

NAUKA O PODACIMA U OKVIRU KURSEVA MASTER AKADEMSKIH STUDIJA NA ELEKTROTEHNIČKOM FAKULTETU UNIVERZITETA U BEOGRADU

Marko Mišić¹, Dražen Drašković², Boško Nikolić³, Jelica Protić⁴

¹²³⁴Univerzitet u Beogradu, Elektrotehnički fakultet, Beograd, Republika Srbija

¹marko.misic@etf.bg.ac.rs, ²drazen.draskovic@etf.bg.ac.rs, ³nbosko@etf.bg.ac.rs, ⁴jelica.protic@etf.bg.ac.rs

Kratak sadržaj: Nauka o podacima već duže vreme privlači pažnju zaposlenih u industriji i akademskim krugovima, ali i studenata računarstva i srodnih oblasti. S obzirom na veoma veliku potražnju za kadrovima koji poznaju ovu oblast, studenti pokušavaju da steknu što više znanja kroz kurseve na osnovnim i master akademskim studijama. U okviru ovog rada su predstavljene mogućnosti za usavršavanje studenata iz ove oblasti kroz predmete na postojećim, akreditovanim master akademskim studijama Elektrotehnike i računarstva na Elektrotehničkom fakultetu Univerziteta u Beogradu.

Ključne reči: nauka o podacima, mašinsko učenje, master studije

DATASCIENCE IN THE GRADUATE ACADEMIC STUDIES COURSES AT THE SCHOOL OF ELECTRICAL ENGINEERING, UNIVERSITY OF BELGRADE

Abstract: Data science has attracted the attention of employees in industry and academia for a long time, as well as students of computer science and related fields. Given the very high demand for staff who are competent in the field, students are trying to gain as much knowledge as possible through courses in undergraduate and master academic studies. In this paper, opportunities for training students in this field through courses in the existing accredited master academic studies of Electrical Engineering and Computing at the School of Electrical Engineering, University of Belgrade are presented.

Key Words: datascience, machine learning, graduate studies

1. UVOD

Nauka o podacima je interdisciplinarna oblast koja već duži niz godina predstavlja jednu od ključnih oblasti računarstva [1]. Obuhvata obradu velikih količina strukturiranih i nestrukturiranih podataka kroz primenu algoritama mašinskog učenja i istraživanja podataka, upotrebu velikih baza podataka i distribuiranih računarskih sistema, ali i koncepte iz verovatnoće i statistike, računске fizike i biologije i drugih sličnih oblasti, pa čak i društvenih nauka poput psihologije i ekonomije [2]. Značaj ove oblasti je prepoznala i stručna organizacija ACM kroz izdavanje dokumenta koji predstavlja preporuke prilikom kreiranja studijskih programa iz oblasti nauke o podacima [3].

Prema *Glassdoor* listama 50 najboljih poslova za 2019., 2020. i 2021. godinu koje objavljuje i komentariše magazin *Forbes* [4], posao istraživača podataka (*data scientist*) se nalazi u prva tri najplaćenija posla za sve tri navedene godine sa veoma visokim iznosima godišnjih plata. Postoji i niz drugih predviđanja renomiranih istraživačkih i marketinških agencija da će potreba za kadrom koji poznaje tehnike nauke o podacima rasti u budućnosti. Stoga i ne čudi veliko interesovanje studenata za usmerenja koja školuju odgovarajuće profile ili koja uključuju veći broj predmeta iz oblasti koje čine jezgro nauke o podacima. U najvećem broju slučajeva univerziteti i fakulteti ne poseduju studijske programe direktno usmerene ka nauci o podacima, već se teme iz ove oblasti obrađuju kroz studijske programe iz oblasti računarstva, ali i automatike, matematike i sličnih oblasti. Međutim, među studentima, naročito postdiplomskih studija postoji potreba za dodatnim usmeravanjem i usavršavanjem iz pomenutih oblasti. U tom smislu, mnogi se opredeljuju za odgovarajuće profile na master akademskim studijama.

Privatne visokoškolske ustanove su se nešto brže uključile na obrazovno tržište ponudom studijskih programa koji obuhvataju predmete iz oblasti nauke o podacima, naročito na master akademskim studijama. Državni fakulteti su reagovali sporije, uvođenjem studijskih programa na osnovnim i master akademskim studijama koji obuhvataju elemente nauke o podacima u manjoj ili većoj meri. Verovatno najveći zamah je dobijen kroz program „Master 4.0“ [5] koji je iniciralo Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije, a podržala Inicijativa „Digitalna Srbija“, kao i mnoge druge organizacije i kompanije. Elektrotehnički fakultet Univerziteta u Beogradu je u partnerstvu sa Fakultetom organizacionih nauka implementirao master studijski program „Napredne informacione tehnologije u digitalnoj transformaciji“ sa četiri modula koji obuhvataju analizu podataka, distribuirano računarstvo, interaktivno računarstvo i organizaciju i upravljanje. Program je privukao pažnju zainteresovanih studenata, ali mahom diplomaca drugih visokoškolskih ustanova, a u maloj meri svršenih studenata Elektrotehničkog fakulteta. Jedan od razloga treba tražiti i u trajanju ovog programa od tri semestra, nasuprot dva semestra koliko traju postojeće, akreditovane master akademske studije na Elektrotehničkom fakultetu, kao i na Fakultetu organizacionih nauka. U

okviru ovog programa je prepoznat i problem što studenti imaju obavezne predmete sa svakog od navedenih modula, što odbija studente koji već imaju predznanja iz određenih oblasti, ali i one koji nemaju dovoljno predznanja. Takođe, postojeće, prethodno akreditovane master akademske studije Elektrotehnike i računarstva sa 12 modula pružaju prilično veliku izbornost, bez obaveznih predmeta, kao i ponudu predmeta sa različitih modula čijim izborom se student može specijalizovati za teme iz oblasti nauke o podacima koje ga interesuju.

U okviru ovog rada će biti analizirani moduli i predmeti sa master akademskih studija Elektrotehnike i računarstva na Elektrotehničkom fakultetu Univerziteta u Beogradu koji omogućavaju usavršavanje iz oblasti nauke o podacima. Identifikovana su tri modula: Softversko inženjerstvo, Računarska tehnika i informatika i Signali i sistemi, kao moduli koji najviše odgovaraju nauci o podacima prema predmetima koji se na njima nude. Takođe, identifikovan je veći broj predmeta na navedenim i drugim modulima čijim izborom se studenti mogu usavršiti na poljima istraživanja podataka, mašinskog učenja i statističke obrade podataka. Moduli i predmeti će biti analizirani sa stanovišta nauke o podacima. U radu će biti analizirani i izbori grupa predmeta koje su studenti učinili u periodu od 2019. do 2021. godine, kako bi se dobio uvid u to u kojoj meri su na sistematičan način pristupili izboru i usmeravanju ka nauci o podacima.

2. PREDMETI IZ OBLASTI NAUKE O PODACIMA NA MASTER AKADEMSKIM STUDIJAMA

Kompetencije koje treba da ima inženjer koji se bavi naukom o podacima se mogu prema [2] podeliti u nekoliko grupa. Najpre se ističu veštine iz oblasti mašinskog učenja i obrade teksta, sa naglaskom na nestrukturirane ili delimično struktuirane podatke. Poznavanje verovatnoće i statistike je izuzetno bitno, sa naglaskom na Bajesovu statistiku kroz poznavanje teorije verovatnoće, distribucija, testiranje hipoteza i multivarijantnu analizu. Podrazumevaju se bazična znanja iz oblasti algoritama i struktura podataka, baza podataka, distribuiranog i paralelnog programiranja i skript jezika poput programskog jezika *Python* za obradu skupova podataka umerene veličine. Međutim, naprednija znanja iz oblasti nerelacionih baza podataka i računarstva u oblaku su neophodna za obradu velikih skupova podataka. Takođe, poželjna su i znanja iz oblasti modelovanja problema i definisanja problema predikcije, kao i vizuelizacije, zaštite podataka, razvoja softvera i sl. Na kraju, s obzirom na čestu osetljivost podataka koji se obrađuju, moraju se uzeti u obzir i različiti profesionalni, etički i sociološki aspekti.

Master akademske studije Elektrotehnike i računarstva na Elektrotehničkom fakultetu Univerziteta u Beogradu traju dva semestra i organizovane su kroz 12 modula: Audio i video tehnologije, Biomedicinski i ekološki inženjering, Elektroenergetski sistemi, Elektronika i digitalni sistemi, Energetska efikasnost, Informaciono komunikacione tehnologije, Mikrotalasna tehnika, Nanoelektronika i fotonika, Primenjena matematika, Računarska tehnika i informatika, Signali i sistemi i Softversko inženjerstvo [6]. Zamišljene su kao logičan nastavak osnovnih akademskih studija, gde student nastavlja svoje obrazovanje na istom ili srodnom modulu. Svi predmeti su izborni, a student bira pet predmeta podeljenih u grupu A i grupu B. Iz grupe A se biraju tri predmeta, a čine je predmeti usko tematski u vezi sa izabranim modulom. Preostala dva predmeta se mogu izabrati sa liste B istog modula ili lista A i B bilo kog drugog modula. Jedan predmet se može zameniti sa dva predmeta iz grupe opšteobrazovnih predmeta.

Analizom navedenih modula i predmeta koji se na njima drže izdvajaju se moduli Softversko inženjerstvo (SI), Računarska tehnika i informatika (RTI) i Signali i sistemi (SIS) kao oni sa najviše zastupljenih predmeta iz oblasti nauke o podacima. U manjoj meri se izdvajaju predmeti sa modula Primenjena matematika (PM) i Mikrotalasna tehnika (MT). Spisak identifikovanih predmeta iz oblasti nauke o podacima je dat u Tabeli 1. Za svaki predmet je naveden modul kome pripada, dat je kratak opis predmeta, kao i način sprovođenja praktičnog rada. Može se primetiti da su predmeti sa modula Signali i sistemi više okrenuti konceptima koji uključuju mašinsko učenje, neuralne mreže i statistiku, a izvode se uglavnom kroz prototipsko programiranje u softverskom paketu *Matlab* ili besplatnom rešenju *Octave*. Na modulima Softversko inženjerstvo i Računarska tehnika i informatika postoji određeni balans u okviru predmeta između teorijskog i praktičnog rada. Većina predmeta predstavlja teorijske koncepte, pre svega iz oblasti mašinskog učenja sa primenama, teorije grafova, zaštite podataka i algoritama i struktura podataka, sa jakom praktičnom primenom. Na većini predmeta praktičan rad kroz laboratorijske vežbe i projektne zadatke nosi i od 60% ocene. Zastupljeno je uglavnom programiranje u programskim jezicima *Python* i *Java* uz druge softverske alate, što doprinosi sticanju znanja koja se mogu direktnije primeniti u produkcionim okruženjima. Takođe postoji i predmet iz oblasti naprednih sistema baza podataka, kao i predmet koji se bavi sociološkim, profesionalnim i etičkim aspektima obrade podataka.

Na Elektrotehničkom fakulteta je bilo pokušaja da se oformi modul za inteligentnu obradu podataka koji bi na amalgamski način obuhvatao različite oblasti elektrotehnike i računarstva i primenu inteligentne obrade podataka u njima. Međutim, u okviru kolektiva nije bilo dovoljno sluha da se oformi jedan takav interedisciplinarni program, kakvi postoje na stranim univerzitetima. Dobar primer je ETH Cirihi koji ima master program zasnovan na nauci o podacima koji predstavlja zajedničku saradnju departmana za matematiku, računarske nauke i elektrotehnike [7], a namenjen je diplomcima koji su prethodno završili računarske nauke, elektrotehniku, mašinstvo, matematiku ili fiziku. Stoga su studenti upućeni na izbor neke od kombinacija navedenih predmeta, u skladu sa pravilima za izbor predmeta na master akademskim studijama, kako bi se usavršavali iz oblasti nauke o podacima.

Tabela 2. Identifikovani predmeti iz oblasti nauke o podacima

Predmet	Modul	Kratak opis predmeta	Praktičan rad
Analiza socijalnih mreža	SI	Modelovanje i analiza socijalnih i kompleksnih mreža. Predstavljanje podataka kroz strukturu grafa. Kvantitativna i kvalitativna analiza podataka. Mere centralnosti. Detekcija komuna. Dinamičko ponašanje mreže. Mrežni modeli. Predikcije.	Python, Gephi
Genomska informatika	RTI	Metodi računarske analize podataka dobijenih iz genoma i bioloških sekvenci. Poređenje sekvenci, skriveni Markovljevi modeli, dizajn algoritama.	Python
Kompjuterska vizija	SIS	Principi formiranja slike, izdvajanje lokalnih obeležja, prepoznavanje objekata i lica na slici, detekcija kretanja, pretraživanja velikih baza slika.	Matlab
Mašinsko učenje	SIS	Problemi klasifikacije. Nadgledano i nenadgledano učenje. Učenje potkrepljivanjem. Linearna i logistička regresija. Metode numeričke optimizacije. Stabla odlučivanja. Metoda nosećih vektora.	Matlab, Octave
Metode <i>soft computing</i> -a	SIS	Projektovanje neuralnih mreža, fazi sistema i genetskih algoritama. Rešavanje problema upravljanja sistemima, obrade signala i prepoznavanja oblika.	Matlab
Obrada prirodnih jezika	SI	Mašinsko učenje u obradi prirodnih jezika. Jezički modeli. Stemovanje i lematizacija. Parsiranje. Klasifikacija tekstova na osnovu tematike i sentimenta. Leksička i distribucionalna semantika. Semantička sličnost. Prepoznavanje imenovanih entiteta.	Java, Python
Optimizacioni algoritmi u inženjerstvu	MT	Slučajno i sistematsko pretraživanje gradijentni metod, tradicionalni metodi, simpleks algoritmi, genetički algoritam, simulirano kaljenje, diferencijalna evolucija i algoritam kretanja jata. Višekriterijumska optimizacija.	Python, C/C++
Pronalaženje skrivenog znanja	RTI	Nadgledano mašinsko učenje. Naivni bayesovski klasifikator. Linearna regresija. Logistička regresija. Metode potpornih vektora. K najbližih suseda. K srednjih vrednosti. Stabla odlučivanja.	Weka, Java, Python
Razvoj bezbednog softvera	SI	Analiza sigurnosnih zahteva. Arhitektura i dizajn bezbednog softvera. Modelovanje pretnji i pravci napada. Procena rizika. Sigurnosni propusti internet aplikacija, skript jezika, API-a i programskih jezika. Sigurnosna revizija programskog koda.	Softverski alati
Sociološki i profesionalni aspekti računarstva	RTI	Istorija računarstva i uticaj računara na društvo. Razvoj i uticaj Interneta. Digitalni jaz. Zaštita privatnosti. GDPR. Socijalni aspekti sigurnosti. Licenciranje softvera. Zaštita intelektualne svojine. Softverska piraterija. Profesionalni etički kodeksi.	/
Softversko inženjerstvo velikih baza podataka	SI	Modeli i arhitekture velikih DB: NoSQL, NotOnlySQL, LogicalDataWarehouse, InMemoryComputing, BigData. Načini integracije. Računarstvo u oblaku.	Softverski alati
Statistička klasifikacija signala	SIS	Osnovne tehnike statističke klasifikacije signala: testiranje hipoteza, parametarska klasifikacija, neparametarska klasifikacija. Izdvajanje i analiza obeležja. Redukcija dimenzija podataka. Bayesovi klasifikatori.	Matlab
Teorija algoritama	SI	Napredne strukture podataka. Obrada stringova i skupova. Napredni grafovski algoritmi. Geometrijski algoritmi. Paralelni algoritmi. Kompresija. Kriptografski algoritmi. Genetski algoritmi. Dinamičko i linearno programiranje.	/
Veštačka inteligencija	PM	Pregled razvoja i oblasti primene veštačke inteligencije. Inteligentni sistemi i rešavanje problema. Znanje, rasuđivanje, planiranje. Neizvesno znanje i planiranje. Automatsko rezonovanje i delovanje. Mašinsko učenje i obučavanje.	/

3. ANALIZA IZBORA STUDENATA

Studenti vrše izbor predmeta na master akademskim studijama prilikom upisa studija, uz mogućnost kasnije zamene. U okviru Tabele 2 se može videti broj studenata po predmetima i modulima u periodu 2019-2021. Većina predmeta je bila birana od strane studenata u navedenom periodu, dok je samo za predmet Teorija algoritama ustanovljeno da se ne drži, odnosno da je u proteklom periodu privremeno povučen iz ponude.

Tabela 2. Broj studenata po predmetima iz oblasti nauke o podacima u periodu 2019-2021. Kolona Uk. predstavlja ukupan broj prvi put upisanih studenata koji su prijavili predmet, uključujući i studente sa drugih modula.

Predmet	2019.				2020.				2021.			
	RTI	SI	SIS	Uk.	RTI	SI	SIS	Uk.	RTI	SI	SIS	Uk.
Analiza socijalnih mreža	7	85	0	93	24	112	2	138	10	68	6	86
Genomska informatika	13	12	6	34	25	41	6	74	14	9	1	27
Kompjuterska vizija	0	1	44	46	0	1	54	55	0	1	52	54
Mašinsko učenje	0	4	39	44	2	2	43	51	1	6	43	56
Metode <i>soft computing</i> -a	1	1	45	47	0	1	22	24	0	0	37	37
Obrada prirodnih jezika	11	64	9	84	6	67	17	90	1	13	5	19
Optimizacioni algoritmi u inženjerstvu	0	0	0	0	0	3	12	21	1	0	19	27
Pronalaženje skrivenog znanja	33	50	0	84	46	67	0	114	26	58	1	88
Razvoj bezbednog softvera	6	25	0	31	4	39	0	43	12	94	0	106
Sociološki i profesionalni aspekti računarstva	31	63	0	96	20	72	1	93	18	79	1	100
Softversko inženjerstvo velikih baza podataka	1	46	0	47	6	34	2	42	4	37	0	41
Statistička klasifikacija signala	0	0	35	35	0	0	32	32	0	0	40	40
Teorija algoritama	0	0	0	0	0	0	0	0	0	0	0	0
Veštačka inteligencija	0	0	0	4	0	1	0	4	0	2	0	6
Ukupno upisanih	53	140	70	471	59	154	72	484	40	146	74	460

Najpre se može primetiti da studenti uglavnom biraju predmete sa njihovog modula. Studenti modula Signali i

sistemi ne biraju suviše predmete sa preostala dva modula, a slično važi i u suprotnom slučaju. Nešto više interesovanja iskazuju ka predmetima Obrada prirodnih jezika, Analiza socijalnih mreža i Genomska informatika koji se više naslanjaju na materiju koja se izučava na ovom modulu. Studenti modula Softversko inženjerstvo i Računarska tehnika i informatika biraju predmete sa oba navedena modula. S obzirom na sličnost istoimenih programa na osnovnim studijama, kao i Katedru za računarsku tehniku i informatiku koja je odgovorna za oba odeljaka, to se moglo i očekivati. Pritom, studenti koji su završili Računarsku tehniku i informatiku na osnovnim studijama, često na master studijama upisuju Softversko inženjerstvo. U proseku su studenti najčešće birali predmete Analiza socijalnih mreža, Sociološki i profesionalni aspekti računarstva i Pronalaženje skrivenog znanja. Od predmeta sa modula Signali i sistemi, studenti dva navedena modula su birali pre svega Mašinsko učenje, u nešto većoj meri studenti modula Softversko inženjerstvo.

Dodatno je izvršena analiza kombinacija predmeta koje su svi studenti na master akademskim studijama Elektrotehnika i računarstvo birali. Identifikovano je čak 1289 kombinacija. Bar dva studenta je izabralo 87 kombinacija, što znači da studenti relativno nesinhronizovano vrše izbor predmeta i koriste izbornost u maksimalnoj mogućoj meri. Ipak, može se primetiti da studenti modula SI prave nešto usaglašenije izbore. Dve kombinacije predmeta je izabralo 8 studenata, dve 6 studenata, a jednu 5 studenata, kao što se može videti u Tabeli 3. Može se primetiti da u prve četiri kombinacija dominiraju predmeti iz oblasti nauke o podacima identifikovani u Tabeli 1 sa po tri zastupljena predmeta. Najzastupljeni su predmeti Sociološki i profesionalni aspekti računarstva, Analiza socijalnih mreža, i Pronalaženje skrivenog znanja koje su studenti često birali i prema Tabeli 2.

Takođe, određeno je i koliko je studenata uzelo 3, 4 ili 5 predmeta sa liste identifikovane Tabelom 1. Rezultati su prikazani u Tabeli 4. Može se primetiti da su studenti modula SI prilično koherentno birali predmete iz oblasti nauke o podacima i da je više od polovine upisanih studenata na modul izabralo bar tri predmeta iz oblasti svake od analiziranih školskih godina. Slično se može reći i za studente modula SIS, kao i modula RTI.

Tabela 3. Prvi pet najčešćih kombinacija izabranih predmeta

No.	Modul	Broj studenata	Izabrani predmeti
1	SI	8	Analiza socijalnih mreža, Odabrana poglavlja iz numeričke analize, Programiranje korisničkih interfejsa, Sociološki i profesionalni aspekti računarstva, Specijalne funkcije
2	SI	8	Analiza socijalnih mreža, Genomska informatika, Obrada prirodnih jezika, Programiranje korisničkih interfejsa, Funkcionalno programiranje
3	SI	6	Analiza socijalnih mreža, Odabrana poglavlja iz numeričke analize, Programiranje korisničkih interfejsa, Pronalaženje skrivenog znanja, Sociološki i profesionalni aspekti računarstva
4	SI	6	Zaštita računarskih sistema i mreža, Programiranje korisničkih interfejsa, Pronalaženje skrivenog znanja, Softversko inženjerstvo velikih baza podataka, Sociološki i profesionalni aspekti računarstva
5	SI	5	Evolucija softvera, Odabrana poglavlja iz numeričke analize, Programiranje korisničkih interfejsa Razvoj bezbednog softvera, Sociološki i profesionalni aspekti računarstva

Tabela 4. Broj studenata koji su izabrali tačno 3, 4 ili 5 iz oblasti nauke o podacima po modulima i godinama

Modul	2019.				2020.				2021.			
	Br. stud.	3	4	5	Br. stud.	3	4	5	Br. stud.	3	4	5
RTI	42	20	3	0	55	24	4	0	33	18	3	0
SI	136	45	24	1	152	67	38	2	138	60	19	0
SIS	67	25	15	4	72	18	20	4	72	23	21	5

4. ZAKLJUČAK

Analiza sprovedena u okviru ovog rada je pokazala da postoji značajno interesovanje studenata za predmete iz oblasti nauke o podacima na master akademskim studijama Elektrotehnika i računarstvo na Elektrotehničkom fakultetu Univerziteta u Beogradu. U svetlu mogućnosti za uvođenje novog studijskog programa ili modula postojećih master studija koji bi se bavio naukom o podacima, ovi rezultati mogu poslužiti kao dobra procena interesovanja studenata za ovu oblast iz koje bi proisteklo definisanje upisnih kvota. Svakako bi u sklopu takvog novog programa bilo potrebno definisati nekoliko novih predmeta iz ove oblasti, naročito u vezi matematičkih osnova oblasti, ali i u vezi manipulacije podacima, skladištenja podataka i sl.

5. LITERATURA

- [1] Wil Van Der Aalst, *Data science in action, Process mining*. Springer, Berlin, Heidelberg, 2016., pp. 3-23
- [2] Vasant Dhar, *Data science and prediction*, Communications of the ACM, 56(12), pp. 2013., pp 64-73
- [3] *Computing Competencies for Undergraduate Data Science Curricula*, ACM Data Science Task Force, January 2021.
- [4] *50 Best Jobs in America for 2019-2021*, Glassdoor, pristupano 13.01.2022., https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm
- [5] *Master 4.0 [IT + BIZ]*, Inicijativa „Digitalna Srbija“, pristupano 13.01.2022., <https://www.dsi.rs/master-40-it-biznis/>
- [6] *Master akademske studije*, Elektrotehnički fakultet, 2021., <https://www.etf.bg.ac.rs/sr/studiranje/master-akademske-studije>
- [7] *Master in Data Science*, ETH Zurich, <https://inf.ethz.ch/studies/master/master-ds.html>