



Природно-математички факултет
Универзитет у Крагујевцу
Радора Домановића 12, 34000 Крагујевац

ИНФОРМАТИКА

Основне академске студије

Студијски програм: Основне академске студије информатике		
Назив предмета: УВОД У НАУКУ О ПОДАЦИМА		
Статус предмета: Обавезни на модулима: Рачунарске науке и Информационо-комуникационе технологије, изборни на модулу Софтверско инжењерство		
Број ЕСПБ: 6		
Услов: Уписан одговарајући семестар		
Циљ предмета Стицање конкурентних и модерних знања о фундаменталним процесима, техникама и методама које чине науку о подацима и могућностима њихове практичне примене.		
Исход предмета Савладане методе и технике оспособиће студента за: Припрему података за анализу (data wrangling) – вештину довођења података у облик погодан за визуелизацију и моделовање; Ефикасно управљање подацима; Истраживачку анализу података, постављање статистичких хипотеза и извођење закључака; Предвиђање помоћу модела базираних на подацима; Комуникацију резултатима кроз визуелизацију података и ефективне сумарне извештаје.		
Садржај предмета <i>Теоријска настава</i> Увод. Дефиниције. Интердисциплинарност. Подаци. Процеси: спецификација и разумевање проблема, припрема података за анализу (data wrangling), истраживачка анализа (exploratory data analysis) и евалуација, моделирање засновано на подацима, анализа резултата, комуникација резултатима. Примена у различитим областима. Визуелизација података. Типови података. Граматика графике. Технике динамичке и интерактивне визуелизације података. Припрема података. Квалитет података. Статистичко размишљање и закључивање. Примери на одабраној софтверској платформи. Одабрани примери расподела: Binomial, Geometric, Poisson, Exponential, Gaussian, Student's, Snedecor's F distribution, Beta, Weibull... Тестирање хипотеза о параметрима основних скупова и облику расподеле; тестови нормалности, анализа варијансе, непараметарски тестови. Редукција димензионалности и факторска анализа података. PCA - анализа главних компоненти. Моделирање података. Регресије. Вишеструка регресија (MLR). Stepwise регресије. Логистичка регресија. Моделирање помоћу вештачких неуронских мрежа. Врсте растојања. Концепти сличности. Концепти класификације и кластеризације података. Кластеризација. K-Means. K-Medoids.. Хијерархијска кластеризација. Остали приступи кластеризацији. Кластеризација категоријских података. FRM (Frequency Resency Monetary) анализа. Комуникацију резултатима кроз визуелизацију података и ефективне сумарне извештаје. Преглед блиских тема. Data Mining у текстуалним документима. Sentiment Analysis. Association rules. Обрада и препознавање слика. Временске серије. Internet of Things. Big Data. Етички проблеми. Будућност науке података. <i>Практична настава</i> Примена програмског језика R у науци о подацима. Рад на вежбама ће подразумевати примену стеченог знања на решавање конкретних актуелних проблема у различитим областима.		
Литература 1. Wickham, Hadley, and Garrett Grolemund, R za statističku obradu podataka, Mikro knjiga, 2017. 2. Hastie, Trevor,, Robert Tibshirani, and J. H Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer, 2009. 3. Pathak, Manas A., Beginning Data Science with R. Springer, 2014. 4. Schutt, Rachel, and Cathy O'Neil. Doing data science: Straight talk from the frontline. "O'Reilly Media, Inc.", 2013.		
Број часова активне наставе	Теоријска настава:	3 Практична настава: 2
Методе извођења наставе Проблемски-оријентисана настава, практична настава и вежбе уз софтверску подршку, самостални рад студената, домаћи задаци и консултације.		
Оцена знања (максимални број поена 100)		

Za slušanje predmeta je nepходно predznanje iz verovatnoće i statistike (izborni predmet iz zimskog semestra)!

Ocena znanja za predmet „Uvod u nauku o podacima“

Način polaganja ispita:

- **Predispitne obaveze: 70 poena**
 - Kolokvijum – 24 poena
 - Seminarski rad – 46 poena
- **Završni ispit: 30 poena**
-

Uslov za izradu seminarskog rada je da student osvoji minimalno 12 poena na kolokvijumu.

Seminarski rad se radi u timovima od 2 ili 3 člana. Timovi od 3 člana dodatno postavljaju modele u produkciju.

Odbrana seminarskih radova uključuje i pitanja iz prakse i pitanja iz teorije.

Uslov za izlazak na završni deo ispita je osvojenih 36 poena na predispitnim obavezama.

Seminarski rad

Pored kolokvijuma, predispitne obaveze iz predmeta Uvod u nauku o podacima stiču se analizom jedne baze podataka dostupne na sajtu [kaggle.com/datasets](https://www.kaggle.com/datasets). Možete da koristite i druge resurse koji daju podatke.

Za one najhrabrije, tema se može izabrati i sa liste završnih projekata kursa Mašinskog učenja Stanford univerziteta. Svaka tema je pokrivena radom koji opisuje proces koji treba oponašati i uključuje odgovarajući skup podataka i prateće kodove. Ovi seminarski radovi se visoko kotiraju!



1. Zašto je potrebna Nauka o Podacima?

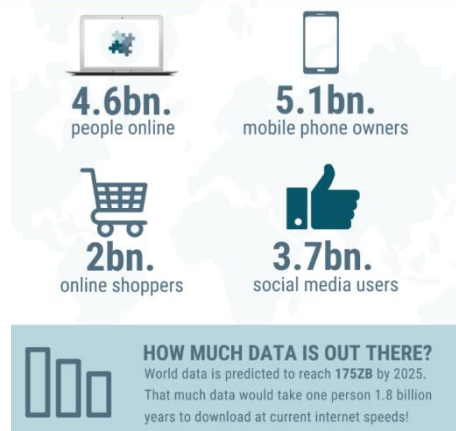
Let's start by looking if internet usage has changed since the end of 2017. By the end of 2019 there was around 4.6 billion internet users around the world, compared to 3.8 billion in June 2017. This is getting closer to two thirds of the world's population being online.

40 zettabytes of data—that's 40 trillion gigabytes.

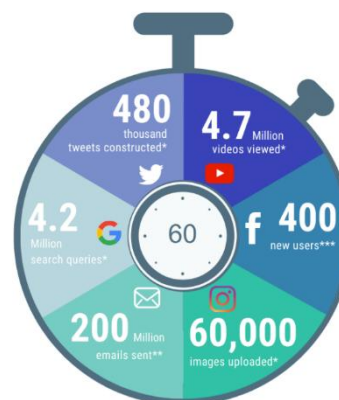
Više informacija o broju korisnika i podacima možete naći na sledećem [linku](#).

Ko sve koristi podatke:

- Cambridge Analytica i Facebook
- Google login
- Cookies
- Delta Holding investira u Ananas
- IBM...

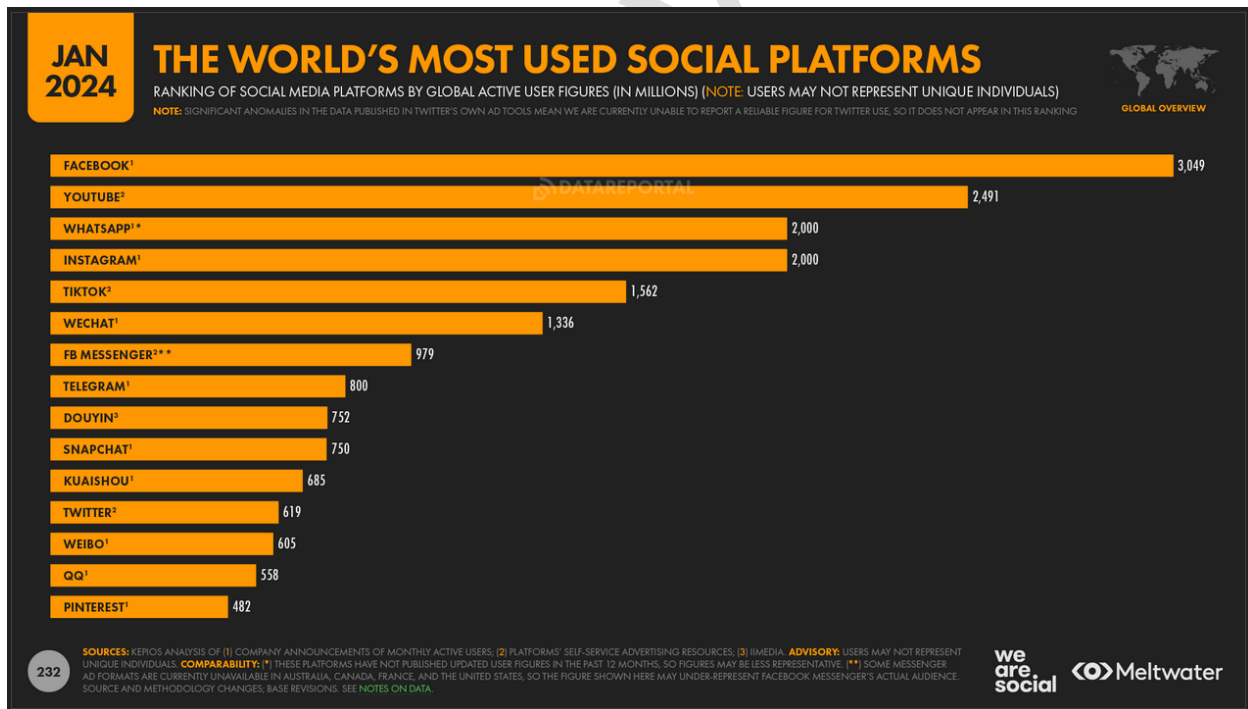
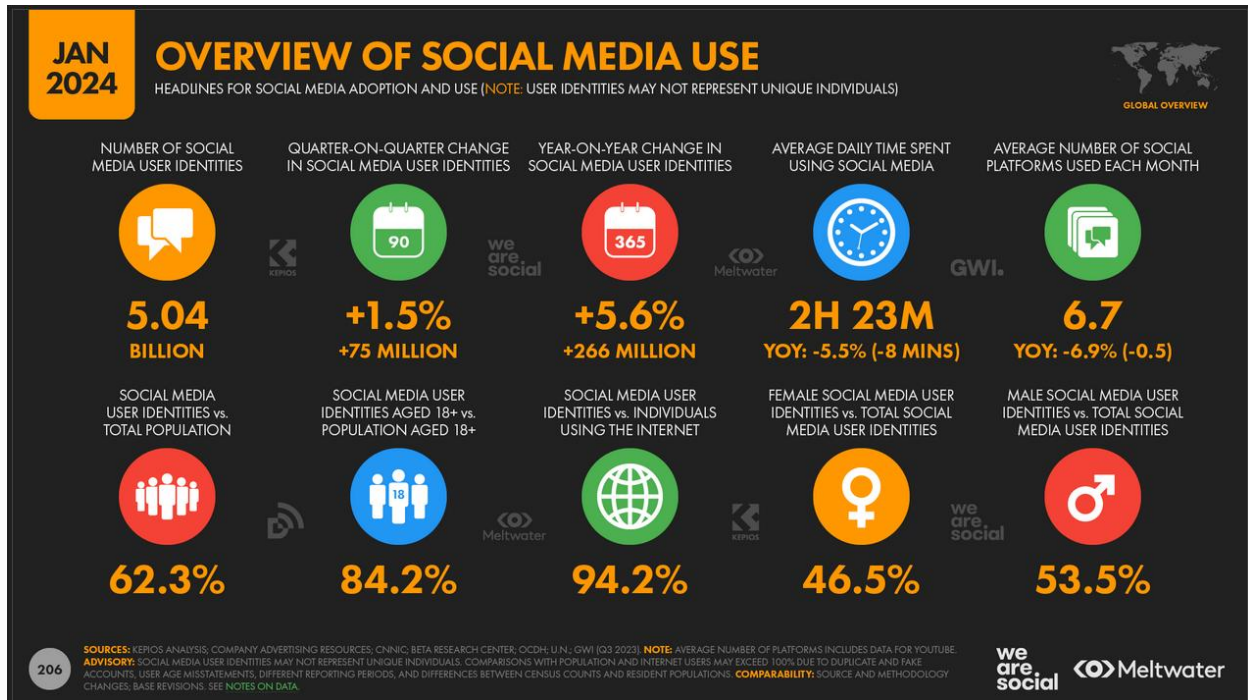


WHAT HAPPENS ONLINE EVERY MINUTE?



Learn more about Data Quality at
nodegraph.se

Sources: nodegraph.se, *internetlivestats.com, **iftw.com, ***unicomcorp.com



Neki opisi poslova sa Interneta

Data Scientist – Product Data Science @ Snap One

Required Qualifications

- Bachelor's Degree and 3+ years of work experience in data science
- Experience developing and monitoring consumer-facing AI
- Experience in applying frequentist or Bayesian statistics to test hypotheses and judge confidence levels
- Experience with Tensorflow or Pytorch
- Experience in Python, SQL, and Git
- Demonstrable experience making highly informational, easily interpreted visualizations
- Effective communication

Preferred Qualifications

- Intermediate Degree in a quantitative field
- Experience with signal process or time series analysis
- Experience in Databricks
- Experience in PowerBI
- Work experience in IoT or ML at the edge
- Experience with MLOps

Junior Data Analyst @ Cube Team

- Izrada i vizuelizacija izveštaja kroz BI alate
- Izrada poslovnih baza u skladu sa zahtevima
- Analiza i provera tačnosti isporučenih podataka
- Kontrola, klasifikacija i korigovanje ustanovljenih nepravilnosti u podacima
- Komunikacija sa biznis timom i prikupljanje zahteva za proširenje i unapređenje podataka
- Davanje detaljnih instrukcija inženjerskom timu za uspostavljanje novih ili modifikaciju postojećih procesa prikupljanja i obrade podataka
- Analiza i dokumentovanje specifikacija za prikupljanje i obradu podataka

Senior Data Analyst @ Microsoft

Qualifications

- Demonstrated experience in navigating complex datasets to come up with clean and actionable insights
- BS/MS in Computer Science, Statistics, Mathematics or related field experience in data analysis and reporting, business intelligence, or business and financial analysis.
- Excellent communication to be able to communicate insights to senior leaders.
- Programmatic Ads across Display, Native, and Video formats (Working with DSPs, SSPs, Ad Exchanges, etc.) experience preferred
- Have you acquired experience in database querying using SQL.
- Do you possess experience with data analytics tools like R, Python, Excel, Tableau, SAS, Power BI

2. Data Science - primene

Data Science je blisko povezan sa pojmom Industry 4.0, ali i više od toga!

U okviru pojma „[Industrija 4.0](#)“ tom području se pojavio niz novih izraza: *Smart factory, Cyber Physical Systems, Industrial Internet of Things*.

Spominju se svojstva tih sistema i njihovih podsistema: *Visualization, Digitalisation, Real time analysis, Identification, Realtime Location, Collaboration, Decentralisation, Autonomy, Agile systems, Big data, Sensors, Cloud computing, Virtual network, itd.*

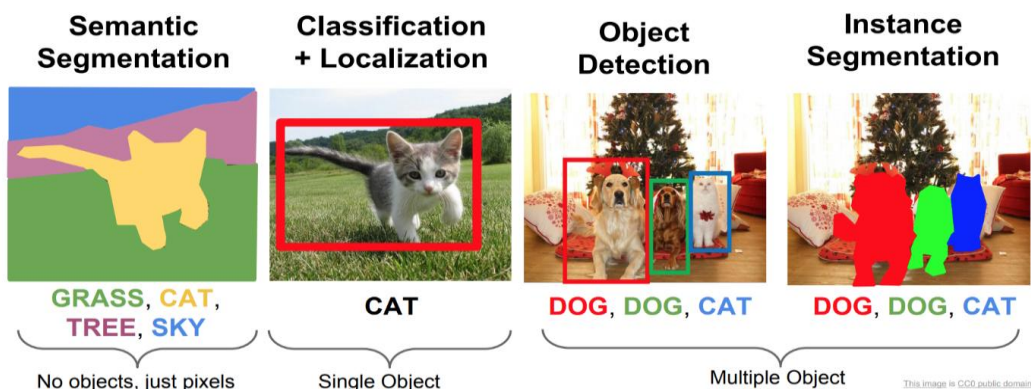
1. Detekcija prevara



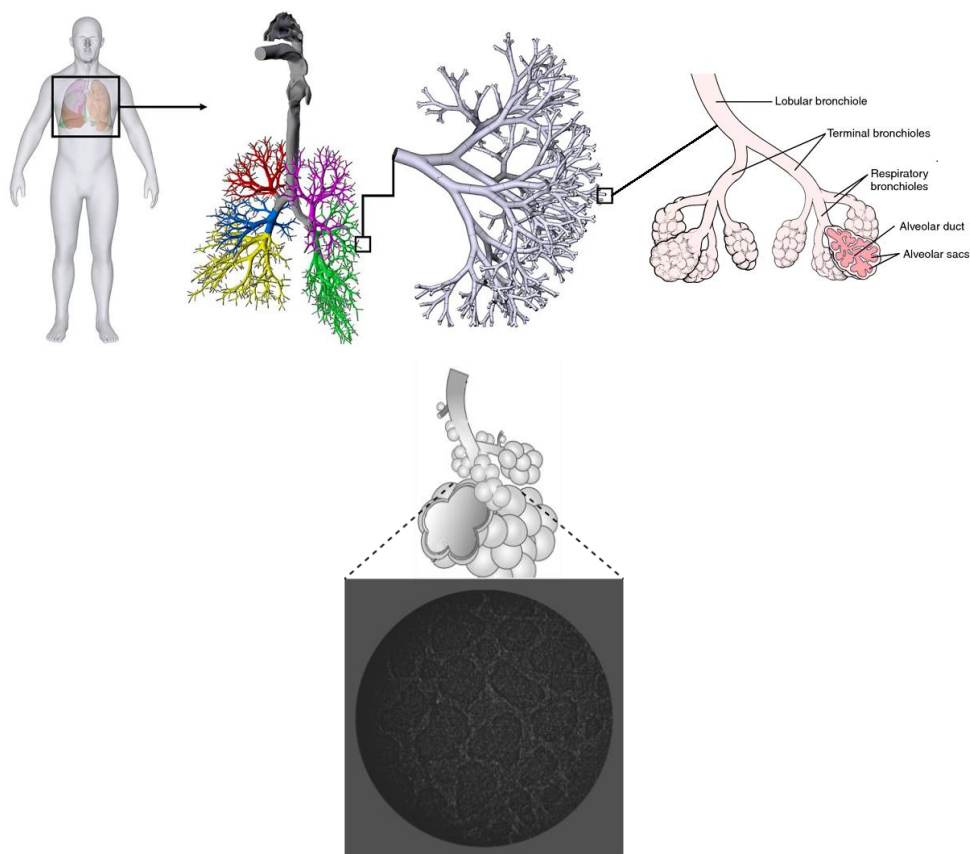
Credit Card Fraud Detection

Using the Machine Learning Classification Algorithms to detect Credit Card Fraudulent Activities

2. Obrada različitih vrsta slika i video snimaka (Everseen Beograd)



Slika ##. Zadaci u obradi slika.



Slika ##. Segmentacija plućnih alveola i karotidne arterije.

[Taxinomis projekt](#) (BioIRC).

3. Autonomna vozila: Tesla, Uber, [TTTech Auto](#), [Veoneer](#)...



3. Sistemi za preporuku (Recommender systems)



najk patike muske

Oko 19.000 rezultata (0,52 sekunde/m)

Ornac - <https://www.planetasport.rs/>

Patike - Obuća - Muškarci - PlanetaSport.rs

Najveći izbor brendova za sportski asortiman u Srbiji. Proverite i poručite online! Poručite online sve za sport — Naša cela Planeta je orijentisana na sport. Najbolje cene su onlajn. **Patike** i sportska oprema. Akcije i popusti. Isporuka širom Srbije.

♀ Kragujevac · 3 lokacije u blizini

Adidas Terrex za muškarce

Veliki izbor muških modela
Adidas Terrex cipela.

Adidas Terrex za decu

Dečiji modeli Adidas Terrex
cipela. Pogledaj ponudu!

Adidas Terrex za žene

Pogledajte ženske Adidas
Terrex cipele!

Dečije Air Max Command

Dečije patike Nike Air Max Command.
Pogledaj modele i poruči online!

Ornac - <https://www.n-sport.net/>

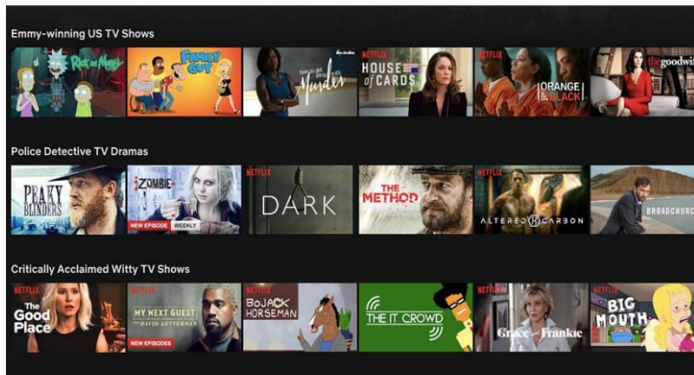
Nike patike - nova kolekcija - N Sport online prodavnica

Pogledajte raznovrsne modele **Nike patika** u našoj ponudi. Najpovoljnije cene. Najnovija ponuda i širok asortiman. Dostava je besplatna. Kupi online!

najk patike muske – Сликe

crvene polovne buzz cortez duboke air max sport vision plave



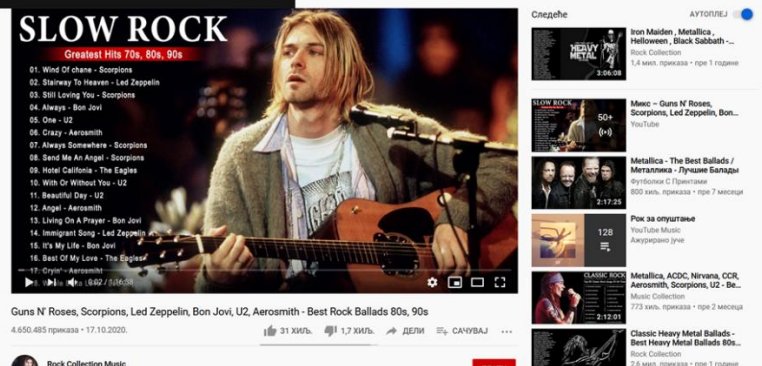


NETFLIX

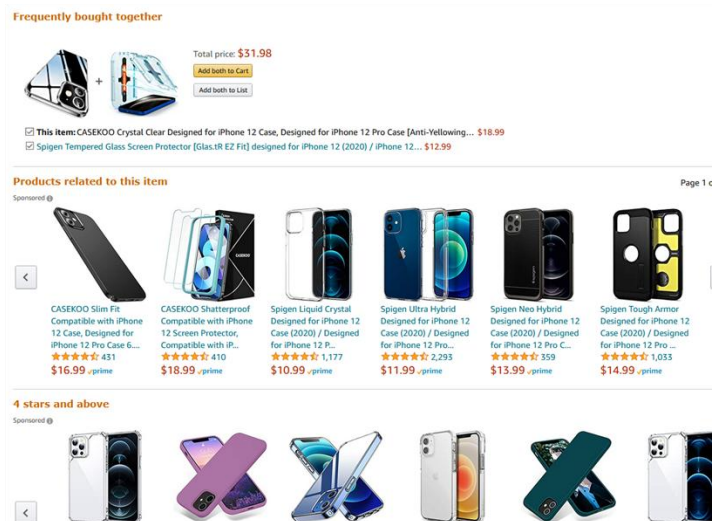
75% корисника филмове бира на основу њихове препоруке

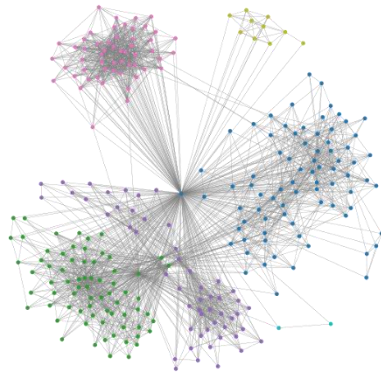
YouTube

P. Covington, J. Adams, E. Sargin, Deep Neural Networks for YouTube Recommendations. 2016

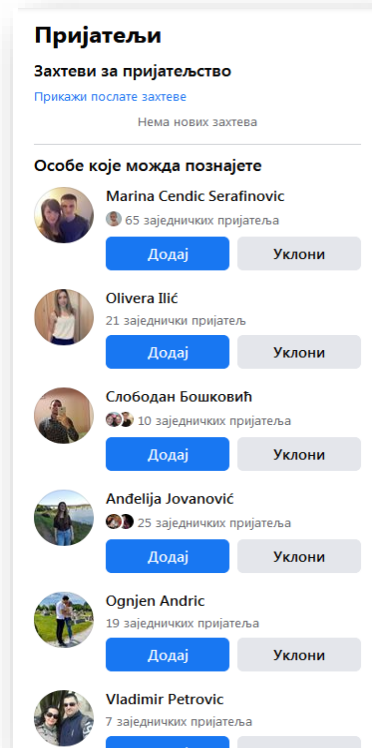
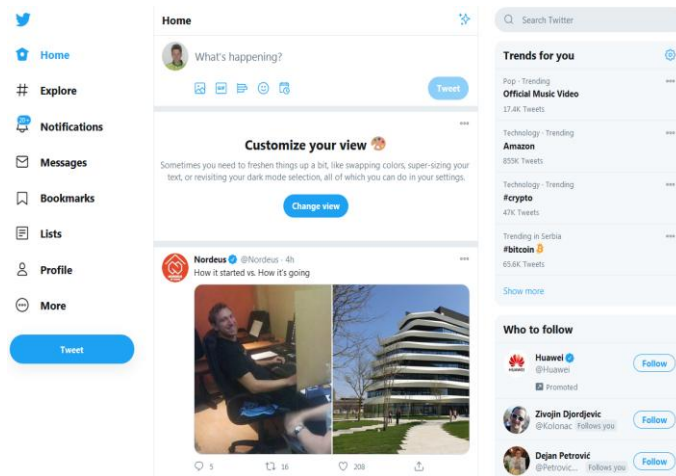


35% прихода приписује свом систему препоруку





Системи за препоруку који се базирају на графовима.



4. Gaming industrija

Nordeus, Logifuture RS Kragujevac

5. Pretražite neke interesantne pojmove: AI Startups / images

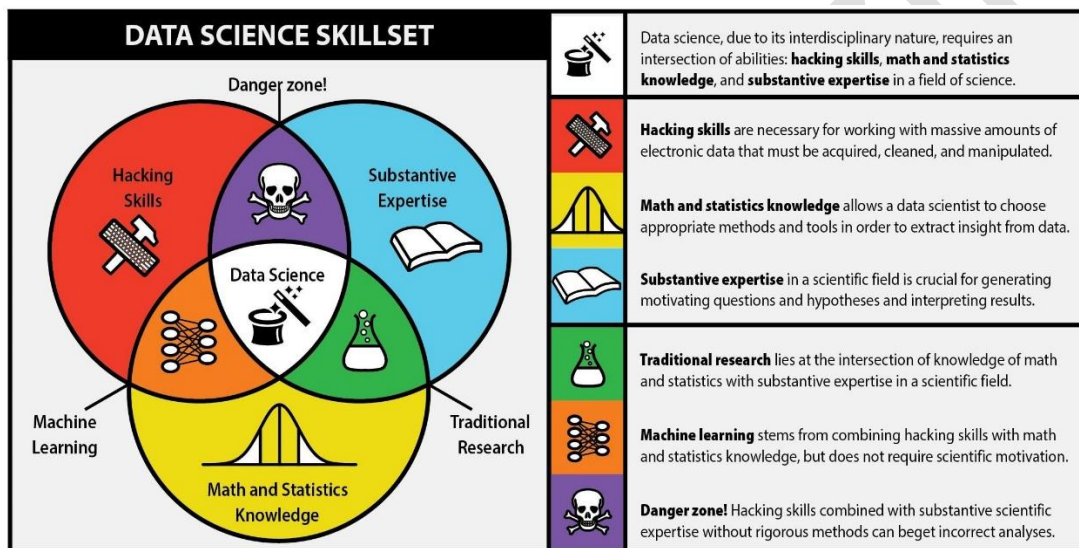
4. Data Scientist

Data Scientist je neko ko zna više statistiku od osobe koja se bavi računarima (computer scientist) i više zna o računarima od statističara.”

Josh Blumenstock

“Data Scientist = statističar + programer + trener + storyteller + umetnik”

Shlomo Aragon



Uloge u timu:

1. Data Architect

Arhitekta podataka se bavi upravljanjem podacima, što obuhvata dizajniranje, kreiranje, primenu i upravljanje arhitekturom podataka organizacije. Arhitekta podataka definišu kako će se podaci čuvati, koristiti, integrisati i upravljati od različitih entiteta podataka i IT sistema, kao i bilo koje aplikacije koje koriste ili obrađuju te podatke na neki način.

2. Data Engineer

Data Architect vs. Data Engineer – oni rade u tandemu! **Arhitekta podataka** pokušava da uvede „red u podatkovnom haosu“. Arhitekta podataka vizualizuje kompletan okvir i stvara nacrt koji inženjer podataka koristi za izgradnju „digitalnog okvira“. **Data inženjeri** pomažu arhitektama podataka da naprave radni okvir za pretraživanje i pronalaženje podataka, koji Data Scientists i analitičari mogu kasnije da koriste za svoj rad.

Pozicija za ljude koji imaju dobro razumevanje distribuiranog programiranja, infrastrukture i arhitekture. Oni rade na razvoju infrastrukture, tokova i skladištenja podataka. Dobro vladaju instalacijom distribuiranih sistema kao što su Hadoop MapReduce/Spark klasteri, znaju da kodiraju u programima kao što su Scala/Python/Java i znaju Unix skripting i SQL.

Neke od tehnologija koje se koriste:

- Big Data Frameworks - Spark/Hadoop
- MongoDB
- Integracija podataka
- Tableau
- Cloud
- Deployment rešenja

Big Data Specialist role

Data Engineer @ MaxBet

Our requirements:

- AWS Cloud Services: Use of key AWS services such as Amazon S3, Amazon RDS, Amazon Redshift, Amazon EMR, Amazon Kinesis, etc.
- Minimum 2 years of experience in Data Engineer position;
- Knowledge of AWS Lambda for serverless architectures.
- SQL and Relational Database: Advanced knowledge of SQL for data management. Experience in database modeling.
- Knowledge of Oracle SQL
- Data Warehousing: Understanding and experience working with data warehouses, such as Amazon Redshift.
- ETL (Extract, Transform, Load): Using tools for ETL processes such as AWS Glue.
- Programming languages: knowledge of Python and experience in Java or Scala programming languages for script and application development is an advantage.
- Tools: AWS Glue and a tool for ETL processes that enables automatic discovery, transformation, and loading of data.
- Amazon RDS (Relational Database Service): a service for managing and scaling relational databases, including SQL Server, PostgreSQL, and Oracle.
- DMS (Database Migration Service): a tool for data migration between different types of databases.
- API Gateway: Working with API Gateway to manage, monitor, and secure access to APIs, enabling integrated communication between different parts of the system.
- Kinesis: Amazon Kinesis is a group of AWS services for real-time processing and analysis of big data.
- Advantages: SQL Server Management Studio (SSMS): Use to manage, monitor, and maintain SQL Server databases

Senior Data Engineer @ Wireless Media

Required experience:

- University degree in engineering, mathematics, computer science or equivalent experience
- Commercial experience in a data-driven role
- Experience in data pipeline software engineering and implementing best practice in python - linting, unit tests, integration tests, git flow/pull request process, object- oriented development, data validation, algorithms and data structures, technical troubleshooting and debugging, bash scripting
- Experience preparing data for analytics and following a data science workflow
- Experience with analytics (descriptive, predictive, EDA), feature engineering and python visualization libraries – e.g. matplotlib, seaborn or other
- Comfortable with notebook and source code development – Jupyter, Pycharm/VScode
- Ability to write clean, maintainable and robust code in Python, SQL & Spark
- Experience on Big Data platforms and distributed computing, ideally Spark
- Knowledge of software engineering concepts and best practices
- Familiarity with the latest OSS, cloud, container (Docker), query languages and database technologies
- Confirmed experience building data pipelines in production and ability to work across structured, semi-structured and unstructured data
- Ability to communicate with both business/commercial and technical stakeholders
- Strong command of the English language (both verbal and written)

Technical skills:

- Python (Core analytical libraries: pandas, numpy, scikit-learn, scipy)
- Spark (PySpark)
- SQL
- Git
- Databricks / Azure
- Analytical pipelines (kedro / sklearn.pipeline)
- Apache Airflow / Kedro
- CI/CD (ideally with CircleCI)
- Docker and containers
- Models productionalization (mlflow)

3. Data Scientist

Posao obuhvata rad nad velikim skupovima podataka sa machine learning (ML) algoritmima, kako da razvijaju prediktivne modele, znaju teoriju (matematiku i statistiku) iza modela i znaju da interpretiraju i objasne ponašanje modela jednostavnim jezikom. Potrebno poznavanje R/Python i SQL-a.

4. Data Analyst

Da li ima posla u Srbiji za poziciju Data Scientist?

Interview Blog: [The 2021 Data Science Interview Report](#)