

PREDIKSI PELUANG MAHASISWA UNTUK LOLOS SELEKSI PROGRAM S2

Oleh : Muhammad Rizal Nurcahyo, A11.2019.11740

A. DESKRIPSI JURNAL

Paper yang saya baca berjudul “PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN PERBANDINGAN ALGORITMA C5.0 DENGAN REGRESSION LINEAR”. Sesuai dengan judul paper tersebut, didalamnya membahas mengenai perbandingan akurasi dari training dan testing data antara algoritma C5 dengan algoritma Regresi Linier. Peneliti tertarik untuk membandingkan akurasi dari algoritma C5 dan Regresi Linier karena prediksi kelulusan mahasiswa yang menerapkan algoritma lain sudah banyak dilakukan.

Adapun tahapan penelitian yang dilakukan adalah sebagai berikut :

1) Mengumpulkan Data

Peneliti menggunakan data dari fakultas teknologi dan ilmu komputer. Data Mahasiswa/Alumni yang diambil memiliki jumlah 715 baris serta 9 atribut/kolom dari tahun 2017-2020.

2) Analisis Data

Langkah yang dilakukan dalam menganalisis data:

- a) Data Preprocessing : dilakukan data cleaning untuk menghilangkan atribut yang tidak bernilai atau missing data (NaN).
- b) Data Transformation : dilakukan normalisasi data dengan tujuan agar data yang bersifat numerik berada range yang tidak terlalu jauh dan data yang bersifat kategorikal akan dikonversi menjadi numerik.
- c) Data mining : proses mencari pola data menggunakan algoritma yang sudah dipilih.

3) Implementasi Python

Melakukan pengujian data menggunakan program python.

4) Evaluasi

Tahap pengambilan keputusan dari pemecahan masalah berkaitan dengan penggunaan hasil data mining.

Hasil :

Hasil perbandingan akurasi training dan testing pada algoritma C5 dengan Regresi Linier adalah sebagai berikut :

Model	Akurasi Training (%)	Akurasi Testing (%)
Algoritma C5	98.85	93.72
Regresi Linier	33.31	40.30

B. PROGRES CAPAIAN DATASET

- 1) Jenis dataset yang digunakan : Public (sumber : Kaggle.com)
- 2) Dataset yang digunakan : Graduate Admission Prediction
- 3) Jumlah dataset : 1
- 4) Format dataset : CSV
- 5) Atribut dataset :

No.	Nama Atribut	Format Dataset
1	Serial No.	Numerik
2	GRE Score	Numerik
3	TOEFL Score	Numerik
4	University Rating	Numerik
5	SOP	Numerik
6	LOR	Numerik
7	CGPA	Numerik
8	Research	Numerik
9	Chance of Admit	Numerik

C. PROGRESS CAPAIAN PROSES EKSPERIMEN

- 1) Mencoba untuk melakukan Preprocessing Data

```

In [89]: from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer.fit(x[:, 1:3])
x[:, 1:3] = imputer.transform(x[:, 1:3])

In [90]: x
Out[90]: array([[ 1. , 337. , 118. , ..., 4.5 , 9.65, 1. ],
 [ 2. , 324. , 107. , ..., 4.5 , 8.87, 1. ],
 [ 3. , 316. , 104. , ..., 3.5 , 8. , 1. ],
 ...,
 [498. , 330. , 120. , ..., 5. , 9.56, 1. ],
 [499. , 312. , 103. , ..., 5. , 8.43, 0. ],
 [500. , 327. , 113. , ..., 4.5 , 9.04, 0. ]])

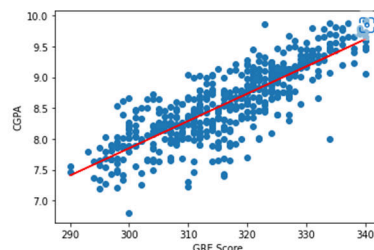
In [77]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state = 1)

In [78]: x_train
Out[78]: array([[239. , 310. , 104. , ..., 3.5 , 8.37, 0. ],
 [439. , 318. , 110. , ..., 3.5 , 8.54, 1. ],
 [476. , 300. , 101. , ..., 2.5 , 7.88, 0. ],
 ...,
 [397. , 325. , 107. , ..., 3.5 , 9.11, 1. ],
 [236. , 326. , 111. , ..., 4. , 9.23, 1. ],
 [ 38. , 300. , 105. , ..., 2. , 7.8 , 0. ]])

```

2) Mencoba memodelkan data dengan linear regression sederhana

```
Out[98]: [<matplotlib.lines.Line2D at 0x215f8253eb0>]
```



```

In [61]: reg = linear_model.LinearRegression()
reg.fit(dataset[['GRE_Score']], dataset.CGPA)

```

```
Out[61]: LinearRegression()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [62]: reg.predict([[340]])
```

D:\datamining\app\env\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
warnings.warn(

```
Out[62]: array([9.61691085])
```

```
In [63]: reg.coef_
```

```
Out[63]: array([0.04422266])
```

```
In [64]: reg.intercept_
```

```
Out[64]: -5.418795105597956
```

Saya mencoba untuk menampilkan Kemungkinan lolos seleksi menggunakan atribut GRE Score (Skor tes masuk pascasarjana) dan CGPA (Rata-rata IPK). Terlihat grafik naik kekanan yang berarti semakin tinggi skor maka kemungkinan diterima S2 lebih besar.

