

**РАЧУНАРСКА ОБРАДА И АНАЛИЗА ПОДАТАКА У ИСТРАЖИВАЊУ ОРАЛНОГ
ЗДРАВЉА****COMPUTER-BASED PROCESSING AND ANALYSIS OF DATA IN ORAL HEALTH
RESEARCH**

Мирослав Томић, Факултет техничких наука, Нови Сад

Област – ЕЛЕКТРОТЕХНИКА И РАЧУНАРСТВО

Кратак садржај – У овом раду описано је рачунарско испитивање података у вези са оралним здрављем код трудница. Описане су статистичке и истражне анализе података, након чега је дат опис чишћења и трансформације података. Описани су употребљавани предикциони модели и које су припреме извршене у циљу њиховог покретања. Приказани су добијени резултати и предочено је њихово могуће тумачење.

Кључне ријечи: Орално здравље, Орално здравље трудница, Анализа података, Предикциони модели

Abstract – This paper describes computer-based investigation of data in relation with oral health in pregnant women. Statistical and exploratory data analyses are described, followed by a description of data cleansing and transformation. The predictive models used are described, together with preparations that were made in order to run them. The obtained results are presented, as well as their possible interpretation.

Keywords: Oral health, Oral health of pregnant women, Data analysis, Prediction models

1. УВОД

Дуго времена опште здравље човјека није довођено у везу са оралним здрављем због недостатка информација о њиховој вези. Орално здравље утиче на опште здравље изазивајући знатне болове и патњу, мијењајући исхрану људи, њихов говор и њихов квалитет живота [1]. Орално здравље такође утиче и на друге хроничне болести [2]. Оно такође утиче на физичко и психичко стање људи и на то како расту, уживају у животу, говоре, жваћу, пробају храну и друже се [3]. Озбиљан каријес умањује квалитет живота дјеце која доживљавају бол, нелагодност, унакаженост, акутне и хроничне инфекције [1].

За опште здравље жене која је била трудна и здравље њеног новорођенчета, важно је и питање оралног здравља [4,5]. Трудноћа је природни процес праћен значајним физиолошким и хормонским промјенама у женском тијелу, укључујући и усну шуљину [6]. Неки од честих оралних проблема у трудноћи су труднички гингивитис, бенигне гингивалне лезије, покретљивост зуба, зубни каријес и пародонтитис [7].

НАПОМЕНА:

Овај рад је проистекао из мастер рада чији ментор је био доц. др Владимир Иванчевић.

Ови проблеми могу довести до губитка зуба. До сада су сагледани различити фактори који могу утицати на здравље труднице, али није прецизно утврђено који су то фактори који директно утичу.

Колико је превенција важна говори и податак да је Свјетска здравствена организација креирала глобалну стратегију за превенцију и контролу незаразних болести која уводи нову стратегију за превенцију и контролу оралних болести [2]. Нажалост у нашој околини програми за превенцију се не промовишу довољно и не уређују редовно. Актуелна уредба о националном програму превентивне стоматолошке здравствене заштите [8] датира из 2009. године, а програм који му је претходио, програм превентивне стоматолошке заштите становника Србије [9], датира из 1996. године.

Циљ овог рада је предикција недостатка зуба код трудница помоћу предиктивних модела. За достизање овог циља примјењена је истражна и статистичка анализа над подацима који су прикупљени путем анкетних упитника. Након тога примјењени су предиктивни модели: логистичка регресија (енгл. *Logistic Regression*) [10], стабла одлучивања (енгл. *Decision Trees*) [10] и случајна шума (енгл. *Random Forest*) [10].

Опис ранијих истраживања је у 2. поглављу. Описи скупа података и његовог чишћења, трансформације и анализе су у 3. и 4. поглављу, а опис припреме предикционих модела у 5. поглављу. Резултати и дискусија су у 6. поглављу, а закључак у 7. поглављу.

2. ПРЕГЛЕД ПРЕТХОДНИХ ИСТРАЖИВАЊА

Значајан број радова посвећује пажњу трудничким навикама и њиховој анализи. У студијама је истицано да постоји велика преваленција у нарушеном оралном здрављу код трудница и да је стање забрињавајуће [11-17]. Погрешна увјерења о заштити зуба током трудноће, велики број поварених зуба, проблеми са каријесом, лоша орална хигијена, депресија, старост и социоекономски показатељи су неки од фактора који су пронађени у литератури као утицајни [12-17].

Разне студије су показале да је потребно повећати промоцију оралног здравља [4,14,18-21]. Савјетује се подизање свијести трудница о важности оралног здравља као и важности одласка на преглед код стоматолога.

У прегледу литературе могле су се уочити ствари које су заједничке за већину радова, а то су: социоекономски фактори, орална хигијена и недовољна превенција одласцима на преглед код стоматолога. У већини радова се истиче и да је лоше орално здравље мајке повезано са новорођенчетом. Промоција оралног здравља и подизање свијести трудница и пружалаца стоматолошких услуга се уочавају као најбоља средства за борбу против лошег оралног здравља.

3. ОПИС СКУПА ПОДАТАКА

Скуп података је настао у оквиру посебног претходног истраживања на основу анкетног упитника који је креиран за потребе унапређења превенције каријеса раног детињства (КРД) у АП Војводини. За спровођење овог истраживања је издато одобрење Етичког одбора Института за јавно здравље Војводине. Узорачки оквир су чиниле све породице које су се породиле у периоду 1. јун – 30. новембар 2015. године.

Питања из анкетног упитника представљају обиљежја у скупу података. На нека питања је било могуће дати више одговора па су она дијељена на више обиљежја приликом уноса у скуп података. Анкетни упитник се састојао од 35 питања, а изворни скуп података од 43 обиљежја. Изворни скуп података је био сачуван у формату *xlsx* и садржао је укупно 4310 редова. Сва обиљежја, сем обиљежја везаног за старост које је нумеричког типа, категоријског су типа. Приликом анкетирања ниједна испитаница није дала одговор на свако питање. У упитнику је постојала варијанта у којој се питања прескачу у зависности од одговора. Поред овога постоји варијанта уноса података и за близанце. Овакве варијанте су могле условити појаву недостајућих вриједности у подацима.

У разговору са доменским експертом закључено је да нису потребна сва обиљежја из изворног скупа података јер нису сва релевантна за посматрани проблем. Такође закључено је да се три питања која су у вези са пушењем могу трансформисати у једно. Овим поступком добијен је подскуп од 22 обиљежја.

4. ЧИШЋЕЊЕ, ТРАНСФОРМАЦИЈА И АНАЛИЗА СКУПА ПОДАТАКА

За потребе имплементације коришћен је програмски језик *Python v3.7* у оквиру окружења *JupyterLab v3.0.16*. Коришћене су и додатне библиотеке, *Pandas v1.2.4* и *NumPy v1.19.2* за манипулацију над подацима. За визуелизацију података су употребљаване библиотеке *Matplotlib v3.4.2* и *Seaborn v0.11.1*.

4.1. Чишћење и трансформација података

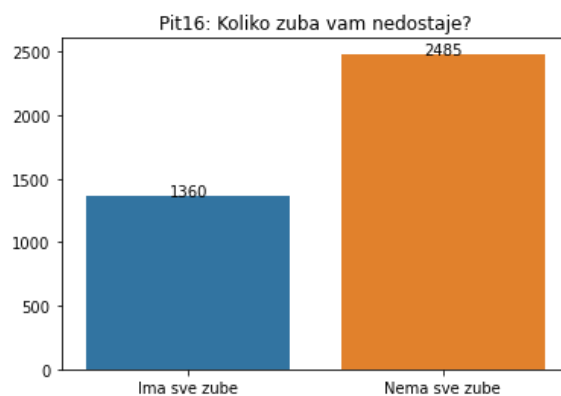
Све вриједности одговора које су погрешно унијете било је потребно замијенити недостајућим вриједностима. Подаци који представљају одговор на неко питање су заправо редни број одговора и морају бити у одређеном опсегу. Било је могуће прескакање неких питања у зависности од одговора. Требало је обезбједити да се на тим мјестима нађу недостајуће вриједности. Обиљежја која су говорила да ли је трудница пушила у неком триместру трудноће су сведена на

једно које говори да ли је пушила током читаве трудноће. Након чишћења и трансформације код појединих обиљежја се промијенио број недостајућих вредности.

Дефинисана је још једна трансформација пресликавања која се извршавала за потребе визуелизације података и покретања модела. Пресликавање се односи на разврставање одговора у двије класе. Класом 0 су означавани одговори који индицирају позитивне факторе односно добре навике труднице. Класом 1 су означавани одговори који представљају негативне факторе односно лоше навике труднице. Пошто је од интереса уочавање негативних фактора зато је и одабрано да се такви одговори пресликавају на класу 1, па се ова класа назива још и позитивном класом.

4.2. Статистичка и истражна анализа

Пошто скуп података посједује већи број обиљежја биће истакнута битна запажања. Циљна варијабла у скупу података је питање број шеснаест јер се њиме провјерава да ли трудници недостају зуби.



Слика 1. Дистрибуција циљне варијабле након пресликавања

На основу слике 1. може се закључити да у скупу података постоји знатно више, скоро па дупло више, трудница којима недостаје бар један зуб у односу на оне које имају све зубе. Првенствено са аспекта оралног здравља ово запажање је веома забрињавајуће. Због много већег броја трудница којима недостаје бар један зуб јавља се проблем небалансираности скупа података. Требало је адекватно приступити ријешавању овог проблема.

Код анализе осталих обиљежја је такође уочено да постоји небалансиран однос. За сваки одговор које су испитанице дале било је више оних које нису имале све зубе, посматрано у односу на циљну варијаблу. Међутим у оним одговорима који представљају добре навике уочено је да је однос скоро једнак.

5. ПРИПРЕМА ПРЕДИКЦИОНИХ МОДЕЛА

За имплементацију предикционих модела коришћене су и библиотека *scikit-learn v0.24.1* и библиотека *statsmodels v0.12.2*. Модел стабала одлучивања и случајне шуме су имплементирани помоћу библиотеке *scikit-learn*. За различите технике балансирања скупа података је коришћена библиотека *imbalanced-learn v0.8.0*.

Како је циљ овог рада извршити предикцију недостатка зуба код трудница у односу на њене навике тако и предикциони модели представљају централну ствар од интереса. Прије покретања модела било је неопходно спровести одређене поступке припреме који су подијељени у четири фазе.

Прву фазу представља подјела скупа података. Да би се предикциони модели могли обучити неопходан је тренинг скуп података. За евалуацију модела је потребно обезбједити валидациони скуп и на крају је потребно модел покренути над непознатим тестним подацима. Након примјене ове фазе, од прочишћеног скупа података у коме су одбачени непотребни подаци, настају три скупа: тренинг, валидациони и тест скуп података. У једном сценарију покретања (сценарио 1) је коришћена унакрсна валидација за евалуацију модела, а у другом (сценарио 2) валидациони скуп.

У другој фази је дефинисан метод евалуације модела, одабрана је референтна метрика одзив (енгл. *Recall*) у односу на коју се мјере перформансе модела.

Трећа фаза укључује дефинисање предикционих модела. За предикцију недостатка зуба код трудница одабрана су три предикциона модела: логистичка регресија, стабла одлучивања и случајна шума.

У четвртој фази је дефинисано на који начин су модели покретани. Пошто скуп података садржи недостајуће вриједности било је неопходно креирати стратегије за њихово третирање. Примјењене су четири стратегије: одбацивање свих редова који садрже недостајуће вриједности (стратегија 1), замјена недостајућих вриједности медијаном (стратегија 2), замјена недостајућих вриједности заокруженом средњом вриједношћу одговора (стратегија 3) и замјена недостајућих вриједности најчешћом вриједношћу одговора (стратегија 4). Такође дефинисан је и начин на који се ријешава проблем небалансираности скупа података. Првенствено задржани су и небалансирани подаци да би модели били покретани и над њима, а након тога су примјењене и три технике балансирања, а то су: *under-sampling* [22], *over-sampling* [22] и *SMOTEENN* [22].

6. ПРИКАЗ РЕЗУЛТАТА И ДИСКУСИЈА

За модел логистичке регресије и случајне шуме постоји могућност испитивања важности обиљежја. Ово може помоћи у прегледу фактора који би могли утицати на орално здравље труднице. У табелама 1. и 2. су приказани најбољи резултати покретаних модела, у односу на метрику одзив.

Модел случајне шуме омогућава анализу вриједности средњег смањења нечистоће гдје се гледа које од обиљежја је током развијања стабла вршило највећи утицај да определијели којој класи примјерак припада. У оба сценарија као важни фактори су се издвојили: фактори самопроцјене стања зуба труднице, фактори посјете стоматологу, фактори који се односе на навике прања зуба и економски показатељи.

Код модела логистичке регресије могуће је анализирати *p*-вриједности. Уколико је ниво значајности ове вриједности за неко обиљежје мањи од 0,05 онда се то

обиљежје узима као статистички значајно. У овом случају као статистички значајни фактори се издвајају: фактори самопроцјене стања зуба труднице, фактори посјете стоматологу, трудничке навике исхране, ношење протезе, фактор старости, едукација и фактори везани за пушење.

Видљиво је да се резултати оба модела у одређеној мјери поклапају. Поред тога резултати се поклапају једним дијелом са оним што је пронађено у литератури и већ раније навођено.

Табела 1. Приказ најбољих резултата сценарио 1

| Модел | Стратегија – техника балансирања | Мјера над валидационом скупом | Мјера над тест скупом |
|----------------------|----------------------------------|-------------------------------|-----------------------|
| | | одзив | одзив |
| Логистичка регресија | стратегија 2 – без балансирања | 0,9704 | 0,8538 |
| Случајна шума | стратегија 4 – без балансирања | 0,9125 | 0,8656 |

Табела 2. Приказ најбољих резултата сценарио 2

| Модел | Стратегија – техника балансирања | Мјера над валидационом скупом | Мјера над тест скупом |
|----------------------|----------------------------------|-------------------------------|-----------------------|
| | | одзив | одзив |
| Логистичка регресија | стратегија 3 – без балансирања | 0,8480 | 0,8300 |
| Случајна шума | стратегија 4 – без балансирања | 0,9520 | 0,8814 |

7. ЗАКЉУЧАК

Примјеном истражне анализе уочено је да велики проценат популације има проблема са оралним здрављем што је могуће и тврдити јер је узорак веома велики у односу на оно што је пронађено у литератури [11,13,15-19,21]. У случају овог упитника има дупло више трудница које су одговориле да им недостаје бар један зуб у односу на оне које имају све зубе.

Помоћу предикционих модела пронађени су потенцијално важни фактори који могу утицати на орално здравље. Код свих предикционих модела су се истакли следећи фактори: трудничке навике посјете стоматологу и самопроцјена стања зуба. Други фактори су се такође показали као потенцијално важни, а то су: навике прања зуба, економски показатељи, ниво едукације, навике конзумирања цигарета, као и поједине навике исхране.

Као примарни циљ је била истакнута предикција да ли трудница може имати проблема са недостатком зуба. У односу на остварене резултате може се рећи да су модели показали изузетно добре перформансе и да би могли дати допринос у раном откривању потенцијалних проблема са недостатком зуба. Труднице би могле попунити већ на самом почетку трудноће упитник о својим навикама. Модели би се могли покренути над прикупљеним подацима, а затим би се могла извршити предикција. Резултати предикције путем модела би трудници могли указати на потенцијалне проблеме те би је то могло потаћи да посјети стоматолога и да превенцијом не дође до проблема са оралним здрављем.

8. ЛИТЕРАТУРА

- [1] A. Sheiham, "Oral health, general health and quality of life," *Bull. World Health Organ.*, vol. 83, pp. 644–644, Sep. 2005, doi: 10.1590/S0042-96862005000900004.
- [2] P. E. Petersen, "The World Oral Health Report 2003: continuous improvement of oral health in the 21st century – the approach of the WHO Global Oral Health Programme," *Community Dent. Oral Epidemiol.*, vol. 31, no. s1, pp. 3–24, 2003, doi: 10.1046/j..2003.com122.x.
- [3] "measuring-oral-health-and-quality-of-life.pdf." Accessed: Sep. 18, 2021. [Online]. Available: <https://www.adelaide.edu.au/arcpoh/downloads/publications/reports/miscellaneous/measuring-oral-health-and-quality-of-life.pdf>
- [4] S. Vanka, A. Vanka, A. Bhambal, V. Saxena, S. Saxena, and S. Kumar, "Association of pregnant women periodontal status to preterm and low-birth weight babies: A systematic and evidence-based review," *Dent. Res. J.*, vol. 9, pp. 368–80, Jul. 2012, doi: 10.4103/1735-3327.102757.
- [5] V. Marla, R. Chettiar, D. Roy, and H. Ajmera, "The Importance of Oral Health during Pregnancy: A review," vol. 5, Mar. 2018, doi: 10.5935/MedicalExpress.2018.mr.002.
- [6] M. Pirie, I. Cooke, G. Linden, and C. Irwin, "Dental manifestations of pregnancy," *Obstet. Gynaecol.*, vol. 9, no. 1, pp. 21–26, 2007, doi: 10.1576/toag.9.1.021.27292.
- [7] L. W. Mills and D. T. Moses, "Oral Health During Pregnancy," *MCN Am. J. Matern. Nurs.*, vol. 27, no. 5, pp. 275–280, Oct. 2002.
- [8] "Uredba o Nacionalnom programu preventivne stomatološke zdravstvene zaštite: 22/2009-11." <https://www.pravno-informacioni-sistem.rs/SIGlasnikPortal/eli/rep/sgrs/vlada/uredba/2009/22/4/reg> (accessed Sep. 18, 2021).
- [9] M. Vulović *et al.*, *Program preventivne stomatološke zaštite stanovnika Srbije*. Zavod za udžbenike i nastavna sredstva, 1996.
- [10] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, Illustrated edition. Cambridge, MA: The MIT Press, 2012.
- [11] E. Kateeb and E. Momany, "Dental caries experience and associated risk indicators among Palestinian pregnant women in the Jerusalem area: a cross-sectional study," *BMC Oral Health*, vol. 18, no. 1, p. 170, Oct. 2018, doi: 10.1186/s12903-018-0628-x.
- [12] S. L. Russell, J. R. Ickovics, and R. A. Yaffee, "Exploring Potential Pathways Between Parity and Tooth Loss Among American Women," *Am. J. Public Health*, vol. 98, no. 7, pp. 1263–1270, Jul. 2008, doi: 10.2105/AJPH.2007.124735.
- [13] M. Deghatipour *et al.*, "Oral health status in relation to socioeconomic and behavioral factors among pregnant women: a community-based cross-sectional study," *BMC Oral Health*, vol. 19, no. 1, p. 117, Jun. 2019, doi: 10.1186/s12903-019-0801-x.
- [14] E. G. Kim, S. K. Park, and J.-H. Nho, "Factors Related to Maternal Oral Health Status: Focus on Pregnant and Breastfeeding Women," *Healthcare*, vol. 9, no. 6, Art. no. 6, Jun. 2021, doi: 10.3390/healthcare9060708.
- [15] J. A. Gil-Montoya, X. Leon-Rios, T. Rivero, M. Expósito-Ruiz, I. Perez-Castillo, and M. J. Aguilar-Cordero, "Factors associated with oral health-related quality of life during pregnancy: a prospective observational study," *Qual. Life Res.*, May 2021, doi: 10.1007/s11136-021-02869-3.
- [16] M. Radnai *et al.*, "The oral health status of postpartum mothers in South-East Hungary," *Community Dent. Health*, vol. 24, no. 2, pp. 111–116, Jun. 2007.
- [17] M. Wandera, I. M. Engebretsen, I. Okullo, J. K. Tumwine, A. N. Åstrøm, and the PROMISE-EBF Study Group, "Socio-demographic factors related to periodontal status and tooth loss of pregnant women in Mbale district, Uganda," *BMC Oral Health*, vol. 9, no. 1, p. 18, Jul. 2009, doi: 10.1186/1472-6831-9-18.
- [18] J.-N. Vergnes *et al.*, "Frequency and Risk Indicators of Tooth Decay among Pregnant Women in France: A Cross-Sectional Analysis," *PLOS ONE*, vol. 7, no. 5, p. e33296, Jul. 2012, doi: 10.1371/journal.pone.0033296.
- [19] N. Thomas, P. Middleton, and C. Crowther, "Oral and dental health care practices in pregnant women in Australia: A postnatal survey," *BMC Pregnancy Childbirth*, vol. 8, p. 13, Feb. 2008, doi: 10.1186/1471-2393-8-13.
- [20] A. Villa *et al.*, "Oral health and oral diseases in pregnancy: a multicentre survey of Italian postpartum women," *Aust. Dent. J.*, vol. 58, no. 2, pp. 224–229, 2013, doi: 10.1111/adj.12058.
- [21] S. A. Khamis, K. Asimakopoulou, T. Newton, and B. Daly, "The effect of dental health education on pregnant women's adherence with toothbrushing and flossing — A randomized control trial," *Community Dent. Oral Epidemiol.*, vol. 45, no. 5, pp. 469–477, 2017, doi: 10.1111/cdoe.12311.
- [22] "imbalanced-learn documentation — Version 0.8.0." <https://imbalanced-learn.org/stable/> (accessed May 25, 2021).

Кратка биографија:



Мирослав Томић рођен је 1997. године у Бијељини, РС, БиХ. Завршио је средњу техничку школу „Михајло Пупин“ у Бијељини. Факултет техничких наука у Новом Саду, студијски програм Рачунарство и аутоматика, уписао је 2016. године. Након завршених основних студија уписао је мастер студије на истом факултету.