

Tema III projekta je kreiranje aplikacije korišćenje Apache Spark okruženja za obradu podataka. Prilikom izrade projekta možete birati jednu od naredne dve ponuđene teme.

Tema A: Kreiranje Spark aplikacije za treniranje i testiranje modela mašinskog učenja:

- Pronaći set podataka koji treba da posluži kao izvor podataka za ovaj projekat.
- Uraditi pripremu podataka (pr. prečišćavanje i normalizacija podataka).
- Kreirati model mašinskog učenja (kao što su klasifikacija, regresija i klasterizacija).
- Kreirani model testirati.
- Sačuvati i prikazati rezultate.

Tema B: 2. Kreiranje Spark aplikacije koja u realnom vremenu (near real-time) obrađuje podatke iz nekog izvora (stream-a) podataka:

- Za odabrani izvorni stream podataka implemetirati prikupljan je podataka.
- U zavisnosti od tipa podataka, obezbediti near real-time obraditu prikupljenih podataka i prikaz rezultata obrade
- Obezbediti skladištenje prikupljenih podataka i rezultata njihove obrade.

Projekat rade isti timovi koji su radili I i II projekat. Za prijavu projekta timovi mogu koristiti sledeći [link](#) (Popunite podatke u koloni 3. projekta).

Rok za izradu projekta: 25.06.2023. godine.