engineering - CrowdRE*). Ваш задатак је да анотирате текстове са његовом намером (пријава бага или давање нових развојних захтева) и темом коју текст обухвата, коришћењем овог упутства.

Анотација се врши на нивоу реченице. Једна реченица може садржати једну тему, где је свака тема везана за једну намеру. Реченице су добијене поделом текста на исте. Поред реченица, доступни су вам и текстови (одговори/коментари) којем исте припадају. Кроз упутство су дате инструкције за идентификовање намере и теме дефинисаних моделом анотације. Инструкције су дате у форми објашњења и примера. Процес анотације илуструје **Слика А.1**.

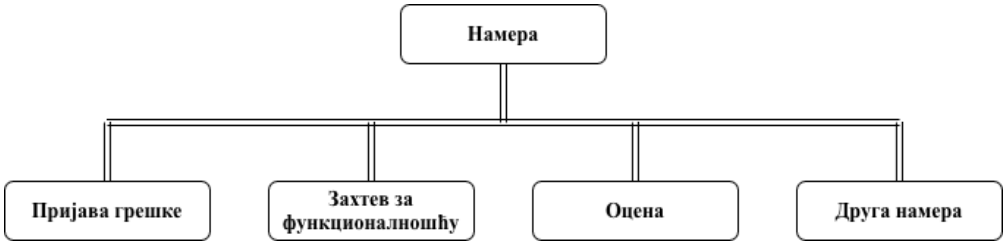
Слика А.1 Процес анотације

1. Кораци анотације и општи савети

1. Пажљиво прочитати реченицу, ако је потребно и **више пута**.
2. Анализирати реченицу како би се утврдила намера писца, и доделити реченици **једну** од предефинисаних вредности за атрибут **Намера**, описаних у одељку **2. Намера**.
3. Анализирати реченицу како би се утврдила тема описана у реченици, и доделити реченици **једну** од предефинисаних вредности за атрибут **Тема**, описаних у одељку **3. Тема**.
4. Намера и тема могу бити исказани на **имплицитан и експлицитан начин**. **Експлицитно** писање наводи информације директно и прецизно, чинећи их лаким за разумевање. **Имплицитно** писање је индиректно и сугестивно и дозвољава вишеструка тумачења. Имплицитност захтева од анотатора да изведе закључак и попуни празнине како би разумео ауторово намеравано значење.
5. Уколико се реченици не може придружити намера или тема без разумевања **контекста**, анотатори се упућују да погледају остале реченице које чине коментар, те да након тога придруже адекватне атрибуте реченици коју анотирају, као што је описано у одељку **4. Контекст**.

2. Намера

Атрибут **Намера** представља намеру или циљ коментарисања. Категорије намера о којима се пише су *Пријава грешке*, *Захтев за функционалношћу* и *Оцена*. За реченице на основу чијег садржаја није могуће утврдити намеру, дефинисана је вредност *Друга намера*. Хијерархију намера илуструје **Слика А.2**. У наставку су детаљније објашњене и илустроване појединачне намере.



Слика А.2 Вредности атрибута Намера

Пријава грешке (енгл. *Bug report*) има за циљ да информише програмера о проблему или дефекту комплетног софтвера или неке његове функционалности које је корисник/писац пронашао. Примере реченица, чија је намера *Пријава грешке*, садржи **Табела А.1**.

Табела А.1 Примери реченица чија је намера Пријава грешке

Реченица	Намера	Тема
Ne mogu da predjem na sledeci zadatak.	Пријава грешке	Поузданост
Dugme za beleške je prekriveno dok je dugme za night mode nekako iznad slike (u beloj kutiji??).	Пријава грешке	Кориснички интерфејс

Захтев за функционалношћу (енгл. *Feature request*) представља захтеве корисника/писца да се додају нове функционалности или садржај у софтвер, или да се уклоне, измене или побољшају постојеће. Примере реченица, чија је намера *Захтев за функционалношћу*, садржи **Табела А.2**.

Табела А.2 Примери реченица чија је намера Захтев за функционалношћу

Реченица	Намера	Тема
Povećati font određenih delova teksta.	Захтев за функционалношћу	Кориснички интерфејс
Odgovori na pitanja bi trebalo da budu malo manje apstraktni	Захтев за функционалношћу	Задаци

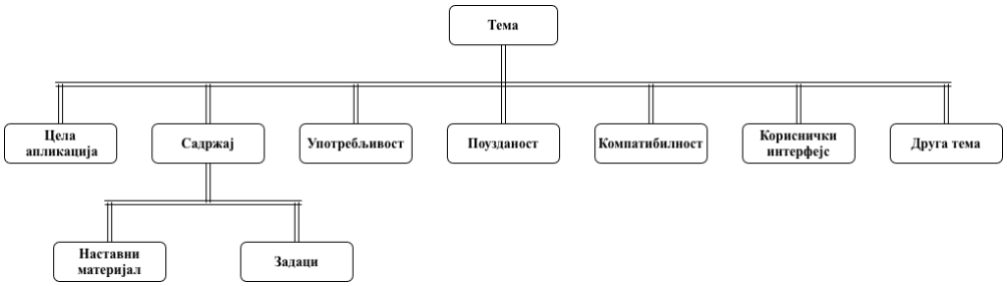
Оцена (енгл. *Rating*) има за циљ да корисник/писац изрази општу захвалност или опште незадовољство апликацијом. Фокусира се на опште расуђивање и укључује похвале, омаловажавања и критике. Примере реченица, чија је намера *Оцена*, садржи **Табела А.3**.

Табела А.3 Примери реченица чија је намера Оцена

Реченица	Намера	Тема
Tamna tema je super.	Оцена	Кориснички интерфејс
Obrisao bih ga	Оцена	Цела апликација

3. Тема

Атрибут **Тема** процењује софтверски производ, његове аспекте и контекст специфичне теме о којима корисник/писац дели своје мишљење. Основне теме о којима се пише су *Цела апликација*, њен *Садржај*, *Употребљивост*, *Поузданост*, *Компатибилност* и *Кориснички интерфејс*. Тема *Садржај* обухвата две категорије: *Наставни материјал* и *Задаци*. За реченице на основу чијег садржаја није могуће утврдити тему, дефинисана је вредност *Друга тема*. Хијерархију тема илуструје **Слика А.3**. Анотатор прво придружује реченици неку од тема на нижем хијерархијском нивоу. Уколико не постоји основа за такав начин анотације, реченици се додељује тема вишег хијерархијског нивоа. У наставку су детаљније објашњене и илустроване појединачне теме.



Слика А.3 Вредности атрибута Тема

Цела апликација (енгл. *Whole App*) се односи на реченице које посматрају софтвер у целини, без његових функционалности. Примере реченица, чија је тема *Цела апликација*, садржи **Табела А.4**.

Табела А.4 Примери реченица чија је тема Цела апликација

Реченица	Тема	Намера
Svidja mi se tutor kakav je sada, mozda cu imati neke zamerke kasnije u toku ucenja.	Цела апликација	Оцена
Zabavna je aplikacija za koriscenje	Цела апликација	Оцена

Наставни материјал (енгл. *Instructional items*) се односни на наставне материјале доступне у софтверу - инструкционе видее, хинтове, текстове и илустрације. Примере реченица, чија је тема *Наставни материјал*, садржи **Табела А.5**.

Табела А.5 Примери реченица чија је тема Наставни материјал

Реченица	Тема	Намера
Edukativan sadržaj je vrlo čitko i jasno napisan i jednostavno ga je pratiti.	Наставни материјал	Оцена
Mogli bi neki hintovi i na WEB delu	Наставни материјал	Захтев за функционалношћу

Задаци (енгл. *Assessment items*) се односи на елементе за проверу знања - питања са понуђеним одговорима, задатке са превлачењем и изазове за програмирање. Примере реченица. чија је тема *Задаци*, садржи **Табела А.6**.

Табела А.6 Примери реченица чија је тема Задаци

Реченица	Тема	Намера
Dodao bih kviz koji sadrži celokupno dosadašnje gradivo.	Задаци	Захтев за функционалношћу
Svi zadaci su veoma dobri i poucni, ali ih je previše.	Задаци	Оцена

Садржај (енгл. *Content*) се односи на сав садржај доступан кроз софтвер, а који се не може категорисати у једну од претходне две категорије (*Наставни материјал* и *Задаци*). Примере реченица, чија је тема *Садржај*, садржи **Табела А.7**.

Табела А.7 Примери реченица чија је тема Садржај

Реченица	Тема	Намера
Koriscenje animacija kao primera koriscenja funkcionalnosti	Садржај	Захтев за функционалношћу
Malo manje teksta	Садржај	Захтев за функционалношћу

Употребљивост (енгл. *Usability*) се односи на лакоћу којом корисник може научити и користити софтвер. Ово укључује интуитивност, ефикасност, лакоћу разумевања, навигације и довршавања задатака. Примере реченица. чија је тема *Употребљивост*, садржи **Табела А.8**.

Табела А.8 Примери реченица чија је тема Употребљивост

Реченица	Тема	Намера
Moguće automatsko preuzimanje materijala.	Употребљивост	Захтев за функционалношћу
Treba omogućiti lakši pregled i navigaciju kroz lekcije.	Употребљивост	Захтев за функционалношћу

Поузданост (енгл. *Reliability*) се односи на способност софтвера да функционише исправно и доследно током времена. Ово такође укључује способност софтвера да се опорави од кварова и настави са радом без прекида. Примере реченица, чија је тема *Поузданост*, садржи **Табела А.9**.

Табела А.9 Примери реченица чија је тема Поузданост

Реченица	Тема	Намера
U jednom trenutku su mi se pojavljivala samo 2 pitanja od 4, samo su se vrtela da tva pitanja u krug.	Поузданост	Пријава грешке
Jedino što mi se učinilo kao da ne valja, je 101% vičnosti.	Поузданост	Пријава грешке

Компатибилност (енгл. *Compatibility*) се односи на могућност софтвера да ради са другим софтверима. Ово омогућава да се подаци и функционалности деле између различитих софтверских апликација и платформи, омогућавајући им да ефикасно комуницирају и размењују информације. Примере реченица. чија је тема *Компатибилност*, садржи **Табела А.10**.

Табела А.10 Примери реченица чија је тема Компатибилност

Реченица	Тема	Намера
Trebalo bi omogućiti rukovanje tutorom sa drugim razvojnim okruženjima.	Компатибилност	Захтев за функционалношћу
VS Code i Tutor ne saradjuju medjusobno	Компатибилност	Пријава грешке

Кориснички интерфејс (енгл. *User Interface - UI*) је тачка интеракције између корисника и софтвера. То је начин на који корисници комуницирају са софтвером и контролишу га, као и начин на који софтвер

представља информације и повратне информације кориснику. Кориснички интерфејс укључује графичке елементе, као што су иконе, дугмад и менији, као и команде засноване на тексту које корисник уноси за интеракцију са софтвером. Примере реченица, чија је тема *Кориснички интерфејс*, садржи **Табела А.11**.

Табела А.11 Примери реченица чија је тема Кориснички интерфејс

Реченица	Тема	Намера
Na nekim mestima se Tutor piše velikim, a na nekim malim početnim slovom.	Кориснички интерфејс	Пријава грешке
Tamna tema mi je jako lijepa, ali mi se ne svidja plava boja menija sa strane.	Кориснички интерфејс	Оцена

4. Контекст

За реченице за које није могуће придружити намеру или тему без разматрања контекста, потребно је погледати и остале реченице које чине коментар. Примере реченица, чију намеру или тему није могуће прецизно утврдити без разматрања контекста, садржи **Табела А.12**.

Табела А.12 Примери реченица чију намеру или тему није могуће прецизно утврдити без разматрања контекста

Реченица	Целокупан коментар	Намера	Тема
Tu boju bih promijenila u neku drugu :)	Tamna tema mi je jako lijepa, ali mi se ne svidja plava boja menija sa strane. Tu boju bih promijenila u neku drugu :)	Захтев за функционалношћу	Кориснички интерфејс
Ostalo super.	Jedino što mi se učinilo da ne valja, je 101% vičnost. Ostalo super.	Оцена	Цела апликација

Б. Анотационе инструкције за домен емоција усмерених на учење

У циљу евалуације рада интелигентног система за учење *Clean CaDET* вршимо анализу повратних информација студената на српском језику. Повратне информације представљају текстуални одговор студената на питање ”Како бисте описали ваше искуство учења са *Clean CaDET* платформом?”. Наш задатак је да развијемо решење, засновано на вештачкој интелигенцији, које ће вршити аутоматску детекцију емоција заснованих на учењу на основу различитих аспеката. Ваш задатак је да анотирате текстове са емоцијама које оне изражавају, коришћењем овог упутства.

Анотација се врши на нивоу реченице. Једна реченица може садржати једну емоцију, где је свака емоција везана за један аспект. Реченице су добијене поделом текста на исте. Поред реченица, доступни су и текстови (одговори/коментари) ком исте припадају. Кроз упутство су дате инструкције за идентификовање аспеката и емоција дефинисаних моделом анотације. Инструкције су дате у форми објашњења и примера. Процес анотације илуструје **Слика Б.1**.

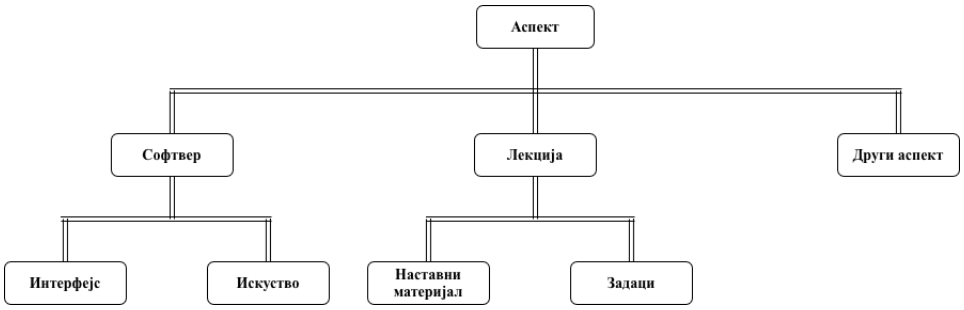
Слика Б.1 Процес анотације

1. Кораци анотације и општи савети

1. Пажљиво прочитати реченицу, ако је потребно и **више пута**.
2. Анализирати реченицу како би се утврдио предмет о којем се пише, и доделити реченици **једну** од предефинисаних вредности за атрибут **Аспект**, описаних у одељку **2. Аспект**.
3. Утврдити која емоција је испољена у реченици, и доделити реченици **једну** од предефинисаних вредности за атрибут **Емоција**, описаних у одељку **3. Емоција**.
4. Аспект и емоција могу бити исказани на **имплицитан и експлицитан начин**. **Експлицитно** писање наводи информације директно и прецизно, чинећи их лаким за разумевање. **Имплицитно** писање је индиректно и сугестивно и дозвољава вишеструка тумачења. Имплицитност захтева од анотатора да изведе закључке и попуни празнине како би разумео ауторово намеравано значење. Примери имплицитности се могу наћи у одељку **4. Имплицитност**.
5. Уколико се реченици не може придружити аспект или емоција без разматрања **контекста**, анотатори се упућују да прегледају остале реченице које чине коментар, те да након тога придруже адекватне атрибуте реченици коју анотирају, као што је описано у одељку **5. Контекст**.

2. Аспект

Атрибут **Аспект** осликава предмет коментарисања. Основни аспекти о којима се пише су *Софтвер* и *Лекција*. Аспект *Софтвер* обухвата две категорије: *Интерфејс* и *Искуство*. Аспект *Лекција* обухвата категорије: *Наставни материјал* и *Задаци*. За реченице на основу чијег садржаја није могуће утврдити аспект, дефинисана је вредност *Други аспект*. Хијерархију аспеката илуструје **Слика Б.2**. Анотатор прво разматра да ли се реченици може доделити неки од аспеката на нижем хијерархијском нивоу. Уколико не постоји основа за такав начин анотације, коментару се додељује аспект вишег хијерархијског нивоа. У наставку су детаљније објашњени и илустровани појединачни аспекти.



Слика Б.2 Вредности атрибута Аспект

Интерфејс (енгл. *Interface*) се односи на начин на који корисник комуницира са софтверском апликацијом и како се креће кроз њу. Укључује изглед, дугмад, текст, слике и друге визуелне елементе који чине интерфејс. Кориснички интерфејс је дизајниран да олакша корисницима да остваре своје задатке и пронађу информације које су им потребне. Примере реченица, чији аспект је *Интерфејс*, садржи **Табела Б.1**.

Табела Б.1 Примери реченица чији је аспект Интерфејс

Реченица	Аспект	Емоција
Sad me već malo iritira ovaj prozorčić	Интерфејс	Фрустрација
Kod zadatka analiziranja dijagrama, lekcija semantička kohezija, dijagram nije potpuno vidljiv i potrebno je smanjiti zoom stranica da bi se video ceo.	Интерфејс	Фрустрација

Искуство (енгл. *Experience*) се односи на целокупно искуство које корисник има приликом интеракције са софтверском апликацијом. Укључује факторе као што су једноставност коришћења, ефикасност и опште задовољство. Примере реченица, чији аспект је *Искуство*, садржи **Табела Б.2**.

Табела Б.2 Примери реченица чији је аспект Искуство

Реченица	Аспект	Емоција
Poprilicno je interesantan nacin da se nastava odrzi	Искуство	Ангажованост
Sjajno iskustvo!	Искуство	Задовољство

Уколико се реченица која је предмет анотације односи на софтвер, али се не може категорисати у једну од претходне две категорије (*Интерфејс* и *Искуство*), таквој реченици доделити аспект **Софтвер** (енгл. *Software*). Примере реченица, чији аспект је *Софтвер*, садржи **Табела Б.3**.

Табела Б.3 Примери реченица чији је аспект Софтвер

Реченица	Аспект	Емоција
Zadovoljan sam radom Tutora.	Софтвер	Задовољство
Super aplikacija.	Софтвер	Задовољство

Наставни материјал (енгл. *Instructional items*) се односи на доступност и квалитет наставних материјала - инструкционих видеа, текстова и илустрација. Примере реченица, чији аспект је *Наставни материјал*, садржи **Табела Б.4**.

Табела Б.4 Примери реченица чији је аспект Наставни материјал

Реченица	Аспект	Емоција
Ilustracije su dobre.	Наставни материјал	Ангажованост
Tesko gradivo	Наставни материјал	Фрустрација

Аспект **Задаци** (енгл. *Instructional items*) се односи на доступност и квалитет елемената за проверу знања - питања са понуђеним одговорима, задатака са превлачењем и изазова за програмирање. Примере реченица, чији аспект је *Задаци*, садржи **Табела Б.5**.

Табела Б.5 Примери реченица чији је аспект Задаци

Реченица	Аспект	Емоција
Zadaci su zanimljivi.	Задаци	Ангажованост
Svideo mi se izazov!	Задаци	Задовољство

Уколико се реченица која је предмет анотације односи на лекцију, али се не може категорисати у једну од претходне две категорије (*Наставни материјал* и *Задаци*), таквој реченици доделити аспект **Лекција** (енгл. *Lesson*). Примере реченица. чији аспект је *Лекција*, садржи **Табела Б.6**.

Табела Б.6 Примери реченица чији је аспект Лекција

Реченица	Аспект	Емоција
Lekcija je bila kratka, jasna i koncizna	Лекција	Ангажованост
Bila mi je naporna lekcija.	Лекција	Фрустрација

3. Емоција

Атрибут **Емоција** процењује емоцију коју писац исказује у свом коментару док процењује или суди о предмету писања. Овај атрибут може попримити вредности *Досада*, *Збуњеност*, *Фрустрација*, *Задовољство*, *Изненађеност* и *Ангажованост*, као што илуструје **Слика Б.3**. За реченице на основу чијег садржаја се није могла утврдити ниједна од горе наведених емоција, или се иста није могла са сигурношћу утврдити, дефинисана је вредност *Неутрално*. У наставку су детаљније објашњене и илустроване појединачне емоције.

Слика Б.3 Вредности атрибута Емоција

Досада (енгл. *Boredom*) је негативно емоционално стање које карактерише осећање апатије и незаинтересованости за задатак, и често је праћено недостатком пажње, ниском мотивацијом и општим осећајем неангажованости. Може се препознати употребом кључних речи или фраза: *досадно*, *монотоно*, *незанимљиво*. Одликује је недостатак ентузијазма у писању. Примере реченица, чија емоција је *Досада*, садржи **Табела Б.7**.

Табела Б.7 Примери реченица чија је емоција Досада

Реченица	Емоција	Аспект
Zadaci sa čekiranjem su monotoni.	Досада	Задаци
Sada vec osecam da ima previse zadataka i postajem umoran	Досада	Задаци
Ne mogu da se koncentrišem na lekciju.	Досада	Лекција
Tutor me uspavljuje.	Досада	Софтвер
Lekcija je neprivlačna.	Досада	Лекција

Збуњеност (енгл. *Confusion*) је афективно стање које карактерише осећај неизвесности, двосмислености и тешкоћа да се разуме информација или ситуација. Може се препознати употребом кључних речи или фраза: *збуњен*, *несигуран*, *изгубљен*. Одликује се употребом језика који указује на недостатак јасноће, као што је изражавање сумње у разумевање или постављање питања која захтевају додатна појашњавања. Примере реченица, чија емоција је *Збуњеност*, садржи **Табела Б.8**.

Табела Б.8 Примери реченица чија је емоција Збуњеност

Реченица	Емоција	Аспект
Konfuzni su mi zahtevi lekcije.	Збуњеност	Лекција
Nisam razumela sta se trazi u zadatku, treba bolje formulisati pitanje	Збуњеност	Задаци
Izgubljen sam u ovoj lekciji	Збуњеност	Лекција
Imam problema da shvatim ideju ovog dijagrama	Збуњеност	Наставни материјал
Mučim se da shvatim ovu lekciju	Збуњеност	Лекција

Фрустрација (енгл. *Frustration*) је негативан емоционални одговор на препреке или сметње у постизању циљева. То је стање несигурности и незадовољства које произилази из нерешених проблема или неиспуњених

потреба. Може се препознати употребом кључних речи или фраза: *фрустриран*, *изнервиран*, *обесхрабрен*. Одликује се употребом оштрог језика. Примере реченица. чија емоција је *Фрустрација*, садржи **Табела Б.9**.

Табела Б.9 Примери реченица чија је емоција Фрустрација

Реченица	Емоција	Аспект
Ponovo problemi sa izazovima	Фрустрација	Задаци
Nerviraju me izazovi	Фрустрација	Задаци
Zašto su izazovi ovako teški?!	Фрустрација	Задаци
Izbor boja me izluđuje	Фрустрација	Интерфејс
Slanje odgovora traje večno	Фрустрација	Задаци

Задовољство или одушевљење (енгл. *Delight*) је позитиван емоционалан одговор на аспекте окружења за учење, као што су занимљив и пријатан садржај или успешно извршење задатка. Често је резултат доживљавања или предвиђања пожељног или пријатног догађаја, или задовољења личне потребе или циља. Карактеришу га осећања радости и усхићења. Може се препознати употребом кључних речи или фраза: *одушевљен*, *задовољан*, *узбуђен*. Одликује се употребом позитивног језика, ентузијастичним тоном или енергичним стилом писања. Примере реченица, чија емоција је *Задовољство*, садржи **Табела Б.10**.

Табела Б.10 Примери реченица чија је емоција Задовољство

Реченица	Емоција	Аспект
Super, lekcije su lake za razumevanje!	Задовољство	Лекција
Kvizovi su zabavni!	Задовољство	Задаци
Ja sam oduševljen Tutorom!	Задовољство	Софтвер
Ovi hintovi su takvo olakšanje	Задовољство	Наставни материјал
Tutor je najbolja stvar za učenje	Задовољство	Софтвер

Изненађеност (енгл. *Surprise*) је осећај неочекиваности или чуђења у погледу наставног материјала или задатка. То је изненадно откриће нечег неочекиваног и карактерисано је осећајем неочекиваности или шока. Може се препознати употребом кључних речи или фраза: *изненађен*, *зачуђен*, *шокиран*, *неочекиван*. Одликује се употребом узвичног језика, повећањем енергије или узбуђења у писању или изненадна промена тона. Примере реченица, чија емоција је *Изненађеност*, садржи **Табела Б.11**.

Табела Б.11 Примери реченица чија је емоција Изненађеност

Реченица	Емоција	Аспект
Nisam se ranije sretao sa ovakvim ucenjem	Изненађеност	Искуство
Zanimljiv uvod za korišćenje tutora	Изненађеност	Софтвер
Nisam očekivao ovakvo pitanje!	Изненађеност	Задаци
Šokiran sam kako Tutor dobro radi.	Изненађеност	Софтвер
Neverovatno iskustvo!	Изненађеност	Искуство

Ангажованост (енгл. *Engagement*) је стање укључености или интересовања за материјал или задатак за учење. Карактерише се осећањем усредсређености, посвећености и потпуног урањања и фокусирања на предмет учења. Бројни фактори, укључујући личну мотивацију, потешкоће задатка и присуство ометања, утичу на ангажовање. Може се препознати употребом кључних речи или фраза: *заинтересован*, *фокусиран*, *укључен*, *очаран*, *фасциниран*. Одликује се доследним стилом писања, употребом описног језика или позитивног тона. Примере реченица, чија емоција је *Ангажованост*, садржи **Табела Б.12**.

Табела Б.12 Примери реченица чија је емоција Ангажованост

Реченица	Емоција	Аспект
Jako interesantna interakcija.	Ангажованост	Искуство
Zanimljivi zadaci, bilo mi je interesantno da resavam.	Ангажованост	Задаци
Potpuno sam fokusiran na lekciju	Ангажованост	Лекција
Jedva čekam da rešim još koji zadatak	Ангажованост	Задаци
Tutor mi pomaže da pazim na svaki detalj	Ангажованост	Софтвер

4. Имплицитност

Имплицитност се односи на стил писања у којем се информација наговештава, сугерише или закључује, а не директно наводи.

Примере реченица, у којима је *Аспект* исказан на имплицитан начин, садржи **Табела Б.13**.

Табела Б.13 Примери реченица у којима је Аспект исказан на имплицитан начин

Реченица	Аспект	Емоција
Trenutno se osjecam zbunjeno.	Искуство	Збуњеност
Nerviram se trenutno.	Искуство	Фрустрација
Zabavno mi je.	Искуство	Задовољство
Nisam trenutno baš motivisan da radim.	Искуство	Досада

Примере реченица, у којима је *Емоција* исказана на имплицитан начин, садржи **Табела Б.14**.

Табела Б.14 Примери реченица у којима је Емоција исказана на имплицитан начин

Реченица	Емоција	Аспект
Kodovi u lekcijama postaju nepregledni.	Фрустрација	Лекција
Suviše "filozofska" pitanja.	Досада	Задаци
Nisam motivisan da učim uz Tutor.	Досада	Софтвер
Dosta pitanja se ponavlja.	Досада	Задаци
Gubim fokus na dugim tekstovima.	Досада	Наставни материјал
Može li se pitanje bolje razložiti?	Збуњеност	Задаци
Ne vidim kako se ovo pitanje uklapa u lekciju.	Збуњеност	Задаци
Mislim da u ovom hintu fali više detalja.	Збуњеност	Наставни материјал
Ovo traje predugo	Фрустрација	Искуство
Pitanje nema smisla	Фрустрација	Задаци
Nigde neću stići s ovom lekcijom	Фрустрација	Лекција
Hint je ono što mi je trebalo	Задовољство	Наставни материјал
Tutor je korak u pravom smeru	Задовољство	Софтвер
Drago mi je što koristim Tutor	Задовољство	Софтвер
Neočekivan ishod izazova	Изненађење	Задаци
Iskustvo odstupa od onoga što sam očekivao	Изненађење	Искуство
Iznenađujuć rezultat izvršavanja koda	Изненађење	Задаци
Mislim da je učenje sa Tutorom zaista zanimljivo	Ангажованост	Софтвер
Želim da zaronim dublje u lekciju	Ангажованост	Лекција
Sve bolje razumem gradivo iz Tutora	Ангажованост	Софтвер

5. Контекст

За реченице за које није могуће придружити аспект или емоцију без разматрања контекста, потребно је погледати и остале реченице које чине коментар. Примере реченица, чији аспект или емоцију није могуће прецизно утврдити без разматрања контекста, садржи **Табела Б.15**.

Табела Б.15 Примери реченица чији аспект или емоцију није могуће прецизно утврдити без разматрања контекста

Реченица	Целокупан коментар	Аспект	Емоција
Drugacije je i samim time zanimljivo.	Nisam se ranije sretao sa ovakvim ucenjem. Drugacije je i samim time zanimljivo.	Искуство	Ангажованост
To mi se svidelo	Nisam ocekivao ovako dugacak zadatak. To mi se svidelo	Задаци	Задовољство

План третмана података

Назив пројекта/истраживања

Аутоматска анализа студентских повратних информација на српском језику ради унапређења образовног софтвера и разумевања емоција у учењу

Назив институције/институција у оквиру којих се спроводи истраживање

Универзитет у Новом Саду, Факултет техничких наука

Назив програма у оквиру ког се реализује истраживање

J. Slivka, N. Luburić, D. Vidaković, A. Kovačević, G. Sladić, K.G. Grujić and S. Prokić, "Clean Code and Design Educational Tool - Clean CaDET", Funded by the Science Fund of the Republic of Serbia, Grant No. 6521051, AI - Clean CaDET.

Рачунарство и аутоматика - докторска дисертација

1. Опис података

1.1 Врста студије

Укратко описати тип студије у оквиру које се подаци прикупљају
Студија је обухватила прикупљање и ручну анотацију студентских коментара ради формирања примарног корпуса. Над тим подацима спроведен је вишефазни експеримент са циљем евалуације развијених модела машинског учења и статистичког тестирања постављених хипотеза.

1.2 Врсте података
а) **квантитативни**
б) **квалитативни**

1.3. Начин прикупљања података
а) **анкете, упитници, тестови**
б) клиничке процене, медицински записи, електронски здравствени записи
в) генотипови: навести врсту _____
г) административни подаци: навести врсту _____
д) узорци ткива: навести врсту_____
ђ) снимци, фотографије: навести врсту_____
е) **текст, навести врсту: Анонимизовани коментари студената на српском језику**
ж) мапа, навести врсту _____
з) остало: описати _____

1.3 Формат података, употребљене скале, количина података

1.3.1 Употребљени софтвер и формат датотеке:
а) Excel фајл, датотека _____

b) SPSS фајл, датотека _____
c) PDF фајл, датотека _____
d) Текст фајл, датотека **.json**
e) JPG фајл, датотека _____
f) Остало, датотека _____

1.3.2. Број записа (код квантитативних података)

а) број варијабли **4 циљне варијабле класификације**
б) број мерења (испитаника, процена, снимака и сл.) **1850 анотираних реченица, 7 анотатора**

1.3.3. Поновљена мерења
а) да
б) **не**

Уколико је одговор да, одговорити на следећа питања:
а) временски размак између поновљених мера је _____
б) варијабле које се више пута мере односе се на _____
в) нове верзије фајлова који садрже поновљена мерења су именоване као _____

Напомене:
_____

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?
а) ***Да***
б) Не
Ако је одговор не, образложити _____
_____

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?
а) експеримент, навести тип _____
б) корелационо истраживање, навести тип _____
ц) анализа текста, навести тип _____
д) **остало, навести шта: Теренско истраживање (студија случаја) у реалном образовном окружењу, кроз прикупљање студентских одговора на питања отвореног типа у оквиру евалуације платформе Clean CaDET.**

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

За потребе спровођења контролисаног, вишеанотаторског процеса коришћен је специјализовани алат *LightTag*.

2.2 Квалитет података и стандарди

2.2.1. Третман недостајућих података
а) Да ли матрица садржи недостајуће податке? Да **Не**

Ако је одговор да, одговорити на следећа питања:
а) Колики је број недостајућих података? _____
б) Да ли се кориснику матрице препоручује замена недостајућих података? Да Не
в) Ако је одговор да, навести сугестије за третман замене недостајућих података

_____

2.2.2. На који начин је контролисан квалитет података? Описати

У процесу је учествовало шест независних анотатора и аутор дисертације, при чему су сваки домен анотирале по четири особе (три анотатора и аутор). Коначне ознаке су одређене већинским гласањем, док је у случају изједначеног броја гласова пресуђивао аутор дисертације.

2.2.3. На који начин је извршена контрола уноса података у матрицу?
***LightTag* алат аутоматски извози забележене податке у коначан *JSON* формат. Алат онемогућава унос недозвољених вредности и осигурава да ниједан ентитет не остане неанотиран.**

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у **Zenodo** *репозиторијум.*
3.1.2. URL адреса **https://zenodo.org/records/13119214**
3.1.3. DOI **https://doi.org/10.5281/zenodo.13119214**

3.1.4. Да ли ће подаци бити у отвореном приступу?
а) ***Да***
б) *Да, али после ембарга који ће трајати до* _____
в) *Не*

Ако је одговор не, навести разлог _____

3.1.5. Подаци неће бити депоновани у репозиторијум, али ће бити чувани. Образложење

_____

_____

3.2 Метаподаци и документација података
3.2.1. Који стандард за метаподатке ће бити примењен? ***DataCite Metadata Schema***

3.2.1. Навести метаподатке на основу којих су подаци депоновани у репозиторијум.
Наслов, Аутор, Датум публикације, Опис, Верзија, Лиценца и Кључне речи.

Ако је потребно, навести методе које се користе за преузимање података, аналитичке и процедуралне информације, њихово кодирање, детаљне описе варијабли, записа итд.

_____

_____

_____

_____

3.3 Стратегија и стандарди за чување података
3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? **Трајно**
3.3.2. Да ли ће подаци бити депоновани под шифром? Да **Не**
3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? Да **Не**
3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена?
Да **Не**
Образложити

_____

_____

4. Безбедност података и заштита поверљивих информација

Овај одељак МОРА бити попуњен ако ваши подаци укључују личне податке који се односе на учеснике у истраживању. За друга истраживања треба такође размотрити заштиту и сигурност података.
4.1 Формални стандарди за сигурност информација/података
Истраживачи који спроводе испитивања с људима морају да се придржавају Закона о заштити података о личности (*https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html*) и одговарајућег институционалног кодекса о академском интегритету.

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? Да **Не**
Ако је одговор Да, навести датум и назив етичке комисије која је одобрила истраживање

_____

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Да **Не**
Ако је одговор да, наведите на који начин сте осигурали поверљивост и сигурност информација везаних за испитанике:

а) Подаци нису у отвореном приступу
б) Подаци су анонимизирани
ц) Остало, навести шта

_____
_____

5. Доступност података

5.1. Подаци ће бити
*а) **јавно доступни***
б) доступни само уском кругу истраживача у одређеној научној области
ц) затворени

Ако су подаци доступни само уском кругу истраживача, навести под којим условима могу да их користе:

_____

_____

Ако су подаци доступни само уском кругу истраживача, навести на који начин могу приступити подацима:

_____

_____

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.
Ауторство - некомерцијално - делити под истим условима.

6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података

Драган Видаковић, vdragan@uns.ac.rs

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима

_____

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима

_____