

**LAPORAN TUGAS BESAR PRAKTIKUM KECERDASAN BUATAN**  
**“KLASIFIKASI GAJI DAN PENGELUARAN MENGGUNAKAN METODE REGRESSION**  
**LINEAR”**



Disusun Oleh:

Dikri Ramadhan      2106005

M. Algifari          2106014

Kelas A

**JURUSAN ILMU KOMPUTER**  
**PROGRAM STUDI TEKNIK INFORMATIKA**  
**INSTITUT TEKNOLOGI GARUT**  
**2023**

## **KATA PENGANTAR**

Puji syukur kami panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga kami dapat menyelesaikan laporan tugas besar praktikum Kecerdasan Buatan ini. Laporan ini disusun dalam rangka memenuhi salah satu tugas akademik yang diberikan oleh dosen pengampu dan asisten dosen mata kuliah Kecerdasan Buatan.

Tugas besar ini bertujuan untuk mengimplementasikan dan menerapkan metode klasifikasi menggunakan regresi linear. Regresi linear merupakan salah satu teknik dalam analisis statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan variabel independen. Dalam konteks klasifikasi, regresi linear dapat digunakan untuk memprediksi nilai variabel target berdasarkan variabel-variabel prediktor yang ada.

Dalam penyusunan laporan ini, kami menyadari bahwa hasil laporan praktikum ini masih jauh dari kata sempurna. Oleh karena itu, kami sangat mengharapkan kritik dan saran yang membangun guna perbaikan kedepannya.

Akhir kata, kami mengucapkan terima kasih kepada seluruh pihak yang telah membantu dan mendukung dalam penyelesaian Tugas Besar ini. Semoga laporan ini bermanfaat dan dapat memberikan kontribusi dalam pemahaman dan pengembangan ilmu.

Garut, Mei 2023

Penyusun

# DAFTAR ISI

<b>KATA PENGANTAR.....</b>	<b>i</b>
<b>DAFTAR ISI.....</b>	<b>ii</b>
<b>DAFTAR GAMBAR.....</b>	<b>iii</b>
1.1. Identifikasi Masalah.....	1
1.2. Tujuan Masalah.....	2
1.3. Batasan Masalah .....	3
1.4. Manfaat .....	4
<b>BAB II GAP ANALYSIS .....</b>	<b>6</b>
2.1. Implementasi metode regresi .....	6
2.2. implementasi metode naive bayes.....	7
2.3. Implementasi metode knn .....	8
2.4. Hasil pengujian akurasi metode naive bayes , knn dan regresi linear.....	9
2.6. Pengumpulan Data .....	11
2.7. Preparation .....	11
2.8. Modelling & Training .....	13
2.9. Akurasi.....	15
2.10. Evaluasi.....	15
<b>BAB III KESIMPULAN DAN SARAN .....</b>	<b>16</b>
3.1 Kesimpulan .....	16
3.2 Saran .....	16
<b>DAFTAR PUSTAKA .....</b>	<b>17</b>

## DAFTAR GAMBAR

Gambar 1. 1 Pengumpulan data .....	11
Gambar 1. 2 Preparation data.....	11
Gambar 1. 3 Menampilka data.....	12
Gambar 1. 4 kolom dan baris .....	12
Gambar 1. 5 Data dikonversi ke array .....	12
Gambar 1. 6 Pembuatan Modelling & Training.....	13
Gambar 1. 7 Data points .....	14
Gambar 1. 8 Regression line .....	14
Gambar 1. 9 Pembuatan grafik .....	14
Gambar 1. 10 Grafik linear regression.....	15
Gambar 1. 11 Pembuatan Akurasi.....	15
Gambar 1. 12 Pembuatan Evaluasi .....	15

# **BAB I**

## **LATAR BELAKANG**

Klasifikasi data merujuk pada proses pengelompokan data menjadi kategori atau kelas yang memiliki karakteristik atau atribut yang serupa. Dalam konteks ini, klasifikasi data digunakan untuk mengorganisir dan menyusun data agar lebih mudah dikelola, dipahami, dan dieksploitasi. Tujuan utama klasifikasi data adalah untuk mengidentifikasi pola, hubungan, atau karakteristik yang tersembunyi dalam dataset. Proses klasifikasi data melibatkan penggunaan algoritma atau metode yang mempelajari pola dan tren dalam data yang ada, serta menghasilkan model yang dapat digunakan untuk mengklasifikasikan data baru. Algoritma pembelajaran mesin seperti Naive Bayes, Decision Tree, Random Forest, dan Support Vector Machines sering digunakan untuk melakukan klasifikasi data.

Regresi linear adalah teknik analisis data yang memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui. Secara matematis memodelkan variabel yang tidak diketahui atau tergantung dan variabel yang dikenal atau independen sebagai persamaan linier.

### **1.1. Identifikasi Masalah**

Masalah yang mungkin terkait dengan klasifikasi gaji dan pengeluaran menggunakan metode regresi linear adalah sebagai berikut:

1. Asumsi linieritas: Metode regresi linear bergantung pada asumsi bahwa hubungan antara variabel independen (misalnya pendapatan) dan variabel dependen (misalnya pengeluaran) bersifat linier. Namun, jika hubungan tersebut tidak linier, model regresi linear mungkin tidak memberikan hasil yang akurat.
2. Multikolinearitas: Multikolinearitas terjadi ketika ada korelasi yang kuat antara variabel independen. Dalam kasus ini, metode regresi linear dapat menghasilkan koefisien regresi yang tidak stabil atau tidak dapat diinterpretasikan dengan baik. Hal ini dapat menyebabkan kesalahan dalam klasifikasi gaji dan pengeluaran.
3. Outlier: Adanya outlier dalam data dapat mempengaruhi model regresi linear. Outlier adalah observasi yang secara signifikan berbeda dari pola umum data.

Outlier dapat memiliki pengaruh yang besar pada hasil regresi, sehingga dapat mempengaruhi klasifikasi gaji dan pengeluaran.

4. Heteroskedastisitas: Heteroskedastisitas terjadi ketika varians dari kesalahan (error) dalam model regresi tidak konstan. Jika heteroskedastisitas terjadi, interpretasi koefisien regresi dan uji signifikansi statistik dapat menjadi tidak valid.
5. Data tidak representatif: Klasifikasi gaji dan pengeluaran yang akurat membutuhkan data yang representatif dari populasi yang ingin diprediksi. Jika data yang digunakan tidak mewakili populasi dengan baik, model regresi linear mungkin menghasilkan hasil yang tidak akurat atau tidak dapat digeneralisasi.
6. Kesalahan pengukuran: Kesalahan pengukuran dalam variabel independen atau dependen dapat mempengaruhi hasil regresi. Jika terdapat kesalahan pengukuran yang signifikan, model regresi linear mungkin tidak memberikan estimasi yang akurat untuk klasifikasi gaji dan pengeluaran.

Penting untuk mempertimbangkan masalah-masalah ini dan melihat apakah asumsi dan persyaratan metode regresi linear cocok dengan data dan konteks yang spesifik sebelum menggunakan metode ini untuk klasifikasi gaji dan pengeluaran.

## 1.2. Tujuan Masalah

Tujuan dari klasifikasi gaji dan pengeluaran dengan metode regresi linear adalah untuk mengidentifikasi dan memprediksi hubungan antara gaji seseorang dengan jumlah pengeluaran mereka. Dengan menggunakan metode regresi linear, kita dapat membangun model matematis yang menggambarkan hubungan linier antara variabel gaji dan pengeluaran.

Berikut adalah beberapa tujuan klasifikasi gaji dan pengeluaran dengan metode regresi linear:

### 1. Pemahaman hubungan

Tujuan utama adalah untuk memahami hubungan antara gaji dan pengeluaran. Dengan menganalisis data gaji dan pengeluaran yang ada, kita dapat mengidentifikasi pola dan tren yang menggambarkan korelasi antara kedua variabel tersebut.

### 2. Prediksi pengeluaran

Dengan membangun model regresi linear yang baik, kita dapat menggunakan variabel gaji seseorang sebagai prediktor untuk memperkirakan jumlah pengeluaran mereka di masa depan. Ini dapat memberikan wawasan berharga bagi

individu, bisnis, atau lembaga keuangan dalam perencanaan keuangan dan pengambilan keputusan.

3. Evaluasi efek gaji

Dengan melakukan analisis regresi linear, kita dapat mengukur seberapa besar pengaruh gaji terhadap pengeluaran. Hal ini dapat membantu dalam mengevaluasi sejauh mana peningkatan atau penurunan gaji akan mempengaruhi pola pengeluaran individu atau kelompok.

4. Identifikasi anomali

Dalam beberapa kasus, analisis regresi linear dapat membantu dalam mengidentifikasi anomali atau situasi yang tidak biasa dalam data gaji dan pengeluaran. Dengan melihat perbedaan antara nilai aktual dan nilai yang diprediksi oleh model regresi, kita dapat mengenali situasi yang tidak sesuai dengan pola umum dan melakukan investigasi lebih lanjut.

5. Pengembangan strategi keuangan

Informasi yang dihasilkan dari analisis regresi linear dapat digunakan untuk mengembangkan strategi keuangan yang lebih efektif. Misalnya, dengan memahami bagaimana pengeluaran berubah seiring dengan pertumbuhan gaji, individu atau lembaga keuangan dapat merancang rencana tabungan atau alokasi anggaran yang lebih baik.

6. Pengambilan keputusan

Hasil analisis regresi linear dapat digunakan sebagai dasar untuk pengambilan keputusan yang lebih informasional. Misalnya, perusahaan dapat menggunakan model regresi untuk menentukan tingkat kenaikan gaji yang dapat mempengaruhi pengeluaran karyawan secara signifikan tanpa menyebabkan ketidakseimbangan keuangan.

### 1.3. Batasan Masalah

Batasan masalah terkait dengan penggunaan metode regresi linear dalam klasifikasi gaji dan pengeluaran adalah:

1. Metode ini hanya cocok untuk hubungan linier antara pendapatan dan pengeluaran.
2. Penyimpangan dari asumsi dasar metode regresi linear dapat mempengaruhi hasil dan interpretasi model.
3. Klasifikasi dapat terbatas oleh kualitas dan representativitas data yang digunakan.
4. Tidak mempertimbangkan faktor non-linier dan efek interaksi yang kompleks.

5. Bergantung pada ketersediaan dan kualitas fitur yang relevan.
6. Tidak memperhitungkan faktor kontekstual dan non-numerik yang mempengaruhi gaji dan pengeluaran.

Dalam penggunaan metode regresi linear, penting untuk memahami dan mengatasi batasan-batasan ini jika diperlukan.

#### 1.4. Manfaat

Klasifikasi gaji dan pengeluaran dengan metode regresi linear dapat memberikan sejumlah manfaat, baik dalam konteks analisis data maupun pengambilan keputusan. Berikut adalah beberapa manfaat utama dari penggunaan metode tersebut:

##### 1. Prediksi Pengeluaran

Dengan menggunakan data gaji sebagai variabel independen dan pengeluaran sebagai variabel dependen, metode regresi linear dapat digunakan untuk memprediksi pengeluaran berdasarkan tingkat gaji seseorang. Hal ini dapat berguna dalam perencanaan keuangan pribadi atau dalam pengambilan keputusan bisnis terkait alokasi anggaran.

##### 2. Analisis Keterkaitan

Regresi linear juga dapat digunakan untuk menganalisis keterkaitan antara gaji dan pengeluaran. Dengan melihat koefisien regresi, kita dapat mengetahui seberapa besar pengaruh variabel gaji terhadap pengeluaran. Misalnya, apakah ada hubungan linear positif di mana semakin tinggi gaji seseorang, semakin tinggi pula pengeluarannya.

##### 3. Identifikasi Outlier

Melalui metode regresi linear, kita dapat mengidentifikasi adanya outlier atau data yang menyimpang dari pola umum. Outlier dapat memberikan informasi penting, seperti adanya pengeluaran yang tidak wajar dibandingkan dengan gaji yang dimiliki. Identifikasi ini dapat membantu dalam mendeteksi masalah dalam pengeluaran atau melakukan tindakan pencegahan.

##### 4. Pengambilan Keputusan

Dengan menggunakan model regresi linear yang telah terlatih, kita dapat melakukan pengambilan keputusan yang lebih terinformasi. Misalnya, jika seseorang memiliki gaji tertentu, model dapat memberikan perkiraan pengeluaran yang diharapkan berdasarkan data yang telah ada. Hal ini dapat membantu dalam perencanaan anggaran, investasi, atau penentuan harga produk dan layanan.



## 5. Evaluasi Kebijakan Keuangan

Metode regresi linear juga dapat digunakan untuk mengevaluasi kebijakan keuangan yang telah dilakukan. Dengan menganalisis hubungan antara gaji dan pengeluaran sebelum dan sesudah kebijakan diimplementasikan, kita dapat mengevaluasi efektivitas kebijakan tersebut dan melihat apakah ada perubahan yang signifikan dalam pola pengeluaran.

## **BAB II**

### **GAP ANALYSIS**

#### 2.1. Implementasi metode regresi

langkah kami yang dilakukan pada tahapan ini yaitu :

##### 1. Persiapan Data:

- Kumpulkan data yang mencakup variabel gaji (variabel independen) dan pengeluaran (variabel dependen) dari sampel individu.
- Pastikan data tersebut telah dibersihkan, tidak ada nilai yang hilang, dan siap untuk dianalisis.

##### 2. Pembagian Data:

- Bagi data menjadi dua subset: data pelatihan (training data) dan data uji (test data).
- Data pelatihan akan digunakan untuk membangun model regresi linear, sedangkan data uji akan digunakan untuk menguji kinerja model yang dikembangkan.

##### 3. Pemilihan Fitur:

- Pilih fitur yang relevan untuk digunakan dalam model regresi linear, misalnya hanya menggunakan gaji sebagai fitur independen.

##### 4. Pembangunan Model:

- Terapkan metode regresi linear pada data pelatihan dengan menggunakan algoritma seperti Ordinary Least Squares (OLS) atau Metode Gradient Descent.
- Sesuaikan model dengan data pelatihan untuk menemukan koefisien regresi yang optimal.

##### 5. Evaluasi Model:

- Gunakan data uji untuk menguji kinerja model yang dikembangkan.
- Hitung metrik evaluasi seperti Mean Squared Error (MSE) atau R-squared untuk mengevaluasi sejauh mana model dapat mengklasifikasikan pengeluaran berdasarkan gaji.

##### 6. Interpretasi Model:

- Analisis koefisien regresi untuk memahami hubungan antara gaji dan pengeluaran.

- Jika koefisien regresi positif, itu menunjukkan bahwa semakin tinggi gaji, semakin tinggi juga pengeluaran. Sebaliknya, jika koefisien negatif, itu menunjukkan bahwa semakin tinggi gaji, semakin rendah pengeluaran.

#### 7. Prediksi:

- Setelah model telah diuji dan diverifikasi, Anda dapat menggunakan model tersebut untuk melakukan prediksi pengeluaran berdasarkan gaji yang baru.
- Masukkan gaji baru ke dalam model regresi linear untuk mendapatkan perkiraan pengeluaran yang sesuai.
- Penting untuk mencatat bahwa implementasi metode regresi linear ini bersifat umum. Setiap implementasi spesifik dapat berbeda tergantung pada tujuan, data yang digunakan, dan algoritma yang diterapkan.

### 2.2. implementasi metode naive bayes

langkah yang dilakukan pada tahapan ini yaitu :

1. Persiapan Data: Siapkan dataset yang berisi atribut-atribut terkait pengeluaran gaji, seperti pendapatan, status pernikahan, tingkat pendidikan, pengalaman kerja, dan lainnya. Pastikan dataset sudah terstruktur dan bersih dari missing values atau outliers.
2. Preprocessing Data: Lakukan preprocessing data yang diperlukan, seperti mengkonversi atribut kategorikal menjadi representasi numerik, normalisasi data, atau menghapus atribut yang tidak relevan.
3. Pembagian Data: Bagi dataset menjadi dua bagian, yaitu data latih (training data) dan data uji (test data). Data latih akan digunakan untuk melatih model Naive Bayes, sementara data uji akan digunakan untuk menguji performa model.
4. Perhitungan Probabilitas: Hitung probabilitas masing-masing atribut dan kelas target (misalnya, pengeluaran rendah, sedang, atau tinggi) dalam dataset latih. Untuk atribut numerik, gunakan distribusi probabilitas yang sesuai, seperti Gaussian atau distribusi lainnya.
5. Perhitungan Probabilitas Bersyarat: Hitung probabilitas bersyarat dari atribut-atribut terhadap kelas target. Misalnya, jika atribut adalah pendapatan dan kelas target adalah pengeluaran, hitung probabilitas bersyarat pengeluaran tertentu terjadi dengan mempertimbangkan nilai pendapatan yang terkait.

6. **Prediksi:** Gunakan model Naive Bayes yang telah dilatih untuk melakukan prediksi pengeluaran gaji pada data uji. Hitung probabilitas kelas target untuk setiap data uji dan pilih kelas dengan probabilitas tertinggi sebagai prediksi.
7. **Evaluasi Model:** Evaluasi performa model Naive Bayes menggunakan metrik evaluasi seperti akurasi, presisi, recall, atau F1-score. Bandingkan prediksi model dengan nilai sebenarnya pada data uji untuk mengukur keakuratan dan keandalan model.
8. **Peningkatan Model:** Jika model memiliki performa yang kurang memuaskan, pertimbangkan untuk melakukan tuning parameter atau menerapkan teknik pemrosesan data lanjutan seperti feature selection atau feature engineering.
9. **Implementasi dan Prediksi:** Setelah model telah dilatih dan dievaluasi, Anda dapat mengimplementasikan model Naive Bayes untuk memprediksi pengeluaran gaji pada data baru yang belum dikenal.

Penting untuk memahami bahwa implementasi metode Naive Bayes ini memerlukan pemahaman yang baik tentang konsep dasar probabilitas dan teori Bayes.

### 2.3. Implementasi metode knn

langkah yang dilakukan pada tahapan ini yaitu :

1. **Persiapan Data:** Kumpulkan data yang terdiri dari pengeluaran dan gaji sebagai fitur, serta label klasifikasi yang menunjukkan apakah pengeluaran tersebut tinggi atau rendah berdasarkan batasan yang ditentukan.
2. **Normalisasi Data:** Jika skala pengukuran fitur berbeda, normalisasikan data agar setiap fitur memiliki skala yang serupa. Misalnya, menggunakan metode normalisasi seperti Min-Max Scaling atau Z-Score Scaling.
3. **Bagi Data:** Bagi dataset menjadi dua subset, yaitu set pelatihan (training set) dan set pengujian (test set). Set pelatihan digunakan untuk melatih model KNN, sedangkan set pengujian digunakan untuk menguji kinerja model.
4. **Tentukan Nilai K:** Pilih nilai K yang tepat untuk metode KNN. Nilai K menentukan jumlah tetangga terdekat yang akan digunakan untuk melakukan klasifikasi. Nilai K dapat dipilih melalui validasi silang atau metode lain yang sesuai.

5. Latih Model: Gunakan set pelatihan untuk melatih model KNN. Model KNN akan "mengingat" data pelatihan dan menghitung jarak antara setiap sampel dalam set pengujian dengan sampel pelatihan.
6. Hitung Jarak: Gunakan metrik jarak (misalnya Euclidean distance atau Manhattan distance) untuk mengukur jarak antara setiap sampel dalam set pengujian dengan sampel pelatihan.
7. Pilih Tetangga Terdekat: Pilih K tetangga terdekat dari setiap sampel dalam set pengujian berdasarkan jarak yang dihitung sebelumnya.
8. Klasifikasi Mayoritas: Tentukan label klasifikasi mayoritas dari tetangga terdekat. Misalnya, jika sebagian besar tetangga termasuk dalam kategori "pengeluaran rendah", sampel tersebut akan diklasifikasikan sebagai "pengeluaran rendah".
9. Evaluasi Model: Gunakan set pengujian untuk menguji kinerja model KNN. Hitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengevaluasi seberapa baik model melakukan klasifikasi. Tuning dan Validasi: Jika kinerja model tidak memuaskan, pertimbangkan untuk melakukan tuning parameter, seperti nilai K, atau menggunakan metode validasi silang untuk mengoptimalkan kinerja model.
10. Setelah melalui langkah-langkah di atas, Anda akan memiliki model KNN yang dilatih untuk melakukan klasifikasi pengeluaran gaji berdasarkan data yang ada. Model ini dapat digunakan untuk memprediksi pengeluaran berdasarkan gaji baru yang diberikan.

#### 2.4. Hasil pengujian akurasi metode naive bayes , knn dan regresi linear

Hasil uji dan evaluasi algoritme akan ditampilkan menggunakan perhitungan confusion matrix pada setiap algoritme yang diuji. Tahap evaluasi dilakukan untuk mengetahui seberapa baik kedua algoritme diterapkan pada dataset BBJ 2020. Dengan adanya confusion matrix, perbandingan antara algoritme dapat dilakukan dengan mudah.

Confusion matrix adalah tabel yang digunakan untuk menampilkan performa klasifikasi dari suatu model. Biasanya, confusion matrix terdiri dari empat kategori hasil klasifikasi, yaitu true positive (TP), true negative (TN), false positive (FP), dan false negative (FN).

Dengan menggunakan confusion matrix, kita dapat menghitung metrik evaluasi seperti precision, recall, dan akurasi:

- Precision =  $TP / (TP + FP)$ : Mengukur sejauh mana prediksi positif yang dilakukan oleh model adalah benar.
- Recall =  $TP / (TP + FN)$ : Mengukur sejauh mana model mampu menemukan semua kasus positif.
- Akurasi =  $(TP + TN) / (TP + TN + FP + FN)$ : Mengukur sejauh mana model secara keseluruhan mampu mengklasifikasikan data dengan benar.

Dengan menggunakan hasil perhitungan confusion matrix dan metrik evaluasi ini, kita dapat membandingkan kinerja Naive Bayes, KNN, dan Regresi Linear pada dataset BBJ 2020. Hal ini akan memberikan pemahaman yang lebih baik tentang seberapa baik algoritme-algoritme ini bekerja dalam konteks klasifikasi pengeluaran gaji pada dataset tersebut.

## 2.5. Perbandingan metode Naive Bayes , KNN dan Regresi Linear

pada tahapan dini dilakukan perbandingan nilai precision, recall dan akurasi pada masing-masing algoritma disetiap kasus. Setelah itu, dilakukan rekapitulasi hasil dari masing-masing algoritma sehingga dapat diambil kesimpulan mengenai algoritma terbaik untuk setiap kasus.

Judul	Akurasi	kesenjangan
Penerapan metode Regresi Linear untuk klasifikasi pengeluaran gaji	33%	Berdasarkan hasil penelitian kami yang menggunakan metode regresi linear dengan menghasilkan akurasi 33%. Dapat disimpulkan bahwa nilai akurasi penelitian kami lebih kecil dari hasil penelitian jurnal lain, dengan selisih nilai akurasi sebesar 47%
Penerapan metode Naive bayes untuk klasifikasi pengeluaran gaji	80%	Berdasarkan hasil penelitian kami yang menggunakan metode Naive Bayes dengan menghasilkan akurasi 80%. Dapat disimpulkan bahwa nilai akurasi penelitian kami lebih kecil dari

		hasil penelitian jurnal lain, dengan selisih nilai akurasi sebesar 47%
Penerapan metode KNN untuk klasifikasi pengeluaran gaji	78,97%	Berdasarkan hasil penelitian kami yang menggunakan metode KNN dengan menghasilkan akurasi 78,97%. Dapat disimpulkan bahwa nilai akurasi penelitian kami lebih kecil kecil dari hasil penelitian jurnal lain, dengan selisih nilai akurasi sebesar 45,97%

## 2.6. Pengumpulan Data



```

# Import library yang akan digunakan
# Matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.cluster import KMeans
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.metrics import accuracy_score

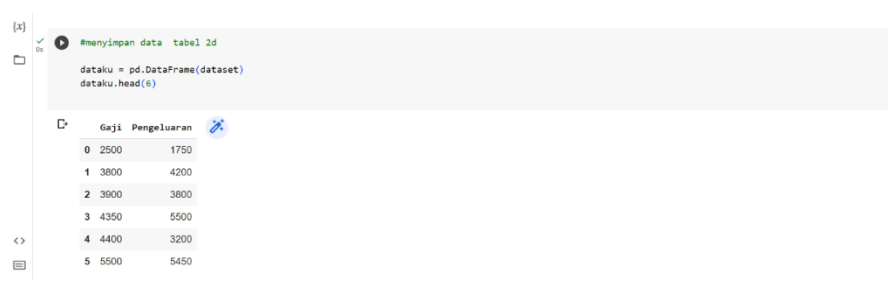
# Menetapkan data dan memanggil dataset
dataset = pd.read_csv('keuangan.csv')
dataset.keys()

Index(['Gaji', 'Pengeluaran'], dtype='object')

```

Gambar 1. 1 Pengumpulan data

## 2.7. Preparation



```

# Menyimpan data tabel 3d
dataku = pd.DataFrame(dataset)
dataku.head(6)

```

	Gaji	Pengeluaran
0	2500	1750
1	3800	4200
2	3900	3800
3	4350	5500
4	4400	3200
5	5500	5450

Gambar 1. 2 Preparation data

Untuk menampilkan frame perintah yang di gunakan seperti gambar dibawah ini

```
# Menampilkan dataframe
print(dataku)
```

	Gaji	Pengeluaran
0	2500	1750
1	3800	4200
2	3900	3800
3	4350	5500
4	4400	3200
5	5500	5450
6	5600	5950
7	5750	4100
8	6850	6050
9	6900	8500
10	7250	9500
11	7350	6050
12	7500	8500
13	7800	9500
14	8200	8300
15	8500	6500
16	8550	8400
17	8750	6000
18	9100	10500
19	9100	8500

Gambar 1. 3 Menampilka data

Untuk mengetahui kolom dan baris caranya seperti gambar dibawah ini

```
#mengetahui jumlah klon dan baris
print(dataku.shape)
```

```
(20, 2)
```

```
# print(dataku.describe)
```

```
<bound method NDFrame.describe of      Gaji  Pengeluaran
0      2500      1750
1      3800      4200
2      3900      3800
3      4350      5500
4      4400      3200
5      5500      5450
6      5600      5950
7      5750      4100
8      6850      6050
9      6900      8500
10     7250      9500
11     7350      6050
12     7500      8500
13     7800      9500
14     8200      8300
15     8500      6500
16     8550      8400
17     8750      6000
18     9100     10500
19     9100      8500
```

Gambar 1. 4 kolom dan baris

Data yang di dapat lalu dikonversikan kedalam data berbentuk array

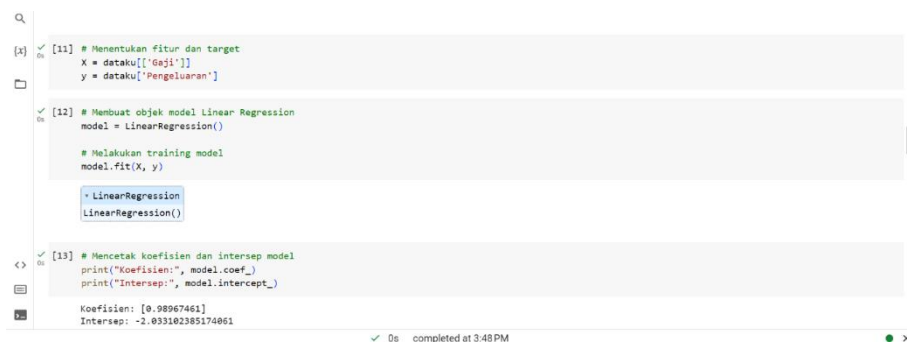
```
#konversi ke data array
X = np.asarray(dataset)
print(X)
```

```
[[ 2500  1750]
 [ 3800  4200]
 [ 3900  3800]
 [ 4350  5500]
 [ 4400  3200]
 [ 5500  5450]
 [ 5600  5950]
 [ 5750  4100]
 [ 6850  6050]
 [ 6900  8500]
 [ 7250  9500]
 [ 7350  6050]
 [ 7500  8500]
 [ 7800  9500]
 [ 8200  8300]
 [ 8500  6500]
 [ 8550  8400]
 [ 8750  6000]
 [ 9100 10500]
 [ 9100  8500]]
```

Gambar 1. 5 Data dikonversi ke array



## 2.8. Modelling & Training



```
[X] ✓ [11] # Menentukan fitur dan target
0s X = dataku['Gaji']
    y = dataku['Pengeluaran']

[12] ✓ [12] # Membuat objek model Linear Regression
0s model = LinearRegression()

    # Melakukan training model
    model.fit(X, y)

    LinearRegression
    LinearRegression()

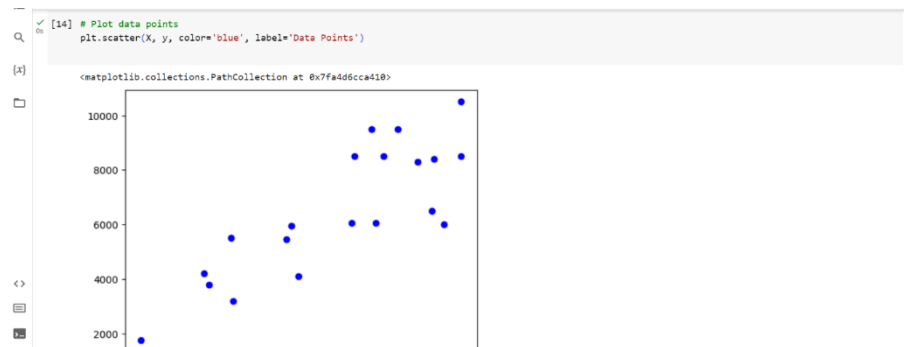
[13] ✓ [13] # Mencetak koefisien dan intersep model
0s print("Koefisien:", model.coef_)
    print("Intersep:", model.intercept_)

Koefisien: [0.98967461]
Intersep: -2.033102385174861

✓ 0s completed at 3:48 PM
```

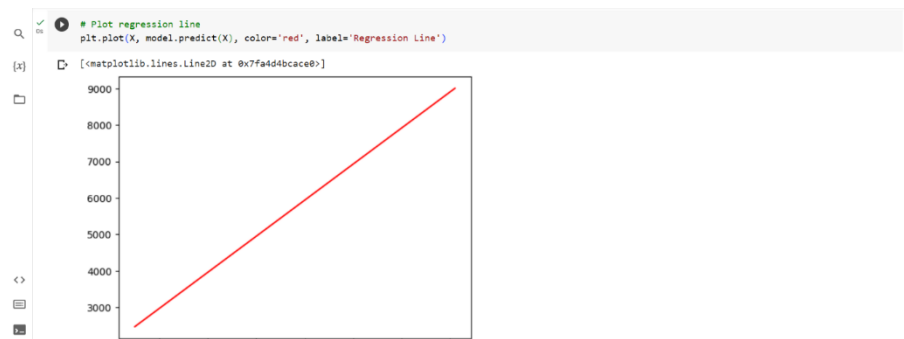
*Gambar 1. 6 Pembuatan Modelling & Training*

## Proses pembuatan plot data point



Gambar 1. 7 Data points

Untuk gambar di bawah ini proses pembuatan regrestion line (garis regresi).



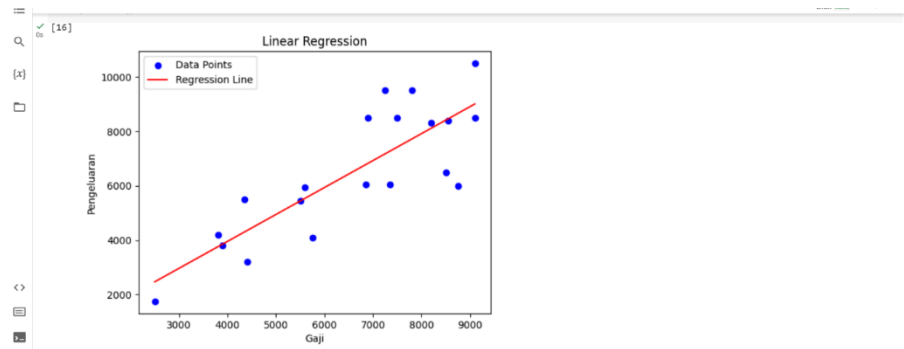
Gambar 1. 8 Regression line

Gambar dibawah ini proses pembuatan grafik dari linear regression, dengan menggabungkan dari pembuatan data points dan regression line



Gambar 1. 9 Pembuatan grafik

## Gambar dari grafik linear regression



Gambar 1. 10 Grafik linear regression

## 2.9. Akurasi

```
[16] ✓
# Contoh label prediksi
predicted_labels = [0, 1, 1, 0, 1, 0]

# Contoh label aktual
actual_labels = [0, 1, 0, 1, 0, 1]

# Menghitung akurasi
accuracy = accuracy_score(actual_labels, predicted_labels)

# Menampilkan akurasi
print("Akurasi:", accuracy)
```

Akurasi: 0.3333333333333333

Gambar 1. 11 Pembuatan Akurasi

## 2.10. Evaluasi

```
[18] ✓ # Melakukan prediksi nilai berdasarkan nilai input
y_pred = model.predict(X)

# Menghitung mean squared error (MSE)
# menghitung seberapa dekat predik dengan nilai sebenarnya
mse = mean_squared_error(y, y_pred)
print("Mean Squared Error (MSE):", mse)

Mean Squared Error (MSE): 1699698.8584926191

[19] ✓ # Menghitung R-squared score
#menghitung seberapa baik regresi linear dlm melakukan prediksinya
r2 = r2_score(y, y_pred)
print("R-squared Score:", r2)

R-squared Score: 0.6843623542861564
```

Gambar 1. 12 Pembuatan Evaluasi

## **BAB III**

### **KESIMPULAN DAN SARAN**

#### **3.1 Kesimpulan**

Metode Naive Bayes memberikan akurasi yang sedikit lebih tinggi daripada KNN ( $k=7$ ), namun KNN ( $k=27$ ) memberikan akurasi yang lebih baik daripada Naive Bayes pada dataset penerimaan beasiswa BBJ 2020.

Regresi Linear adalah metode statistik yang digunakan untuk memodelkan hubungan linier antara variabel-variabel. Metode ini cocok untuk memprediksi nilai kontinu dan memahami hubungan antara variabel numerik. Dalam konteks penerimaan beasiswa BBJ 2020, tidak ada kesimpulan yang diberikan tentang Regresi Linear.

#### **3.2 Saran**

Untuk Naive Bayes, periksa asumsi independensi variabel dan pertimbangkan metode berbasis kernel atau Gaussian jika dataset memiliki fitur kontinu. Untuk KNN, pilih nilai  $K$  yang optimal dan lakukan pemrosesan data yang tepat, seperti normalisasi fitur. Untuk Regresi Linear, evaluasi asumsi dasar dan pertimbangkan seleksi fitur jika diperlukan.

Selain itu, penting untuk memperhatikan karakteristik data dan konteks masalah yang sedang dihadapi. Eksplorasi metode lain, seperti Decision Tree, Random Forest, atau Support Vector Machines, juga dapat dilakukan untuk membandingkan performa dan memilih metode terbaik sesuai kebutuhan.

## DAFTAR PUSTAKA

- Mulyono. (2019, Desember 2). Retrieved from Analisis Regresi Sederhana:  
<https://bbs.binus.ac.id/management/2019/12/ANALISIS-REGRESI-SEDERHANA/>
- Suksmawati, Affi Nizar, et al. "DESAIN SISTEM PENENTUAN NILAI BONUS GAJI AKHIR TAHUN MENGGUNAKAN NAIVE BAYES DAN FUZZY MAMDANI." Prosidia Widya Saintek 1.1 (2022): 1-8.
- Suksmawati, A. N., Istiadi, I., Rahman, A. Y., Nurdiansyah, F., & Rachmawati, Y. (2022). DESAIN SISTEM PENENTUAN NILAI BONUS GAJI AKHIR TAHUN MENGGUNAKAN NAIVE BAYES DAN FUZZY MAMDANI. Prosidia Widya Saintek, 1(1), 1-8.
- SUKSMAWATI, Affi Nizar, et al. DESAIN SISTEM PENENTUAN NILAI BONUS GAJI AKHIR TAHUN MENGGUNAKAN NAIVE BAYES DAN FUZZY MAMDANI. Prosidia Widya Saintek, 2022, 1.1: 1-8.