

## ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

<b>I ПОДАЦИ О КОМИСИЈИ</b>
<p>1. Датум и орган који је именовео комисију</p> <p>7. април 2016. године, Наставно-научно веће Економског факултета у Суботици, Универзитет у Новом Саду</p> <p>2. Састав комисије са назнаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:</p> <p>1. Проф. др Зита Бошњак, редовни професор, Информациони системи и инжењеринг, 14. децембар 2006. године, Универзитет у Новом Саду, Економски факултет у Суботици.</p> <p>2. Проф. др Саша Бошњак, редовни професор, Информациони системи и инжењеринг, 13. јун 2011. године, Универзитет у Новом Саду, Економски факултет у Суботици.</p> <p>3. др Александар Ковачевић, доцент, Примењене рачунарске науке и информатика, 27. јануар 2012. године, Универзитет у Новом Саду, Факултет техничких наука.</p>
<b>II ПОДАЦИ О КАНДИДАТУ</b>
<p>1. Име, име једног родитеља, презиме: Оливера, Бранко, Грљевић</p> <p>2. Датум рођења, општина, држава: 22.12.1981. године, Зрењанин, Република Србија</p> <p>3. Назив факултета, назив студијског програма дипломских академских студија – мастер и стечени стручни назив</p> <p>Економски факултет у Суботици, Информациони инжењеринг, дипломирани економиста – мастер.</p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија</p> <p>2009. године, Пословна информатика</p> <p>5. Назив факултета, назив магистарске тезе, научна област и датум одбране: /</p> <p>6. Научна област из које је стечено академско звање магистра наука: /</p>
<b>III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>
<b>„СЕНТИМЕНТ У САДРЖАЈИМА СА ДРУШТВЕНИХ МРЕЖА КАО ИНСТРУМЕНТ УНАПРЕЂЕЊА ПОСЛОВАЊА ВИСОКОШКОЛСКИХ ИНСТИТУЦИЈА“</b>
<b>IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>
<p>Докторска дисертација кандидаткиње Оливере Грљевић под називом „Сентимент у садржајима са друштвених мрежа као инструмент унапређења пословања високошколских институција“ структуриран је из пет поглавља, који су изложени на укупно 257 страница, А4 формата. Докторска дисертација обухвата 55 илустрација, 39 графикана, 71 табелу и 12 прилога. При</p>

изради докторске дисертације кандидаткиња је консултовала 185 референци. У наставку је приказан садржај дисертације:

## **1. УВОД** (од 1. до 18. странице)

- 1.1. Друштвене мреже и друштвени медији – појам, место и улога у савременом друштву
- 1.2. Анализа података о мишљењима и сентимент
- 1.3. Опис истраживања и приступа
- 1.4. Предметна терминологија

## **2. СЕНТИМЕНТ АНАЛИЗА** (од 19. до 32. странице)

- 2.1. Сентимент анализа и мишљења о производима и услугама
- 2.2. Сентимент анализа у финансијском сектору
- 2.3. Примена сентимент анализе у пословној интелигенцији
- 2.4. Остале примене сентимент анализе
- 2.5. Сентимент анализа у високом образовању

## **3. ИМПЛЕМЕНТАЦИЈА СЕНТИМЕНТ АНАЛИЗЕ** (од 33. до 52. странице)

- 3.1. Методологија сентимент анализе
- 3.2. Изазови у сентимент анализи
  - 3.2.1. Доменска и језичка зависност
  - 3.2.2. Екстракција атрибута
  - 3.2.3. Детекција и руковање негацијом, сарказмом и иронијом
  - 3.2.4. Утврђивање поларитета
  - 3.2.5. Утврђивање субјективности документа
- 3.3. Методе, технике и приступи у сентимент анализи
  - 3.3.1. Алгоритми надгледаног учења у сентимент анализи
    - 3.3.1.1. Стабла одлучивања
    - 3.3.1.2. Линеарни класификатори: Неуронске мреже и Машине потпорног вектора
    - 3.3.1.3. Класификатор заснован на правилима
    - 3.3.1.4. Пробабалистички класификатори
    - 3.3.1.5. Класификатор к-најближих суседа
  - 3.3.2. Примена ненадгледаног учења у сентимент анализи
  - 3.3.3. Класификација заснована на лексиконима
  - 3.3.4. Сентимент анализа заснована на аспектима
    - 3.3.4.1. Сумирање мишљења

## **4. ПРИМЕНА СЕНТИМЕНТ АНАЛИЗЕ У ВИСОКОМ ОБРАЗОВАЊУ** (од 53. до 146. странице)

- 4.1. Методолошки кораци истраживања
  - 4.1.1. Прикупљање и препроцесирање података за основни скуп из друштвених медија
  - 4.1.2. Моделовање садржаја за потребе аотације корпуса
  - 4.1.3. Креирање аотираниог корпуса за сентимент анализу помињања високошколских институција
    - 4.1.3.1. Процес аотације корпуса
    - 4.1.3.2. Процена степена сагласности аотатора
    - 4.1.3.3. Статистичка експлорација и приказ аотираниог корпуса
  - 4.1.4. Формирање речника сентимент појмова
  - 4.1.5. Откривање преовлађујућег сентимента применом интелигентних метода и техника
    - 4.1.5.1. Алгоритми машинског учења коришћени за откривање сентимента
    - 4.1.5.2. Сентимент анализа заснована на лексиконима
  - 4.1.6. Интерпретација и приказ резултата
    - 4.1.6.1. Евалуација класификатора изграђених алгоритмима машинског учења
    - 4.1.6.2. Евалуација модела сентимент анализе засноване на лексиконима
    - 4.1.6.3. Дискусија резултата
- 4.2. Могућност унапређења пословања високошколских институција применом закључака анализе

**5. ЗАКЉУЧАК** (од 147. до 157. странице)

**ЛИТЕРАТУРА** (од 158. до 173. странице)

**ПРИЛОЗИ** (од 174. до 257. странице)

- А. Преузимање садржаја са сајта Оцени професора
- Б. Разделник реченица
- Ц. XML схема анотације
- Д. Упутство за анотацију корпуса
- Е. Процена степена сагласности анотатора
- Ф. Сентимент речи идентификоване у корпусу
- Г. Интензификатори сентимента идентификовани у корпусу
- Х. Доменско специфичне фразе идентификоване у корпусу
- И. Кључне речи негације идентификоване у корпусу
- Ј. Модели класификације сентимента засновани на алгоритмима надгледаног учења
  - Ј.1. Naïve Bayes класификатор
  - Ј.2. Support vector machines класификатор
  - Ј.3. k-Nearest Neighbor класификатор
- К. Сентимент анализа на основу лексикона: Примењени алгоритми
- Л. Евалуационе мере бинарне класификације сентимента засноване на лексиконима

## V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

У уводном поглављу кандидаткиња даје преглед и опис основних концепата и истраживачких области обухваћених докторском дисертацијом. Дефинише концепте друштвених мрежа и друштвених медија, указује на њихов значај у савременом пословању и утицај који имају на понашање потрошача, као и на значај сентимент анализе доступних садржаја на сајтовима друштвених мрежа и медија. У уводу докторске дисертације кандидаткиња описује структуру истраживања и даје преглед истраживачких циљева и хипотезу. Услед двосмислености појединих појмова и приступа на које се истраживање највише ослања, они су сажети и формално дефинисани у уводном поглављу. Њиховим формализовањем омогућен је увид у богат скуп међусобно повезаних потпроблема који сачињавају сентимент анализу.

У другом поглављу дат је преглед употребе сентимент анализе у различитим областима. На овај начин предмет истраживања је стављен у шири контекст и указано је на праксом потврђене користи које различите области пословања имају од примене сентимент анализе. Истраживања из области сентимент анализе, до сада претежно рађена за енглеско говорно подручје, потврђују да садржаји са друштвених медија представљају одличан ресурс за анализу патерна и трендова, за боље разумевање потрошачких навика и преференција и проактивно иновирање производа или услуга. Овакви резултати из других области навели су на закључак да би и високо школство имало користи од примене сентимент анализе над садржајима које студенти генеришу. У другом поглављу кандидаткиња даје и кратак преглед досадашњих истраживања у домену високог образовања. Она се разликују од истраживања, које је предмет докторске дисертације, по извору података и предмету анализе.

У трећем поглављу су систематизоване методолозије и технике у процесу сентимент анализе, као и изазови са којима су се досадашња истраживања суочила. Изучавањем постигнућа у области сентимент анализе, као и развијених метода, техника и приступа омогућено је стицање увида о њиховој примењивости на област високог образовања.

У четвртном поглављу кандидаткиња детаљно приказује реализовано емпиријско истраживање и даје опис добијених резултата. У истраживању је праћена проширена МАТТЕР методолозија, која подразумева прикупљање релевантних података, њихово препроцесирање за потребе аотације корпуса (мануелно обогаћивање прикупљених података релевантним информацијама), моделовање феномена, аотацију садржаја на основу развијене схеме аотације, евалуацију усклађености реализованих аотација, статистичку експлорацију и приказ резултујућег корпуса, формирање речника сентимент појмова за потребе сентимент анализе и коначно спровођење сентимент анализе над припремљеним корпусом.

Релевантни подаци су прикупљени са сајта Оцени професора на којем студенти врше рецензирање наставног особља у Републици Србији. У фази моделовања феномена идентификована је потреба за аотирањем информација о предмету рецензије тзв. аспекту (професор, однос према студентима, предавања, предмет, организација предмета, материјали и други аспект), оријентацији исказаног сентимента, интензитету сентимента, сентимент речима, негацији и опсегу њеног важења. Поступак аотације реализовала су четири квалификована анотатора, а проценом степена сагласности аотација потврђена је њихова усклађеност и квалитет развијеног корпуса. На основу аотираних студентских рецензија кандидаткиња је развила пет сентимент речника: речник позитивних сентимент појмова, речник негативних сентимент појмова, речник интензификатора, речник неутрализатора и речник негације и других речи које мењају сентимент оријентацију наредне речи. Сентимент анализа је реализована применом алгоритама машинског учења: Naïve Bayes, Support Vector Machines и k-Nearest Neighbor. Такође, кандидаткиња је развила систем за откривање сентимента у студентским рецензијама наставног особља на основу развијених сентимент речника. У четвртном поглављу дата је и детаљна дискусија резултата рада развијених модела сентимент анализе. Кандидаткиња завршава поглавље илустровањем могућности примене развијених модела у пословању високошколских институција на примеру Економског факултета у Суботици.

У последњем поглављу изведени су закључци на основу резултата истраживања, указано је на

постигнуте циљеве истраживања, научни допринос дисертације, као и на будуће правце истраживања.

**VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ**

Грљевић О., Бошњак З., Бошњак С.: *Social Media Analytics in Higher Education and Challenges*, International Multidisciplinary Scientific Conferences on Social Sciences and Arts – Psychology & Psychiatry, Sociology & Healthcare, Education, September 3-9 2014, Albena, Bulgaria, Volume I, стр. 883-891, ИСБН 978-619-7105-22-3, ИССН 2367-5659, DOI: 10.5593/sgemsocial2014B11, WOS: 000357835000113 (Kobson – Web of Science, Thomson Reuters). **(M23)**

Грљевић О., Шереш Л., Дебељачки Р.: *Perspectives of Social Media Analytics Application in Higher Education in Serbia*, Facing ICT Challenges in the Era of Social Media monograph, издавач Peter Lang international Academic Publishers, 2014, ИСБН 978-3-631-65383-8. **(M14)**

Грљевић О., Бошњак З.: *Development of Serbian Higher Education Corpus*, 16<sup>th</sup> IEEE International Symposium on Computational Intelligence and Informatics, CINTI 2015, Будимпешта, Мађарска, ИСБН 978-1-4673-8520-6, пп. 177-181. **(M33)**

Грљевић О., Шереш Л., Дебељачки Р.: *Perspectives of Social Media Analytics Application in Higher Education in Serbia*, Proceedings of 4<sup>th</sup> International Conference on Information Technologies and Information Society, 7-9 новембар 2012, Долењске Топлице, Словенија, ИСБН 978-961-93391-1-4. **(M33)**

Грљевић О., Шереш Л.: *Могућност интеграције пословне интелигенције и социјалних медија*, Зборник радова XXXIX Симпозијума о операционим истраживањима, SYM-OP-IS 2012, 25-28. септембар 2012, Тара, Србија, ИСБН 978-86-7488-086-9, стр. 7-11. **(M63)**

Грљевић О., Бошњак З.: *Значај анализе садржаја социјалних медија*, Зборник радова XVIII YU INFO конференције, 29.2 – 3.3.2012., Копаоник, Србија, ИСБН 978-86-85525-09-4, стр. 84-89. **(M63)**

Грљевић О., Бошњак З.: *Друштвени медији у образовању*, SYM-OP-IS 2014 XLI Симпозијум о операционим истраживањима, 16-19. септембар 2014, Дивчибаре, Србија, ИСБН 978-86-7395-325-0, стр. 221-227. **(M63)**

Грљевић О., Бошњак З.: *Utilization of Internet Social Media Sites: Students' Perspective*, CBU International Conference Proceedings – Innovation, Technology Transfer & Education, Volume 2, 2014, Праг, Република Чешка, ИССН: 1805-9961, doi: <http://dx.doi.org/10.12955/cbup.v2.484>, стр. 259-268. **(M33)**

Грљевић О., Бошњак З., Бошњак С.: *Students' Behavior on Social Media Sites – A Data Mining Approach*, 11<sup>th</sup> IEEE International Symposium on Intelligent Systems and Informatics, Sisy2013, 26-28. септембар 2013, Суботица, Србија, IEEE Catalog Number CFP1384C, ИСБН 978-1-4799-0304-7, стр. 347-353. **(M33)**

## VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

На основу спроведеног истраживања кандидаткиња закључује да адекватно екстраховање и припрема садржаја о високошколским институцијама са сајтова друштвених мрежа и медија и њихова консеквентна анализа (анализа сентимента и мишљења) могу указати на скривене шаблоне понашања корисника сајтова. Ово је уједно и полазна хипотеза истраживања, која се на основу описаног истраживања и добијених резултата може прихватити. Адекватно екстраховање података са сајта Оцени професора створило је основу за креирање евалуационог корпуса над којим се успешно реализовала сентимент анализа развијеним моделима. Квалитетна припрема података и процес аотације корпуса осигурали су поузданост евалуационог корпуса. Резултати примене развијених модела указују на широке могућности примене развијеног приступа које се превасходно односе на праћење онлајн репутације институције у односу на конкуренцију, идентификовање преференција постојећих студената и праћење преференција студената на конкурентским институцијама, затим праћење онлајн репутације током времена.

Основни циљ истраживања било је сагледавање начина унапређења пословања високошколских институција ослањајући се на резултате анализе садржаја друштвених медија. Примарни циљ је разложен на подциљеве:

1. Утврдити применљивост метода, техника и приступа анализе садржаја са друштвених медија на област високог образовања, као и могуће ефекте које би њихова примена дала на услуге и сервисе које високошколске институције нуде.

Услед недостатка развијених ресурса, констатована је неопходност креирања евалуационог корпуса који ће омогућити примену алгоритама надгледаног учења у сентимент анализи. Адекватно креиран корпус, који подразумева мануелну аотацију скупа текстуалних података, отвара и могућност примене класификације сентимента на основу лексикона који се деривирају из корпуса. Такође, аотирањем негације у фази развоја корпуса омогућено је њено уважавање при идентификовању сентимент поларитета, као и развој будућих система за детектовање негације у садржајима. Атрибути за реализацију сентимент анализе могу се дефинисати на основу присуства и фреквентности речи у корпусу.

Досадашња истраживања указују на следеће ефекте примене сентимент анализе у домену високог образовања: унапређење интелигентних туторских система, идентификовање предности и недостатака предмета, идентификовање проблема са којима се студенти суочавају и њихово благовремено решавање током трајања семестра, предвиђање студената који ће напустити курс који похађају, односно факултет, и подстицање ангажовања студената на настави. Иако се у другим доменима корисничке рецензије интензивно користе за свеобухватно праћење онлајн репутације и задовољства корисника нема истраживања која се баве питањем студентског задовољства и праћењем онлајн репутације високошколске институције анализом студентских рецензија наставног особља, што представља предмет истраживања докторске дисертације.

2. Идентификовати ограничења и потешкоће у анализи података које се могу појавити као последица специфичног предмета анализе и услед специфичности језика и указати на потенцијална решења.

Реализацијом полазних корака истраживања кандидаткиња је идентификовала следећа ограничења и потешкоће у имплементацији сентимент анализе:

- Као и већину кориснички генерисаног садржаја на вебу, неформални стил писања и непоштовање правописних правила карактерише и студентске рецензије наставног особља. У већини рецензија, које су прикупљене са сајта Оцени професора, идентификован је

изостанак празног места након интерпункцијских знакова, употреба вишеструких узвичника и тачака на половини или на крају реченице, изостанак интерпункцијског знака на крају реченице, дијакритички знакови се нису користили и начињени су бројни граматички пропусти. За коректно спровођење сентимент анализе неопходно је извршити корекцију правописних грешака. Анализом доступних алата, који имају уграђен речник српског језика и који би могли да се употребе за кориговање правописних грешака, кандидаткиња је закључила да они у пракси раде са бројним пропустима и грешкама, те је за потребе истраживања употребила онлајн алат Hascheck који је намењен кориговању правописних грешака у садржајима писаним на хрватском језику. За разлику од других алата, Hascheck омогућава кориговање појединачних појављивања речи у скупу, као и корекцију речи на свим местима појављивања, чиме се штеди време, даје могућност уношења исправног облика речи уколико софтвер у својој бази података нема адекватан облик. Ограничење овог алата се огледа у малој количини података коју може да процесуира у једној анализи. Иако се применом наведеног алата успешно реализовала припрема података, кандидаткиња закључује да за српски језик још увек није развијен адекватан алат за кориговање правописних грешака или да такав алат није јавно доступан.

- У сентимент анализи основна припрема података подразумева сегментацију текста: раздвајање садржаја на реченице (енгл. *sentence splitting*) и раздвајање реченица на речи, односно токенизацију речи. Употреба интерпункцијских знакова, превасходно тачке, представља основу рада развијених алата за поделу садржаја на реченице. Овакви алати су развијени за доминантне светске језике, попут енглеског, немачког и других, и укључују специфичности ових језика, које се првенствено огледају у употреби тачке у скраћеницама након којих не следи крај реченице. У српском језику користи се сет скраћеница карактеристичан искључиво за наше говорно подручје. Такође, датуми и редни бројеви се пишу другачијим навођењем. Наведене специфичности српског језика онемогућиле су исправну примену већ развијених алата за сегментацију садржаја на реченице над садржајима писаним на српском језику. За потребе истраживања које је предмет докторске дисертације кандидаткиња је развила алат за раздвајање садржаја на реченице, који обухвата истакнуте специфичности, а има могућност и додатног проширења.
- Лематизација и тзв. *stemming* такође представљају уобичајене кораке припреме података. Циљ оба поступка јесте уклањање флексија, односно деривираних облика речи, чиме се свака реч своди на свој основни облик – лему. Кандидаткиња је констатовала недостатак алата за лематизацију садржаја писаних на српском језику. Из овог разлога и услед чињенице да је развој оваквог алата уско повезан са лингвистиком, наведени корак препроцесирања садржаја није реализован и може бити предмет будућих истраживања. Очекивано је да би лематизација садржаја подигла прецизност рада развијених модела, јер би се смањила комплексност писаног садржаја свођењем сваке речи на њен основни облик, чиме би постојало мање варијација у анализираним садржајима.

3. Формирати аотирани корпус као основу за спровођење сентимент анализе помињања високошколских институција.

Креирање сопственог евалуационог корпуса мотивисано је недостатком већ развијених корпуса за наше говорно подручје. Такође, услед разлика у доменима примене сентимент анализе и типовима текстова над којима се примењује (рецензија, порука на форуму, коментар на новински чланак и друго), не постоји стандардизована схема аотације која би послужила као основа за креирање референтног корпуса за специфичан задатак сентимент анализе.

Да би се схватио проблем и обезбедила основа за примену сентимент анализе над прикупљеним садржајима, кандидаткиња је дефинисала формални модел који описује елементе исказаног сентимента у прикупљеним релевантним текстовима. Из неструктурираних података кандидаткиња је извела хијерархију аспеката, односно тема о којима студенти пишу у рецензијама, са три аспекта на вишем хијерархијском нивоу

(Професор, Предмет и Други аспект) и четири аспекта на нижим нивоима (Предавања и Однос према студентима, односно Материјали и Организација предмета). Извршила је анализу начина на који се сентимент исказује према идентификованим аспектима и установила да студенти у рецензијама исказују три типа сентимента: позитиван, негативан и неутралан, као и да је присутан различит интензитет позитивног и негативног сентимента. Отуда, модел аотације прикупљених података на нивоу реченице подразумева аотацију аспекта, исказаног сентимента и његовог интензитета, док модел аотације на нивоу речи подразумева означавање индивидуалних сентимент израза, негације и њеног опсега. На основу развијених модела креирала је аотациону схему, која је представљала основу за мануелно означавање студентских рецензија, као и детаљно упутство за аотацију, које може послужити за будуће проширење или адаптацију развијеног корпуса. Резултат мануелне аотације представља корпус, који је омогућио детаљну анализу проблема, анализу лингвистичких феномена и створио основу за процену релевантности појединих потпроблема. За евалуацију квалитета аотације употребила је *Fleiss kappa* и *agr* метрике. Постигнута сагласност аотатора указује на конзистентност у аотацији и високу поузданост развијеног корпуса за домен високог образовања. Анализом корпуса кандидаткиња изводи следеће закључке и резултате:

- Корпус је сачињен из 3.863 рецензија и 6.896 реченица написаних на српском језику, које су преузете са сајта Оцени професора.
- Рецензије студената су уско фокусиране. У 98.22% рецензија предмет рецензирања представља домен високог образовања, док је у 98.7% рецензија исказан сентимент према неком од предефинисаних аспеката.
- По обиму, већина рецензија је средње дужине (од 6 до 30 речи), што захтева идентификовање ширег спектра веома специфичних карактеристика за реализацију класификације.
- Дистрибуција рецензија према поларитету и аспектима је асиметрична. Корпусом доминирају позитивни коментари и аспект Професор. Негативне рецензије су по обиму дуже од позитивних и употребљен вокабулар је разноврснији, што потенцијално може узроковати потешкоће у постизању доброг *Recall*-а за класу негативних рецензија. Неутралан сентимент се ретко појављује у корпусу и ова класа сентимента представља посебан изазов, као и у већини ранијих истраживања. Пошто је у већини рецензија исказан негативан или позитиван сентимент, неутрална категорија се не мора укључити у систем за аутоматско препознавање сентимента.
- Корпусом доминирају сентимент изрази снажног интензитета, чинећи 65% идентификованог сентимента.
- Током аотације корпуса идентификоване су позитивне и негативне сентимент речи. Постојећи сентимент речници се нису могли користити у анализи сентимента због доменске зависности прикупљеног садржаја и специфичног стила писања. Унутар домена уочена је контекстуална зависност појединих речи или фраза. Такође, идентификован је сет фраза, које студенти често користе у рецензијама, а које носе јасан позитиван или негативан сентимент. На основу идентификованих сентимент речи и фраза у корпусу кандидаткиња је креирала речнике позитивних и негативних сентимент појмова. Они представљају основу за спровођење класификације сентимента на основу лексикона.
- Кандидаткиња констатује да су сентимент речи претежно придеви на основу чега закључује да би ПОС означавање садржаја било корисно за аутоматско креирање сентимент речника.
- Током аотације су издвојени интензификатори и неутрализатори сентимента. Реч је о специфичном сету речи које се употребљавају у домену високог образовања, а које, зависно од контекста у ком се користе, могу појачати или умањити интензитет исказаног сентимента у садржају. У функцији интензификатора сентимента јављају се прилози (за количину и начин), придеви, речце и везници. На основу издвојених речи и фраза кандидаткиња је креирала два речника: речник интензификатора и речник неутрализатора, који су употребљени у каснијим анализама садржаја.



- Стил писања у рецензијама омогућава једноставно детектовање негације услед недвосмислене употребе кључних речи негације. Међутим, кандидаткиња констатује проблем вишеструке негације, која са једне стране може узроковати промену сентимент поларитета и значења реченице, док су у другачијем контексту овакве промене изузете – сентимент поларитет упркос појави негације остаје непромењен. На основу идентификованих облика негације креирала је речник негације који је употребљен у класификацији сентимента на основу лексикона.
- Зависне условне реченице се користе за исказивање позитивних коментара изразитог интензитета.
- Модални глаголи (нпр. моћи, требати) и компаративи придева и прилога (нпр. боље) користе се за исказивање негативног сентимента. Такође, модални глаголи у потенцијалу (нпр. требало би, могло би) користе се за исказивање препорука у коментарима.

Анализом корпуса кандидаткиња указује на комплексност лингвистичког феномена који се моделује: мишљења се исказују експлицитно и имплицитно, поједине речи јачају интензитет сентимента, слабе га или мењају поларитет, присутни су вишеструки искази сентимента усмерени ка различитим аспектима, као и сентимент различитог поларитета. Идентификоване специфичности и суптилности у стилу писања и начину на који се исказује сентимент изузетно су битне за адекватно лингвистичко моделовање исказаног мишљења.

4. Идентификовати најадекватнији начин за сентимент анализу српског језика, односно сентимент анализу помињања високошколских институција, кроз систематичну анализу и поређење одабраних алгоритама.

Анотације садржаја на нивоу реченице и речи, дате за сваку рецензију, резултовале су неопходним информацијама за реализацију класификације рецензија према сентимент поларитету или различитим аспектима, за идентификовање аспеката са најснажнијим и најслабијим интензитетом сентимента, идентификовање карактеристика које студенти преферирају у сваком аспекту, компарацију сентимента студената са различитих високошколских институција, као и мониторинг и анализу промена у мишљењима и исказаном сентименту током времена. Све ово је омогућило да се стекну вредна сазнања о ставовима и мишљењу студената о високошколској институцији коју похађају. Развијени модели користе само одређене идентификоване специфичности о стилу писања и структури реченица, које се односе на интензификаторе и неутрализаторе сентимента и негацију. Међутим, детаљна запажања створила су основу за генерисање специфичних правила рада аутоматизованих система за препознавање сентимента, која ће бити предмет будућих истраживања.

Проблем идентификовања сентимента у студентским рецензијама дефинисан је као бинарни проблем класификације (позитиван сентимент наспрот негативном) у моделима базираним на примени алгоритама машинског учења или као троструки проблем класификације (позитиван, негативан и неутралан сентимент) у моделима базираним на примени сентимент лексикона.

Како је идентификовање сета атрибута, који ће се користити у класификационим моделима, важан корак, кандидаткиња је реализовала низ експеримената да би идентификовала оптималан сет атрибута. Кандидаткиња закључује следеће:

- У класификацији сентимента применом алгоритама машинског учења употреба биграма као атрибута класификације даје супериорије резултате у поређењу са униграмима, триграмима и четириграмима.
- Филтрирање стоп речи даје значајан помак у тачности класификације.

- Најбоље перформансе у класификацији сентимента применом алгоритама машинског учења на нивоу рецензија постижу се применом SVM алгорита – 84.94% тачности и Ф-мера износи 89.66% за позитивну класу, а 72.32% за негативну. *kNN* алгорита, где је  $k=3$ , резултује нешто нижом тачности од 84.53%, али приближно успешно класификује рецензије у скупу за обучавање и у скупу за тестирање и резултује нешто вишом Ф-мером за негативну класу – 74.28%. У контексту пословања релевантнија је негативна класа која указује на незадовољство студената и потребна побољшања у пословању. Из тог разлога кандидаткиња закључује да је *kNN* алгорита боље решење за дати домен.
  - У класификацији сентимента применом алгоритама машинског учења на нивоу реченице најбоље перформансе даје SVM алгорита – 80.13% тачности у класификовању непознатих примера реченица. Ф-мера за позитивну класу износи 84.53%, док за негативну износи 72.21%.
  - У класификацији сентимента применом сентимент лексикона најбољи резултати се постижу увођењем засебних речника за позитивне и негативне сентимент речи, интензификаторе, неутрализаторе, и негацију и друге речи које мењају сентимент наредне речи у реченици.
  - Мануелно кориговање речника значајно побољшава резултате класификације.
  - Тачност класификације на нивоу рецензија износи 80.71%. Ф-мера износи 89.75% за позитивну класу, 77.28% за негативну и 26.04% за неутралну.
  - Тачност класификације на нивоу реченица износи 73.70%. Ф-мера износи 86.05% за позитивну класу, 71.80% за негативну и 4.49% за неутралну.
  - Незаступљеност неутралне класе у корпусу може се сматрати примарним узроком ниске прецизности њене класификације.
  - На нивоу рецензија боље перформансе се постижу применом класификације засноване на лексиконима, док се на нивоу реченица боље перформансе постижу применом алгоритама машинског учења.
  - Категорија лажно негативних вредности представља проблем подједнако код приступа заснованом на алгоритмима машинског учења, као и код приступа заснованом на сентимент лексиконима. Реч је о ситуацијама када систем идентификује присуство одређене негативне речи у анализираном садржају, али не препозна свеукупно позитиван контекст.
5. Израдити методолошки оквир за екстраховање садржаја о помињањима високошколских институција Републике Србије на друштвеним медијима и спровођење поступка сентимент анализе за потребе управљања репутацијом.

Кандидаткиња детаљно описује процес прикупљања кориснички генерисаног садржаја са портала Оцени професора, неопходно препроцесирање података које претходи креирању корпуса за домен високог образовања поступком аотације прикупљеног садржаја и анализу креираног корпуса применом развијених сентимент речника и алгоритама машинског учења. Наведени кораци представљају свеобухватан методолошки оквир који омогућава спровођење поступка сентимент анализе применом надгледаног учења и лексикона.

Спроведено истраживање је створило неопходне услове за давање одговора на основни циљ истраживања, односно сагледавање могућности за унапређења пословања високошколских институција ослањањем на резултате анализа садржаја друштвених медија. Кандидаткиња испитује могућности употребе развијених модела за праћење онлајн репутације и идентификовање извора незадовољства студената на примеру Економског факултета у Суботици:

- Идентификује глобално задовољство студената сумирањем позитивних и негативних рецензија на нивоу институције.

- Врши компарацију онлајн репутације институција.
- Врши анализу позитивних и негативних рецензија по аспектима како би идентификовала аспект пословања високошколске институције који студенти фаворизују, односно аспект који у највећој мери критикују.
- Пореди институције према исказаном задовољству и незадовољству по аспектима.
- Анализира асоцијативна правила, биграме и триграме унутар позитивних и негативних рецензија, чиме се стиче увид у доминантне и важне фразе којима студенти исказују своје задовољство или незадовољство, а које указују на подручја пословања која дају компаративну предност, односно која треба унапредити, респективно.
- Прати исказани сентимент студената кроз време са циљем идентификовања осцилација у задовољству и незадовољству студената.

На основу спроведених анализа кандидаткиња закључује да су студенти Економског факултета у Суботици претежно задовољни искуством студирања. У поређењу са конкурентским факултетима, према коментарима остављеним на сајту Оцени професора, закључује да је Економски факултет у Суботици боље оцењен од стране својих студената по питању Односа професора према студентима, те да је предност Економског факултета у Суботици квалитетан однос наставника према студентима и постојање квалитетних наставних материјала и уџбеника. Међутим, студенти Економског факултета у Суботици исказују виши степен незадовољства по питању Предавања и Предмета, што се огледа у лошијим коментарима. Као највећи извор незадовољства студената идентификовала је аспект Предавања. Студенти оштрим речником и интензивним сентиментом критикују квалитет појединих предавања, описујући их као неинтересантна и застарела. Надаље, из коментара се види да професори недовољно укључују студенте у наставни процес, што утиче на (не)заинтересованост студената и (не)пажњу на часу. Студенти оштро критикују и незаинтересованост појединих професора за предавања, дидактичке методе, тј. начин на који се знање преноси, и форсирање учења градива напамет и „од речи до речи“. Праћењем кретања сентимент поларитета у рецензијама кроз време идентификовала је јасне цикличне осцилације у негативним коментарима, које су повезане са испитним роковима, док у позитивним коментарима сличне шаблоне кретања није запазила.

Реализацијом постављених истраживачких циљева и спроведеним експериментима кандидаткиња је показала да адекватно екстраховање и припрема садржаја о високошколским институцијама са сајтова друштвених медија и њихова консеквентна анализа (анализа сентимента и мишљења) указују на скривене шаблоне понашања корисника сајтова. Ослушкивањем сентимента студената у коментарима о институцији на друштвеним мрежама и медијима, високошколским институцијама се пружа могућност деривирања битних сазнања и превођења стечених сазнања у акције, те се са једне стране могу предузети корективне мере како би се превазишли идентификовани проблеми, а са друге стране, отвара се могућност употребе стечених сазнања за предузимање мера ка унапређењу студентског искуства и свеукупне слике јавности о институцији. Примена развијених модела сентимент анализе створила је услове за праћење онлајн репутације институција, за компарацију са другим институцијама, идентификовање преференција и незадовољства студената и указала на широке могућности примене сентимент анализе у домену високог образовања.

## VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА

Докторска дисертација кандидаткиње Оливере Грљевић у потпуности је испунила циљеве дефинисане у пријави теме. Емпиријско истраживање је у складу са насловом докторске дисертације. Кандидаткиња је дала одговор на постављена истраживачка питања и реализовала дефинисане циљеве истраживања применом квантитативне и квалитативне методе анализе и тумачења података. У четвртом поглављу кандидаткиња детаљно описује методологију истраживања и на разумљив и концизан начин презентује, анализира и тумачи резултате према фазама истраживања. Истраживање резултује адекватним научним и примењивим резултатима, који су у докторској дисертацији презентовани кроз текст, табеле и слике.

Кандидаткиња је успешно реализовала постављене циљеве истраживања и тестирала хипотезу, стога Комисија доноси позитивну оцену начина приказа и тумачења резултата истраживања.

## IX КОНАЧНА ОЦЕНА ДОКТОРСKE ДИСЕРТАЦИЈЕ:

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме

Докторска дисертација је у потпуности **написана у складу са образложењима** које је кандидаткиња навела у пријави теме.

2. Да ли дисертација садржи све битне елементе

Докторска дисертација **садржи све битне елементе**: увод, три поглавља у којима је разрађен предмет истраживања и описани истраживачки поступци и примењене методе, те закључак у којем су приказани и објашњени резултати истраживања. Поред ових елемената, дате су библиографске референце и укупно 12 прилога којима се поткрепљује свака фаза рада и пружа увид у извршене експерименте.

3. По чему је дисертација оригиналан допринос науци

Резултати истраживања докторске дисертације доприносе бољем разумевању комплексности сентимент анализе. Теоријски допринос докторске дисертације огледа се у систематизацији релевантних теоријских концепата који објашњавају актуелна достигнућа и могућности у имплементацији сентимент анализе. Услед чињенице да је примена сентимент анализе незаступљена истраживачка област у Србији, систематизацијом владајућих ставова, метода, техника и приступа креирана је адекватна библиотечка јединица на српском језику.

У Србији до сада није спроведено слично истраживање те је научно-истраживачки допринос докторске дисертације вишеструк, а преваходно се односи на:

1. Систематизацију методолошких корака и формирање методолошког оквира за примену сентимент анализе;
2. Развој алата за препроцесирање садржаја на српском језику, тзв. разделника реченица;
3. Развој евалуационог корпуса на српском језику за домен високог образовања, чиме су створени неопходни предуслови за имплементацију различитих приступа сентимент анализи и њихову евалуацију. Јавно доступни корпуси на српском језику потпуно изостају, те ово представља први допринос ове врсте;
4. Израду упутства за анотацију корпуса, које може послужити за будуће проширење и адаптацију корпуса;
5. Креирање сентимент речника, речника интензификатора и неутрализатора, као и речника речи које мењају значење и сентимент наредне речи у реченици;
6. Детаљну лингвистичку анализу садржаја, која ће послужити за генерисање специфичних правила на основу којих ће радити будући системи за аутоматско препознавање сентимента;
7. Развој модела и алата за сентимент анализу садржаја на српском језику;

8. Систематизацију могућности примене развијених модела у праћењу онлајн репутације и за унапређење пословања високошколских институција, као и постављање смерница за корективне акције.
4. Недостаци дисертације и њихов утицај на резултат истраживања  На основу претходног изнетог, може се закључити да докторска дисертација нема уочених недостатака који би могли утицати на постизање, интерпретацију или презентовање дефинисаних циљева истраживања.
<b>X ПРЕДЛОГ:</b>
На основу укупне оцене дисертације, комисија предлаже:
На основу напред дефинисаних ставова и у складу са анализом постигнутих резултата и доприноса докторске дисертације, Комисија предлаже Наставно-научном већу Економског факултета у Суботици да прихвати <b>позитивну оцену докторске дисертације</b> „Сентимент у садржајима са друштвених мрежа као инструмент унапређења пословања високошколских институција“ и да кандидаткињи Оливери Грљевић одобри јавну одбрану докторске дисертације пред усвојеном Комисијом.

У Суботици, 2016. године

ЧЛАНОВИ КОМИСИЈЕ:

Проф. др Зита Бошњак, редовни професор  
Економски факултет у Суботици,  
Универзитет у Новом Саду

---

Проф. др Саша Бошњак, редовни професор  
Економски факултет у Суботици,  
Универзитет у Новом Саду

---

Проф. др Александар Ковачевић, доцент  
Факултет техничких наука,  
Универзитет у Новом Саду

---

НАПОМЕНА: Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај.