

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата **Милоша Котлара**, мастер инжењера електротехнике и рачунарства, за израду докторске дисертације „Детекција аномалија коришћењем мета података у аутоматизованим системима за машинско учење” (на енглеском „*Anomaly detection using meta features in automated machine learning systems*”)

Одлуком 5003/17-1 бр. од 24.09.2021. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата Милоша Котлара за израду докторске дисертације и научне заснованости теме „Детекција аномалија коришћењем мета података у аутоматизованим системима за машинско учење” (на енглеском „*Anomaly detection using meta features in automated machine learning systems*”)

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Милош Котлар је рођен 07.11.1993. године у Београду. Електротехнички факултет уписао је 2012. године и дипломирао на модулу софтверско инжењерство 2016. године са просечном оценом 9,24 одбранивши дипломски рад код проф. др Вељка Милутиновића. Дипломске академске – мастер студије на Електротехничком факултету у Београду уписао је 2016. године и мастерирао на модулу софтверско инжењерство 2017. године са просечном оценом 9,50 одбранивши мастер рад код доц. др Марије Пунт. Докторске студије уписао је 2017. године на модулу рачунарска техника и информатика и положио је све испите са просечном оценом 10,00. Од 2016. године био је запослен као софтверски инжењер у фирми „ABB”, Цирих, Швајцарска. Од 2019. године запослен је као водећи истраживач у фирми „TraceLabs”, Љубљана, Словенија.

Милош Котлар је научно-истраживачко искуство започео 2016. године у сарадњи са Електротехничком факултету у Београду на радовима у области *dataflow* рачунара и машинског учења. Даље искуство је стекао учествовањем на два пројекта као водећи истраживач у фирми „TraceLabs”, финансираним од стране ЕУ у оквиру H2020 програма. Такође, члан је радне групе за формирање *EPCIS* стандарда међународне организације за стандарде *GS1*. Пројекти у којима је учествовао припадају области машинског учења, проналажења скривеног знања, графови знања, дистрибуираних система, и архитектуре рачунара.

1.2. Сечено научноистраживачко искуство

Кандидат је докторске академске студије на изборном подручју Електротехника и рачунарство на ЕТФ-у уписао 2017/2018. школске године. Од тренутка уписа до тренутка подношења Захтева за одобрење докторске дисертације положио је све испите (са просечном оценом 10,0) и обавио све обавезе предвиђене докторским академским студијама на изборном подручју Електротехника и рачунарство. Списак положених испита и обављених обавеза је дат у следећој табели:

бр.	Назив испита / обавезе	Оцена	ЕСПБ	Датум
1	Одабрана поглавља из архитектуре рачунарских система	10	9	22.09.2019.
2	Сензорске мреже	10	9	20.09.2019.
3	Интернет програмирање	10	9	18.09.2019.
4	Увод у научни рад	10	6	08.07.2019.
5	Одабрана поглавља из програмских преводилаца	10	9	01.07.2019.
6	Софтверска техника	10	9	30.08.2018.
7	Архитектуре за подршку софтверским системима	10	9	06.07.2018.
8	Пројектовање рачунара у ВЛСИ техници	10	9	05.07.2018.
9	Вештачка интелигенција и експертски системи	10	9	02.07.2018.
10	Мултимедијални системи - одабрана поглавља	10	9	14.02.2018.
11	Студијски истраживачки рад II	-	15	30.09.2019.
12	Научно стручни рад	-	3	24.09.2019.
13	Студијски истраживачки рад I	-	15	28.09.2018.

Милош Котлар се бави развојем метода за детекцију аномалија коришћењем мета података, као и анализом различитих архитектура за извршавање оваквих система. Циљ истраживања кандидата је да се предложи скуп функција за израчунавање мета података који ће се користити за детекцију аномалија у подацима и који ће испуњавати критичне захтеве за примену у системима за аутоматизовано машинско учење. Како би се предложио и евалуирао скуп функција за израчунавање мета података циљ је да се постојећа решења упореде кроз различите аспекте комплексности. Милош Котлар је предложио решење за детекцију аномалија коришћењем мета података који се заснивају на доменском знању.

Милош Котлар је аутор на два научна рада објављена у часописима са импакт фактором са SCI листе у категоријама M21 и M22, а такође је коаутор у још једном раду објављеном у часопису категорије M22, затим аутор је поглавља у књизи, као и у пет радова објављених на међународним и домаћим конференцијама категорија M33, M34 и M53. Такође, коедитор је две књиге издавача Springer и IGI Global.

Дана 08.10.2021. године одржана је јавна усмена одбрана теме докторске дисертације (докторски испит) пред Комисијом у саставу: др Милош Цветановић, ванредни професор са Електротехничког факултета; др Горан Квашчев, ванредни професор са Електротехничког факултета; др Синиша Влајић, редовни професор са Факултета организационих наука. Оцена Комисије је да је кандидат Милош Котлар задовољио све критеријуме и одбранио предложену тему докторске дисертације.

У наставку је списак радова који показују резултате досадашњег рада и афинитета који су у складу са предложеном дисертацијом кандидата Милоша Котлара.

Списак радова:

Категорија M21:

Milos Kotlar, Marija Punt, Zaharije Radivojevic, Milos Cvetanovic, Veljko Milutinovic, „Novel Meta-Features for Automated Machine Learning Model Selection in Anomaly Detection”, *IEEE Access*, vol. 9, pp. 89675-89687, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2021.3090936, 2021. (IF 2019 = 3.745).

Категорија M22:

Milos Kotlar, Dragan Bojic, Marija Punt, Veljko Milutinovic, „Survey of deployment locations and underlying hardware architectures for contemporary deep neural networks”, *International Journal of Distributed Sensor Networks*, vol. 8, ISSN: 15501329, DOI: 10.1177/1550147719868669, 2019. (IF 2019 = 1.787).

Milan Banković, et al., „Teaching graduate students how to review research articles and respond to reviewer comments”, *Advances in Computers*, vol. 116, Elsevier, ISSN: 652458, DOI: 10.1016/bs.adcom.2019.07.001, 2020. (IF 2020 = 2.655).

Поглавље у књизи:

Milutinovic V., Kotlar M., Stojanovic M., Dundic I., Trifunovic N., Babovic Z., „Implementing Neural Networks by Using the DataFlow Paradigm”, *DataFlow Supercomputing Essentials*, Springer, DOI: 10.1007/978-3-319-66125-4_1, 2017.

Категорија M33:

Milos Kotlar, Zoran Babovic, Veljko Milutinovic, „Implementation of perceptron algorithm using dataflow paradigm”, *13th Symposium on Neural Networks and Applications (NEUREL)*, DOI: 10.1109/NEUREL.2016.7800110, 2016.

Trifunovic N., et al., „Experiences in the teaching of a new computing paradigm: DataFlow supercomputing”, *25th Telecommunication Forum (TELFOR)*, pp. 1-4, DOI: 10.1109/TELFOR.2017.8249482, 2017.

Kotlar M., Bojic D., Punt M., Milutinovic V., „A Survey of Deep Neural Networks: Deployment Location and Underlying Hardware”, *14th Symposium on Neural Networks and Applications (NEUREL)*, pp. 1-6, DOI: 10.1109/NEUREL.2018.8587006, 2018.

Kotlar M., Milutinovic V., „Comparing Controlflow and Dataflow for Tensor Calculus: Speed, Power, Complexity, and MTBF”, *Lecture Notes in Computer Science*, vol. 12151, pp. 329-346, ISSN: 0302-9743, DOI: 10.1007/978-3-030-02465-9_22, Springer, 2018. (IF 2018 = 1.390).

Категорија M34:

Milos Kotlar, Zoran Babovic, Veljko Milutinovic, „DataFlow Supercomputing for BigData”, *19th European Conference on Mathematics for Industry*, pp.136-137, DOI: 10.15304/cc.2016.968, 2016.

Категорија M53:

Trifunovic N., et al., „Experiences in the teaching of a new computing paradigm: Lessons Learned”, *Telfor Journal*, vol. 10, pp. 108-111, DOI: 10.5937/telfor1802108T, 2018.

1.3. Оцена подобности кандидата за рад на предложеној теми

Приликом писања овог Извештаја, Комисија је размотрила све претходно наведене чињенице и резултате у досадашњем раду Милоша Котлара. Са посебном пажњом су размотрена искуства у истраживачком раду и узимајући све наведено, Комисија је стекла увид у његову способност за научно истраживачки рад и рад на предложеној теми докторске дисертације, која је, такође, препозната и вреднована и од стране научне заједнице.

2. Предмет и циљ истраживања

Генерисање велике количине података условљено развојем крајњих уређаја (енг. *Edge Devices*) и интернет ствари (енг. *Internet of Things*) [1-2] довело је до убрзаног развоја технологије и алгоритама који се користе у системима за анализу и обраду података. Са великом количином података у системима за њихову анализу и обраду, заснованим на алгоритмима за машинско учење, перформансе система искључиво зависе од квалитета података, одабраног модела и параметара модела. Аномалије у подацима представљају инстанце које се разликују од дистрибуције података, утичу на квалитет података и могу да се детектују коришћењем алгоритама машинског учења. У зависности од случаја коришћења, аномалије могу да се детектују и анализирају ако представљају нови подскуп података са сличним карактеристикама, као и да се детектују и елиминишу ако представљају грешке у подацима. Аномалије у подацима могу да се разликују по типу локалитета и димензионалном простору у коме се детектују. Ако постоји балансиран узорак података са обележеним аномалијама, модели засновани на алгоритмима за класификацију се могу користити за детектују аномалија у подацима. У супротном се могу користити модели засновани на алгоритмима за машинско учење. У оба случаја неопходно је креирати модел за детекцију аномалија у подацима. Предлог модела и параметара модела за детекцију аномалија искључиво зависи од експертизе креатора система или доменског експерта. У случајевима када не постоји узорак података са обележеним аномалијама, што је чест случај у подацима из реалног света, предлог модела за детекцију аномалија није тривијалан. Предлог модела за детекцију аномалија се може аутоматизовати, при чему такав систем за аутоматизовано машинско учење (енг. *AutoML*) предлаже модел за детекцију аномалија у подацима на основу података, мета података, одговарајуће оптимизационе метрике и претходно стеченог знања [3-4].

Предмет истраживања у склопу ове докторске дисертације представља развој проширеног система за израчунавање мета података који ће се користити за предлагање модела за детекцију аномалија у небалансираним подацима за одговарајућу оптимизациону метрику у системима за аутоматизовано машинско учење. Идеја је да се систем за израчунавање мета података заснива на функцијама које користе доменско знање. Подразумева се да се за

детекцију аномалија користе постојећи алгоритми, тако да се проблем своди на креирање скупа функција за израчунавање мета података који ће карактеризовати аномалије у подацима. Коришћењем функција које користе доменско знање, креатор података или доменски експерт ће бити у могућности да израчуна мета податке у случајевима када не постоји узорак података са обележеним аномалијама. Претходна истраживања су показала да је детекција аномалија заступљена у различитим областима и за различите типове података, тако да наведени аспекти треба да буду посебно анализирани за различите типове локалитета и различите димензионалне просторе аномалија [5-6]. У циљу добијања резултата који ће бити релевантни, један од праваца истраживања је прикупљање и коришћење скупова података из индустрије и отворене литературе уз евалуацију алгоритама машинског учења који имају широку примену у детекцији аномалија. Како се у аутоматизованим системима за машинско учење користе функције које мере удаљеност између скупова података на основу мета података, идеја је да се уради евалуација постојећих и предлог нових функција за мерење удаљености између мета података како би се показало који тип функција за мерење удаљености даје добре резултате за предложени скуп мета података.

Циљ истраживања у склопу ове докторске дисертације је да се предложи скуп функција за израчунавање мета података који ће испуњавати критичне захтеве за наведену примену. Како би се предложио и евалуирао скуп функција за израчунавање мета података циљ је да се постојећа решења упореде кроз различите аспекте комплексности [7]. Предложени скуп функција за рачунање мета података ће моћи да се користи у будућим аутоматизованим системима за машинско учење за детекцију аномалија. Циљ је да се евалуирају функције за мерење удаљености између мета података и да се провери да ли одређени тип или група функција даје боље резултате у односу на остале функције. Циљ је да се дизајнирају експерименти и добију резултати који ће моћи да се користе у будућим истраживањима у области аутоматизованог машинског учења. Такође, биће анализирана ефикасност система на различитим архитектурама и местима извршавања како би се показало у којим окружењима предложено решење може да се имплементира и под којим условима.

Значај истраживања представља могућност имплементације аутоматизованих система за детекцију аномалија заснованим на предложеном скупу функција за израчунавање мета података који ће испуњавати критичне захтеве за наведену примену. У случајевима када не постоји узорак података са обележеним аномалијама, или подаци нису присутни, креатор података или доменски експерт ће моћи ефикасно да карактеризује аномалије у подацима на основу доменског знања.

- [1] A. Oussous, F. Benjelloun, A. Lahcen, and S. Belfkih, „Big data technologies: A survey,” *Journal of King Saud University - Computer and Information Sciences, Elsevier*, vol. 30, no. 4, pp. 431–448, 2018.
- [2] W. Gunther, M. Mehrizi, M. Huysman, and F. Feldberg, „Debating big data: A literature review on realizing value from big data,” *The Journal of Strategic Information Systems, Elsevier*, vol. 26, no. 3, pp. 191–209, 2017.
- [3] A. Filchenkov and A. Pendryak, „Datasets meta-feature description for recommending feature selection algorithm,” *Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search Conference*, pp. 11–18, 2015.
- [4] X. He, K. Zhao, and X. Chu, „Automl: A survey of the state-of-the-art,” *Knowledge-Based Systems, Elsevier*, vol. 212, 2021.
- [5] G. Aguiar, E. Santana, S. Mastelini, R. Mantovani, and S. Junior, „Towards meta-learning for multi-target regression problems,” *8th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 377–382, 2019.
- [6] C. Aggarwal, „Outlier analysis,” *Data mining, Springer*, pp. 237–263, 2015.
- [7] Milos Kotlar, Marija Punt, Zaharije Radivojevic, Milos Cvetanovic, Veljko Milutinovic, „Novel

3. Полазне хипотезе

Предложени критични захтеви које функције за израчунавање мета података треба да испуне, а односе се на скалабилност и перформансе, су:

- Неутралност – мета подаци треба да ефикасно описују карактеристике аномалија за различите области и типове података
- Скалабилност – функције за израчунавање мета података треба да буду једноставне за израчунавање и без сложених операција
- Повезаност – мета подаци треба да ефикасно описују карактеристике аномалија за различите оптимизационе метрике
- Једноставност – креатор података или доменски експерт треба да има могућност да израчуна мета податке у ситуацијама када подаци нису обележени или нису присутни

Постојећа решења за предлог модела на основу мета података се заснивају на статистичким функцијама и теорији информација које захтевају велике количине података и тако не испуњавају критичне захтеве за ову примену. Чињеница је да још није предложен проширив скуп функција за израчунавање мета података за предлог модела за детекцију аномалија на основу доменског знања, који ће смањити комплексност и потребне ресурсе за израчунавање [1].

Полазне хипотезе у склопу ове докторске дисертације су:

- Могуће је предложити скуп мета података заснованом на доменском знању за карактеризацију аномалија у подацима
- Предложени скуп мета података може да задовољи наведене критичне захтеве
- Предложени скуп мета података постиже исте или боље резултате у односу на постојеће скупове мета података за различите типове података, различите типове аномалија, као и за различите области из којих подаци долазе
- Предложени скуп мета података је могуће проценити искључиво на основу доменског знања и без присуства података
- За предложени скуп мета података је могуће одредити тип функција за мерење удаљености између скупова података који у већини случајева даје боље резултате од осталих типова функција које се најчешће користе за мерење удаљености између скупова података

Постојећи алгоритми за детекцију аномалија и скупови података из отворене литературе и индустрије ће се користити у дисертацији [2]. Такође, постојећи предлози аутоматизованих система за машинско учење ће се користити у дисертацији [3].

[1] H. Song, Z. Jiang, A. Men, and B. Yang, „A hybrid semi-supervised anomaly detection model for high-dimensional data,” *Computational intelligence and neuroscience, Hindawi*, vol. 1, 2017.

[2] M. Goldstein, „Unsupervised anomaly detection benchmark,” *Harvard University*, 2015.

[3] A. Filchenkov and A. Pendryak, „Datasets meta-feature description for recommending feature selection algorithm,” *Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference*, pp. 11–18, 2015.

4. Научне методе истраживања

У првом делу истраживања биће извршен детаљан преглед постојећих решења из отворене литературе за детекцију аномалија. Затим ће бити извршен детаљан преглед метода за предлог модела за детекцију аномалија на основу мета података, која обухватају решења где се мета подаци добијају из података или из модела. Систематична класификација ће идентификовати кључне карактеристике решења и биће анализиране добре и лоше стране сваког приступа, како би се одабрао оптималан приступ којим ће бити реализоване жељене карактеристике мета података. Посебно ће бити анализирана решења за предлог модела за детекцију аномалија која се заснивају на доменском знању података. Након тога ће бити анализирана природа аутоматизованих система за машинско учење. У складу са тим ће бити пројектован аутоматизовани систем за одабира модела за детекције аномалија у подацима.

У следећем делу истраживања ће бити имплементиран прототип функција за израчунавање мета података које се заснивају на доменском знању. Имплементирани прототип функција ће се евалуирати коришћењем података из отворене литературе и индустрије и извршиће се провера постављених критеријума са аспекта критичних захтева. Перформансе имплементираних система ће бити поређене са постојећим решењима коришћењем скупова података из отворене литературе и индустрије. Резултати ће бити приказани кроз аспекте типа података и области из које подаци долазе, као и типа локалитета аномалија и типа димензионалног простора у коме се детектују. Посебно ће бити анализирани случајеви када подаци нису доступни па се мета подаци израчунавају на основу доменског знања. Након тога ће се перформансе пројектованог аутоматизованог система за одабир модела евалуирати са становишта локација извршавања и различитих архитектура.

5. Очекивани научни допринос

Предвиђено је да ће у склопу дисертације бити развијен аутоматизовани систем за детекцију аномалија на основу функција за израчунавање мета података које се заснивају на доменском знању, при чему ће мета подаци омогућити довољно добро предлагање модела за одабрану оптимизациону метрику. Прототип система ће бити пројектован имајући у виду критичне захтеве за ову примену, а то су скалабилност и перформансе. Очекује се да ови захтеви буду испуњени креирањем мета података на основу доменског знања [1].

Листа доприноса:

- Евалуација постојећих скупова мета података који се користе у аутоматизованим системима за машинско учење и предлог њихове класификације са аспекта критичних захтева важних за проблем детекције аномалија и аутоматизованих система за машинско учење
- Предлог новог скупа мета података заснованог на доменском знању који ће пружати боље карактеристике са аспекта критичних захтева важних за проблем одабира модела за детекцију аномалија
- Креиран и карактерисан нови репозиторијум који ће обухватити скупове података релевантне за евалуацију предлога модела у аутоматизованим системима за машинско учење на основу доступних података у отвореној литератури и индустрији допуњеним са мета подацима заснованим на доменском знању
- Креиран софтверски систем за одабир модела за детекцију аномалија на основу мета података погодан за коришћење у аутоматизованим системима за машинско учење, који је проширив новим скуповима мета података, алгоритмима за детекцију аномалија и функцијама за рачунање удаљености између скупова података
- Дизајнирани експерименти погодни за мерење перформанси предлога модела за детекцију аномалија на основу предложеног скупа мета података

- Евалуација перформанси креираног софтверски система са становишта различитих локација извршавања и архитектура

Подразумева се да кључна примена резултата истраживања буде у области предлога модела за детекцију аномалија. Такође, резултати истраживања могу бити искоришћени и у другим применама где се користе аутоматизовани системи за машинско учење. Један од заједничких карактеристике таквих система је захтев за довољно добрим предлогом модела на основу доменског знања и функција за израчунавање мета података [2].

[1] H. Song, Z. Jiang, A. Men, and B. Yang, „A hybrid semi-supervised anomaly detection model for high-dimensional data,” *Computational intelligence and neuroscience, Hindawi*, vol. 1, 2017.

[2] J. Suez and E. Corchado, „A meta-learning recommendation system for characterizing unsupervised problems: On using quality indices to describe data conformations,” *IEEE Access*, vol. 7, pp. 63247–63263, 2019.

6. План истраживања и структура рада

Планирано је да се истраживање спроведе у пет фаза:

Прва фаза обухвата анализу постојећих решења за одабир модела на основу мета података, као и система за аутоматизовано машинско учење. Такође, она укључује прикупљање скупова података из отворене литературе и индустрије. Друга фаза обухвата детаљну анализу и дефинисање критичних захтева за примену у аутоматизованим системима за машинско учење на основу којих се жели реализовати архитектура система за аутоматизовано детектовање аномалија у подацима. Трећа фаза предвиђа дефинисање скупа функција за израчунавање мета података и провера испуњености критичних захтева. Четврта фаза предвиђа имплементацију кључних алгоритама за детекцију аномалија, функција за мерење удаљености, као и функција за рачунање мета података, односно прототип целог система и мерење перформанси. Пета фаза предвиђа анализу добијених резултата и доношење закључка.

Ако резултати и закључци пете фазе буду указивали да почетне хипотезе нису оправдане, биће поновљене друга, трећа, четврта и пета фаза са коригованим хипотезама.

7. Закључак и предлог

Приликом писања извештаја, Комисија је размотрила релевантне чињенице о подобности теме и кандидата. Узимајући у обзир резултате досадашњег рада кандидата **Милоша Котлара**, Комисија је закључила да кандидат јесте способан за научно истраживачки рад. Детаљном анализом предложене теме под називом „**Детекција аномалија коришћењем мета података у аутоматизованим системима за машинско учење**” (на енглеском „**Anomaly detection using meta features in automated machine learning systems**”), Комисија је у потпуности сагласна да оваква тема задовољава све услове подобности за докторску дисертацију.

Комисија констатује да предложена тема припада ужој научној области **Рачунарска техника и информатика** и да је Универзитет у Београду - Електротехнички факултет (ЕТФ) матичан за ову научну област. За менторе докторске дисертације комисија предлаже **др Марију Пунт**, доцента са ЕТФ-а и **др Захарију Радивојевића**, ванредног професора са ЕТФ-а, и прилаже уз овај Извештај и списак радова **др Марије Пунт** и **др Захарија Радивојевића**, који их квалификују за менторе докторске дисертације.

Имајући у виду све претходно наведено, Комисија има част и задовољство да Наставно-научном већу ЕТФ-а предложи да прихвати тему докторске дисертације „Детекција аномалија коришћењем мета података у аутоматизованим системима за машинско учење” (на енглеском „Anomaly detection using meta features in automated machine learning systems”) и да кандидату Милошу Котлару омогући приступ њеној изradi.

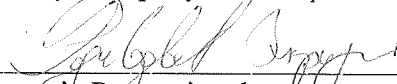
У Београду, 19.10.2021.

ЧЛАНОВИ КОМИСИЈЕ



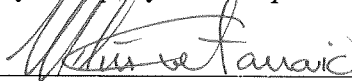
др Марија Пунт, доцент

Универзитет у Београду – Електротехнички факултет



др Захарије Радивојевић, ванредни професор

Универзитет у Београду – Електротехнички факултет



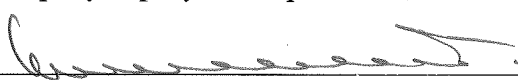
др Милош Цветановић, ванредни професор

Универзитет у Београду – Електротехнички факултет



др Синиша Влајић, редовни професор

Универзитет у Београду – факултет Организационих наука



др Горан Квашчев, ванредни професор

Универзитет у Београду – Електротехнички факултет