

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидаткиње Тамаре Живковић, мастер инжењера електротехнике и рачунарства, за израду докторске дисертације под насловом „Предвиђање дефеката у софтверу применом модела машинског учења оптимизованих метахеуристикама“.

Одлуком 883/35 од 13.06.2023. године, именовани смо за чланове Комисије за оцену подобности теме и кандидаткиње Тамаре Живковић, за израду докторске дисертације и научне заснованости теме „Предвиђање дефеката у софтверу применом модела машинског учења оптимизованих метахеуристикама” (енг. „*Software Defect Prediction Using Machine Learning Models Optimized by Metaheuristics*”).

На основу материјала приложеног уз Захтев кандидаткиње, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидаткињи

1.1 Кратка биографија

Тамара Живковић, мастер инжењер електротехнике и рачунарства, рођена је 28.11.1982. у Ваљеву, Република Србија. Основну школу и гимназију природно-математичког смера завршила је у Ваљеву са одличним успехом. Од ране младости исказивала је велико интересовање за природне науке. Поред интересовања за природне науке показала је интересовање и за дебату, књижевност и стране језике. Носилац је *Cambridge Advanced English (CAE)* сертификата.

Електротехнички факултет у Београду уписала је 2001. године и у децембру 2007. године дипломирала на Одсеку за телекомуникације и информационе технологије. Уписала је мастер академске студије 2008. године, које је завршила у октобру 2010. године. Докторске студије је уписала 2012. године на Електротехничком факултету Универзитета у Београду, модул Софтверско инжењерство, и положила је све испите предвиђене планом и програмом.

Списак положених испита на докторским студијама приказан је у табели која следи:

Назив предмета (са шифром предмета)	Оцена	ЕСПБ	Датум полагања испита
Теорија графова са применама у електротехници и рачунарству (13Д081ТГ)	10	9	3.2.2019.
Софтверска техника (13Д111СТ)	10	9	8.2.2020.
Медицинска информатика (13Д111МИ)	10	9	9.6.2016.
Интернет програмирање (13Д111ИП)	10	9	27.6.2016.
Сензорске мреже (13Д111СМ)	10	9	8.7.2016.
Софтверски системи за рад у реалном времену (13Д111ССР)	10	9	31.1.2020.
Виртуелна стварност (13Д111ВС)	10	9	14.1.2019.
Вештачка интелигенција и експертски системи (13Д111ВИЕ)	10	9	8.2.2019.
Линеарна алгебра (13Д08ЛА)	10	9	13.1.2019.
Француски језик (13Д091ФЈ)	10	6	17.1.2019.

1.2 Стручна делатност

Од априла 2008. године била је запослена у београдском одсеку компаније *P3 communications*, чије је седиште у Ахену, Немачка. Радилa је на пројектима тестирања софтвера мобилних телефона у државама Европе, Азије и Африке, на позицији тест инжењера. Од јануара 2012. године до данас запослена је у америчкој компанији *Merit Solutions*. Ради на позицији менаџера квалитета софтвера. Носилац је међународно признатих сертификата из области тестирања софтвера: *ISTQB Certified Tester Foundation Level*, *ISTQB Certified Tester Advanced Level Test Manage* и *ISTQB Certified Tester Foundation Level Agile Tester*. Фокус њеног истраживачког рада је на областима тестирања софтвера и примене вештачке интелигенције у тој области.

1.3 Основни подаци о објављеним радовима

Кандидаткиња се већ неколико година бавила анализом и решавањем проблема из области вештачке интелигенције и тестирања софтвера. Кандидаткиња из поменуте области има радове који су објављени као поглавља у монографијама, радове на међународним конференцијама, као и радове који су представљени на домаћим стручним конференцијама.

Поглавља у међународним монографијама (M13):

1. Zivkovic, M., **Zivkovic, T.**, Venkatachalam, K., Bacanin, N., "Enhanced Dragonfly Algorithm Adapted for Wireless Sensor Network Lifetime Optimization," *Chapter in Data Intelligence and Cognitive Informatics: ICDICI 2020, Springer, Algorithms for Intelligent Systems, Data Intelligence and Cognitive Informatics: ICDICI 2020, International Conference on Data Intelligence and Cognitive Informatics*, pp. 803 - 817, Springer, Singapore, Tirunelveli, India, Jan, 2021.
2. Bezdan, T., Zivkovic, M., Antonijevic, M., **Zivkovic, T.**, Bacanin, N., "Enhanced Flower Pollination Algorithm for Task Scheduling in Cloud Computing Environment," *Chapter in*

Lecture Notes in Networks and Systems, Machine Learning for Predictive Analysis: ICTIS 2020, Information and Communication Technology for Intelligent Systems, Springer, LNNS, volume 141, Machine Learning for Predictive Analysis: Lecture Notes in Networks and Systems, ICTIS 2020: International Conference on Information and Communication Technology for Intelligent Systems, volume 141, pp. 163-171, Springer, Singapore, Ahmedabad. India, Oct, 2020.

Радови на међународним конференцијама (M33):

1. Zivkovic, M., Stoean, C., Petrovic, A., Bacanin, N., Strumberger, I., **Zivkovic, T.**, "A Novel Method for COVID-19 Pandemic Information Fake News Detection Based on the Arithmetic Optimization Algorithm," *23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pp. 259 - 266, December 2021.
2. **Zivkovic, T.**, Zivkovic, M., "Survey of Learning Environments for Software Testing Education," *ECBS 2021: 7th Conference on the Engineering of Computer Based Systems, Association for Computing Machinery*, pp. 1 - 9, May 2021.
3. Živković, M., Bačanin, N., **Živković, T.**, Štrumberger, I., Tuba, E., Tuba, M., "Enhanced Grey Wolf Algorithm for Energy Efficient Wireless Sensor Networks," *IEEE Zooming Innovation in Consumer Technologies (ZINC)*, May 2020.
4. **Živković, T.**, Živković, M., "Comparative Analysis of Techniques for Testing Object Oriented Programs," *IEEE Zooming Innovation in Consumer Technologies (ZINC)*, May 2020.
5. **Živković, T.**, Živković, M., Algaddafi, K., Branović, I., Popović, R., "3D Virtual Laboratory for Wireless Sensor Networks," *21. telekomunikacioni forum TELFOR 2013*, Beograd, 2013.

Радови на домаћим конференцијама (M63):

1. Živković, M., **Živković, T.**, Bačanin, N., Strumberger, I., „Nature-Inspired Approaches in Software Testing Optimization," *Sinteza 2021 - International Scientific Conference on Information Technology and Data Related Research*, pp. 28 - 33, jun 2021.
2. **Živković, T.**, Živković, M., „Pregled primena virtuelnih okruženja u obrazovanju," *Sinteza 2019 - International Scientific Conference on Information Technology and Data Related Research*, pp. 99 - 106, april 2019.
3. Živković, M., **Živković, T.**, „Challenges in Testing of Web Applications," *Sinteza 2018 - International Scientific Conference on Information Technology and Data Related Research*, april 2018.
4. Živković, M., **Živković, T.**, „Wireless sensor networks integration into Internet of Things," *Sinteza 2017 - International Scientific Conference on Information Technology and Data Related Research*, april 2017.
5. **Živković, T.**, Živković, M., „Testiranje mobilnih aplikacija," *SYM-OP-IS*, Tara, septembar 2016.
6. Živković, M., Popović, R., **Tanasković, T.**, „Pregled komercijalno raspoloživih senzorskih platformi sa aspekta potrošnje energije," *20. Telekomunikacioni forum TELFOR 2012*, Beograd, novembar 2012.
7. **Tanasković, T.**, Živković, M., „Integracija bežičnih senzorskih mreža u cloud," *Konferencija ETRAN 2012*, Zlatibor, jun 2012.
8. **Tanasković, T.**, Živković, M., „Aplikacija za praćenje aktivnosti mobilnih telefona u mobilnim mrežama - CellDetective," *19. telekomunikacioni forum TELFOR 2011*, Beograd, novembar 2011.

9. **Tanasković, T., Živković, M., „Principi sigurnosti i zaštite veb aplikacija,“ 19. telekomunikacioni forum TELFOR 2011, Beograd, novembar 2011.**

Доприноси објављени у поменутих радовима представљају део основних резултата истраживања који ће бити представљени у докторској дисертацији кандидаткиње. Ти доприноси ће бити допуњени резултатима који до сада нису објављени. Образовање које је кандидаткиња стекла на дипломским и мастер студијама, као и објављени радови, у потпуности квалификују кандидаткињу за рад на предложеној теми.

Дана 26. јуна 2023. године, кандидаткиња Тамара Живковић је на Електротехничком факултету Универзитета у Београду усмено полагала испит о подобности предложене теме докторске дисертације и кандидата (докторски испит), пред комисијом у саставу:

- др Бошко Николић, редовни професор, Електротехнички факултет, Универзитет у Београду,
- др Драган Памучар, ванредни професор, Факултет организационих наука, Универзитет у Београду,
- др Бојана Михаиловић, доцент, Електротехнички факултет, Универзитет у Београду.

На јавној усменој одбрани предложене теме Комисија је оценила предложену тему дисертације и излагање кандидаткиње оценом задовољно.

1.4. Оцена подобности кандидаткиње за рад на предложеној теми

Разматрајући претходно приказане резултате досадашњег рада кандидаткиње Тамаре Живковић и узимајући у обзир све положене испите и обавезе током докторских студија, успешну одбрану теме докторске дисертације, као и до сада објављене научне радове на научно-стручним скуповима, Комисија је утврдила да је кандидаткиња показала да је стекла неопходно искуство за самосталан научно-истраживачки рад. Комисија је мишљења да је кандидаткиња способна да током израде докторске дисертације оствари резултате и доприносе који могу бити приказани и верификовани од стране међународне научне заједнице. Комисија сматра да кандидаткиња Тамара Живковић испуњава све неопходне и формалне услове да приступи изради докторске дисертације на предложеној теми.

2. Предмет и циљ истраживања

Предмет рада је примена модела машинског учења, чији су хиперпараметри оптимизовани метахеуристичким алгоритмима, за предвиђање дефеката у софтверу. Откривање дефеката у софтверу је једна од важних фаза развоја софтверских пројеката и, ако се правилно изведе, може бити кључ успеха или неуспеха. Данас, овај процес се некада занемарује због брзог темпа развоја и рокова који се уводе код савремених пројеката, јер недостатак времена доводи до смањеног броја сати посвећених њиховом тестирању. Овај лош приступ тестирању често резултира касним откривањем дефеката, а чак и од стране крајњих корисника у продукцији. Исправка дефеката који се открију касније у развојном процесу или у употреби од стране корисника је скупља. Зато, да би се ефикасно тестирале апликације из различитих области, тестери софтвера морају константно учити и прилагођавати се новим ситуацијама.

Анализа и предвиђање дефеката доприносе процени напретка пројекта и помажу менаџеру пројекта у планирању процеса тестирања. Такође, помажу у процени квалитета софтверског производа. На крају, побољшавају поузданост и функционалност производа. Модули склони дефектима могу бити идентификовани помоћу специфичних метрика, које су идентификоване у претходним студијама. Важне информације, попут броја дефеката, места у коду где су откривени, озбиљности последица и дистрибуције, могу се пронаћи и анализирати како би се побољшала ефикасност процеса тестирања. То, заузврат, може побољшати укупан квалитет будућег софтверског издања.

У литератури су идентификоване и коришћене бројне метрике за предвиђање дефеката у софтверу. С друге стране, ефикаснији приступ је препознати најважније метрике и фокусирати се на њихове перформансе и унапређења ради предвиђања грешака, уместо да се разматра велики број метрика. Наведене метрике се могу базирати на параметрима као што је број линија кода, цикломатска комплексност, број линија са коментарима, број оператора и операнада, и многе друге.

Додатне метрике које се могу користити за предвиђање грешака су изведене из графа тока контроле (енг. *Control Flow Graph*, скр. *CFG*) и често се користе за процену ефективности структурног тестирања, односно тестирања методама беле кутије. Ове метрике, изведене из *CFG*, укључују број исказа, грана, одлука и путања у програмском коду. Такође, оне се могу користити за мерење покривености кода (енг. *Code Coverage Techniques*).

Класификација је техника која се користи за категоризацију софтверских модула као модула склоних грешкама или модула који нису склони грешкама користећи метрике засноване на класификационим методама. Овај приступ су проучавали бројни истраживачи, па су класификациони алгоритми често коришћена метода машинског учења за предвиђање грешака у софтверу. Атрибути софтверског кода се класификују као дефектни или недефектни, користећи модел класификације заснован на подацима о софтверским метрикама из историјских развојних пројеката. Класификациони алгоритам је способан да предвиди који модули имају већу вероватноћу да буду склони грешкама, омогућавајући ефикасније коришћење ресурса за тестирање. Када се грешка открије током системског тестирања, одговарајућа категорија грешке модула се означава са 1, или са 0 у супротном.

Проблем предикције дефеката спада у проблеме бинарне класификације. Досадашња истраживања показала су да за таква предвиђања добре резултате дају алгоритми и модели машинског учења, а претрагом литературе може се уочити да је ово веома динамична и жива област.

Предмет рада обухвата и истраживање о претходно коришћеним приступима машинског учења како би се предвидело да ли је одређени модул подложен грешкама или не. Велика количина доступних података представља кључни изазов за сваки модел машинског учења, јер способност тренирања модела са више података генерално повећава способност модела да изврши процес генерализације. Јавно доступни репозиторијуми података имају кључну улогу у непристрасној компаративној анализи различитих приступа. Недавна истраживања се фокусирају на класификацију засновану на подацима о грешкама и софтверским метрикама.

Циљ рада је реализација система за предикцију дефеката у софтверским модулима, употребом модела машинског учења чији су хиперпараметри оптимизовани помоћу метахеуристичких

алгоритама. Систем је потребно реализовати у програмском језику који пружа најбоље перформансе. Систем као улаз добија скуп података са метрикама које описују софтверски модул. На почетку, неопходно је одредити одговарајућа подешавања хиперпараметара модела, пошто се за сваки конкретан проблем класификације мора подесити и модел који се користи. Затим ће се метахеуристички алгоритми користити за одређивање најбољег скупа хиперпараметара модела. Имплементирани систем ће, као резултат, радити бинарну класификацију (да ли је модул дефектан или не) на основу скупа софтверских метрика за дати модул. Биће анализирани резултати различитих метахеуристичких алгоритама, а на крају, биће приказана интерпретација најбољих модела генерисаних метахеуристичкама применом SHAP (SHapley Additive exPlanations) метода. Осим тога, биће обављена и статистичка анализа и биће дати статистички тестови који ће показати који модели су најбољи за овај тип проблема.

Осврт на релевантне библиографске изворе

Значај истраживања огледа се у примени реализованог решења у планирању процеса тестирања. Модули склони грешкама могу бити идентификовани помоћу специфичних метрика, а након тога, тестирање се може фокусирати на њих. Тиме се може побољшати ефикасност процеса тестирања, као и укупан квалитет софтвера. О значају теме говори у прилог и чињеница да су области предикције дефеката, као и проблем класификације уопштено и даље веома активне, имајући у виду бројна релевантна истраживања која су и даље актуелна, а о чему сведочи и велики број прикупљених референци из ових области.

Референце:

- [1] Alsaeedi, Abdullah, and Mohammad Zubair Khan. „Software defect prediction using supervised machine learning and ensemble techniques: a comparative study.“ *Journal of Software Engineering and Applications* 12.5 (2019): 85-100.
- [2] Qiao, Lei, et al. „Deep learning based software defect prediction.“ *Neurocomputing* 385 (2020): 100-110.
- [3] Akimova, Elena N., et al. „A survey on software defect prediction using deep learning.“ *Mathematics* 9.11 (2021): 1180.
- [4] Guo, Shikai, et al. „Feature transfer learning by reinforcement learning for detecting software defect.“ *Software: Practice and Experience* 53.2 (2023): 366-389.
- [5] Sharma, Umamaheswara, and Ravichandra Sadam. „How far does the predictive decision impact the software project? The cost, service time, and failure analysis from a cross-project defect prediction model.“ *Journal of Systems and Software* 195 (2023): 111522.
- [6] Oktian, Yustus Eko, and Howon Kim. „XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems.“ *Sustainability* 14.14 (2022): 1-21.
- [7] da Cruz, Mauro AA, et al. „Detecting Compromised IoT Devices Through XGBoost.“ *IEEE Transactions on Intelligent Transportation Systems* (2022).
- [8] Abaei, G., Selamat, A. „A survey on software fault detection based on different prediction approaches“. *Vietnam Journal of Computer Science*, 1 (2014): 79–95
- [9] Ai, J. X. W. „Defect prediction with semantics and context features of codes based on graph representation learning“. *IEEE Transactions on Reliability* (2021)

- [10] Bai, X., Zhou, H., & Yang, H. „An hvsm-based gru approach to predict cross-version software defects“. *International Journal of Performability Engineering*, 16 (2020)
- [11] Catal, C. „A comparison of semi-supervised classification approaches for software defect prediction“. *Journal of Intelligent Systems*, 23 (2014): 75–82.
- [12] Ghaffarian, S. M., & Shahriari, H. R. „Software vulnerability analysis and discovery using machine-learning and data-mining techniques: A survey.“ *ACM Computing Surveys (CSUR)*, 50 (2017): 1–36.
- [13] Hall, T., Beecham, S., Bowes, D., Gray, D., & Counsell, S. „A systematic literature review on fault prediction performance in software engineering.“ *IEEE Transactions on Software Engineering*, 38 (2011); 1276–1304.
- [14] Lessmann, S., Baesens, B., Mues, C., & Pietsch, S. „Benchmarking classification models for software defect prediction: A proposed framework and novel findings“. *IEEE Transactions on Software Engineering*, 34 (2008): 485–496.
- [15] Park, B.-J., Oh, S.-K., & Pedrycz, W. „The design of polynomial function-based neural network predictors for detection of software defects“. *Information Sciences*, 229 , (2013): 40–57.
- [16] Thota, M. K., Shajin, F. H., Rajesh, P. et al. „Survey on software defect prediction techniques“. *International Journal of Applied Science and Engineering*, 17 , (2020): 331–344.
- [17] Tong, H., Liu, B., & Wang, S. „Software defect prediction using stacked denoising autoencoders and two-stage ensemble learning“. *Information and Software Technology*, 96 , (2018): 94–111
- [18] Wahono, R. S. „A systematic literature review of software defect prediction“. *Journal of Software Engineering*, 1 , (2015): 1–16.

На основу анализе литературе, утврђено је да се модели машинског учења интензивно примењују не само на проблем предикције дефеката, већ постоји широк спектар могућих примена, од система детекције интрузије и злоупотребе кредитних картица у плаћањима преко интернета, па све до медицинских апликација. Може се такође приметити да добре резултате у области класификације постижу различити модели машинског учења, попут стабла одлучивања, чији је најбољи представник *XGBoost (Extreme Gradient Boosting)*. Међутим, теорема „нема бесплатног ручка“ („*no free lunch theorem*“) каже да не постоји универзалан алгоритам који се може применити на све проблеме, као и да не постоји универзални модел машинског учења, који за све скупове података може дати добре резултате. Због тога је сваки модел машинског учења потребно прилагодити за специфичан скуп података. Сваки модел машинског учења, укључујући и моделе засноване на алгоритму *XGBoost*, има велики број хиперпараметара које треба прилагодити за сваки специфичан проблем, као што су број епоха за тренирање, број слојева и њихове типове, број неурона по слојевима, стопа учења, и многе друге. Прилагођавање хиперпараметара је *NP (Non-deterministic polynomial-time)* тешки оптимизациони проблем.

За *NP* тешке оптимизационе проблеме, може да се напише детерминистички алгоритам, али би за његово извршавање било потребно много времена и због тога се овакви проблеми сматрају практично нерешивим. Из тих разлога је за њихово решавање потребно користити метахеуристичке методе и алгоритме. Метахеуристике не гарантују да ће дати оптимално решење, али зато постижу субоптимално задовољавајуће решење у разумном временском периоду.

У складу са свим наведеним, предмет истраживања докторске дисертације обухвата:

- свеобухватни преглед алгоритама класификације и модела који су расположиви у литератури;
- унапређење модела машинског учења за проблем предикције дефеката;
- оптимизацију модела машинског учења применом метахеуристичких метода оптимизације;
- анализу основних и унапређених модела машинског учења применом метахеуристика са истицањем предности и недостатака сваке методе;
- интерпретацију резултата развијених модела најбољих перформанси и генерисање значајаних информација за стручњаке из домена тестирања софтвера.

3. Полазне хипотезе

Основне хипотезе су:

- Применом до сада недовољно коришћених модела машинског учења за проблем предикције дефеката, могуће је генерисати боља решења у смислу брзине конвергенције и квалитета (перформанси, MSE , $R2$), чиме ће се унапредити резултати које су друге методе оптимизације постигле за исту инстанцу проблема.
- Адаптирати и оптимизовати постојеће имплементације модела машинског учења за проблем класификације дефеката софтвера.
- Адаптирати постојеће имплементације метахеуристика које до сада нису примењиване у овом домену, за оптимизацију машинског учења за проблем класификације.
- Унапредити постојеће имплементације метахеуристика ројева, које су раније већ биле примењене у овом домену, за оптимизацију модела машинског учења за проблем класификације.
- Могуће је уочити софтверске модуле који су подложни дефектима анализом метрика које описују софтвер.

4. Научне методе истраживања

Да би се развиле хипотезе које су у складу са предметом и циљем истраживања, биће примењене методе истраживања које су уобичајене у пракси и научно потврђене: аналитичко–синтетичка метода, дедуктивно–индуктивна метода, компаративно–квантитативна анализа, емпиријска метода, статистичка анализа, моделирање и експериментални део за доказивање хипотеза.

Аналитичким методом ће се обавити анализа садржаја релевантне научне и стручне литературе у овој области, као и релевантних чињеница публикованих путем електронских и других медија.

Синтетичким методом ће се интегрисати рашчлањени анализирани елементи појаве у једну целину ради дефинисања одређених правила у понашању појаве која је предмет истраживања.

Другим речима, сви прикупљени подаци биће груписани методом синтезе, како би се дошло до поузданих закључака.

Коришћење дедуктивно–индуктивне методе је у функцији усмеравања истраживања од општег ка појединачном, односно од појединачног ка општем у циљу доласка до адекватних закључака.

У оквиру компаративно-квантитативне анализе, а у корелацији са циљевима истраживања и појавом којом се истраживање бави, као најадекватнији приступ истраживању изабрана је метода анализе. За анализу прикупљених конкретних резултата примене метода, техника и метахеуристика за решавање проблема класификације, користиће се метода квантитативне анализе. Методом компаративне анализе упоредиће се добијени резултати оптимизације и извешће се закључак о стварном унапређењу решавања овог проблема применом алгоритама интелигенције ројева.

Применом емпиријске методе прикупљени примарни подаци ће бити разматрани и у светлу дискутованих постојећих теоријских и емпиријских налаза, односно секундарних извора података, како би се тестирале полазне истраживачке хипотезе. Практичним експериментисањем (пуштањем алгорита) долази се до примарних података.

Статистичка анализа подразумева коришћење статистичких података и метода у обради добијених резултата, у складу са постављеним истраживачким циљевима. За одређивање правца и степена међузависности резултата различитих метахеуристика, метода и техника, користиће се статистичко-економетријске методе тренда и корелације.

Под моделирањем се подразумева конструкција математичких модела који представљају валидну представу проблема који се изучава и решава. За решавање оптимизационих проблема посматрају се математички модели и формулације, као и функције циљева проблема истраживања. Такође се, у експерименталном делу имплементације, сагледавају технике метахеуристика у решавању проблема.

5. Очекивани научни допринос

Главни покретач истраживања које ће бити спроведено за потребе ове докторске дисертације је изводљиво и реално очекивање, да ће имплементирани и адаптирани алгоритми интелигенције ројева, који су примењени за оптимизацију неуронских мрежа за проблем предикције дефеката, генерисати боље моделе од модела машинског учења који се тренутно могу наћи у литератури, тј. моделе који постижу боље индикаторе перформанси од постојећих модела (*MSE*, *RMSE*, *R2*, итд.).

На основу до сада наведеног, могу се конкретно издвојити очекивани научни доприноси:

- детаљне таксономије модела машинског учења и алгоритама који су дати у литератури за проблем класификације и предикције дефеката;
- унапређење решавања проблема предвиђања дефеката применом реализованих модела машинског учења помоћу метахеуристичких метода оптимизације;
- упоредна анализа резултата различитих модела машинског учења за ову класу проблема, као и између развијених унапређених модела, оптимизованих различитим метахеуристикама које се користе у литератури,

- интерпретација најбољих модела и генерисања корисног знања за експерте из домена тестирања софтвера, које се може употребити за доношење бољих управљачких одлука,
- одређивање софтверских метрика које имају највише утицаја на предвиђање.

6. План истраживања и структура рада

Истраживање ће се састојати из следећих фаза:

- У првој фази биће идентификоване метрике које се користе да опишу софтвер. Модули склони грешкама могу бити идентификовани помоћу ових специфичних метрика, које су идентификоване у претходним студијама предвиђања грешака. Важне информације, попут броја грешака, њихове локације, озбиљности и дистрибуције, могу се извући како би се побољшала ефикасност процеса тестирања. Затим ће бити извршена анализа најчешће коришћених приступа проблему предвиђања грешака, сагледане њихове предности и недостаци и извршена њихова класификација.
- У другој фази биће спроведена анализа постојећих решења од интереса у области предвиђања дефеката, која за сваки систем подразумева: навођење сврхе система, структуре система, технологије у којима је систем реализован, алгоритме које систем примењује, области вештачке интелигенције које систем користи, уочавање предности и недостатака система, категоризација система према прецизности детекције дефеката, као и ограничења под којима систем ради и његова коначна евалуација.
- У трећој фази биће урађена анализа модела машинског учења. Посебна пажња биће посвећена математичкој основи и подтиповима ових модела, које ће се испитивати за потребе ове докторске дисертације, попут *XGBoost* модела. Осим наведеног, биће направљен и осврт на један од најзначајнијих проблема у домену машинског учења, проблем оптимизације хиперпараметара и могућностима за његово решавање.
- У четвртој фази биће приказана таксономија оптимизационих алгоритама. Посебан фокус ће бити на категоризацији метахеуристичких алгоритама. Детаљније ће бити описане и анализиране метахеуристике интелигенције ројева. Уз основне принципе по којима ови алгоритми функционишу, биће приказане специфичности значајнијих представника ове класе алгоритама.
- Главни део истраживања ће представљати имплементације метахеуристика и модела машинског учења за решавање проблема предвиђања дефеката. Овде ће бити приказани и скупови података који ће бити коришћени. Циљ система је да пружи могућност поузданог предвиђања да ли је софтверски модул подложен дефектима или не, на основу скупа метрика које се користе да опишу софтвер. Идејно решење ће бити имплементирано и биће приказани најзначајнији модели који су развијени.
- На самом крају ће бити урађена интерпретација најбољих модела, генерисаних метахеуристикама применом *SHAP* метода. Додатно, биће извршена статистичка анализа и биће дати статистички тестови који ће показати који модели су најбољи за овај тип проблема.

Дисертација ће се састојати из увода, седам поглавља и закључка.

У уводном делу докторске дисертације укратко ће бити изложен проблем који ће бити разматран, као и методе које ће се користити за унапређење решавања разматраног проблема. Осим наведеног, у уводном делу ће бити приказани и циљ истраживања, хипотезе, као и методе истраживања.

Прво поглавље биће посвећено концепту проблема предвиђања дефеката. Биће описан значај квалитета софтвера, као и метрике које се користе да опишу софтвер. Модули склони грешкама могу бити идентификовани помоћу ових специфичних метрика, које су идентификоване у претходним студијама предвиђања грешака.

У другом поглављу дисертације детаљно ће бити приказани појам и таксономија проблема оптимизације и укратко ће бити приказане класе комплексности.

Треће поглавље бавиће се прегледом модела машинског учења. Посебна пажња биће посвећена математичкој основи и подтиповима ових модела, које ће се испитивати за потребе ове докторске дисертације. Осим наведеног, биће направљен и осврт на један од најзначајнијих проблема у домену машинског учења, проблем оптимизације хиперпараметара и могућностима за његово решавање.

У четвртом поглављу биће приказана таксономија оптимизационих алгоритама. Посебан фокус ће бити на категоризацији метахеуристичких алгоритама.

У петом поглављу, детаљније ће бити описане и анализиране метахеуристике интелигенције ројева. Уз основне принципе по којима ови алгоритми функционишу, биће приказане специфичности значајнијих представника ове класе алгоритама.

У оквиру шестог поглавља биће детаљно приказане имплементације метахеуристике и модела машинског учења које су коришћене за решавање проблема предвиђања дефеката. Овде ће бити приказани и скупови података који ће бити коришћени, и то скупови *NASA-MDP*, *PROMISE* и *GHPR*, као и модели који су развијени.

У седмом поглављу биће приказана интерпретација најбољих модела, генерисаних метахеуристичким применом *SHAP* метода. Осим тога, у овом поглављу ће бити приказана статистичка анализа и биће дати статистички тестови који ће показати који модели су најбољи за који тип проблема.

На крају дисертације биће изнет закључак са могућим правцима развоја у овој области.

7. Закључак и предлог

На основу свега изложеног, Комисија констатује да Тамара Живковић, мастер инж. електротехнике и рачунарства, испуњава потребне услове да може приступити изради докторске дисертације, под насловом „**Предвиђање дефеката у софтверу применом модела машинског учења оптимизованих метахеуристикама**”. Комисија предлаже Наставно-научном већу Електротехничког факултета у Београду да одобри Тамари Живковић израду докторске дисертације под горе наведеним насловом и менторством доцента др Дражена Драшковића.

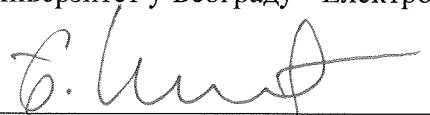
У Београду,
26.6.2023. године

Чланови Комисије:



др Дражен Драшковић, доцент

Универзитет у Београду - Електротехнички факултет



др Бошко Николић, редовни професор

Универзитет у Београду - Електротехнички факултет



др Драган Памучар, ванредни професор

Универзитет у Београду - Факултет организационих наука



др Бојана Михаиловић, доцент

Универзитет у Београду - Електротехнички факултет