Univerzitet u Beogradu

Elektrotehnički fakultet



Analiza socijalnih mreža

Izveštaj

|  |  |
| --- | --- |
| Profesor: | Student: |
| prof. dr. Marko Mišić,  prof. dr. Jelica Protić,  as. ms Predrag Obradović | Andrijana Miković 2023/3042 |

Beograd, avgust 2024.

Sadržaj

[Sadržaj 0](#_Toc175077091)

[1. Uvod 1](#_Toc175077092)

[2. Analiza problema 3](#_Toc175077093)

[2.1. Ulazni fajlovi 3](#_Toc175077094)

[3. Baza podataka 5](#_Toc175077095)

[3.1. Opšte informacije 5](#_Toc175077096)

[4. Statistička obrada podataka 7](#_Toc175077097)

[4.1 . Analiza produktivnosti autora 7](#_Toc175077098)

[4.2 . Analiza broja koautora po svakom autoru 8](#_Toc175077099)

[4.3 . H-indeks 9](#_Toc175077100)

[4.4 . Produktivnost katedri 10](#_Toc175077101)

[4.5 . Najproduktivnija godina **Error! Bookmark not defined.**](#_Toc175077102)

[4.6 . Broj radova po časopisima 11](#_Toc175077103)

[5. Модели машинског учења Error! Bookmark not defined.](#_Toc175077104)

[5.1. Векторизација **Error! Bookmark not defined.**](#_Toc175077105)

[5.2. Логистичка регресија **Error! Bookmark not defined.**](#_Toc175077106)

[5.3. КНН – к најближих комшија **Error! Bookmark not defined.**](#_Toc175077107)

[5.4. Стабло одлучивања **Error! Bookmark not defined.**](#_Toc175077108)

[5.5. Мултиномијални наивни Бајесов класификатор **Error! Bookmark not defined.**](#_Toc175077109)

[6. Модели неуралних мрежа 17](#_Toc175077110)

[*6.1.* Токенизација и *GloVe* **Error! Bookmark not defined.**](#_Toc175077111)

[*6.2.* *LSTM* **Error! Bookmark not defined.**](#_Toc175077112)

[*6.3.* *BiLSTM* **Error! Bookmark not defined.**](#_Toc175077113)

[*6.4.* *CNN* 27](#_Toc175077114)

[6.5. Креирање слојева код модела неуралних мрежа 27](#_Toc175077115)

[7. Резултати 28](#_Toc175077116)

[7.1. Метрика 28](#_Toc175077117)

[7.1.1. Тачност 28](#_Toc175077118)

[7.1.2. Прецизност 28](#_Toc175077119)

[7.1.3. Одзив 29](#_Toc175077120)

[7.1.4. Ф1-мера 29](#_Toc175077121)

[7.1.5. Матрица конфузије 29](#_Toc175077122)

[7.1.6. Multi-class Log Loss 29](#_Toc175077123)

[7.2. Резултати добијени алгоритмима машинског учења 30](#_Toc175077124)

[7.3. Резултати добијени неуронским мрежама 33](#_Toc175077125)

[8. Закључак 38](#_Toc175077126)

[Литература Error! Bookmark not defined.](#_Toc175077127)

[Списак слика Error! Bookmark not defined.](#_Toc175077128)

[Списак табела Error! Bookmark not defined.](#_Toc175077129)

1. Uvod

Epidemije, kao fenomen neočekivano brzog širenja zaraznih bolesti unutar određenih populacija ili geografskih regiona tokom ograničenog vremenskog perioda, vekovima su prepoznate kao jedan od najvažnijih izazova za zdravlje čovečanstva. Od najranijih civilizacija pa do savremenog doba, epidemije su uzrokovale masovne gubitke ljudskih života, duboko pogađajući društva i ekonomije širom sveta. Istorijski značajne epidemije, poput kuge u srednjem veku, koja je prepolovila populaciju Evrope, ili pandemija kolere u XIX veku, koja se proširila preko kontinenata usled porasta međunarodnih trgovinskih puteva, ostavile su trajne posledice na čovečanstvo.

Globalizacija u XX veku je olakšala međunarodnu razmenu ljudi i dobara. Kao i mogućnost saradnje naučnika iz raznih oblasti, pa i medicine. Ovakva globalizacija je takođe dovela do toga da se zarazne bolesti brže šire. Što je moglo da se primeti i pandemijom koronavirusa SARS-CoV-2 (Severe Acute Respiratory 3 Syndrome Coronavirus 2), slika 1.1, koja je zadesila svet početkom 2020. godine. Ova pandemija je na dramatičan način demonstrirala koliko je savremeno društvo ranjivo na pojavu novih i neočekivanih epidemija, uprkos napretku u medicini i javnom zdravlju. Istovremeno, pandemija je ukazala na ključnu ulogu koju imaju oblasti medicine koje se bave proučavanje epidemiologije i zaraznih bolesti, u borbi protiv pandemija u savremenom svetu.

A close-up of a virus

Description automatically generated

Slika 1.1. SARS-CoV-2

Predmet istraživanja u okviru ovog projektnog zadatka je kvantitativna i kvalitativna analiza naučne produkcije Medicinskog fakulteta Univerziteta u Beogradu, sa posebnim fokusom na oblasti medicinske nauke koje su direktno povezane sa proučavanjem i suzbijanjem epidemija zaraznih bolesti. U analizu će biti uključeni naučni radovi i kolaboracije zaposlenih na Katedri za imunologiju, Katedri za epidemiologiju, Katedri za infektivne bolesti i Katedri za mikrobiologiju. Kroz ovu analizu, biće ispitani obim i kvalitet naučne produkcije, kao i stepen saradnje među zaposlenima na pomenutim katedrama, sa ciljem da se utvrdi kako doprinosi istraživača sa Medicinskog fakulteta Univerziteta u Beogradu doprinose globalnom razumevanju i borbi protiv epidemija..

Cilj ovog rada je da kroz obradu prikupljenih podataka, i analizu podataka iz oblasti medicinskih nauka prikaže stanje naučne produkcije na Fakultetu, mapira mreže saradnje među istraživačima, i identifikuje ključne pojedince i grupe koji doprinose u borbi protiv epidemija. Analiza će obuhvatiti publikacije kao što su knjige, disertacije, naučni radovi u časopisima, kao i radovi predstavljeni na međunarodnim i domaćim naučnim konferencijama, s posebnim akcentom na one publikacije koje su objavljene u naučnim časopisima sa impakt faktorom. Posebna pažnja će biti posvećena razmatranju naučne produkcije sa aspekta njene vidljivosti i uticaja u globalnoj naučnoj zajednici, što uključuje analizu citiranosti radova, mreže kolaboracija.

1. Analiza problema

U ovom poglavlju biće analizirani ulazni fajlovi koji se koriste pri analizi naučne produkcija iz oblasti epidemiologije i zaraznih bolesti na Medicinskom fakultetu u Beogradu. Koji predsavlja problem konstruisanja mreže saradnje na osnovu ulaznih podataka, kao i donošenje zaključka o osobinama i načinu funkcionisanja ljudi, kao pojedinca, i kao predstavnika svoje katedre. Ovi fajlovi su formirani na osnovu upita u bazu *Scopus.*

* 1. Ulazni fajlovi

Podaci prikupljeni za analiz, dostupni su u vidu odgovarajućih fajlova. Primarni skup podataka je predstavljen u okviru fajla *autori.xlsx,* i on sadrži informacije o autorima koji su se vodili kao zaposleni u nastavnim zvanjima na katedrama za imunologiju, epidemiologiju, infektivne bolesti i mikrobiologiju, Medicinskog fakultea Univerziteta u Beogradu, u oktobru 2020. godine, što je prikupljeno sa internet prezentacija ovih katedra. Autori naučnih publikacija koji se javljaju u okviru skupa podataka, a ne nalaze se u ovom primarnom skupu, nisu od interesa za analizu.

Fajl *autori.xlsx* sadrži sledeće informacije:

* *Ime*
* *Prezime*
* *Katedra* – na kojoj je posmatrani naučnik angažovan u oktobru 2020.
* *H indeks* – kao mera produktivnosti i citiranosti naučnika
* *Broj radova*

Ostali podaci iz primarnog skupa podataka, predstavljaju podatke o radovima dobijene iz indeksne baze načnih radova *Scopus,*  zaključno sa oktobrom 2020. Radovi su podeljeni u četri dadoteke gde svaka odgovara jednoj od prethudno pomenute katedre, *epidemiologija.xlsx, imunologija.xlsx, infektivne\_bolesti.xlsx, mikrobiologija.xlsx.*

Podaci se sastoje od sledećih kolona:

1. *Author* – ime zaposlenog kome je pridružena odgovarajuća publikacija u Scopus bazi,

* *Authors* – svi autori rada,
* *Title* – naslov publikacije,
* *Year* – godina izdavanja,
* *Source title* – ime časopisa u kome je rad objavljen,
* *Volume* – volumen časopisa u kome je rad objavljen,
* *Issue* – redni broj sveske časopisa u kome je rad objavljen,
* *Art. No.* – identifikacioni broj rada,
* *Page start* – početna stranica rada u časopisu,
* *Page end* – poslednja stranica rada u časopisu,
* *Page count* – ukupan broj stranica rada,
* *Cited by* – broj citata drugih autora prema Scopus bazi,
* *Link* – link ka izvoru,
* *Document Type* – tip rada,
* *Source* – indeksna baza koja predstavlja izvor podataka (Scopus).

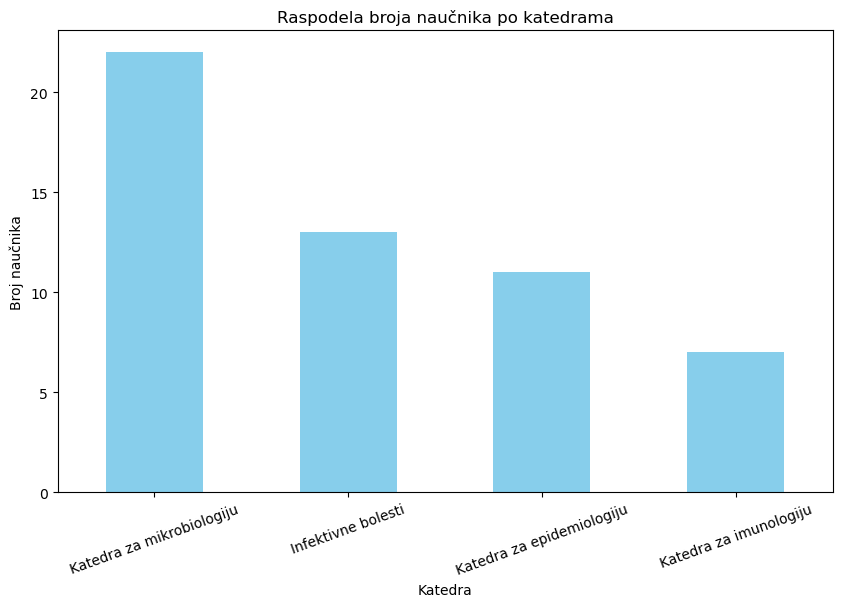
U analizi su od interesa samo publikacije tipa Article, Article in Press, Review, Book Chapter, Letter, Note, dok se ostale mogu zanemariti

1. Baza podataka

U ovom poglavlju biće dat opis baze podataka korišćene u ovom istraživanju.

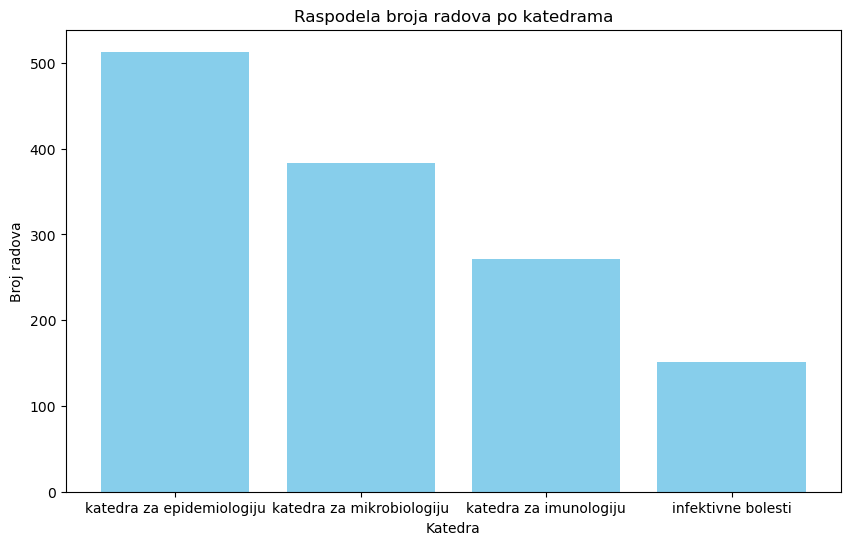
* 1. Opšte informacije

Fajl koji sadrži informacije o naučnicima, od interesa inicijalno sadrži informacije o 59 naučnika, analizom sadržaja fajla primećujemo da od toga postoji 6 naučnika koji ne sadrže informaciju u H indeksu i broju radova. Predmet ovog istraživanja je analiza naučne produkcije, samim tim nam zaposleni za koje nemamo informacije o radovima u posmatranom periodu nisu od interesa, i otklanjamo ih iz inicijalnog skupa podataka. Nakon čega je analizirana raspodela posmatranih naučnika po katedrama, koja je prikazana na slici 3.1.1. Sa koje primećujemo da aktivni naučnici koji se bave proučavanjem epidemiologije i zaraznih bolesti nisu ravnomerno raspoređeni po katedrama od interesa.



Slika 3.1.1. Raspodela broja naučnika po katedrama

Vršimo analizu radova prikupljenih iz *Scopus* baze, primećujemo da postoje duplikati. Konačan skup podataka je dobijen otklanjanjem među duplikatima, onih vrednosti koji sadrže manje informacija, fokusirajući se na broj citiranosti. Nakon uklanjanja podataka navedenim principom u bazi su ostala još dva para duplikata. Kod jednog od parova duplikata imamo dva puta identične informacije, tako da je nasumično vršeno uklanjanje jednog zapisa, dok je za drugi par izvršena provera dostupnosti izvora u *Scopus* bazi, i otklonjen je onaj duplikat čiji izvorni link ne vodi na traženi rad. Na slici

****

Slika 3.1.2. Raspodela broja radova po katedrama

Možemo primetiti da iako se na katedri za epidemiologiju nalazi manje autora u odnosu na katedru za mikrobiologiju (slika 3.1.1.), oni su učestvovali u pisanju većeg broja radova, čije vrednosti možemo videti u tabeli 3.1.1.

Tabela 3.1.1. Raspodela broja radova po katedrama

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Infektivne bolesti | Katedra za epidemiologiju | Katedra za imunologiju | Katedra za mikrobiologiju |
| Broj radova | 151 | 513 | 271 | 383 |

1. Statistička obrada podataka

Izvršena je statistička obrada podataka, sa fokusom na istraživačka pitanja data u postavci projekta [1].

4.1 . Analiza produktivnosti autora

Analiziran je očišćeni skup radova, u tabeli 4.1.1. vidimo 10 istraživača sa najvećim brojem radova na osnovu analiziranih podataka.

Tabela 4.1.1. Prikaz autora sa najviše radova celovito brojanje

|  |  |  |
| --- | --- | --- |
| Autor | Katedra | Broj radova |
| Tatjana Pekmezović | Katedra za epidemiologiju | 262 |
| Vladimir Trajković | Katedra za imunologiju | 155 |
| Vera Pravica | Katedra za imunologiju | 87 |
| Ljiljana Marković-Denić | Katedra za epidemiologiju | 84 |
| Tatjana Gazibar | Katedra za epidemiologiju | 77 |

Uviđamo da Tatjana Pekmezović prednjači sa brojem radova, nakon nje se nalazi Vladimir Trajković sa 107 radova manje u odnosu na svoju koleginicom. Tatjana pripada katedri za epidemiologiju, koja kao što smo videli u analizi u prethodnom poglavlju broji najviše radova, njihova 3 autora se nalaze u prvih 5 po broju radova. Dok zanimljiva je činjenica da je druga katedra po broju radova iz prethodne analize katedra za mikrobiologiju, iako se njihovi autori ne nalaze u prvih 5. Prethodna tabela je dobijena celovitim brojanjem radova po autoru, dok ako uzmemo u obzir i broj koautora sa Medicinskog fakulteta koji su se bavili istim radom, dobijemo tabelu 4.1.2.

Tabela 4.1.2. Prikaz autora sa najviše radova frakcijalno brojanje

|  |  |  |
| --- | --- | --- |
| Autor | Katedra | Broj radova |
| Tatjana Pekmezović | Katedra za epidemiologiju | 34.90 |
| Vladimir Trajković | Katedra za imunologiju | 19.14 |
| Ljiljana Marković-Denić | Katedra za epidemiologiju | 13.53 |
| Vera Pravica | Katedra za imunologiju | 11.73 |
| Tatjana Gazibar | Katedra za epidemiologiju | 9.87 |

Broj koautora je dobijen kao suma broja koautora sa fakulteta, i broja koautora van njega. Primećujemo da top 5 autora po broju radova su isti i sa celovitim i frakcijalnim brojanjem, samo su Vera Pravica i Ljiljana Marković-Denić premotavale mesta.

4.2 . Analiza broja koautora po svakom autoru

Analizom polja *Authors* u kome se nalaze informacije o koautorima na posmatranom radu, odvojenih zarezom. Utvrđen je prosečan broj koautora po svakom autoru koji nam je od interesa i čije ime i prezime se nalazi u *Author* koloni, treba napomenuti da nije vršena analiza stringa koji predstavlja koautore, već je uzeta pretpostavka o ispravnosti, da je svaki autor odvojen zarezom i da se posmatrani autor sa Medicinskog fakulteta univerziteta u Beogradu ne nalazi u toj listi. U tabeli 4.2.1. se nalazi 5 autora sa najvećim brojem koautora u proseku na svojim radovima.

Tabela 4.2.2. Prikaz autora sa najviše koautora u proseku na svojim radovima

|  |  |  |
| --- | --- | --- |
| Autor | Katedra | Broj koautora |
| Elenora Dubljanin | Katedra za mikrobiologiju | 390.74 |
| Isidora Vujičić | Katedra za epidemiologiju | 343.81 |
| Valentina Arsić-Arsenijević | Katedra za mikrobiologiju | 71.82 |
| Maja Stanojević | Katedra za mikrobiologiju | 32.34 |
| Vladimir Trajković | Katedra za imunologiju | 32.00 |

Možemo primetiti da se prva dva autora iz tabele 4.2.2. ne nalaze u tabelama koje prikazuju broj autora sa najvećim brojem radova (4.1.1. i 4.1.2.), dok se ostali nalaze u sva tri posmatrana rezultata. Što znači da ova dva autora imaju manji broj radova, a oni su pisani u saradnji sa velikim brojem koautora. Sa slike 4.2.1. možemo primetiti da najveći broj autora ima manje od 10 koautora na svojim radovima, za koje možemo pretpostaviti da su oni više doprineli tim radovima.

A graph with blue squares

Description automatically generatedSlika 4.2.1. Raspodela broja prosečnog broja koautora po autoru

4.3 . H-indeks

H-indeks je način da se u jednom podatku objedini informacija o broju publikacija autora i broju citata. Naučnik poseduje j-indeks od h ukoliko ima najmanje h radova, od kojih je svaki barem h puta citiran. [2]. U primarnom skupu podataka koji sadrži informaciju o autorima od interesa, imamo informaciju o njihovom H-indeksu, za te podatke znamo samo da su dobijeni sa prezentacija odgovarajućih katedri, vršeno je upoređivanje vrednosti tog H-indeksa sa izračunatim H-indeksom na osnovu prethodne definicije podacima o radovima iz *Scopus* baze. U tabeli 4.3.1. su prikazani dobijeni i izračunati H-indeks samo za autore kod kojih postoji razlika između te dve vrednosti.

Tabela 4.3.1. Prikaz autora za koje postoji razlika između izračunatog i dobijenog H-indeksa

|  |  |  |
| --- | --- | --- |
| Autor | H-indeks dobijen | H-indeks izračunat |
| Vera Pravica | 35 | 33 |
| Nataša Vučković-Opavski | 1 | 0 |
| Aleksandar Džamić | 6 | 5 |
| Vladimir Trajković | 37 | 36 |
| Emina Milošević | 8 | 7 |

Primećujemo da je izračunati indeks manji od dobijenog, razlog odstupanja može biti nepoklapanje baza radova na osnovu kojih je računat H-indeks za autore, u ovom istraživanju je korišćena *Scopus* baza, dok nemamo informaciju na osnovu koje baze je dobijen izloženi H-indeks na sajtu fakulteta.

4.4 . Produktivnost

Produktivnost katedri na osnovu naučne produkcije i citiranosti u časopisima možemo da vidimo u tabeli 4.4.1.

Tabela 4.4.1. Produktivnost katedri

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Katedra | Ukupan broj radova | Ukupan broj citata | Rang na osnovu rada | Rang na osnovu citiranosti | Produktivnost |
| Katedra za Mikrobiologiju | 383 | 19081 | 2 | 1 | 3 |
| Katedra za Epidemiologiju | 513 | 6601 | 1 | 3 | 4 |
| Katedra za Imunologiju | 271 | 16380 | 3 | 2 | 5 |
| Infektivne bolesti | 151 | 717 | 4 | 4 | 8 |

Po ukupnom broju radova najproduktivnija katedra je Katedra za Epidemiologiju, dok ona po broju citata zauzima treće mesto, po broju citata najproduktivnija katedra je Katedra za Mikrobiologiju, koja je druga po ukupnom broju radova, zbog čega je prva po produktivnosti ako se gledaju obe vrednosti.

A ako posmatramo produktivnost istraživača sa Medicinskog fakulteta Univerziteta u Beogradu, na nivou celog fakulteta, iz podataka možemo videti da su autori bili najproduktivniji 2014 godine, kada su učestvovali u pisanju 107 radova, koji su citirani ukupno 2068 puta, dok produktivnost po godinama na nivou katedri možemo videti u tabeli 4.5.1.

Tabela 4.4.2. Najproduktivnije godine po katedri

|  |  |  |  |
| --- | --- | --- | --- |
| Katedra | Godina | Broj radova | Broj citata |
| Katedra za Mikrobiologiju | 2014 | 50 | 2170 |
| Katedra za Epidemiologiju | 2018 | 50 | 2170 |
| Katedra za Imunologiju | 2012 | 21 | 3118 |
| Infektivne bolesti | 2019 | 16 | 32 |

Iz tabele 4.4.2. vidimo da je katedra za epidemiologiju doprinela u preko trećini radova, u najproduktivnijoj godini na nivou fakulteta.

4.5 . Raspodela po časopisima

Na osnovu datog skupa podata, možemo videti u kojim časopisima ovi autori pretežno objavljuju radove koji su od interesa za oblast epidemiologije i zaraznih bolesti. U tabeli 4.6.1. vidimo pet najpopularnijih časopisa na osnovu ulaznog skupa podataka.

Tabela 4.5.1. Časopisi sa najviše objavljenih radova

|  |  |
| --- | --- |
| Časopis | Broj objavljenih radova |
| Srpski arhiv za celokupno lekarstvo | 80 |
| Vojnosanitetski pregled | 37 |
| Archives of biological sciences | 34 |
| Medicinski pregled | 27 |
| Vojnosanitetski pregled. Military-medical and pharmaceutical review | 24 |

Iz tabele 4.6.1. možemo videti da autori pretežno objavljuju radove u domaćim časopisima. A ako posmatramo učestalost objavljivanja na nivou pojedinačnih katedri, možemo iz tabele 4.5.2. videti da sve katedre, osim katedre za imunologiju objavljuju u proseku 2 rada po časopisu. Takođe ako pogledamo nazive časopisa koji po broju objavljenih radova zauzimaju 2 i 5 mesto, možemo zaključiti da predstavljaju isti časopisi gde u jednom slučaju imamo samo naziv na srpskom jeziku, a na drugim mestima imamo naziv na srpskom i na engleskom, kao naziv časopisa. Iz toga zaključujemo da u časopisu Vojnosanitetski pregled (engl. *Military-medical and pharmaceutical review*) je objavljeno 61 rad, i predstavlja drugi časopis po broju obavljenih radova.

Učestalost objavljivanja u časopisima po katedrama, je izračunata kao odnos ukupnog broja radova i broja jedinstvenih časopisa u kojima su objavljivani radovi sa te katedre. U tabeli 4.5.2. vidimo da u proseku katedre objavljuju 2 rada po časopisu, a ta vrednost je nešto manja za katedru za imunologiju.

Tabela 4.5.2. Učestalost objavljivanja po katedrama

|  |  |  |  |
| --- | --- | --- | --- |
| Katedra | Ukupan broj radova | Broj časopisa | Učestalost |
| Katedra za Mikrobiologiju | 513 | 233 | 2.15 |
| Katedra za Epidemiologiju | 271 | 178 | 2.20 |
| Katedra za Imunologiju | 383 | 164 | 1.65 |
| Infektivne bolesti | 151 | 66 | 2.28 |

4.6 . Broj autora po radu

Autori na svojim radovima mogu da sarađuju sa jednim ili više koautora. Koji su nekad pripadnici iste institucije, a mogu i da budu eksterne kolege. Na ovaj način se povezuje naučna zajednica i zajedničkim snagama dolaze do najboljih mogućih zaključaka. Posmatrali smo koliko autori sa Medicinskog fakulteta sarađuju međusobno, a u kojoj meri sa autorima van ove institucije. Treba naglasiti da nije vršena analiza niza koji predstavlja koautore u datim podacima. Uzeta je pretpostavka da se u njemu nalaze samo autori koji nama nisu od značajna, tj. koji nisu bili zaposleni na fakultetu Medicinskih nauka Univerziteta u Beogradu, u Oktobru 2022. godine. U tabeli 4.6.1. možemo videti sa koliko autora sa fakulteta, i van njega u proseku sarađuju autori u odnosu na katedre.

Tabela 4.6.1. Broj koautora na radovima po katedrama

|  |  |  |
| --- | --- | --- |
| Katedra | Broj autora sa fakulteta | Broj autora van fakulteta |
| Katedra za Mikrobiologiju | 1.85 | 56.24 |
| Katedra za Epidemiologiju | 1.37 | 24.82 |
| Katedra za Imunologiju | 1.35 | 21.39 |
| Infektivne bolesti | 2.0 | 8.26 |

1. Osnovne karakteristike mreže

U ovom poglavlju biće analizirane osnove karakteristike mreže autora. Konstruisan je neusmereni težinski graf, gde čvor predstavlja autora, veza između njih znači da su zajedno radili na pisanju rada, dok je težina te veze broj napisanih radova. Pored identifikacionih podataka autora (Ime i Prezime) čvor čuva i informaciju kojoj katedri autor pripada.

* 1. Gustina mreže

Gustina mreže se računa kao količnik ukupnog broja grana i maksimalnog mogućeg broja grana u mreži. Koristi se kao gruba procena mreže, realne mreže su retke, tj. imaju malu gustinu. Mreža autora konstruisana od ulaznih podataka ima gustinu 0.175 što se slaže sa činjenicom da je upitanju realna mreža naučnih kolaboracija.

* 1. Prosečna distanca i dijametar

Prosečna distanca u mreži je prosečna dužina svih najkraćih puteva između čvorova u mreži , dok dijametar predstavlja najdužu putanju, od svih najkraćih putanja između dva čvora u mreži. Koncept distanci govori o načinu širenja i dosegu informacija, što su ove vrednosti manje to je mreža povezanija i više autora je međusobno direktno sarađivalo na radu. Prosečna distanca kreirane mreže je 2.36 dok je dijametar 5. Što nam govori da u proseku je potrebno pogledati saradnike, saradnika naših saradnika da bi došli do svake osobe na fakultetu.

* 1. Povezanost i centralizovanost mreže

Posmatrana mreža sadrži jednu džinovsku povezanu komponentu, to znači da ne postoji autor na Medicinskom fakultetu, koji nije sarađivao sa barem jednim kolegom. Takođe znači da ne postoji podela takva da neka grupa autora sarađuje samo jedni sa drugima, a ne i sa ostatkom fakulteta. Centralnost mreže autora je 0.1942 računata je na osnovu formule (5.3.1.),

(5.3.1.)

Gde *G(V,E)* predstavlja mrežu autora, *X(Y,Z)* mreža sa istim brojem čvorova koja ima topologiju zvezde. A čvor označen sa \* je onaj sa najvećim stepenom u mreži. Veći procenat centralizacije mreže ukazuje na postojanje čvorova koji su u boljoj poziciji u odnosu na ostale, što ovde nije slučaj.

* 1. Koeficijent klasterizacije

Koeficijent klasterizacije opisuje tendenciju čvora ka tome da stvori kliku. Klika je podgraf koji se sastoji od najvećeg mogućeg skupa čvorova koji su međusobno povezani svaki sa svakim. U tabeli 5.4.1. možemo videti vrednost lokalnog i globalnog koeficijenta klasterizacije posmatrane mreže, kao i *Erdos-Renzi* i *scale-free* mreže istih dimenzija.

Tabela 5.4.1. Koeficijenti klasterizacije

|  |  |  |  |
| --- | --- | --- | --- |
|  | Graf autora | *Erdos-Renzi* | Scale-free |
| Globalni koeficijent | 0.54 | 0.17 | 0.19 |
| Lokalni koeficijent | 0.60 | 0.17 | 0.27 |

Primećujemo da su koeficijenti klasterizacije mreže autora veći u odnosu na koeficijente kod posmatranih mrežnih modela. Globalni koeficijent nam pokazuje da mreža autora ima tendencije da formira grupe, gde koautori posmatranog čvora takođe sarađuju međusobno, dok lokalni koeficijent nam pokazuje da autori u našoj mreži imaju tendenciju da sarađuju međusobno. Kod *Scalee-free* mreže koeficijent klasterizacije je viši u odnosu na *Erdos-Renzi* zbog postojanja habova. Na slici 5.4.1. možemo videti raspored koeficijenta klasterizacije mreže autora.

A graph with blue bars

Description automatically generatedSlika 5.4.1. Raspodela lokalnog koeficijenta klasterizacije

Sa slike 5.4.1. možemo videti da većina autora ima umereni koeficijent klasterizacije 0.4-0.6, dok postoji i veliki broj čvorova koji ima koeficijent klasterizacije oko 0.9 što znači da oni formiraju jako povezane grupe.

* 1. Tendencija povezivanja sa istim koautorima

Na osnovu koeficijenta klasterizacije smo videli da autori imaju tendenciju da stvaraju grupe, i da sarađuju sa kolegama na svojim istraživanjima. Grana u grafu nam znači da postoji barem jedan rad na kome su dva povezana autora sarađivala, a šta ako ih ima više? Težina grane nam predstavlja broj radova na kojima su dva autora sarađivala, i na osnovu nje možemo videti da li oni imaju tendenciju da sarađuju sa istim koautorima, to je prikazano na slici 5.5.1.

A graph of a number of blue bars

Description automatically generatedSlika 5.5.1. Distribucija težine grane

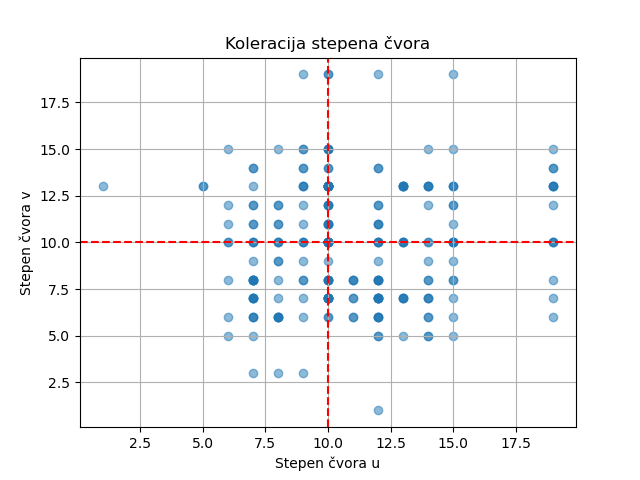
Primećujemo da distribucija težine grane ima izgled opadajući eksponencijalne funkcije, gde većina autora sarađuje sa različitim kolegama na jednom radu. Ali vidimo i da postoje autori koji su sarađivali međusobno na 20 radova.

* 1. Mali svet

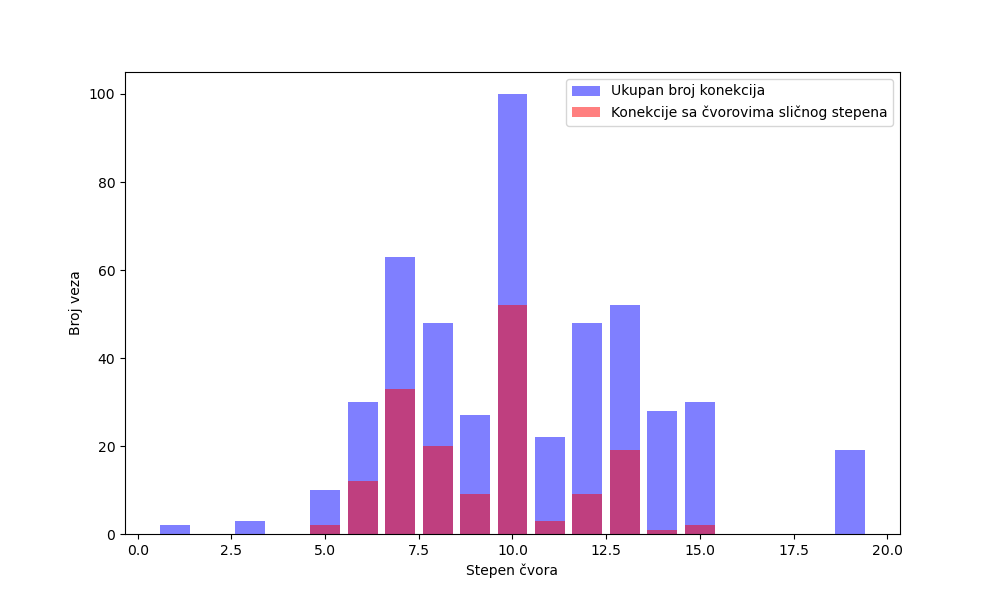
Fenomen malog sveta je definisan kao koncept da je svaka osoba indirektno povezana sa drugom preko malog broja indirektnih veza u velikim socijalnim mrežama. Posmatrana mreža autora ne predstavlja veliku socijalnu mrežu, ali ispoljava osobine mreže malog sveta:

* Mala prosečna udaljenost
* Visok stepen klasterizacije
  1. Asortivnost

Asortivnost pokazuje da li mreža ispoljava osobine homofilije ili heterofilije. Izračunava se u kontekstu stepena čvora, ako čvorovi istog stepena imaju tendenciju da se povezuju sa čvorovima sličnog stepena, mreža ispoljava osobine homofilije, u suprotnom heterofilije. Koeficijent asortivnosti uzima vrednost u skupu [-1, 1], gde vrednost bliža 1 predstavlja da mreža ima osobine homofilije, dok vrednost bliža -1 heterofilije. Stepen asortivnosti mreže autora je 0.046, koleraciju između stepena povezanih čvorova vidimo na slici 5.7.1.

Slika 5.7.1. Koleracija stepena čvora

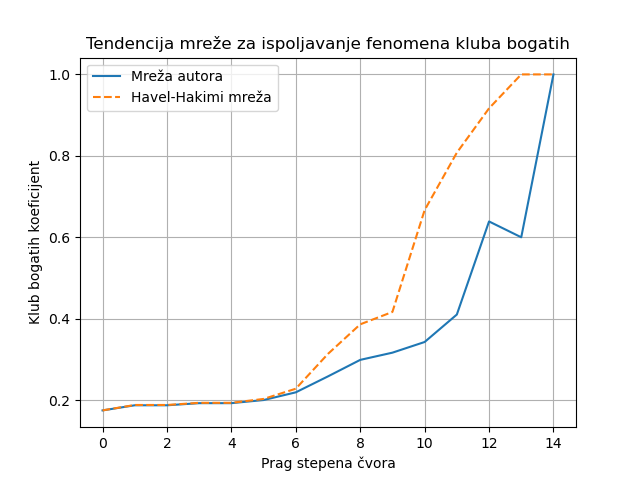
Sa slike 5.7.1. vidimo relativno uniformnu raspodelu tačaka korelacije u sva četiri kvadranta, ali i sam koeficijent asortivnosti je blizak nuli, pa je homofilija jako nisko izražena. Njeno pojavljivanje možemo bolje videti na slici 5.7.2. gde je prikazana tendencija čvora da stupi u veze sa drugim autorom sličnog stepena.



Slika 5.7.2. Raspodela broja veza čvora određenog stepena

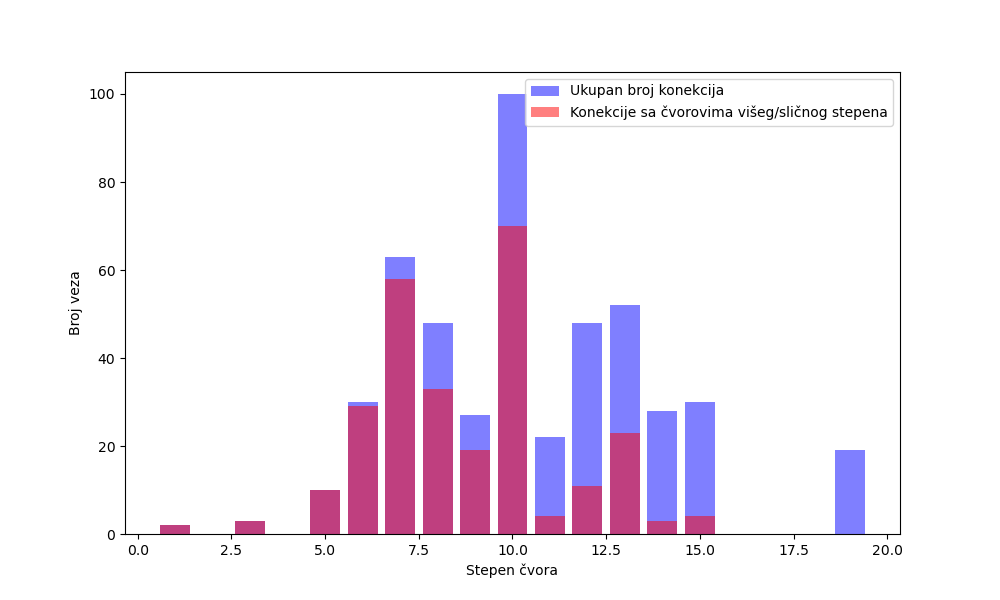
* 1. Fenomen kluba bogatih

Koeficijent kluba bogatih je mera koja kazuje koliko dobro povezani čvorovi imaju tendenciju da se međusobno povezuju. Na slici 5.8.1. je prikazana raspodela koeficijenta kluba bogatih mreže autora, u odnosu na Havel-Hakim mrežu istih dimenzija, koja ima izraženi koeficijent kluba bogatih, zato što se sama mreža pravi tako što se spajaju čvorovi najvišeg stepena međusobno, pa se ostali čvorovi sortiraju na osnovu stepena i ponavlja se proces sve dok se ne dobije povezan graf.



Slika 5.8.1. Tendencija mreže za ispoljavanje fenomena kluba bogatih

Sa slike 5.8.1. vidimo da mreža autora ima manje izraženu tendenciju za ispoljavanje fenomena kluba bogatih u odnosu na Havel-Hakim mrežu. Ali je funkcija svakako rastuća pa postoje veze između čvorova visokog stepena, na slici 5.8.2. je prikazano u zavisnosti od stepena čvora odnos ukupnog broja konekcija i konekcija sa čvorovima istog ili višeg stepena od posmatranog.



Slika 5.8.2. Odnos ukupnog broja veza čvora određenog stepena, i veza sa čvorovima istog ili višeg stepena

Sa slike 5.8.2. vidimo da čvorovi višeg stepena se konektuju međusobno ali fenomen kluba bogatih nije izražen u ovoj mreži. Možemo primetiti da postoje habovi, zato što se čvorovi nižeg stepena pretežno vezuju za čvorove višeg stepena.

* 1. Distribucija stepena čvorova

Kao što se može videti i na 5.8.2. čvorovi koji predstavljaju autore imaju različite vrednosti stepena, a na 5.9.1. možemo pogledati kako su te vrednosti distribuirane u našem grafu.

A graph with blue dots

Description automatically generated

Slika 5.9.1. Distribucija stepena čvora

Sa slike 5.9.1. ne možemo proceniti tip funkcije koja bi mogla da se formira kroz date tačke i da ih najbliže opiše. Iskoristićemo paket u programskom jeziku Python, koji je korišćen kroz ovaj zadatak, paket je opisan u [3]. I uporediti powerlow funkciju sa drugim funkcijama za opisivanje sistema. Na osnovu rezultata odgovarajuće funkcije za upoređivanje dobili smo da powerlow približnije opisuje podatke od eksponencijalne funkcije, kao i od *trunced-power-low*, ali pošto parametri powerlow funkcije imaju visoke vrednosti. Eksponent od 12.058 sa velikim standardnim odstupanjem od 3.69 što se ne nalazi u očekivanom opsegu. Powerlow raspodela se iskazuje tek od čvorova sa stepenom 13 što predstavlja jako mali broj podataka. Čvorovi ne prate powerlow raspodelu iako za nju dobijamo da je bolja od ostalih, već prestavljaju neku na izgled nasumičnu distribuciju što odgovara korišćenju realnog seta podataka.

1. Centralnost mreže

Mere centralnosti otkrivaju koji čvorovi su bitni za određivanje povezanosti mreže. Daju uvid o položaju pojedinačnih aktera u mreži, računa se za pojedinačne čvorove ali i za celu mrežu.

* 1. Mere centralnosti

Posmatraćemo 4 različite mere centralnosti, i videti koji autori dolaze do isticaja putem ovih vrednosti. Centralnost po stepenu (DC), autori koji se ističu po ovoj vrednosti se smatraju komunikativnim, i sarađivali su sa velikim brojem različitih kolega na svojim radovima, predstavlja broj direktnih suseda čvora u mreži. Centralnost po bliskosti (CC), prikazuje čvorove koji nisu centralni, ali su dosta povezani i bitni, influenseri. Relaciona centralnost (BC), ukazuje na aktera koji je bitan u mreži za povezivanja. Dok centralnost po sopstvenom vektoru (EVC) koristi koncept uticaja i moći. Čvor je uticajniji ukoliko i njegovi susedi imaju veliki broj suseda. Čvor je moćniji ukoliko njegovi susedi nemaju veliki broj svojih suseda. U tabeli 6.1.1. prikazano je 5 autora koji se najviše ističu po centralnosti po stepenu.

Tabela 6.1.1. Prikaz autora opadajući po stepenoj centralnosti

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Autor | Katedra | DC | CC | BC | EVC |
| Ivana Lazarević | Katedra za mikrobiologiju | 0.36 | 0.55 | 0.17 | 0.30 |
| Goran Stevanović | Infektivne bolesti | 0.29 | 0.51 | 0.07 | 0.27 |
| Maja Cupić | Katedra za mikrobiologiju | 0.29 | 0.51 | 0.07 | 0.18 |
| Vera Mijac | Katedra za mikrobiologiju | 0.27 | 0.50 | 0.07 | 0.12 |
| Milos Korać | Infektivne bolesti | 0.27 | 0.49 | 0.05 | 0.27 |

Primećujemo da među komunikativnim autorima, onima koji imaju veliku centralnost po stepenu nalaze se autori sa dve katedre, što su dve katedre koje su imale najveći prosečni broj autora sa fakulteta u svojim radovima.

U tabeli 6.1.2. prikazano je koji autori se ističu po centralnosti po bliskosti.

Tabela 6.2.1. Prikaz autora opadajući po centralnosti po bliskosti

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Autor | Katedra | DC | CC | BC | EVC |
| Ivana Lazarević | Katedra za mikrobiologiju | 0.36 | 0.55 | 0.17 | 0.30 |
| Aleksandra Knežević | Katedra za mikrobiologiju | 0.19 | 0.51 | 0.03 | 0.14 |
| Maja Cupić | Katedra za mikrobiologiju | 0.29 | 0.51 | 0.07 | 0.18 |
| Goran Stevanović | Infektivne bolesti | 0.29 | 0.51 | 0.06 | 0.27 |
| Vera Pravica | Katedra za imunologiju | 0.23 | 0.51 | 0.05 | 0.12 |

Centralnost po bliskosti prikazuje čvorove koji predstavljaju centre okupljanja, možemo primetiti iz tabela 6.1.1. i 6.1.2. da Ivana Lazarević predstavlja i centar okupljanja na nivou fakulteta i sarađuje sa dosta različitih kolega. Možemo pogledati i koji autori predstavljaju centre okupljanja na svakoj različitoj katedri u tabeli 6.1.3.

Tabela 6.1.3. Prikaz najboljih predstavnika svake od katedri po centralnosti po bliskosti

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Autor | Katedra | DC | CC | BC | EVC |
| Ivana Lazarević | Katedra za mikrobiologiju | 0.36 | 0.55 | 0.17 | 0.30 |
| Goran Stevanović | Infektivne bolesti | 0.29 | 0.51 | 0.06 | 0.27 |
| Vera Pravica | Katedra za imunologiju | 0.23 | 0.51 | 0.05 | 0.12 |
| Sandra Spasić-Grujičić | Katedra za epidemiologiju | 0.15 | 0.44 | 0.03 | 0.07 |

U tabeli 6.1.4. prikazano je 5 najuticajnijih autora po relacionoj centralnosti, to su autori koji povezuju različite delove mreže.

Tabela 6.1.4. Prikaz autora opadajući po relacionoj centralnosti.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Autor | Katedra | DC | CC | BC | EVC |
| Ivana Lazarević | Katedra za mikrobiologiju | 0.36 | 0.55 | 0.17 | 0.30 |
| Tatjana Pekmezović | Katedra za epidemiologiju | 0.23 | 0.44 | 0.10 | 0.77 |
| Vera Mijac | Katedra za mikrobiologiju | 0.27 | 0.50 | 0.07 | 0.12 |
| Ljiljana Marković-Denić | Katedra za epidemiologiju | 0.19 | 0.43 | 0.07 | 0.16 |
| Goran Stevanović | Infektivne bolesti | 0.29 | 0.51 | 0.07 | 0.27 |

Iz tabele 6.1.4. primećujemo da Ivana Lazarević i povezuje različite delove mreže. A ako pogledamo uticajne autore po centralnosti po sopstvenom vektoru, prikazane u tabeli 6.1.5.

Tabela 6.1.4. Prikaz autora opadajući po centralnosti po sopstvenom vektoru

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Autor | Katedra | DC | CC | BC | EVC |
| Ivana Lazarević | Katedra za mikrobiologiju | 0.36 | 0.55 | 0.17 | 0.30 |
| Goran Stevanović | Infektivne bolesti | 0.29 | 0.51 | 0.07 | 0.27 |
| Miloš Korać | Infektivne bolesti | 0.27 | 0.49 | 0.05 | 0.27 |
| Nataša Nikolić | Infektivne bolesti | 0.25 | 0.46 | 0.01 | 0.26 |
| Nikola Mitrović | Infektivne bolesti | 0.25 | 0.46 | 0.01 | 0.26 |

Iz tabele 6.1.4. primećujemo da katedra za Infektivne bolesti dolazi do izražaja u slučaju centralnosti po sopstvenom vektoru. Centralnost po sopstvenom vektoru ističe autore koji su ili uticajniji, ukoliko i njegovi susedi imaju veliki broj suseda, ili moćniji ukoliko njegovi susedi nemaju veliki broj suseda. Ivana i Goran se nalaze u vrhu i po centralnosti po bliskosti, tako da oni predstavljaju uticajnije autore, dok ostale tri kolege prestavljaju moćnije autore.

* 1. Kompozitna centralnost

Cilj istraživanja je posmatranje naučne produkcije iz oblasti epidemiologije i zaraznih bolesti na Medicinskom fakultetu Univerziteta u Beograd. Od značaja nam je da autori sarađuju međusobno i dele znanja iz svojih referentnih oblasti. Hoćemo da naglasimo moćnije autore, koji su sarađivali sa dosta kolega, i prestavljaju spone između različitih katedri, zato je stavljen akcenat na relacionu centralnost i centralnost po sopstvenom vektoru. Takođe su naglašeni autori sa visokom stepenom centralnosti zbog sarađivanja sa dosta svojih kolega. Za heuristiku kompozitne centralnosti uzeta he formula (6.1)

(6.1)

Na osnovu ove karakteristike izdvajaju se autori prikazani u tabeli 6.2.1. na nivou celog fakulteta.

Tabela 6.1.4. Prikaz autora opadajući po centralnosti po sopstvenom vektoru

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Autor | Katedra | DC | CC | BC | EVC | Comp |
| Ivana Lazarević | Katedra za mikrobiologiju | 0.36 | 0.55 | 0.17 | 0.30 | 0.30 |
| Goran Stevanović | Infektivne bolesti | 0.29 | 0.51 | 0.07 | 0.27 | 0.24 |
| Miloš Korać | Infektivne bolesti | 0.27 | 0.49 | 0.05 | 0.27 | 0.22 |
| Ksenija Bojović | Infektivne bolesti | 0.25 | 0.47 | 0.06 | 0.24 | 0.21 |
| Maja Cupić | Katedra za mikrobiologiju | 0.25 | 0.46 | 0.01 | 0.26 | 0.21 |

* 1. Saradnja katedri

A red and blue squares with numbers

Description automatically generated

Slika 5.3.1.Grafički prikaz saradnje između različitih katedri

1. Резултати

У овом поглављу биће представљени кључни резултати семинарског. Анализираћемо перформансе различитих модела класификације на описаном скупу података у поглављу 3, као и перформансе неуронских мрежа. Приказати њихове резултате коришћењем одговарајућих метрика.

* 1. Метрика

За евалуацију резултата класификације коришћене су различите метрике:

* Тачност
* Прецизност
* Одзив
* Ф1-мера
* Матрица конфузије
* *Multi-class Log Loss*

Свака метрика пружа јединствени увид у перформансе класификационих модела, омогућавајући дубље разумевање њихових способности и недостатака.

* + 1. Тачност

Тачност даје информацију о генералној успешности предикције модел, кроз све класе. Рачуна се по следећој формули:

(6.1.2.)

где представља број тачних предикција, а , укупан број предикција.

* + 1. Прецизност

Прецизност је метрика која процењује тачност позитивних предвиђања модела. Даје информацију о томе колико је модел склон тачном класификовању позитивних примера. Рачуна се по следећој формули:

(7.1.2.)

где *Tp*  представља број правих позитивних вредности, тј. узорака које је модел класификације означио као позитивне, и они то заиста и јесу, а *Fp* представља број лажно позитивних, модел је негативне вредности погрешно класификовао.

* + 1. Одзив

Одзив је метрика која процењује способност модела да тачно идентификује све стварно позитивне примере. Пружа информацију о томе колико је модел ефикасан у откривању позитивних примера. Рачуна се по следећој формули:

(7.1.3)

где *Tp*  представља број правих позитивних вредности и *Fn* представља број лажно негативних вредности.

* + 1. Ф1-мера

Ф1-мера комбинује прецизност и одзив модела у једну вредност, и представља њихову хармоничну средину. Користи се када је потребно узети у обзир равнотежу између тачности позитивних предвиђања и способности модела да идентификује све позитивне примере. Рачуна се по формули 6.1.4. где *P* представља вредностпрецизности, а *O* одзива.

(7.1.4.)

* + 1. Матрица конфузије

Табела која сумира број тачно позитивних, тачно негативних, лажно позитивних и лажно негативних предикција за сваку класу. Даје детаљан пресек перформанси модела за сваку класу.

* + 1. Multi-class Log Loss

Процењује перформансе класификације модела, чији излаз представља вероватноћу између 0 и 1. У поставци проблема за такмичење, захтевана је евалуација решења коришћењем ове метрике. Добијене по следећој формули:

(7.1.6.)

где је *N* број посматрања у тест подскупу, *M* број класа, у овом случају 3,  *y*  да ли тренутни улаз одговара датој класи или не, узима вредност 0 или 1, и  *p*  представља вероватноћу да дати улаз одговара посматраној класи. Коришћена је за процену перформанси само неуронских мрежа, зато што оне директно враћају вероватноћу. Могуће је прилагодити и алгоритме машинског учења тако да враћају вероватноћу припадања одређеној класи, али ова вредност није значајна у процесу класификације датих алгоритама.

* 1. Резултати добијени алгоритмима машинског учења

Резултате класификације алгоритама машинског учења можемо видети у табели 7.2.1.

Најбоље се показао Мултиномијални наивни Бајесов класификатор, по свим параметрима, одмах после њега је Логистичка регресија. Док стабло одлучивања има доста слабе перформансе.

Табела 7.2.1. Резултати класификације алгоритама машинског учења

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Алгоритми | Тачност [%] | Прецизност [%] | Одзив [%] | Ф1-мера [%] |
| КНН | 68.9 | 74.45 | 68.897 | 67.588 |
| Логистичка регресија | 80.4 | 80.568 | 80.388 | 80.352 |
| Стабло одлучивања | 58.5 | 58.412 | 58.495 | 58.421 |
| МНБ | 81.3 | 82.203 | 81.324 | 81.246 |

У матрицама конфузије датим у наставку, лабеле су кодиране на следећи начин:

* 0 = Едгар Алан По
* 1 = Хаурд Филипс Лафкрафт
* 2 = Мери Шели

На слици 7.2.1. налази се матрица конфузије КНН алгоритма, можемо приметити да он има велике вредности дуж дијагонале, што се поклапа са перформансама алгоритма. Такође видимо да лажно детектује 0 класу, што значи да преписује По-у дела која нису његова, до овога долази због благог дизбаланса класа, где постоји више текстова овог аутора у односу на друга два.

Slika na kojoj se nalazi tekst, snimak ekrana, dijagram, linija

Opis je automatski generisan

Слика 7.2.1. Матрица конфузије КНН алгоритма

На слици 7.2.3. налази се матрица конфузије логистичке регресије, на њој примећујемо доста мање вредности лажних позитива у односу на КНН алгоритам.

Slika na kojoj se nalazi tekst, snimak ekrana, dijagram, kvadrat

Opis je automatski generisan

Слика 7.2.2. Матрица конфузије Логистичке регресије

Матрица конфузије стабла одлучивања приказана је на слици 7.2.3., на њој видимо знатно више вредности лажних позитива, као и ниже вредности на дијагонали. Овакви резултати одговарају перформансама алгоритма из табеле, који се показао као најлошији избор за решавање датог проблема.

Slika na kojoj se nalazi tekst, snimak ekrana, dijagram, kvadrat

Opis je automatski generisan

Слика 7.2.3. Матрица конфузије стабла одлучивања

На слици 7.2.4. приказана је матрица конфузије мултиномијалног наивни Бајесов класификатор, који се показао најбоље у класификације. Али има јако сличне перформансе са логистичком регресијом, што се види и са упоређивањем њихових матрица конфузије.

Slika na kojoj se nalazi tekst, snimak ekrana, dijagram, kvadrat

Opis je automatski generisan

Слика 7.2.4. Матрица конфузије МултиномијалногНБ

* 1. Резултати добијени неуронским мрежама

Резултате класификације неуронским мрежама можемо видети у табели 7.3.1.

Најбоље се показао Мултиномијални наивни Бајесов класификатор, по свим параметрима, одмах после њега је Логистичка регресија. Док стабло одлучивања има доста слабе перформансе.

Табела 7.3.1. Резултати класификације неуронским мрежама

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Алгоритми | Тачност [%] | Прецизност [%] | Одзив [%] | Ф1-мера [%] |
| *LSTM* | 74.41 | 74.46 | 74.41 | 74.38 |
| *BiLSTM* | 74.95 | 74.95 | 74.95 | 74.94 |
| *CNN* | 69.87 | 70.11 | 69.87 | 69.76 |

На слици 7.3.1. налази се матрица конфузије *LSTM* модела, примећујемо високе вредности дуж дијагонале, али не занемарив број лажних позитива.

Slika na kojoj se nalazi tekst, snimak ekrana, kvadrat, dijagram

Opis je automatski generisan

Слика 7.3.1. Матрица конфузије *LSTM*

На слици 7.3.2. приказани су графици *Log-Loss* функције, на тренинг и валидационом сету, можемо приметити да обе функције опадају што представља индикацију да не постоји *overfitting.*

Slika na kojoj se nalazi tekst, linija, snimak ekrana, Plot

Opis je automatski generisan

Слика 7.3.2. *Log-Loss* фунција *LSTM*

На слици 7.3.3. налази се матрица конфузије *BiLSTM* модела, који има сличне вредности као и *LSTM* модел, тако да бидирекционални приступ не доводи до успешније класификације.

Slika na kojoj se nalazi tekst, snimak ekrana, kvadrat, dijagram

Opis je automatski generisan

Слика 7.3.3. Матрица конфузије *BiLSM*

На слици 7.3.4. видимо *Log-Loss* функције за овај модел, које су опадајуће. Такође примећујемо више вредности валидационе Log-Loss функције што указује на веће грешке и већи проблем *overfitting-a* у односу на *LSM* модел.

Slika na kojoj se nalazi tekst, linija, Plot, snimak ekrana

Opis je automatski generisan

Слика 7.3.4. *Log-Loss* функција *BiLSTM*

Матрица конфузије *CNN* мреже приказана је на слици 7.3.5., на њој видимо ниже вредности на дијагонали за не доминанту класу у односу на претходна два алгоритма.

Slika na kojoj se nalazi tekst, snimak ekrana, kvadrat, dijagram

Opis je automatski generisan

Слика 7.3.5. Матрица конфузије *CNN* мреже

Слика 7.3.3. Матрица конфузије *BiLSM*

На слици 7.3.6. видимо *Log-Loss* функције за овај модел, које су опадајуће. Такође примећујемо више вредности валидационе Log-Loss функције што указује на веће грешке и већи проблем *overfitting-a,* али чињеница да ова функција има вредност мању од 0.5 указује на успешност алгоритма.

Slika na kojoj se nalazi tekst, Plot, dijagram, linija

Opis je automatski generisan

Слика 7.3.6. *LogLoss* фунција *CNN*

1. Закључак

На основу резултата приказаних у претходном поглављу, примећујемо да најбоље перформансе има мултиномијалног наивни Бајесов класификатор код кога све метрике су у распону 81.2% - 82.2%, након њега следи Логистичка регресија чије метрике достижу 80%. Након ова два алгоритма машинског учења, следе неуралне мреже, *LSTM I BiLSTM* дају јако сличне резултате, мало испод 75% али када погледамо график log-loss функције код *BiLSTM* модела примећујемо већи *overfitting.* Након њих следе *CNN* модел и КНН алгоритам, неурална мрежа има нижу прецизност, док се по осталим параметрима показала за нијансу боља од КНН алгоритма. Стабло одлучивања има најлошије резултате класификације.

Literatura

1. ,,Analiza socijalnih mreža“, Projektni zadatak
2. ,,H-indeks“, https://nauka.cpn.rs/sadrzaj/vrednovanje-bibliografskih-podataka/

(20.08.2024.)

[3] ,,powerlow: A Python Package for Analysis of Heavy-Tailed Distributions”, Jeff Alsstott, Bullmore E. and Plenz D., PLOS JOURNALS, Januar 2014

Spisak slika

[Slika 1.1. SARS-CoV-2 1](#_Toc175223328)

[Slika 3.1.1. Raspodela broja naučnika po katedrama 5](#_Toc175223329)

[Slika 3.1.2. Raspodela broja radova po katedrama 6](#_Toc175223330)

[Slika 4.2.1. Raspodela broja prosečnog broja koautora po autoru 9](#_Toc175223331)

[Slika 5.4.1. Raspodela lokalnog koeficijenta klasterizacije 16](#_Toc175223332)

[Slika 5.5.1. Distribucija težine grane 17](#_Toc175223333)

[Slika 5.7.1. Koleracija stepena čvora 18](#_Toc175223334)

[Slika 5.7.2. Raspodela broja veza čvora određenog stepena 19](#_Toc175223335)

[Slika 5.8.1. Tendencija mreže za ispoljavanje fenomena kluba bogatih 20](#_Toc175223336)

[Slika 5.8.2. Odnos ukupnog broja veza čvora određenog stepena, i veza sa čvorovima istog ili višeg stepena 21](#_Toc175223337)

[Slika 5.9.1. Distribucija stepena čvora 22](#_Toc175223338)

[Слика 6.1.1. Учитавање и примена *GloVe* фајла **Error! Bookmark not defined.**](#_Toc175223339)

[Слика 6.5.1. Додавање *Embedding layer* 27](#_Toc175223340)

[Слика 7.2.1. Матрица конфузије КНН алгоритма 31](#_Toc175223341)

[Слика 7.2.2. Матрица конфузије Логистичке регресије 31](#_Toc175223342)

[Слика 7.2.3. Матрица конфузије стабла одлучивања 32](#_Toc175223343)

[Слика 7.2.4. Матрица конфузије МултиномијалногНБ 33](#_Toc175223344)

[Слика 7.3.1. Матрица конфузије *LSTM* 34](#_Toc175223345)

[Слика 7.3.2. *Log-Loss* фунција *LSTM* 34](#_Toc175223346)

[Слика 7.3.3. Матрица конфузије *BiLSM* 35](#_Toc175223347)

[Слика 7.3.4. *Log-Loss* функција *BiLSTM* 36](#_Toc175223348)

[Слика 7.3.5. Матрица конфузије *CNN* мреже 36](#_Toc175223349)

[Слика 7.3.3. Матрица конфузије *BiLSM* 37](#_Toc175223350)

[Слика 7.3.6. *LogLoss* фунција *CNN* 37](#_Toc175223351)

Spisak tabela

[Tabela 3.1.1. Raspodela broja radova po katedrama 6](#_Toc175223359)

[Tabela 4.1.1. Prikaz autora sa najviše radova celovito brojanje 7](#_Toc175223360)

[Tabela 4.1.2. Prikaz autora sa najviše radova frakcijalno brojanje 8](#_Toc175223361)

[Tabela 4.2.2. Prikaz autora sa najviše koautora u proseku na svojim radovima 8](#_Toc175223362)

[Tabela 4.3.1. Prikaz autora za koje postoji razlika između izračunatog i dobijenog H-indeksa 9](#_Toc175223363)

[Tabela 4.4.1. Produktivnost katedri 10](#_Toc175223364)

[Tabela 4.4.2. Najproduktivnije godine po katedri 11](#_Toc175223365)

[Tabela 4.5.1. Časopisi sa najviše objavljenih radova 11](#_Toc175223366)

[Tabela 4.5.2. Učestalost objavljivanja po katedrama 12](#_Toc175223367)

[Tabela 4.6.1. Broj koautora na radovima po katedrama 13](#_Toc175223368)

[Tabela 5.4.1. Koeficijenti klasterizacije 15](#_Toc175223369)

[Табела 7.2.1. Резултати класификације алгоритама машинског учења 30](#_Toc175223370)

[Табела 7.3.1. Резултати класификације неуронским мрежама 33](#_Toc175223371)