

Модули:

Безбедност друмског саобраћаја

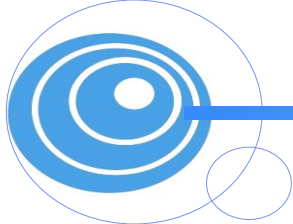
и

Менаџмент и економија у транспорту и комуникацијама

Софтверски алати Big Data аналитике

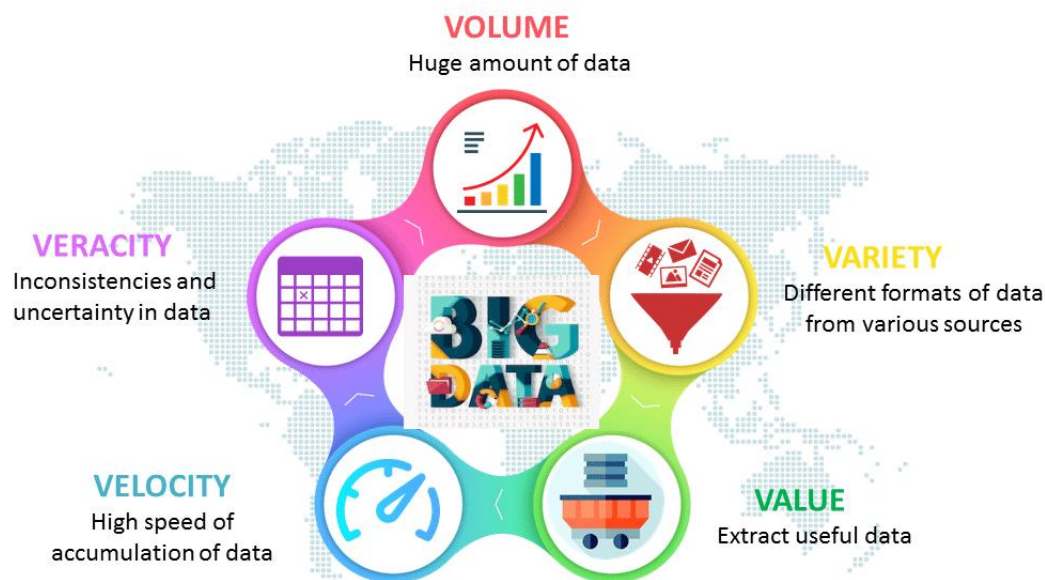
Наставници (*e-mail*): др Слађана Јанковић, ванредни професор (*s.jankovic@sf.bg.ac.rs*)

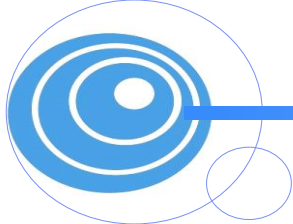
др Ана Узелац, доцент (*ana.uzelac@sf.bg.ac.rs*)



Појам Big Data

- Под појмом Big Data подразумева се информациони ресурс велике количине, велике брзине увећавања и велике разноврсности података, који превазилази могућности уобичајено коришћеног софтвера за складиштење, обраду и управљање подацима.



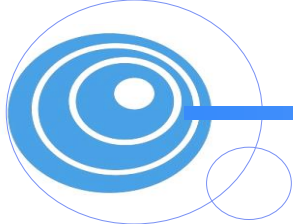


Зашто Big Data технологије?

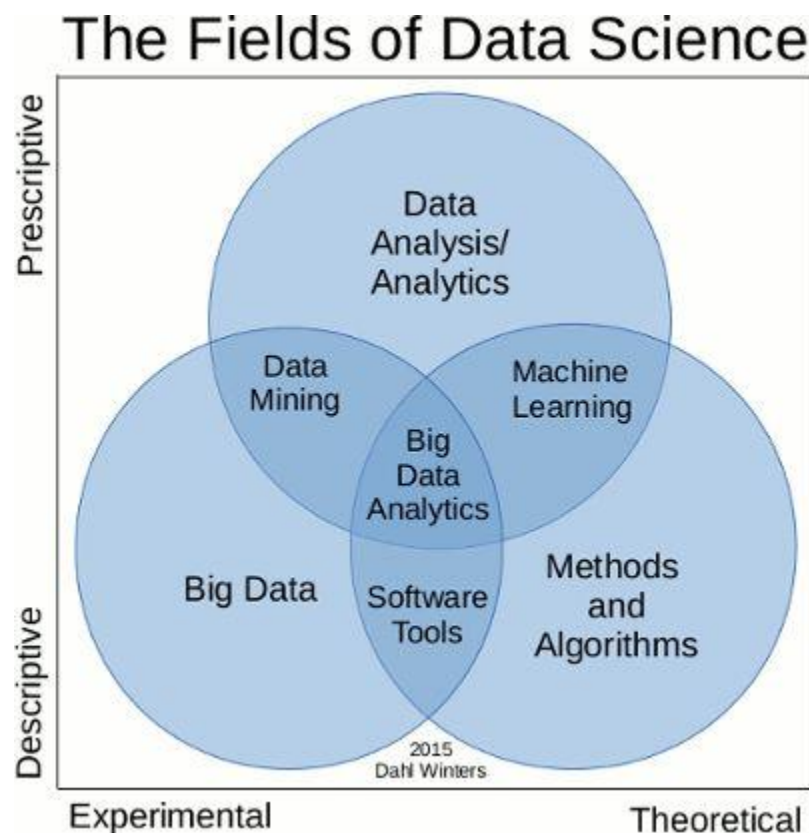
- Са експлозијом сензора, паметних уређаја, технологија друштвеног умрежавања, подаци у једној организацији постају све комплекснији.
- Осим традиционалних релационих база података, све више су присутни сирови подаци, полуструктурирани и неструктурирани подаци са web страница, web log фајлови (укључујући click-stream податке), search indexes, social media forums, e-mail, документи, сензорски подаци са активних и пасивних система, и др.
- Процењује се да је свега око 20 процената података који се данас складиште у свету структурирано.

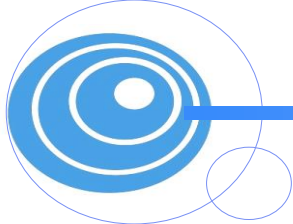


- Док количина података које организације складиште расте, проценат података које су организације способне да анализирају и употребе у свом пословању – опада! Та чињеница пружила је шансу Big Data технологијама.

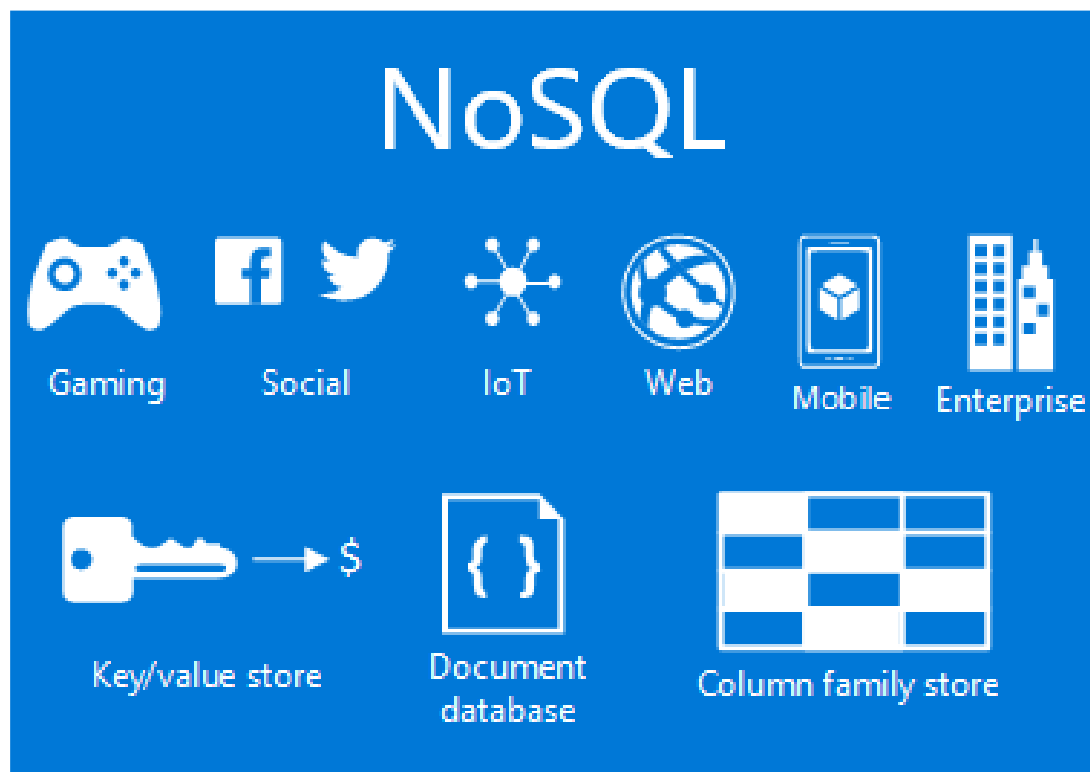


Big Data аналитика

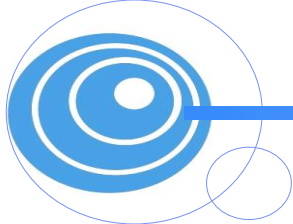




Садржај предмета



Нерелационе базе података (NoSQL)



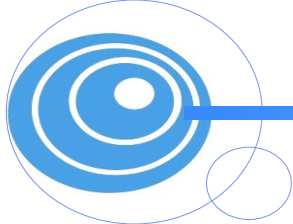
Садржај предмета



- Упознавање са *Hadoop* платформом за складиштење и обраду Big Data скупова података

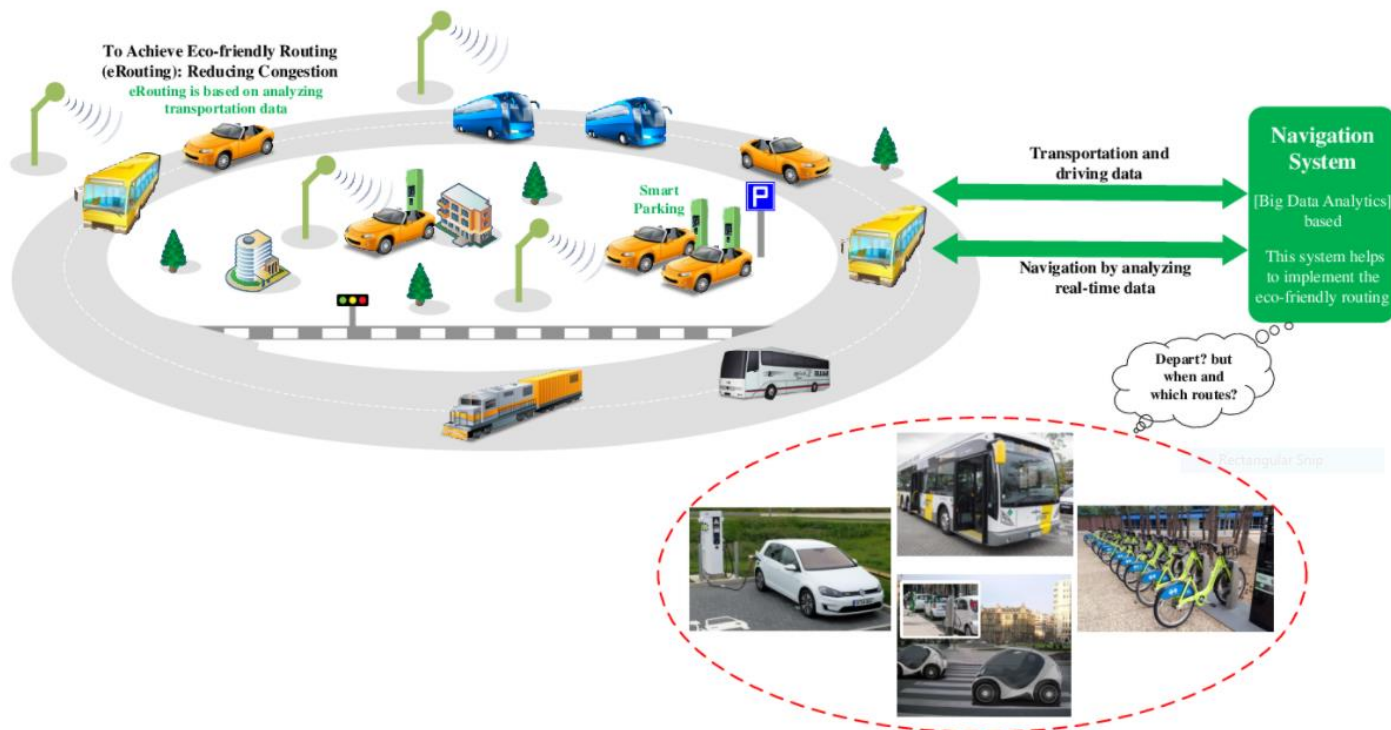


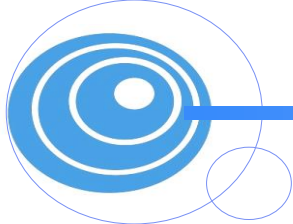
- Овладавање нерелационим системом за управљање базама података *MongoDB*



Садржај предмета

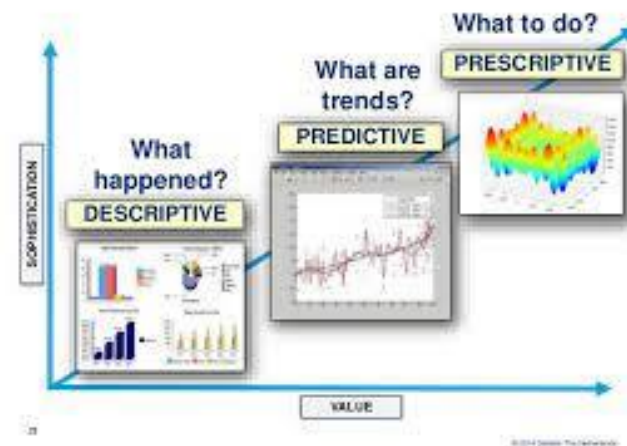
- Улога Big Data аналитике у саобраћају





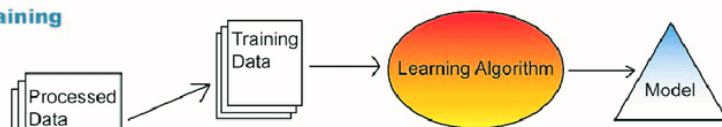
Садржај предмета

- Методе Big Data аналитике
- Дескриптивна анализа Big Data скупова података
- Предиктивна анализа базирана на моделима надгледаног машинског учења (алгоритми класификације и регресије)



Supervised Learning

A) Training

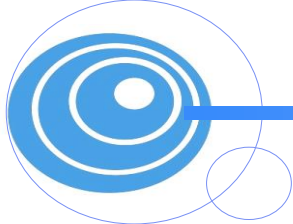


B) Validation



C) Prediction

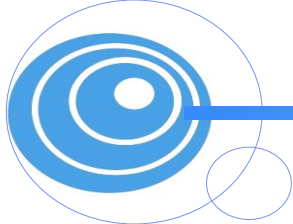




Садржај предмета

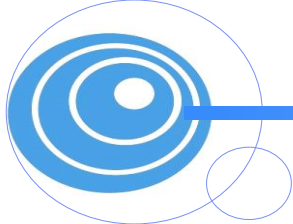


- Овладавање *data mining* софтверским алатом *Weka*.
- Коришћење софтверског алата *Weka* за изградњу и примену модела надгледаног и ненадгледаног машинског учења.
- Основне структуре података у *Python*-у.
- Рад са библиотекама у *Python*-у које служе за процесирање података.
- Рад са библиотекама у *Python*-у које служе за визуелизацију података.



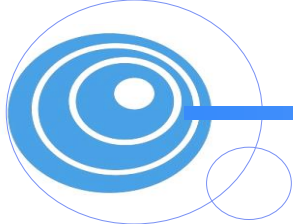
Циљ предмета

- Овладавање најпопуларнијим софтверским алатима који се користе за складиштење и обраду података који имају Big Data обележја:
 - велике количине,
 - изражену хетерогеност и
 - велику брзину пристизања.
- Овладавање изабраним софтверским алатима који се користе у Big Data аналитици.



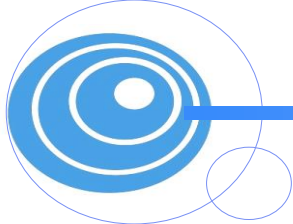
Циљ предмета

- Студенти треба да се обуче да:
 - препознају скупове података који имају Big Data обележја;
 - препознају реалне проблеме из саобраћаја и транспорта који могу бити решени коришћењем нерелационих база података и метода Big Data аналитике;
 - пројектују и креирају нерелационе базе података на две популарне платформе: *Hadoop* и *MongoDB*;
 - врше предиктивну анализу применом технике машинског учења у софтверском алату *Weka*;
 - раде са библиотекама у *Python*-у које служе за процесирање и визуелизацију података.

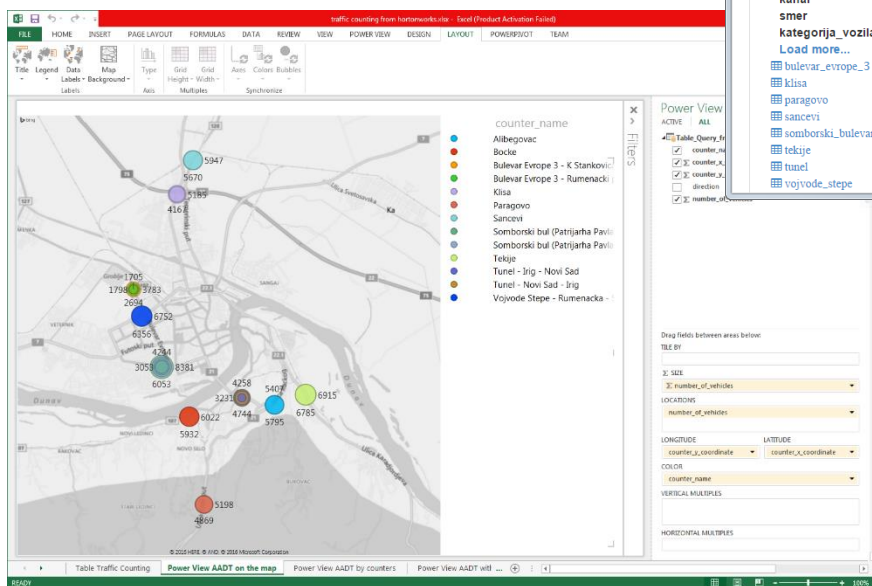


Провера знања

- **Пројектни задатак (обавезан)** – самосталан рад студента уз консултације са предметним наставницима: **60 бодова**. Пројектни задатак може да се реализује на један од следећа два начина:
 - креирање нерелационе базе података и упита над њом на **Hadoop** или **MongoDB** платформи по избору студента, или
 - предиктивна анализа применом технике машинског учења у софтверском алату **Weka**.
- **Завршни испит (није обавезан)** – семинарски рад или публиковани научно-стручни рад студента у сарадњи са предметним наставницима: **40 бодова**. Завршни испит обухвата: обраду, анализу и визуелизацију података коришћењем програмског језика **Python**.



Пример нерелационе базе података у области друмског саобраћаја креиране на *Hadoop* платформи, упита над њом и визуелизације података из базе



Ambari - Sandbox

localhost:8080/#/main/views/HIVE/1.0.0/Hive

Ambari Sandbox 0 ops 2 alerts

Dashboard Services Hosts 1 Alerts Admin

Hive Query Saved Queries History UDFs

Database Explorer

brojanje_saobracao

Search tables...

Databases

- brojanje_saobracao
 - alibegovac
 - bocke
 - brojac_id STRING
 - naziv_brojaca STRING
 - x_koordinata STRING
 - y_koordinata STRING
 - redni_broj_vozila STRING
 - datum STRING
 - vreme STRING
 - kanal STRING
 - smer STRING
 - kategorija_vozila STRING
 - Lead more...
 - bulevar_evrope_3
 - kisa
 - paragovo
 - sancevi
 - somborski_bulevar
 - tekije
 - tunel
 - vojvode_stepe

Query Editor

Worksheet

```
1 SELECT naziv_brojaca, x_koordinata, y_koordinata, smer,  
2 ROUND(AVG(brzina_vozila), 2) AS PROSEČNA_BRZINA, MAX(brzina_vozila) AS MAX_BRZINA,  
3 MIN(brzina_vozila) AS MIN_BRZINA, MAX(brzina_vozila)-MIN(brzina_vozila) AS RAZLIKA_MAX_I_MIN_BRZINA  
4 FROM bocke  
5 WHERE naziv_brojaca="Bocke" AND brzina_vozila > 10  
6 GROUP BY naziv_brojaca, x_koordinata, y_koordinata, smer;
```

Stop execution Explain Save as... Kill Session New Worksheet

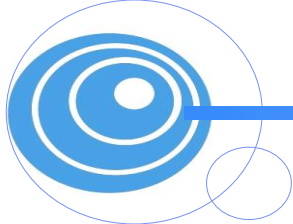
9%

Query Process Results (Status: RUNNING)

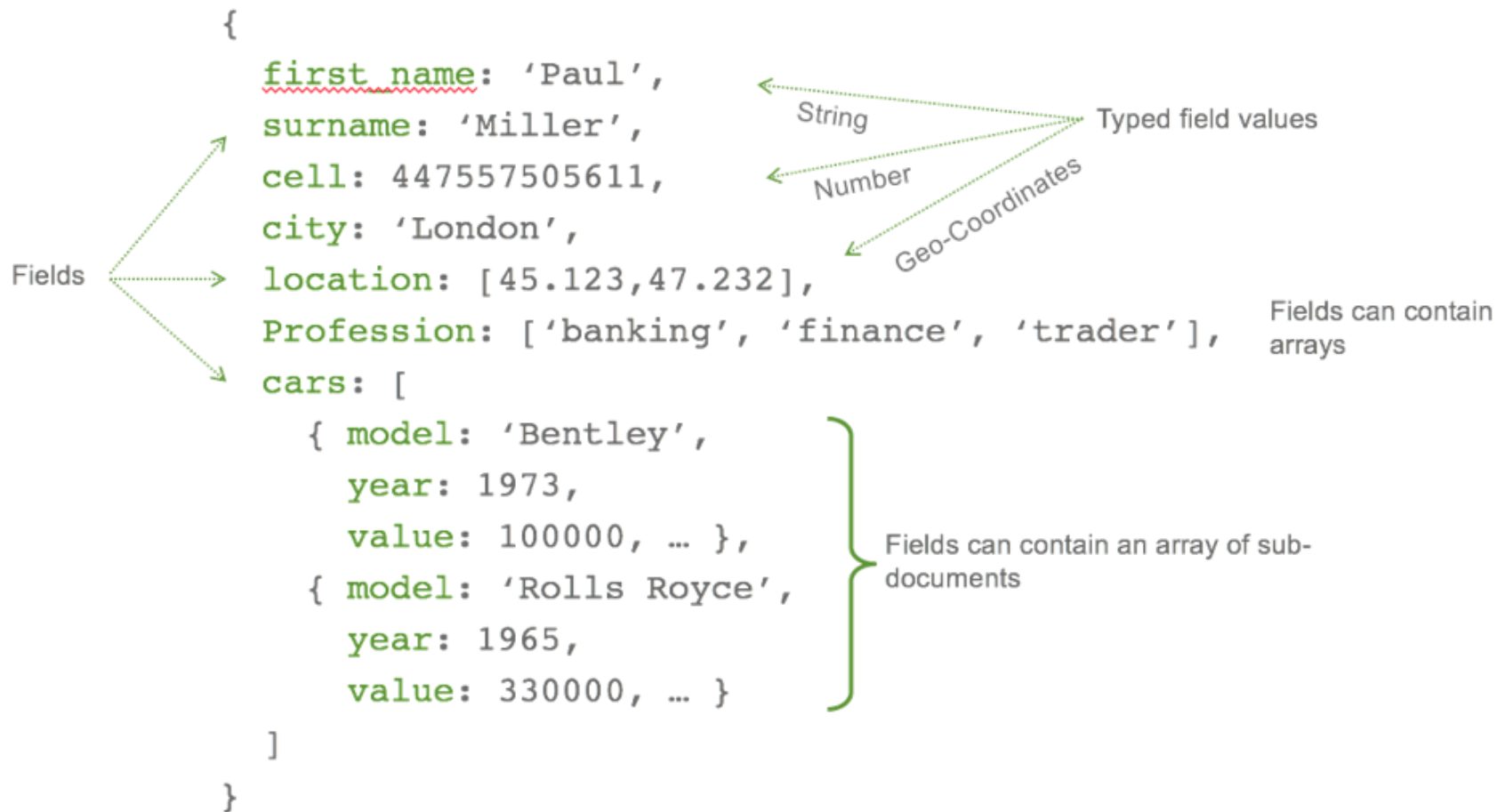
Logs Results

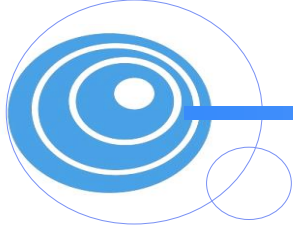
INFO: Tez session hasn't been created yet. Opening session

INFO:



Пример MongoDB документа





Софтверски алат Weka – пример примене у области друмског саобраћаја

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose None | Apply | Stop

Current relation: Relation: Training dataset - PMDS, Instances: 1237, Attributes: 3, Sum of weights: 1237

Selected attribute: Name: PMDS, Missing: 0 (0%), Distinct: 1236, Type: Numeric, Unique: 1235 (100%)

Statistic	Value
Minimum	0
Maximum	28838.06
Mean	7399.954
StdDev	3823.469

Class: PMDS (Num) | Visualize All

Attributes: All | None | Invert | Pattern

No.	Name
1	brojac
2	mesec
3	PMDS

Status: OK | Log

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier: Choose RandomTree -K 0 -M 1.0 -V 0.001 -S 1

Test options: Use training set, Supplied test set, Cross-validation Folds 10, Percentage split % 66

Classifier output:

```
| mesec = oktobar : 15773.16 (5/320329.7)
| mesec = novembar : 14883.44 (5/391495.92)
| mesec = decembar : 14030.89 (5/771358.02)
```

Size of the tree : 483

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.9858
Mean absolute error	334.8252
Root mean squared error	643.0268
Relative absolute error	11.1024 %
Root relative squared error	16.8127 %
Total Number of Instances	1237

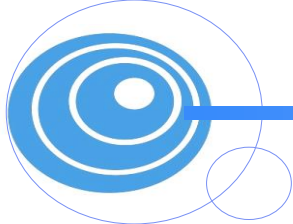
(Num) PMDS

Start | Stop

Result list (right-click for options):

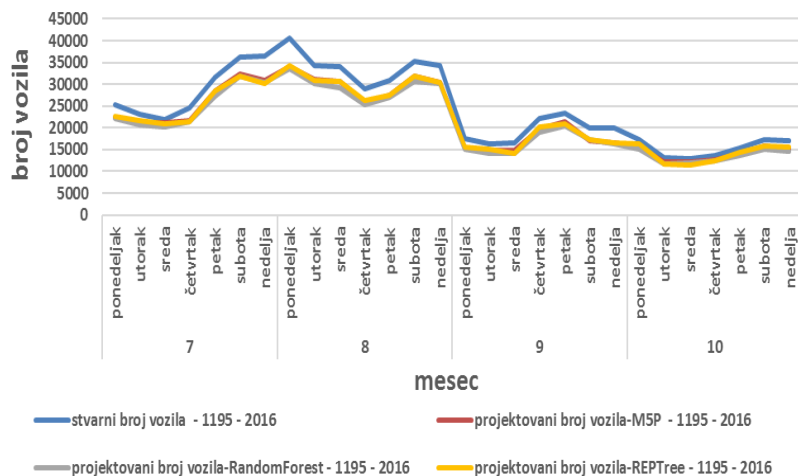
- 18:19:13 - trees.RandomForest
- 18:19:43 - trees.RandomTree
- 18:20:15 - trees.REPTree

Status: OK | Log

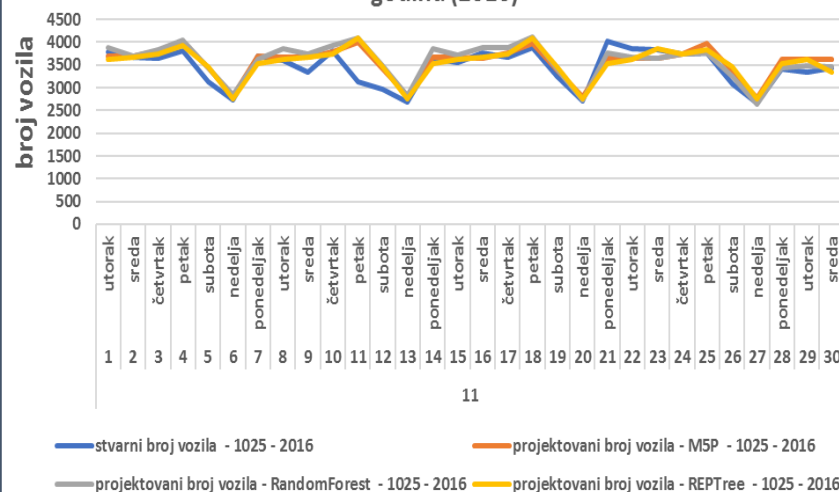


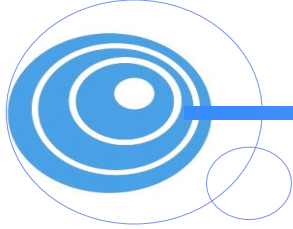
Примери резултата предиктивне анализе у области друмског саобраћаја

Ukupan mesečni stvarni i projektovani protok vozila po danima u nedelji za izabrani brojač saobraćaja (ID: 1195), izabranu godinu (2016) i izabrane mesece (jul, avgust, septembar i oktobar)



Stvarni i projektovani dnevni protok vozila za izabrani brojač saobraćaja (ID: 1025), izabrani mesec (novembar) i izabranu godinu (2016)





Литература

- Sinan Ozdemir, *Principles of Data Science*, Packt, Birmingham – Mumbai, 2016.
- Sridhar Alla, *Big Data Analytics with Hadoop 3*, Packt, Birmingham – Mumbai 2018.
- Doug Bierer, *MongoDB 4 Quick Start Guide*, Packt, Birmingham – Mumbai 2018.
- Ian Witten, Eibe Frank, Mark Hall, Christopher Pal, *Data Mining: Practical Machine Learning Tools and Techniques, 4th edition*, Morgan Kaufmann, 2016.
- Joel Grus, *Data Science from Scratch: First Principles with Python 2nd Edition*, O'Reilly, 2019.
- Ivan Marin, Ankit Shukla, et al., *Big Data Analysis with Python*, Packt, Birmingham – Mumbai, 2019.
- Remco R. Bouckaert, et al., *WEKA Manual for Version 3-9-3*, University of Waikato, Hamilton, New Zealand, 2018.
- Tom White, *Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale 4th Edition*, O'Reilly Media, 2015.