

Klasifikasi Mahasiswa Berdasarkan Kesuksesan Akademik

Pengantar Sains Data



Our Groups

1. Radityo Ar Rasyid (5002211034)
2. Komang Ryandhi S. (5002211109)
3. Maulana Yusuf Ikhwan R (5002211167)



Table of contents

01 Introduction

Pengantar tentang klasifikasi mahasiswa, tujuan dan metode yang digunakan.

02 Business and Data Understanding

Pemahaman terkait konteks dan lembaga terkait serta dataset yang disediakan.

03 Exploratory Data Analysis

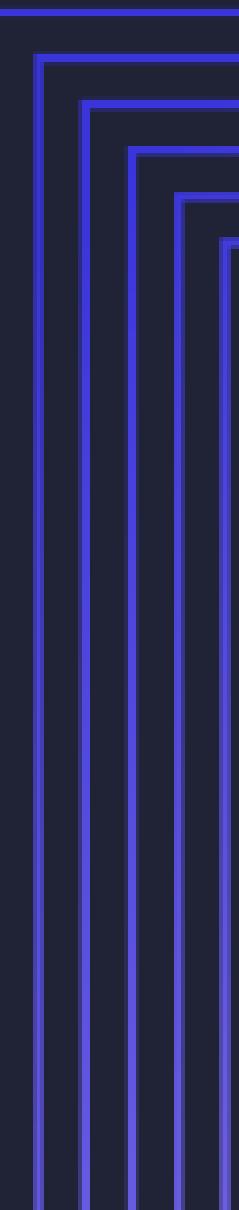
Melakukan analisis data penunjang model.

04 Data Preparation and Modelling

Melakukan rekayasa data dan modelling.

05 Recommendation

Rekomendasi hal yang harus dilakukan.



01

Introduction

Pengantar klasifikasi mahasiswa, tujuan dan metode yang digunakan.

Klasifikasi Tingkat Kesuksesan Mahasiswa

Gambaran Umum

Mahasiswa mempunyai latar belakang yang berbeda-beda berdasarkan kesuksesan akademik mereka. Tiap latar belakang mahasiswa tersebut dapat mempengaruhi mahasiswa itu apakah akan lulus atau tidak. Maka perlu dilakukan klasifikasi untuk mengetahui faktor apa saja yang bisa mempengaruhi tingkat kelulusan mahasiswa.



Tujuan Klasifikasi

Proyek ini bertujuan untuk mengidentifikasi dan memprediksi kesuksesan akademik mahasiswa berdasarkan berbagai faktor, seperti jalur pendaftaran, jenis kursus, waktu kuliah, dan latar belakang pribadi serta keluarga.

Bagaimana caranya?

Kami menggunakan kombinasi beberapa model machine learning seperti Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, dan XGBoost. Voting classifier digunakan untuk menggabungkan prediksi dari beberapa model ini guna meningkatkan akurasi.

Kategori Klasifikasi Mahasiswa

Graduate: Mahasiswa yang berhasil menyelesaikan studi mereka tepat waktu.

Dropout: Mahasiswa yang keluar atau berhenti sebelum menyelesaikan studi.

Enrolled: Mahasiswa yang masih terdaftar dan aktif dalam program studi mereka.

02

Business & Data Understanding

Apa yang diharapkan terhadap project ini?

Overview Dataset

APPLICATION MODE

Jenis jalur pendaftaran mahasiswa.

PREVIOUS QUALIFICATION

Nilai yang pernah dicapai mahasiswa.

COURSE

Jenis-jenis kursus yang diambil mahasiswa.

NATIONALITY

Latar belakang asal negara mahasiswa.

DAYTIME EVENING

Waktu kuliah (siang atau sore).

MOTHER'S & FATHER'S QUALIFICATION

Latar belakang orang tua mahasiswa secara finansial.

Impact of Classification

- Mengetahui ciri-ciri mahasiswa dengan akademik yang baik
- Memprediksi kesuksesan akademik mahasiswa berdasarkan latar belakangnya.

03

Exploratory Data Analysis

Insight apa saja yang terkandung pada data

Exploratory Data Analysis

id	Marital status	Application mode	Application order	Course	Daytime/evening attendance	Previous qualification	Previous qualification (grade)	Nacionality	Mother's qualification	...
0	1	1	1	9238	1	1	126.0	1	1	...
1	1	17	1	9238	1	1	125.0	1	19	...
2	1	17	2	9254	1	1	137.0	1	3	...
3	1	1	3	9500	1	1	131.0	1	19	...
4	1	1	2	9500	1	1	132.0	1	19	...
Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)	Curricular units 2nd sem (evaluations)	Curricular units 2nd sem (approved)	Curricular units 2nd sem (grade)	Curricular units 2nd sem (without evaluations)	Unemployment rate	Inflation rate	GDP	Target	
0	6	7	6	12.428571	0	11.1	0.6	2.02	Graduate	
0	6	9	0	0.000000	0	11.1	0.6	2.02	Dropout	
0	6	0	0	0.000000	0	16.2	0.3	-0.92	Dropout	
0	8	11	7	12.820000	0	11.1	0.6	2.02	Enrolled	
0	7	12	6	12.933333	0	7.6	2.6	0.32	Graduate	

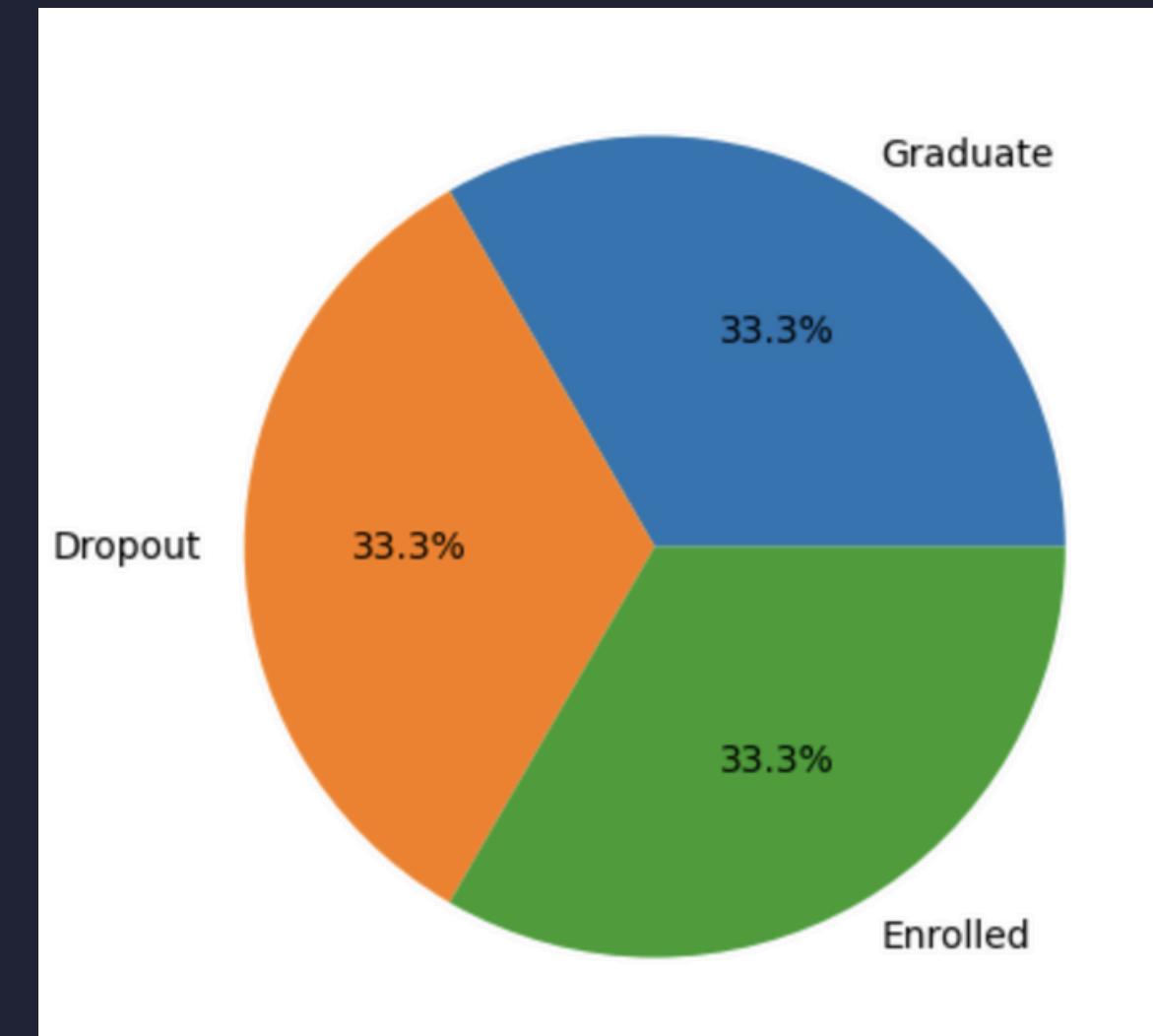
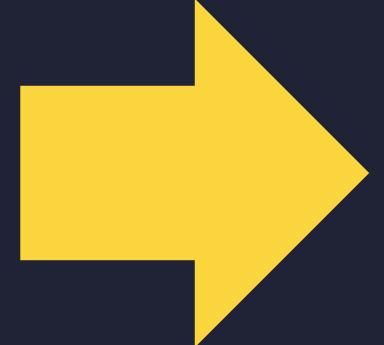
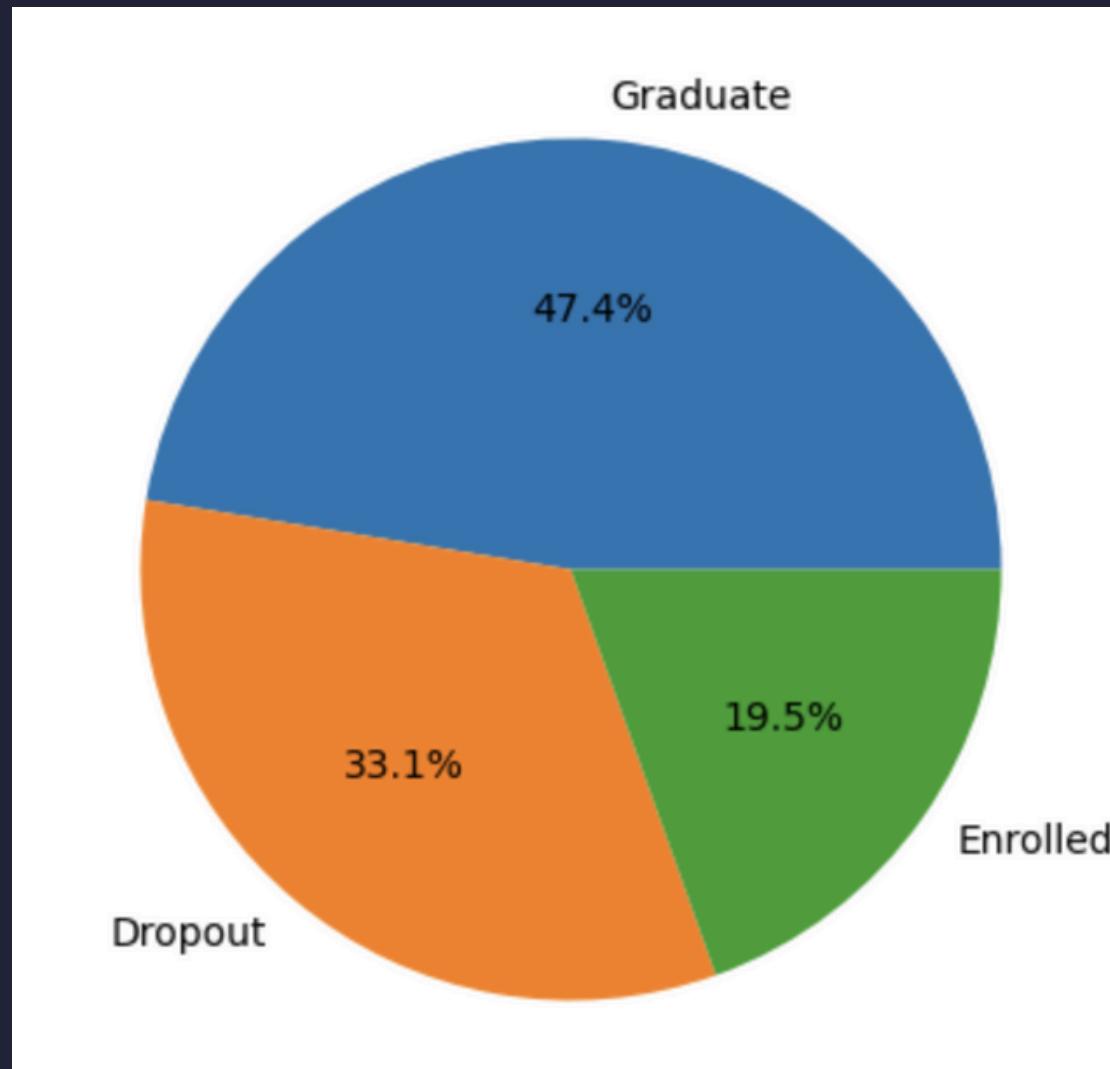


Exploratory Data Analysis

Porsi Mahasiswa yang lulus, DO, masuk

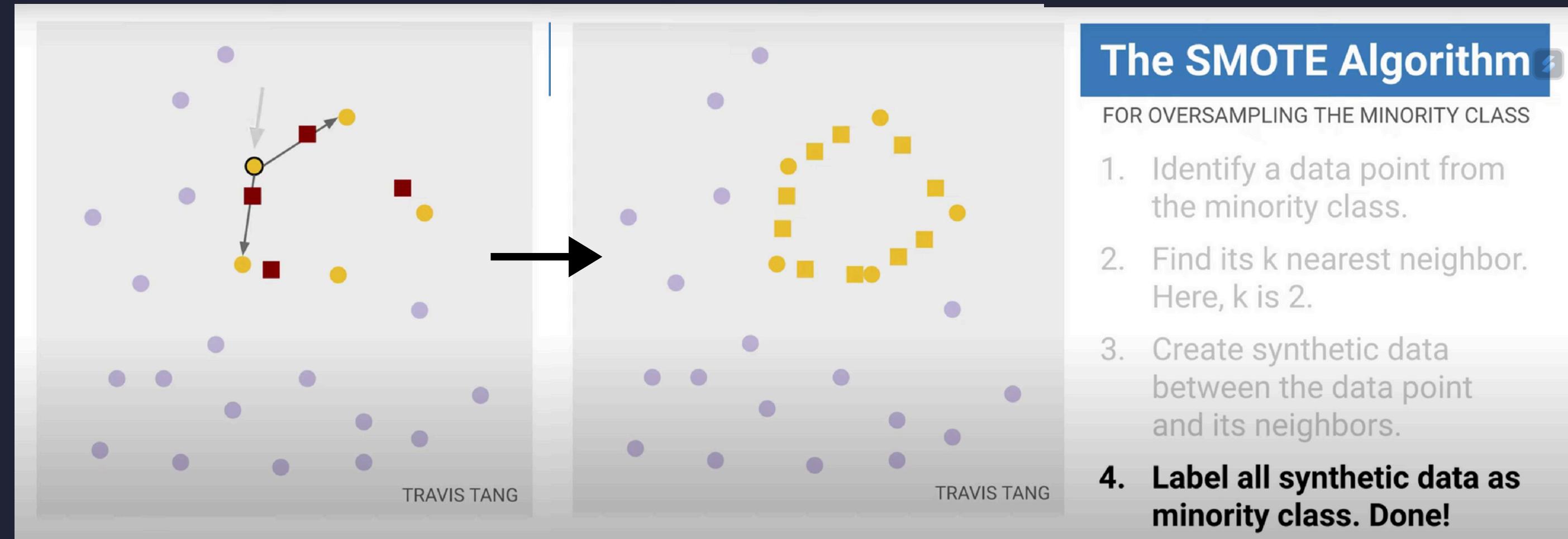
Setelah Oversampling

Sebelum Oversampling



Handling Imbalance Data

Oversampling



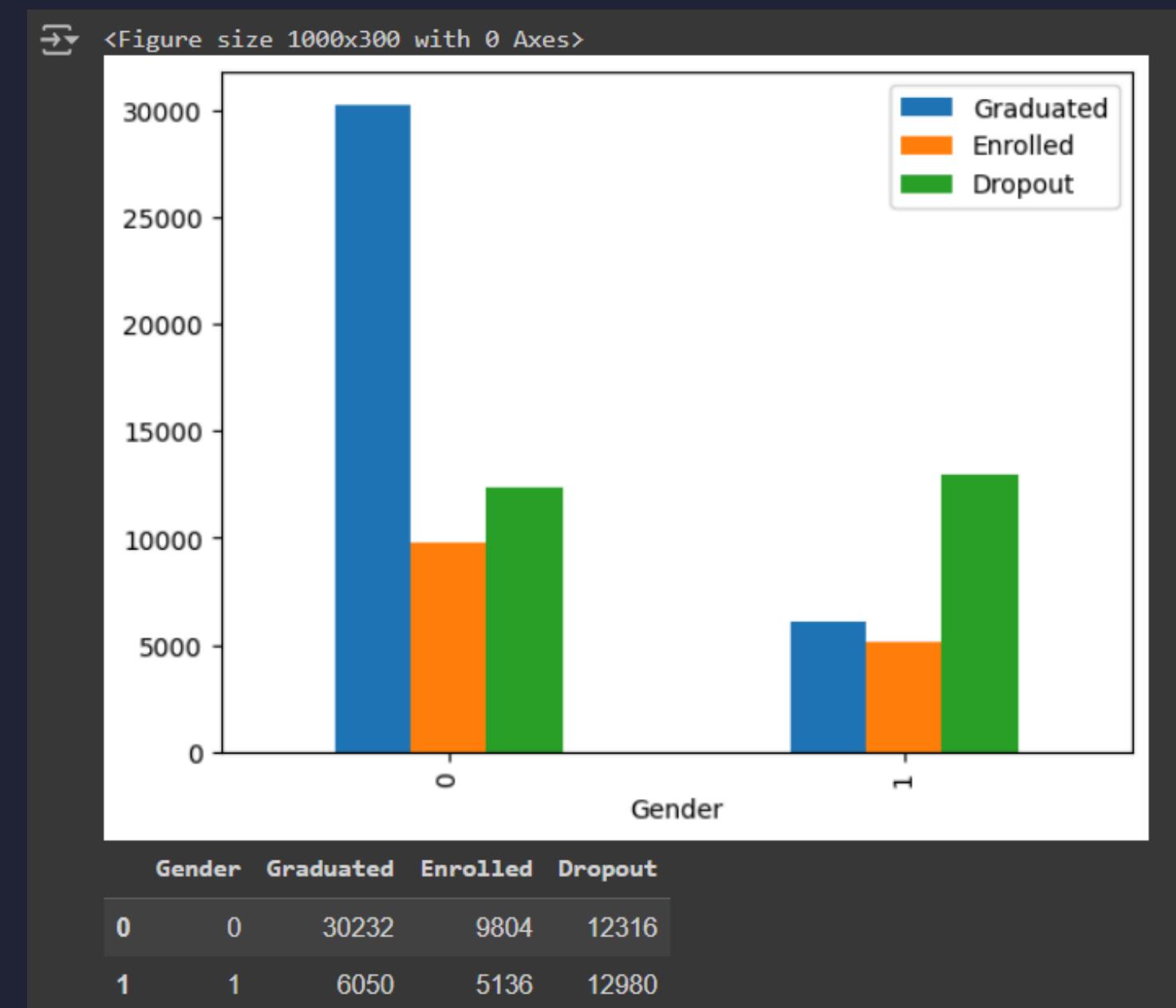
- Imbalance data adalah kondisi ketika data suatu kolom ada satu nilai jumlahnya mendominasi dibanding nilai lain
- Penanggulangannya dengan Oversampling SMOTE Algorithm

Source: Travis Tang



Exploratory Data Analysis

Porsi Mahasiswa berdasarkan Gender



04

Data Preparation & Modeling

Table of contents

01 Feature Engineering

Melakukan encoding

03 Hasil Model

Mencari hasil prediksi yang ditentukan model

02 Model Training

Melatih model dengan data yang kita miliki
agar dapat memprediksi

Feature Engineering

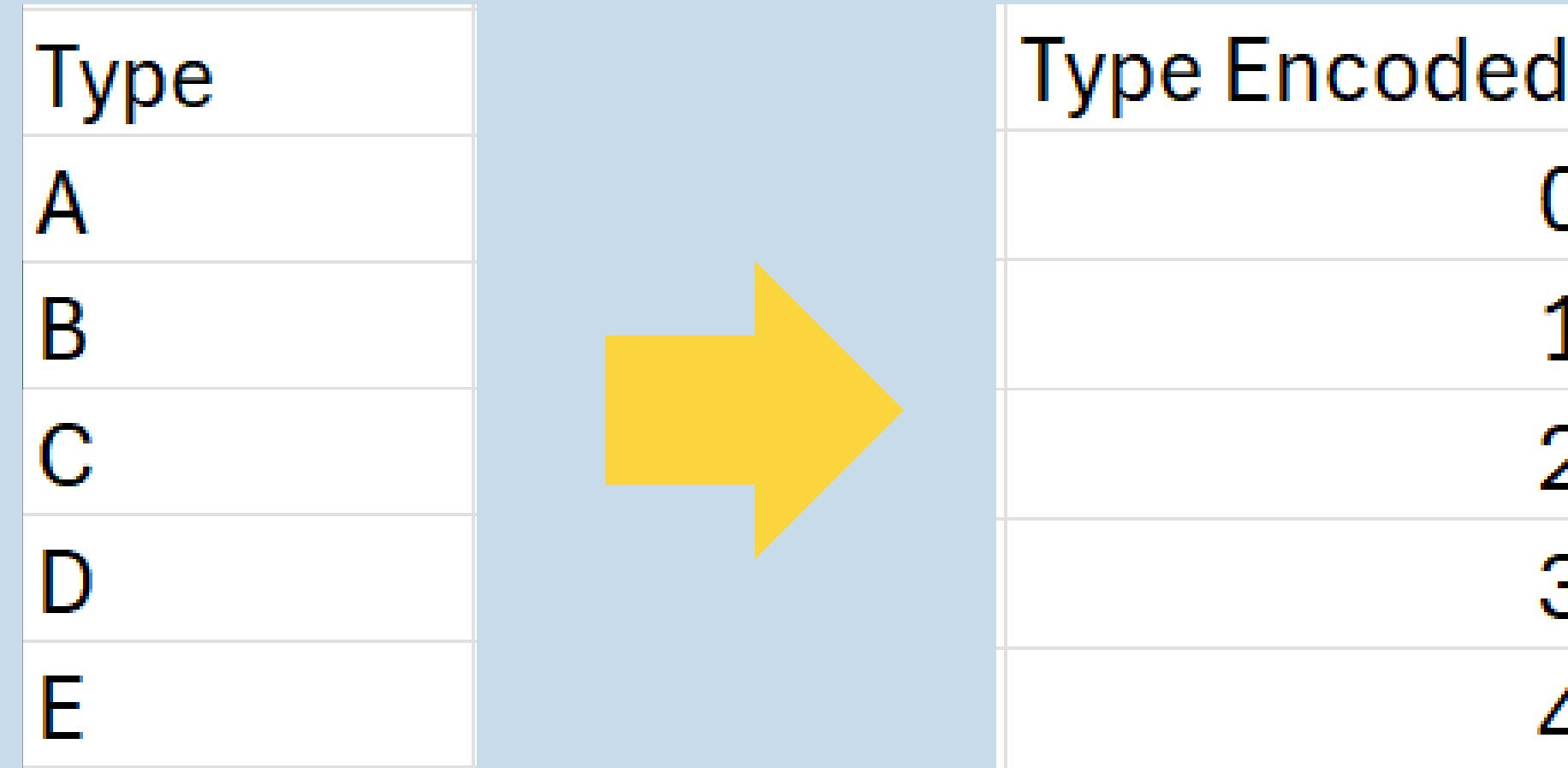
Tujuannya adalah membuat fitur (kolom) baru dari data yang sudah ada. Dengan ini, model akan mendapat lebih banyak wawasan sehingga model lebih baik dalam menentukan dropout dll

Feature Engineering yang dilakukan:

1. Label Encoding
2. Standard scaller

Feature Engineering

Label Encoding



Mengubah data kategorik ordinal menjadi data numerik

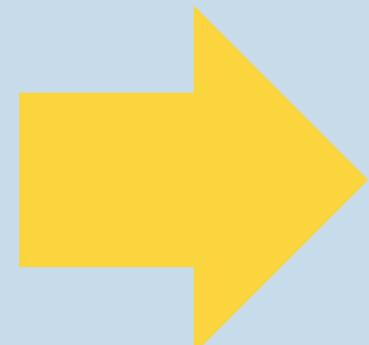
Feature Engineering

Label Encoding

```
label_encoder = LabelEncoder()  
y_resampled = label_encoder.fit_transform(y_resampled)  
✓ 0.0s
```

Kolom Target

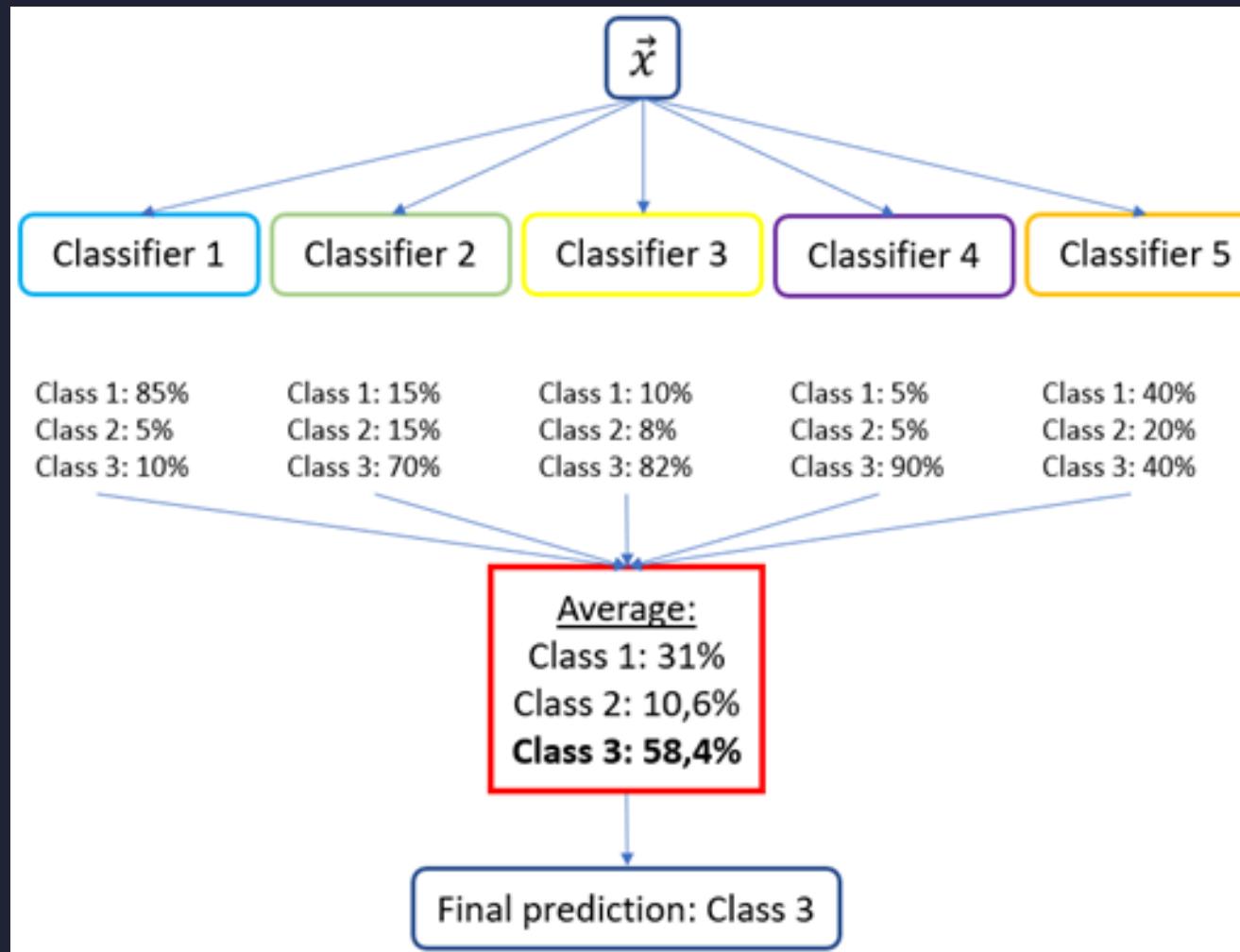
0	Graduate
1	Dropout
2	Enrolled
3	Graduate



```
array([2, 0, 0, ..., 1, 1, 1])
```

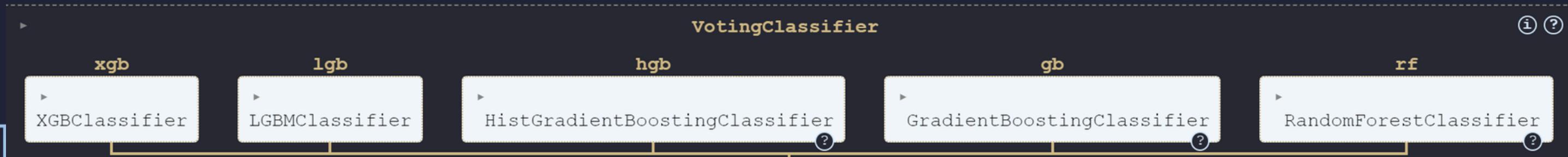
Mengubah data kategorik ordinal menjadi data numerik

Model Training



Model Training adalah proses melatih model dengan diberi data yang sudah kita transformasi.

Digunakan Voting Classifier, metode penggabungan seluruh model dan dilakukan voting untuk seluruh model yang dipakai.



Sumber: Michael Fuchs

05

Recommendation

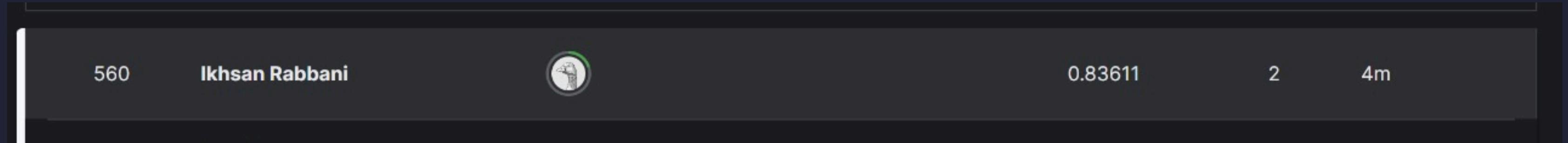
Rekomendasi apa yang dapat diberikan?

Rekomendasi

Potensi Peningkatan:

- Eksplorasi model-model lain atau teknik ensemble lainnya.
- Melakukan hyperparameter tuning yang lebih mendalam.
- Melakukan feature engineering yang berbeda terhadap kolom penting

Submission Result



Model mendapatkan **skor akurasi 0,83** karena data yang digunakan berkualitas cukup baik, dengan nilai yang hilang dan outliers yang telah diatasi. Variabel prediktor yang digunakan juga memiliki korelasi yang signifikan dengan variabel target. Model yang dipilih, seperti Random Forest dan XGBoost bisa menangani data yang kompleks. Selain itu, tuning parameter dengan GridSearchCV memastikan model bekerja optimal, dan validasi silang membantu menghindari overfitting, sehingga hasil prediksi lebih akurat.

Terima Kasih

