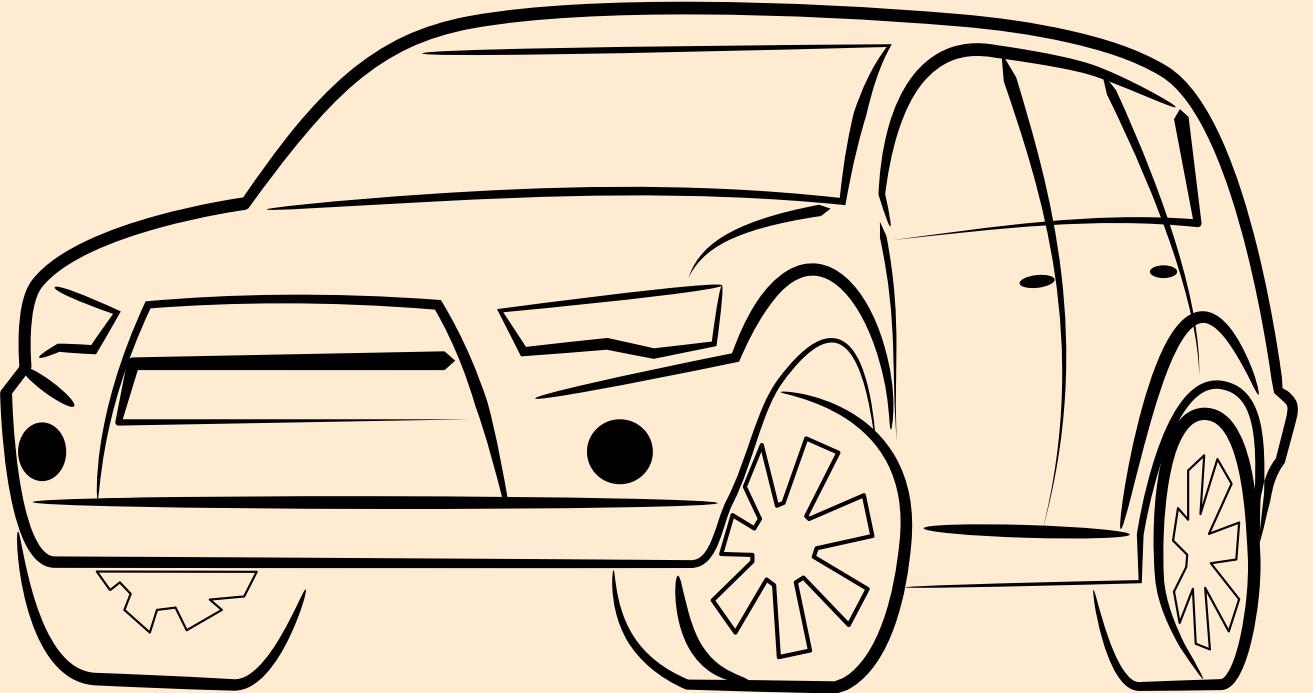


Classification Car Evaluation with Python



Kode Program: AIM0116F2202_ML_01

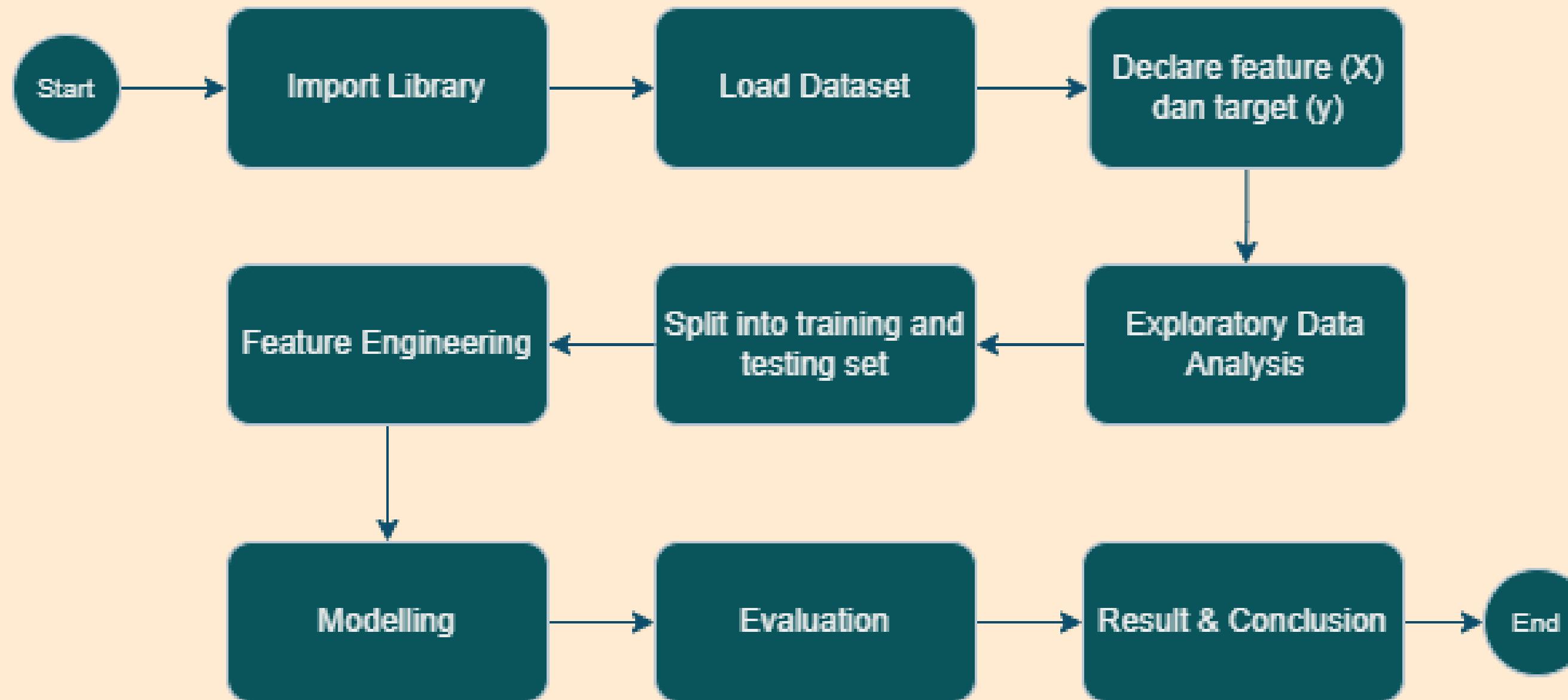
Fayza Apriliza – Broyden

Description

Program ini adalah salah satu penerapan machine learning yang bertujuan untuk mengklasifikasi tingkat keamanan mobil ke dalam beberapa kelas, yakni unacc, acc, good, v-good. Pada program ini menggunakan dataset Car Evaluation yang diunduh UCI ML Repository.

Sumber dataset: https://raw.githubusercontent.com/Wayan123/dataset-ml/main/car_evaluation.csv

Flowchart Program



Modifikasi Program

Beberapa modifikasi yang dilakukan:



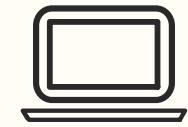
PREPROCESSING DATA

Handling imbalanced
data & Feature
Selection



TEST SIZE

Data dibagi menjadi training
dan testing set dengan
perbandingan 80:20.



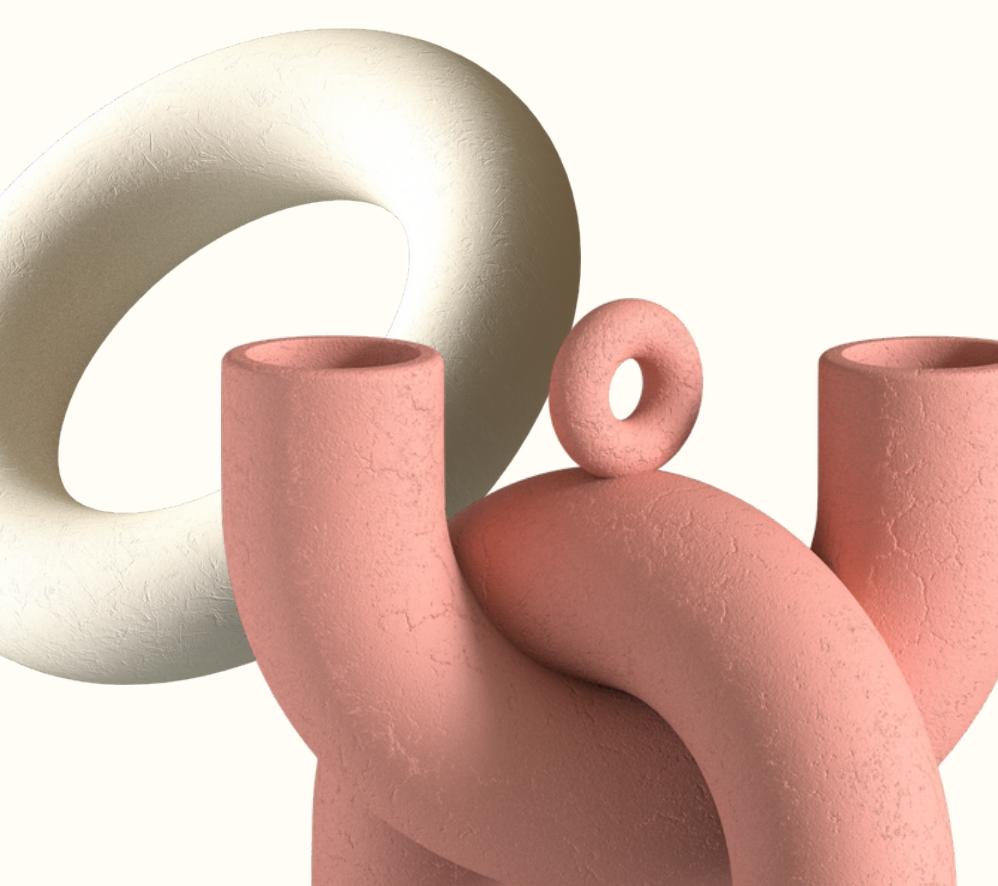
MODELLING

Terdapat 7 variasi model
dengan parameter
berbeda.

Demonstrasi Program Modifikasi

(Kode Program: bit.ly/Broyden-FayzaApriliza-UprakAMD)

Analisis Perbandingan Hasil Modifikasi



1. Data After Undersampling

```
Before undersampling: Counter({'unacc': 975, 'acc': 301, 'good': 58, 'vgood': 48})  
After undersampling: Counter({'acc': 301, 'good': 58, 'unacc': 48, 'vgood': 48})
```

2. Data After Feature Selection (Drop 'doors')

```
#   Column    Non-Null Count   Dtype     
---  --      --          --       --  
0   buying    455 non-null   int64  
1   maint     455 non-null   int64  
2   persons   455 non-null   int64  
3   lug_boot  455 non-null   int64  
4   safety    455 non-null   int64  
dtypes: int64(5)
```

3. Accuracy

```
Model accuracy score with criterion gini index: 0.9364
```

```
Model accuracy score with criterion gini index: 0.9306 (+ feature selection)
```

```
Model accuracy score with criterion entropy: 0.9220
```

```
Model accuracy score with criterion entropy index: 0.9162 (+ feature selection)
```

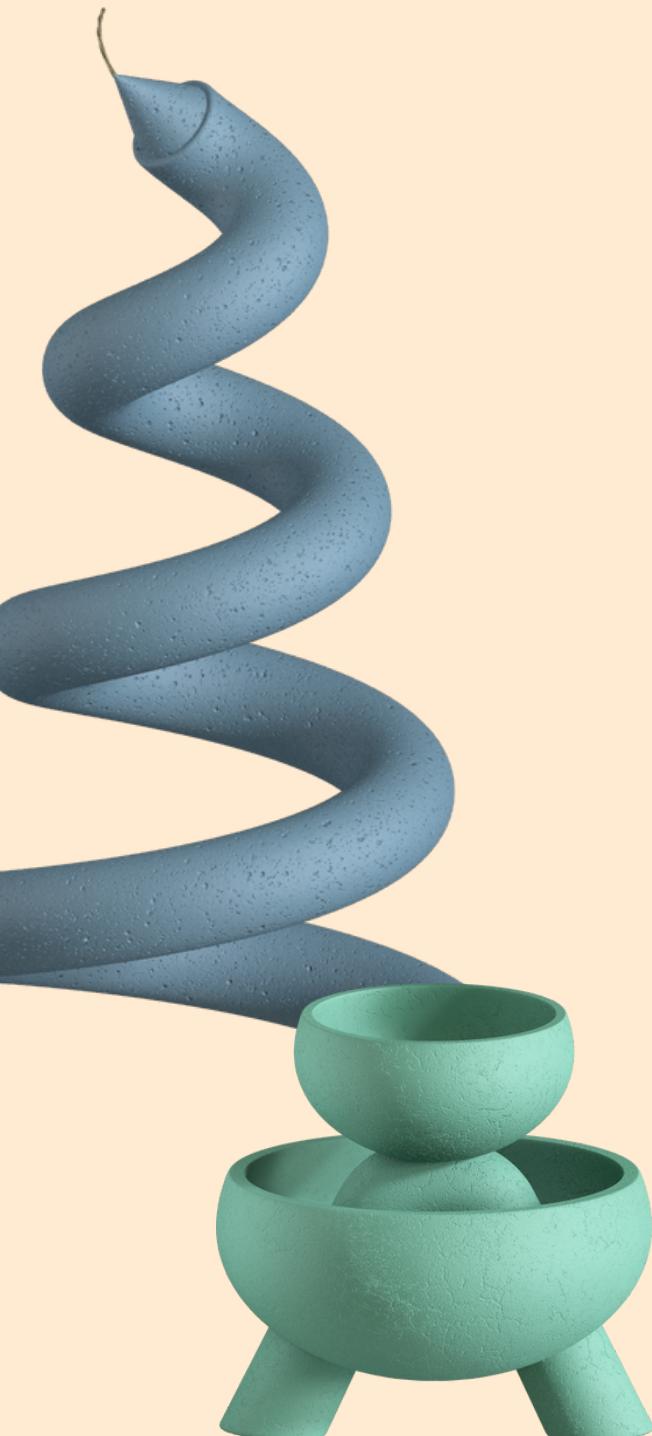
```
Model accuracy score with GridSearchCV: 0.9393
```

```
Model accuracy score with gridsearchcv and feature selection index: 0.9364
```

```
Model accuracy score with Random Forest: 0.8237
```

Kesimpulan

Program modifikasi dibuat dengan 2 base model klasifikasi, yakni random forest dan decision tree. Dengan dilakukan undersampling, set test size, feature selection, dan tune parameter mengalami perbedaan akurasi yang signifikan dibandingkan program sebelumnya. Diperoleh model terbaik dengan akurasi sebesar 94% dengan gap (selisih akurasi train dan test) tidak terlalu besar. Artinya, model tidak overfitting ataupun underfitting.



Thank you!

Kode Program: AIM0116F2202_ML_01

Fayza Apriliza - Broyden

