

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

Akademski godina 2023./2024.

OSNOVNI PODACI O PREDMETU		
Naziv predmeta	Analitika podataka velikog obujma	
Studijski program	Sveučilišni prijediplomski studij Informatika	
Status predmeta	obvezatan za modul IIS	
Semestar	2.	
Bodovna vrijednost i nastavno opterećenje	ECTS koeficijent opterećenosti studenata	6
	Broj sati (P+V+S)	30+30+X
Nositelj predmeta	prof. dr. sc. Sanda Martinčić-Ipšić	
E-mail	smarti@uniri.hr	
Ured	O-409	
Vrijeme konzultacija	<i>Ponedjeljkom od 10:00 do 11:00 uz prethodni dogovor e-mailom</i>	
Asistent	Karlo Babić	
E-mail	karlo.babic@uniri.hr	
Ured	O-419	
Vrijeme konzultacija	<i>Srijedom od 15:30 do 16:00 uz prethodni dogovor e-mailom</i>	
DETALJNI OPIS PREDMETA		
<i>Ciljevi predmeta</i>		
Razviti sustav za analitiku podataka velikog obujma postavljanjem arhitekture i analitičkog rješenja za postavljeni problem: pronalaženje sličnih i/ili učestalih proizvoda, usluga ili korisnika, analitike podataka u beskonačnom toku, praćenje oglašavanja, preporučivanje sadržaja odnosno proizvoda ili usluga, praćenje korisničkog mišljenja, pretvaranje nestrukturiranog sadržaja u strukturirani ili polustrukturirani format, predviđanja novih veza, praćenje trendova u vremenskim serijama i sličnih.		
<i>Uvjeti za upis predmeta</i>		
Odslušani predmeti Infrastruktura za podatke velikog obujma i Dubinska analiza podataka.		
<i>Očekivani ishodi učenja za predmet</i>		
Očekuje se da nakon izvršavanja svih programom predviđenih obveza studenti budu sposobni:		
<ol style="list-style-type: none"> 11. Osmisliti organizaciju skupova podataka u heterogenom okruženju obrade podataka velikog obujma s aspekta distribuiranih, polustrukturiranih i nestrukturiranih podataka primjerenih postavljenom problemu uključujući relacijske baze i skladišta podataka, NoSQL baze podataka, podatkovna jezera, te baze za zapis znanja ontologije i grafove znanja. 12. Konceptualno i logički oblikovati NoSQL bazu podataka primjerenu za jedan od načina zapisa ključeva-vrijednosti, stupaca, grafova ili dokumenata te predložiti implementaciju za zadani problem. 13. Predložiti i vrednovati implementaciju NoSQL baze uz formulaciju upita za pretraživanja u odgovarajućoj paradigmi za zadani domenski problem analitike podataka velikog obujma. 14. Kritički vrednovati tehnologije za rad s heterogenim, distribuiranim, polustrukturiranim i nestrukturiranim podacima, odnosno tokovima podataka (engl. streaming data). 		

- 15. Vrednovati sustav za analitiku tokovnih podataka u stvarnom vremenu.
- 16. Razviti i implementirati komponente inteligentnog informacijskog sustava na temelju velikih skupova podataka koristeći programske jezike i računalne alate koji efikasno udovoljavaju zahtjevima problema.
- 17. Osmisliti, razviti i vrednovati rješenje za odabrani problem analitike podataka velikog obujma poput: pronalaženje sličnih i/ili učestalih proizvoda, usluga ili korisnika, analitike podataka u beskonačnom toku, praćenje oglašavanja, preporučivanje sadržaja odnosno proizvoda ili usluga, praćenje korisničkog mišljenja, pretvaranje nestrukturiranog sadržaja u strukturirani ili polustrukturirani format, predviđanja novih veza, praćenje trendova u vremenskim serijama i sličnih.

Sadržaj predmeta

Na predmetu se obrađuju sljedeći sadržaji:

- Uvod u analitiku podataka velikog obujma, principi, platforme i ekosustavi. Primjeri aplikacija. Izvori i vrste podataka velikog obujma. Definicije.
- Podatkovno inženjerstvo. Principi obrade i pohrane heterogenih i distribuiranih podataka velikog obujma. Organizacija računalnih sustava za aplikacije temeljene na podacima velikog obujma. CAP, BASE i ACID teoremi. (I1, I4)
- NoSQL baze podataka temeljena na ključu-vrijednosti, stupcu, dokumentu ili grafu za zapis nestrukturiranih podataka. (I1, I2, I3)
- Pretraživanja informacija u NoSQL bazama podatka. (I1, I2, I3)
- Podatkovna jezera. Integracija baza i skladišta podatka s podacima velikog obujma u jedinstvenu infrastrukturu za poslovnu analitiku. Nove paradigme: data lakehouse, data mash. (I1, I4)
- Integracija podataka za analitičke potrebe. Grafovi znanja.(I1, I4)
- HADOOP. Paketni načini obrade podataka velikog obujma Map-Reduce. SQL kao Map-Reduce. (I4, I6)
- Procesiranje i analitika beskonačnih tokova podataka. Obrada tokovnih podataka u stvarnom vremenu. Principi sažimanja i analitike na toku podataka- Porazdijeljeni tokovni podaci. (I4, I5)
- Tokovi podatka bazirani na logovima (Kafka). Pretraživanje velikih količina nestrukturiranih (tekstualnih) podataka (Elastic search). (I4, I5)
- Principi strojnog učenja za podatke velikog obujma: alati i servisi u oblaku (I6, I7)
- Analitika podataka velikog obujma za probleme pronalaženja sličnih i/ili učestalih proizvoda, usluga ili korisnika. (I7)
- Analitika praćenja oglašavanja, preporučivanje sadržaja odnosno proizvoda ili usluga. (I7)
- Analitika praćenje korisničkog mišljenja, pretvaranje nestrukturiranog sadržaja u strukturirani ili polustrukturirani format. (I7)
- Analitika predviđanja novih veza, praćenje trendova u vremenskim serijama i sličnih. (I7)
- Trendovi analitike podataka velikog obujma.

<i>Način 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 checked="" 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.
<i>Obavezna literatura (u trenutku prijave prijedloga studijskog programa)</i>	
<ol style="list-style-type: none"> 1. Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 3rd edition, 2020. 2. Nathan Marz and James Warren, Big Data: Principles and best practices of scalable realtime data systems, Manning Pub. 2015. 3. Andrew G. Psaltis: Streaming Data - Understanding the real-time pipeline, Manning Pub, 2017. 4. Guy Harrison, Next Generation Databases: NoSQL, NewSQL, and Big Data, Apress, 2015. 5. Luc Perkins, Eric Redmond, Jim Wilson, Seven Databases in Seven Weeks, 2nd Edition, O'Reilly, 2018. 6. Joe Reis, Matt Housley, Fundamentals of Data Engineering, O'Reilly, 2022. 7. Kleppmann, M. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. O'Reilly, 2017. (part III) 8. Kleppmann, M. Making Sense of Steam Procesing . O'Reilly, 2016. 9. Sadržaji pripremljeni za učenje putem sustava za učenje. 	
<i>Dopunska literatura (u trenutku prijave prijedloga studijskog programa)</i>	
<ol style="list-style-type: none"> 1. Dan McCreary, Ann Kelly Making Sense of NoSQL, 1st Edition, Manning , 2013. 2. Dean Wampler, Fast Data Architectures for Streaming Applications, O'Reilly, 2016, http://www.oreilly.com/data/free/fast-data-architectures-for-streaming-applications.csp (17.2.2022.) 3. Scalable Systems for Big Data Analytics: A Technology Tutorial, IEEE, 2014. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6842585 4. Dylan Scott, Viktor Gamov, Dave Klein, Kafka in Action, Manning, 2022. https://livebook.manning.com/book/kafka-in-action/ 5. Jimmy Lin, Chris Dyer, Data-Intensive Text Processing with MapReduce, Morgan& Claypool, 2010 6. Alex Gorelik, The Enterprise Big Data Lake . O'Reilly, 2019. 	
<i>Načini praćenja kvalitete koji osiguravaju stjecanje izlaznih znanja, vještina i kompetencija</i>	
<p>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).</p>	
<i>Jezik izvođenja nastave</i>	Hrvatski jezik
<i>Mogućnost izvođenja na stranom jeziku</i>	Nema

OBVEZE, PRAĆENJE RADA I VREDNOVANJE STUDENATA

Konstruktivno povezivanje

VRSTA AKTIVNOSTI	ECTS	ECTS - PRAKTIČNI RAD	ISHODI UČENJA	SPECIFIČNA AKTIVNOST	METODA PROCJENJIVANJA	BODOVI MAX.
Pohađanje nastave i aktivnosti u nastavi	2	1	I1–I7	Prisutnost studenata i odgovaranje na pitanja nastavnika	Evidencija prisustva	0
Kontinuirana provjera znanja	1	1	I1–I7	Zadaci na vježbama, industrijske radionice,	Kontinuirano rješavanje zadataka: 15 Industrijske radionice (4 po 5): 20	35
Praktični rad	2	2	I1–I7	Izrada samostalnog rada koje uključuje izgradnju sustava za analitiku podataka velikog obujma na postavljenoj arhitekturi i infrastrukturi.	Ovisno o stupnju točnosti i potpunosti Programski kod Tehnički opis rješenja Prezentacija rješenja	35
Završni ispit	1	0	I1–I7	Pisani ispit iz teorije	Prema unaprijed razrađenim kriterijima (0-30 bodova)	30
UKUPNO	6	4				100

Obveze i vrednovanje studenata – puna nastavna satnica

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).

Studenti koji studiraju u punoj nastavnoj satnici dužni su redovito pohađati nastavu, aktivno sudjelovati tijekom nastave te izvršavati aktivnosti predmeta u okviru sustava Merlin koje će nastavnici najavljivati putem foruma.

...

2. Kontinuirana provjera znanja

Kontinuiranim praćenjem tijekom semestra studentima će se bodovati zadaće vezane uz I1-I7. Student će raditi s različitim rješenjima za pohranu i obradu podataka velikog obujma, studenti će sudjelovati u četiri industrijske radionice (praktične radionice).

3. Praktični rad

Student će izraditi samostalnim praktični rad, koje uključuje izgradnju sustava za analitiku podataka velikog obujma na postavljenoj infrastrukturi ili servisima u oblaku za izabran problem (pronalaženje sličnih i/ili učestalih proizvoda, usluga ili korisnika, analitike podataka u beskonačnom toku, praćenje oglašavanja, preporučivanje sadržaja odnosno proizvoda

ili usluga, praćenje korisničkog mišljenja, pretvaranje nestrukturiranog sadržaja u strukturirani ili polustrukturirani format, predviđanja novih veza, praćenje trendova u vremenskim serijama i sličnih) ispitat će se I3, I5 i I7 gdje će student pokazati praktičnu i teorijsku primjenu kroz predano rješenje, tehničku dokumentaciju te predstavljanje.

4. Ispit

Teorijski dio predmeta provjerava se na završnom pismenom ili online ispitu na kojem se od studenta traži da odgovaranjem na pitanja pokaže poznavanje koncepata analitike podataka velikog obujma (I1-I7).

Obveze i vrednovanje studenata – prilagođena nastavna satnica

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).

Studenti koji studiraju u sklopu prilagođene nastavne satnice mogu izostati s najviše 50% sati nastave (predavanja i vježbi), a dužni su aktivno sudjelovati tijekom nastave (u učionici ili *online*) te izvršavati aktivnosti predmeta u okviru sustava Merlin koje će nastavnici najavljivati putem foruma.

2. Obveze i aktivnosti vrednovanja

Obveze i vrednovanje studenata koji studiraju u sklopu prilagođene nastavne satnice, jednake su onima studenata koji studiraju u sklopu pune nastavne satnice.

Ocjenjivanje

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

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).

Ako je ispit prolazan, skupljeni bodovi će se pribrojati prethodnima i prema ukupnom rezultatu formirat će se pripadajuća ocjena. U suprotnom, student ima pravo pristupa ispitu još 2 puta (ukupno do 3 puta tijekom akademske godine).

Konačna ocjena ostvarenosti ishoda učenja na predmetu

Konačna ocjena ostvarenosti ishoda učenja na predmetu je zbroj ocjenskih bodova postignutih u kontinuiranom praćenju i vrednovanju i ocjenskih bodova postignutih na ispitu, a donosi se na sljedeći način:

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

Ispitni termini

26.06.2024.
10.07.2024.
04.09.2024.
11.09.2024.

SATNICA IZVOĐENJA NASTAVE – ljetni (II.) semestar akademske godine 2023./2024.

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

predavanja: četvrtak 12-14 u 365

vježbe: četvrtak 14-16 u 365

Tj.	Datum	Vrijeme	Prostor*	Tema	Nastava	Izvođač
1.	7.3.2024.	12:00-13:30	O-365	Uvod. Analitika podataka velikog obujma, principi, platforme i ekosustavi, primjeri i definicije.	P1	prof. dr. sc. Sanda Martinčić-Ipšić
1.	7.3.2024.	14:00-15:30	O-365	Podatkovno inženjerstvo. Principi obrade i pohrane heterogenih i distribuiranih podataka velikog obujma.	V1	prof. dr. sc. Sanda Martinčić-Ipšić
2.	14.3.2024.	12:00-13:30	O-365	Mongo DB 1	V2	Karlo Babić
2.	14.3.2024.	14:00-15:30	O-365	Mongo DB 2	V3	Karlo Babić
3.	21.3.2024.	12:00-13:30	O-365	NoSQL baze podataka temeljene na dokumentima ključu-vrijednosti i dokumentu	P2	prof. dr. sc. Sanda Martinčić-Ipšić
3.	21.3.2024.	14:00-15:30	O-365	NoSQL baze podataka temeljene na ključu-vrijednosti, stupcima i grafu.	P3	prof. dr. sc. Sanda Martinčić-Ipšić
4.	28.3.2024.	12:00-13:30	O-365	Podatkovna jezera. Integracija baza i skladišta podatka s podacima velikog obujma u jedinstvenu infrastrukturu za poslovnu analitiku.	P4	prof. dr. sc. Sanda Martinčić-Ipšić
4.	28.3.2024.	14:00-15:30	O-365	Neo4J	V4	Karlo Babić
5.	4.4.2024.	12:00-13:30	O-365	Nove paradigme: data lakehouse, data mesh. Integracija podataka za analitičke potrebe.	P5	prof. dr. sc. Sanda Martinčić-Ipšić
5.	4.4.2024.	14:00-15:30	O-365	Neo4J	V5	Karlo Babić
6.	11.4.2024.	12:00-13:30	O-365	HADOOP. Paketni načini obrade podataka velikog obujma Map-Reduce. SQL kao Map-Reduce.	P6	prof. dr. sc. Sanda Martinčić-Ipšić
6.	11.4.2024.	14:00-15:30	O-365	Pandas	V6	Karlo Babić
7.	18.4.2024.	12:00-13:30	O-365	Procesiranje i analitika beskonačnih tokova podataka. Obrada tokovnih podataka u stvarnom vremenu. Principi sažimanja i analitike na toku podataka- Porazdjeljeni tokovni podaci.	P7	prof. dr. sc. Sanda Martinčić-Ipšić
	18.4.2024.			ROK Prijava teme praktičnog rada		
7.	18.4.2024.	14:00-15:30	O-365	Spark1	V7	Karlo Babić
8.	25.4.2024.	12:00-13:30	O-365	Principi strojnog učenja za podatke velikog obujma: alati i servisi u oblaku.	P8	prof. dr. sc. Sanda Martinčić-Ipšić/Croz
8.	25.4.2024.	14:00-17:30	O-365	Spark 2	V8	Karlo Babić /Croz
9.	2.5.2024.	12:00-13:30	O-365	Sustavi preporučivanja 1	P9	prof. dr. sc. Sanda Martinčić-Ipšić
9.	2.5.2024.	14:00-15:30	O-365	Spark 3	V9	Karlo Babić
10.	9.5.2024.	12:00-13:30	O-365	Kafka (industrijska radionica)	P10	prof. dr. sc. Sanda Martinčić-Ipšić/Croz

10.	9.5.2024.	14:00-15:30	O-365	Kafka (industrijska radionica)	V10	Karlo Babić /Croz
11.	16.5.2024.	12:00-13:30	O-365	Kafka (industrijska radionica)	P11	prof. dr. sc. Sanda Martinčić-Ipšić/Croz
11.	16.5.2024.	14:00-15:30	O-365	Kafka (industrijska radionica)	V11	Karlo Babić /Croz
12.	23.5.2024.	12:00-13:30	O-365	Elastic search. (industrijska radionica)	P12	prof. dr. sc. Sanda Martinčić-Ipšić/QED
12.	23.5.2024.	14:00-17:30	O-365	Elastic search. (industrijska radionica)	V12	Karlo Babić/QED
13.	30.5.2024.	12:00-13:30	O-365	Amazon Web Servisi , Redshift Dynamo DB	P13	prof. dr. sc. Sanda Martinčić-Ipšić/QED
13.	30.5.2024.	14:00-17:30	O-365	Amazon Web Servisi , Redshift Dynamo DB	V13	Karlo Babić/QED
14.	6.6.2024.	12:00-13:30	O-365	Pretraživanje velikih količina nestrukturiranih (tekstualnih) podataka. Otkrivanje mišljenja korisnika	P14	prof. dr. sc. Sanda Martinčić-Ipšić
14.	6.6.2024.	14:00-17:30	O-365	NLP I analiza mišljenja korisnika	V14	Karlo Babić
	10.6.2024.			Rok za predaju praktičnog rada		
15.	13.6.2024.	12:00-13:30	O-365	Prezentacije praktičnih studentskih radova.	P15	prof. dr. sc. Sanda Martinčić-Ipšić
15.	13.6.2024.	14:00-15:30	O-365	Prezentacije praktičnih studentskih radova.	V15	Karlo Babić

*Napomena: upisati broj prostorije ili *online*

P – predavanja

V – vježbe