

Big Data Systems

Project: Big Mobility & IoT Data Analytics

- Pročitati članke vezane za Mobility Analytics i Mobility Data Science
 - a. A Survey on Deep Learning for Human Mobility - <https://arxiv.org/abs/2012.02825>
 - b. A Survey on Big Data Processing Frameworks for Mobility Analytics - <https://www.ds.unipi.gr/prof/cdoulik/papers/sigrec21.pdf>
 - c. Mobility Data Science - https://drops.dagstuhl.de/opus/volltexte/2022/16919/pdf/dagrep_v012_i001_p001_2021.pdf
- Izabrati jedan od navedenih velikih skupova podataka (~1GB) koji se odnose na kretanje ljudi i vozila u IoT sistemima pametnih gradova, transporta, zdravstva, poljoprivrede,... Ukoliko na Webu nađete neki sličan veliki dataset potrebno prethodno dobiti odobrenje za korišćenje u projektu od profesora. Prijaviti izabrani dataset odgovorom na ovo obaveštenje na Teams.
 - Human Mobility datasets - https://github.com/scikit-mobility/DeepLearning4HumanMobility/blob/master/Datasets/human_mobility.md
 - The Piraeus AIS Dataset for Large-scale Maritime Data Analytics - <https://doi.org/10.5281/zenodo.5562629>
 - Oslo City Bikes - <https://oslobysyssel.no/en/open-data>
 - SUMO simulator - <https://sumo.dlr.de>
<https://sumo.dlr.de/docs/Tutorials/OSMWebWizard.html>
<https://sumo.dlr.de/docs/Simulation/Output/FCDOOutput.html>
 - Autobusi u Nišu - Podaci o praćenju i kretanju autobusa u gradu
- Download-ovati, prečistiti, eventualno re-formatirati i postaviti podatke na HDFS:
- Razviti aplikaciju korišćenjem Apache Spark framework-a i Python (Scala, Java) API za batch obradu i analizu ovih podataka koja uključuje sledeće funkcionalnosti:
 1. Odrediti broj i karakteristike odgovarajućih vrednosti atributa/događaja na određenoj lokaciji (oblasti) u određenom vremenskom periodu (koji se zadaju kao parametri aplikacije), koji zadovoljavaju zadati uslov.
 2. Naći minimalne, maksimalne, srednje vrednosti, standardne devijaciju i ostale statističke parametre određene(-ih) atributa grupisane po lokaciji/oblasti i vremenu.
- Izvršiti testiranje, evaluaciju i analizu performansi samostalne Python/Java aplikacije, kao i Spark aplikacije na klasteru Spark Docker container-a (BDE) na sopstvenom računaru.
- Pripremiti prezentaciju sa opisom projekta i značajnim detaljima implementacije. Ovu prezentaciju i izvorni kod projekta postaviti na GitHub.