

Analisis Data Pegawai untuk Memprediksi Gaji Berdasarkan Faktor-Faktor Spesifik dengan Pendekatan Machine Learning

Syafrial Fachri Pane^{#1}, Amri Yanuar^{#2}, Bachtiar Ramadhan^{#3}, Nur Tri Ramadhanti Adiningrum^{#4}

[#]Diploma IV/ Program Studi Teknik Informatika, Universitas Logistik Bisnis Internasional Jl. Sariasih Nomor 54, Bandung 40501, Jawa Barat, Indonesia

¹syafrial.fachri@poltekpos.ac.id ²amriyanuar@poltekpos.ac.id

³1204077_bachtiar@students.poltekpos.ac.id ⁴1204061_nur@students.poltekpos.ac.id

Abstract— Providing an appropriate salary is one of the important factors to boost employee performance. Unfortunately, the current development of the company does not have a decision media to predict employee salaries based on data quality. In predicting, not many factors are taken into consideration, so this study aims to predict salaries based on specific factors. The study was conducted using a larger number of factors than other studies. The more factors considered to predict salary, the better the model will be. The factors taken include the independent factor in the form of Age, JobLevel, TotalWorkingYears, YearsAtCompany and the dependent factor in the form of MonthlyIncome. The data analysis technique used is multivariate linear regression analysis. The model successfully passed the test in the model validation step which resulted in an accuracy value of 90.9%. So it can be concluded that the model can perform well. Furthermore, this research produces data visualization using the Django framework.

Keywords— Salary Prediction, Multivariate Linear Regression, Specific Factors, Web Base

Abstrak— Pemberian gaji yang sesuai merupakan salah satu faktor penting untuk mendorong kinerja karyawan. Sayangnya, perkembangan perusahaan saat ini belum memiliki media pengambilan keputusan untuk memprediksi gaji karyawan berdasarkan kualitas data. Dalam memprediksi gaji tidak banyak faktor yang dipertimbangkan, sehingga penelitian ini bertujuan untuk memprediksi gaji berdasarkan faktor-faktor tertentu. Penelitian ini dilakukan dengan menggunakan jumlah faktor yang lebih banyak dibandingkan penelitian lainnya. Semakin banyak faktor yang dipertimbangkan untuk memprediksi gaji, semakin baik modelnya. Faktor yang diambil antara lain faktor independen berupa Umur, JobLevel, Jumlah Tahun Kerja, Tahun Di Perusahaan dan faktor dependen berupa Pendapatan Bulanan. Teknik analisis yang dipakai adalah regresi linier multivariat. Model berhasil lolos uji pada tahap validasi model yang menghasilkan nilai akurasi sebesar 90,9%. Sehingga dapat disimpulkan bahwa model dapat bekerja dengan baik. Selanjutnya penelitian ini menghasilkan visualisasi data menggunakan framework Django.

Kata Kunci— Prediksi Gaji, Regresi Linier Multivariat, Faktor Spesifik, Web Base

I. PENDAHULUAN

Perkembangan ilmu pengetahuan serta teknologi di era Revolusi Industri 4.0 semakin berkembang dengan pesat. Revolusi Industri 4.0 sendiri mulai terjadi melalui bentuk rekayasa intelegensia dan juga *internet of thing* sebagai kerangka penggerak dan konektivitas antara manusia dengan mesin[1]. Sehingga, didapatkan kombinasi antara internet dan teknologi digital dengan industri konvensional, dimana hal tersebut memiliki tujuan untuk meningkatkan efisiensi, produktivitas serta pelayanan konsumen dengan signifikan[2]. Era revolusi ini akan mendisrupsi berbagai kegiatan diberbagai bidang seperti pada bidang teknologi, ekonomi, sosial, dan politik[1]. Saat ini, kehidupan sedang berada pada awal revolusi yang mampu mengubah cara bekerja, hidup, dan berkomunikasi satu sama lain[3].

Perubahan karakteristik pekerjaan adalah salah satu pengaruh tersendiri dari datangnya revolusi industri 4.0[4]. Perubahan karakteristik pekerjaan akan mencabut pekerjaan yang ada sebelumnya dan menggantikannya dengan karakteristik pekerjaan yang baru [5]. Karakteristik baru pada pekerjaan juga membutuhkan para pekerja dengan kompetensi baru[6]. Tentunya perusahaan harus siap untuk bersaing dengan perusahaan yang lain[7]. Selanjutnya, perusahaan perlu memiliki keunggulan dan manajemen yang efektif untuk menghadapi persaingan tersebut [7]. Dengan demikian salah satu aspek yang berdampak besar terhadap keberhasilan dan kemajuan dari suatu perusahaan ialah 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 upah atau gaji yang tepat oleh perusahaan untuk karyawannya adalah salah satu faktor yang berpengaruh secara internal terhadap kemajuan sebuah perusahaan. Selain itu, perusahaan pun diharuskan untuk bersedia mengeluarkan gaji tambahan 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

A. Machine Learning

Machine learning dapat didefinisikan sebagai suatu aplikasi pada komputer berupa bentuk algoritma matematika yang diangkat dengan model pembelajaran yang bersumber dari data dan dapat mengeluarkan prediksi untuk masa yang akan datang[9].

B. Regresi Linier Multivariat

Regresi linier multivariat adalah perpanjangan dari model regresi linier sederhana[10]. Model regresi linier multivariat dapat digunakan untuk melakukan prediksi hubungan antara satu variabel independen berdasarkan nilai dari dua atau lebih suatu variabel dependen [10]. Regresi linier berganda juga menghasilkan persamaan matematis [10]. Jika terdapat lebih dari dua variabel, maka jenis hubungan linier tersebut dapat dibuktikan pada persamaan regresi linier multivariat yang dikutip pada persamaan 1 serta persamaan 2 sebagai berikut :

$$Y'_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni} \quad (1)$$

$$Y' = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (2)$$

Keterangan :

Y = nilai-nilai hasil pengamatan

Y' = nilai regresi

$i = 1, 2, 3, k$

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

C. Scikit-Learn

Scikit-learn adalah *library* pada bahasa pemrograman python yang menyediakan banyak jenis algoritma *machine learning*. *Scikit-learn* menggunakan antarmuka berorientasi yang konsisten yang memudahkan untuk membandingkan metode. *Scikit-learn* mengintegrasikan berbagai algoritma *machine learning* untuk *supervised learning* dan *unsupervised learning*[11].

D. Framework Django

Django ialah sebuah web *framework* dengan basis bahasa pemrograman python yang dirancang untuk membangun suatu aplikasi web yang aman, kaya fitur dan dinamis[12]. *Framework* ini selalu mendapatkan *update* yang membuat web *framework* Django menjadi pilihan banyak *developer* aplikasi web [12].

E. Penelitian Sebelumnya

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

TABEL I
PENELITIAN SEBELUMNYA

Area Penelitian	Dataset	Metode
Penentu gaji dokter hewan [13]	Data survey dokter hewan	<i>Multiple Regression</i>
Penentu kepuasan kerja ahli gizi [14]	Data kuisisioner ahli gizi	<i>Multivariate Regression</i>
Prediksi gaji siswa [15]	Data siswa Pendidikan tinggi	Regresi OLS
Prediksi gaji tahunan[16]	Data global HealthEconomics.com	<i>Multiple Regression</i>
Analisi kontrak dan gaji tertinggi [17]	Data ketenagakerja-an	Regresi logistik
Analisi tingkat kompetitif gaji guru [18]	Data sensus Amerika Serikat	<i>Logarithmic Regressions</i>
Analisis gaji dan bakat guru[19]	Data Schools and Staffing Survey (SASS)	<i>Quantile Regression</i>
Prediksi gaji karyawan [20]	Data survey Google Form.	Regresi linier
Menentukan gaji karyawan [21]	Data rekap gaji dan kontrak karyawan	K-Means Clustering
Prediksi gaji dengan kecerdasan emosional [22]	Data survey penelitian 785 subjek	Regresi Multivariat
Prediksi gaji setelah tahun tertentu[23].	Data karyawan dari perusahaan.	Regresi Linear, Regresi Polinomial
Analisis empiris	Dataset gaji pegawai	<i>Multiple Linear</i>

harga rumah dan prediksi gaji[24]	dan harga rumah.	<i>Regression</i>
Prediksi gaji dalam penerapan model regresi [25]	Data gaji dan lama pengalaman bekerja	<i>Simple Linear Regression</i>
Analisis dan prediksi kepuasan gaji [26]	Data karyawan staf	<i>Simple Regression</i>
Perancangan sistem prediksi kenaikan gaji [27]	Data Enterprise Resource Planning (ERP)	<i>Linear regression, artificial neural networks, random forest regression</i>
Sistem prediksi gaji untuk meningkat-kan motivasi [28].	Data mahasiswa yang lulus dengan gajinya.	<i>K-NN, Naïve Bayes, Decision trees, Multilayer perception, SVM</i>
Prediksi gaji di pasar kerja TI [29].	Data <i>e-Recruitment</i> khusus untuk pekerjaan TI	<i>Logistic regression, KNN, Multi-layer perceptrons, SVM, Random forest, Vote, Vote3</i>
Analisis prediktif gaji SDM[30]	Data ketenagakerjaan	<i>Logistic regression SVM</i>
Mesin prediksi untuk memprediksi gaji[31]	Data kepegawaian	<i>Decision tree classifier, Random forest classifier</i>
Analisis prediktif untuk pendapatan alumni[32]	Data survei studi alumni	<i>Quantile Regression, Quantile Random Forest, Quantile Gradient Boosting, Linear Regression, Random Forest Classifier, Gradient Boosting Classifier</i>
Komputasi cloud fasilitas sinyal digital biomedis. [33]	Data biomedis	<i>Framework Django</i>
Aplikasi web prediksi diabetes [34]	Data klinis penyakit diabetes	<i>Framework Django</i>
Deployment klasifikasi penyakit paru-paru [35]	Dataset <i>x-ray</i> tubuh bagian atas untuk Covid-19, <i>Pneumonia</i> , dan <i>Normal</i> .	<i>Framework Django</i>

Pada penelitian ini, digunakan metode dengan jenis deskriptif kualitatif dengan menggunakan model regresi linier multivariat dengan bahasa pemrograman Python. Penelitian deskriptif yaitu metode penelitian yang digunakan untuk mendapatkan pengetahuan yang seluas-luasnya terhadap objek penelitian. Penelitian dekriptif ini menyajikan suatu gambar yang terperinci terhadap satu

kondisi khusus. Penelitian deskriptif bertujuan mendeskripsikan suatu keadaan dengan apa adanya dan menginterpretasikan objek sesuai dengan peristiwa, ataupun apapun yang berhubungan dengan variabel-variabel yang dapat dijelaskan dalam bentuk berupa angka maupun kata.

Dalam penelitian ini, dataset yang digunakan dalam proses regresi linier multivariat adalah dataset kepegawaian yang bersumber dari Kaggle. Dataset yang digunakan adalah *employee_attrition_train.csv* sebanyak 1029 baris dan 35 kolom, dan *employee_attrition_test.csv* sebanyak 441 baris dan 34 kolom. Dataset tersebut dapat diakses pada [link](https://www.kaggle.com/collearninglounge/employee-attrition) berikut:

<https://www.kaggle.com/collearninglounge/employee-attrition>

III. IMPLEMENTASI

Kebutuhan untuk pembuatan model *machine learning* dan aplikasi prediksi gaji pegawai adalah sebagai berikut :

- Aplikasi *software* : XAMPP 3.2.4, Lucidchart, Visual Studio Code, Jupyter Notebook.
- *Hardware* : laptop merk LENOVO dengan kriteria : Prosesor Intel™ Core i3-4030U, RAM 4 GB, 64-bit OS. Software : Ms. Office, Windows 10 Pro.

A. Implementasi Model Machine Learning

A.1. Himpunan Data

Pada tahap himpunan data, hal yang dilakukan yaitu memahami dan mempersiapkan data. Tahap ini dapat disebut dengan istilah *Data Preprocessing*. Metode yang digunakan dalam *Data Preprocessing* pada model ini adalah *Data Cleaning*. Berikut tahapan himpunan data:

```
# Basic Library
import pandas as pd
import numpy as np

# Data Visualization
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import seaborn as sns
from scipy.stats import skew

# Model Building
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm
```

Gambar 1 Import Library

```
df_train = pd.read_csv('E:\employee_attrition_train.csv')
df_train
```

Gambar 2 Import Dataset

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.

```
# Encoder BusinessTravel Variable
# converting type of columns to 'category'
df_train['BusinessTravel'] = df_train['BusinessTravel'].astype('category')
# Assigning numerical values and storing in another column
df_train['BusinessTravel'] = df_train['BusinessTravel'].cat.codes

# Encoder Department Variable
df_train['Department'] = df_train['Department'].astype('category')
# Assigning numerical values and storing in another column
df_train['Department'] = df_train['Department'].cat.codes
```

Gambar 3 Skrip Encoder Variabel

Pada dataset yang digunakan, terdapat nilai NaN/Null. Oleh karena itu, dilakukan proses pengisian nilai tersebut dengan nilai rata-rata (*mean*) variabelnya.

```
Age = df_train['Age']
df_train.Age = df_train.Age.fillna(value=df_train.Age.mean())
```

Gambar 4 Skrip Proses Pergantian Nilai Null menjadi Nilai Mean

Setelah semua data berbentuk integer, lakukan cek korelasi antar atribut untuk memilih atribut yang berkorelasi sedang- kuat terhadap atribut gaji (MonthlyIncome).

```
df_train_clean.corr().abs()
```

Gambar 5 Skrip Cek Tabel Korelasi

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.

```
df_train_clean = df_train_clean.drop(['Attrition', 'BusinessTravel', 'DailyRate', 'Department',
'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
'JobInvolvement', 'JobRole', 'JobSatisfaction',
'MaritalStatus', 'MonthlyRate', 'NumCompaniesWorked',
'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
'TrainingTimesLastYear', 'WorkLifeBalance',
'YearsInCurrentRole', 'YearsSinceLastPromotion',
'YearsWithCurrManager'], axis=1)
```

Gambar 6 Skrip Drop Atribut

A.2. Proses Data Mining & Pengetahuan

Pada tahap proses *Data Mining & Pengetahuan*, proses yang dilakukan adalah pemilihan jenis metode yang sesuai dengan karakter data. Proses ini dikenal dengan istilah *Modelling*. Pada model ini digunakan proses *Data Mining Prediction*. Proses *Data Mining & Knowledge* dilakukan dengan membandingkan pengaruh variabel dependen dengan mengacu pada korelasi antara masing-masing variabel dependen dan variabel independen. Model implementasi yang digunakan adalah Regresi Linier Multivariat dari modul *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.

```
# Menentukan variabel X dan variabel Y
x_train = df_train_clean[['Age', 'JobLevel', 'TotalWorkingYears', 'YearsAtCompany']]
y_train = df_train_clean[['MonthlyIncome']]
```

Gambar 7 Skrip Penentuan Variabel Dependen dan Independen

```
regressor = LinearRegression()
persamaan = regressor.fit(x_train, y_train)
print(regressor.coef_)
print(regressor.intercept_)
```

Gambar 8 Skrip Penentuan Variabel Dependen dan Independen

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 :

$$Y = -1728 - 5,054X_1 + 3871,7530X_2 + 46,9405X_3 - 9,8460X_4 \quad (3)$$

Y = MonthlyIncome (Variabel Terikat)
 X_1 = Age (Variabel Bebas pertama)
 X_2 = JobLevel (Variabel Bebas kedua)
 X_3 = TotalWorkingYears (Variabel Bebas ketiga)
 X_4 = YearsAtCompany (Variabel Bebas keempat)

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

$$\text{MonthlyIncome} = -1728 - 5,054(\text{Age}) + 3871,7530(\text{JobLevel}) + 46,9405(\text{TotalWorkingYears}) - 9,8460(\text{YearsAtCompany}) \quad (4)$$

A.3. Evaluasi Data

1) Validasi Model

Validasi model *machine learning* menggunakan model OLS dan *dmatrixes*.

```
X = df_train_clean[['Age', 'JobLevel', 'TotalWorkingYears', 'YearsAtCompany']]
X = sm.add_constant(X) # adding a constant

olsmod = sm.OLS(df_train['MonthlyIncome'], X).fit()
print(olsmod.summary())
```

Gambar 9. Skrip 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 10. Hasil OLS Regresi

2) Uji F (ANOVA)

Uji kelayakan model disebut juga dengan uji F, merupakan pengujian yang dilakukan untuk mengetahui apakah variabel bebas yang digunakan dalam melihat persamaan tersebut mempunyai pengaruh begitu signifikan terhadap nilai variabel terikat[36].

Uji ANOVA pada regresi berganda digunakan untuk melihat apakah model tersebut bekerja lebih baik atau tidak dari model yang lebih sederhana sebelumnya.

```
print('F-statistic:', olsmod.fvalue)
print('Probability of observing value at least as high as F-statistic:', olsmod.f_pvalue)
```

Gambar 11. Skrip Uji F (ANOVA)

```
F-statistic: 2570.622889791836
Probability of observing value at least as high as F-statistic: 0.0
```

Gambar 12. Output Uji F (ANOVA)

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 penjelas yang signifikan terhadap variabel dependen (model cocok).

Berdasarkan gambar 12, 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.

3) Uji-t

Uji koefisien regresi atau dikenal sebagai uji-t memiliki tujuan untuk menguji taraf signifikan konstanta variabel independen yang ditemukan pada persamaan

regresi secara individu[36].

Hipotesa yang ada yaitu :

H_0 = Tidak ada pengaruh signifikan terhadap Variabel independen.

H_1 = Ada pengaruh signifikan terhadap Variabel independen.

$\alpha = 0,05$ (Taraf Signifikansi/Threshold)

Berdasarkan gambar 10, uji-t dapat diambil hipotesa sebagai berikut :

- Nilai variabel X_1 (Age) berada di atas taraf signifikansi/terima H_1 .
- Nilai variabel X_2 (JobLevel) berada di bawah tarafsignifikansi/terima H_0 .
- Nilai variabel X_3 (TotalWorkingYears) di bawah tarafsignifikansi/terima H_0 .
- Nilai variabel X_4 (YearsAtCompany) di atas taraf signifikansi/terima H_1 .

Kesimpulannya adalah variabel bebas (independen) JobLevel dengan variabel terikat (dependen) yaitu MonthlyIncome. Sedangkan variabel independent Age dan YearsAtCompany adalah variabel yang mempengaruhi variabel dependen yaitu MonthlyIncome.

4) R-Square

R-square dapat diartikan sebagai koefisien determinasi. *R square* menunjukkan proporsi pengaruh antara variabel bebas (independen) terhadap variabel terikat (dependen)[37].

Berdasarkan gambar 10, nilai koefisien determinasi (*R-square*) adalah 0,909 (90,9%). Maka, MonthlyIncome dipengaruhi oleh faktor Age dan YearsAtCompany sebesar 0,909 (90,9%). Nilai residual koefisien determinasi sebesar 0,091 (9,1%) dipengaruhi oleh faktor lainnya yang tidak diketahui.

5) Pengujian Asumsi

Berikut ini adalah berbagai pengujian asumsi yang dilakukan untuk melihat validitas model :

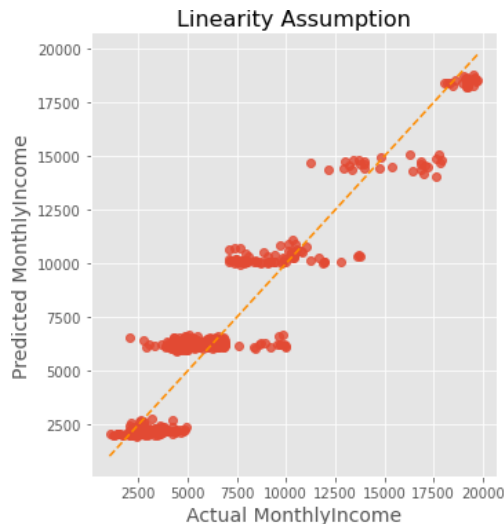
- Linearitas
Linearitas merupakan uji yang digunakan untuk mengetahui apakah model yang dipakai merupakan model linear atau tidak[38]. Uji ini dilakukan dengan melihat kurva regresi, yaitu kurva yang digunakan untuk melihat hubungan linear antara variabel bebas dengan variabel terikat[38].


```
# Plotting the observed vs predicted values
sns.lmplot(x='MonthlyIncome', y='MonthlyIncome Prediction', data=df_test_new, fit_reg=False, size=5)

# Plotting the diagonal line
line_coords = np.arange(vis_test[['MonthlyIncome', 'MonthlyIncome Prediction']].min().min()-10,
                        vis_test[['MonthlyIncome', 'MonthlyIncome Prediction']].max().max()+10)
plt.plot(line_coords, line_coords, # X and y points
         color='darkorange', linestyle='--')

plt.xlabel('Predicted MonthlyIncome', fontsize=14)
plt.ylabel('Actual MonthlyIncome', fontsize=14)
plt.title('Linearity Assumption', fontsize=16)
plt.show()
```

Gambar 13. Skrip Linearitas dan Grafik Asumsi Linier



Gambar 14. Grafik Asumsi Linier

Pada Gambar 14, *scatter plot* tersebar disekitar garis diagonal, sehingga dapat diasumsikan bahwa terdapat hubungan linier diantara variabel bebas (independen) dan terikat (dependen).

- Normalitas

Uji normalitas merupakan uji untuk melihat apakah variabel-variabel yang ada terdistribusi normal atau tidak[38]. Model dikatakan baik apabila data terdistribusi secara normal ataupun mendekati normal[38].

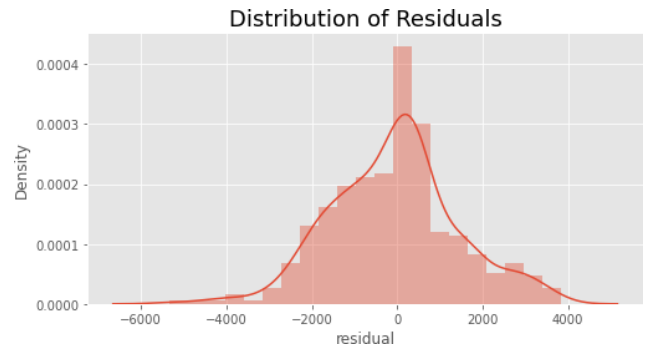
```
from statsmodels.stats.diagnostic import normal_ad

# Performing the test on the residuals
p_value = normal_ad(df_test_new2['residual'])[1]
print('p-value from the test Anderson-Darling test below 0.05 generally means non-normal:', p_value)

# Plotting the residuals distribution
plt.subplots(figsize=(8, 4))
plt.title('Distribution of Residuals', fontsize=18)
sns.distplot(df_test_new2['residual'])
plt.show()

# Reporting the normality of the residuals
if p_value < 0.05:
    print('Residuals are not normally distributed')
else:
    print('Residuals are normally distributed')
```

Gambar 15. Skrip Uji Normalitas



Gambar 16. Diagram Distribusi Residual

Berdasarkan asumsi di atas, dapat diketahui hipotesa sebagai berikut :

- H_0 = Residual terdistribusi secara 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 di bawah nilai threshold yang ditentukan yaitu 0,05 dimana berarti H_0 ditolak dan H_1 diterima. Sehingga asumsi normalitas terpenuhi.

- Multikolinieritas

Dalam analisis regresi linier berganda, jika terdapat dua atau lebih variabel independen yang berkorelasi sangat kuat, maka terdapat multikolinieritas[39].

Uji ini digunakan untuk melihat korelasi antar variabel bebas. Untuk melihat apakah terjadi multikolinieritas atau tidak, dibutuhkan kriteria yang ditetapkan melalui Variance Inflation Factors (VIF). Jika VIF bernilai lebih dari 10, maka ditemukan problematika pada multikolinearitas. Nilai VIF dapat diperoleh dengan melakukan regresi terhadap variabel bebas[40].

```
corr = vis_test[['Age', 'JobLevel', 'TotalWorkingYears', 'YearsAtCompany', 'MonthlyIncome']].corr()
print('Pearson correlation coefficient matrix of each variables:\n', corr)

# Generate a mask for the diagonal cell
mask = np.zeros_like(corr, dtype=np.bool)
np.fill_diagonal(mask, val=True)

# Initialize matplotlib figure
fig, ax = plt.subplots(figsize=(4, 3))

# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True, sep=100)
cmap.set_bad('grey')

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(corr, mask=mask, cmap=cmap, vmin=-1, vmax=1, center=0, linewidths=.5)
fig.suptitle('Pearson correlation coefficient matrix', fontsize=14)
ax.tick_params(axis='both', which='major', labelsize=10)
# fig.tight_layout()
```

Gambar 17. Skrip Uji Multikolinieritas

```
Pearson correlation coefficient matrix of each variables:
```

	Age	JobLevel	TotalWorkingYears	YearsAtCompany
Age	1.000000	0.449794	0.623246	0.294041
JobLevel	0.449794	1.000000	0.772222	0.511125
TotalWorkingYears	0.623246	0.772222	1.000000	0.638163
YearsAtCompany	0.294041	0.511125	0.638163	1.000000
MonthlyIncome	0.439321	0.944295	0.771358	0.488875

	MonthlyIncome
Age	0.439321
JobLevel	0.944295
TotalWorkingYears	0.771358
YearsAtCompany	0.488875
MonthlyIncome	1.000000

Gambar 18. Koefisien Korelasi Pearson

Dari hasil asumsi di atas, dapat dikatakan variabel-variabel saling berkorelasi.

```
from patsy import Dmatrices
from statsmodels.stats.outliers_influence import variance_inflation_factor

#find design matrix for linear regression model using 'rating' as response variable
y, X = dmatrices('MonthlyIncome ~ Age+JobLevel+TotalWorkingYears+YearsAtCompany', data=vis_test, return_type='dataframe')

#calculate VIF for each explanatory variable
vif = pd.DataFrame()
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif['variable'] = X.columns

#view VIF for each explanatory variable
vif
```

Gambar 19. Skrip Multikolinearitas

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

Gambar 20. Tabel VIF

Berdasarkan gambar 20, dapat dilihat nilai variabel Age, JobLevel, TotalWorkingYears, dan YearsAtCompany memiliki nilai VIF < 10 dan dengan menggunakan taraf signifikansi sebesar 0,05 maka dapat disimpulkan bahwa tidak terdapat multikolinearitas pada variabel prediktor.

- Autokorelasi

Autokorelasi adalah pengujian yang memiliki tujuan untuk melakukan pengujian model apakah terdapat hubungan antara kesalahan pengganggu di periode t dengan kesalahan di periode t1. Jika diindikasikan terdapat hubungan, maka model tersebut mengalami autokorelasi. Model yang baik adalah model yang terhindar dari autokorelasi [38].

Pada langkah ini, skor Durbin- Watson dihitung menggunakan fungsi `durbin_watson()` dari `statsmodel` yang dibuat, kemudian skor dengan ketentuan sebagai berikut :

1. Apabila skor < 1,5 maka terdapat autokorelasi positif dan asumsi tak terpenuhi.
2. Apabila skor antara 1,5 – 2,5 maka tidak terdapat autokorelasi dan asumsi puas.

3. Apabila skor > 2,5 maka terdapat autokorelasi negatif dan asumsi tak terpenuhi.

```
from statsmodels.stats.stattools import durbin_watson

durbinWatson = durbin_watson(df_test_new2['residual'])

print('Durbin-Watson:', durbinWatson)
if durbinWatson < 1.5:
    print('Signs of positive autocorrelation', '\n')
    print('Assumption not satisfied')
elif durbinWatson > 2.5:
    print('Signs of negative autocorrelation', '\n')
    print('Assumption not satisfied')
else:
    print('Little to no autocorrelation', '\n')
    print('Assumption satisfied')
```

Gambar 21. Skrip Uji Autokorelasi

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

```
Assumption satisfied
```

Gambar 22. Output Uji Autokorelasi

Dari hasil output pada gambar 20, dapat diasumsikan bahwa terdapat sedikit atau tidak ada autokorelasi, yang berarti asumsi puas.

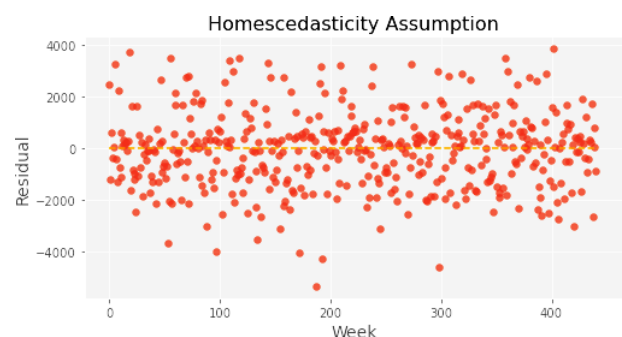
- Homoskedastisitas

Homoskedastisitas adalah pengujian yang memiliki tujuan untuk melakukan pengujian apakah model regresi linier multivariat terdapat gejala heteroskedastisitas atau tidak dengan cara melihat ada atau tidaknya suatu pola tertentu pada grafik *scatterplots*. Model regresi dapat dikatakan model baik apabila model tersebut mengalami homoskedastisitas atau tidak terjadinya heteroskedastisitas[41].

```
# Plotting the residuals
plt.subplots(figsize=(8, 4))
plt.scatter(x=df_test_new2.index, y=df_test_new2.residual, alpha=0.8)
plt.plot(np.repeat(0, len(df_test_new2.index)+2), color='darkorange', linestyle='--')

plt.ylabel('Residual', fontsize=14)
plt.xlabel('Week', fontsize=14)
plt.title('Homoscedasticity Assumption', fontsize=16)
plt.show()
```

Gambar 23. Skrip Plot Penyebaran Residual



Gambar 24. Plot Penyebaran Residual

Dari grafik scatterplot (gambar 22), terlihat bahwa titik residual menyebar acak. Titik residual tersebar baik di atas maupun di bawah titik nol. Dengan demikian, dapat diketahui bahwa model yang digunakan tidak terdapat gejala heteroskedastisitas.

B. Implementasi Antarmuka Aplikasi

B.1. Halaman Prediksi Gaji

Tampilan untuk aplikasi untuk melakukan prediksi gaji pegawai terlihat pada Gambar 25.

Gambar 25. Halaman Prediksi Gaji

B.2. Halaman hasil prediksi gaji

Tampilan untuk aplikasi menampilkan hasil prediksi gaji pegawai terlihat pada Gambar 26.

Gambar 26. Halaman Hasil Prediksi gaji

B.3. Halaman data pegawai

Tampilan aplikasi untuk menampilkan data pegawai yang terdiri dari atribut Id, Age, JobLevel, MonthlyIncome, dan Action. Halaman data pegawai terlihat pada Gambar 27.

Id	Age	JobLevel	MonthlyIncome	TotalWorkingYears	YearsAtCompany	Action
2	38.0	2	8463	6	5	Edit Delete
3	45.0	3	9724	25	1	Edit Delete
4	36.0	2	5914	16	13	Edit Delete
5	34.0	1	2579	8	8	Edit Delete
6	38.0	1	4230	6	5	Edit Delete
7	39.0	1	2232	7	3	Edit Delete

Gambar 27. Halaman Data Pegawai

B.4. Halaman tambah data pegawai

Tampilan aplikasi untuk menginputkan data pegawai yang baru. Form input terdiri dari Age, Job Level, Monthly Income, dan Total Working Years, dan Years At Company. Halaman tambah data pegawai terlihat pada Gambar 28.

Gambar 28. Halaman tambah data Pegawai

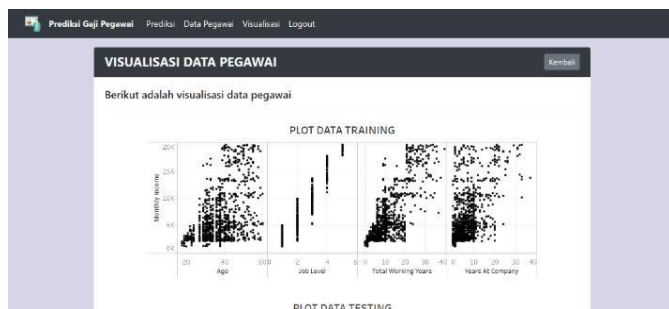
B.4. Halaman edit data pegawai

Tampilan untuk aplikasi mengedit data pegawai yang baru. Form input untuk mengedit data terdiri dari Age, Job Level, Monthly Income, dan Total Working Years, dan Years At Company. Halaman edit data pegawai terlihat pada Gambar 29.

Gambar 29. Halaman Edit Data Pegawai

B.5. Halaman visualisasi

Tampilan aplikasi untuk visualisasi dari model yang dibuat. Halaman visualisasi terlihat pada Gambar 30.



Gambar 30. Halaman Visualisasi

IV. KESIMPULAN DAN SARAN

A. 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 yang dirancang menggunakan *machine learning* dengan pendekatan regresi ini berhasil lolos uji validasi model, sehingga dapat dipastikan model yang dibuat dapat 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

B. Saran

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

- Pembuatan model prediksi yang digunakan dapat lebih beragam untuk membandingkan performa antara modelsatu dengan model yang lainnya.
- 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.

DAFTAR REFERENSI

- [1] B. Prasetyo and U. Trisyanti, "Revolusi Industri 4.0 Dan Tantangan Perubahan Sosial," *Prosiding SEMATEKSOS 3*, pp. 22–27, 2018, doi: 10.12962/j23546026.y2018i5.4417.
- [2] H. Prasetyo and W. Sutopo, "Perkembangan Keilmuan Teknik Industri Menuju Era Industri 4.0," 2017. doi: https://id.ec.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: 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: 10.6007/ijarbss/v8-i9/4593.
- [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: 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," in *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 2019, vol. Part F148155, pp. 244–253. doi: 10.1145/3326365.3326398.
- [7] Y. Adrianova Eka 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, 2020, doi: 10.31932/jutech.v1i2.1289.
- [8] A. Saputra Tamrin, P. Rumapea, and R. Mambo, "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, Accessed: Jun. 16, 2022. [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/JAP/article/view/16283/15786>
- [9] J. Homepage, A. Roihan, P. Abas Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, pp. 75–82, 2019, doi: 10.31294/ijcit.v5i1.7951.
- [10] A. Iskandar, Muttaqin, S. V. Dewi, and Jamaludin, *Statistika Bidang Teknologi Informasi*. Kita Menulis, 2021. Accessed: Jun. 14, 2022. [Online]. Available: <https://kitamenulis.id/2021/04/23/statistika-bidang-teknologi-informasi/>
- [11] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive text mining and database fusion approach," *Environ Health Perspect*, 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. 2, pp. 17–22, 2018, doi: 10.52771/bangkitindonesia.v7i2.90.
- [13] R. E. Kreisler, M. E. Spindel, and M. Rishniw, "Determinants of Salary for Veterinarians Employed in the Field of Shelter Medicine in the United States," *Top Companion Anim Med*, vol. 40, Aug. 2020, doi: 10.1016/j.tcam.2020.100428.
- [14] N. A. Elshahry, A. Alathamneh, I. Mahmoud, and F. Hammad, "Association of salary and intention to stay with the job satisfaction of the dietitians in Jordan: A cross-sectional study," *Health Policy Open*, vol. 3, Dec. 2022, doi: 10.1016/j.hpopen.2021.100058.
- [15] D. Webbink and J. Hartog, "Can students predict starting salaries? Yes!," *Econ Educ Rev*, vol. 23, no. 2, pp. 103–113, 2004, doi: 10.1016/S0272-7757(03)00080-3.
- [16] S. Gosh, K. Rascati, A. Shah, and P. Peebles, "PHP88 - Predictors of Annual Salary for Health Economics, Outcomes Research, and Market Access Professionals," *Value in Health*, vol. 21, p. S101, 2018, doi: <https://doi.org/10.1016/j.jval.2018.04.682>.
- [17] R. Marrero-Rodríguez, S. Morini-Marrero, and J. M. Ramos-Henriquez, "Tourism jobs in demand: Where the best contracts and high salaries go at online offers," *Tour Manag Perspect*, vol. 35, Jul. 2020, doi: 10.1016/j.tmp.2020.100721.
- [18] M. L. Blackburn, "Are U.S. teacher salaries competitive? Accounting for geography and the retransformation bias in logarithmic regressions," *Econ Educ Rev*, vol. 84, Oct. 2021, doi: 10.1016/j.econedurev.2021.102169.

- [19] G. A. Gilpin, "Teacher salaries and teacher aptitude: An analysis using quantile regressions," *Econ Educ Rev*, vol. 31, no. 3, pp. 15–29, Jun. 2012, doi: 10.1016/j.econeducrev.2012.01.003.
- [20] K. K. Rekayasa, M. A. Saputra, N. Prasetyo, I. Zulfikar, T. Rijanandi, and F. Dharma Adhinata, "Pengalaman Bekerja Menggunakan Metode Regresi Linear," *Journal of Dinda*, vol. 2, no. 2, pp. 58–63, 2022, [Online]. Available: <http://journal.itelkom-pwt.ac.id/index.php/dinda>
- [21] Munti and Y. S. Novi, "Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-Means Clustering," *Jurnal Inovasi Teknik Informatika*, vol. 1, no. 1, pp. 1–11, 2018.
- [22] M. Sanchez-Gomez, E. Bresó, and G. Giorgi, "Could emotional intelligence ability predict salary? A cross-sectional study in a multioccupational sample," *Int J Environ Res Public Health*, vol. 18, no. 3, pp. 1–10, Feb. 2021, doi: 10.3390/ijerph18031322.
- [23] Sayan Das, Rupashri Barik, and Ayush Mukherjee, "Salary Prediction Using Regression Technique," *International Conference On Industry Interactive Innovations In Science, Engineering And Technology*, 2020, doi: <https://dx.doi.org/10.2139/ssrn.3526707>.
- [24] U. Bansal, A. Narang, A. Sachdeva, I. Kashyap, and S. P. Panda, "Empirical analysis of regression techniques by house price and salary prediction," in *IOP Conference Series: Materials Science and Engineering*, Jan. 2021, vol. 1022, no. 1, doi: 10.1088/1757-899X/1022/1/012110.
- [25] S. Gupta, "A Regression Modeling Technique on Data Mining," *Int J Comput Appl*, vol. 116, no. 9, pp. 975–8887, 2015, [Online]. Available: <http://www.nag.co.uk/stats/GDGE>
- [26] J. Prakash Sharma and N. Bajpai, "Salary Satisfaction as an Antecedent of Job Satisfaction: Development of a Regression Model to Determine the Linearity between Salary Satisfaction and Job Satisfaction in a Public and a Private Organization," *European Journal of Social Sciences*, vol. 18, no. 3, 2011.
- [27] Y. Gormez, H. Arslan, S. Sari, and M. Danis, "SALDA-ML: Machine Learning Based System Design to Predict Salary Increase," *Advances in Artificial Intelligence Research*, vol. 2, no. 1, pp. 15–19, Jan. 2022, doi: 10.54569/air.1029836.
- [28] K. Pomthep and S. Pokpong, "Implement Of Salary Prediction System To Improve Student Motivation Using Data Mining Technique," *International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, 2016.
- [29] I. Martín, A. Mariello, R. Battiti, and J. Alberto Hernández, "Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study," *International Journal of Computational Intelligence Systems*, vol. 11, pp. 1192–1209, 2018, doi: <http://dx.doi.org/10.2991/ijcis.11.1.90>.
- [30] R. Voleti and B. Jana, "Predictive Analysis of HR Salary using Machine Learning Techniques," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 1, 2022, [Online]. Available: www.ijert.org
- [31] D. Sananda, H. Airiddha, and D. Kousik, "Design of a novel Prediction Engine for predicting suitable salary for a job," in *Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2018, pp. 275–279. doi: <https://doi.org/10.1109/ICRCICN.2018.8718711>.
- [32] D. A. Gomez-Cravioto, R. E. Diaz-Ramos, N. Hernandez-Gress, J. L. Preciado, and H. G. Ceballos, "Supervised machine learning predictive analytics for alumni income," *J Big Data*, vol. 9, no. 1, pp. 1–31, Dec. 2022, doi: 10.1186/s40537-022-00559-6.
- [33] M. R. Jennings *et al.*, "Code-free cloud computing service to facilitate rapid biomedical digital signal processing and algorithm development," *Comput Methods Programs Biomed*, vol. 211, Nov. 2021, doi: 10.1016/j.cmpb.2021.106398.
- [34] N. Ahmed *et al.*, "Machine learning based diabetes prediction and development of smart web application," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229–241, Jun. 2021, doi: 10.1016/j.ijcce.2021.12.001.
- [35] T. Kumaraguru, P. Abirami, K. M. Darshan, S. P. Angeline Kirubha, S. Latha, and P. Muthu, "Smart access development for classifying lung disease with chest x-ray images using deep learning," in *Materials Today: Proceedings*, 2021, vol. 47, pp. 76–79. doi: 10.1016/j.matpr.2021.03.650.
- [36] A. L. Dita and A. Sarini, "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.
- [37] E. Khumaedi, J. I. Manajemen, and D. Bisnis, "Pengaruh Disiplin Dan Motivasi Kerja Terhadap Kinerja Pegawai Pada Dinas Sentra Operasi Terminal Pt. Angkasa Pura II," 2016.
- [38] N. Sitti, K. Sekolah, T. Ilmu, and E. Gempol, "Analisis Ekuitas Merek Produk Notebook Asus Terhadap Keputusan Pembelian Konsumen Pada Distributor Diva Jaya Cabang Sidoarjo," *Jurnal Akuntansi Dan Manajemen*, vol. 3, no. 2, pp. 73–83, 2018.
- [39] "Analisis Regresi untuk Penelitian."
- [40] R. G. Ali and J. Nugraha, "Penerapan Metode Regresi Ridge Dalam Mengatasi Masalah Multikolinearitas Pada Kasus Indeks Pembangunan Manusia Di Indonesia Tahun 2017," in *Prosiding Sendika*, 2019, vol. 5, no. 2, pp. 226–235. [Online]. Available: www.statistik.data.kemdikbud.go.id
- [41] Nurfajar, M. Syafiq, 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, pp. 1411–710, 2018, [Online]. Available: <http://ejurnal.stieipwaja.ac.id/index.php/jpw>

Syafrial Fachri Pane, kelahiran kota Medan. Memperoleh gelar Ahli Madya Teknik Informatika di Politeknik Pos Indoensia, gelar Sarjana Teknik Informatika di Universitas Pasundan. Kemudian memperoleh gelar Magister Teknik Informatika di Telkom University. Bidang yang ditekuni adalah ilmu data. Saat ini sebagai pengajar di Universitas Logistik Bisnis Internasional.

Amri Yanuar, kelahiran kota Bandung. Memperoleh gelar Sarjana Teknik Industri di Universitas Pasundan dan gelar Magister Management of Technology di Universiti Teknologi Malaysia. Bidang yang ditekuni adalah bidang logistic. Saat ini sebagai pengajad di Universitas Logistik dan Bisnis Internasional.

Bachtiar Ramadhan, sebagai mahasiswa Program Studi Teknik Informatika Universitas Logistik dan Bisnis Internasional. Kini sedang menekuni bidang ilmu data.

Nur Tri Ramadhanti Adiningrum, sebagai mahasiswa Program Studi Teknik Informatika Universitas Logistik dan Bisnis Internasional. Kini sedang menekuni bidang ilmu data.