

UTS MACHINE LEARNING

Nama : Septya Kurnia Azzahra

NPM : 227006516065

LINK :

<https://colab.research.google.com/drive/1R9sD40JDD7JSFAr14IHAh1YMpjVCXoiL?usp=sharing>

Bagian A: Persiapan Data

1.

```
Jumlah baris dan kolom: (4424, 37)

Nilai kosong:
Marital status      0
Application mode    0
Application order   0
Course              0
Daytime/evening attendance 0
Previous qualification 0
Previous qualification (grade) 0
Nationality         0
Mother's qualification 0
Father's qualification 0
Mother's occupation 0
Father's occupation  0
Admission grade     0
Displaced           0
Educational special needs 0
Debtor              0
Tuition fees up to date 0
Gender              0
Scholarship holder  0
Age at enrollment   0
International        0
Curricular units 1st sem (credited) 0
Curricular units 1st sem (enrolled) 0
Curricular units 1st sem (evaluations) 0
Curricular units 1st sem (approved) 0
Curricular units 1st sem (grade) 0
Curricular units 1st sem (without evaluations) 0
Curricular units 2nd sem (credited) 0
Curricular units 2nd sem (enrolled) 0
Curricular units 2nd sem (evaluations) 0
Curricular units 2nd sem (approved) 0
Curricular units 2nd sem (grade) 0
Curricular units 2nd sem (without evaluations) 0
Unemployment rate   0
Inflation rate       0
GDP                  0
Target              0
dtype: int64

Jumlah data duplikat: 0

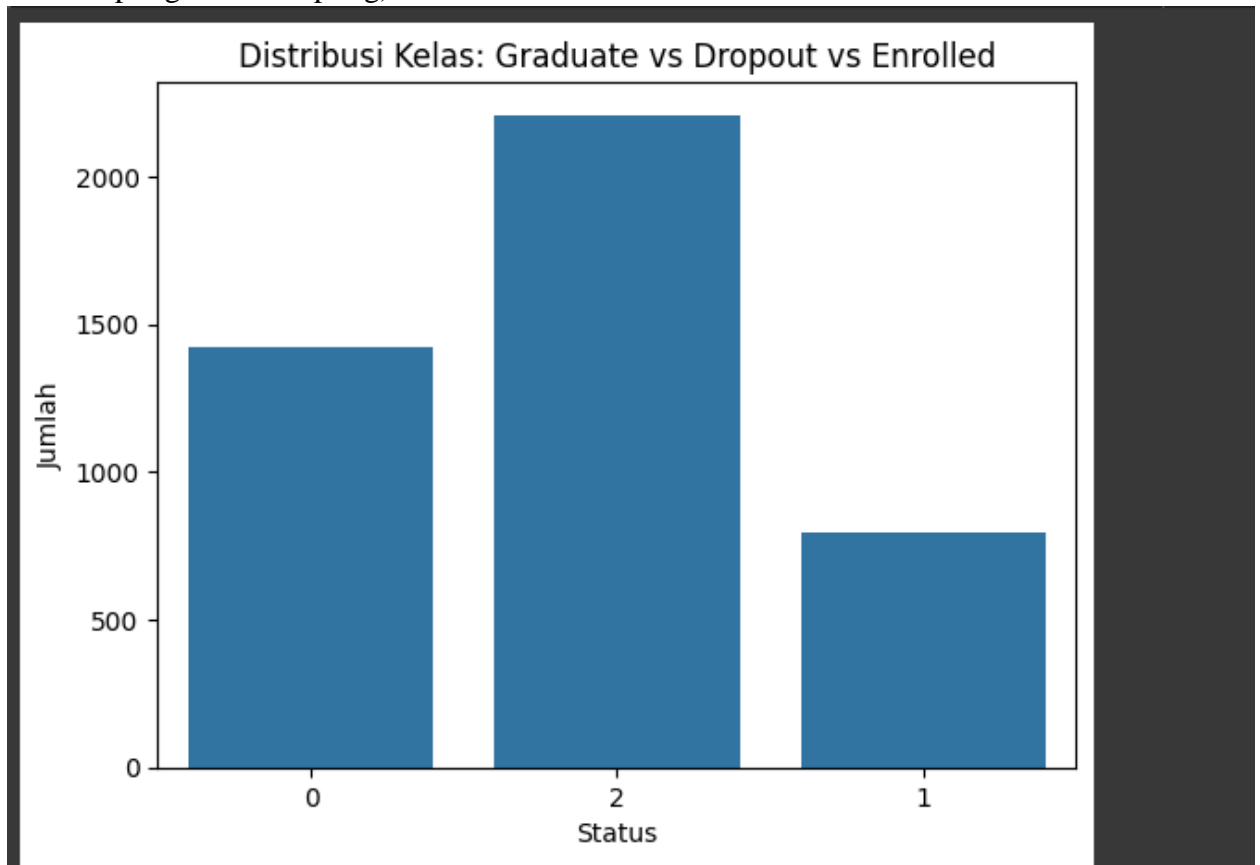
Range nilai 'Previous qualification (grade)': 95.0 - 190.0
```

Jumlah data duplikat: 0 Range nilai 'Previous qualification (grade)': 95.0 - 190.0

Nilai kosong : 0

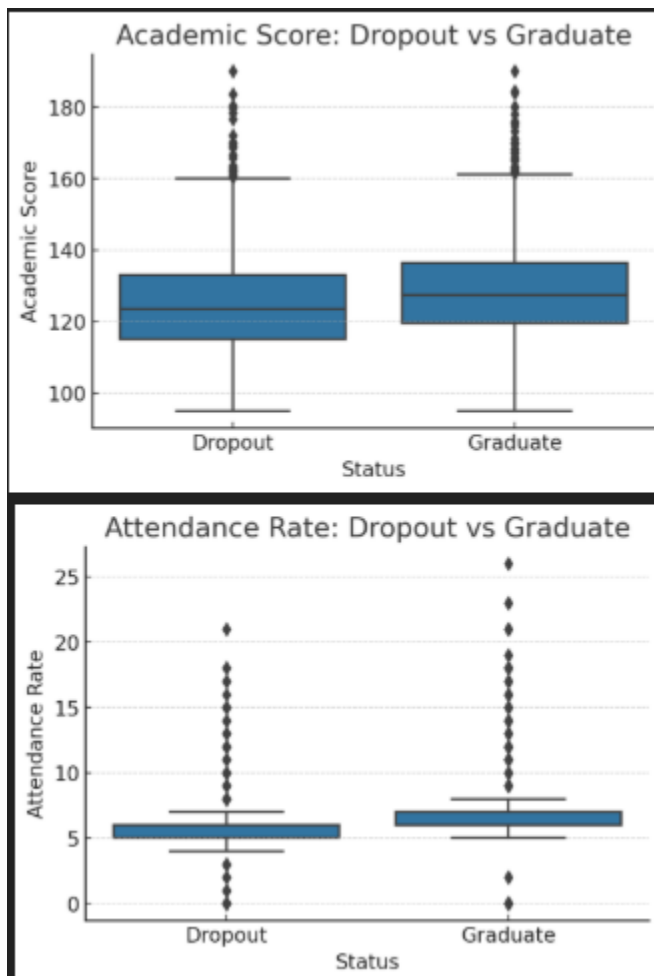
2. A. Hasil menunjukkan bahwa distribusi label **tidak seimbang**, dengan proporsi tertinggi berasal dari kelas Graduate. Ketidakseimbangan ini dapat mempengaruhi performa model

dan perlu dipertimbangkan pada saat training (misalnya dengan oversampling/undersampling).



B. Untuk memahami perbedaan karakteristik siswa **Dropout** dan **Graduate**, dilakukan analisis terhadap dua fitur utama:

- academic_score (nilai masuk)
- attendance_rate (tingkat kehadiran)
- **Academic Score – Dropout vs Graduate**
 - Siswa **Graduate** cenderung memiliki skor akademik yang lebih tinggi.
 - **Dropout** memiliki rentang yang lebih rendah dan lebih menyebar.
- **Attendance Rate – Dropout vs Graduate**
 - Tingkat kehadiran **Graduate** lebih tinggi secara umum.
 - **Dropout** cenderung memiliki kehadiran rendah, dengan outlier ekstrem.



3.



Application mode	Application order	Course	Previous qualification	Previous qualification (grade)	Mother's qualification	Father's qualification	Mother's occupation	Father's occupation	academic_score	...	Nacionality_3.4896936324176555	Nacionality_4.357531662143938	Nacionality_5.65928870673336
-0.095470	2.490896	-4.209520	-0.35023	-0.804841	-0.036018	-0.669778	-0.225661	-0.080456	0.022229	...	False	False	False
-0.209869	-0.554068	0.192580	-0.35023	2.076819	-1.189759	-1.256427	-0.301375	-0.317984	1.071926	...	False	False	False
-1.010660	2.490896	0.103404	-0.35023	-0.804841	1.117723	0.959802	-0.074233	-0.080456	-0.150419	...	False	False	False
-0.095470	0.207173	0.444115	-0.35023	-0.804841	1.181819	0.959802	-0.225661	-0.317984	-0.509526	...	False	False	False
1.162916	-0.554068	-0.408389	-0.35023	-2.473171	1.117723	1.024985	-0.074233	-0.080456	1.002867	...	False	False	False

rows x 60 columns

4. Untuk fitur numerik seperti:

- Application order
- Previous qualification (grade)
- academic_score
- attendance_rate

digunakan metode **standarisasi (Z-score normalization)** menggunakan `StandardScaler` dari `sklearn`.

	Application order	Previous qualification (grade)	academic_score	attendance_rate	
0	2.490896	-0.804841	0.022229	-2.528560	
1	-0.554068	2.076819	1.071926	-0.109105	
2	2.490896	-0.804841	-0.150419	-0.109105	
3	0.207173	-0.804841	-0.509526	-0.109105	
4	-0.554068	-2.473171	1.002867	-0.109105	