



SUBMIT

(Jurnal Ilmiah Teknologi Informasi dan Sains)

Vol. 4 No. 2 (2024) 5 - 8

ISSN Media Elektronik: 2798-6861

PREDIKSI GAJI BERDASARKAN LAMA BEKERJA MENGGUNAKAN METODE REGRESI LINEAR SEDERHANA

¹Friska Andriani, ²Yanuarini Nur Sukmaningtyas

^{1,2}Informatika, Fakultas Teknik, Universitas Islam Majapahit

Email: friskaandriani74@gmail.com

(Naskah masuk: 20 Juni 2024, diterima untuk diterbitkan: 14 Desember 2024)

Abstrak

Gaji adalah salah satu faktor utama yang memotivasi karyawan, sehingga perusahaan harus memperhatikan dengan baik sistem penggajian, termasuk upah dan tunjangan sosial yang diharapkan oleh karyawan. Oleh karena itu, pengukuran kinerja karyawan yang akurat dan sistem penggajian yang transparan sangatlah penting untuk memastikan bahwa pembayaran gaji sesuai dengan kontribusi yang diberikan oleh masing-masing karyawan. Penelitian ini bertujuan untuk mengembangkan model prediksi gaji karyawan berdasarkan lama bekerja menggunakan metode regresi sederhana. Dengan memanfaatkan pendekatan ini, diharapkan dapat memberikan wawasan yang lebih baik tentang bagaimana variabel-variabel tertentu mempengaruhi penggajian dan kontribusi karyawan. Berdasarkan pengujian menggunakan fungsi regresi linier diperoleh nilai akurasi sebesar 83% untuk memprediksi besarnya gaji (y) dari tahun kerja (X), diketahui bahwa semakin lama bekerja seorang karyawan, maka gaji karyawan tersebut juga akan semakin meningkat setiap tahunnya. Hasil yang didapatkan yaitu tingkat error MAE 281517.77. Dalam hal ini, nilai MAE yang cukup besar menunjukkan bahwa kesalahan absolut rata-rata antara nilai prediksi dan nilai aktual cukup tinggi.

Kata kunci: *Regresi Linear, Prediksi Gaji, Karyawan*

SALARY PREDICTION BASED ON LENGTH OF SERVICE USING SIMPLE LINEAR REGRESSION METHOD

Abstract

Salary is one of the main factors that motivate employees, so companies must pay close attention to the payroll system, including wages and social benefits expected by employees. Therefore, accurate measurement of employee performance and a transparent payroll system are essential to ensure that salary payments match the contributions made by each employee. This research aims to develop a prediction model of employee salary based on length of service using the simple regression method. By utilizing this approach, it is expected to provide better insight into how certain variables affect payroll and employee contributions. Based on testing using the linear regression function, an accuracy value of 83% is obtained to predict the amount of salary (y) from years of work (X), it is known that the longer an employee works, the employee's salary will also increase every year. The result obtained is an MAE error rate of 281517.77. In this case, the large MAE value indicates that the average absolute error between the predicted and actual values is quite high.

Keywords: *Linear Regression, Salary Prediction, Employee*

1. PENDAHULUAN

Sumber daya manusia (SDM) adalah faktor kunci yang menentukan keberhasilan sebuah perusahaan di era globalisasi yang penuh persaingan ketat. Memiliki keuangan yang kuat, bahan baku murah, dan teknologi canggih saja tidak cukup, SDM yang berkualitas sangat penting untuk mencapai tujuan perusahaan. Gaji menjadi salah satu faktor utama yang memotivasi karyawan, sehingga sistem penggajian yang mencakup upah dan tunjangan sosial harus diperhatikan dengan baik oleh perusahaan. (Zulfikar et al., 2022)

Produktivitas karyawan memiliki dampak langsung terhadap kinerja perusahaan secara keseluruhan. Gaji yang sesuai dengan kontribusi dan prestasi kerja karyawan dapat meningkatkan motivasi dan profesionalisme mereka. Sebaliknya, gaji yang tidak memadai dapat menurunkan produktivitas. Oleh karena itu, pengukuran kinerja karyawan yang akurat dan sistem penggajian yang transparan sangat penting untuk memastikan bahwa pembayaran gaji sesuai dengan kontribusi yang diberikan oleh setiap karyawan. (Ramadhan et al., 2023)

Di Indonesia, perkembangan industri yang pesat dipicu oleh kemajuan ilmu pengetahuan dan teknologi, sehingga persaingan antar perusahaan semakin meningkat. Pengembangan SDM melalui pelatihan dan peningkatan keterampilan menjadi kunci untuk menjaga dan meningkatkan produktivitas perusahaan. Selain faktor internal seperti pengaturan gaji dan sistem insentif, faktor eksternal seperti pengembangan karir dan kondisi kerja yang mendukung juga berperan penting. Perusahaan harus memastikan bahwa tenaga kerja mereka tidak hanya terampil tetapi juga termotivasi untuk mencapai tujuan bersama, dengan perhatian pada SDM dan dukungan infrastruktur yang tepat sebagai kunci peningkatan produktivitas dan daya saing perusahaan. (Ratnasari & Mahmud, 2020)

Dalam penelitian ini, penggunaan metode yang disarankan adalah dengan menggunakan perhitungan regresi sederhana. Pemilihan metode ini dilakukan agar nilai variabel terikat dapat diketahui dengan menghitung pengaruh beberapa variabel bebas. Pengembangan sistem ini diharapkan dapat memberikan kemudahan dalam melakukan prediksi kisaran gaji yang akan ditentukan berdasarkan berapa lama karyawan tersebut bekerja.

Penelitian ini bertujuan untuk mengembangkan model prediksi gaji karyawan berdasarkan lama bekerja menggunakan metode regresi sederhana. Dengan memanfaatkan pendekatan ini, diharapkan dapat memberikan wawasan yang lebih baik tentang bagaimana variabel-variabel tertentu mempengaruhi penggajian dan kontribusi karyawan. Tujuan akhirnya

adalah membangun sistem penggajian yang lebih objektif dan adil, yang pada gilirannya diharapkan dapat meningkatkan motivasi, kinerja, dan kepuasan kerja karyawan dalam jangka panjang. (Yayan et al., 2020)

2. METODE PENELITIAN



Gambar 1. Tahap Metode Penelitian

Pada Gambar 1 menjelaskan tahap-tahap dalam penelitian dengan menggunakan metode regresi linier sederhana, yaitu pengumpulan data, prediksi gaji, pembahasan, dan hasil.

3.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini diperoleh dari *kaggle*. Dataset ini terdiri dari 870 data dan 3 kolom. Dataset yang digunakan dalam penelitian ini mengandung informasi mengenai Region, Salary (gaji) dan Tahun.

3.2. Pre-processing Data

Pre-processing data adalah langkah awal dalam analisis data yang bertujuan untuk mempersiapkan dan membersihkan data mentah sebelum digunakan dalam model atau analisis lebih lanjut. Proses ini melibatkan berbagai teknik untuk mengatasi masalah seperti data yang hilang, data yang tidak konsisten, atau data yang tidak relevan. Dengan melakukan pre-processing data, kita dapat meningkatkan kualitas data, yang pada akhirnya membantu meningkatkan akurasi dan efektivitas model atau analisis yang dilakukan.

3.3. Pemrograman Python

Python merupakan bahasa pemrograman tingkat tinggi yang diracik oleh Guido van Rossum. Python banyak digunakan untuk membuat berbagai macam program, seperti: program CLI, Program GUI (desktop), Aplikasi Mobile, Web, IoT, Game, Program untuk Hacking, dsb. Python juga dikenal dengan bahasa pemrograman yang mudah dipelajari, karena struktur sintaknya rapi dan mudah dipahami.

3.4. Regresi Linear

Regresi linear adalah metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan variabel dependen (respons). Regresi linear bertujuan untuk menemukan persamaan linear yang paling sesuai dengan data yang ada.

Ada dua jenis regresi linier: regresi linier sederhana dan regresi linier berganda. Perbedaananya terletak pada jumlah variabel independennya. Regresi linier tunggal merupakan metode membangun model hubungan antara suatu variabel terikat dengan variabel bebas, sedangkan regresi linier berganda merupakan metode membangun model hubungan antara suatu variabel terikat dengan beberapa variabel bebas. Analisis regresi sederhana adalah hubungan antara dua variabel, terutama satu variabel bebas dan satu variabel terikat, sedangkan regresi berganda atau berganda adalah hubungan paling sedikit tiga variabel, terutama dua variabel bebas dan satu variabel terikat. Rumus regresi Linear Sederhana yaitu :

$$\hat{Y} = a + bX \tag{1}$$

Dengan keterangan :

\hat{Y} = penduga

a = Intersep

X = Variabel bebas

b = Slop (kemiringan garis regresi)

3. HASIL DAN PEMBAHASAN

3.1. Pengolahan Data

Dataset yang digunakan dalam penelitian ini adalah gaji karyawan dari tahun 1997-2022 yang dapat dilihat pada tabel 1. Dalam tahap ini, yang akan dilakukan adalah memahami dan mempersiapkan data yang dikenal dengan istilah Data Preprocessing. Metode yang digunakan dalam Data Preprocessing pada model ini adalah Data Cleaning.

Tabel 1. Pengolahan Data

	REGION	SALARY	YEAR
0	ACEH	128000	1997
1	SUMATERA UTARA	151000	1997
2	SUMATERA BARAT	119000	1997
3	RIAU	151500	1977
4	JAMBI	119500	1997

Pada tahap ini dilakukan pemeriksaan *missing value* untuk mengetahui di dalam dataset terdapat data yang bernilai *null* (*missing value*).

```
REGION      870
SALARY      0
YEAR        0
dtype: int64
```

Gambar 2. Missing Value

Pada Gambar 2 terdapat *missing value* dalam variabel Region karena dilakukan perubahan tipe data Objek menjadi Int64. Untuk mengatasi *missing value* kita dapat melakukan *drop* pada baris yang memiliki *missing value* atau mengisi data (mengganti).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 870 entries, 0 to 869
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  ---
0   REGION  0 non-null        Int64
1   SALARY  870 non-null     int64
2   YEAR    870 non-null     int64
dtypes: Int64(1), int64(2)
memory usage: 21.4 KB
```

Gambar 3. DataFrame

Pada Gambar 3 menampilkan DataFrame setelah dilakukan perubahan tipe data dari Objek menjadi Int64.

3.2. Pertumbuhan Gaji

Semakin lama seseorang bekerja di sebuah perusahaan, umumnya mereka akan mengalami peningkatan gaji yang signifikan seiring dengan bertambahnya pengalaman dan tanggung jawab. Secara umum, pertumbuhan gaji cenderung mengikuti kurva yang naik seiring bertambahnya masa kerja, meskipun kecepatan dan besarnya kenaikan bisa berbeda-beda tergantung pada kebijakan perusahaan dan industri.

Tabel 2. Rata-Rata Berdasarkan Tahun

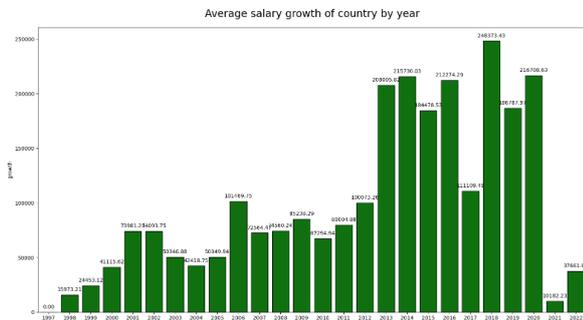
	YEAR	REGION	SALARY
0	1997	<NA>	134964.285714
1	1998	<NA>	150937.500000
2	1999	<NA>	175390.625000
3	2000	<NA>	216506.250000
4	2001	<NA>	290487.500000

Pada tahap ini dilakukan penghitungan rata-rata UMP(Upah Minimum Provinsi) untuk memperoleh gambaran menyeluruh mengenai tingkat upah minimum yang berlaku di berbagai daerah, serta untuk menganalisis disparitas upah antar provinsi yang dapat dilihat pada tabel 2.

Tabel 3. Menghitung Pertumbuhan Gaji

	YEAR	REGION	SALARY	growth
0	1997	<NA>	134964.285714	0.000000
1	1998	<NA>	150937.500000	15973.214286
2	1999	<NA>	175390.625000	24453.125000
3	2000	<NA>	216506.250000	41115.625000
4	2001	<NA>	290487.500000	73981.250000

Pada tabel 3 menghitung pertumbuhan gaji dari periode ke periode dan menambahkan informasi ini sebagai kolom baru ke DataFrame, lalu menampilkan beberapa baris pertama untuk memverifikasi hasilnya.



Gambar 4. Grafik Rata-Rata Pertumbuhan Gaji

Pada Gambar 4 dapat disimpulkan bahwa perkembangan pada tahun 2018 merupakan tahun dengan rata-rata UMP tertinggi, yang kemudian diikuti oleh tahun 2020 dan 2014, dan yang terendah pada tahun 2021 kemudian diikuti oleh tahun 1998 dan 1999. nilai yang rendah ini mungkin disebabkan oleh pandemi covid pada tahun 2021 dan krisis moneter tahun 1998-1999.

3.3. Pemodelan Linear Regression

Perhitungan dalam *linear regression* menggunakan library *sklearn*. Dataset yang berjumlah 870 dibagi menjadi data training dan data testing, 80% data training (696 data) dan 20% data testing (174 data). Dalam pemodelan ini menggunakan 2 variabel.

```
print(model.coef_)
print(model.intercept_)

[112596.17755904]
-225117426.24620375
```

Gambar 5. Pemodelan

3.4. Pengujian Hasil

```
model.score(X_test, y_test)

0.8389099082629923
```

Gambar 6. Akurasi

Berdasarkan gambar 6 diperoleh hasil akurasi sebesar 0.83. Sehingga dapat disimpulkan bahwa akurasi tersebut signifikan terhadap variabel yang digunakan.

```
MAE: 281517.7729002532
MSE: 136234849887.56837
RMSE: 369100.05403354845
```

Gambar 7. Hasil Akurasi

Berdasarkan gambar 7 diperoleh nilai MAE (*Mean Absolute Error*) sebesar 281517.7729002532. Dalam hal ini, nilai MAE yang cukup besar menunjukkan bahwa kesalahan absolut rata-rata antara nilai prediksi dan nilai aktual cukup tinggi. Sedangkan nilai MSE (*Mean Squared Error*) sebesar 136234849887.56. Dan nilai RMSE (*Root Mean Square Error*) sebesar 369100.05.

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Berdasarkan pengujian menggunakan fungsi regresi linier untuk memprediksi besarnya gaji (y) dari tahun kerja (X), diketahui bahwa semakin lama bekerja seorang karyawan, maka gaji karyawan tersebut juga akan semakin meningkat setiap tahunnya. Hasil ini menunjukkan bahwa lamanya bekerja memiliki pengaruh positif terhadap besarnya gaji karyawan. Dan dapat dilihat dari hasil nilai akurasi sebesar 83%.

4.2. Saran

Sumber data yang digunakan masih belum maksimal. Pada penelitian ini, hanya data asli yang diperoleh dari Kaggle. Diharapkan untuk kedepannya dapat menggunakan data asli langsung berasal dari perusahaan.

DAFTAR PUSTAKA

Yayan Adrianova, E. T., & Anyan, A. (2020). IMPLEMENTASI MODEL REGRESI LINEAR SEDERHANA UNTUK PREDIKSI GAJI BERDASARKAN PENGALAMAN LAMA BEKERJA. *education and technology*.

Ramadhan, B., Firdaus, D., & Adiningrum, N. T. R. (2023). ANALISIS DATA PEGAWAI UNTUK MEMREDIKSI GAJI BERDASARKAN FAKTOR-FAKTOR SPESIFIK DENGAN PENDEKATAN MACHINE LEARNING. *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, 5(2), 131-139.

Zulfikar, I., Saputra, M. A., Prasetyo, N., Rijanandi, T., & Adhinata, F. D. (2022). Prediksi Gaji Berdasarkan Pengalaman Bekerja Menggunakan Metode Regresi Linear. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, 2(2), 58-63.

Ratnasari, I., & Mahmud, A. (2020). Pengaruh Gaji dan Insentif terhadap Kinerja Karyawan bagian produksi PT. Uniplastindo Interbuana Pandaan. *JAMIN: Jurnal Aplikasi Manajemen dan Inovasi Bisnis*, 2(2), 66-79.