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

**ИНФОРМАТИКА**

Основне академске студије

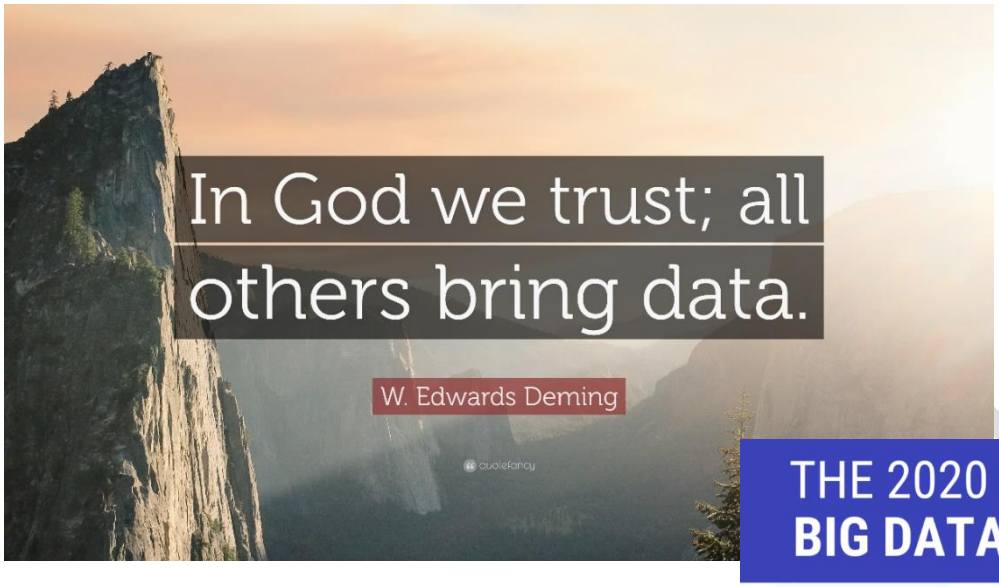
| |
|---|
| **Студијски програм:** Основне академске студије информатике |
| **Назив предмета: УВОД У НАУКУ О ПОДАЦИМА** |
| **Статус предмета:** Обавезни на модулима: Рачунарске науке и Информационо-комуникационе технологије, изборни на модулу Софтверско инжењерство |
| **Број ЕСПБ: 6** |
| **Услов:** Уписан одговарајући семестар |
| **Циљ предмета**<br>Стицање конкурентних и модерних знања о фундаменталним процесима, техникама и методама које чине науку о подацима и могућностима њихове практичне примене. |
| **Исход предмета**<br>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**<br>*Теоријска настава*<br>**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 Recency 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.<br>*Практична настава*<br>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**<br>1. Wickham, Hadley, and Garrett Grolemund, R za statističku obradu podataka, Mikro knjiga, 2017.<br>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.<br>3. Pathak, Manas A., Beginning Data Science with R. Springer, 2014.<br>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**<br>Проблемски-оријентисана настава, практична настава и вежбе уз софтверску подршку, самостални рад студената, домаћи задаци и консултације. |

**Оцена знања (максимални број поена 100)**

| **Предиспитне обавезе** | **70** поена | **Завршни испит** | **30** поена |
|---|---|---|---|
| колоквијуми | 20 + 20 | писмени испит | 20 |
| семинар | 30 | усмени испит | 10 |




## 1. Zašto je potrebna **Nauka o Podacima**?

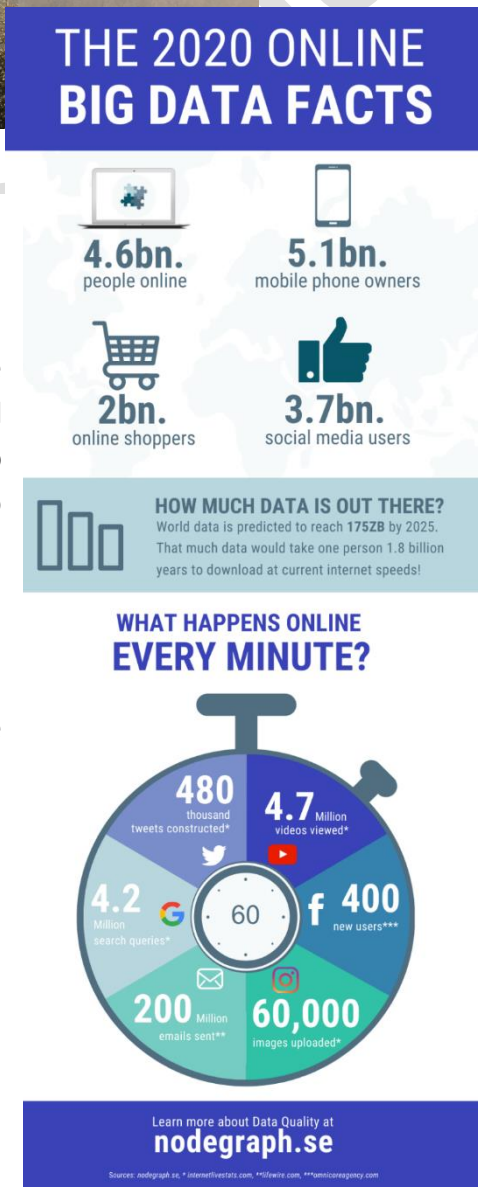
Let's start by looking if internet usage has changed since the end of 2017. By the end of 2019 there was around 4.6 billion internet users around the world, compared to 3.8 billion in June 2017. This is getting closer to two thirds of the world's population being online.

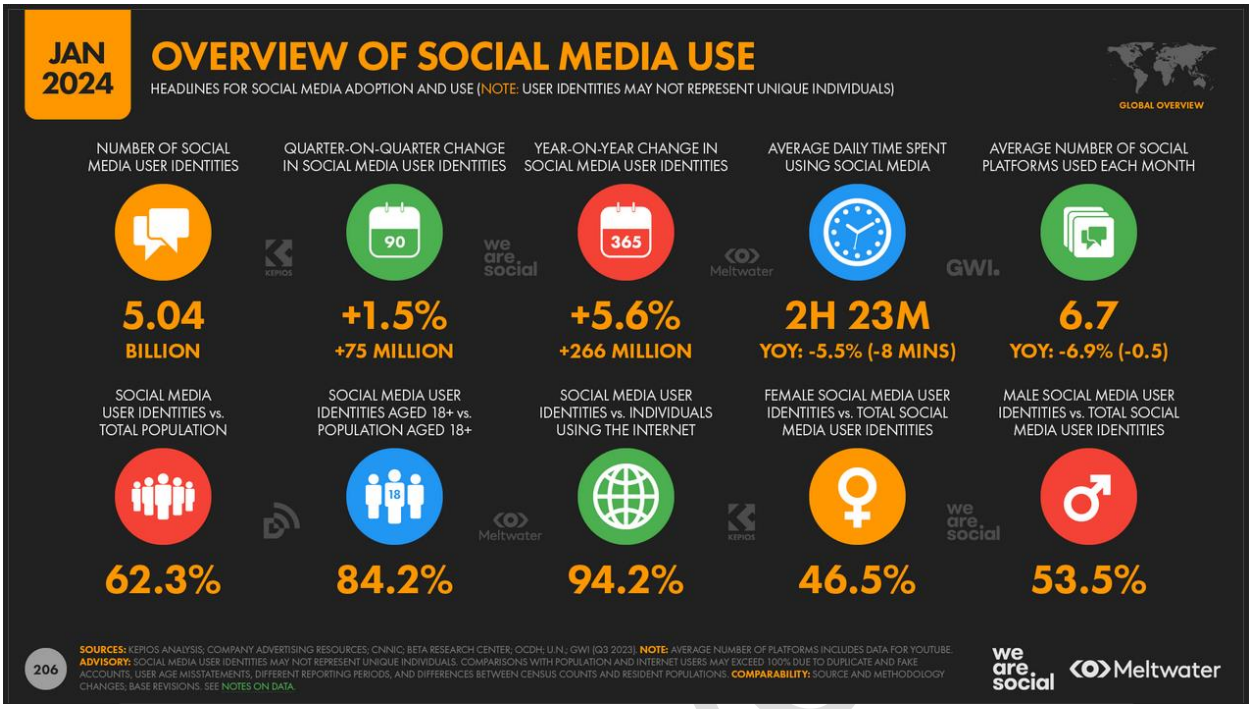
40 zettabytes of data—that's 40 trillion gigabytes.

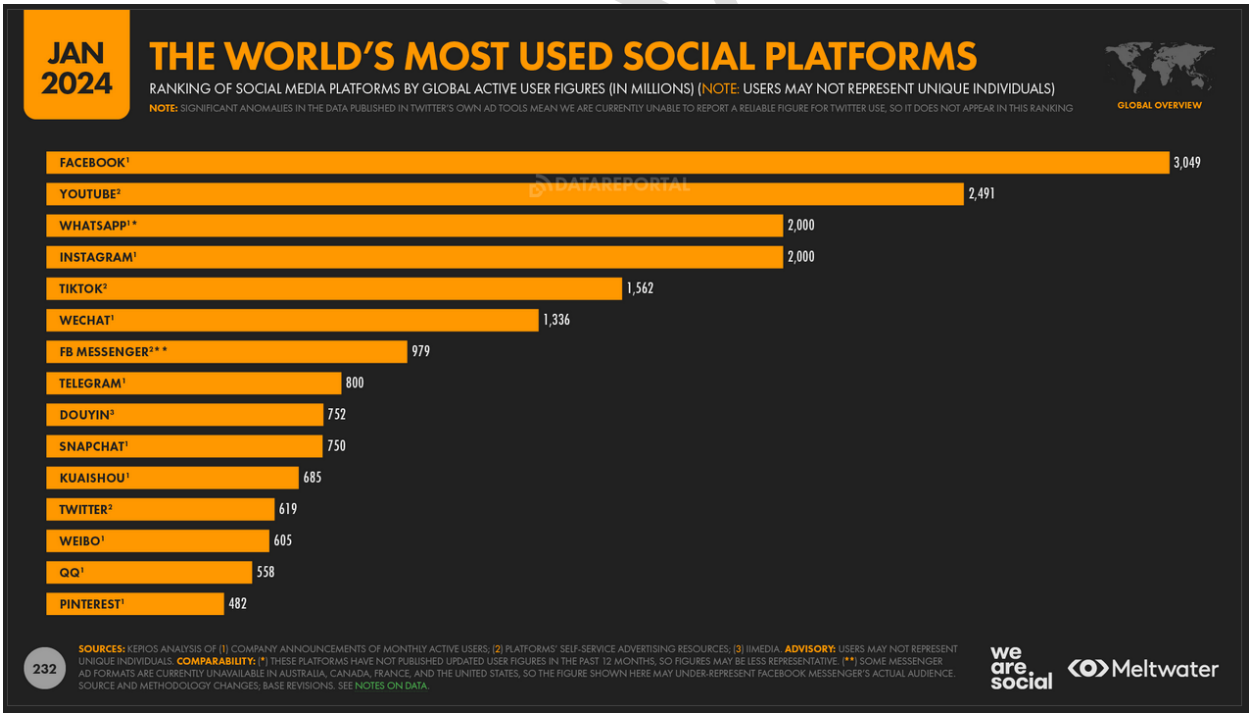
Više informacija o broju korisnika i podacima možete naći na sledećem linku.

Ko sve koristi podatke:

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




JAN 2024
OVERVIEW OF SOCIAL MEDIA USE
HEADLINES FOR SOCIAL MEDIA ADOPTION AND USE (NOTE: USER IDENTITIES MAY NOT REPRESENT UNIQUE INDIVIDUALS)
GLOBAL OVERVIEW
NUMBER OF SOCIAL MEDIA USER IDENTITIES
5.04
BILLION
QUARTER-ON-QUARTER CHANGE IN SOCIAL MEDIA USER IDENTITIES
+1.5%
+75 MILLION
YEAR-ON-YEAR CHANGE IN SOCIAL MEDIA USER IDENTITIES
+5.6%
+266 MILLION
AVERAGE DAILY TIME SPENT USING SOCIAL MEDIA
2H 23M
YOY: -5.5% (-8 MINS)
AVERAGE NUMBER OF SOCIAL PLATFORMS USED EACH MONTH
6.7
YOY: -6.9% (-0.5)
SOCIAL MEDIA USER IDENTITIES vs. TOTAL POPULATION
62.3%
SOCIAL MEDIA USER IDENTITIES AGED 18+ vs. POPULATION AGED 18+
84.2%
SOCIAL MEDIA USER IDENTITIES vs. INDIVIDUALS USING THE INTERNET
94.2%
FEMALE SOCIAL MEDIA USER IDENTITIES vs. TOTAL SOCIAL MEDIA USER IDENTITIES
46.5%
MALE SOCIAL MEDIA USER IDENTITIES vs. TOTAL SOCIAL MEDIA USER IDENTITIES
53.5%
206
SOURCES: KEPIOS ANALYSIS; COMPANY ADVERTISING RESOURCES; CNNIC; BETA RESEARCH CENTER; OCDH; U.N.; GWI (Q3 2023). NOTE: AVERAGE NUMBER OF PLATFORMS INCLUDES DATA FOR YOUTUBE. ADVISORY: SOCIAL MEDIA USER IDENTITIES MAY NOT REPRESENT UNIQUE INDIVIDUALS. COMPARISONS WITH POPULATION AND INTERNET USERS MAY EXCEED 100% DUE TO DUPLICATE AND FAKE ACCOUNTS, USER AGE MISSTATEMENTS, DIFFERENT REPORTING PERIODS, AND DIFFERENCES BETWEEN CENSUS COUNTS AND RESIDENT POPULATIONS. COMPARABILITY: SOURCE AND METHODOLOGY CHANGES; BASE REVISIONS. SEE NOTES ON DATA.


JAN 2024
THE WORLD'S MOST USED SOCIAL PLATFORMS
RANKING OF SOCIAL MEDIA PLATFORMS BY GLOBAL ACTIVE USER FIGURES (IN MILLIONS) (NOTE: USERS MAY NOT REPRESENT UNIQUE INDIVIDUALS)
NOTE: SIGNIFICANT ANOMALIES IN THE DATA PUBLISHED IN TWITTER'S OWN AD TOOLS MEAN WE ARE CURRENTLY UNABLE TO REPORT A RELIABLE FIGURE FOR TWITTER USE, SO IT DOES NOT APPEAR IN THIS RANKING
GLOBAL OVERVIEW
FACEBOOK[1] 3,049
YOUTUBE[2] 2,491
WHATSAPP[1]* 2,000
INSTAGRAM[1] 2,000
TIKTOK[2] 1,562
WECHAT[1] 1,336
FB MESSENGER[2]** 979
TELEGRAM[1] 800
DOUYIN[3] 752
SNAPCHAT[1] 750
KUAISHOU[1] 685
TWITTER[2] 619
WEIBO[1] 605
QQ[1] 558
PINTEREST[1] 482
232
SOURCES: KEPIOS ANALYSIS OF [1] COMPANY ANNOUNCEMENTS OF MONTHLY ACTIVE USERS; [2] PLATFORMS' SELF-SERVICE ADVERTISING RESOURCES; [3] IIMEDIA. ADVISORY: USERS MAY NOT REPRESENT UNIQUE INDIVIDUALS. COMPARABILITY: (*) THESE PLATFORMS HAVE NOT PUBLISHED UPDATED USER FIGURES IN THE PAST 12 MONTHS, SO FIGURES MAY BE LESS REPRESENTATIVE. (**) SOME MESSENGER AD FORMATS ARE CURRENTLY UNAVAILABLE IN AUSTRALIA, CANADA, FRANCE, AND THE UNITED STATES, SO THE FIGURE SHOWN HERE MAY UNDER-REPRESENT FACEBOOK MESSENGER'S ACTUAL AUDIENCE. SOURCE AND METHODOLOGY CHANGES; BASE REVISIONS. SEE NOTES ON DATA.

## 2. Nauka o podacima - definicija:

*Ne postoji opšte prihvaćena definicija pojma „Data Science“, ali hajde da pokušamo sa nekoliko.*

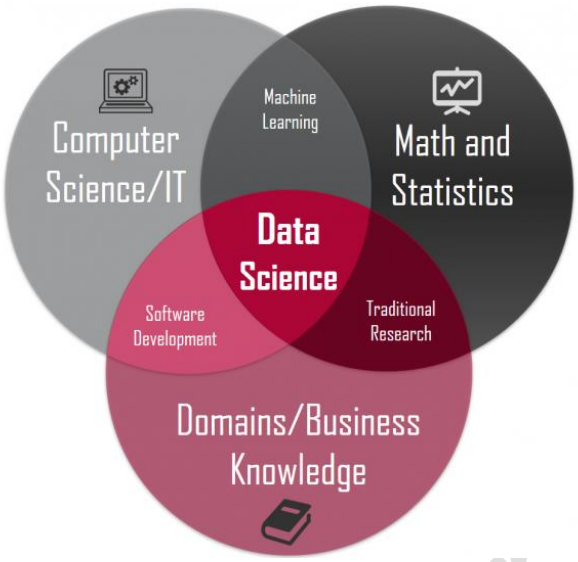
1. Nauka o podacima je oblast istraživanja koja kombinuje domensko znanje, programerske veštine, kao i znanje iz matematike i statistike za dobijanje znanja iz podataka. Data Scientists primenjuju algoritme mašinskog učenja na brojeve, tekst, slike, video simke, audio snimke i druge tipove podataka da bi razvili sisteme zasnovane na veštačkoj inteligenciji koji rešavaju zadatke koji obično zahtevaju ljudsku inteligenciju.

2. Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data. Data science is related to data mining, machine learning and big data.

   Data science is a "concept to unify statistics, data analysis and their related methods" in order to "understand and analyze actual phenomena" with data. It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, domain knowledge and information science. Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.

   Izvor: Wikipedia

3. Data science provides meaningful information based on large amounts of complex data or big data. Data science, or data-driven science, combines different fields of work in statistics and computation to interpret data for decision-making purposes.




Slika ##. Definicija Nauke o podacima kroz sliku.

## Data Science vs Data Analytics

Data Scientist stupa na scenu ranije od analitičara podataka. On istražuje velike skupove podataka, istražuje njihove potencijale, identifikuje trendove i značaj podataka, a zatim radi vizuelizaciju podataka i predstavlja ih drugima. Analitičar podataka podatke vidi nešto kasnije.

Analitičar podataka analizira određeni skup strukturiranih ili numeričkih podataka koristeći zadata pitanja. Verovatnije je da će Data Scientist da se bavi većim količinama strukturiranih i nestrukturiranih podataka. Oboje će takođe formulisati, testirati i proceniti učinak postavljenog pitanja za podatke u kontekstu sveukupne strategije.

Ciljevi:

- **Data Science** se bavi kreiranjem novih metoda za analizu, obradu i modeliranje podataka kako bi se otkrile nove informacije ili stvorila prediktivna moć. Cilj nauke o podacima je često da se generišu nove tehnike za obradu velikih i složenih skupova podataka.
- **Data Analytics** se više fokusira na procesiranje i analizu postojećih podataka kako bi se došlo do korisnih uvida i rešenja za konkretna poslovna pitanja. Cilj analitike podataka je da se nađu odgovori na specifična pitanja i da se donesu odluke na osnovu podataka.

## **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

...

## Data Science vs Statistics

Nauka o podacima je široka, interdisciplinarna oblast koje spaja primenjeno poslovno upravljanje (business management), računarstvo, ekonomiju, matematiku, programiranje i softverski inženjering zajedno sa statistikom. Izazovi u nauci podataka zahtevaju prikupljanje, obradu, upravljanje, analizu i vizuelizaciju velikih količina podataka, a Data scientist koristi alate iz različitih oblasti, uključujući i statistiku, da bi postigli te ciljeve.

## Data Mining vs Data Science

Istraživanje podataka (engl. Data Mining, traženje podataka, prekopavanje podataka, rudarenje podataka, dejta majning) je proces otkrivanja šablona u velikim skupovima podataka, korišćenjem metoda mašinskog učenja, statistike i sistema baza podataka. Cilj Data mining-a je obrada podataka sa specifičnim poslovnim ciljem. Nasuprot tome, nauka o podacima ima za cilj stvaranje proizvoda i rezultata zasnovanih na podacima - obično u poslovnom kontekstu. LINK

**Ciljevi i Obim:**

- **Data Mining** se fokusira na ekstrakciju korisnih informacija i uzoraka iz velikih skupova podataka. Cilj rudarenja podataka je identifikovanje značajnih obrazaca, veza, anomalija i trendova unutar podataka.
- **Data Science** je širi koncept koji obuhvata Data Mining, ali takođe uključuje razvoj novih algoritama, statističkih metoda, i prediktivnih modela. Nauka o podacima koristi nalaze iz rudarenja podataka kao jedan od svojih resursa, ali ide dalje u stvaranju novih metoda za analizu i interpretaciju podataka.

**Metode**:

- **Data Mining** koristi tehničke pristupe kao što su klasterovanje, klasifikacija, regresija i asocijativna pravila da bi otkrio obrasce u podacima.
- **Data Science** kombinuje ove tehnike sa naprednim algoritmima, mašinskim učenjem, dubokim učenjem i drugim statističkim metodama da bi se ne samo identifikovali obrasci, već i napravili prediktivni modeli i novi načini analize.

## Data Science vs Artificial Intelligence

**Prostor:** Veštačka inteligencija ograničena je samo na primenu algoritama ML, dok nauka o podacima uključuje različite osnovne operacije nad podacima.

**Tip podataka:** Veštačka inteligencija radi sa podacima koji su standardizovani u obliku vektora, dok sa druge strane, Nauka o podacima će imati mnogo različitih vrsta podataka kao što su strukturirani, polustrukturirani i nestrukturirani podaci.

**Svrha:** Primarna svrha veštačke inteligencije je automatizacija procesa i uvođenje autonomije u model podataka. Primarni cilj Data Science-a je pronaći obrasce koji su skriveni u podacima.

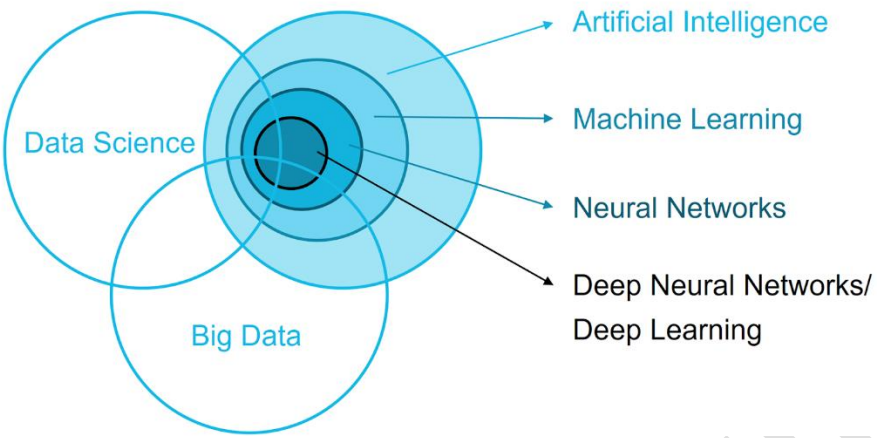
**Tehnike:** Veštačka inteligencija koristi algoritme u računarima za rešavanje problema, dok nauka o podacima uključuje mnogo različitih statističkih metoda.

**Različiti modeli:** U veštačkoj inteligenciji se grade modeli za koje se očekuje da budu slični ljudskom razumevanju i saznanju. U nauci o podacima, modeli su konstruisani tako da daju statističke uvide za donošenje odluka.

## Data Science vs Machine Learning

Budući da je nauka o podacima širok pojam za više disciplina, mašinsko učenje je deo nauke o podacima. Mašinsko učenje koristi razne tehnike, poput regresije i klasifikacije itd. Sa druge strane, podaci koji se koriste u Data Science-u mogu ili ne moraju evoluirati iz mašine ili mehaničkog procesa. Glavna razlika između ove dve oblasti je u tome što se nauka o podacima kao širi pojam ne fokusira samo na algoritme i statistiku, već se brine i o celokupnoj metodologiji.

Nauka o podacima može se videti kao povezivanje više nad-disciplina, uključujući analitiku podataka, softverski inženjering, inženjering podataka, mašinsko učenje, prediktivnu analitiku, analitiku podataka i još mnogo toga. Obuhvata pronalaženje, prikupljanje, unošenje i transformaciju velike količine podataka (Big Data). Nauka o podacima odgovorna je za približavanje strukture velikim podacima, traženje ubedljivih obrazaca i savetovanje donosilaca odluka da efikasno uvedu promene u skladu sa poslovnim potrebama. Analitika podataka i mašinsko učenje dva su od mnogih alata i procesa koje nauka podataka koristi.



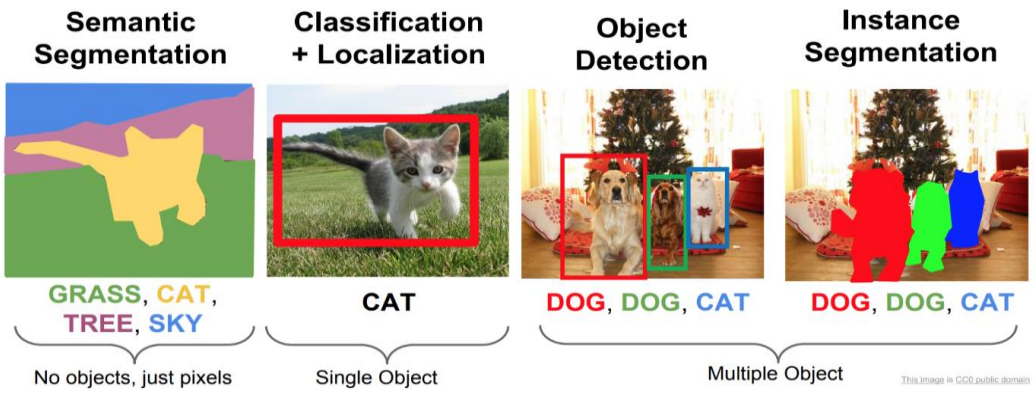

## 3. Data Science - primene

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

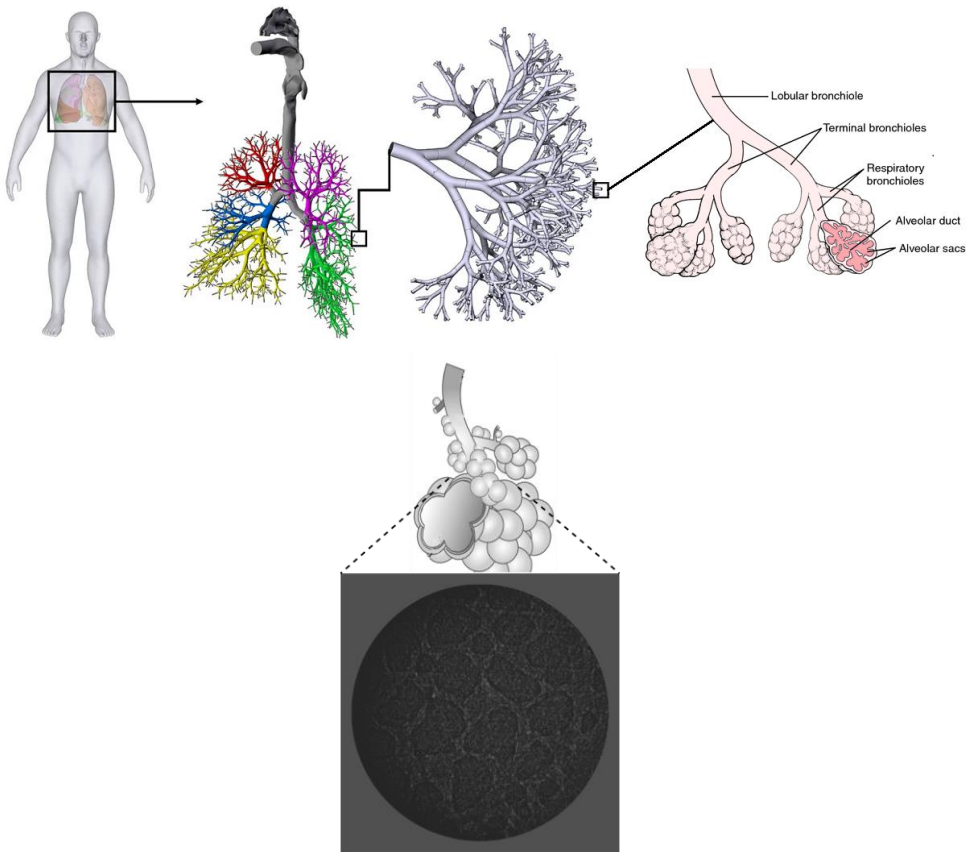
U okviru pojma „**Industrija 4.0**“ tom području se pojavio niz novih izraza: *Smart factory, Cyber Phisical Systems, Industrial Internet of Things*.
Spominju se svojstva tih sistema i njihovih podsistema: *Visualization, Digitalisation, Real time analysis, Identification, Realtime Location, Collaboration, Decentralisation, Autonomy, Agile systems, Big data, Sensors, Cloud computing, Virtual network, itd*.

1. **Detekcija prevara**

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



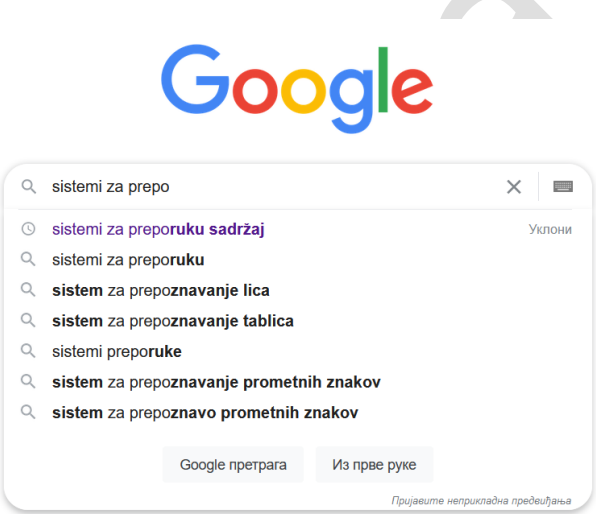

Slika ##. Zadaci u obradi slika.

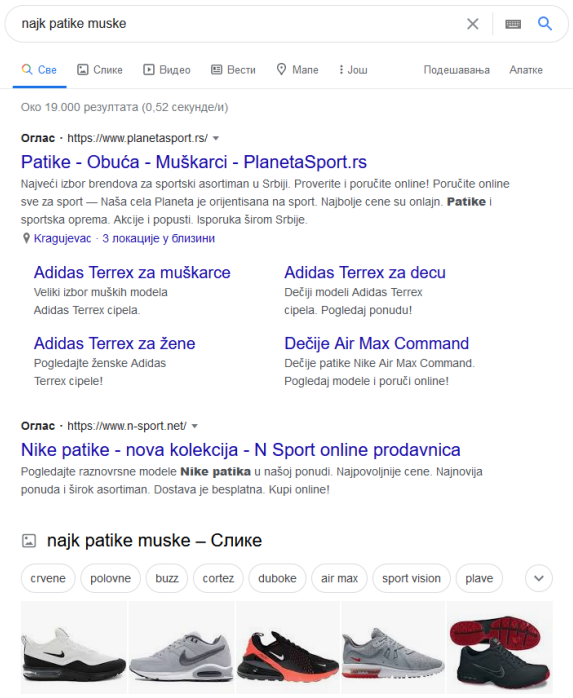


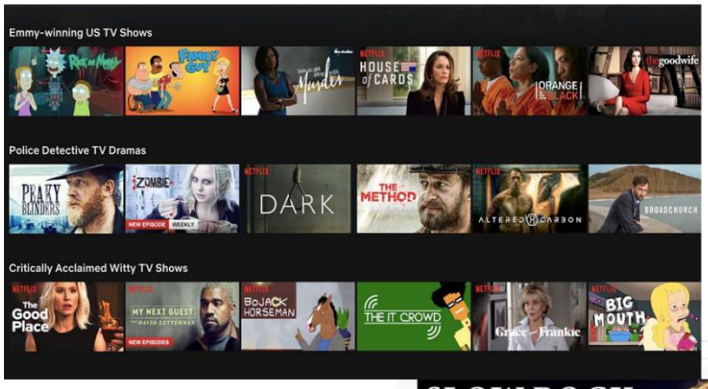

Slika ##. Segmentacija plućnih alveola i karotidne arterije.
Taxinomisis projekat (BioIRC).

3. **Autonomna vozila:** Tesla, Uber, TTTech Auto, Veoneer...


4. **Sistemi za preporuku (Recommender systems)**

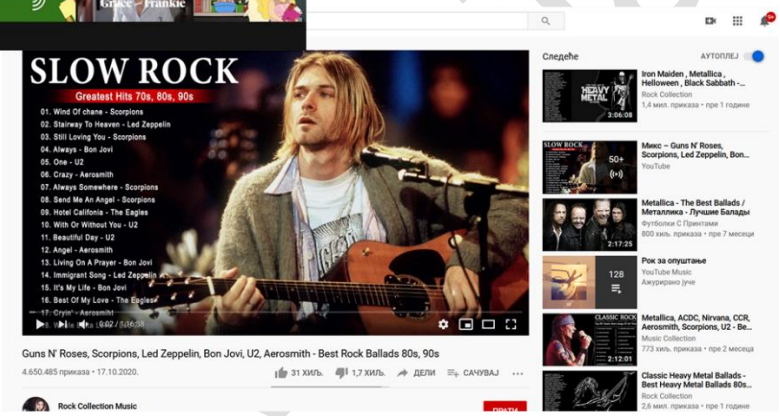





NETFLIX

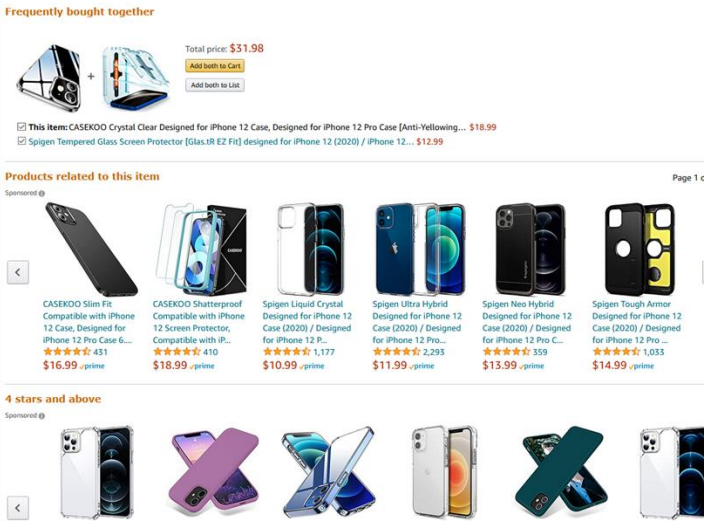
**75%** корисника филмове бира на основу њихове препоруке

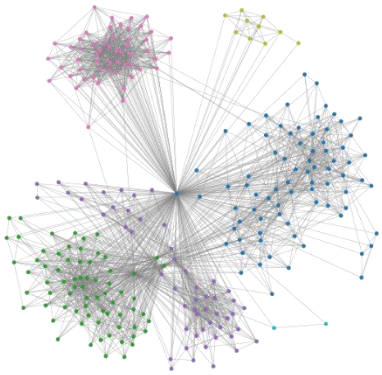



YouTube

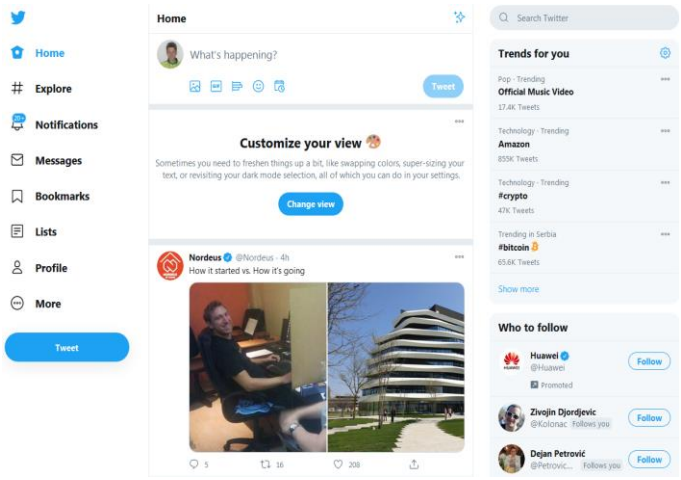
P. Covington, J. Adams, E. Sargin, Deep Neural Networks for YouTube Recommendations. 2016

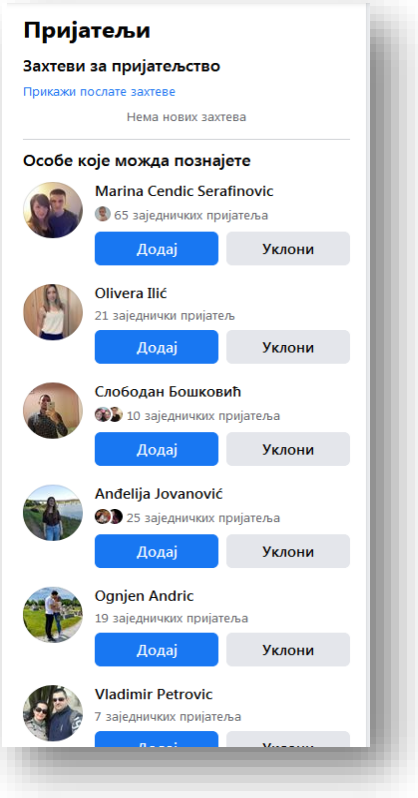
**35%** прихода приписује свом систему препоруку





Системи за препоруку који се базирају на графовима.




5. **Gaming industrija**

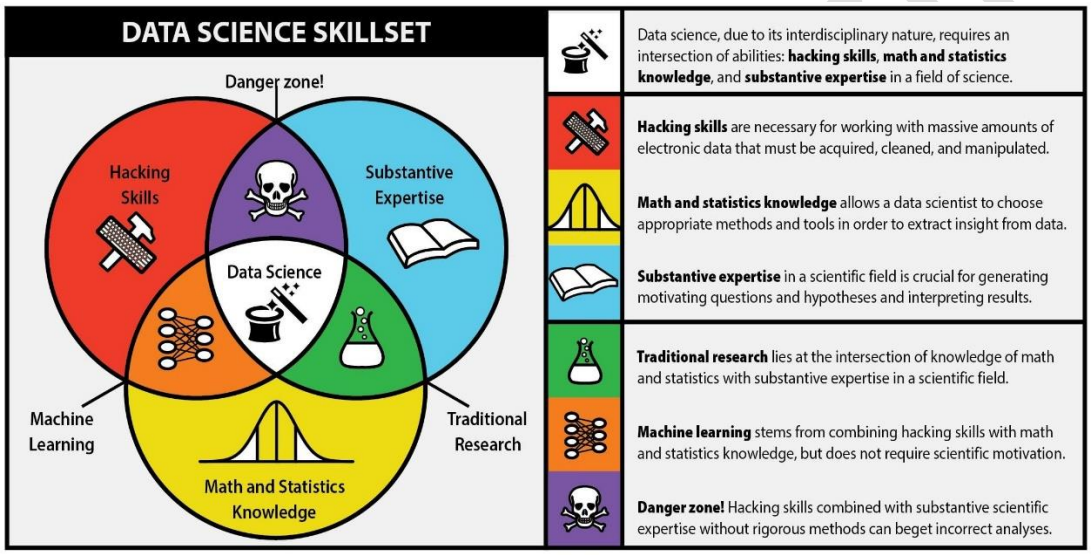
   Nordeus, Logifuture RS Kragujevac

6. **Pretražite neke interesantne pojmove:** AI Startups / images

**DOMAĆI:** Pronađite 3 startapa koji koriste algoritme mašinskog učenja i veštačku inteligenciju za poboljšanje svojih proizvoda.

# 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.”<br>Josh Blumenstock | “Data Scientist = statističar + programer + trener + storyteller + umetnik”<br>Shlomo Aragmon |




# Uloge u timu:

## 1. Data Architect

Arhitekta podataka se bavi upravljanjem podacima, što obuhvata dizajniranje, kreiranje, primenu i upravljanje arhitekturom podataka organizacije. Arhitekte 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
- Inetegracija 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
- Databriscks / Azure
- Analytical pipelines (kedro / sklean.pipeline)
- Apache Airflow / Kedro
- CI/CD (ideally with CircelCI)
- Docker and containers
- Models productionalization (mlflow)

## 3. Data Scientist

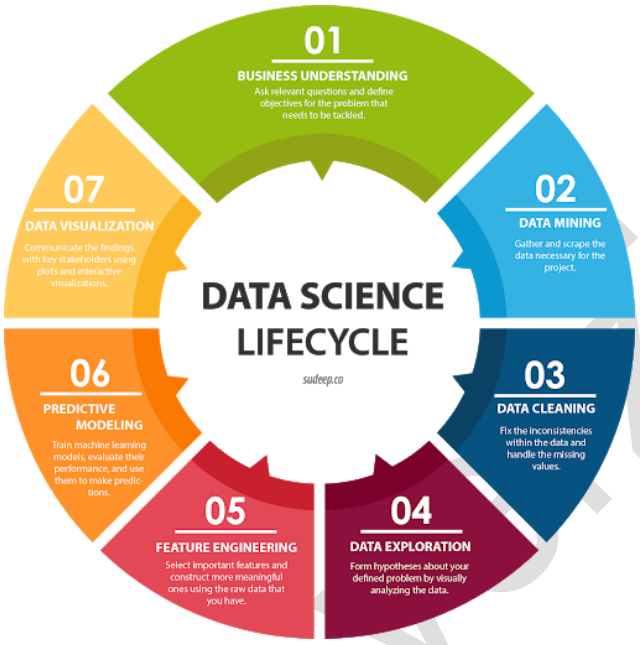
Posao obuhvata rad nad velikim skupovima podataka sa machine learning (ML) algoritmima, kako da razvijaju prediktivne modele, znaju teoriju (matematiku i statistiku) iza modela i znaju da interpretiraju i objasne ponašanje modela jednostavnim jezikom. Potrebno poznavanje R/Python i SQL-a.

## 4. Data Analyst

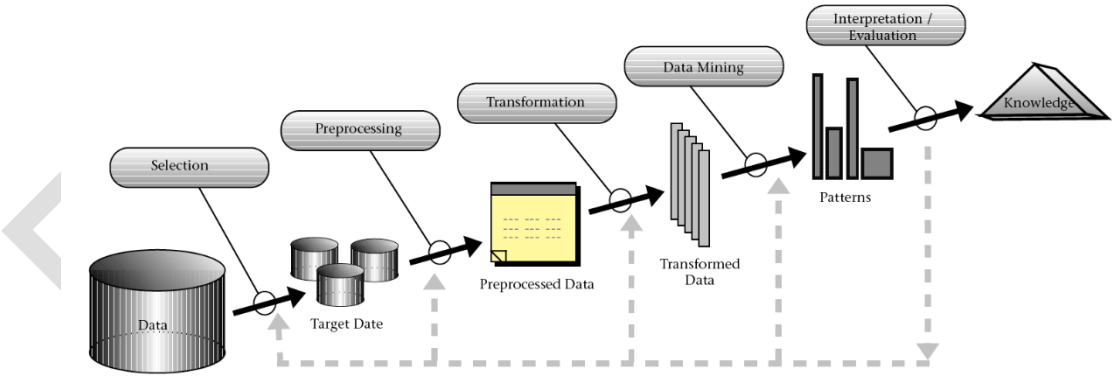
Da li ima posla u Srbiji za poziciju Data Scientist?

**Interview Blog:** **The 2021 Data Science Interview Report**

## 5. Kako izgleda jedan Data Science projekat




## 6. Šta ćemo sve da naučimo?




Rad sa podacima, transformacije, vizuelizacija...
Exploatory Data Analysis
Statističko učenje
I još mnogo toga...

**Literatura:**


- T. Hastie, R. Tibshirani, J. Friedman: The Elements of Statistical Learning
- C. Bishop: Pattern Recognition and Machine Learning
- K. Murphey: Machine Learning: A Probabilistic Perspective
- I. Goodfellow, Y. Bengio, A. Courville: Deep Learning
- R. Sutton, A. Barto: Reinforcement Learning: An Introduction
- S. Shalev-Schwartz, S. Ben-David: Understanding Machine Learning: From Theory to Algorithms