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.fach](mailto:1syafrial.fachri@poltekpos.ac.id)[ri@poltekpos.ac.id](mailto:ri@poltekpos.ac.id) [2amriyanuar@poltekpos.ac.id](mailto:2amriyanuar@poltekpos.ac.id) [31204077\_bachtiar@students.poltekpos.ac.id](mailto:31204077_bachtiar@students.poltekpos.ac.id) [41204061\_nur@students.poltekpos.ac.id](mailto: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 mendongkrak 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 data yang digunakan adalah analisis 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***

1. 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 *internet of thing* sebagai pergerakan dan konektivitas antara manusia dengan mesin[1]. Sehingga, diperoleh penggabungan antara teknologi digital serta internet dengan industri konvensional, yang memiliki tujuan untuk meningkatkan produktivitas, efisiensi 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 menggantikanya 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 astu aspek yang berdampak besar terhadap keberhasilan dan kemajuan sebuah 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 memliki 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.

1. TINJAUAN PUSTAKA
2. *Machine Learning*

*Machine* *learning* diartikan sebagai suatu aplikasi komputer serta algoritma matematika yang diangkat dengan cara model pembelajaran yang bersumber dari data dan dapat mengeluarkan suatu prediksi untuk masa yang akan datang[9].

1. *Regresi Linier Berganda*

Regresi berganda adalah perpanjangan dari regresi linier sederhana[10]. Model regresi linier berganda dapat digunakan untuk memprediksi 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 dapat dibuktikan dalam persamaan regresi linier berganda yang dikutip pada persamaan 1 serta persamaan 2 sebagai berikut :

 (1)

 (2)

Keterangan :

= nilai-nilai hasil pengamatan

= nilai regresi

i = 1,2,3, k

Pada persamaan di atas, terdapat satu variabel dependen yaitu Y’ dan terdapat n variabel independent yaitu X1, X2,… Xn [10].

*C. Scikit-Learn*

*Scikit-learn* adalah modul pada bahasa pemrograman python yang menyediakan banyak jenis algoritma *machine* *learning*. Bentuk yang terdapat pada *library* pythonadalah jenis bentuk yang tersedia di dalam *scikit-learn*. *Scikit-learn* menggunakan bentuk *task-oriented interface* yang konsisten sehingga dapat memudahkan perbandingan antar 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 membuat suatu aplikasi web yang aman, kaya fitur dan dinamis[12]. Django dikembangkan oleh Django Software Foundation. *Framework* ini terus-menerus mendapatkan perbaikan yang membuat web *framework* Django menjadi pilihan utama bagi 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 | *Multiple Regression* |
| Penentu kepuasan kerja ahli gizi [14] | Data kuisioner ahli gizi | *Multivariate Regression* |
| Prediksi gaji siswa [15] | Data siswa Pendidikan tinggi | Regresi OLS |
| Prediksi gaji tahunan[16] | Data global HealthEconomics.com | *Multiple Regression* |
| Analisi kontrak dan gaji tertinggi [17] | Data ketenagakerja-an | Regresi logistik |
| Analisi tingkat kompetitif gaji guru [18] | Data sensus Amerika Serikat | *Logarithmic Regressions* |
| Analisis gaji dan bakat guru[19] | Data Schools and Staffing Survey (SASS) | *Quantile Regression* |
| 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 penelitan 785 subjek | Regresi Multivariat |
| Prediksi gaji setelah tahun tertentu[23]. | Data karyawan dari perusahaan. | Regresi Linear, Regresi Polinomial |
| Analisis empiris harga rumah dan prediksi gaji[24] | Dataset gaji pegawai dan harga rumah. | *Multiple Linear Regression* |
| Prediksi gaji dalam penerapan model regresi [25] | Data gaji dan lama pengalaman bekerja | *Simple Linear Regression* |
| Analisis dan prediksi kepuasan gaji [26] | Data karyawan staf | *Simple Regression* |
| Perancangan sistem prediksi kenaikan gaji [27] | Data Enterprise Respurce Planning (ERP) | *Linear regression, artificial neural networks, random forest regression* |
| Sistem prediksi gaji untuk meningkat-kan motivasi [28]. | Data mahasiswa yang lulus dengan gajinya. | *K-NN, Naïve Bayes, Decision trees, Multilayer perception, SVM* |
| Prediksi gaji di pasar kerja TI [29]. | Data *e-Recruitment* khusus untuk pekerjaan TI | *Logistic regression, KNN, Multi-layer perceptrons, SVM, Random forest, Vote, Vote3* |
| Analisis prediktif gaji SDM[30] | Data ketenagakerjaan | *Logistic regression SVM* |
| Mesin prediksi untuk memprediksi gaji[31] | Data kepegawaian | *Decision tree classifier, Random forest classifier* |
| Analisis prediktif untuk pendapatan alumni[32] | Data survei studi alumni | *Quantile Regression, Quantile Random Forest, Quantile Gradient Boosting, Linear Regression, Random Forest Classifier, Gradient Boosting Classifier* |
| Komputasi cloud fasilitas sinyal digital biomedis. [33] | Data biomedis | *Framework Django* |
| Aplikasi web prediksi diabetes [34] | Data klinis penyakit diabetes | *Framework Django* |
| Deployment klasifikasi penyakit paru-paru [35] | Dataset *x-ray* tubuh bagian atas untuk Covid-19, *Pneumonia*, dan *Normal*. | *Framework Django* |

1. METODE PENELITIAN

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 mendeskprisikan suatu keadaan dengan apa adanya dan menginterpretasikan objek sesuai dengan peristiwa, ataupun segala sesuatu yang berhubungan dengan variabel-variabel yang dapat dijelaskan baik dengan angka-angka maupun kata-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* berikut :https[://www.](http://www.kaggle.com/colearninglounge/employee-)kag[gle.com/colearninglounge/employee-](http://www.kaggle.com/colearninglounge/employee-) attrition

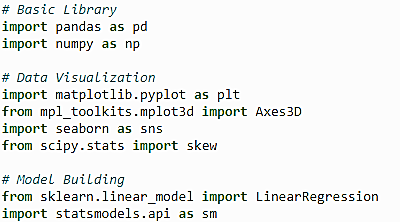
1. IMPLEMENTASI

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 IntelTM Core i3-4030U, RAM 4 GB, 64-bit OS. Software : Ms. Office, Windows 10 Pro.

1. *Implementasi Model Machine Learning*
   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:

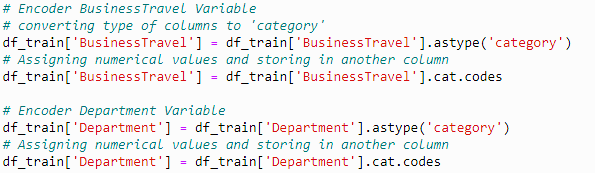


Gambar 1 Import Library



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.



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.



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).



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.



Gambar 6 Skrip Drop Atribut

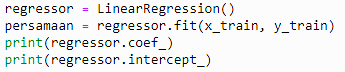
* 1. *Proses Data Mining & Pengetahuan*

Pada tahap proses *Data Mining* & Pengetahuan, proses yang dilakukan adalah pemilihan jenis metode yang sesuai dengan karakter data atau dikenal dengan istilah *Modelling*. Pada model ini digunakan proses *Data Mining Prediction*. Proses *Data Mining* & Pengetahun 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* meggunakan *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.



Gambar 7 Skrip Penentuan Variabel Dependen dan Independen



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 :



(3)

MonthlyIncome (Variabel Dependen) Age (Variabel Independen-1) JobLevel (Variabel Independen-2)



TotalWorkingYears (Variabel Independen-3)

= YearsAtCompany (Variabel Independen-4)

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

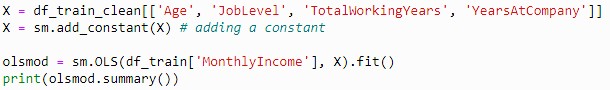


(4)

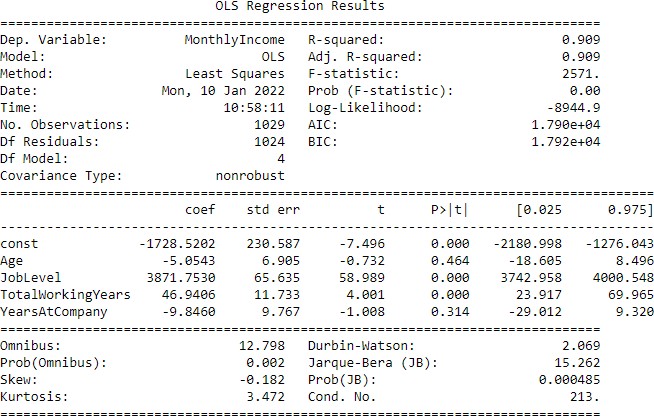
* 1. *Evaluasi Data*

1. Validasi Model

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



Gambar 9. Skrip Validasi OLS



Gambar 10. Hasil OLS Regresi

1. Uji F (ANOVA)

Uji kelayakan model atau dikenal dengan uji F, yaitu pengujian yang dilakukan untuk mengetahui apakah variabel independen yang digunakan dalam persamaan tersebut secara bersama-sama berpengaruh signifikan pada nilai variabel dependen[36].

F-test atau ANOVA (*Analysis of Variance*) dalam regresi linier berganda dapat dimanfaatkan untuk mengetahui apakah model yang dibuat dapat bekerja lebih baik daripada model yang lebih sederhana.



Gambar 11. Skrip Uji F (ANOVA)



Gambar 12. Output Uji F (ANOVA)

Hipotesa yang didapat dari tabel uji F adalah:

H0 = Variabel independen secara simultan bukan merupakan penjelas yang signifikan terhadap variabel dependen (model tidak cocok).

H1 = Variabel independen secara simultan merupakan penjelas yang siginifikan terhadap variabel dependen (model cocok).

Berdasarkan gambar 12, dapat diketahui bahwa Fs > P- value, yang artinya hipotesa yang dapat diambil adalah terima H1 dan tolak H0. Dapat dikatakan, variabel independen (Age, JobLevel, TotalWorkingYears, YearsAtCompany) dan MonthlyIncome berpengaruh signifikan terhadap permintaan. Pada taraf signifikansi 5% (0,05), H0 ditolak karena nilai probabilitasnya yaitu 0,00 yang berarti dibawah dari 5%. Maka dapat disimpulkan, model yang dipakai cocok.

1. Uji-t

Uji parsial (koefisien regresi) atau dikenal dengan Uji-t bertujuan untuk menguji taraf signifikan konstanta dajugan variabel independen yang terdapat dalam persamaan regresi secara individu[36].

Hipotesa yang ada sebagai berikut :

H0 = Variabel independen tidak berpengaruh signifikan. H1 = Variabel independen berpengaruh signifikan.

= 0,05 (Taraf Signikansi/Threshold)

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

* + Nilai variabel  (Age) berada di atas taraf signifikansi/terima H1.
  + Nilai variabel  (JobLevel) berada di bawah taraf signifikansi/terima H0.
  + Nilai variabel  (TotalWorkingYears) di bawah taraf signifikansi/terima H0.
  + Nilai variabel  (YearsAtCompany) di atas taraf signifikansi/terima H1.

Dapat diambil kesimpulan bahwa variabel independen JobLevel dan TotalWorkingYears adalah variabel yang tidak mempengaruhi variabel dependen yaitu MonthlyIncome. Sedangkan variabel independent Age dan YearsAtCompany adalah variabel yang mempengaruhi variabel dependen yaitu MonthlyIncome.

1. R-Square

R square dapat diartikan sebagai koefisien determinasi. R square menunjukan suatu persentase pengaruh antara variabel X1 dengan X2 terhadap variabel Y[37].

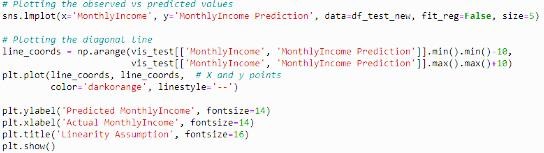
Berdasarkan gambar 10, nilai koefisien determinasi (R- Square) adalah 0,909 atau 90,9%. Maka, MonthlyIncome dipengaruhi oleh faktor Age dan YearsAtCompany sebesar 0,909 atau 90,9%. Nilai sisa dari koefisien determinasi adalah 0,091 atau 9,1% dipengaruhi oleh faktor lain yang tidak diketahui.

1. Pengujian Asumsi

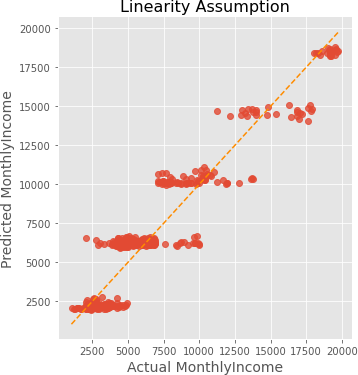
Untuk memvalidasi model *machine learning*, dilakukan analisis residual. Berikut adalah berbagai pengujian atau asumsi yang akan lakukan untuk mengetahui validitas model :

* Linearitas

Pengujian linearitas merupakan pengujian yang dilakukan untuk mengetahui apakah model yang dibuktikan merupakan model linear atau bukan [38]. Uji linearitas dilakukan menggunakan regresi kurva, yaitu sebuah gambaran hubungan linier antara variabel X dengan variabel Y [38].



Gambar 13. Skrip Linearitas dan Grafik Asumsi Linier

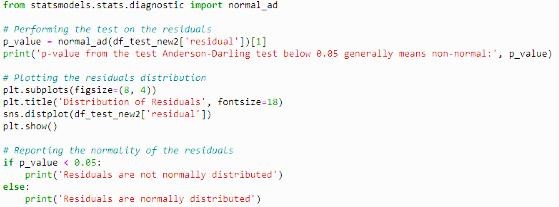


Gambar 14. Grafik Asumsi Linier

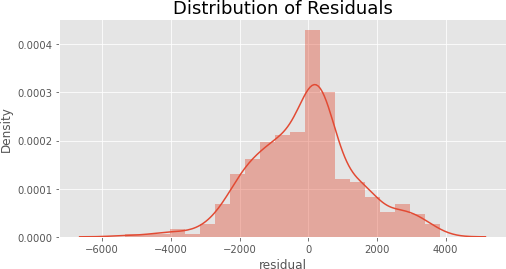
Pada Gambar 14, plot sebar memperlihatkan sisa yang tersebar merata pada sekitar garis diagonal, sehingga dapat diasumsikan bahwa teradapat hubungan linier antara variabel independen dan dependen.

* Normalitas

Uji normalitas yaitu pengujian suatu model regresi berupa apakah variabel dependen, variabel independen ataupun keduanya memiliki distribusi yang normal ataukah tidak[38]. Model regresi dikatakan baik jika distribusi data adalah normal atau mendekati normal[38].



Gambar 15. Skrip Uji Normalitas



Gambar 16. Diagram Distribusi Residual

Berdasarkan asumsi di atas, dapat diketahui hipotesa sebagai berikut :

H0 = Residual terdistribusi normal.

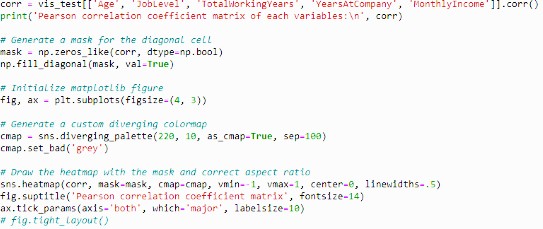
H1 = Residual terdistribusi secara tidak normal.

Dari hasli 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, yang berarti H0 ditolak H1 diterima. Sehingga asumsi normalitas terpenuhi.

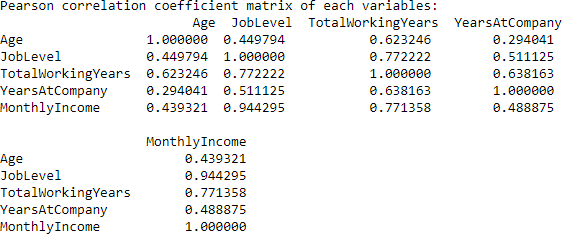
* Multikolinieritas

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

Uji multikolinieritas digunakan untuk menampilkan korelasi antar variabel prediktor. Apabila terjadi multikolinearitas pada model regresi, maka hal tersebut menyebabkan parameter regresi yang dihasilkan akan memiliki nilai error yang sangat besar. Kriteria yang digunakan utnuk mengetahui adanya multikolinearitas antara variabel prediktor adalah dengan menggunakan nilai *Variance Inflation Factors* (VIF). Apabila nilai VIF lebih besar dari 10 mengindikasikan bahwa terdapat masalah multikolinearitas. Nilai VIF dapat diperoleh dengan cara meregresikan variabel independent[40].

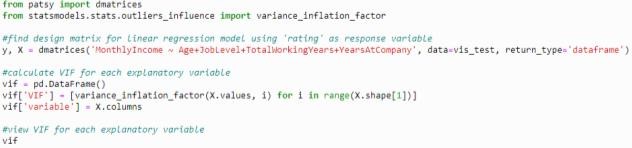


Gambar 17. Skrip Uji Multikolinieritas

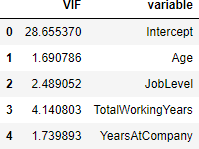


Gambar 18. Koefisen Korelasi Pearson

Dari hasil asumsi di atas, dapat dikatakan bahwa prediktor yang digunakan dalam regresi berkorelasi antara satu dengan yang lainnya.



Gambar 19. Skrip Multikolinearitas



Gambar 20. Tabel VIF

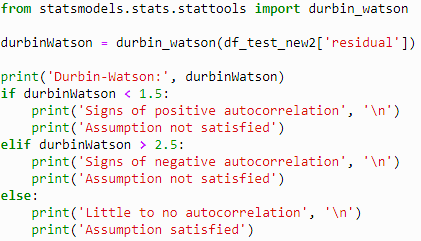
Berdasarkan gambar 20, dapat dilihat nilai variabel Age, JobLevel, TotalWorkingYears, YearsAtCompany memiliki nilia kurang dari 10 sehingga dengan menggunakan tingkat taraf signifikansi sebesar 0,05 dapat disimpulkan bahwa pada data tersebut tidak terdapat multikolinearitas pada variabel-variabel prediktor.

* Autokorelasi

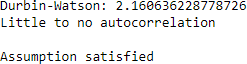
Autokorelasi merupakan pengujian yang ditujukan untuk menguji dalam sebuah model regresi linier berganda apakah terdapat korelasi antara kesalahan pengganggu pada periode t dengan kesalahan di periode t1 (sebelumnya). Jika terdapat korelasi, maka hal itu dikenal sebagai autokorelasi. Model regresi yang baik adalah model yang terbebas dari autokorelasi[38].

Pada langkah ini, dilakukan perhitungan skor Durbin- Watson menggunakan *durbin\_watson()* fungsi dari *statsmodel* yang dibuat, kemudian menilainya dengan kondisi sebagai berikut :

1. Apabila skor Durbin-Watson kurang dari 1,5 maka terdapat autokorelasi positif dan asumsi tidak terpenuhi.
2. Apabila skor Durbin-Watson antara 1,5 – 2,5 maka tidak ada autokorelasi dan asumsi puas.
3. Apabila skor Durbin-Watson lebih dari 2,5 maka terdapat autokorelasi negatif dan asumsi tidak puas



Gambar 21. Skrip Uji Autokorelasi

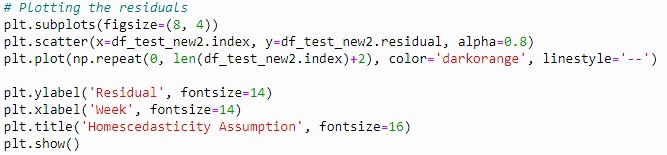


Gambar 22. Output Uji Autokorelasi

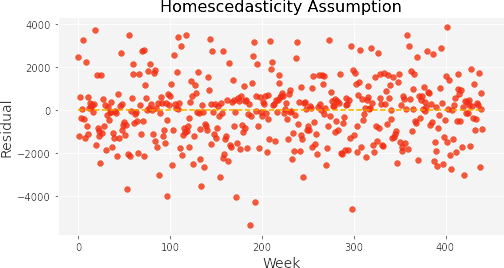
Dari hasil output pada gambar 20, dapat diasumsikan bahwa terdapat sedikir atau tidak ada autokorelasi, yang berarti asumsi puas.

* + Homoskedatisitas

Homoskedastisitas merupakan pengujian yang ditujukan untuk menguji apakah didalam model regresi linier berganda terdapat gejala heteroskedasitas atau tidak dengan cara melihat ada atau tidaknya pola tertentu pada grafik *scatterplots*. Model regresi dapat dikatakan model baik jika model tersebut mengalami homoskedastisitas atau tidak terjadinya heteroskesdatisitas[41].



Gambar 23. Skrip Plot Penyebaran Residual



Gambar 24. Plot Penyebaran Residual

Dari grafik scatterplot (gambar 22), 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.

1. *Implementasi Antarmuka Aplikasi*
   1. *Halaman Prediksi Gaji*

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



Gambar 25. Halaman Prediksi Gaji

* 1. *Halaman hasil prediksi gaji*

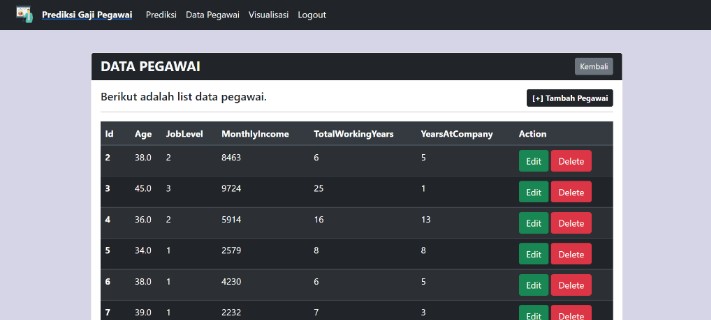
Tampilan untuk aplikasi menampilkan hasil prediksi gaji pegawai terlihat pada Gambar 26.



Gambar 26. Halaman Hasil Prediksi gaji

* 1. *Halaman data pegawai*

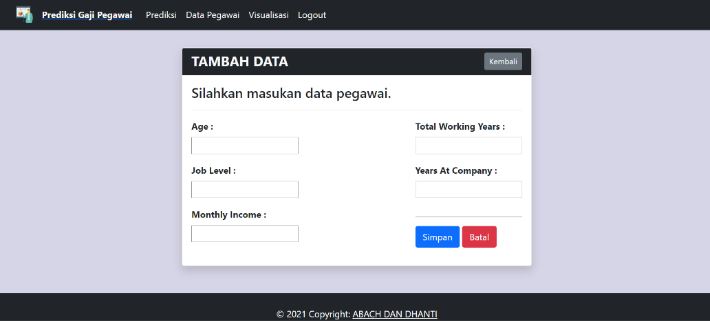
Tampilan aplikasi untuk menampilkan data pegawai yang terdiri dari atribut Id, Age, JobLevel, MonthlyIncome, dan Action. Halaman data pegawai terlihat pada Gambar 27.



Gambar 27. Halaman Data Pegawai

* 1. *Halaman tambah data pegawai*

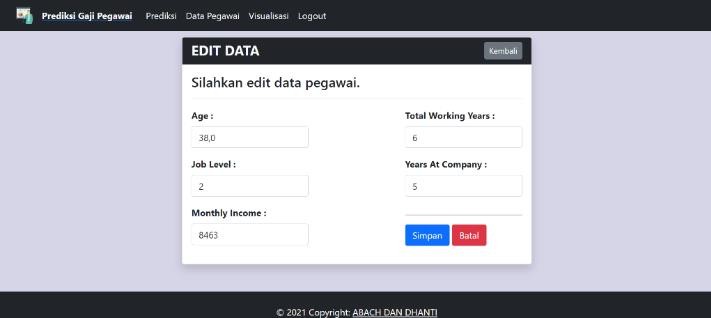
Tampilan aplikasi untuk menginputkan data pegawai yang baru. Form input terdiri dari Age, Job Level, Monthly Income, dan Total Working Years, dan Years At Company. Halaman tambah data pegawai terlihat pada Gambar 28.



Gambar 28. Halaman tambah data Pegawai

* 1. *Halaman edit data pegawai*

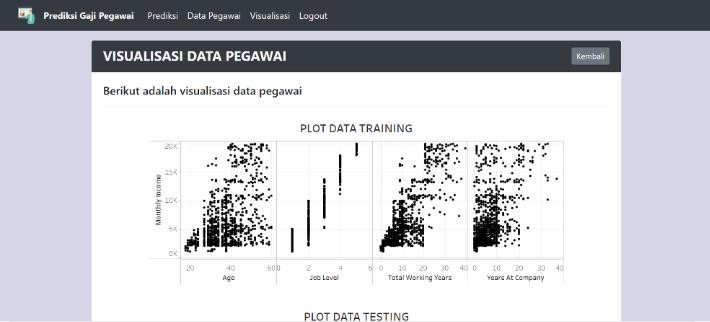
Tampilan untuk aplikasi mengedit data pegawai yang baru. Form input untuk mengedit data terdiri dari Age, Job Level, Monthly Income, dan Total Working Years, dan Years At Company. Halaman edit data pegawai terlihat pada Gambar 29.



Gambar 29. Halaman Edit Data Pegawai

* 1. *Halaman visualisasi*

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



Gambar 30. Halaman Visualisasi

1. KESIMPULAN DAN SARAN
2. *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 seluruh pengujian dalam langkah validasi model, sehingga dapat disimpulkan bahwa model yang dibuat 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

1. *Saran*

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

* Pembuatan model prediksi yang digunakan dapat lebih beragam untuk membandingkan performa antara model satu 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 PUSTAKA

[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://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: 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. Elsahoryi, 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. Peeples, “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.econedurev.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.ittelkom-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. Breso, 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 Predicition 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 In-crease,” *Advances in Artificial Intelligence Research*, vol. 2, no. 1, pp. 15–19, Jan. 2022, doi: 10.54569/aair.1029836.

[28] K. Pornthep 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.stieipwija.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 pengajat 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.