Одлуком Наставно-научног већа Електротехничког факултета у Источном Сарајеву, Универзитета у Источном Сарајеву, број 03-865/17 од 04.07.2017. године, именована је Комисија за оцену и одбрану урађене докторске дисертације кандидата мр Маријане Ћосовић под називом **"Модели машинског учења за класификацију аномалија у БГП протоколу"**

(у даљем тексту: Комисија) у следећем саставу:

1. **Проф. др Стеван Станковски**, (ФТН Нови Сад), ужа научна област: *Мехатроника, роботика и аутоматизација* - предсједник;
2. **Проф. др Слободан Обрадовић**, (ЕТФ Источно Сарајево), ужа научна област: *Рачунарске науке* - ментор и члан;
3. **Доц. др Мирјана Максимовић**, (ЕТФ Источно Сарајево), ужа научна област: Телекомуникације - члан;
4. **Проф. др Синиша Ранђић**, (ФТН Чачак) , ужа научна област: *Рачунарска техника* и *телекомуникације* - члан;
5. **Проф. др Срђан Дамјановић**, (ФПЕ Бијељина), ужа научна област: *Информационе науке и биоинформатика* - члан,

Комисија је прегледала и оценила докторску дисертацију и о томе подноси Наставно-научном већу Електротехничког факултета у Источном Сарајеву, Универзитета у Источном Сарајеву следећи

И З В Е Ш Т А Ј
о оцени урађене докторске дисертације

1. Значај и допринос докторске дисертације са становишта актуелног стања у одређеној научној области
Интернет као глобални систем повезаних компјутерских мрежа користи протокол пограничних рутера (енг. *Border Gateway Protocol*) за повезивање свих мрежа објављивањем постојања сваке мреже и процедура којима се истим приступа. С обзиром да Интернет није централизован систем, повезаност његових делова базирана је на исправном функционисању БГП протокола на граничним рутерима. БГП протокол је стандардни протокол за рутирање на Интернету и базиран је на моделу поверења. Као такав мета је напада и штетних ефеката аномалија. Дисертација разматра моделе машинског учења за детекцију аномалија у БГП протоколу. Основни проблем који машинско учење решава јесте предикција. Једна од техника

машинског учења за детекцију аномалија је класификација, која припада надгледаном учењу, и бави се класификовањем података у дефинисан, коначан број класа. У овој дисертацији решава се питање бинарне класификације: аномалија постоји или не постоји. Кориштени су алгоритми вектора подршке (енг. *Support Vector Machines*), наивног Бајеса (енг. *Naïve Bayes*), стабла одлучивања, неуронске мреже, ансамбл методе. Разматрају се различити модели развијени на бази алгоритама машинског учења у сврху што боље предикције аномалија. Такође се разматрају конкретни сценарији различитих врста аномалија, које утичу на рад БГП протокола.

У склопу истраживања користи се Хероку платформа као сервис (енг. *Platform as a Service*) за постављање, извршавање и управљање апликацијама. Апликација пружа даљински приступ свеобухватном систему, који подржава цели процес неопходан за класификацију аномалија: трансформацију података, екстракцију својстава, селекцију својстава и, напослетку, обраду матрице својстава моделима машинског учења.

2. Оцену да је урађена докторска дисертација резултат оригиналног научног рада кандидата у одговарајућој научној области

Урађена докторска дисертација представља оригинално научно дело. Поред развоја модела за детекцију аномалија базираних на техникама машинског учења примењени су и упоређивани разни механизми машинског учења са циљем креирања бољих модела за откривање аномалија. Такође су анализирани конкретни сценарији аномалија у БГП протоколу, упоређивани алгоритми селекције својстава, селекције најрелеватнијих својстава класе ‘аномалија’. Развијени софтверски алат – десктоп и веб апликација – пружају приступ систему за трансформацију података из MRT у ASCII формат, екстракцију својстава из БГП порука, селекцију својстава из матрице својстава, примену алгоритама узорковања за трансформацију података, те обраду матрице својстава моделима машинског учења.

3. Преглед остварених резултата рада кандидата у одређеној научној области

Кандидаткиња је током спроведеног истраживања остварила значајне резултате од којих су неки потврда ранијих истраживања, а неки су потпуна новина у области класификације аномалија у БГП протоколу. Истраживањем су идентификовани случајеви различитих врста аномалија, које настају у БГП протоколу. Такође су идентификована својства, која су екстрактована из БГП порука, који су индикатори појављивања аномалија, дакле идентификовани су фактори ризика за настанак аномалије. Даљи остварени резултати се односе на смањење трајања времена обраде матрице својстава. Селекција најрелеватнијих својстава се односи на одабир подскупа својстава, који редукује матрицу својстава и време обраде исте тиме не компромитујући квалитет података.

4. Оцену о испуњености обима и квалитета у односу на пријављену тему (по поглављима)

Кандидаткиња је у потпуности испоштовала план и програм рада на дисертацији а у складу са пријавом докторске дисертације.

Докторска дисертација кандидата садржи укупно 167 страница, укључујући насловну страну, захвалницу, посвету, предговор, сажетак на српском и на енглеском језику и садржај. Текст дисертације је обима 151 страница, у оквиру којих је приказано 53 слике, 39 табела и 174 референце. Дисертација је организована кроз седам поглавља, укључујући уводно поглавље, закључак и доприносе дисертације. На крају дисертације приложена су два додатка, списак

кориштене литературе, попис кориштених скраћеница, слика, табела и кратка биографија аутора.

Увод
У првом, уводном поглављу представљени су предмет, циљ и резултат истраживања, те структура рада.

Детекција аномалија
Након уводног дела слиједи друго поглавље у коме се даје детаљан преглед система за детекцију аномалија и њихове примене у БГП протоколу.

Машинско учење
Технике машинског учења и њихове основне поделе представљене су у трећем поглављу. Бинарна класификација је један од најчешћих задатака машинског учења. Алгоритми машинског учења: методе вектора подршке, стабла одлучивања, неуронске мреже, наивни Бајес, као и ансамбл технике, користе се у експерименталним истраживањима за доказ постављењих хипотеза су детаљно описани. Примена и поређење разних алгоритама машинског учења су разматрани у сврху креирања бољих модела за откривање аномалија. Представљене су и мере за евалуацију класификацијских модела, на основу којих су вршена поређења.

Рутирање на Интернету
У четвртом поглављу се даје детаљан преглед из области структуре Интернета, где су објашњени различити протоколи за рутирање. У наставку истог поглавља је представљен БГП, протокол рутирања, који се користи за рутирање на цијелом Интернету. Такође су представљене БГП упдате поруке, које се користе у дисертацији, као и екстракција својстава из истих. Због обимности кориштених БГП упдате порука које се генеришу сваких пет минута, и лакше манипулације подацима, користили смо базу података у којој смо генерисали својства БГП упдате порука. Паралелно је развијена алтернативна метода генерисања података, која је имала за циљ брже екстрактовање података због евентуалне примене у стварном времену. У наставку су представљене технике за селекцију својстава као и селекцију најрелеватнијих својстава класе 'аномалија'. Технике за селекцију својстава које се користе у експерименталном истраживању рада су методе филтрирања (енг. *filter*) и методе омотача (енг. *wrapper*). С обзиром да подаци у домену детекције аномалија често нису балансирани (различитим класама припада несразмјеран број података), то представља проблем при класификацији, те су због тога наведене и разматране технике балансирања података.

Класификација аномалија
У петом поглављу је извршена бинарна класификација БГП аномалија над скуповима података, који припадају различитим врстама аномалија. Прије класификације аномалија, извршено је претпроцесирање података, да би подаци били ослобођени неконзистентности и комплетирани, у случајевима у којима је то било неопходно. Дискретизација и нормализација података је извршена како на оригиналним скуповима података тако и на подацима који су претходно обрађени алгоритмима узорковања података. Кориштени су алгоритми преузорковања, подузорковања и комбинација алгоритама преузорковања и подузорковања. Извршена је селекција својстава користећи филтер методе и методе омотача. У овом поглављу су дефинисани и модели мета-класификатора машинског учења. Модели машинског учења су базирани на SVM, NB, C4.5 и неуронским мрежама са неколико скривених слојева. Такође су кориштене ансамбл методе за класификацију (енг. *bagging*,

boosting і *random forest*).

БГП тоол

У шестом поглављу представљене су десктоп апликације, које су развијене у току рада на докторској дисертацији. Прва апликација је конверзија MRT формата у ASCII формат. У овој апликацији се такође обавља и екстракција својстава из ASCII порука. Друга апликација се бави примјеном алгоритама машинског учења и формирање платформе за поређење различитих алгоритама машинског учења над одабраним подацима. Веб апликација је надоградња постојећих десктоп апликација а реализована је ради лакшег корисничког приступа апликацији уз потребан Веб претраживач и Интернет конекцију. За развој апликација кориштен је програмски језик Python. Python је платформски независан објектно оријентисан, интерпретерски и интерактивни програмски језик широке примјене: од системске администрације, преко развоја интернет апликација, све до нумеричке примјене у научним истраживањима. Широкој употреби допринијеле су читљивост и прегледност синтаксе Python, те чињеница да се брзо и лако учи и усваја.

Закључак

Седмо поглавље рада садржи сумиране резултате истраживања. Приказани су доприноси дисертације и наведене смјернице за будући рад.

Докторска дисертација је и по обиму и по квалитету у потпуности испунила циљеве и задатке постављене у пријави дисертације.

5. Научне резултате докторске дисертације

Истраживања спроведена у оквиру докторске дисертације у циљу потврде постављених хипотеза су у потпуности испуниле очекиване резултате и постављене циљеве.

Основни научни доприноси дисертације су:

- развој модела за детекцију аномалија базираних на техникама машинског учења;
- примена и поређење разних механизама машинског учења у сврху креирања бољих модела откривања аномалија;
- развијен софтверски алат – десктоп и веб апликација за:
  - трансформацију MRT у ASCII формат;
  - екстракцију података из БГП порука;
  - селекцију својстава из матрице својстава;
  - примјену алгоритама узорковања за трансформацију података;
  - примјену различитих алгоритма машинског учења за детекцију аномалија;
- анализа података конкретних сценарија аномалија у БГП протоколу;
- поређење алгоритама селекције својстава и селекција најрелевантнијих својстава за класу ‘аномалија’.

6. Примењивост и корисност резултата у теорији и пракси

Материјал који дисертација обрађује је актуелан како са стране поља примене тако и по питању алата који се користе. Развијени теоретски модели су подлога за практично решење дисертације у коме се нуди лакоћа кориштења обзиром да се ради о веб апликацији. Спој детекције аномалија у БГП протоколу и машинског учења је простор у коме се могу реализовати многа будућа практична решења тј. надоградња постојећег решења. Корисност и примењивост како теоретских тако и практичних резултата дисертације је у томе што је установљен практичан начин за анализу разних догађаја аномалија у једноставном окружењу веб апликације.

7. Начин презентирања резултата научној јавности

Мр Маријана Ћосовић је објавила највећи део резлутата своје докторске дисертације у међународним часописима и зборницима домаћих и међународних конференција.

- **Marijana Ćosović**, Slobodan Obradović, "BGP anomaly detection with balanced datasets," Tehnički vjesnik/Technical Gazette, vol. 25, no. 3, in press, June 2018.
- **Marijana Ćosović,** Slobodan Obradović, and Emina Junuz, "Deep learning for detection of BGP anomalies," in Proceedings of International work-conference on Time Series (ITISE 2017), Granada, Spain, Sept. 2017, accepted for presentation.
- **Marijana Ćosović,** Slobodan Obradović, "Ensemble methods for classifying BGP anomalies," Industrial Technologies, ISSN: 13149911, vol. 4, no. 1, pp. 12–20, June 2017.
- **Marijana Ćosović,** Slobodan Obradović, and Ljiljana Trajković, "Classifying anomalous events in BGP datasets," in Proceedings of the 29th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2016), Vancouver, Canada, May 2016, pp. 697–700.
- Elvisa Bećirović, **Marijana Ćosović**, "Machine learning techniques for short-term load forecasting," in Proceedings of the 4th International Symposium on Environmentaly Friendly Energies and Applications, Belgrade, Serbia, Sept. 2016, pp. 1–4.
- **Marijana Ćosović**, Slobodan Obradović, and Ljiljana Trajković, "Performance evaluation of BGP anomaly classifiers," in Proceedings of the International Conference on Digital Information, Networking and Wireless Communication, Moscow, Russia, Feb. 2015, pp. 115–120.
- **Marijana Ćosović**, Slobodan Obradović, and Ljiljana Trajković, "Using databases for BGP data analysis," in Proc. UNITECH 2014, Gabrovo, Bulgaria, Nov. 2014, vol. 2, pp. 367–370.
- **Marijana Ćosović**, Slobodan Obradović, "Sigurnost u BGP Protokolu," INFOTEH-JAHORINA Vol. 13, Ref. KST-3-6, March 2014. ISBN 978-99955-763-3-2, p. 496-500.
- **Marijana Ćosović**, Slobodan Obradović, and Ljiljana Trajković, "Feature selection techniques for machine learning," in Proc. International Scientific Conference, UNITECH 2013, Gabrovo, Bulgaria, Nov. 2013, no. 1, pp. 85-89.
- **Marijana Ćosović**, Slobodan Obradović, and Ljiljana Trajković, "Algorithms for investigation of abnormal BGP events," in Proc. International Scientific Conference, UNITECH 2013, Gabrovo, Bulgaria, Nov. 2013, no. 2, pp. 253-257.
- **Marijana Ćosović**, Mirjana Maksimović, and Slobodan Obradović, "Role of data mining techniques in wireless sensor networks," XI International Conference ETAI, Ohrid, Makedonija, Septembar 2013, ISBN-978-9989-630-68-2.

8. ЗАКЉУЧАК И ПРЕДЛОГ

На основу увида у докторску дисертацију мр Маријане Ћосовић под називом **“Модели машинског учења за класификацију аномалија у БГП протоколу“,** Комисија је једногласно закључила да је кандидат изабрала актуелну и оригиналну тему истраживања, коју је спровела поштујући све принципе научног рада и користећи савремене методе испитивања и анализе резултата. Дисертација садржи оригиналне теоријске и практичне резултате у области примене машинског учења у БГП протоколу. Резултати добијени у овом истраживању омогућавају да се боље сагледају и креирају превентивне стратегије за оптимално функционисање рутирања и веза на Интернету.

На основу наведеног, Комисија предлаже Вијећу Електротехничког факултета у Источном Сарајеву и Сенату Универзитета у Источном Сарајеву, да докторску дисертацију под насловом

“Модели машинског учења за класификацију аномалија у БГП протоколу“

аутора мр Маријане Ћосовић, дипл. инж. ел. прихвати и одобри њену јавну одбрану, којом ће стећи звање доктора техничких наука.

Место: Источно Сарајево
Датум: 01. 09. 2017. године

Комисија:

1. **Стеван Станковски, у** звању редовни професор, Мехатроника, роботика и аутоматизација, Универзитет у Новом Саду, Факултет техничких наука у Новом Саду, председник Комисије;

_____

2. **Слободан Обрадовић**, у звању ванредни професор, Рачунарске науке, Универзитет у Источном Сарајеву, Електротехнички факултет у Источном Сарајеву, члан Комисије;

_____

3. **Мирјана Максимовић**, у звању доцент, Телекомуникације, Универзитет у Источном Сарајеву, Електротехнички факултет у Источном Сарајеву, члан Комисије;

_____

4. **Синиша Ранђић**, у звању редовни професор, Рачунарска техника и телекомуникације, Универзитет у Крагујевцу, Факултет техничких наука у Чачку, члан Комисије;

_____

5. **Срђан Дамјановић**, у звању ванредни професор, Информационе науке и биоинформатика, Универзитет у Источном Сарајеву, Факултет пословне економије у Бијељини, члан Комисије;

_____