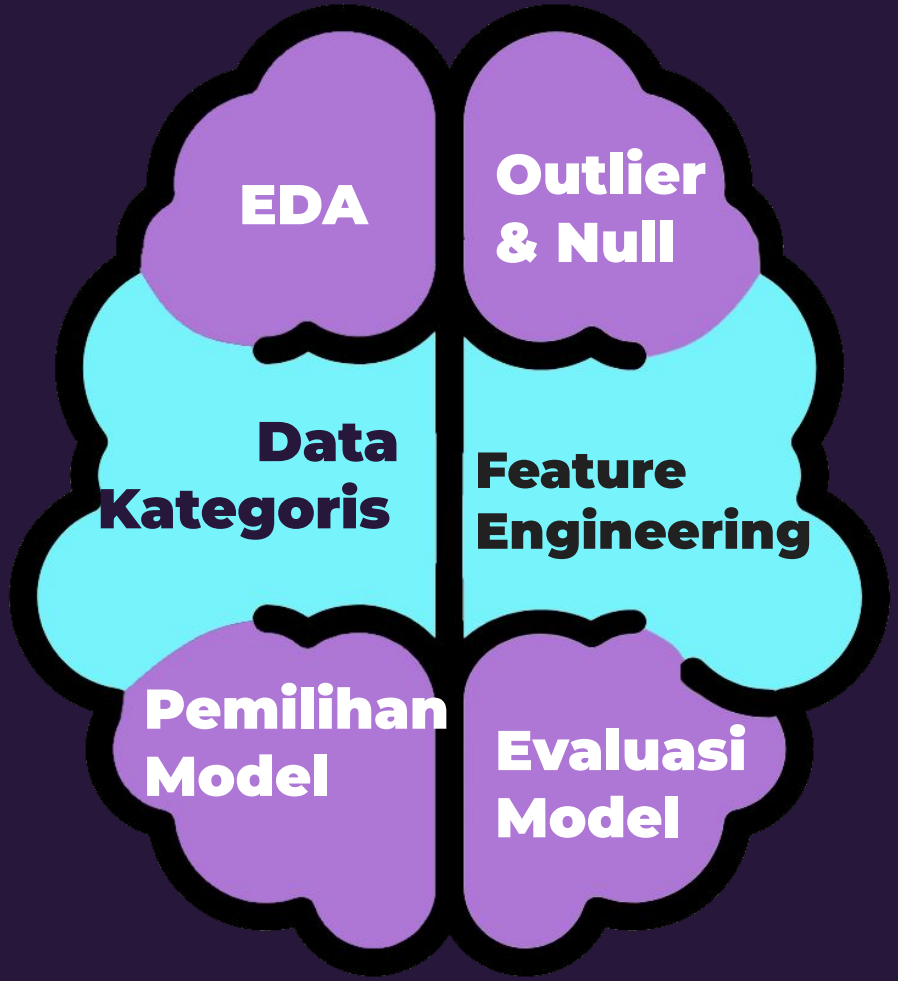
MultiClass Music Genre Prediction 😘😘

Bagus Cipta Pratama (23/516539/PA/22097)
Hafid Sasayuda Ambardi (23/519841/PA/22308)
Joseph Greffen Komala (23/519909/PA/22318)


EDA
Outlier & Null
Data Kategoris
Feature Engineering
Pemilihan Model
Evaluasi Model

Tantangan :

01 Melakukan EDA

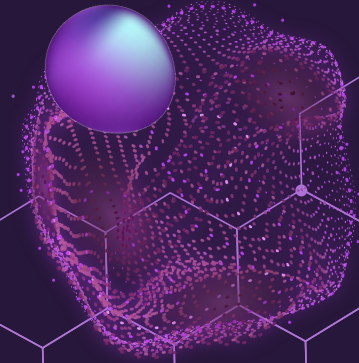
02 Munculnya outlier dan null values

03 Pengelolaan data kategoris

04 Feature engineering

05 Memilih model yg efektif

Exploratory Data Analysis (EDA)

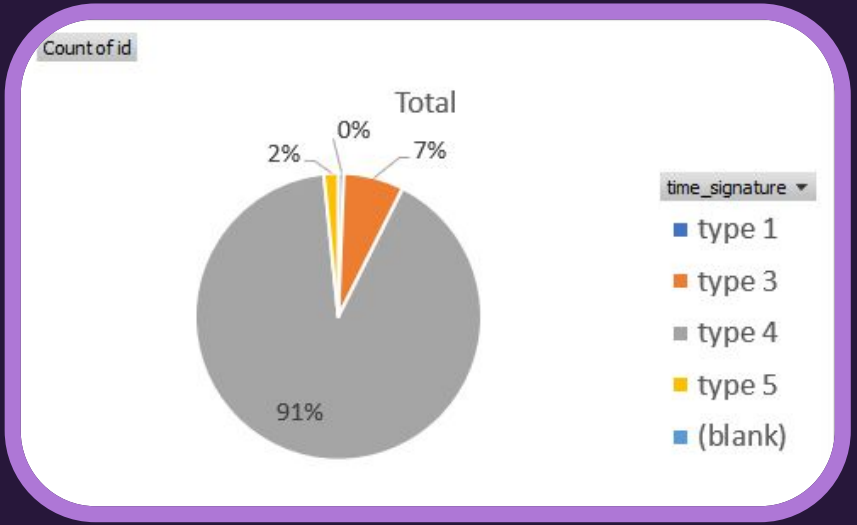


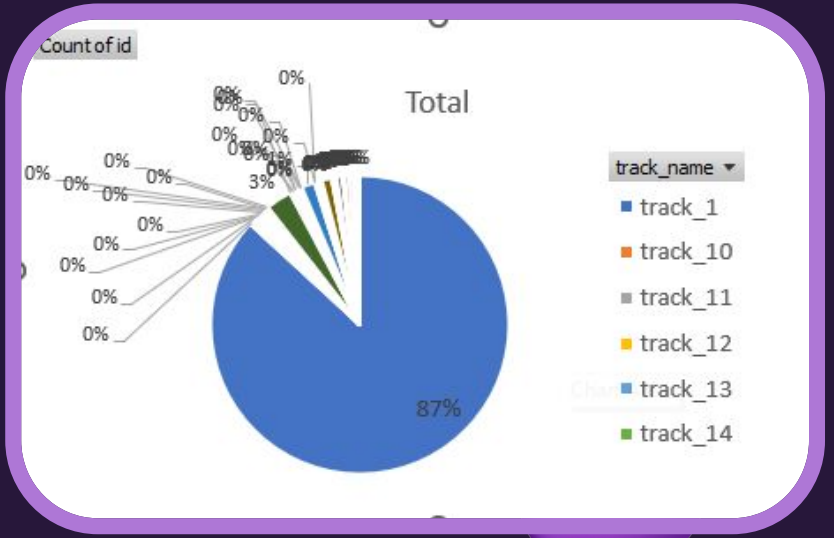
Exploratory Data Analysis (EDA)

Time Signature × **Artist Name**

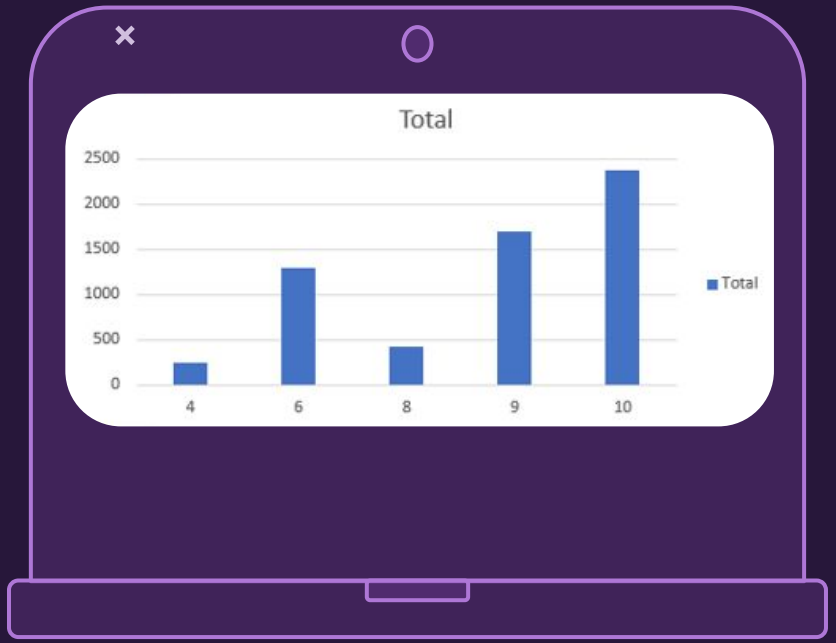
5265 dari 6052
Artis Unik

Track Name





Exploratory Data Analysis (EDA)



Class

Pre-Processing

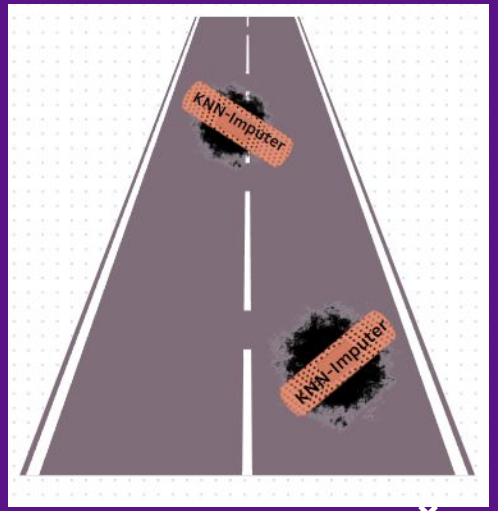
Our Action

Outlier
Pembatasan dengan IQR

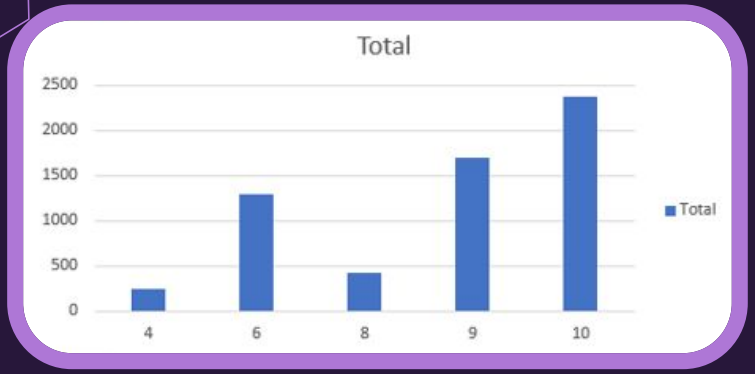
Null Values
KNN-Imputer

Mode
Ordinal Encoding

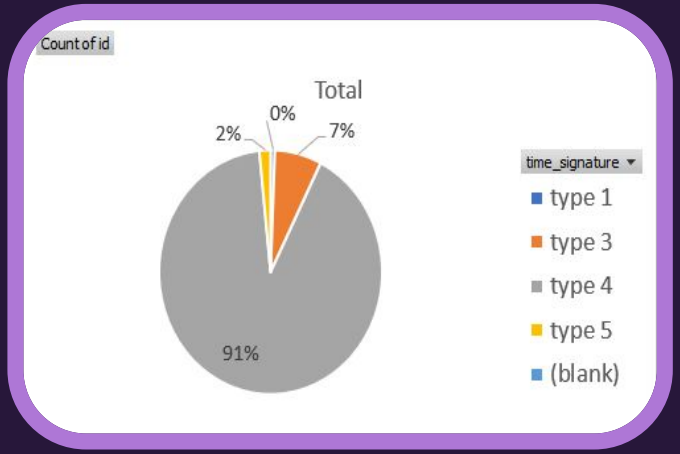
Artist Name
Menghapus



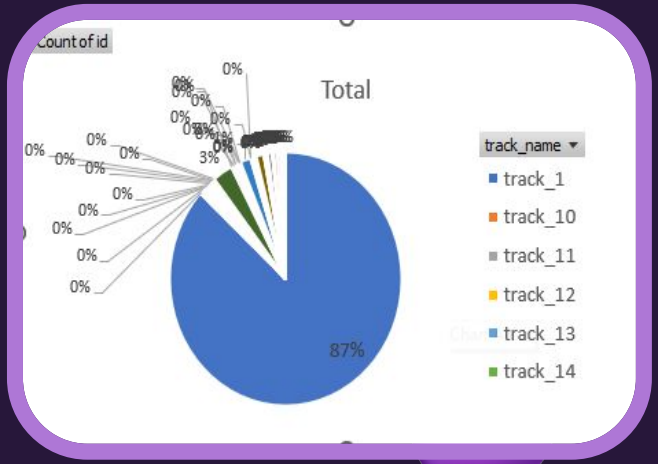
Before Pre-Processing



Class

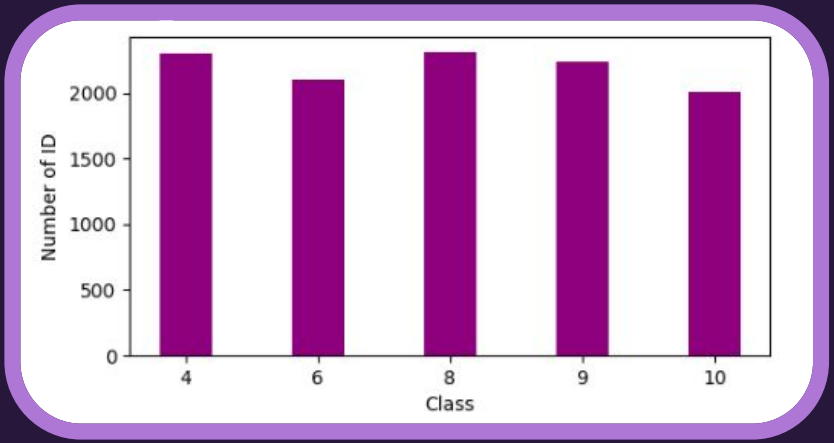


Time Signature

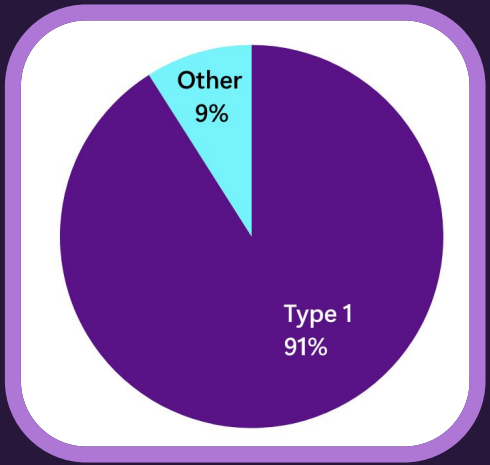


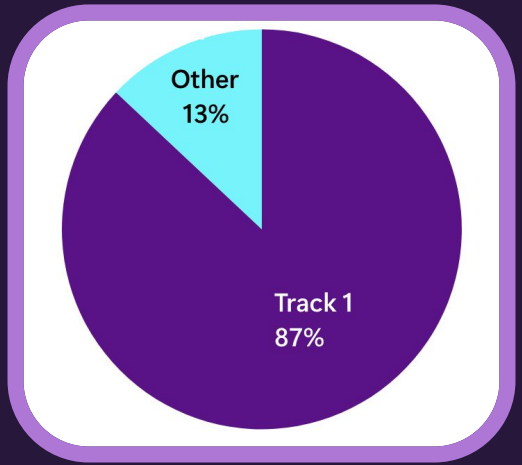
Track Name

Our Action



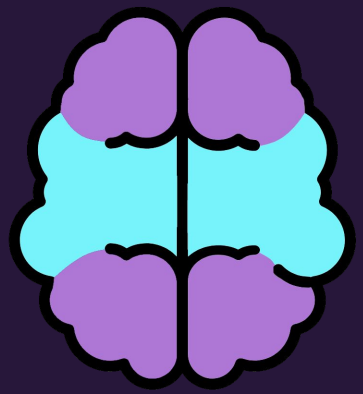
Imbalanced
SMOTE-TOMEK links



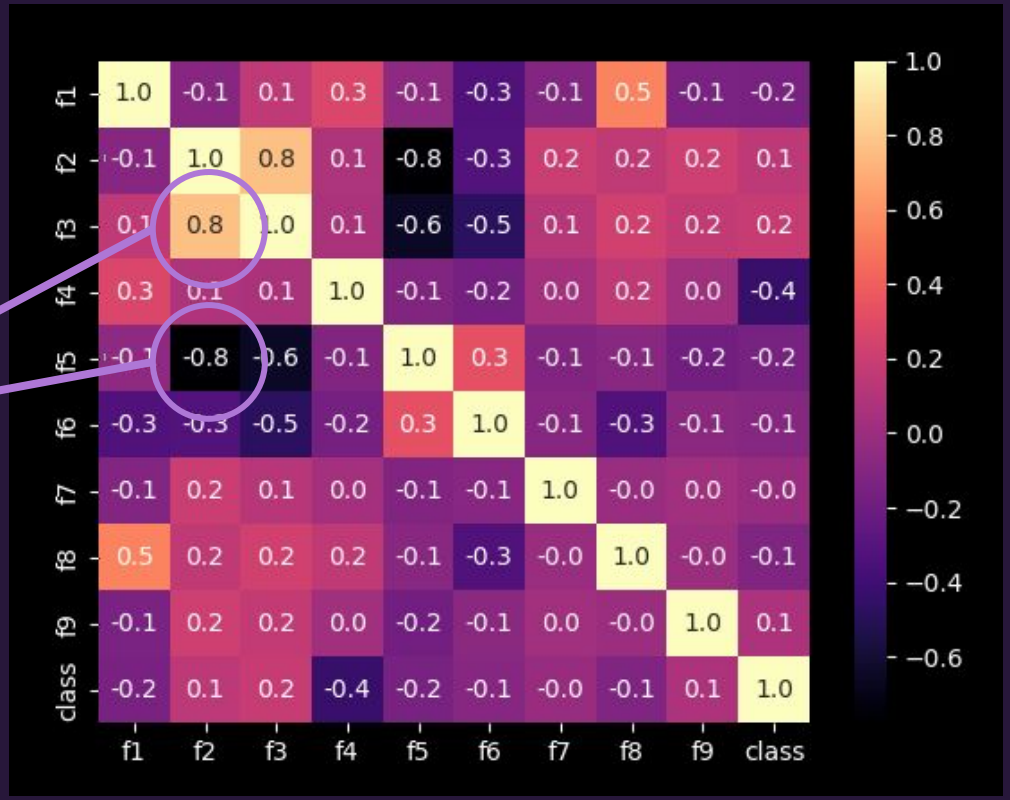


Track Name & Time Signature
Melakukan *grouping sparse classes*

FEATURE ENGINEERING



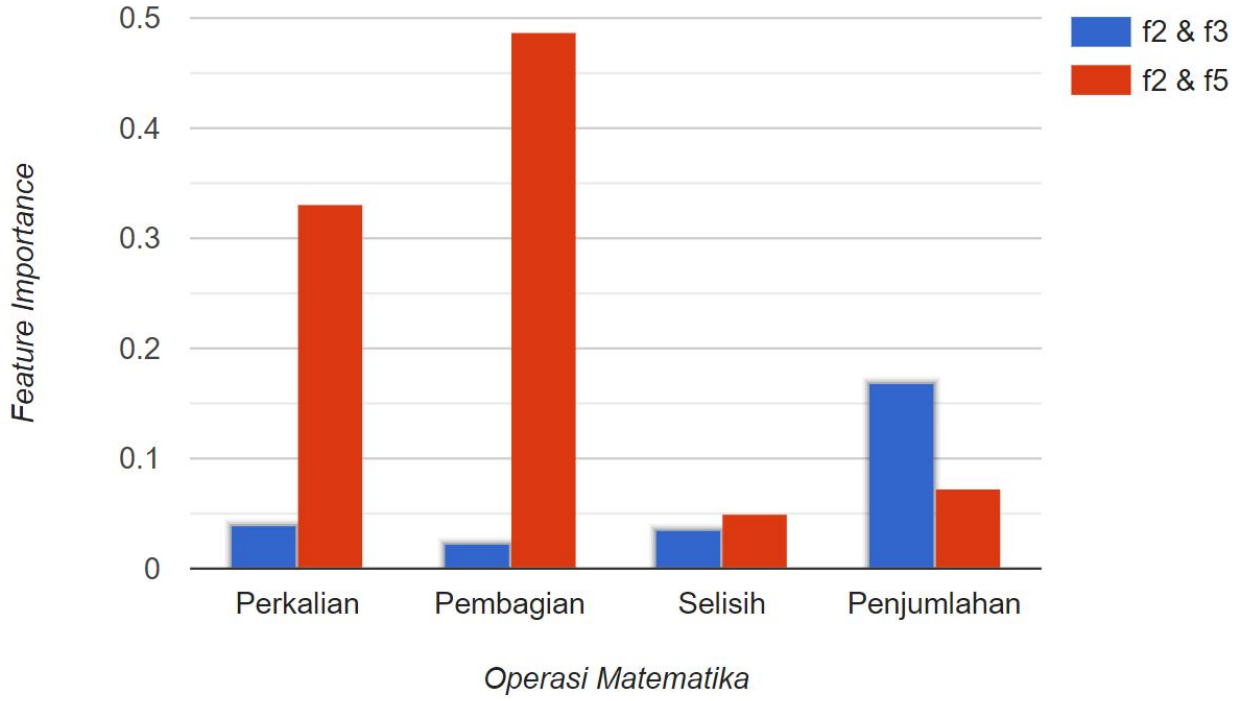
CORRELATION HEATMAP



0.8

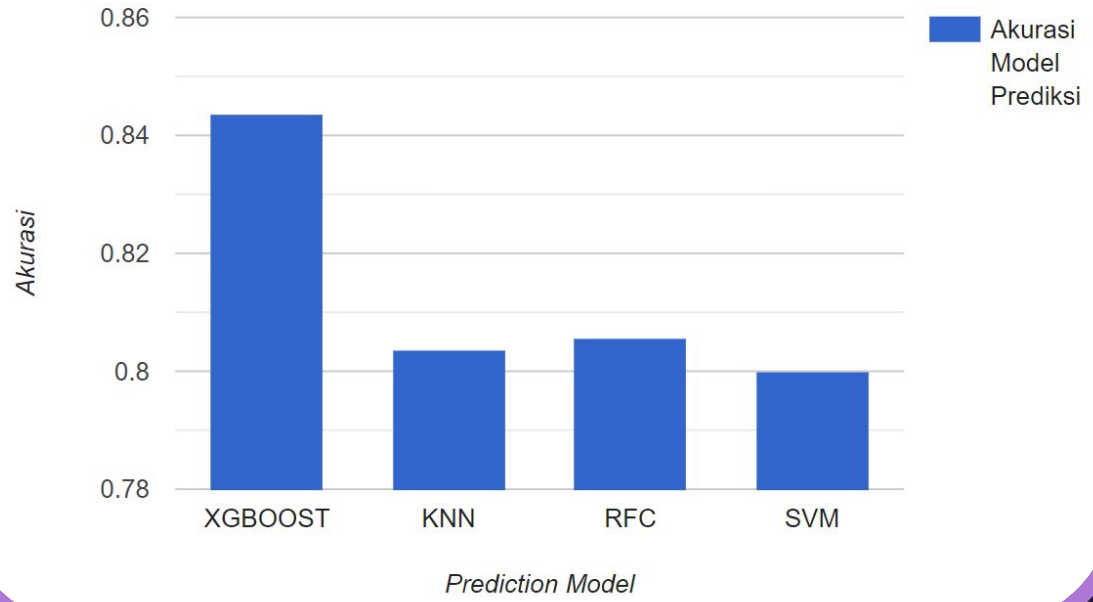
F2 F3 dan F2 F5
Korelasi Terbesar

FEATURE ENGINEERING

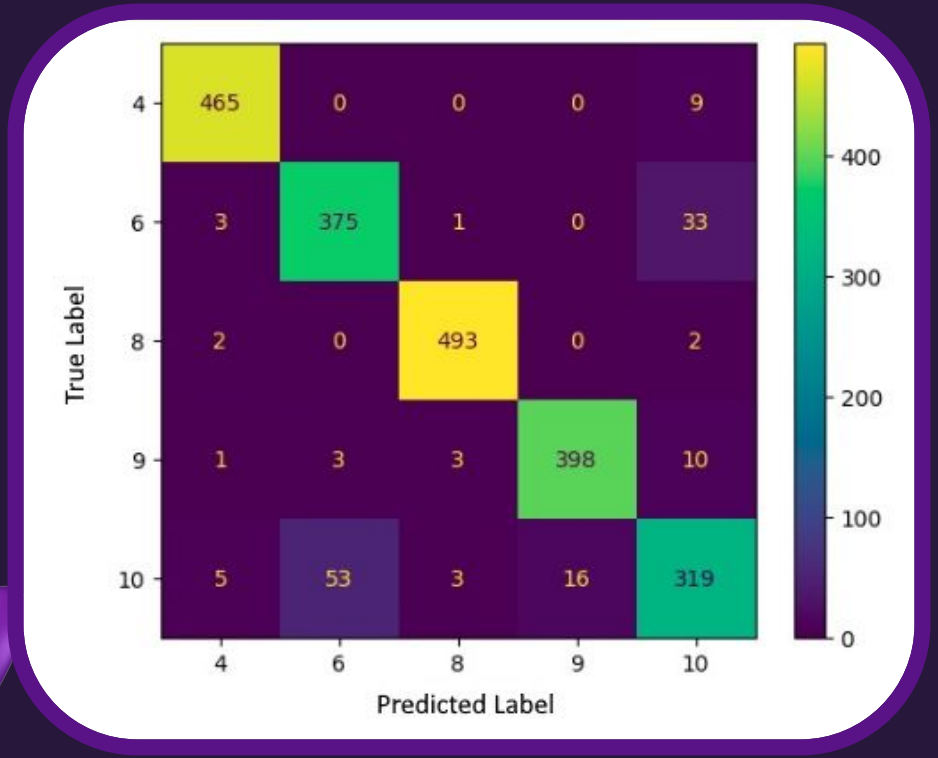


XGBOOST
SVM
RFC
KNN

PREDICTED RESULTS



Confusion Matrix



KESIMPULAN

Data

Pre-Processing

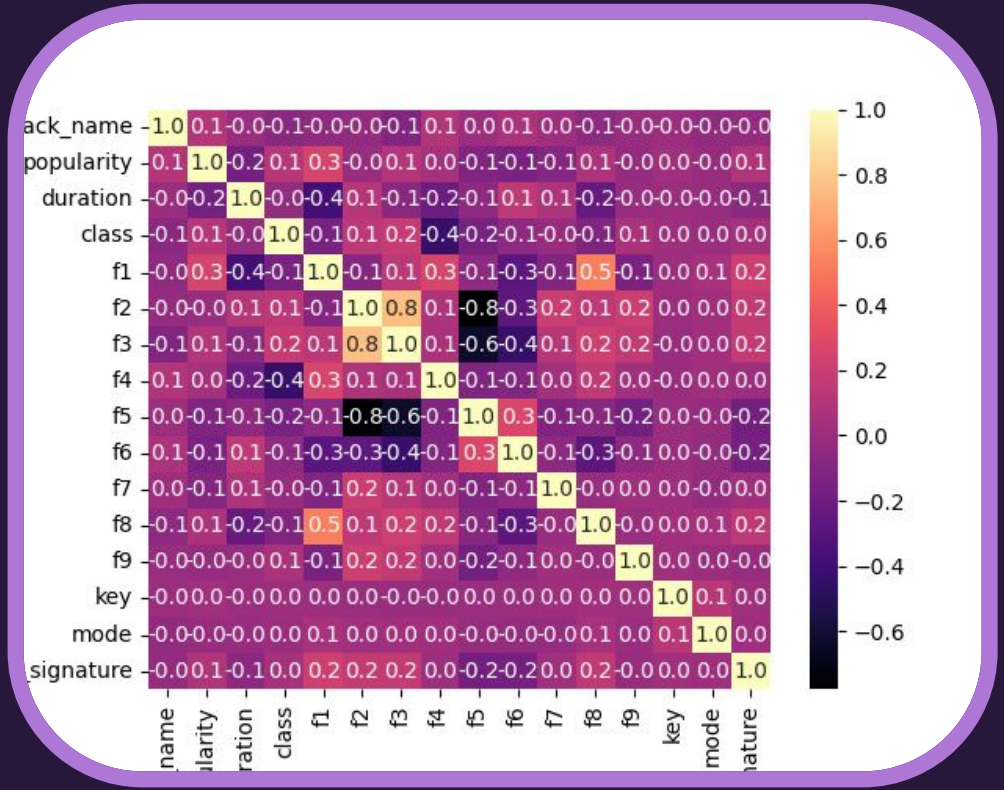
Modeling

Berguna

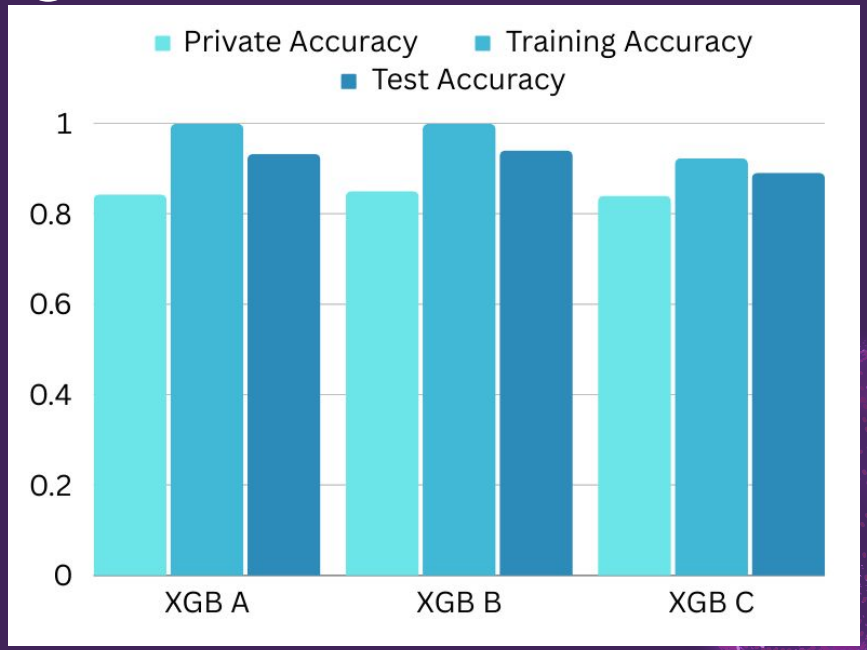
Grafik untuk QnA

*hanya di tampilkan saat dibutuhkan di sesi QnA

CORRELATION HEATMAP

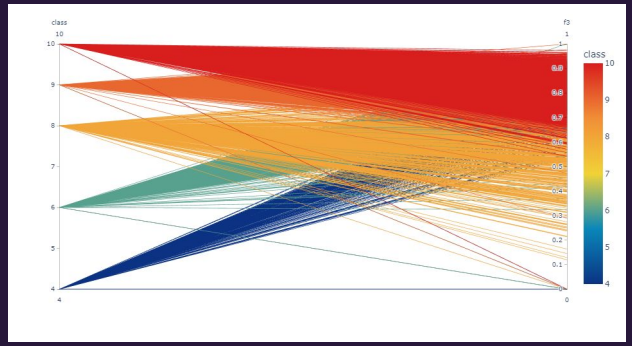
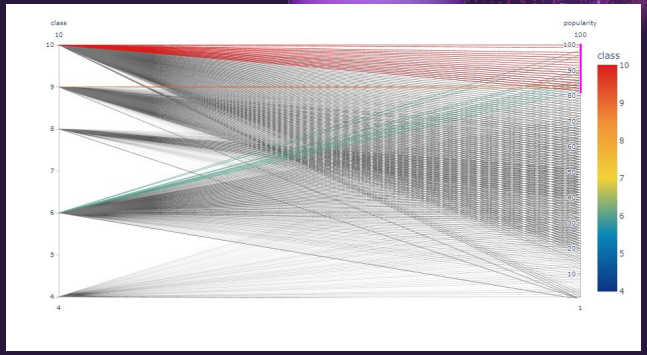
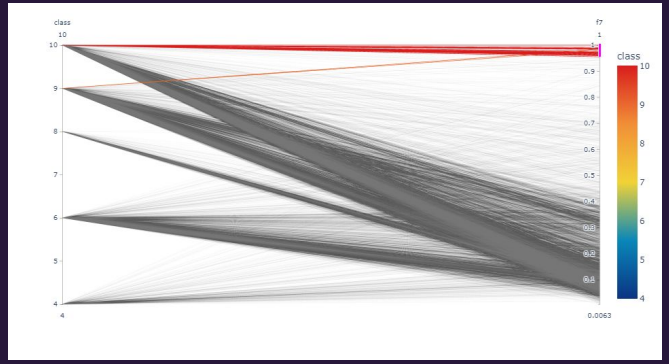


BERBAGAI MODEL XGBOOST



	Private Accurac	Training Accura	Test Accuracy
Upsampling Ove	0.843	0.99933888	0.932794659064
Downsampling	0.85	0.999	0.94
Upsampling Unc	0.84	0.9225647573878	0.890642101422

Different Param

XGBoost A	XGBoost B	XGboost C
'subsample': 1	1	1
'n_estimators': 100	200	200
'max_depth': 6	6	6
'colsample_bytree': 0.75	1	1
'learning_rate': 0.01	0.1	0.1

EDA MENARIK