

Binary Classification Using KNN for BPJS Fraud Prediction

BUSINESS UNDERSTANDING

Pengembangan model data mining yang ditujukan untuk membantu rumah sakit dalam melakukan pendekripsi terhadap penipuan yang terjadi di rumah sakit terkait klaim pelayanan. Analisis dilakukan dengan algoritma klasifikasi supervised learning, yaitu KNN (K-Nearest Neighbors) pada data BPJS dan metodologi CRISP-DM (Standar Kompetensi Kerja Nasional: KepMen Ketenagakerjaan No 299 thn 2020).

Data Preparation

- Data Cleaning
- Feature Selection
- Data Transformation
- Data Labelling

Data Understanding

Dataset yang digunakan dataset BPJS Kesehatan Hackathon, terdiri dari 200217 record dengan 53 kolom.

Modelling

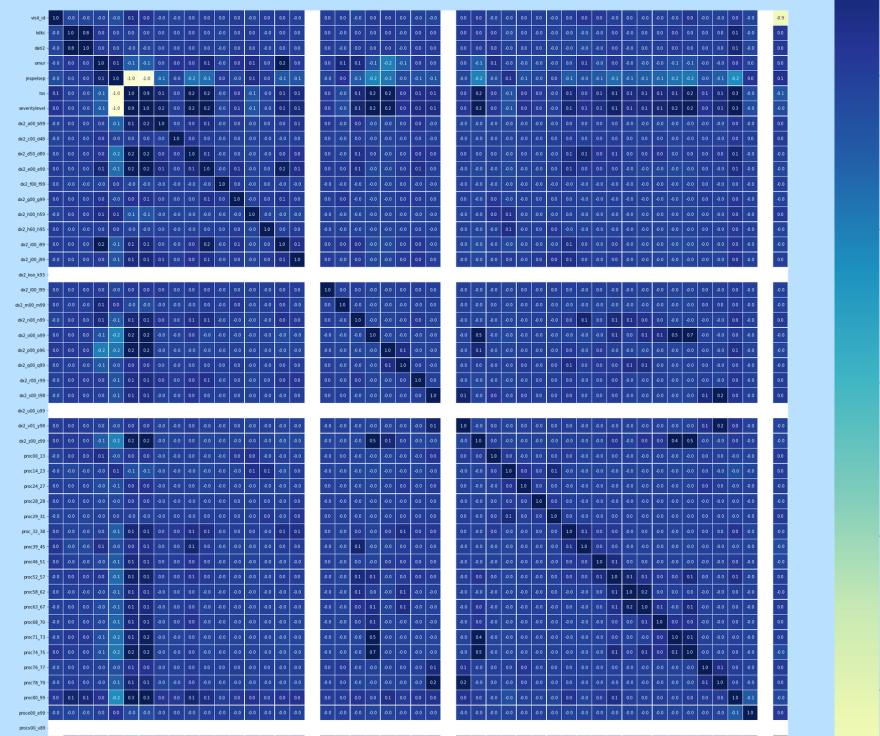
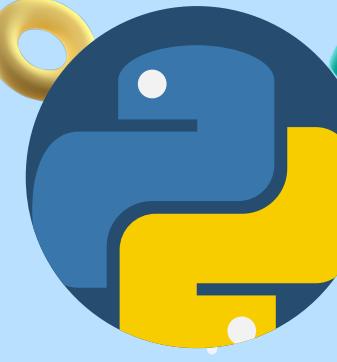
Membagi data menjadi training set dan testing set untuk pemodelan Model yang diterapkan dalam menyelesaikan masalah klasifikasi dan regresi adalah KNN (K-Nearest Neighbors).

Evaluation

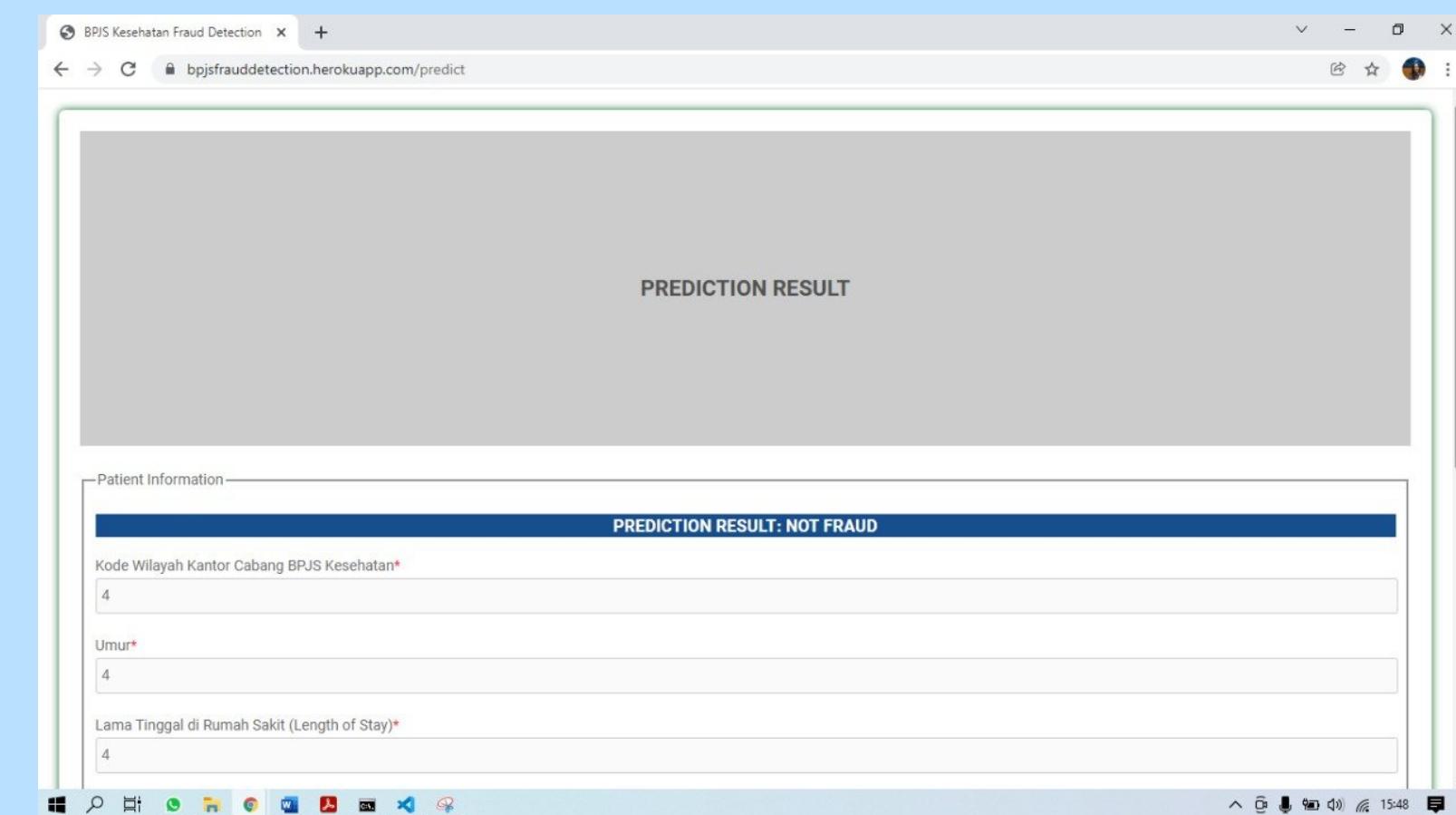
- Classification report
- Confusion matrix
- Accuracy score

Deployment

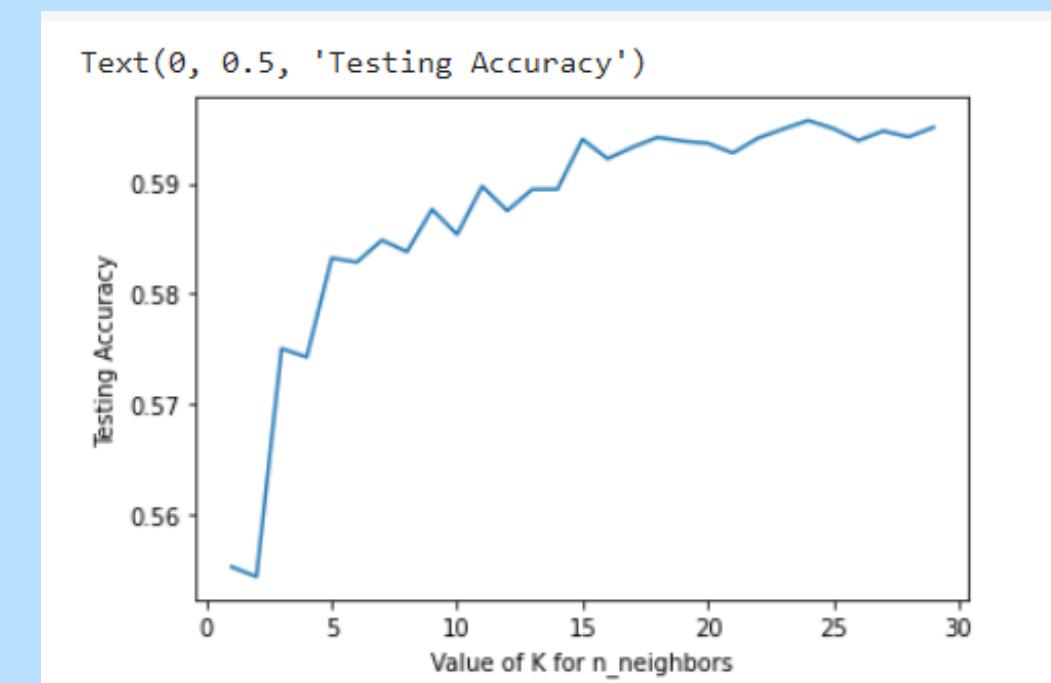
Deploy website model fraud detection menggunakan Heroku.



KORELASI ANTAR ATRIBUT MENGGUNAKAN HEATMAP



HASIL DEPLOYMENT



VISUALISASI TESTING ACCURACY

