

Nama: Farrel Ghazy Affrudin

NIM: 452029611053

Kelas: TI9 A1.

Prediksi keberhasilan siswa menggunakan pendekatan  
klasifikasi pada data set student performance dan machine Learning

### # Pendahuluan.

Disini saya ingin mencoba pembuatan model machine Learning  
guna melakukan prediksi keberhasilan ~~siswa~~ siswa. Pembuatan ~~model~~ model  
bertujuan untuk melakukan prediksi dengan berdasarkan pada data set  
public yang telah di berikan. Nantinya ~~prediksi~~ Output prediksi  
yang dikeluarkan adalah "lulus" atau "gagal".

### # Tentang Data Set Public

- o Sumber = UCI Machine Learning Repository
- o Judul Dataset = Student Performance
- o Link = <https://archive.ics.uci.edu/ml/datasets/Student+Performance>
- o Deskripsi ini = Data ini berisi tentang ~~prediksi~~ prediksi keberhasilan 2 sekolah  
di Portugal, dan didalamnya ada faktor faktor demografi  
siswa dan juga catatan belajar siswa.

### # Rumusan Masalah & Tujuan

- o Masalah = Mendeteksi ~~prediksi~~ ~~hasil~~ keberhasilan siswa dan mempercepat  
pengungkapan ~~hasil~~ kegagalan siswa.
- o Tujuan = ~~untuk~~ membuat model yang dapat memprediksi apakah  
seorang siswa akan "lulus" atau "gagal" berdasarkan demografi  
dan nilai awal mereka. Sehingga bisa mengadakan  
bimbingan sebelum ujian.

## # Pendahuluan Machine Learning.

Metode yang digunakan = Supervised Learning - ~~Klasifikasi~~ Klasifikasi

Tujuan utama adalah memilih dan mengategorikan siswa

~~Menjadi~~ Menjadi 2 kelompok "Lulus" (nilai akhir  $\geq 10$ ) "Gagal" (nilai akhir  $\leq 10$ ).  
Karena ~~data~~ variabel target bersifat kategorikal dan bukan Continuous, dan data juga memiliki label awal (nilai G3) maka dari itu klasifikasi, odds metode yang paling layak dan dapat diandalkan.

Mengapa tidak Regresi? Meskipun data angka (0-10) tujuan kita bukan memprediksi angka persis nilai Akhir. ~~Melainkan~~ melainkan hanya Status kelulusannya. Regresi terlalu ketat dan sangat Rawan terhadap Outlier jika keputusan yang di ambil Binar (Lulus / tidak Lulus)

Mengapa tidak Clustering? Clustering merupakan Unsupervised Learning (data tidak memiliki label) karena data juga sudah memiliki label target yang jelas yaitu G3 kita tidak perlu menebak skor ambur.

## # Fitur dan matrik Evaluasi

o Fitur (input) = Study time (waktu belajar), failures (kegagalan masalah), absences (jumlah Absen), G1, G2 (nilai pra-sebelumnya)

o Target (label) status kelulusan 1 = "Lulus", 0 = "Gagal"

o Matrik Evaluasi: Accuracy (ketepatan ~~tesis~~ prediksi keseluruhan) dan Recall (mendeteksi siswa yang akan gagal).

## # Kesimpulan

• Dengan adanya model ini kita dapat menggunakan machine learning untuk memprediksi ~~hasil~~ kelulusan dan mencari Peluang Probabilitas untuk mencapai kelulusan itu. Jika di implementasikan di Sekolah ini juga bisa membantu meningkatkan kedisiplinan siswa dalam ujian selanjutnya. Selain itu masih ada tugas yang sudah terlihat potensi nya untuk ~~yang~~ memperkirakan banyak hal mulai dari faktor yang paling berpengaruh, Prediksi kelulusan siswa dan juga pengamalan keputusan yang lebih tepat dan akurat.

## # Lampiran

• ~~Link~~ Link Dataset : <https://archive.ics.uci.edu/ml/Student+Performance>

• Link Github ~~Project~~ <https://github.com/FarrelGhozal/Proyek-ML-Prediksi-Kelulusan.git>

Repository Proyek : <https://github.com/FarrelGhozal/Proyek-ML-Prediksi-Kelulusan.git>



# Code Program.

```
# mengimpor library  
import pandas as pd  
import numpy as np
```

```
df = pd.read_csv("Student-mat.csv", sep = ";")
```

```
df.head()
```

```
# mengecek nilai
```

```
df.info()
```

```
df.describe()
```

```
# membuat kolom target "lulus"
```

```
df['lulus'] = (df['G3'] >= 10).astype(int)
```

```
df['lulus'].value_counts()
```

```
# memilih fitur dan target.
```

```
features = [  
    'Study time',  
    'Failures',  
    'Absences',  
    'G1',  
    'G2',  
]
```

```
X = df[features]
```

```
Y = df['lulus']
```

```
# distribusi target kelulusan
```

```
Import matplotlib.pyplot as plt.
```

```
Y.value_counts().plot(kind = 'bar')
```

```
plt.title("Distribusi kelulusan")
```

```
plt.xlabel("Status (0 = tidak lulus, 1 = lulus)")
```

```
plt.ylabel("Jumlah Siswa")
```

```
plt.show()
```

\* # Analisis hubungan antar absensi dan kelulusan.

```
plt.boxplot([
```

```
df[df['lulus'] == 0] ['absences'],  
df[df['lulus'] == 1] ['absences']
```

```
], labels = ['Tidak Lulus', 'Lulus'])
```

```
plt.title("absensi vs status kelulusan")
```

```
plt.ylabel("jumlah Absensi")
```

```
plt.show.
```

```
plt.scatter(df['62'], df['63'])
```

```
plt.xlabel("Nilai 62")
```

```
plt.ylabel("Nilai Akhir 63")
```

```
plt.title("Hubungan Nilai Sebelum dan nilai Akhir")
```

```
plt.show()
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(
```

```
X, y, test_size=0.2, random_state=42, stratify=y.
```

```
)
```

```
# melatih model logit.
```

```
from sklearn.linear_model import LogisticRegression
```

```
model = LogisticRegression(max_iter=1000)
```

```
model.fit(X_train, y_train)
```

```
# Evaluasi model.
```

```
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

```
y_pred = model.predict(X_test)
```

```
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
print (Confusion-Matrix (y_test, y_pred))  
print (Classification Report (y_test, y_pred))
```

\* Fungsi untuk melakukan prediksi siswa

```
def Prediksi_kelulusan (Study time, failures, absences, G1, G2):  
    data = np.array ([[Study time, failures, absences, G1, G2]])  
    Prob = model.predict_proba (data) [0] [1]  
    Pred = model.predict (data=0) [0]  
    return  
        "prediksi: "lulus" if Pred = 1 else "tidak lulus"  
        , "Maka probabilitas lulus" round (Prob * 100, 2)  
}
```

\* Contoh peggunaan ~~function~~ Fungsi Prediksi.

```
Prediksi_kelulusan (  
    StudyTime = 2,  
    failures = 2,  
    absences = 4,  
    G1 = 12,  
    G2 = 10  
)
```