

Binarna klasifikacija: Predviđanje zadovoljstva putnika

Martina Gačina, Nikola Sunara, Fran Vojković, Alen Živković

Travanj 2020

Opis problema

Opis problema i

- Bavimo se problemom predviđanja zadovoljstva putnika avionske kompanije.
- Klasifikacijski problem gdje zadovoljstvo može biti: neutralno ili nezadovoljno te zadovoljno
- Proučavamo dataset sa Kaggle-a (Airline Passenger Satisfaction).
- Dataset sadrži kategorije i kontinuirane značajke.
- Proveli smo čišćenje podataka od nedostajućih vrijednosti i previše kolinearnih značajki te odbacili smo nekoliko značajki.

Cilj istraživanja i pregled dosadašnjih istraživanja

Cilj istraživanja i pregled dosadašnjih istraživanja i

- Radimo na novom datasetu, stoga ima nekoliko dosadašnjih istraživanja.
- Dosad se koristio Random Forest, LightGBM, Catboost, XGBoost, linearna regresija itd.
- Nije provedena adekvatna analiza značajki kao ni njihovo odbacivanje.
- Cilj našeg istraživanja je konstruirati bolji model za klasifikaciju zadovoljstva danog putnika sa većom pouzdanošću.

Plan istraživanja i očekivani rezultati projekta

Plan istraživanja i očekivani rezultati projekta i

- Koristit ćemo sljedeće metode nadziranog učenja za naš model:
 - Random Forest
 - Linearnu regresiju
 - Naivnog Bayesa
- Razmišljamo i o dodatnim metodama koje ćemo koristiti po potrebi.
- Planiramo evaluirati naš model na nekoliko načina, npr. pomoću roc auc metrike.

Plan istraživanja i očekivani rezultati projekta ii

- Očekujemo da ćemo konstruirati bolji model od već postojećih za naš dataset
- Nadamo se da će točnost našeg modela biti barem 90 %

Hvala na pažnji 😊