

Nama: Farrel Ghazy Affifudin

NIM: 452029611053

Kelas: T19 A1.

Prediksi keberhasilan siswa mengandalkan pendekatan
Klasifikasi pada data set student performance dan machine learning

* Pendekatan.

Dalam tugas ini akan mencoba pembuatan model machine learning
guna melakukan prediksi keberhasilan ~~siswa~~ siswa. Pembuatan ~~model~~
bertujuan untuk melakukan prediksi dengan berdasarkan pada data set
public yang telah diberikan. Nestinga ~~feature~~ Output prediksi
yang dihasilkan 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 Penjelasan isi = Data ini berisi tentang ~~pendata~~ kinerja 2 sekolah
di Portugal, dan di dalamnya ada faktor-faktor demografis
siswa dan juga ciri-ciri pelajar siswa.

* Rumusan Masalah & Tujuan

- o Masalah = mendekripsi ~~fungsi~~ keterkaitan siswa dan mempercepat
pemungutan ~~hasil~~ kegagalan siswa.
- o Tujuan = ~~tujuan~~ membuat model yang dapat memprediksi apakah
seorang siswa akan "lulus" atau "gagal" berdasarkan demografi
dan nilai awal mereka. sehingga bisa mengambil
tindakan sebelum tiba-tiba.

* Pendekatan Machine Learning.

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

Tujuan akhirnya adalah memprediksi dan mengategorikan siswa

~~Hanya~~ membuat 2 kelompok "Lulus" (Nilai Akhir ≥ 10) "Gagal" (Nilai Akhir ≤ 10).
Karena ~~data~~ Variabel target bersifat kategorikal dan berasal Continuous. dan
data juga memiliki table hasil (Nilai G3) maka dari itu klasifikasi Odd or
Metode yang paling logis dan dapat diandalkan.

Mengapa tidak Regresi Meskipun data angka (0-20) tujuan kita
bulan memprediksi angka persis nilai akhir. ~~Ma~~ Melainkan hanya
Status kelulusannya. Regresi terlalu ketat dan sangat rawan terhadap
Outlier atau pecahan yang dibil bilinear (Citus / tidak Lulus)

Mengapa tidak Clustering = Clustering merupakan ~~Supervised learning~~
(Data tidak memiliki label) karena data juga sudah memiliki hasil
target yang jelas yaitu G3 kita tidak perlu membalik sebarang.

* Fitur dan matrix evaluasi

- Fitur (Input) = Study time (waktu belajar), failures (kali gagal matkul), absences (jumlah absen), G1,G2 (nilai prak sebelumnya)
- Target (label) status kelulusan $I = \text{"Lulus"}, 0 = \text{"Gagal"}$
- Matrix Evaluasi: Accuracy (determines ~~the~~ precision keakuratan) dan Recall
(mendeteksi siswa yang lama gagal).

* Kesimpulan

- Dengan adanya model ini kita dapat menggunakan Machine Learning untuk memprediksi ~~atau~~ kelulusan dan mencari peluang probabilitas untuk mencapai kelulusan itu. B. jika di implementasikan di sekolah ini juga bisa membantu mengontrol kenyamanan siswa dalam ujian sebenarnya. Selain itu masih ada tugas yang masih sudah terlambat potensi nya untuk ~~ada~~ memperbaiki bangku hal malai dari faktor yang pada berpengaruh, prediksi kelulusan siswa dan juga pengambilan keputusan yang lebih tepat dan akurat.

* Komponen

- ~~Link~~ Link Dataset : <https://archive.ics.uci.edu/ml/Student+Performance>.
- Link Github Projek https://github.com/FarruGhoz/Projek_ML_Predksi_Kelulusan_97.

Repository Projek : https://github.com/FarruGhoz/Projek_ML_Predksi_Kelulusan_97

Kode Program.

* memanggil 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~~ fitur dan target.

features = [
 'Study time',
 'Failures',
 'absences',
 "G1",
 'G2',

]

X = df[features]

y = df['lulus']

* distribusi target ke lulusan

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 antara absensi dan kelulusan.

plt. boxplot([

df[df['absen'] == 0] ['kelulusan'],
df[df['absen'] == 1] ['kelulusan']

], 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=92, 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(Xtest)

print (Accuracy: ", accuracy_score(ytest, y_pred))

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

•
•

```
def Predict_Lulusan(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 {  
        "predisi": "lulus" if Pred == 1 else "tidak lulus",  
        "probabilitas_lulus": round((prob[1] * 100), 2)  
    }
```

y

•
•
•

Predict_Lulusan(

studyTime = 2,

failures = 2,

absences = 4,

G1 = 12,

G2 = 10

)