

SVEUČILIŠTE U RIJECI
 FAKULTET INFORMATIKE I DIGITALNIH TEHNOLOGIJA
 Radmile Matejčić 2, Rijeka

Akadska godina 2022./2023.

OSNOVNI PODACI O PREDMETU		
Naziv predmeta	Dubinska analiza podataka	
Studijski program	Diplomski studij Informatika / Diplomski studij Informatika (nastavnički smjer)	
Status predmeta	obvezatan za modul IIS	
Semestar	1.	
Bodovna vrijednost i način izvođenja nastave	ECTS koeficijent opterećenosti studenata	6
	Broj sati (P+V+S)	30+30+0
Nositelj predmeta	Prof. dr. sc. Maja Matetić / Izv. prof. dr. sc. Marija Brkić Bakarić	
E-mail	majam@uniri.hr / mbrkic@uniri.hr	
Ured	Radmile Matejčić 2, kabinet O-407 / kabinet O-408	
Vrijeme konzultacija	uz prethodni dogovor e-mailom, četvrtkom 11.00.-12.00.	
Asistent	Dejan Ljubobratović, mag. educ.	
E-mail	dejan.ljubobratovic@uniri.hr	
Ured	Radmile Matejčić 2, kabinet O-522	
Vrijeme konzultacija	Petkom od 11:30 do 12:30 ili po dogovoru e-mailom	
DETALJNI OPIS PREDMETA		
<i>Ciljevi predmeta</i>		
Automatsko prikupljanje podataka i napredne tehnologije baza podataka doveli su do velikih količina pohranjenih podataka. Cilj je predmeta osposobiti studenta za primjenu postupaka dubinske analize podataka u zadatku otkrivanja znanja u podacima u raznim područjima primjene.		
<i>Uvjeti za upis predmeta</i>		
Nema uvjeta za upis predmeta.		
<i>Očekivani ishodi učenja za predmet</i>		
Očekuje se da će nakon uspješno ispunjenih svih programom predviđenih obveza na predmetu student biti sposoban:		
<ol style="list-style-type: none"> 11. Identificirati različite zadatke dubinske analize podataka i objasniti algoritme na kojima se temelje postupci dubinske analize podataka. 12. Razlikovati tipove podataka u dubinskoj analizi podataka i postupke pripreme podataka uporabom tehnika predobrade. 13. Demonstrirati izvedbu praktičnog rada koji zahtijeva primjenu prikladnih metoda dubinske analize podataka, vrednovati i interpretirati rezultate. 14. Kritički istražiti i odabrati algoritme za zadatak podatkovne analize. 15. Razviti vlastitu programsku podršku za analizu određenog skupa podataka temeljenu na algoritmima dubinske analize podataka. 16. Odabrati i primijeniti napredne postupke dubinske analize podataka za rješavanje različitih problema u poslovnoj domeni, interpretirati rezultate i prikazati rješenja zadanih problema. 17. Objasniti važnost linearne algebre za podatkovnu znanost i strojno učenje, parametre koji opisuju brojeve s pomičnim zarezom i različite aritmetike pomičnog zareza i njihov utjecaj na točnost izračuna. 		

18. Analizirati uvjetovanost izvođenja, složenost i stabilnost algoritama numeričke linearne algebre za rješavanje problema, primjerice linearnog problema najmanjih kvadrata i sustava linearnih jednadžbi.		
<i>Sadržaj predmeta</i>		
Sadržaj predmeta čine teme: <ul style="list-style-type: none"> • Definicija i područja primjene dubinske analize podataka. Tipovi podataka. • Priprema podataka za analizu. Balansiranost skupova podataka. • Postupci odabira značajki. • Postupci klasifikacije, grupiranja, asocijacijske analize. • Učenje, vrednovanje i interpretacija modela. • Otkrivanje anomalija. Izbjegavanje lažnih otkrića. • Korištenje alata otvorenog pristupa za dubinsku analizu podataka. • Projekt dubinske analize podataka. 		
<i>Vrsta izvođenja nastave</i>	<input checked="" type="checkbox"/> predavanja	<input checked="" type="checkbox"/> samostalni zadaci
	<input checked="" type="checkbox"/> seminari i radionice	<input type="checkbox"/> multimedija i mreža
	<input checked="" type="checkbox"/> vježbe	<input type="checkbox"/> laboratorij
	<input checked="" type="checkbox"/> obrazovanje na daljinu	<input type="checkbox"/> mentorski rad
	<input type="checkbox"/> terenska nastava	<input type="checkbox"/> ostalo
<i>Komentari</i>	Nastava će se izvoditi kombinirajući rad u učionici i samostalni rad izvan učionice, uz korištenje sustava za e-učenje. Studenti će samostalno ili timski raditi na projektnom zadatku.	
<i>Obavezna literatura (u trenutku prijave prijedloga studijskog programa)</i>		
<ol style="list-style-type: none"> 1. Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, 2nd ed., Pearson, 2019. 2. Shmueli, Galit, Peter C. Bruce, Inbal Yahav, Nitin R. Patel, and Kenneth C. Lichtendahl Jr. Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons, 2017. 3. James W. Demmel: Applied Numerical Linear Algebra, SIAM 1997. 		
<i>Dopunska literatura (u trenutku prijave prijedloga studijskog programa)</i>		
<ol style="list-style-type: none"> 1. Data Mining: Practical Machine Learning Tools and Techniques, Ian Witten, Eibe Frank, Mark Hall, 4th ed., Morgan Kaufmann, 2016. 2. Data Mining: The Textbook, Charu C. Aggarwal Hardcover, Springer, 2015 3. Gareth, James, Witten Daniela, Hastie Trevor, and Tibshirani Robert. An introduction to statistical learning: with applications in R. Springer, 2021. 4. Bruce, Peter, Andrew Bruce, and Peter Gedeck. Practical statistics for data scientists: 50+ essential concepts using R and Python. O'Reilly Media, 2020. 		
<i>Načini praćenja kvalitete koji osiguravaju stjecanje izlaznih znanja, vještina i kompetencija</i>		
Predviđa se periodičko provođenje evaluacije s ciljem osiguranja i kontinuiranog unapređenja kvalitete nastave i studijskog programa (u okviru aktivnosti Odbora za upravljanje i unapređenje kvalitete Fakulteta informatike i digitalnih tehnologija). U zadnjem tjednu nastave provodit će se anonimna		

evaluacija kvalitete održane nastave od strane studenata. Provest će se i analiza uspješnosti studenata na predmetu (postotak studenata koji su položili predmet i prosjek njihovih ocjena).

Mogućnost izvođenja na stranom jeziku | Ne

OBVEZE, PRAĆENJE RADA I VREDNOVANJE STUDENATA

VRSTA AKTIVNOSTI	ECTS	ECTS - PRAKTIČNI RAD	ISHODI UČENJA	SPECIFIČNA AKTIVNOST	METODA PROCJENJIVANJA	BODOVI MAX.
Pohađanje nastave i aktivnosti u nastavi	2	0.5	1-8	Aktivnost na satu	Sudjelovanje u rješavanju problema	5
Kontinuirana provjera znanja - kvizovi	1	0.25	1-8	Dva kviza (provjere na računalu: teorija)	2 kviza: (12+13) bodova ovisno o stupnju točnosti i potpunosti	25
Kontinuirana provjera znanja - kolokviji	1	1	1-8	2 praktična kolokvija (rješavanje zadataka na računalu)	2 praktična kolokvija: (20+20) bodova ovisno o stupnju točnosti i potpunosti	40
Seminarski rad i projektni zadatak (završni ispit)	2	2	1-8	Primjena podatkovne analize u određenoj domeni primjene	Prezentacija seminarskog rada i projekta (na kraju semestra ili na ispitnom roku)	30
UKUPNO	6	3.75				100

Obveze i vrednovanje studenata

1. Pohađanje nastave i aktivnosti u nastavi

Nastava se odvija prema mješovitom modelu u kombinaciji klasične nastave u učionici i *online* nastave uz pomoć sustava za e-učenje prema rasporedu koji je prikazan je tablicom u nastavku. Studenti su dužni koristiti sustav za e-učenje Merlin (<https://moodle.srce.hr/>) gdje će se objavljivati informacije o predmetu, materijali za učenje, zadaci za vježbu, zadaci za domaće zadaće te obavijesti vezane za izvođenje nastave (putem foruma Obavijesti).

2. Kontinuirana provjera znanja – teorija

Tijekom semestra pisat će se u učionici uz nadzor nastavnika dva kviza putem sustava Merlin koji uključuje teorijska pitanja i praktične zadatke. Na ovoj aktivnosti studenti će moći skupiti ukupno maksimalno 25 bodova (definiran je ukupni prag od 6 bodova za zbroj bodova kvizova).

3. Kontinuirana provjera znanja – praktični dio

U okviru vježbi održavaju se dva praktična kolokvija (rješavanje zadataka na računalu) koji se boduju sa ukupno 40 bodova (20+20 bodova, definiran je ukupni prag od 10 bodova za zbroj bodova kolokvija).

4. Završni ispit (Seminarski rad i projekt)

Tijekom semestra studentima će biti ponuđene teme za projektni zadatak. . Seminarski rad i projekt podrazumijeva istraživanje određene teme u okviru projekta. Obrana projekta predstavlja završni ispit. Studenti za koje se utvrdi da su u seminarskom radu koristili veće dijelove doslovno kopiranog teksta dobit će 0 bodova.

Seminarski rad i/ili projektni zadatak studenti mogu rješavati samostalno i/ili u projektnim timovima. Uloge članova projekta:

- a) Priprema podataka za analizu.
- b) Odabir naprednih metoda za dubinsku analizu podataka, kratak opis i objašnjenje metoda.
- c) Izvođenje eksperimenata primjenom metoda dubinske analize podataka kao priprema za izradu projekta.
- d) Prikaz rezultata, interpretacija rezultata.
- e) Izrada izvješća projekta (svaki član piše vlastito izvješće u par rečenica o tome koji su mu bili zadaci na projektu i kako je tim funkcionirao).
- f) Predaja zajedničkog seminara i izvješća na kraju semestra (prva verzija).
- g) Cijeli tim zajednički predstavlja rezultate projekta na kraju semestra ili na ispitnom roku. Svaki član tima izlaže dio kojim se bavio na projektu.

Barem dva dana prije obrane potrebno je na Merlin predati projektnu dokumentaciju (kratki opis primijenjenih metoda, skupa podataka i izvedenih eksperimenata te diskusiju i interpretaciju rezultata).

Seminarski rad i projekt se izlaže po dogovoru na kraju semestra ili u okviru termina završnog ispita i nosi ukupno 30 bodova, a smatra se pozitivno ocijenjenim samo ako student postigne minimalno 50%-ni uspjeh (15-30 bodova).

5. Nadoknada aktivnosti

Nadoknadi kolokvija 27.01.2023. student ima pravo pristupiti samo u slučaju opravdanog izostanka s kviza/kolokvija uz predočenje ispričnice asistentu u roku 10 dana nakon izostanka.

Ocjenjivanje

Kontinuiranim radom tijekom semestra na prethodno opisani način studenti mogu ostvariti najviše 70 ocjenskih bodova, a da bi mogli pristupiti završnom ispitu moraju ostvariti 50% i više bodova (minimalno 35).

Završni ispit nosi udio od maksimalno 30 ocjenskih bodova, a smatra se položenim samo ako na njemu student postigne minimalno 50%-ni uspjeh (ispitni prag je 50% uspješno riješenih zadataka).

Ukoliko je završni ispit prolazan, skupljeni bodovi će se pribrojati prethodnima i prema ukupnom rezultatu formirati će se pripadajuća ocjena. U suprotnom, student ima pravo pristupa završnom ispitu još 2 puta (ukupno do 3 puta).

Konačna ocjena iz predmeta

Donosi se na osnovu zbroja svih bodova prikupljenih tijekom izvođenja nastave prema sljedećoj skali:

A – 90% - 100%	(ekvivalent: izvrstan 5)
B – 75% - 89,9%	(ekvivalent: vrlo dobar 4)
C – 60% - 74,9%	(ekvivalent: dobar 3)
D – 50% - 59,9%	(ekvivalent: dovoljan 2)
F – 0% - 49,9%	(ekvivalent: nedovoljan 1)

Ispitni rokovi

Redoviti:

3.02.2023., 17.02.2023.

Izvanredni:

9.03.2023., 14.09.2023.

RASPORED NASTAVE – zimski (I.) semestar akademske godine 2022./2023.

Nastava će se na predmetu odvijati u zimskom semestru prema sljedećem rasporedu:

predavanja: petkom od 12:00 do 13:30 sati u prostoriji O-365

vježbe: petkom od 14:00 do 15:30 sati u prostoriji O-365

Tj.	Datum	Vrijeme	Prostor*	Tema	Nastava	Izvođač
1.	7.10.2022.	12:00	365	Uvod u dubinsku analizu podataka	P1	izv. prof. Marija Brkić Bakarić
1.	7.10.2022.	14:00	365	Uvod u predmet. R - Osnovne računске operacije, varijable i vektori	V1	Dejan Ljubobratović
2.	14.10.2022.	12:00	365	Pretprocesiranje podataka	P2	izv. prof. Marija Brkić Bakarić
2.	14.10.2022.	14:00	365	R - matrice i podatkovni okviri (dataframes)	V2	Dejan Ljubobratović
3.	21.10.2022.	12:00	365	Temeljni koncepti i tehnike klasifikacije	P3	izv. prof. Marija Brkić Bakarić
3.	21.10.2022.	14:00	365	R - Filtriranje, sortiranje i mijenjanje (mutate) podataka. Vizualizacija.	V3	Dejan Ljubobratović
4.	28.10.2022.	12:00	365	Pretreniranost. Odabir modela. Vrednovanje modela.	P4	izv. prof. Marija Brkić Bakarić
4.	28.10.2022.	14:00	365	Pearsonov koeficijent korelacije, odabir i evaluacija modela, pretreniranost	V4	Dejan Ljubobratović
5.	4.11.2022.	12:00	365	Klasifikacijski postupci, prvi dio	P5	izv. prof. Marija Brkić Bakarić
5.	4.11.2022.	14:00	365	Stabla odlučivanja, Klasifikacija K najbližih susjeda (KNN)	V5	Dejan Ljubobratović
6.	11.11.2022.	12:00	365	Klasifikacijski postupci, drugi dio	P6	izv. prof. Marija Brkić Bakarić
6.	11.11.2022.	14:00	365	1. kolokvij (praktični dio)	V6	Dejan Ljubobratović
7.	18.11.2022.	-	online	Asocijacijska analiza, osnovni algoritmi	P7	prof. dr. sc. Maja Matetić
7.	18.11.2022.	-	online	Bayesove mreže	V7	Dejan Ljubobratović
8.	25.11.2022.	12:00	365	Prvi kviz teorije	P8	izv. prof. Marija Brkić Bakarić
8.	25.11.2022.	14:00	365	Naïve Bayes klasifikacija, Logistička regresija	V8	Dejan Ljubobratović
9.	2.12.2022.	12:00	365	Asocijacijska analiza, napredni algoritmi	P9	prof. dr. sc. Maja Matetić
9.	2.12.2022.	14:00	365	Umjetne neuronske mreže (ANN-Artificial Neural Network)	V9	Dejan Ljubobratović
10.	9.12.2022.	12:00	365	Grupiranje, osnovni algoritmi	P10	prof. dr. sc. Maja Matetić
10.	9.12.2022.	14:00	365	Grupiranje K-srednjih vrijednosti (K-means clustering)	V10	Dejan Ljubobratović
11.	16.12.2022.	12:00	365	Grupiranje, napredni algoritmi	P11	prof. dr. sc. Maja Matetić
11.	16.12.2022.	14:00	365	Aglomerativno hijerarhijsko grupiranje, DBSCAN	V11	Dejan Ljubobratović
12.	23.12.2022.	12:00	365	Otkrivanje anomalija	P12	prof. dr. sc. Maja Matetić
12.	23.12.2022.	14:00	365	Testiranje značajnosti, Testiranje hipoteze	V12	Dejan Ljubobratović
13.	13.1.2023.	12:00	365	Izbjegavanje lažnih otkrića	P13	prof. dr. sc. Maja Matetić
13.	13.1.2023.	14:00	365	2. kolokvij (praktični dio)	V13	Dejan Ljubobratović

14.	20.1.2023.	12:00	365	Drugi kviz teorije	P14	prof. dr. sc. Maja Matetić
14.	20.1.2023.	14:00	365	Metode izbjegavanje lažnih otkrića	V14	Dejan Ljubobratović
15.	27.1.2023.		365	Nadoknade	P15 i V15	prof. dr. sc. Maja Matetić, izv. prof. Marija Brkić Bakarić, Dejan Ljubobratović

*upisati broj prostorije ili online

P – predavanja

V – vježbe

Napomena: Moguće su manje izmjene rasporeda nastave. Za nove verzije rasporeda potrebno je pratiti obavijesti u e kolegiju.