

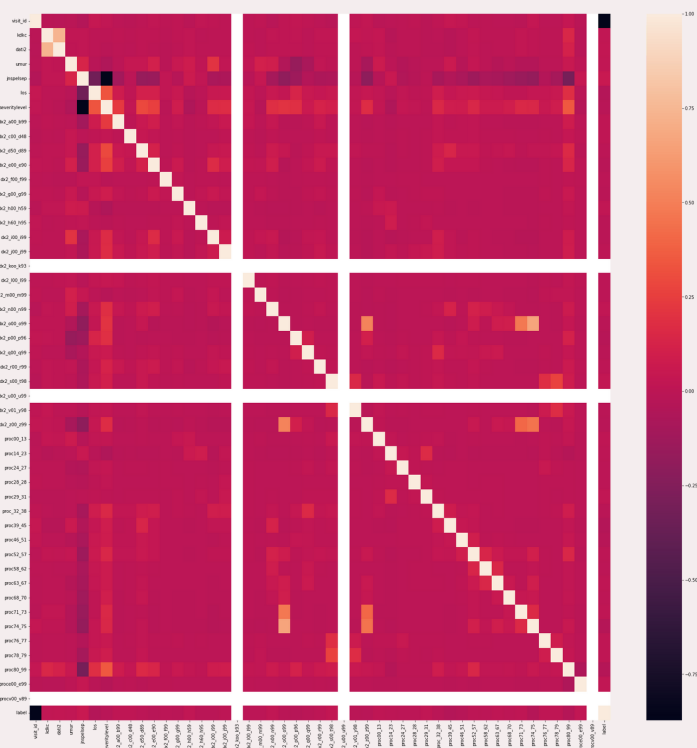


Fraud Detection of Badan Jaminan Kesehatan Masyarakat Hackathon Dataset Using Random Forest Classification

Business Understanding

- **Business Objective:** Rumah sakit sebagai sarana pelayanan kesehatan masyarakat tidak lepas dari terjadinya fraud yang menguntungkan berbagai pihak
- **Goal:** Melakukan prediksi potensi terjadinya fraud pada klaim pelayanan BPJS rumah sakit
- **Plan:** Melakukan tahapan data mining sesuai dengan metodologi data science CRISP-DM

Data visualization



Berdasarkan hasil visualisasi terdapat 3 fitur yang saling tidak berkorelasi, yaitu:

- 'procv00_v89',
- 'dx2_u00_u99'
- 'dx2_koo_k93'

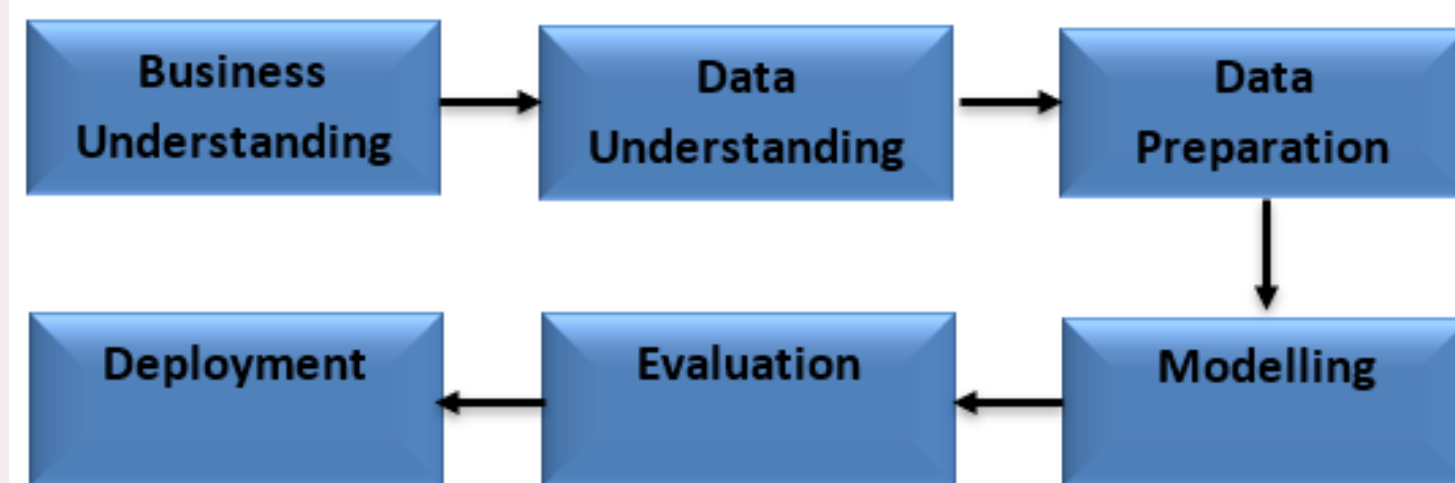
Sedangkan fitur **dati2** dan **kdkc** berkorelasi tinggi

Feature Selection

Pemilihan fitur yang dilakukan berdasarkan kegunaan dan pengaruh fitur terhadap deteksi kecurangan data. Maka fitur yang akan dihapus pada saat pembentukan model adalah.

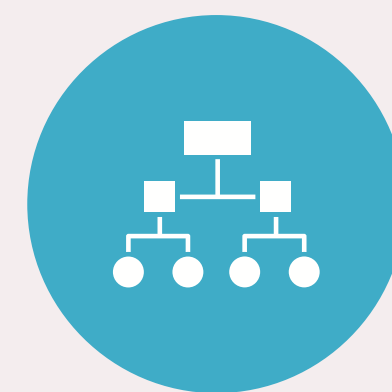
- 1.visit_id
- 2.dati2
- 3.procv00_v89
- 4.dx2_u00_u99
- 5.dx2_koo_k93

CRISP - DM



Penyelesaian masalah dilakukan dengan menggunakan metodologi CRISP-DM dan pemodelan dilakukan dengan menggunakan **Random Forest Classification**.

Pemilihan RFC untuk penyelesaian klasifikasi kecurangan data bpjs dipilih karena banyaknya fitur yang digunakan saat pemodelan.



Modeling

4 cara pengujian untuk pemilihan model terbaik:



Model yang dipilih: **Tuning Hyperparameters**

Dengan kriteria:

1. Membagi data latih 70% dan data uji 30%
2. Mendefinisikan 3 informasi utama dalam pembuatan keputusan
 - Parameter random_state=5, n_estimators=20
 - Algoritma Random Forest Classifier (RFC)
 - Hasil penilaian dengan akurasi dari data latih dan data uji

Result

Evaluasi Model dilakukan dengan melihat confusion matrix dan AUC Score. Untuk nilai akurasi pada data testing yang dilakukan adalah sebesar 68%

```
=== Confusion Matrix ===
[[20166  9755]
 [ 9415 20730]]

=== Classification Report ===
              precision    recall  f1-score   support

     0       0.68      0.67      0.68     29921
     1       0.68      0.69      0.68     30145

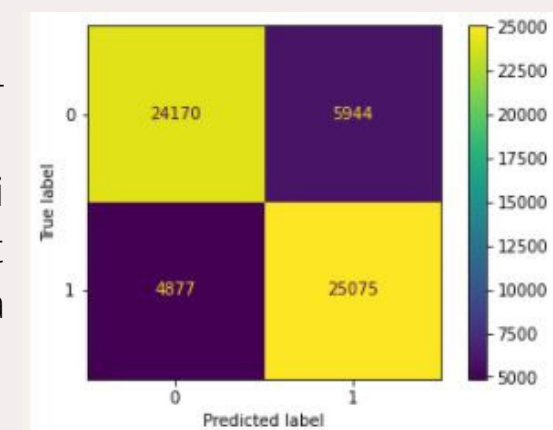
 accuracy          0.68
  macro avg       0.68      0.68      0.68
 weighted avg     0.68      0.68      0.68

=== All AUC Scores ===
[0.73226123 0.73327847 0.73719385 0.73289206 0.73274133]

=== Mean AUC Score ===
Mean AUC Score - Random Forest: 0.733673388293793
```

Hasil uji sebanyak 30% data uji (60.064 sampel).

Terdapat sebanyak 49.245 data yang di prediksi akurat. Sedangkan terdapat 9.921 sampel yang tidak valid pada saat dilakukan pengujian.



Deployment

Web : <https://fraud-detection-rfc-model.herokuapp.com/>
Deployed in : Heroku
Framework : Flask