

**LAPORAN TUGAS BESAR**  
**KECERDASAN BUATAN**  
**PREDIKSI GAJI DENGAN PENGALAMAN BEKERJA**  
**MENGGUNAKAN METODE LINEAR REGRESION**

Dosen Pengampu : Leni Fitriani, ST. M.Kom



**Disusun Oleh:**

Dendi mugia ikhlas 2106007

Gina Suciya 2106066

**PROGRAM STUDI TEKNIK INFORMATIKA INSTITUT**  
**TEKNOLOGI GARUT**

**2023**

## DAFTAR ISI

DAFTAR ISI .....	i
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan masalah.....	1
1.3 Tujuan .....	1
1.5 Manfaat.....	3
BAB II PENELITIAN SEBELUMNYA DAN GAP ANALYSIS .....	4
2.1 Penelitian Sebelumnya .....	4
2.2 Gap Analysis.....	5
BAB III METODE.....	7
3.1 Metode.....	7
BAB IV HASIL DAN PEMBAHASAN.....	9
4.1 Pengumpulan Data(Data Collection).....	9
4.2 Persiapan Data(Data Preparation) .....	9
4.3 Pemodelan(Modelling) .....	11
4.4 Pelatihan (Training).....	12
4.5 Evaluasi (Evaluation) .....	13
BAB V KESIMPULAN .....	15
5.1 Kesimpulan.....	15
5.2 Saran .....	15
DAFTAR PUSTAKA .....	17
LAMPIRAN.....	18

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Dalam dunia kerja, gaji merupakan salah satu aspek penting yang menjadi pertimbangan bagi karyawan maupun perusahaan. Penggajian yang adil dan sejalan dengan tingkat pengalaman kerja menjadi hal yang krusial dalam menjaga kepuasan karyawan dan memotivasi produktivitas mereka. Oleh karena itu, memahami hubungan antara pengalaman kerja dan gaji menjadi sangat relevan dan bermanfaat.

Dalam upaya memahami hubungan tersebut, metode analisis regresi linear sering digunakan. Regresi linear merupakan salah satu metode statistika yang digunakan untuk memodelkan hubungan antara variabel independen dan dependen. Dalam konteks ini, variabel independen adalah tahun pengalaman kerja (YearsExperience) sementara variabel dependen adalah gaji (Salary).

### **1.2 Rumusan masalah**

1. Bagaimana pengalaman kerja seorang individu dapat mempengaruhi tingkat gaji yang diterima?
2. Seberapa akurat prediksi gaji berdasarkan pengalaman kerja menggunakan metode linear regression?
3. Seberapa akurat prediksi gaji dengan linear regression dibandingkan algoritma lain?
4. Seberapa signifikan pengaruh pengalaman kerja terhadap perbedaan gaji antara individu yang memiliki pengalaman kerja yang berbeda?

### **1.3 Tujuan**

Tujuan utama dari program ini adalah untuk mengetahui akurat prediksi gaji dengan linear regression dibandingkan algoritma lain dan memahami hubungan antara tahun pengalaman dan gaji serta memprediksi gaji berdasarkan tahun pengalaman menggunakan model regresi linear. Program ini juga bertujuan untuk mengevaluasi performa model menggunakan skor dan melakukan visualisasi data dan prediksi dengan menggunakan scatter plot dan plot regresi.

## 1.4 Batasan Masalah

1. Metode: Penelitian ini akan menggunakan metode linear regression sebagai alat untuk memprediksi gaji berdasarkan pengalaman kerja.
2. Variabel: Fokus penelitian ini akan terbatas pada pengaruh pengalaman kerja sebagai variabel independen terhadap gaji sebagai variabel dependen. Variabel lain, seperti pendidikan, industri, atau faktor personal lainnya, tidak akan dimasukkan dalam analisis ini.
3. Data: Penelitian ini akan menggunakan data yang telah terkumpul sebelumnya, baik dari sumber-sumber terpercaya atau studi-studi sebelumnya yang relevan. Tidak ada pengumpulan data primer yang akan dilakukan.
4. Validitas prediksi: Penelitian ini akan mengasumsikan bahwa prediksi gaji berdasarkan pengalaman kerja menggunakan metode linear regression memiliki tingkat validitas yang memadai. Namun, tidak ada jaminan bahwa prediksi tersebut akan benar-benar akurat dan dapat dipertimbangkan sebagai keputusan final dalam pengambilan keputusan terkait gaji.
5. Batasan populasi: Penelitian ini akan memfokuskan pada populasi pekerja secara umum dan tidak akan mempertimbangkan faktor-faktor spesifik seperti jenis pekerjaan, tingkat pendidikan, atau wilayah geografis.
6. Asumsi linearitas: Penelitian ini akan berasumsi bahwa hubungan antara pengalaman kerja dan gaji adalah linier, yaitu bahwa perubahan dalam pengalaman kerja akan berdampak secara proporsional pada perubahan gaji.
7. Faktor-faktor lain: Penelitian ini tidak akan mempertimbangkan faktor-faktor eksternal seperti kondisi ekonomi, perubahan kebijakan, atau tren industri yang dapat mempengaruhi hubungan antara pengalaman kerja dan gaji.

### **1.5 Manfaat**

1. **Pemahaman Hubungan:** Program ini memberikan pemahaman yang lebih baik tentang keterhubungan antara pengalaman kerja dan gaji.
2. **Prediksi Gaji:** Program ini dapat memprediksi gaji berdasarkan pengalaman kerja, membantu dalam estimasi gaji yang sesuai.
3. **Evaluasi Model:** Model regresi linear dievaluasi menggunakan skor untuk menilai kualitasnya dalam menjelaskan variasi data.
4. **Visualisasi Data:** Program ini memberikan visualisasi scatter plot dan plot regresi untuk memvisualisasikan data dan hasil prediksi.

## **BAB II**

### **PENELITIAN SEBELUMNYA DAN GAP ANALYSIS**

#### **2.1 Penelitian Sebelumnya**

1. Penelitian terdahulu yang dilakukan oleh Muhammad Rafiqul Adnan , Rama Bagus Satria P , Fajar Aji Prayoga , dan Didin Kusuma Wardani dalam penelitiannya yang berjudul “Program prediksi gaji karyawan menggunakan machine learning dengan algoritma SVM”. [1]

Persamaan penelitian sebelumnya dengan penelitian ini adalah sebagai berikut:

- Objek yang diteliti adalah sama-sama prediksi gaji dengan pengalaman
- Impelementasinya sama-sama menggunakan machine learning

2. Penelitian terdahulu yang dilakukan oleh Tee Zhen Quan, dan Mafas Raheem dalam penelitiannya yang berjudul “Salary Prediction in Data Science Field Using Specialized Skills and Job Benefits – A Literature Review”. [2]

Persamaan penelitian sebelumnya dengan penelitian ini adalah sebagai berikut:

- Objek yang diteliti adalah sama-sama prediksi gaji
- Impelementasinya sama-sama menggunakan machine learning

3. Penelitian terdahulu yang dilakukan oleh Muhammad Arif Saputra , Nike Prasetyo , Irsyad Zulfikar , Teguh Rijanandi , dan Faisal Dharma Adhinata dalam penelitiannya yang berjudul “Prediksi Gaji Berdasarkan Pengalaman Bekerja Menggunakan Metode Regresi Linear”. [3]

Persamaan penelitian sebelumnya dengan penelitian ini adalah sebagai berikut:

- Objek yang diteliti adalah sama-sama prediksi gaji dengan pengalaman
- Impelementasinya sama-sama menggunakan metode Regresi linear

## 2.2 Gap Analysis

No	Penelitian	Hasil	Research gap
1	M. R. Adnan, R. B. S. P, F. A. Prayoga, and D. Kusuma, "Program prediksi gaji karyawan menggunakan machine learning dengan algoritma SVM," pp. 1–6.	Dengan metode svm didapat accuracy paling tinggi sebesar 100%	Dengan metode Linear Regression didapatkan accuracy sebesar 97%, Tetapi meskipun accuracy Linear Regression sedikit lebih kecil dari penelitian sebelumnya, Linear Regression memiliki kelebihan yang sederhana dan mudah dimengerti. Sedangkan untuk svm terlalu kompleks
2	M. Raheem, "Salary Prediction in Data Science Field Using Specialized Skills and Job Benefits – A Literature Review," no. July, 2022.	Dalam peneltian ini menggunakan dua metode yang menghasilkan accuracy 87% untuk metoder random forest	Untuk yang dihasilkan linear regresiion lebih unggul accuracynya yaitu sebesar 97%

		dan 84% untuk decision tree	
3	K. K. Rekayasa, M. A. Saputra, N. Prasetyo, I. Zulfikar, T. Rijanandi, and F. Dharma Adhinata, "Terbit online pada laman web jurnal: <a href="http://journal.itttelkom-pwt.ac.id/index.php/dinda">http://journal.itttelkom-pwt.ac.id/index.php/dinda</a> Journal of Dinda Prediksi Gaji Berdasarkan Pengalaman Bekerja Menggunakan Metode Regresi Linear," <i>Data Inst. Teknol. Telkom Purwokerto</i> , vol. 2, no. 2, pp. 58–63, 2022,	Dalam penelitian ini menghasilkan penghitungan gaji dengan metode regresi linear menghasilkan tingkat eror dari metode MAE sebesar 2,051033333	Untuk ketepatan klasifikasi nilai accuracy linear regression 97% lebih baik dari pada nilai MAE 2,051033333.



## **BAB III**

### **METODE**

#### **3.1 Metode**

Model yang digunakan dalam penelitian ini merupakan model Statistik Linear Regression. Linear Regression (Regresi Linear) adalah suatu regresi linear yang digunakan untuk mengestimasi atau memprediksi hubungan antara dua variabel dalam penelitian kuantitatif. Dimana regresi linear ini mampu membuat satu asumsi tambahan yang mengkorelasikan antara variabel independen dan dependen melalui garis yang paling sesuai dari titik data garis lurus. Artinya, bukan kurva atau semacam faktor penelompokkan.

Meski demikian, regresi linear memiliki keterbatasan, karena dalam data terbaik pun tidak menceritakan kisah yang lengkap. Analisis regresi biasanya digunakan dalam penelitian untuk menetapkan bahwa ada korelasi antar variabel.

Namun, korelasi yang tidak sama sebab akibat adanya hubungan antara dua variabel tidak berarti yang satu menyebabkan yang lainnya terjadi. Bahkan garis dalam regresi linear sederhana yang cocok dengan titik data mungkin tidak menjamin hubungan sebab-akibat.

Menurut Sugiyono (2011), regresi linear sederhana adalah regresi linear yang didasarkan pada hubungan fungsional ataupun kausal satu variabel independen dengan satu variabel dependen.

Kelebihan dari linear regression Ketika hubungan antara variabel independen dan dependen memiliki hubungan linear, algoritma ini adalah yang terbaik untuk digunakan, karena ini adalah yang paling kompleks dibanding algoritma lain yang juga menemukan hubungan antara variabel independen dan dependen. Metode ini mampu digunakan untuk memprediksi nilai yang ada pada masa depan. Hal ini sejalan dengan fungsi dari analisis regresi yang dapat digunakan untuk peramalan dan prediksi.

Kekurangan dari linear regression Pada kenyataannya, dalam data real, jarang masalah di dunia yang menunjukkan hubungan yang jelas antara variabel dependen dan independen. Hal ini dapat membuat model yang tidak cukup bagus, disebabkan karena kesalahan dalam memilih variabel yang digunakan untuk analisis.[4]

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Pengumpulan Data(Data Collection)

Dalam tahapan ini menggunakan data csv yang diperoleh dari github

#### 4.2 Persiapan Data(Data Preparation)

```
import pandas as pd
import matplotlib.pyplot as plt
```

# Load dataset

```
df = pd.read_csv('gaji.csv', usecols=['YearsExperience', 'Salary'])
```

# Melihat data set

```
df.head()
```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

```
: df.shape
```

```
: (30, 2)
```

```
: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   YearsExperience  30 non-null    float64
1   Salary          30 non-null    float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

```
df.describe()
```

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

```
df.isnull().sum()
```

```
YearsExperience    0  
Salary            0  
dtype: int64
```

# Visualisasi data dengan scatter plot

```
plt.scatter(df['YearsExperience'], df['Salary'])  
plt.xlabel('YearsExperience')  
plt.ylabel('Salary')  
plt.title('scatter plot YearsExperience vs Salary')  
plt.show()
```



# Menghitung korelasi antara variabel YearsExperience dan Salary

```
df.corr()
```

	YearsExperience	Salary
YearsExperience	1.000000	0.978242
Salary	0.978242	1.000000

didalam bagian "Persiapan Data", dilakukan pemuatan dataset, eksplorasi data dengan metode seperti head(), shape, info(), describe(), dan isnull().sum(). Selanjutnya, data divisualisasikan menggunakan scatter plot untuk memahami hubungan antara variabel YearsExperience dan Salary.

### 4.3 Pemodelan(Modelling)

Tahapan pemodelan

a. Mengimpor library yang diperlukan:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
```

dalam tahap ini mengimport library yang diperlukan untuk melakukan pemodelan, pelatihan, dan evaluasi diimpor. Library yang digunakan antara lain pandas untuk manipulasi data, sklearn untuk model regresi linear, dan matplotlib untuk visualisasi.

b. Memuat dataset 'gaji.csv' dan memilih kolom 'YearsExperience' dan 'Salary':

```
df = pd.read_csv('gaji.csv', usecols=['YearsExperience', 'Salary'])
```

Dataset dalam format CSV dimuat menggunakan pandas. Hanya kolom 'YearsExperience' dan 'Salary' yang dipilih untuk digunakan dalam analisis.

- c. Persiapan data dengan memisahkan variabel independen (x) dan variabel dependen (y):

```
x = df['YearsExperience'].values.reshape(-1,1)
y = df['Salary'].values.reshape(-1,1)
```

Data pada kolom 'YearsExperience' diambil sebagai variabel independen (x), sedangkan data pada kolom 'Salary' diambil sebagai variabel dependen (y). Data kemudian diubah ke dalam bentuk yang sesuai untuk analisis menggunakan metode reshape().

- d. Pembagian data menjadi data latih dan data uji menggunakan train\_test\_split():

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
```

Data dibagi menjadi dua set: data latih (80%) dan data uji (20%). Fungsi train\_test\_split() dari sklearn digunakan untuk melakukan pembagian secara acak.

- e. Pemodelan menggunakan Regresi Linear:

```
lin_reg = LinearRegression()
```

Model Regresi Linear dari sklearn diinisialisasi menggunakan LinearRegression()

#### 4.4 Pelatihan (Training)

Model Regresi Linear dilatih dengan data latih menggunakan metode fit(). Proses ini melibatkan penyesuaian garis regresi ke data latih untuk mempelajari hubungan antara variabel independen dan dependen

```
from pandas.core.apply import normalize_keyword_aggregation
lin_reg.fit(x_train, y_train)
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None)
```

#### 4.5 Evaluasi (Evaluation)

Model dievaluasi dengan menghitung skor pada data uji. Skor mengindikasikan sejauh mana model dapat menjelaskan variasi dalam data uji. Skor semakin mendekati 1, semakin baik model dalam menjelaskan variasi data.

```
lin_reg.score(x_test, y_test)
```

```
0.9700369972243545
```

Memprediksi nilai Salary berdasarkan YearsExperience pada data uji dan membuat plot regresi

```
y_prediksi = lin_reg.predict(x_test)
plt.scatter(x_test, y_test)
plt.plot(x_test, y_prediksi, c='r')
plt.xlabel('Years Experience')
plt.ylabel('Salary')
plt.title('plot YearsExperience vs Salary')
```

```
Text(0.5, 1.0, 'plot YearsExperience vs Salary')
```



Model yang telah dilatih digunakan untuk memprediksi nilai Salary berdasarkan YearsExperience pada data uji ( $x_{\text{test}}$ ). Prediksi ini dilakukan menggunakan metode `predict()`. Hasil prediksi yang diperoleh pada langkah sebelumnya divisualisasikan dalam bentuk scatter plot dan plot regresi.

Scatter plot menunjukkan titik data asli dari data uji ( $x_{\text{test}}$  dan  $y_{\text{test}}$ ), sedangkan plot regresi menunjukkan garis regresi yang dihasilkan oleh model untuk memperkirakan hubungan antara YearsExperience dan Salary.



## **BAB V**

### **KESIMPULAN**

#### **5.1 Kesimpulan**

Hasil dari penelitian dengan menggunakan metode linear regression didapatkan akurasi sebesar 97%. Hasil akurasi yang diperoleh berdasarkan dataset yang diperoleh dari kaggle dan github telah melalui beberapa tahapan dalam pembuatan model. Penelitian ini juga telah melakukan pengujian dengan melakukan beberapa parameter untuk mendapatkan performa yang terbaik dalam pembuatan pemodelan. Mulai dari kernel yang digunakan secara variasi. Diperoleh kernel linear yang memiliki tingkat akurasi yang cukup tinggi dalam metode Linear Regression. Dalam penelitian ini, model juga dibandingkan dengan model lain yaitu metode SVM, random forest, decision tree. Metode SVM mampu memperoleh nilai akurasi sebesar 100%, dimana metode SVM lebih unggul 3% dibanding dengan metode linear regression. Sedangkan untuk metode random forest dengan accuracy 87% dan decision tree 84% menjadikan hasil accuracy linear regression lebih unggul karena memiliki accuracy 97%. Dengan hasil akurasi yang diperoleh dengan menggunakan metode linear regression, model tersebut sudah cukup baik untuk melakukan prediksi untuk menentukan bahwa pengalaman berperan penting untuk besarnya gaji.

#### **5.2 Saran**

Berdasarkan program yang telah dibuat untuk prediksi gaji menggunakan metode linear regression ada beberapa saran yang ditemukan yaitu :

Memperluas Variabel Independen: Selain tahun pengalaman kerja, bisa juga dipertimbangkan untuk memasukkan variabel lain yang dapat berpengaruh terhadap gaji, seperti tingkat pendidikan, lokasi geografis, atau spesialisasi pekerjaan. Ini dapat membantu meningkatkan akurasi prediksi gaji.

Menyertakan Analisis Statistik Tambahan: Selain visualisasi data dengan scatter plot dan plot regresi, program ini juga dapat menyertakan analisis statistik tambahan seperti interval kepercayaan, uji signifikansi, atau pengujian asumsi

regresi linear. Ini akan memberikan pemahaman yang lebih mendalam tentang hubungan antara pengalaman kerja dan gaji.

## DAFTAR PUSTAKA

- [1] M. R. Adnan, R. B. S. P, F. A. Prayoga, and D. Kusuma, “Program prediksi gaji karyawan menggunakan machine learning dengan algoritma SVM,” pp. 1–6.
- [2] M. Raheem, “Salary Prediction in Data Science Field Using Specialized Skills and Job Benefits – A Literature Review,” no. July, 2022.
- [3] K. K. Rekayasa, M. A. Saputra, N. Prasetyo, I. Zulfikar, T. Rijanandi, and F. Dharma Adhinata, “Terbit online pada laman web jurnal: <http://journal.ittelkom-pwt.ac.id/index.php/dinda> Journal of Dinda Prediksi Gaji Berdasarkan Pengalaman Bekerja Menggunakan Metode Regresi Linear,” *Data Inst. Teknol. Telkom Purwokerto*, vol. 2, no. 2, pp. 58–63, 2022, [Online]. Available: <http://journal.ittelkom-pwt.ac.id/index.php/dinda>
- [4] Caraguna, “Apa Itu Linear Regression dalam Machine Learning?” <https://caraguna.com/apa-itu-linear-regression-dalam-machine-learning/>

## LAMPIRAN

<https://github.com/dendimugia/tugas-besar-praktikumkecerdasanbuatan/blob/e092969acd382f13bdde9c908361b4cf4f87d594/Program%20Prediksi%20Gaji.ipynb>

[https://github.com/srafay/Machine\\_Learning\\_A-Z/blob/868976e3590b9cf731ed1f3ab3b7ef8bcf2f8fb0/Part%202%20-%20Regression/Section%204%20-%20Simple%20Linear%20Regression/Salary\\_Data.csv](https://github.com/srafay/Machine_Learning_A-Z/blob/868976e3590b9cf731ed1f3ab3b7ef8bcf2f8fb0/Part%202%20-%20Regression/Section%204%20-%20Simple%20Linear%20Regression/Salary_Data.csv)