

**SOFTVERSKO REŠENJE ZA DETEKCIJU ANOMALIJA POTROŠNJE ELEKTRIČNE ENERGIJE U ML .NET-U I PYTHON-U****SOFTWARE SOLUTION FOR THE DETECTION OF ELECTRICAL ENERGY CONSUMPTION ANOMALIES IN ML .NET AND PYTHON**

Nina Grbić, *Fakultet tehničkih nauka, Novi Sad*

**Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

**Kratka sadržaj** – Podatke iz Smart Grid sistema moguće je analizirati za detekciju abnormalnih pojava u različitim oblastima poput sajber bezbednosti, detekcije krađe, detekcije kvara i slično. Algoritmi mašinskog učenja mogu poslužiti kao alat za analizu i obradu sirovih podataka koji dolaze iz sistema u realnom vremenu. Kroz ovaj su rad opisani, analizirani i prikazani rezultati implementiranog softverskog rešenja koje vrši detekciju anomalija na osnovu podataka o potrošnji električne energije nad algoritmima u ML .NET-u i Python-u.

**Ključne reči:** Detekcija anomalija, Mašinsko učenje, ML .NET, Python

**Abstract** – Data from the Smart Grid system can be analyzed to detect abnormal phenomena in various areas such as cyber security, theft detection, failure detection, etc. Machine learning algorithms can be used as a tool for analyzing and processing raw data coming from real-time systems. This paper describes, analyzes and presents the results of the implemented software solution that detects anomalies based on data containing electricity consumption over algorithms in ML .NET and Python.

**Keywords:** Anomaly detection, Machine learning, ML .NET, Python

**1. UVOD**

Gubici koji se javljaju u elektroenergetskim sistemima (EES) u velikoj meri utiču na stabilnost mreže, kvalitet napajanja kao i procenat troškova u konačnoj ceni električne energije. Ukupni distributivni gubici Elektroprivrede Srbije (EPS) su među najvećim u Evropi i kao direktna posledica postoji značajna potreba da se tehnički gubici na mreži i krađe električne energije svedu na prihvatljivu meru **Error! Reference source not found.**

Ovo predstavlja jasan pokazatelj da tradicionalan pristup upravljanja električnom mrežom mora da se zameni sa inovativnim tehnologijama koje će optimizovati procese prilikom nadgledanja električne mreže. U softver za monitoring elektroenergetskog sistema moguće je integrisati takvo rešenje da algoritmima mašinskog učenja i analizom podataka se detektuju oblasti ili potrošači sa neuobičajenim karakteristikama ili potrošnjom.

**NAPOMENA:**

**Ovaj rad proistekao je iz master rada čiji mentor je bio dr Sebastijan Stoja, docent.**

Najpopularnija tehnologija za implementaciju algoritama mašinskog učenja jeste Python odnosno njegova moćna biblioteka SciKit-learn. Međutim, Microsoft 2019. godine objavljuje svoju tehnologiju pod nazivom ML .NET sa ciljem da bude konkurentan na tržištu tehnologija za integrisanje modela mašinskog učenja u softverska rešenja. Glavna motivacija ovog rada jeste analiza preciznosti algoritama za detekciju anomalija u krivama potrošnje uz pomoć algoritama mašinskog učenja u ML .NET-u i Python-u.

**2. TEORIJSKE OSNOVE****2.1. Elektroenergetski sistemi****2.1.1. Smart Grid**

Smart Grid se odnosi na električne mreže sledeće generacije zasnovane na digitalnoj tehnologiji za snabevanje potrošača električnom energijom putem dvosmerne digitalne komunikacije. Ovaj sistem isporučuje električnu energiju između potrošača i dobavljača i kontroliše digitalne uređaje kako bi obezbedio uštedu energije uz smanjenje potrošnje energije i troškova, a maksimiziranje transparentnosti i pouzdanosti lanca snabevanja energijom [2].

Pored tradicionalnih proizvodnih objekata i prenosne mreže, Smart Grid se sastoji od tri nove komponente: pametni kontrolni i merni uređaji, digitalni komunikacioni sistemi i računarski softverski program. Pametni uređaji podrazumevaju kompjuterski kontrolisane generatore i druge izvore energije, kao i brojila i inteligentne elektronske uređaje koji prikupljaju informacije o potražnji električne energije, njenoj dostupnosti iz različitih izvora, kapacitetu isporuke svakog dela mreže i protoku energije [3].

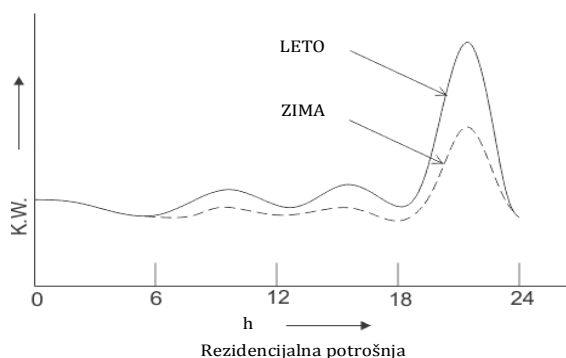
**2.1.2. Pametna brojila**

Pametna brojila (eng. *Smart Meter*) su ključan faktor u Smart Grid sistemu i čini opremu koja prati potrošnju električne energije u predefinisanim intervalima. Digitalna brojila zajedno sa sistemom za automatsko očitavanje merenja (AMR), obradom i arhiviranjem merenih podataka i automatskim upravljanjem merenjima (AMM) čini integrisani sistem za naprednu infrastrukturu za merenje (AMI) koja danas u velikom broju zamenjuje tradicionalna brojila [4].

U naprednim sistemima za upravljanje EES, praćenje potrošnje u realnom vremenu omogućava i ažuriranje krive potrošnje korisnika kako bi se dobila preciznija kriva sa rezultatima sa polja. Vrednosti sa potrošačke krive se koriste prilikom izvršavanja glavnih funkcija proračuna stanja mreže.

### 2.1.3. Krive potrošnje

Kriva potrošnje predstavlja dijagram potražnje za snagom (y osa) u odnosu na vreme (x osa) u hronološkom redu. Ova kriva reprezentuje varijaciju opterećenja nekog elementa elektroenergetskog sistema (npr. potrošača) u odnosu na vreme (Slika 1).



Slika 1. Primer krive potrošnje

### 2.1.4. Gubici električne energije

Gubici električne energije u EES su jasan pokazatelj efikasnosti rada infrastrukture mreže. Gubici na nivou prenosa i distribucije se mogu podeliti na tehničke i netehničke gubitke. Tehnički gubici su neizbežni, jer su rezultat fizičkog procesa prenosa odnosno transformacije električne energije, kvarova ili problema sa opremom i infrastrukturom tokom procesa prenosa i distribucije [5]. Za razliku od tehničkih, netehnički gubici nastaju kao posledica situacija koje je moguće kontrolisati poput nenaplaćenu ili pogrešno naplaćenju potrošnju električne energije, kao i nezakonitu upotrebu električne energije.

## 2.2. Mašinsko učenje

Mašinsko učenje predstavlja proces analize podataka koji korišćenjem posebno podešenih algoritama i iterativne obrade podataka na ulazu, omogućavajući sistemu koji implementira da otkrije skrivena znanja u modelu podataka, bez eksplicitnog programiranja [6]. Glavna ideja mašinskog učenja predstavlja razvoj matematičkog modela koji će analizom i obradom postojećih podataka generisati zaključke i predviđanja o novim podacima. Ako to izrazimo matematički, pokušavamo da aproksimiramo funkciju mapiranja –  $f$  od ulaznih promenljivih  $y$  u izlazne promenjive  $x(1)$ :

$$f(x) = y \quad (1)$$

Kreiranje modela mašinskog učenja se svodi na sledeće korake:

- Analiza i priprema podataka – prikupljanje atributa, analiza značenja atributa i obeležja, redukcija suvišnih atributa ili podataka koji unose šum, normalizacija vrednosti atributa kako bi rezultovali definisanjem krajnjeg skupa podataka koji će se koristiti za treniranje, validaciju i testiranje modela mašinskog učenja.
- Primena odgovarajućeg algoritma i podešenje parametara – konfigurisanje modela biranjem odgovarajućeg tipa algoritma i definisanjem neophodnih parametara ili hiperparametara.
- Evaluacija – eksperimentalna evaluacija se sprovodi radi određivanja performansi rada nad istim skupom podataka, pruža uvid u tačnost i

preciznost matematičkog modela i vrše se iterativne korekcije u konfiguraciji parametara kako bi dobili bolje rezultate [7].

### 2.2.1. Algoritmi mašinskog učenja

Algoritmi mašinskog učenja se najčešće kategorizuju po načinu na koji algoritam vrši predikcije:

- Nadgledano obučavanje – sistem se obučava uz pomoć obeležja podataka, algoritam generiše funkciju koja mapira ulazne podatke na željene izlaze.
- Nenadgledano obučavanje – sistem se obučava uz pomoć ulaznih podataka i bez dobijanja obeležja utvrđuje glavne karakteristike podataka i formira prirodne klustere.
- Polunadgledano obučavanje – obeležja podataka su dostupna samo za normalne podatke, dok za obučavanje koristi se modifikovani model klasifikacije čime detektuju instance podataka koje se ne uklapaju u predviđeni šablon.
- Pojačano obučavanje – sistem obučava softverske agente koje akcije da preduzmu u svom okruženju. Učenje se vrši na osnovu datih ulaznih podataka i signala podrške [8].

### 2.2.2. Detekcija anomalija

Anomalijom se smatra tačkom podataka koja se značajno razlikuje od drugih tačaka u datom skupu. Detekcija anomalija predstavlja proces pronalaženja odstupanja u podacima. Anomalije se u globalnom smislu mogu podeliti u 3 glavne kategorije opisane uz primere u tabeli:

Tabela 1. Osnovna kategorizacija anomalija

Kategorija	Opis
<u>Tačkaste anomalije</u>	Određena instanca podatka koja u velikoj meri odstupa u odnosu na očekivano ponašanje podataka sistema
<u>Kontekstualne anomalije</u>	Instance odstupanja podataka koje u velikoj meri zavise od konteksta podataka.
<u>Zbirne anomalije</u>	Analiza pojedinačne instance ne predstavlja anomaliju, ali njihovo grupisanje se smatra neuobičajenim ponašanjem.

## 3. METODOLOGIJA

### 3.1 Formulacija problema

Kreirano je softversko rešenje u *ML .NET*-u sa njegovim implementiranim algoritmima za detekciju anomalija, kao i rešenje u *Python*-u sa *Isolation Forest* odabranim algoritmom za detekciju anomalija na osnovu radova [9][10]. Od podataka kao ulaz u model mašinskog učenja koristi se fajl koji sadrži 24-časovna očitavanja električne energije niskonaponskih potrošača proizvoljno generisanih. Analiziraju se rezultati dobijeni kao izlaz iz svih funkcija, kao i eksperimentalna evaluacija radi određivanja performansi algoritama nad istim ulazom.

### 3.2. Odabrani algoritmi

#### 3.2.1. ML .NET

*DetectSpikeBySSA* – SSA je skraćena za analizu singularnog spektra i predstavlja složenu metodu analize koja kombinuje delove standardne analize podataka

vremenske serije, multivarijantnu statistiku, multivarijantnu geometriju, dinamičke sisteme i obradu signala [11].

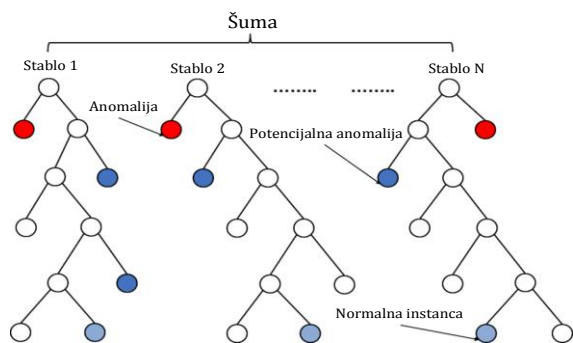
**DetectIidSpike** – U teoriji verovatnoće i statistici, kolekcija slučajnih promenljivih je nezavisna i identično raspoređena ako svaka slučajna promenljiva ima istu distribuciju verovatnoće kao i druge i sve su međusobno nezavisne [12].

**DetectEntireAnomalyBySrCnn** – Algoritam konvolucione neuronske mreže super rezolucije Sr-Cnn (eng. *Super Resolution Convolutional Neural Network*) koristi kompjuterski pogled na problem detekcije anomalija. Sr-Cnn pozajmljuje model super rezolucije iz domena otkrivanja vizuelne istaknutosti i primenjuje ga na otkrivanje anomalija u tipovima podataka poput vremenskih serija [13].

**RandomizedPCA** – Randomizovani PCA (eng. *Principal Component Analysis*) algoritam funkcioniše tako što traži korelacije između varijabli i utvrđuje kombinaciju vrednosti koja najbolje predstavlja razlike u rezultatima. Zatim se ove kombinovane vrednosti obeležja koriste za kreiranje kompaktnijeg skupa obeležja, koji se naziva glavnim komponentama. Randomizovani PCA označava približni model analize glavnih komponenti koji koristi algoritam randomizovane dekompozicije singularnih vrednosti [14].

### 3.2.2. Python

**Isolation Forest** izoluje zapažanja tako što nasumično bira obeležje, a zatim nasumično određuje vrednost između maksimalne i minimalne vrednosti izabranog obeležja.



Slika 2. Vizuelni prikaz detekcije anomalija koristeći *Isolation Forest* algoritam

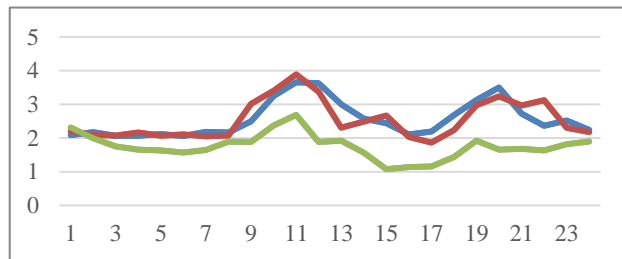
Pošto se rekursivno particionisanje može predstaviti skrukturom stabla, broj particionisanja potrebnih za izolovanje uzorka je jednak dužini putanje od korenskog čvora do završnog čvora. Nasumično particionisanje proizvodi приметно kraće putanje za anomalije. Stoga, ukoliko skup nasumičnih stabala zajedno proizvodi kraće dužine puta za određene uzorke, veća je verovatnoća da će te tačke biti okarakterisane kao anomalije [15].

### 3.3. Podaci

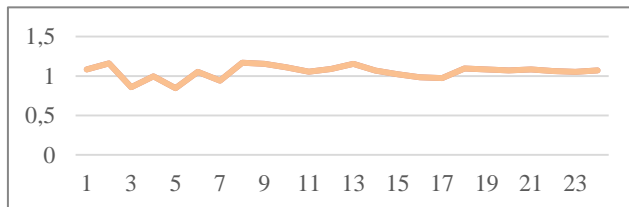
U sistemima koji rade u realnom vremenu, jedan od glavnih poslovnih zahteva predstavlja praćenje ključnih signala kako bismo bili sigurni da prate očekivane obrasce ponašanja. Podaci ove prirode su poznati kao vremenske serije (eng. *time series*) i predstavljaju hronološki uređen niz podataka pojave koja se izučava. Pojavu od interesa najčešće pratimo po danima, mesecima, kvartalima, godinama itd.

Za potrebe analize problema, podaci su izgenerisani uz pomoć ručno kreirane skripte na osnovu podataka standardne krive potrošnje. Kriva potrošnje je predstavljena sa 24 tačke gde svaka tačka reprezentuje očitane vrednost potrošnje u datom satu u kW. Da bi se adekvatno evaluirali razmatrani modeli mašinskog učenja, ručno su kreirani različiti problemi koji su tipični u analizi podataka dnevne potrošnje poput:

- **Odsecanje pikova potrošnje** – predstavlja jedan od načina upravljanja troškovima komunalnih usluga tako što se eliminišu kratkoročni skokovi potražnje odgovorne za visoke troškove potrošnje električne energije u datom momentu.
- **Profilisanje potrošnje** – algoritmima mašinskog učenja moguće je utvrditi i nove tipove potrošača u sistemu. Naknadom analizom rezultata novih potrošačkih krivi, distribucija utvrđuje validnost novodetektovanih potrošača ili potencijalnu malverzaciju podacima očitavanja električne energije.
- U podacima su ručno uneta 3 primera generisanih novih tipova potrošača (Slika 3) a 1 slučaj ručno simuliranog odsecanja pikova potrošnje (Slika 4).



Slika 3. Ručno generisani primeri novih tipova potrošača

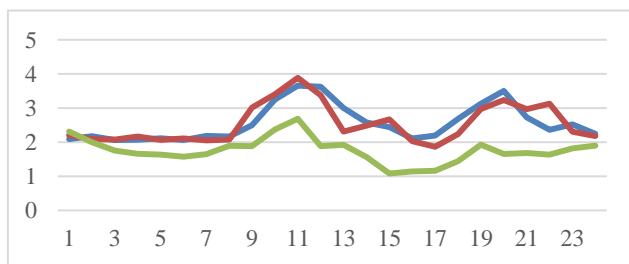


Slika 4. Ručno generisan primer odsecanja potrošnje

Uz pomoć ovih primera evaluiira se uspešnost algoritama za detekciju anomalija u krivama potrošnje.

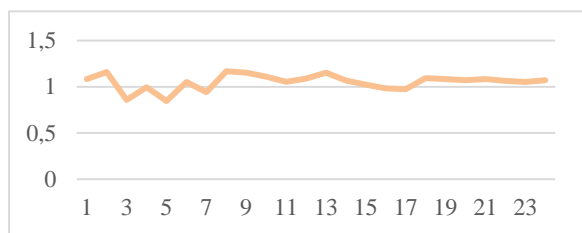
## 4. REZULTATI REŠENJA

U ML .NET-u najuspešnije se pokazala *DetectEntireAnomalyBySrCnn* funkcija. Ona je detektovala sve nove tipove potrošača ali nije detektovala odsecanje potrošnje kao vid anomalije u ulaznim podacima (Slika 5).



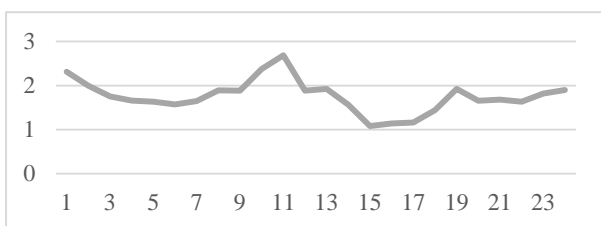
Slika 5. Rezultati funkcije *DetectEntireAnomalyBySrCnn* u ML .NET-u

Funkcija *DetectSpikeBySSA* daje rezultate nasuprot prethodno testirane funkcije i kao rezultat prikazuje samo primer odsecanja potrošnje (Slika 6).

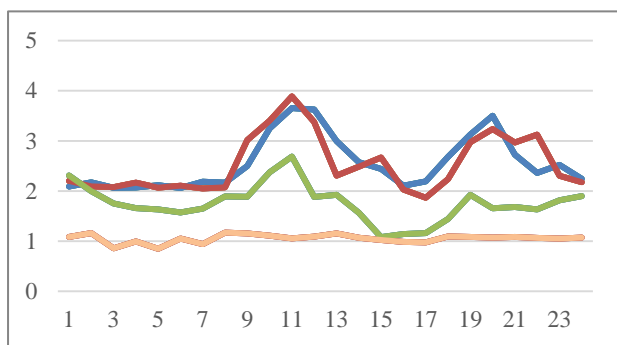


Slika 6. Rezultati funkcije *DetectSpikeBySSA* u *ML .NET-u*

*DetectLidSpike* nije dao pozitivne rezultate ni za jedan primer anomalije u ulaznom skupu podataka i nije dalje razmatran. *DetectSpikeBySSA*, *DetectLidSpike* i *DetectEntireAnomalyBySrCnn* funkcije kao ulaz primaju jednodimenzioni niz, što je zahtevalo transformaciju 24 dimenzije podataka vremenske serije na jednu dimenziju po potrošaču. Prilikom smanjenja dimenzionalnosti, gubimo potencijalno bitne karakteristike podataka što može uticati na performanse i preciznost algoritama. Izuzetak prethodno istaknutog problema *ML .NET* je izbegao u implementaciji *RandomizedPCA* algoritma. Međutim, ova metoda pokazuje najlošije rezultate jer je detektovala samo 1 novi tip potrošača od ukupno 3, a nije detektovala odsecanje pikova potrošnje (Slika 7).



Slika 7. Rezultati funkcije *RandomizedPCA* u *ML .NET-u*  
*Python* je opravdao svoju vodeću poziciju na tržištu tehnologija za integrisanje modela mašinskog učenja u softverska rešenja, i pokazao najbolje rezultate sa detektovanim svim novim tipovima potrošnje, kao i primer odsecanja pikova potrošnje (Slika 8).



Slika 8. Rezultati funkcije *IsolationForest* u *Python-u*

## 5. ZAKLJUČAK

U ovom radu, analizirani su algoritmi za detekciju anomalija u *Python-u* i *ML .NET-u*. Rezultati su pokazali da *Python*-ova implementacija *Isolation Forest* algoritma pruža najbolje rezultate, dok kod *ML .NET*-a možemo uočiti detekcije pojedinih slučajeva ali ni jedan zaseban algoritam ne detektuje sve anomalije sa generisanim

skupom podataka. Dalji pravci unapređenja softverskog rešenja za detekciju anomalija u *ML .NET-u* mogu biti fokusirani na kreiranje hibridnog modela koji će uz pomoć više implementacija algoritama imati model mašinskog učenja koji može da pokrije sve slučajeve neuobičajenog ponašanja. Jedno od zanimljivih rešenja bi bilo i istraživanje benefita hibridne softverske implementacije *Python* i *ML .NET* algoritama za analizu i pripremu podataka, tako i rešavanje problema uz pomoć algoritama mašinskog učenja.

## 6. LITERATURA

- [1] <https://bif.rs/2021/12/ukupni-distributivni-gubici-eps-a-su-medju-najvecim-u-evropi/> (pristupljeno u martu 2023.)
- [2] Raja Masood Larik, Mohd Wazir Mustafa, Sajid Hussain Qazi, „*Smart Grid Technologies in Power System*“, University of Technology Malaysia, Malaysia
- [3] <https://www.techopedia.com/definition/692/smart-grid> (pristupljeno u martu 2023.)
- [4] Srdjan Milošević, Emil Naumovski, „*Sistem za daljinsko očitavanje i upravljanje potrošnjom u PD „Elektrodistribucija Beograd*“, Zbornik Međunarodnog kongresa o KGH, [S.l.], v. 44, n. 1, p. 1-6, oct. 2017
- [5] <https://anyline.com/news/detect-non-technical-losses-energy-utility> (pristupljeno u martu 2023.)
- [6] Bishop, C. M., „*Pattern Recognition and Machine Learning*, Springer“
- [7] Nenad Rakić, „*Primena metoda mašinskog učenja za rangiranje individualnih sposobnosti*“, Prirodni Matematički Fakultet, Univerzitet u Novom Sadu, 2020,
- [8] as. ms Vladimir Jocović, as. ms Adrian Milaković, „*Inteligentni sistemi*“, Elektrotehnički fakultet, Univerzitet u Beogradu
- [9] <https://blog.paperspace.com/anomaly-detection-isolation-forest/> (pristupljeno u martu 2023.)
- [10] Kumar Reddy Shabad, Abdulmuen Alrshide, „*Anomaly Detection in Smart Grids using Machine Learning*, Miami, Florida, USA, 2021.
- [11] <https://learn.microsoft.com/en-us/dotnet/api/microsoft.ml.timeseriescatalog.detectspikebyssa?view=ml-dotnet> (pristupljeno u martu 2023.)
- [12] [https://en.wikipedia.org/wiki/Independent\\_and\\_identically\\_distributed\\_random\\_variables](https://en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables) (pristupljeno u martu 2023.)
- [13] <https://towardsdatascience.com/time-series-anomaly-detection-b10fdb542974> (pristupljeno u martu 2023.)
- [14] Vladimir Rokhlin, Arthur Szlam, Mark Tygert, „*A randomized algorithm for principal component analysis*“, Jul 2009.
- [15] <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html> (pristupljeno u martu 2023.)

## Kratka biografija:



**Nina Grbić** rođena je u Novom Sadu 1996. godine. Osnovne akademske studije na Fakultetu tehničkih nauka Univerziteta u Novom Sadu upisala je 2015. godine. Diplomirala je 2019. godine na smeru Primenjeno softversko inženjerstvo i iste godine upisala master akademske studije na istom smeru.