

ANALISIS DATA PEGAWAI UNTUK MEMPREDIKSI GAJI BERDASARKAN FAKTOR-FAKTOR SPESIFIK DENGAN ALGORITMA *REGRESI LINIER*

Bachtiar Ramadhan¹, Diash Firdaus², Nur Tri Ramadhanti Adiningrum³
Program Studi Teknik Informatika^{1,3}
Program Studi Informatika²
Universitas Logistik dan Bisnis Internasional^{1,3}
Institut Teknologi Nasional²
bachtiamadham26@gmail.com¹, diashfirdaus@gmail.com², nurtrira06@gmail.com³

Abstrak

Perusahaan tidak dapat dipisahkan dari tenaga kerjanya. Salah satu faktor yang memengaruhi kemajuan perusahaan adalah kinerja karyawannya. Menyediakan gaji yang sesuai merupakan salah satu faktor penting untuk meningkatkan kinerja tenaga kerja. Sayangnya, pengembangan perusahaan saat ini tidak memiliki media keputusan untuk memprediksi gaji karyawan berdasarkan kualitas data. Studi ini bertujuan untuk menentukan prediksi gaji karyawan berdasarkan faktor-faktor tertentu. Dalam studi ini, faktor-faktor yang diuji termasuk variabel independen berupa Usia, Tingkat Jabatan, Total Tahun Bekerja, dan Tahun di Perusahaan. Kemudian variabel dependen adalah Pendapatan Bulanan. Dalam melakukan prediksi gaji karyawan terdapat beberapa Teknik analisis, Teknik analisis data yang digunakan adalah analisis *regresi linier multivariat*. Output aplikasi untuk hasil prediksi gaji karyawan akan ditampilkan dalam bentuk *web*. Pengujian model dilakukan dengan Uji f (ANOVA), R-Squared, Linearitas, Normalitas, Multikolinearitas, Autokorelasi dan Homoskedastisitas. Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa *MonthlyIncome* dipengaruhi oleh faktor independen (*Age, YearsAtCompany*) sebesar 0,909 atau 90,9%. Nilai sisa dari akurasi tersebut adalah 0,091 atau 9,1% yang artinya *MonthlyIncome* dipengaruhi oleh faktor lain yang tidak diketahui sebesar 9,1%. Model yang dibuat berhasil melewati semua uji pada tahap validasi model, sehingga dapat disimpulkan bahwa model yang dibuat dapat bekerja dengan baik untuk memprediksi gaji karyawan. Visualisasi data dari hasil model prediksi gaji karyawan dapat digunakan menjadi bentuk aplikasi berbasis web base dengan menggunakan *framework Django*. Dengan aplikasi tersebut, admin dapat melakukan prediksi gaji karyawan dengan mudah dan dengan cepat.

Kata kunci : Prediksi Gaji, Linier Regresi Berganda, *Web Base*

Abstract

The company cannot be separated from the workforce. One aspect that affects the progress of a company is the performance of its employees. Providing an appropriate salary is one of the important factors to boost the performance of the workforce. Unfortunately, the current development of the company does not have a decision media to predict employee salaries based on data quality. This study aims to determine the prediction of employee salaries based on specific factors. In this study, the factors that were tested included independent variables in the form of Age, JobLevel, TotalWorkingYears, and YearsAtCompany. Then the dependent variable is MonthlyIncome. In predicting employee salaries, there are several analysis techniques. The data analysis technique used is multivariate linear regression analysis. The application output for predicting employee salaries will be displayed in a web form. Model testing was carried out by f test (ANOVA), R-Squared, Linearity, Normality, Multicollinearity, Autocorrelation and Homoscedasticity. Based on the validity test, the accuracy value of 0.909 indicates that MonthlyIncome is influenced by an independent factor (Age, YearsAtCompany) of 0.909 or 90.9%. The residual value of this accuracy is 0.091 or 9.1%, which means that MonthlyIncome is influenced by other unknown factors of 9.1%. The model made successfully passed all tests at the model validation stage, so it can be concluded that the model created can work well for predicting employee salaries. Data visualization from the results of the employee salary prediction model can be used to form a web-based application using the Django framework. With this application, admins can predict employee salaries easily and quickly.

Keywords : Salary Prediction, Multivariate Linear Regressio, Web Base

I. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi pada Revolusi Industri 4.0 semakin berkembang pesat. Revolusi Industri 4.0 sendiri mulai terjadi melalui rekayasa intelegensia dan *internet of thing* sebagai tulang punggung pergerakan dan konektivitas antara manusia dengan mesin[1]. Sehingga, terdapat penggabungan teknologi digital dan internet dengan industri konvensional, yang bertujuan untuk meningkatkan produktivitas, efisiensi dan layanan konsumen secara signifikan[2]. Era revolusi ini akan mendisrupsi berbagai kegiatan diberbagai bidang seperti pada bidang teknologi, ekonomi, sosial, dan politik[1]. Saat ini, kehidupan berada diawal revolusi yang secara mendasar mengubah cara hidup, bekerja, dan berhubungan satu sama lain [3].

Perubahan karakteristik pekerjaan adalah salah satu dampak tersendiri dari datangnya revolusi industri 4.0[4]. Karakteristik pekerjaan yang berubah akan mendisrupsi pekerjaan yang telah ada dan menggantikannya dengan pekerjaan dengan karakteristik baru [5]. Karakteristik baru pada pekerjaan juga membutuhkan kompetensi baru kepada para pekerja[6]. Tentunya perusahaan harus siap untuk saling bersaing dengan perusahaan yang lain[7]. Selanjutnya, perusahaan perlu memiliki keunggulan dan manajemen yang efektif untuk menghadapi persaingan tersebut[7]. Dengan demikian salah astu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja

karyawannya[7], walaupun perusahaan tersebut memiliki teknologi yang canggih, namun tidak terdapat tenaga kerja didalamnya, perusahaan tidak akan dapat mencapai tujuannya[7].

Oleh karena itu, penentuan gaji yang tepat oleh perusahaan kepada karyawan adalah salah satu faktor yang berpengaruh secara internal terhadap kemajuan perusahaan. Selain itu, perusahaan juga harus bersedia mengeluarkan gaji bonus bagi karyawannya yang telah bekerja dengan maksimal dan sesuai dengan apa yang dibutuhkan oleh sebuah perusahaan. Sangat disayangkan, perkembangan perusahaan saat ini belum memiliki suatu media keputusan untuk melakukan prediksi gaji karyawan berdasarkan kualitas data.

Karakteristik dataset yang digunakan untuk memprediksi gaji karyawan terdiri dari parameter-parameter berdasarkan faktor-faktor spesifik. Selanjutnya faktor-faktor tersebut akan diuji validitas dan korelasinya menggunakan pendekatan *machine learning*. Faktor-faktor tersebut akan diambil berdasarkan pedoman interpretasi koefisien korelasi [8]. Untuk menentukan faktor yang dominan terhadap prediksi gaji, maka koefisien korelasi yang akan digunakan adalah tingkat hubungan sedang, kuat, dan sangat kuat. Metode yang digunakan pada *machine learning* yaitu *regression*. *Regression* digunakan untuk melakukan prediksi gaji karyawan. Tentunya hasil prediksi gaji karyawan perlu divisualisasikan secara *realtime* untuk dapat digunakan oleh perusahaan dalam menentukan keputusan dengan cepat. Visualisasi hasil prediksi tersebut akan ditampilkan berbasis *web base* dengan *framework* Django.

II. TINJAUAN PUSTAKA

1. *Machine learning*

Machine learning dapat diartikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan dapat menghasilkan suatu prediksi di masa yang akan datang [9]. *Machine Learning* juga bisa dikatakan sebagai metode komputasi berdasarkan pengalaman untuk meningkatkan performa dari sebuah system dengan membuat prediksi yang akurat. Terdapat beberapa kategori pada *machine learning* diantaranya *supervised*, *unsupervised* dan *semi supervised*. *Machine Learning* dapat digunakan dalam berbagai bidang seperti Pendidikan, Kesehatan, industry, dsb.

2. Regresi Linier Berganda

Regresi berganda adalah perpanjangan dari regresi linier sederhana [10]. Analisis regresi linier berganda dapat digunakan untuk memprediksi hubungan antara satu variabel independen berdasarkan nilai dari dua atau lebih variabel dependen [25]. Analisis regresi linier berganda juga menghasilkan persamaan matematis [10]. Jika ada lebih dari dua variabel maka hubungan linier dapat dinyatakan dalam persamaan regresi linier berganda yang dikutip pada persamaan - persamaan sebagai berikut.

$$Y' = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

$$Y'_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_n X_{ni}$$

Keterangan :

Y = nilai-nilai hasil pengamatan

Y' = nilai regresi

i = 1,2,3, k

Pada persamaan di atas ada satu variabel dependen, yaitu Y' dan ada n variabel independen X_1, X_2, \dots, X_n [10].

3. *Scikit-Learn*

Scikit-learn adalah modul pada bahasa pemrograman Python yang menyediakan berbagai jenis algoritma *machine learning*. Bentuk yang terdapat pada library Python merupakan bentuk yang tersedia dalam *scikit-learn*. *Scikit-learn* memanfaatkan *task-oriented interface* yang konsisten sehingga memudahkan dalam membandingkan antarmetode. *Scikit-learn* mengintegrasikan berbagai algoritma *machine learning* untuk *supervised learning* dan *unsupervised learning* [11].

4. *Framework* Django

Django ialah sebuah *web framework* berbasis bahasa pemrograman Python yang didesain untuk membuat suatu aplikasi *web* yang dinamis, kaya fitur dan aman. [12] Django yang dikembangkan oleh Django Software Foundation terus mendapatkan perbaikan sehingga membuat *web framework* yang satu ini menjadi pilihan utama bagi banyak pengembang aplikasi *web*. Salah satu kelebihan Django adalah support dengan *Dynamic HTML*. Selain itu Django juga merupakan *framework* terpopuler kedua setelah Laravel pada tahun 2022 dilanjutkan dengan flask, expressJS, Spring, dsb.

5. Penelitian Sebelumnya

Pada penelitian sebelumnya terdapat beberapa jurnal yang telah dirangkum seperti pada table 1.

TABEL I
 PENELITIAN SEBELUMNYA

No.	Area Penelitian	Dataset	Metode
1	Prediksi Gaji [7]	Data Gaji Pegawai	Analisis Regresi Linear
2	Prediksi Saham [9]	Data Historis Harga Saham	CRISP-DM
3	Prediksi Sembako [14]	Data Sembako	Regresi Linear Berganda
4	Prediksi Harga Rumah [15]	<i>Web Scrapping 2 Website</i>	Regresi Linear
5	Prediksi Inventaris Barang [16]	Data Inventaris Barang	Regresi Linear
6	Prediksi Kasus Covid-19 [17]	Databooks	<i>Backpropagation</i> dan Regresi Linear
7	Prediksi Harga Emas [18]	Data Harga Emas	Regresi Linear, <i>backpropagation</i> , fuzzy mandani
8	Prediksi Pendapatan Daerah [19]	Data Besaran Pendapatan	Regresi Linear Sederhana
9	Prediksi Tingkat Produksi Kopi [20]	Data Produksi Kopi	Regresi Linear Sederhana
10	Peramalan Penerimaan Mahasiswa Baru [21]	Data Mahasiswa	Regresi Linear Sederhana
11	Prediksi Gaji [22]	Data Gaji	Regresi Linear dan Regresi Polinomial
12	Analisis Empiris [23]	Data Prediksi dan Harga Rumah	<i>Simple Linear Regression, Multiple Linear Regression</i>
13	Korelasi Gaji dan Efisiensi Inovasi <i>Enterprise</i> [24]	Data Perusahaan Manufaktur	<i>Multiple Regression</i>
14	Aplikasi Data Cuaca [25]	Dataset BMKG	Metode Regresi Linear Bivariat Simple
15	Estimasi Produktivitas Tanaman Padi [26]	Data Dinas Pertanian Kehutanan Perkebunan dan Peternakan	Regresi Linear Berganda
16	Analisis Kesejahteraan Pedagang Kaki Lima [27]	Data Pedagang Kaki Lima	Regresi Linear dan Logistik Ordinal

III. METODE PENELITIAN

Di dalam penelitian ini, digunakan metode dengan jenis penelitian deskriptif kualitatif yang menggunakan regresi linier multivariat dengan menggunakan bahasa pemrograman *Python*. Penelitian deskriptif adalah penelitian yang digunakan untuk menemukan pengetahuan yang seluas-luasnya terhadap objek penelitian pada suatu masa tertentu. Penelitian deskriptif ini menyajikan satu gambar yang terperinci mengenai satu situasi khusus. Penelitian deskriptif bertujuan untuk menjelaskan atau mendeskripsikan suatu keadaan apa adanya dan menginterpretasi objek sesuai dengan apa adanya peristiwa, ataupun segala sesuatu yang terkait dengan variabel-variabel yang bisa dijelaskan baik dengan angka-angka maupun kata-kata.

1. Kerangka Penelitian

Dalam penelitian ini, data yang digunakan dalam proses regresi linier multivariat diambil dari Kaggle. Data yang digunakan adalah data training sebanyak 1029 baris dan 35 kolom, dan data test sebanyak 441 baris dan 34 kolom. Kebutuhan untuk pembuatan model *machine learning* dan aplikasi prediksi gaji pegawai bisa dilihat pada table 2.

TABEL III
 EXPERIMENT SETUP

1	<i>Editor</i>	<i>Visual Studio Code</i>
2	<i>Dataset</i>	Data Gaji Pegawai
3	<i>CPU</i>	Prosesor Intel™ Core i3-4030U
4	<i>RAM</i>	4GB
5	<i>Operating System</i>	Windows 10
6	<i>Supporting Tools</i>	Google Colab Sklearn library Matplotlib Library Python3

IV. HASIL DAN PEMBAHASAN

1. Implementasi Model *Machine learning*

a. Himpunan Data

Pada tahap ini, hal yang dilakukan adalah memahami dan mempersiapkan data yang dikenal dengan istilah *Data Preprocessing*. Metode yang digunakan dalam *Data Preprocessing* pada model ini adalah *Data Cleaning*. Karena *machine learning* tidak bisa membaca tipe data *object*, maka perlu adanya perubahan tipe data tersebut dengan integer. Pada tahapan ini, digunakan proses encoder kategori. Pada dataset yang digunakan, terdapat nilai NaN/Null. Oleh karena itu, dilakukan proses pengisian nilai tersebut dengan nilai rata-rata (*mean*) variabelnya. Setelah semua data berbentuk integer, lakukan cek korelasi antar atribut untuk memilih atribut yang berkorelasi sedang-kuat terhadap atribut gaji (*MonthlyIncome*). Kemudian, langkah selanjutnya adalah drop atribut yang memiliki nilai korelasi dibawah kriteria sedang-kuat dan hanya menyisakan atribut *Age*, *JobLevel*, *TotalWorkingYears*, *YearsAtCompany* sebagai variabel independen dan *MonthlyIncome* sebagai variabel dependen.

b. Proses *Data Mining* dan Pengetahuan

Pada tahap ini, hal yang dilakukan adalah melakukan pemilihan metode yang sesuai dengan karakter data yang dikenal dengan istilah *Modelling*. Pada model ini digunakan proses *Data Mining Prediction*. Proses *Data Mining & Pengetahuan* yang dilakukan ini dengan melakukan perbandingan pengaruh variabel dependen dengan mengacu korelasi antara tiap-tiap variabel dependen dan variabel independen. Penerapan model yang digunakan adalah *Linear Regression Multivariate* menggunakan *Scikit Learn*. Untuk membuat model *machine learning*, ditentukan terlebih dahulu variabel dependen dan independennya. *Age*, *JobLevel*, *TotalWorkingYears*, *YearsAtCompany* sebagai variabel independen dan *MonthlyIncome* sebagai variabel dependen. Dari model linier regresi tersebut didapat koefisien independen yaitu -5,054 untuk *Age*, 3871,7530 untuk *JobLevel*, 46,9405 untuk *TotalWorkingYears*, -9,8460 untuk *YearsAtCompany* dan variabel dependen -1728 untuk *MonthlyIncome*. Berikut persamaan linear yang dikutip pada persamaan 3.

$$Y = -1728 - 5,054X_1 + 3871,7530X_2 + 46,9405X_3 - 9,8460X_4$$

$Y = \text{MonthlyIncome}$ (Variabel Dependen)

$X_1 = \text{Age}$ (Variabel Independen-1)

$X_2 = \text{JobLevel}$ (Variabel Independen-2)

$X_3 = \text{TotalWorkingYears}$ (Variabel Independen-3)

$X_4 = \text{YearsAtCompany}$ (Variabel Independen-4)

Maka dapat disimpulkan persamaan regresi linier multivariabel yang dikutip pada persamaan 4 sebagai berikut.

$$\text{MonthlyIncome} = -1728 - 5,054(\text{Age}) + 3871,7530(\text{JobLevel}) + 46,9405(\text{TotalWorkingYears} - 9,8460(\text{YearsAtCompany}))$$

c. Evaluasi Data

a. Validasi Model

Validasi model *machine learning* menggunakan model OLS dan *dmatrixes*. Berikut di bawah ini tabel validasi OLS.

OLS Regression Results						
Dep. Variable:	MonthlyIncome	R-squared:	0.909			
Model:	OLS	Adj. R-squared:	0.909			
Method:	Least Squares	F-statistic:	2571.			
Date:	Mon, 10 Jan 2022	Prob (F-statistic):	0.00			
Time:	10:58:11	Log-Likelihood:	-8944.9			
No. Observations:	1029	AIC:	1.790e+04			
Df Residuals:	1024	BIC:	1.792e+04			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-1728.5202	230.587	-7.496	0.000	-2180.998	-1276.043
Age	-5.0543	6.905	-0.732	0.464	-18.605	8.496
JobLevel	3871.7530	65.635	58.989	0.000	3742.958	4000.548
TotalWorkingYears	46.9406	11.733	4.001	0.000	23.917	69.965
YearsAtCompany	-9.8460	9.767	-1.008	0.314	-29.012	9.320
Omnibus:	12.798	Durbin-Watson:	2.069			
Prob(Omnibus):	0.002	Jarque-Bera (JB):	15.262			
Skew:	-0.182	Prob(JB):	0.000485			
Kurtosis:	3.472	Cond. No.	213.			

Gambar 1. Hasil OLS Regresi

b. Uji F (ANOVA)

Uji kelayakan model (koefisien regresi) atau disebut dengan uji F, yaitu untuk mengetahui apakah variabel independen yang terdapat dalam persamaan tersebut di atas secara bersama-sama berpengaruh signifikan pada nilai variabel dependen [28]. F-test atau ANOVA (Analysis of Variance) dalam regresi multi-linier dapat digunakan untuk menentukan apakah model yang dibuat berkinerja lebih baik daripada model yang lebih sederhana. Hipotesa yang didapat dari tabel uji F adalah:

H_0 = Variabel independen secara simultan bukan penjelas yang signifikan terhadap variabel dependen (model tidak cocok).

H_1 = Variabel independen secara simultan merupakan penjelas yang signifikan terhadap variabel dependen (model cocok).

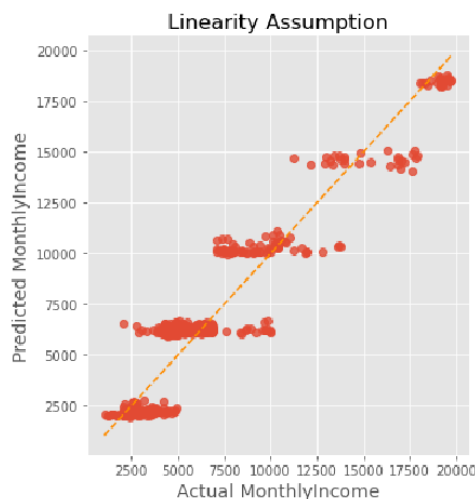
Berdasarkan gambar 1, dapat diketahui bahwa $F_s > P$ -value, yang artinya hipotesa yang dapat diambil adalah terima H_1 dan tolak H_0 . Dapat dikatakan, variabel independen (*Age*, *JobLevel*, *TotalWorkingYears*, *YearsAtCompany*) dan *MonthlyIncome* berpengaruh signifikan terhadap permintaan. Pada taraf signifikansi 5% (0,05), H_0 ditolak karena nilai probabilitasnya yaitu 0,00 yang berarti dibawah dari 5%. Maka dapat disimpulkan, model yang dipakai cocok.

c. R-Squared

R square dapat diartikan sebagai koefisien determinasi. R square menunjukkan suatu persentase pengaruh antara variabel X1 dengan X2 terhadap variabel Y [29]. Berdasarkan gambar 10, nilai koefisien determinasi (R-Square) adalah 0,909 atau 90,9%. Maka, *MonthlyIncome* dipengaruhi oleh faktor *Age* dan *YearsAtCompany* sebesar 0,909 atau 90,9%. Nilai sisa dari koefisien determinasi adalah 0,091 atau 9,1% dipengaruhi oleh faktor lain yang tidak diketahui.

d. Linearitas

Pengujian linearitas merupakan pengujian yang dilakukan untuk mengetahui model yang dibuktikan apakah merupakan model linear atau tidak [30]. Uji linearitas dilakukan dengan menggunakan regresi kurva, yaitu gambaran hubungan linier antara variabel X dengan variabel Y [30].

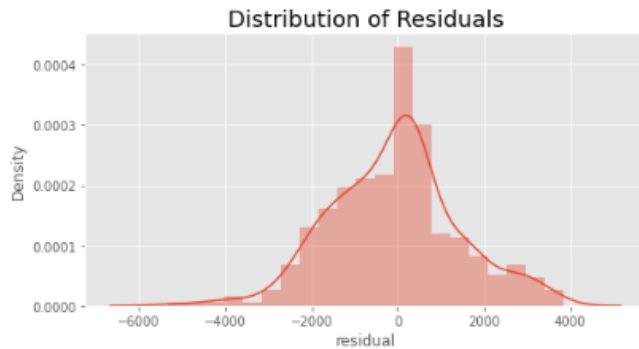


Gambar 2. Grafik Asumsi Linier

Plot sebar menunjukkan sisa yang tersebar merata di sekitar garis diagonal, sehingga dapat diasumsikan bahwa ada hubungan linier antara variabel independen dan dependen.

e. Normalitas

Uji normalitas merupakan pengujian suatu model regresi berupa variabel dependen, variabel independen ataupun keduanya apakah mempunyai distribusi normal atau tidak [30]. Model regresi dikatakan baik jika distribusi data normal atau mendekati normal [30].



Gambar 3. Diagram Distribusi Residual

Berdasarkan asumsi di atas, dapat diketahui hipotesa sebagai berikut.

H_0 = Residual terdistribusi normal.

H_1 = Residual terdistribusi secara tidak normal.

Dari hasil perhitungan, dapat diketahui bahwa nilai p-value yang dihitung menggunakan metode Anderson Darling adalah 0,00032261. Angka tersebut berada dibawah nilai threshold yang ditentukan yaitu 0,05, yang berarti H_0 ditolak H_1 diterima. Sehingga asumsi normalitas terpenuhi.

f. Multikolinearitas

Dalam analisis regresi linier ganda, jika ada dua atau lebih variabel independen yang berkorelasi sangat kuat, maka dikatakan terdapat multikolinieritas [31]. Uji multikolinieritas dilakukan untuk melihat korelasi antar variabel prediktor. Apabila terjadi multikolinieritas pada model regresi menyebabkan parameter regresi yang dihasilkan akan memiliki *error* yang sangat besar. Kriteria yang digunakan untuk mengetahui adanya multikolinieritas antara variabel prediktor adalah dengan menggunakan nilai *variance inflation factors* (VIF). Apabila nilai VIF lebih besar dari 10 mengindikasikan bahwa ada masalah multikolinieritas. Nilai VIF diperoleh dengan cara meregresikan variabel independen [32].

	VIF	variable
0	28.655370	Intercept
1	1.690786	Age
2	2.489052	JobLevel
3	4.140803	TotalWorkingYears
4	1.739893	YearsAtCompany

Gambar 4. Tabel VIF

Berdasarkan gambar 4, dapat dilihat nilai variabel Age, JobLevel, TotalWorkingYears, YearsAtCompany memiliki nilai kurang dari 10 sehingga dengan menggunakan tingkat signifikansi sebesar 0,05 dapat disimpulkan bahwa pada data tersebut tidak terdapat multikolinieritas pada variabel-variabel prediktor.

g. Autokorelasi

Autokorelasi merupakan pengujian untuk menguji apakah dalam sebuah model regresi linier berganda terdapat korelasi antara kesalahan pengganggu pada periode t dengan kesalahan pada periode t1 (sebelumnya). Jika terjadi korelasi, maka hal itu disebut sebagai autokorelasi. Model regresi yang baik adalah terbebas dari autokorelasi. [30] Pada langkah ini, dilakukan perhitungan skor Durbin-Watson menggunakan `durbin_watson()` fungsi dari `statsmodel` yang dibuat, kemudian menilainya dengan kondisi sebagai berikut.

Jika skor Durbin-Watson kurang dari 1,5 maka terdapat autokorelasi positif dan asumsi tidak terpenuhi.

Jika skor Durbin-Watson antara 1,5 – 2,5 maka tidak ada autokorelasi dan asumsi puas.

Jika skor Durbin-Watson lebih dari 2,5 maka terdapat autokorelasi negative dan asumsi tidak puas.

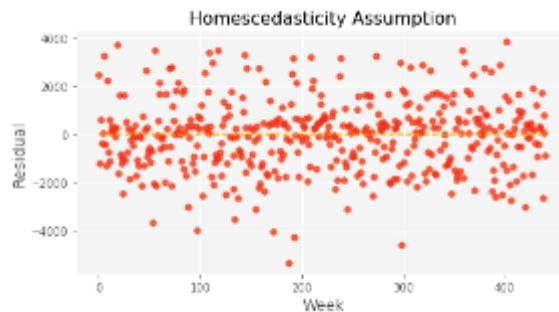
```
Durbin-Watson: 2.160636228778726  
Little to no autocorrelation  
  
Assumption satisfied
```

Gambar 5. Output Uji Autokorelasi

Dari hasil *output* pada gambar 5, dapat diasumsikan bahwa terdapat sedikit atau tidak ada autokorelasi, yang berarti asumsi puas.

h. Homoskedastisitas

Homoskedastisitas merupakan pengujian untuk menguji apakah dalam sebuah model regresi linier berganda terdapat gejala heteroskedastisitas atau tidak dengan cara melihat ada tidaknya pola tertentu pada grafik *scatterplots*. Model regresi dikatakan baik apabila model tersebut homoskedastisitas atau tidak terjadi heteroskedastisitas. [33]



Gambar 6. Plot Penyebaran Residual

Dari grafik *scatterplot* (gambar 6), terlihat titik- titik residual menyebar secara acak, serta tersebar baik di atas maupun di bawah angka 0 (nol) pada sumbu Y. Oleh karena itu, dapat diambil kesimpulan bahwa tidak terdapat gejala heteroskedastisitas pada model regresi yang digunakan.

2. Implementasi Antarmuka Aplikasi

Berikut di bawah ini hasil implementasi prediksi gaji dengan menggunakan *framework* Django.

Gambar 7. Antarmuka Aplikasi Prediksi Gaji Karyawan

V. KESIMPULAN

1. Kesimpulan

Berdasarkan model yang diambil dari model OLS didapatkan nilai akurasi sebesar 0,909. Akurasi tersebut merupakan nilai akurasi yang baik, sehingga dapat dikatakan model *machine learning* dapat berperforma baik untuk memprediksi gaji. Model prediksi yang dirancang dengan menggunakan *machine learning* dengan pendekatan regresi, berhasil melewati semua pengujian dalam langkah validasi model, sehingga pada penelitian ini dapat disimpulkan bahwa model algoritma yang digunakan dapat berperforma baik untuk memprediksi gaji karyawan dengan menggunakan empat variabel independen, yaitu Age, JobLevel, TotalWorkingYears, dan YearsAtCompany.

Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa MonthlyIncome dipengaruhi oleh faktor independen (Age, YearsAtCompany) sebesar 0,909 atau 90,9%. Nilai sisa dari akurasi tersebut adalah 0,091 atau 9,1% yang artinya MonthlyIncome dipengaruhi oleh faktor lain yang tidak diketahui sebesar 9,1%.

Visualisasi data dari hasil model prediksi gaji karyawan dapat digunakan menjadi bentuk aplikasi berbasis *web base* dengan menggunakan *framework* Django. Dengan aplikasi tersebut, admin dapat melakukan prediksi gaji karyawan dengan mudah dan dengan cepat.

2. Saran

Saran yang dapat disampaikan pada peneliti yang akan melanjutkan dan mengembangkan penelitian ini adalah :

- a. Pembuatan model prediksi yang digunakan dapat lebih beragam untuk membandingkan performa antara model satu dengan model yang lainnya.
- b. Sumber data yang digunakan kurang maksimal. Pada penelitian ini, hanya didapatkan *real* yang berasal dari Kaggle. Diharapkan kedepannya dapat menggunakan data *real* langsung dari perusahaan.

REFERENSI

- [1] Prasetyo B and Trisyanti U, "Revolusi Industri 4.0 Dan Tantangan Perubahan Sosial", Journal of Proceedings Series, no. 5, pp. 22-27, Nov. 2018, doi : <http://dx.doi.org/10.12962/j23546026.y2018i5.4417>
- [2] H. Prasetyo and W. Sutopo, "Perkembangan Keilmuan Teknik Industri Menuju Era Industri 4.0", Seminar dan Konferensi Nasional IDEC, vol. 2017, pp. 488-495, May .2017, doi : https://idec.ft.uns.ac.id/wp-content/uploads/2017/11/Prosiding2017_ID069.pdf
- [3] O. C. Pangaribuan and I. Irwansyah, "Media Cetak Indonesia di Era Revolusi Industri 4.0," Jurnal Pewarta Indonesia, vol. 1, no. 2, pp. 134-145, Oct. 2019, doi: <https://dx.doi.org/10.25008/jpi.v1i2.11>
- [4] A. A. Shahroom and N. Hussin, "Industrial Revolution 4.0 and Education," International Journal of Academic Research in Business and Social Sciences, vol. 8, no. 9, pp. 314-319, Oct. 2018, doi: <https://doi.org/10.24114/jh.v10i1.14138>
- [5] S. Kergroach, "Industry 4.0: New Challenges And Opportunities For The Labour Market," Foresight and STI Governance, vol. 11, no. 4, pp. 6-8, 2017, doi: <http://dx.doi.org/10.17323/2500-2597.2017.4.6.8>
- [6] M. I. Manda and S. ben Dhaou, "Responding to the challenges and opportunities in the 4th industrial revolution in developing countries", PervasiveHealth: Pervasive Computing Technologies for Healthcare, Part F148155, pp. 244-253, 2019, doi: <http://dx.doi.org/10.1145/3326365.3326398>
- [7] Y. Adrianova Eka Tuah and Anyan, "Implementasi Model Regresi Linear Sederhana Untuk Prediksi Gaji Berdasarkan Pengalaman Lama Bekerja", Journal Education and Technology, vol. 1, no. 2, pp. 56-70 Dec. 2020, doi : <https://doi.org/10.31932/jutech.v1i2.1289>
- [8] Tamrin A.S, Rumapea Patar, Mambo R, "Pengaruh Profesionalisme Kerja Pegawai Terhadap Tingkat Kepuasan Pelanggan Pada Kantor Pt. Taspen Cabang Manado", Jurnal Administrasi Publik, vol. 3, no. 46, pp. 1-9 2017, doi : , <https://ejournal.unsrat.ac.id/index.php/JAP/article/view/16283>
- [9] P.E.N. Desak, S. Made, "Universitas Udayana Fakultas Matematika Dan Ilmu Pengetahuan Alam Jurusan Matematika", Conference: Prosiding Seminar Nasional Matematika II, vol. 2, pp. 43-54, Oct. 2016, doi: https://simdos.unud.ac.id/uploads/file_penelitian_1_dir/24a473ea40f085c51398cd477f586a3a.pdf
- [10] Akbar Iskandar, Muttaqin, Sarini Vita Dewi, Jamaludin, Irawati HM Cahyo Prianto, Rosmita Sari Siregar, Muhammad Noor Hasan Siregar Dina Chamidah, Marzuki Sinambela, Albinur Limbong Yusra Fadhillah, Janner Simarmata, "Statistika Bidang Teknologi Informasi", 1st ed. Yayasan Kita Menulis, 2021.
- [11] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive text mining and database fusion approach," Environmental Health Perspectives, vol. 127, no. 9, Sep. 2019, doi: 10.1289/EHP4713.
- [12] D. Saputra and R. Fathoni Aji, "Analisis Perbandingan Performa Web Service Rest Menggunakan Framework Laravel, Django Dan Ruby On Rails Untuk Akses Data Dengan Aplikasi Mobile (Studi Kasus:Portal E-Kampus STT Indonesia Tanjungpinang)," Bangkit Indonesia, vol. 2, no. 7, pp. 17-22, Oct 2018.
- [13] E. P. Ariesanto Akhmad, "Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran," Jurnal Aplikasi Pelayaran dan Kepelabuhanan, vol. 10, no. 2, p. 120, Dec. 2020, doi: <https://dx.doi.org/10.30649/japk.v10i2.83>
- [14] K. Puteri and A. Silvanie, "Machine Learning Untuk Model Prediksi Harga Sembako Dengan Metode Regresi Linier Berganda", Jurnal Nasional Informatika, vol. 1, no. 2, pp. 82-94, Oct. 2020, doi : <https://ejournal-ibik57.ac.id/index.php/junif/article/view/134>
- [15] A. Saiful, S. Andryana, and A. Gunaryati, "Prediksi Harga Rumah Menggunakan Web Scrapping Dan Machine learning Dengan Algoritma Linear Regression", Jurnal Teknik Informatika dan Sistem Informasi, vol. 8, no. 1, pp. 41-50, Mar. 2012, doi : <https://jurnal.mdp.ac.id/index.php/jatisi/article/download/701/219/>
- [16] M. W. Pertiwi and R. E. Indrajit, "Metode Regresi Linier Untuk Prediksi Pengadaan Inventaris Barang", Simposium Nasional Ilmu Pengetahuan dan Teknologi (SIMNASIPTEK) 2017, vol. 1, no. 1, pp. 27-30, 2017, doi : <https://seminar.bsi.ac.id/simnasiptek/index.php/simnasiptek-2017/article/view/114>
- [17] W. Wahyudin and H. Purwanto, "PREDIKSI KASUS COVID-19 DI INDONESIA MENGGUNAKAN METODE BACKPROPAGATION DAN REGRESI LINEAR," Journal of Information System, Applied, Management, Accounting and Research, vol. 5, no. 2, p. 331, May 2021, doi: <https://doi.org/10.52362/jisamar.v5i2.420>
- [18] N. Nafi'iyah, "Perbandingan Regresi Linear, Backpropagation Dan Fuzzy Mamdani Dalam Prediksi Harga Emas," Seminar Nasional Inovasi Dan Aplikasi Teknologi Di Industri (SENIATI) 2016, vol. 2, pp. 291-296, Mar. 2016, doi : <https://ejournal.itn.ac.id/index.php/seniati/article/download/840/767/>
- [19] F. Ginting, E. Buulolo, and E. R. Siagian, "Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang)," KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), vol. 3, no. 1, Nov. 2019, doi: <http://dx.doi.org/10.30865/komik.v3i1.1602>
- [20] P. Katemba and K.D. Rosita, "PREDIKSI TINGKAT PRODUKSI KOPI MENGGUNAKAN REGRESI LINEAR", Jurnal Ilmiah Flash, vol. 3, pp. 42-51, Jun. 2017, doi: <http://jurnal.pnk.ac.id/index.php/flash/article/view/136/79>
- [21] T. N. Putri, A. Yordan, and D. H. Lamkaruna, "Peramalan Penerimaan Mahasiswa Baru Universitas Samudra Menggunakan Metode Regresi Linear Sederhana," Jurnal Teknologi Informatika, vol. 2, no. 1, Mar. 2019, doi : <http://jurnal.umm.ac.id/index.php/J-TIFA/article/view/237/149>
- [22] D. Sayan, B. Rupashri, M. Ayush, "Salary Prediction Using Regression Techniques.", Proceedings of Industry Interactive, Innovations in Science, Engineering & Technology, Jan. 2020, doi : <https://dx.doi.org/10.2139/ssrn.3526707>

- [23] U. Bansal, A. Narang, A. Sachdeva, I. Kashyap, and S. P. Panda, "Empirical Analysis Of *Regression* Techniques By House Price And Salary Prediction," IOP ConferenceSeries: Materials Science and Engineering, vol. 1022, no. 1, pp. 1-13, Jan. 2021, doi: <https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012110>
- [24] X. Pan, X. Wan, H. Wang, and Y. Li, "The Correlation Analysis Between Salary Gap and Enterprise Innovation Efficiency Based on the Entrepreneur Psychology," *Frontiers in Psychology*, vol. 11, Aug. 2020, doi: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01749/full>
- [25] S. Dan and B. Pratikno, "REGRESI LINEAR BIVARIAT SIMPEL DAN APLIKASINYA PADA DATA CUACA DI CILACAP", *JMP*, vol. 6, no. 1, pp. 45-52, Jun. 2014, doi : <http://dx.doi.org/10.20884/1.jmp.2014.6.1.2902>
- [26] T. N. Padilah and R. I. Adam, "ANALISIS REGRESI LINIER BERGANDA DALAM ESTIMASI PRODUKTIVITAS TANAMAN PADI DI KABUPATEN KARAWANG", *Jurnal Pendidikan Matematika dan Matematika*, vol. 5, no. 2, pp. 117-128, Dec. 2019, doi : <https://jurnal.umj.ac.id/index.php/fbc/article/view/3333>
- [27] J. Homepage, A. Roihan, P. Abas Sunarya, and A. S.Rafika, "Pemanfaatan *Machine learning* dalam Berbagai Bidang: Review paper," *Indonesian Journal on Computer and Information Technology*, vol. 5, no. 2, pp. 75-82, Maret 2020.
- [28] Dita Anggun Lestari, Sarini Abdullah, "Analisis Tingkat Kesehatan Dan Efisiensi Perbankan Terhadap Profitabilitas Bank Menggunakan Regresi Berganda Dan Anova", *Indonesian Journal of Statistics and Its Applications*, vol. 4, no. 3, pp. 401-418, 2020.
- [29] E. Khumaedi, "Pengaruh Disiplin Dan Motivasi Kerja Terhadap Kinerja Pegawai Pada Dinas Sentra Operasi Terminal PT.Angkasa Pura II," *Jurnal Ilmiah Manajemen dan Bisnis*, vol. 2, no. 1, pp. 66-77, Mar 2016.
- [30] N. Sitti, K. Sekolah, T. Ilmu, and E. Gempol, "Analisis Ekuitas Merek Produk Notebook Asus Terhadap Keputusan Pembelian Konsumen Pada Distributor Diva Jaya Cabangsidoarjo," *Jurnal Akuntansi Dan Manajemen*, vol. 3, no. 2, pp. 73-83, 2018.
- [31] Prof. Dr. Suyono, M.Si. "Analisis Regresi untuk Penelitian". 1st ed. Sleman : deepublish. 2015.
- [32] R. G. Ali and J. Nugraha, "Penerapan Metode Regresi Ridge Dalam Mengatasi Masalah Multikolinearitas Pada Kasus Indeks Pembangunan Manusia Di Indonesia Tahun 2017," *Prosiding Sendika*, vol. 5, no. 22, pp. 226-235, 2019. Available: www.statistik.data.kemdikbud.go.id
- [33] O. : Nurfajar, ; M Syafiq Marzuqi, N. Rohmayati, U. Sultan, and A. Tirtayasa, "Pengaruh Employee Engagement Dan Efikasi Diri Terhadap Kinerja Karyawan Pt Nikomas Gemilang Divisi Pci S5 Serang Banten," *Jurnal Pengembangan Wiraswasta*, vol. 20, no. 1, pp. 35- 46, 2018. [Online]. Available: <http://ejurnal.steipwija.ac.id/index.php/jpw>