

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата Стефана Костића, мастер инжењера електротехнике и рачунарства, за израду докторске дисертације

Одлуком Наставно-научног већа Електротехничког факултета бр. 852 од 09.09.2020. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата Стефана Костића, мастер инжењера електротехнике и рачунарства, за израду докторске дисертације и научне заснованости теме **„Предикција губитка корисника у мобилним телекомуникационим мрежама примене ненадгледаног машинског учења“**.

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Стефан Костић је рођен 29.09.1990. године у Београду. Основну школу „Вељко Дугошевић“ завршава са одличним успехом (просек оцена 5,00) и Вуковом наградом, те као ђак генерације. Носилац је више награда на савезном и републичком нивоу из физике и математике. Средњу школу „Математичка гимназија“ завршава са одличним успехом (просек оцена 5,00) и Вуковом наградом. Основне академске студије на Електротехничком факултету у Београду уписао је 2009. године. Дипломирао је на одсеку за Телекомуникације и информационе технологије, смер Системско инжењерство 2013. године, са просечном оценом 9,18. Дипломски рад на тему „Принципи UWB технологије“ одбранио је са оценом 10. Мастер академске студије уписао је 2013. године на Електротехничком факултету, модул Системско инжењерство и радио комуникације, и завршио их је 2014. године са просечном оценом 9,83. Мастер рад на тему „Принципи компресије аудио и видео сигнала у савременим телекомуникационим системима“ одбранио је са оценом 10.

Докторске академске студије на Електротехничком факултету у Београду, модул Телекомуникације, уписао је 2014. године, где је положио све испите са просечном оценом 10. Области истраживања током докторских студија обухватају припрему и анализу података, алгоритме надгледаног и ненадгледаног машинског учења, као и науку о подацима у целини.

Стефан Костић је од октобра 2013. до маја 2015. године био запослен у фирми „P3 Communications“ као инжењер за верификацију података, где му је свакодневна обавеза била анализа записа на релацији мобилни уређај - базна станица ради откривања потенцијалних проблема у конфигурацији мрежних елемената. Од јуна 2015. до маја 2016. године, Стефан

Костић је био запослен у фирми „Телеком Србија“ као софтверски инжењер. На тој позицији, бавио се подршком пословној интелигенцији у бизнису, као и моделовањем и анализом података које телекомуникациони оператер прикупља у редовном раду. Од јуна 2016. до октобра 2016. године, Стефан Костић је био запослен у фирми „Ibis-Instruments“ као аналитичар великих података где се бавио анализом података из сензорских мрежа. Од новембра 2016. до априла 2019. године, Стефан Костић је био запослен у фирми „Ерсте банка“ као *Data Scientist*, односно као *Lead Data Scientist*. У том периоду, између осталог, бавио се и моделовањем података за регулаторне потребе, креирањем стрес тестова пословања банке, моделовањем података за подршку пословној интелигенцији и продаји, као и различитим пројектима везаним за пословање банке на локалном и регионалном нивоу. Од маја 2019. запослен је у фирми „United Group“ на позицији *Lead Data Scientist* где је наставио са радом на моделовању различитих врста података кабловских и мобилних оператера у земљи и региону југоисточне Европе.

Стефан Костић је аутор рада објављеног у научном часопису међународног значаја са SCI листе категорије M22. Аутор је и два рада објављена на конференцијама од међународног значаја, као и два рада објављена на конференцијама од националног значаја.

1.2. Сечено научноистраживачко искуство

Стефан Костић је докторске студије на Електротехничком факултету уписао школске 2014/2015. године на модулу Телекомуникације. Током докторских студија положио је све наставним планом и програмом предвиђене испите и испунио све обавезе и то:

Р. Бр.	Назив испита	Оцена	ЕСПБ
1	Когнитивне радио мреже	10	9
2	Случајни процеси у телекомуникацијама	10	9
3	Вештачка интелигенција	10	9
4	Специјалне функције	10	6
5	Моделовање саобраћаја у M2M комуникационим системима	10	9
6	Савремене технике контроле грешака у телекомуникацијама	10	9
7	Енглески језик	10	9
8	Функционална анализа и примене	10	9
9	Неуралне мреже	10	9
10	Синхронизација у телекомуникационим мрежама	10	9
Назив обавезе			
1	Научно стручни рад		3
2	Студијски истраживачки рад I		15
3	Студијски истраживачки рад II		15

Списак публикација кандидата Стефана Костића, у складу са нормативима Министарства просвете и науке Србије, наведен је у даљем тексту:

Радови у међународним часописима – M22:

- M22.1 Stefan Kostić, Mirjana Simić, Miroljub Kostić, “*Social Network Analysis and Churn Prediction in Telecommunications Using Graph Theory*,” Entropy (ISSN 1099-4300), 2020, 22, 753. (IF2019=2.494).

Радови на међународним скуповима – M33

- M33.1 Stefan Kostić, Miloš Đuričić, Mirjana Simić, Miroljub Kostić, “*Data Mining and Modeling use Case in Banking Industry*,” IEEE 26th Telecommunications Forum (TELFOR), Belgrade, Serbia, November 2018, pp. 1-4.
- M33.2 Jelena Sretenović, Stefan Kostić, Mirjana Simić. “*Experimental analysis of weight-compensated weighted centroid localization algorithm based on RSSI*,” IEEE 12th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), Niš, Serbia, October 2015, pp. 373-376.

Радови саопштени на скуповима националног значаја – M63

- M63.1 Stefan Kostić, Jelena Sretenović, Mirjana Simić, Miroljub Kostić, “*Detekcija ekstremnih korisnika u telekomunikacionim mrežama pomoću analize socijalnih mreža*,” 60. konferencija za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku (ETRAN), Zlatibor, Srbija, jun 2016, str. TE1.1. 1-4.
- M63.2 Jelena Sretenović, Stefan Kostić, Mirjana Simić, “*Eksperimentalna analiza pozicioniranja primenom centroid metode sa težinskim koeficijentima*”, 59. konferencija za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku (ETRAN), Srebrno jezero, Srbija, jun 2015, str. TE1.4. 1-4.

Дана 01.10.2020. године на Електротехничком факултету у Београду кандидат Стефан Костић имао је јавни усмени испит о подобности теме пред трочланом комисијом у саставу:

1. др Милан Бјелица, редовни професор, Универзитет у Београду - Електротехнички факултет
2. др Ненад С. Митић, редовни професор, Универзитет у Београду – Математички факултет
3. др Бошко Николић, редовни професор, Универзитет у Београду - Електротехнички факултет.

Комисија је закључила да кандидат испуњава све услове за рад на предложеној теми докторске дисертације и оценила испит оценом **задовољно**.

1.3. Оцена подобности кандидата за рад на предложеној теми

На основу претходног приказа резултата досадашњег рада, кандидат Стефан Костић, узимајући у обзир све претходно наведене чињенице, Комисија је стекла увид у способности и научно-истраживачки рад кандидата те констатује да је исти:

- Положио све испите и извршио све обавезе предвиђене наставним планом и програмом и тиме остварио 120 ЕСПБ бодова,
- Јавно одбранио тему докторске дисертације 01.10.2020. године пред трочланом комисијом формираном од стране Наставно-научног већа Електротехничког факултета у Београду,
- Публиковао више радова у научним часописима и научним скуповима из области којој припада предложена тема докторске дисертације чиме је показао квалификације за научни рад и афинитет према теми докторске дисертације.

Стога је став Комисије да кандидат Стефан Костић задовољава све услове за рад на предложеној теми.

2. Предмет и циљ истраживања

Последњих година, за телекомуникационе операторе широм света постало је витално да увек имају актуелан увид у понашање и навике својих корисника. У почетку, раст тржишта телекомуникација, првенствено мобилних, био је експоненцијални, па проблем губитка корисника телекомуникационих услуга унутар једне телекомуникационе мреже тј. оператора уопште није било потребно разматрати. Данас, услед засићења телекомуникационог тржишта, повећана конкуренција између мобилних оператора доводи до тога да се они више фокусирају на задржавање постојећих корисника а мање на привлачењу нових. Предвиђање вероватноће губитка корисника је проблем бинарне класификације у којем су сви корисници неког телекомуникационог оператора подељени у две групе. Једна група је означена као високо ризична група клијената (у смислу губитка корисника) која захтева неку реакцију, док друга група подразумева стандардне кориснике који нису у непосредној опасности да напусте телекомуникационог оператора. Додатно, губитак корисника може бити подељен у две основне категорије: вољни и невољни [1]. Невољни губитак корисника подразумева ситуацију у којој телекомуникациони оператор одлучује да раскине уговор са својим корисником. То се може десити из више различитих разлога, као на пример: малверзације корисника, неизмиривање обавеза корисника према телекомуникационом оператору, и слично. Са друге стране, вољни губитак корисника подразумева свесну одлуку корисника да напусти телекомуникационог оператора. Та одлука може бити проистећи из различитих разлога, при којима је за телекомуникационог оператора најпроблематичнији онај у којем се корисник одлучује за другог телекомуникационог оператора услед боље и/или приступачније понуде. Предмет истраживања ове докторске дисертације представља предикцију вероватноће вољног екстерног губитка корисника у оквиру мобилних телекомуникационих мрежа.

У литератури постоје различити алгоритми за предикцију губитка корисника у мобилним телекомуникацијама [2-5]. Од посебног значаја су они алгоритми базирани на широком спектру улазних података, као што регресиони модел [2], модел стабла одлучивања [3,4], кластеризација [5], неуралне мреже [4]. Неки од уобичајених улазних података у оквиру ових алгоритама су подаци о уговорима, позивима, задовољству корисника, демографији корисника, квалитету услуге који је доступан кориснику и слично.

Последњих година један од честих начина предикције губитка корисника у телекомуникацијама јесте применом науке о подацима. Наиме, у савремено доба, услед брзог развоја рачунара и информационих система у целини, количина података која се може забележити постаје изузетна. Главни изазов више није везан само за смештање велике количине података већ и изузетно велика брзина креирања нових података као и генерисања корисних информација на основу велике количине података. Сам циљ науке о подацима је проналажење образаца које дефинишу ови подаци, узимајући у обзир да ти образци морају бити такви да је њихово појављивање у будућности врло вероватно и очекивано [6]. Један од

начина за лоцирање ових образаца представљају алгоритми машинског учења. Постоје различити механизми машинског учења који се могу користити у науци о подацима. Основна подела би била базирана на надгледане (*supervised*) и ненадгледане (*unsupervised*) алгоритме [7]. Предиктивни ненадгледани алгоритам машинског учења, познат као кластеризација (анализа кластера), ће бити окосница истраживања у предикцији губитка корисника у телекомуникационим мрежама у оквиру ове докторске дисертације. Кластеризација подразумева групу техника које имају за циљ поделу података у групе на начин да опсервације унутар једне групе буду међусобно што сличније, док су истовремено што различитије од остатка опсервација у иницијалном скупу података. Телекомуникациона мрежа биће посматрана као друштвена мрежа, формирана применом теорије графова. У истраживању ће бити коришћена Вардова метода анализе кластера [8], која ће обрађивати податке добијене анализом графа позива телекомуникационог оператора [9]. Метрике теорије графова [10] ће бити коришћене ради квантификовања значаја и утицаја сваког појединачног корисника мреже телекомуникационог оператора.

Циљ истраживања у оквиру ове докторске дисертације је показати да је могуће, на основу постојећих корисника који напуштају посматраног телекомуникационог оператора и њихових образаца комуникације, проактивно предвиђати нове кориснике који имају високу вероватноћу напуштања посматраног телекомуникационог оператора. Резултати истраживања биће квантификовани коришћењем lift мерике. Иако примењена на примеру мобилних телекомуникационих мрежа, предложена методологија се може користити и у било којој другој области где се пријатељске или хомофилне везе могу посматрати као потенцијални узрочник осипања и умањења броја клијената.

Релевантни радови из области истраживања:

- [1] Amin, A.; Al-Obeidat, F.; Shah, B.; Adnan, A.; Loo, J.; Anwar, S. Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*. 2019, 94, 290–301.
- [2] Jain, H.; Khunteta, A.; Srivastava, S. Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. *Procedia Computer Science*, 2020, 167, 101-112.
- [3] Wei, C.P.; Chiu, I.T. Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach. *Expert systems with applications*. 2002, 23, 103–112.
- [4] Hung, S.Y.; Yen, D.C.; Wang, H.Y. Applying Data Mining to Telecom Churn Management. *Expert systems with applications*. 2006, 31, 515–524.
- [5] Bose, I.; Chen, X. Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn. *Journal of Organizational Computing and Electronic Commerce*. 2009, 19, 133–151.
- [6] Cielen, D; Arno M.; Mohamed A. *Introducing data science: big data, machine learning, and more, using Python tools*. Manning Publications Co., 2016.
- [7] Flach, P. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.
- [8] Ward, J.H., Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American statistical association*. 1963, 58, 236–244
- [9] Nanavati, A.A.; Singh, R.; Chakraborty, D.; Dasgupta, K.; Mukherjea, S.; Das, G.; Gurusurthy, S.; Joshi, A. Analyzing the Structure and Evolution of Massive Telecom Graphs. *IEEE Transactions on Knowledge and Data Engineering*. 2008, 20, 703–718.
- [10] Van Steen, M. *Graph Theory and Complex Networks: An Introduction*; Maarten Van Steen: Enschede, The Netherlands, 2010.

3. Полазне хипотезе

Истраживачки рад кандидата Стефана Костића се заснива на следећим хипотезама:

- Истраживачки рад представља анализу телекомуникационе мреже као друштвене мреже, формиране применом теорије графова, и предикције губитка корисника у оквиру такве мреже.
- Граф је креиран на основу свих позива унутар мреже телекомуникационог оператора у периоду од месец дана. При креирању друштвене мреже позива у телекомуникационом оператору, као чворови се могу посматрати сви MSISDN бројеви (*Mobile Station International Subscriber Directory Number*), било као иницијатори, било као примаоци позива, унутар свих CDR (*Call Data Record*) записа анализираних у току посматраног временског периода.
- Чворови креираног графа ће бити предмет кластеризације која ће бити извршена на основу структуралних метрика чворова унутар графа (степен чвора, улазни степен чвора, излазни степен чвора, значајност чвора првог реда, значајност чвора другог реда, сопствени вектор чвора, вредност ауторитета чвора, *hub* вредност чвора и мера артикулације чвора).
- Полазна хипотеза је да је могуће идентификовати неке битне чворове унутар графа (то јест друштвене мреже), који су кључни за предикцију вероватноће да ће клијент напустити посматраног телекомуникационог оператора.

4. Научне методе истраживања

У оквиру овог истраживања биће коришћене следеће методе:

- У циљу анализе чворова друштвене мреже креиране помоћу записа о позивима унутар мреже телекомуникационог оператора, користиће се теорија графова и њени механизми за одређивање метрика чворова у графу.
- Развој модела за откривање неважећих чворова графа (чворова који не одговарају људима, већ телемаркетинг центрима, или корисничким сервисима).
- Развој модела базиран на анализи кластера и то за два одвојена скупа улазних метрика (метрике за неусмерене графове, односно метрике за усмерене графове).
- Имплементација предложених модела у програмском окружењу *SAS Enterprise Guide* и *SAS Enterprise Miner*.
- Креирање јединствене оцене – кластера клијената на основу њиховог утицаја у анализираној друштвеној мрежи.
- Перформансе предложеног модела биће анализирани и поређени са алтернативним методама базираним на машинском учењу.

5. Очекивани научни доприноси

Као резултат овог истраживачког рада очекују се следећи научни доприноси:

- Проучавање и опис широког спектра метрика чворова у графу на основу којих се могу открити структуралне карактеристике графа сачињеног од свих позива унутар мреже посматраног телекомуникационог оператора.
- Дефинисање три скупа метрика (мере за неусмерене графове, мере за усмерене графове и мера артикулације) које ће прецизно описати сваки елемент графа

сачињеног од свих позива унутар мреже посматраног телекомуникационог оператора; биће показано да, иако је граф телекомуникационе мреже по својој природи усмерен (услед природног усмерења позива од једног броја ка другом), мере за неусмерене графове такође могу пружити значајне додатне увиде који могу резултовати предвиђањем вероватноће губитка клијента.

- Развој новог метода базираног на примени теорије графова за откривање неважећих чворова (чворова који не одговарају људима, већ телемаркетинг центрима, или корисничким сервисима) графа сачињеног од свих позива унутар мреже посматраног телекомуникационог оператора.
- Развој модела базираног на реалним подацима који може пружити важан увид у сферу мобилног телекомуникационог пословања и може послужити као водич за осмишљавање нових и бољих стратегија за борбу против губитка корисника телекомуникационих оператора.
- Дефинисање алгоритма за спајање резултата две засебне методе кластеризације (које су базиране на метрикама за неусмерене графове, односно метрикама за усмерене графове) у унифицирану скалу која ће представљати финални модел предикције губитка корисника у телекомуникационим мрежама.
- Због ограниченог обима основних варијабли (број и трајање позива), разматрање могућности комбиновања предложеног метода са другим решењима базираним на додатним варијаблама (на пример, са моделима базираним на задовољству корисника и/или на преосталом времену трајања уговора корисника).

6. План истраживања и структура рада

Истраживање у оквиру израде докторске дисертације биће обављено у више фаза, и то:

- Преглед научне литературе и анализа перформанси постојећих алгоритама за одређивање вероватноће губитка корисника у телекомуникационим мрежама.
- Детаљан опис основних карактеристика науке о подацима и њених примена у индустрији, посебно телекомуникационој.
- Преглед карактеристика и метода технологије ненадгледаног машинског учења, конкретно метода кластеризације (карактеристике кластеризације, поређење метода, одређивање броја кластера у непознатим подацима).
- Опис примена теорије графова у анализи друштвених мрежа, преглед концепта теорије графова и основних метрика које дају увид у структуру графа, као и преглед концепта друштвених мрежа.
- Имплементација алгоритма за предикцију вероватноће губитка корисника у оквиру телекомуникационих мрежа применом кластеризације података. Конкретно, биће приказана предложена методологија (са описом модела, доступних података, метрика које ће се користити као и принципом њиховог груписања), кластеризација података (укључујући и препроцесирање података), кластеризација Вардовом хијерархијском методом, као и резултати предикционог модела.
- Верификација предложеног решења поређењем са алтернативним методама моделовања анализираних података.

Структура докторске тезе ће у великој мери пратити наведене фазе истраживања.

7. Закључак и предлог

Након свих претходно наведених чињеница и стечене слике о теми докторске дисертације и кандидату, Комисија закључује да кандидат Стефан Костић испуњава све услове да спроведе истраживање и да је способан да реализује докторску дисертацију на предложеној тему „Предикција губитка корисника у мобилним телекомуникационим мрежама применом ненадгледаног машинског учења“.

Презентовани садржаји и планирана истраживања припадају областима актуелних научних истраживања, а очекивани научни доприноси су значајни у домену развоја модела ненадгледаног машинског учења и њихове примене у пољу предикције губитка корисника у мобилним телекомуникационим мрежама. Комисија је сагласна да је предложена тема подобна за докторску дисертацију. Комисија констатује да предложена тема припада ужој научној области: Телекомуникације, и да је Електротехнички факултет матичан за ову научну област, те за ментора за израду докторске дисертације предлаже др Мирјану Симић-Пејовић, ванредног професора Електротехничког факултета у Београду.

У закључку овог извештаја, имајући у виду све горе наведено, Комисија предлаже Наставно-научном већу Електротехничког факултета у Београду да кандидату Стефан Костић одобри израду докторске дисертације под насловом: „**Предикција губитка корисника у мобилним телекомуникационим мрежама применом ненадгледаног машинског учења**“.

У Београду, 14.10.2020. год.

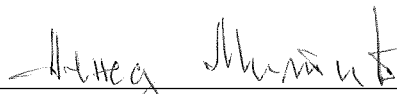
ЧЛАНОВИ КОМИСИЈЕ



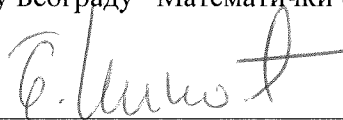
др Мирјана Симић-Пејовић, ванредни професор
Универзитет у Београду – Електротехнички факултет



др Милан Бјелица, редовни професор
Универзитет у Београду – Електротехнички факултет



др Ненад С. Митић, редовни професор
Универзитет у Београду – Математички факултет



др Бошко Николић, редовни професор
Универзитет у Београду – Електротехнички факултет