**LAPORAN AKHIR**

**PENELITIAN INTERNAL**

**PEMODELAN BERBASIS DATA UNTUK MEMPREDIKSI GAJI BERDASARKAN FAKTOR-FAKTOR SPESIFIK DENGAN PENDEKATAN *MACHINE LEARNING***



**Oleh:**

|  |  |
| --- | --- |
| **Syafrial Fachri Pane, ST., M.T.I., EBDP** | **0416048803** |
| **Amri Yanuar, ST., M.OT** | **0412018603** |

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA**

**POLITEKNIK POS INDONESIA**

**TAHUN 2022**

# HALAMAN PENGESAHAN

# HALAMAN KETERLIBATAN MAHASISWA DALAM PENELITIAN

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Nama Mahasiswa** | **NPM** | **Keterlibatan** | **Paraf** |
| 1. | Bachtiar Ramadhan | 1204077 | Pemograman |  |
| 2. | Nur Tri Ramadhanti Adiningrum | 1204061 | Analisi dan pengumpulan data |  |
| 3. | M. Rizky | 1194021 | Data Engineer |  |

Bandung, 14 Maret 2022

Ketua Peneliti

**Syafrial Fachri Pane, ST., M.T.I., EBDP**

**NIK. 117.88.233**

# ABSTRAK

Perkembangan ilmu pengetahuan dan teknologi pada Revolusi Industri 4.0 semakin berkembang pesat. Perubahan karakteristik pekerjaan adalah salah satu dampak tersendiri dari datangnya revolusi industri 4.0. Tentunya perusahaan perlu memiliki keunggulan manajemen yang efektif dalam menghadapi hal tersebut. Dengan demikian salah satu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja karyawannya. Oleh karena itu, penentuan gaji yang tepat oleh perusahaan adalah faktor internal terhadap kemajuan perusahaan. Sangat disayangkan, perkembangan perusahaan saat ini belum memliki suatu media keputusan untuk melakukan prediksi gaji karyawan berdasarkan kualitas data. Namun, untuk membuat keputusan bagaimana cara menentukan gaji karyawan dengan optimal perlu mempertimbangkan faktor-faktor lainya karena faktor tersebut merupakan bobot penilaian untuk mengukur kelayakan karyawan mendapatkan gaji.

Oleh karena itu, tujuan utama penelitian ini adalah membuat model prediksi gaji karyawan berdasarkan data. Karakteristik data yang digunakan terdiri dari umur, job level, total lama bekerja, masa bakti yang disebut faktor-faktor spesifik. Selanjutnya faktor-faktor spesifik akan diuji validitas dan korelasinya menggunakan pendekatan *machine learning* dengan metode *regression*. 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. Rencana Keluaran penelitian ini adalah jurnal nasional terakreditasi SINTA 3, HAKI dan Buku ISBN yang diimplementasikan untuk referensi praktikum pada matakuliah Database di Prodi D4 Teknik Informatika Politeknik Pos Indonesia.

Kata Kunci : Pemodelan, Prediksi, Gaji, Regresi, faktor-faktor spesifik, *Machine Learning*

# PRAKATA

# DAFTAR ISI

[HALAMAN PENGESAHAN i](#_Toc108877830)

[HALAMAN KETERLIBATAN MAHASISWA DALAM PENELITIAN ii](#_Toc108877831)

[ABSTRAK iii](#_Toc108877832)

[PRAKATA iv](#_Toc108877833)

[DAFTAR ISI v](#_Toc108877834)

[DAFTAR TABEL vii](#_Toc108877835)

[DAFTAR GAMBAR viii](#_Toc108877836)

[DAFTAR LAMPIRAN ix](#_Toc108877837)

[BAB I PENDAHULUAN 1](#_Toc108877838)

[1.1 Latar Belakang Penelitian 1](#_Toc108877839)

[1.2 Identifikasi Masalah 2](#_Toc108877840)

[1.3 Rancangan Hipotesis Penelitian 3](#_Toc108877841)

[1.4 Ruang Lingkup Penelitian 3](#_Toc108877842)

[1.5 Sistematika Penulisan 3](#_Toc108877843)

[BAB II TINJAUAN PUSTAKA 5](#_Toc108877844)

[2.1 State of The Art (SoTA) 5](#_Toc108877845)

[2.2 Tinjauan Pustaka 7](#_Toc108877846)

[2.3 Machine Learning 15](#_Toc108877859)

[2.3.1 Multivariate Linear Regression 16](#_Toc108877860)

[2.3.2 Bayesian Multivariate Linear Regression 18](#_Toc108877861)

[2.3.3 Multivariate functional linear regression (mFLR) 20](#_Toc108877862)

[2.3.4 Multivariate Regression Modelling (MRM) 20](#_Toc108877863)

[2.3.5 Multivariate Adaptive Regression Splines 20](#_Toc108877864)

[2.4 Deep Learning 20](#_Toc108877865)

[2.4.1 Convolutional Neural Network 21](#_Toc108877866)

[2.5 Statistika 21](#_Toc108877867)

[2.6 Metode Evaluasi Model Machine Learning 21](#_Toc108877868)

[2.6.1 Root Mean Square Error (RMSE) 21](#_Toc108877869)

[2.6.2 Mean Square Error (MSE) 22](#_Toc108877870)

[2.6.3 Mean Absolute Error (MAE) 22](#_Toc108877871)

[2.6.4 Mean Absolute Percentage Error (MAPE) 22](#_Toc108877872)

[2.7 Taksnomi Studi Literatur 23](#_Toc108877873)

[BAB III 26](#_Toc108877874)

[TUJUAN DAN MANFAAT 26](#_Toc108877875)

[3.1 Tujuan dan Manfaat Penelitian 26](#_Toc108877876)

[3.1.1 Tujuan Penelitian 26](#_Toc108877877)

[3.1.2 Manfaat Penelitian 26](#_Toc108877878)

[BAB IV METODE PENELITIAN 28](#_Toc108877879)

[4.1 Diagram Alur Metodologi Penelitian 28](#_Toc108877880)

[4.1.1 Diagram Alur Utama 28](#_Toc108877881)

[BAB V 30](#_Toc108877882)

[HASIL DAN LUARAN YANG DICAPAI 30](#_Toc108877883)

[5.1 Kegiatan dan Hasil Pelaksanaan 30](#_Toc108877884)

[5.2 Luaran yang Dicapai 32](#_Toc108877885)

[5.3 Hasil Penelitian 34](#_Toc108877886)

[DAFTAR PUSTAKA 48](#_Toc108877887)

[LAMPIRAN-LAMPIRAN 52](#_Toc108877888)

# DAFTAR TABEL

[Tabel 2.1 Penelitian Terkait **Error! Bookmark not defined.**](#_Toc103865754)

[Tabel 4.1 Penjelasan Diagram Alur Metodologi Penelitian 29](#_Toc103867102)

[Tabel 5.1 Luaran dan Target Capaian 33](#_Toc103867130)

# DAFTAR GAMBAR

[Gambar 4.1 Gambar Diagram Alur Metodologi Penelitian 28](#_Toc103865095)

# DAFTAR LAMPIRAN

[Lampiran 1 Susunan Organisasi Tim Peneliti dan Pembagian Tugas 52](#_Toc103866735)

[Lampiran 2 Biodata Ketua dan Anggota Pengusul 53](#_Toc103866736)

[Lampiran 3 Surat Pernyataan Ketua Penelitian 63](#_Toc103866737)

[Lampiran 4 Penggunaan Anggaran 64](#_Toc103866738)

[Lampiran 5 Bukti Penerimaan Artikel Ilmiah (LOA) atau URL dan Screenshoot Halaman Jurnal yang Sudah Dipublikasi 65](#_Toc103866739)

[Lampiran 6 Format Catatan Harian (Logbook) 66](#_Toc103866740)

[Lampiran 7 Poster 67](#_Toc103866741)

# BAB I PENDAHULUAN

## Latar Belakang Penelitian

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 menggantikanya 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 memliki suatu media keputusan untuk melakukan dua prediksi gaji karyawan berdasarkan kualitas data. Berdasarkan hal tersebut, pentingnya studi ini, tidak hanya digunakan untuk penetapan gaji saja, tetapi juga menjadi studi terkait pemodelan prediksi penggajian secara umum dimasa yang akan datang. Urgensi pada penelitian ini adalah pemodelan yang dibuat dapat digunakan sebagai tools untuk menentukan gaji karyawan.

Karakteristik data yang digunakan terdiri dari umur, job level, total lama bekerja, masa bakti yang disebut 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. Target luaran yang ingin dicapai pada penelitian ini adalah jurnal nasional terakreditasi SINTA 3, HAKI dan Buku ISBN yang diimplementasikan untuk referensi pembelajaran dan pratikum pada matakuliah Database di Program Studi Diploma 4 Teknik Informatika Politeknik Pos Indonesia.

## Identifikasi Masalah

Adapun identifikasi masalah pada penelitian ini sebagai berikut.

1. Bagaimana menganalisa karakterisik dan koelasi data terkait gaji karyawan?
2. Bagaimana membuat model prediksi gaji karyawan berdasarkan kualitas data dengan mempertimbangkan faktor-faktor spesifik?
3. Bagaimana merancang *framework* yang dinamis untuk menampilkan hasil prediksi gaji?

## Rancangan Hipotesis Penelitian

Adapun rancangan hipotesis pada penelitian ini sebagai berikut.

**Hipotesis penelitian/kerja:**

H0 : Tidak ada korelasi faktor-faktor spesifik untuk menentukan gaji karyawan.

H1 : faktor-faktor spesifik mempunyai korelasi positif dan berpotensi menjadi faktor-faktor utama menentukan gaji karyawan.

## Ruang Lingkup Penelitian

Adapun ruang lingkup pada penelitian ini sebagai berikut.

1. Dikarenakan kondisi pandemi, data yang digunakan dalam penelitian ini bersumber dari data kepegawaian disuatu perusahaan.
2. Periode waktu 1 tahun.
3. Bahasa pemograman menggunakan *python*.
4. *Software* yang digunakan *Jupyier* atau dan *google colab*.

## Sistematika Penulisan

Berdasarkan latar belakang dan perumusan masalah diatas, maka penyusunan laporan ini dibuat dalam suatu sistematika yang terdiri dalam lima BAB, yaitu:

**BAB I PENDAHULUAN**

Bab ini berisi penjelasan terkait dengan *State-of-The-Art* yang menjelaskan mengenai pemaparan teori umum dengan topik yang dibahas secara global dan mengaitkan dengan referensi yang ada. Identifikasi masalah menjelaskan mengenai masalah dalam pemodelan berbasis data untuk memprediksi gaji berdasarkan faktor-faktor spesifik dengan pendekatan *machine learning* dan memberikan solusi atas masalah tersebut. Ruang lingkup menjelaskan mengenai batasan dalam pemodelan dan aplikasi tersebut. Serta sistematika penulisan menjelaskan tentang isi dari aplikasi tersebut.

**BAB II LANDASAN TEORI**

Bab ini berisi penjelasan mengenai konsep dasar dan pendukung dari sistem yang akan dibangun dengan menggunakan metode tertentu, antara lain *State-of-The-Art*, diagram alur metodologi penelitian, dan penelitian sebelumnya yang berhubungan dengan tema yag di ambil.

**BAB III TUJUAN DAN MANFAAT**

Bab ini berisi penjelasan mengenai solusi dari masalah yang ada dan manfaat dari penelitian yang dilakukan.

**BAB IV METODOLOGI PENELITIAN**

Bab ini berisi penjelasan diagram alur metodologi penelitian beserta tahapan –tahapan diagram alur penelitian untuk menyelesaikan penelitian yang sedang dilakukan sehingga bisa mencapai tujuan yang diharapkan.

**BAB V HASIL DAN LUARAN YANG DICAPAI**

Bab ini berisi penjelasan tentang hasil dan luaran yang dicapai dari penelitian yang dilakukan.

# BAB II TINJAUAN PUSTAKA

## State-of-The-Art

Pada Revolusi Industri 4.0, perkembangan ilmu pengetahuan semakin pesat. Pada saat ini, kehidupan sedang berada pada Revolusi Industri 4.0, dimana pada revolusi ini dapat berdampak dalam perubahan cara hidup, bekerja, dan berhubungan satu sama lain [3]. Salah satu dampak tersendiri dari datangnya Revolusi Industri 4.0 adalah perubahan karakteristik pekerjaan[4]. Karakteristik pekerjaan sebelumnya akan terdisrupsi dengan karakteristik pekerjaan yang baru[5]. Tentunya perusahaan perlu memiliki keunggulan dan manajemen yang efektif untuk saling bersaing dengan perusahaan yang lain[7]. Oleh karena itu, salah satu faktor internal yang dapat berpengaruh adalah penentuan gaji yang tepat oleh perusahaan kepada karyawannya. Sangat disayangkan, perkembangan perusahaan saat ini belum memiliki suatu media keputusan untuk melakukan prediksi gaji karyawan berdasarkan kualitas data. Banyak para peneliti yang telah berkontribusi dalam melakukan analisis untuk menghasilkan sebuah prediksi. Namun, di dalam suatu perusahaan pada umumnya sering terdapat perkembangan dan perubahan data kepegawaian, sehingga diperlukan teknik yang tepat agar dapat memodelkan kondisi untuk menghasilkan keputusan yang optimal. Pendekatan berbasis *machine learning* merupakan salah satu metode untuk membuat prediksi dan mengekstrak informasi dari data semakin diterapkan di berbagai bidang ehingga metode tersebut dapat digunakan untuk memodelkan atau memprediksi suatu hal seperti gaji[8]. Pendekatan tersebut dapat dioptimalkan dengan memperhatikan faktor-faktor spesifik, yang meliputi umur, *job level*, total lama bekerja, dan masa bakti. Semakin banyak data relevan yang dilibatkan, luaran berupa kebijakan perusahaan yang dihasilkan akan semakin komprehensif. Banyak penelitian telah dilakukan untuk menilai efek dari faktor tunggal, sementara studi yang ditujukan untuk menilai efek dari berbagai faktor jarang dilakukan[9]. Pada penelitian sebelumnya, sumber informasi data yang relevan digunakan untuk melakukan prediksi gaji dengan satu faktor yaitu pengalaman lama bekerja. Untuk melakukan prediksi gaji pada perusahaan, tentunya diperlukan faktor-faktor lain untuk menghasilkan keputusan yang tepat. Sehingga diperlukan berbagai faktor yang terlibat dalam memodelkan prediksi gaji karyawan agar hasil keputusan dari prediksi tersebut semakin relevan. Regresi linier adalah model algoritma analisis statistik yang melatih kumpulan data dengan [fungsi linier](https://www.sciencedirect.com/topics/computer-science/linear-function) untuk menganalisis dan menghitung [risiko sistemik](https://www.sciencedirect.com/topics/engineering/systemic-risk)[10]. Hasil dari model ini dapat digunakan sebagai rekomendasi bagi peneliti untuk melakukan penelitian selanjutnya. Peneliti ini menggunakan metode regresi linier multivariat untuk melakukan proses pemilihan data sesuai dengan kriteria yang dipilih dan ditinjau dengan fokus pada kebijakan dalam melakukan keputusan gaji karyawan disuatu perusahaan berdasarkan faktor-faktor spesifik. Karena penting untuk mempertimbangkan pemilihan variabel dalam [analisis multivariat](https://www.sciencedirect.com/topics/mathematics/multivariate-analysis)[11]. Selain itu, 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*.

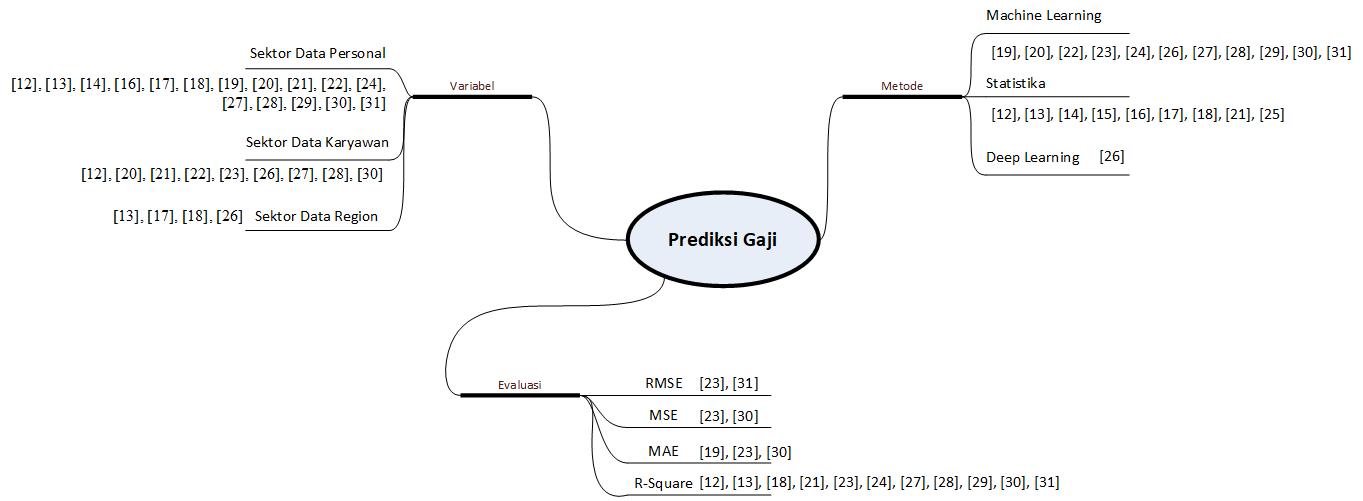
Tim peneliti telah mengkaji sejumlah referensi mengenai perkembangan teknologi pada revolusi industri 4.0 yang dapat merubah karakteristik pekerjaan, serta pemodelan untuk prediksi dengan pendekatan *machine learning*[3]-[7]. Selain itu, terkait dengan kajian literatur terkait dalam melakukan prediksi, faktor-faktor apa saja yang mempengaruhi dan model *machine learning* yang digunakan [12]- [13]. Lebih lanjut, detail dari tiap referensi lainnya ditunjukkan pada tabel.

## Tinjauan Pustaka

Tabel 2.1 Penelitian Terkait

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Area Penelitian** | **Tahun** | **Karakteristik Data** | **Metode** | **Model** | **Hasil Penelitian** | **Evaluasi** | | | |
| **RMSE** | **MSE** | **MAE** | **ACC** |
| 1. | Penentu Gaji untuk Dokter Hewan[12]. | 2020 | Data survey Dokter Hewan. | Statistika | *Multiple Regression* | Analisis dan prediksi berdasarkan data survey dokter hewan denganmodel *multiple regression* (Akurasi 42%). | - | - | - | 0,42 |
| 2. | Penentu kepuasan kerja Ahli Gizi di Yordania[14] | 2022 | Data kuesioner ahli gizi. | Statistika | *Multivariate Regression* | Analisis dan prediksi berdasarkan data kuesioner ahli gizi dengan model *multivariate regression* (Akurasi 80%). | - | - | - | 0,80 |
| 3. | Prediksi gaji siswa empat tahun kemudian [15]. | 2004 | Data siswa pendidikan tinggi tahun 1991. | Statistika | Regresi OLS | Analisis dan prediksi berdasarkan data siswa pendidikan tinggi tahun 1991 dengan model regresi OLS. | - | - | - | - |
| 4. | Prediksi utama gaji tahunan untuk ekonomi kesehatan, penelitian hasil, dan professional akses pasar[16]. | 2018 | Data pelanggan global HealthEconomics.com. | Statistika | *Multivariate Regression* | Prediksi berdasarkan data pelanggan dengan model regresi multivariat. | - | - | - | - |
| 5. | Analisis kontrak terbaik dan gaji tertinggi[17]. | 2022 | Data ketenagakerjaan | Statistika | Regresi Logistik | Analisis kontrak gaji terbaik menggunakan model regresi logistik menghasilkan gaji rata-rata kontrak standar lebih tinggi, dan pengalaman lebih utama dibandingkan Pendidikan. | - | - | - | - |
| 6. | Analisis tingkat kompetitif gaji guru[18]. | 2021 | Data sensus Amerika Serikat 2012-2016. | Statistika | *Logarithmic Regressions* | Analisis data dari Survei Komunitas Amerika menunjukkan bahwa besarnya perbedaan gaji guru meningkat dari waktu ke waktu. | - | - | - | - |
| 7. | Analisis gaji dan bakat guru[19]. | 2021 | Data *Schools and Staffing* Survey (SASS). | Statistika | *Quantile Regression* | Variasi gaji antara distrik sekolah yang berdekatan menunjukkan bahwa bakat guru berkorelasi positif dengan gaji guru (Akurasi 90%). | - | - | - | **0,90** |
| 8. | Prediksi gaji karyawan berdasarkan pengalaman bekerja[20]. | 2022 | Data survey dari *google form*. | *Machine Learning* | Regresi Linear | Analisis dan prediksi gaji karyawan berdasarkan pengalaman lama bekerja (MAE 2,051). | - | - | 2,051 | - |
| 9. | Analisis dan penerapan data mining untuk menentukan gaji karyawan[21]. | 2018 | Data rekap gaji karyawan tetap dan karyawan kontrak PT Indomex Dwijaya Lestari. | *Machine Learning* | K-Means Clustering. | Analisis dan prediksi dalam menentukan gaji karyawan tetap dan kontrak pada PT Indomex Dwijaya Lestari dengan model K-Means Clustering. | - | - | - | - |
| 10. | Prediksi gaji dengan menggunakan Kemampuan Kecerdasan Emosional [22] | 2021 | Data survey penelitan 785 subjek. | Statistika | Regresi Multivariat. | Analisis dan prediksi gaji karyawan dengan metode regresi multivariat. | - | - | - | 0,33 |
| 11. | Prediksi gaji setelah tahun tertentu[23]. | 2020 | Data karyawan dari perusahaan. | *Machine Learning* | Regresi Linear, Regresi Polinomial | Prediksi gaji karyawan dengan metode regresi linear dan regresi polinomial. | - | - | - | - |
| 12. | Analisis empiris Teknik regresi dengan harga rumah dan prediksi gaji[24]. | 2020 | Dataset gaji pegawai dan harga rumah. | *Machine Learning* | *Multiple Linear Regression* | Prediksi gaji karyawan dan harga rumah dengan model *multiple linear regression*. | 1704,3 | 2904523,3 | 1410,9 | **0,92** |
| 13. | Prediksi gaji dalam penerapan model regresi dalam Data Mining[25]. | 2015 | Data gaji dan lama pengalaman bekerja. | *Machine Learning* | *Simple Linear Regression* | Prediksi gaji karyawan dari lama pengalaman bekerja dengan model *simple linear regression*. | - | - | - | - |
| 14. | Analisis dan Prediksi Kepuasan Gaji dalam Organisasi Negeri dan Swasta[26]. | 2011 | Data 250 karyawan staf-manajerial dan non-manajerial. | Statistika | *Simple Regression* | Prediksi kepuasan bekerja dengan metode statistika model *simple regression* (Akurasi 91,5%). | - | - | - | **0,915** |
| 15. | Perancangan Sistem Berbasis Machine Learning untuk Prediksi Kenaikan Gaji[27] | 2022 | Data dari database *enterprise resource planning* (ERP) | *Machine Learning, Deep Learning* | *Linear regression, artificial neural networks, random forest regression* | Perancangan sistem prediksi kenaikan gaji berbasis *machine* *learning* dengan menggunakan arsitektur *micro-services*. | - | - | - | - |
| 16. | Implementasi Sistem Prediksi Gaji Untuk Meningkatkan Motivasi Siswa[28]. | 2016 | 13.541 data mahasiswa yang lulus dengan gajinya. | *Machine Learning* | *K-NN* | Prediksi gaji model K-NN memberikan akurasi terbaik sebesar 84,69% sedangkan Multilayer perceptron memberikan akurasi terendah sebesar 38,08%. | - | - | - | 0,847 |
| *Naïve Bayes* | - | - | - | 0,436 |
| *Decision trees* | - | - | - | 0,74 |
| *Multilayer perceptron* | - | - | - | 0,381 |
| *SVM* | - | - | - | 0,437 |
| 17. | Prediksi Gaji di Pasar Kerja TI [29]. | 2020 | Data e-Recruitment khusus untuk pekerjaan TI di Spanyol. | *Machine Learning* | Linear models | Prediksi kisaran gaji yang akurat dengan menggunakan model *random* *forest* secara umum lebih baik dengan menghasilkan akurasi 84%. | - | - | - | 0,586 |
| Logistic regression | - | - | - | 0,792 |
| *KNN* | - | - | - | 0,591 |
| Multi-layer perceptrons | - | - | - | 0,663 |
| *SVM* | - | - | - | 0,836 |
| *Random forest* | - | - | - | 0,840 |
| *Vote* | - | - | - | 0,844 |
| *Vote3* | - | - | - | 0,837 |
| 18. | Analisis Prediktif Gaji Sumber Daya Manusia[30] | 2022 | Data ketenagakerjaan | *Machine Learning* | *Logistic Regression* | Prediksi gaji dengan model Support Vector Machine yang lebih akurat dibandingkan dengan Logistic Regression. | - | - | - | 0,79 |
| *Support Vector Machine* | - | - | - | 0,83 |
| 19. | Desain Mesin Prediksi baru untuk memprediksi  gaji[31] | 2018 | Data kepegawaian | *Machine Learning* | *Decision tree classifier* | Memprediksi gaji yang sesuai untuk suatu pekerjaan dengan metode *Decision tree classifier* dan *Random forest classifier.* | - | 389.64 | 6,04 | 0,844 |
| *Random forest classifier* | - | 329.12 | 5,04 | **0,873** |
| 20. | Analisis prediktif untuk pendapatan alumni[32] | 2022 | Data survei studi dampak alumni Universitas Tecnolog ico de Monterrey | *Machine Learning* | *Quantile Regression* | Memprediksi pendapatan alumni untuk mendapatkan wawasan tentang prediktor terkuat dan kelas 'berpenghasilan tinggi' menggunakan metode *Quantile Regression (QR)* | 50,431.45 | - | - | 0,44 |
| *Quantile Random Forest* | 47,325.67 | - | - | 0,51 |
| *Quantile Gradient Boosting* | 45,892.69 | - | - | 0,38 |
| *Linear Regression* | - | - | - | 0,48 |
| *Random Forest CLassifier* | - | - | - | 0,50 |
| *Gradient Boosting Classifier* | - | - | - | 0,53 |
| 21. | Komputasi cloud untuk fasilitas sinyal digital biomedis. [33] | 2021 | Data biomedis | *Framework Django* | | Mmenyediakan layanan yang mampu menangani dan memproses data biomedis melalui *code-free interface* dengan *framework django*. | | | | |
| 22. | Aplikasi Web Prediksi Diabetes [34] | 2021 | Data klinis penyakit diabetes | *Framework Django* | | Model *machine learning* di-*deploy* menjadi aplikasi web dikembangkan untuk memprediksi diabetes yang sesuai. | | | | |
| 23. | Deployment Klasifikasi Penyakit Paru-paru [13] | 2021 | Dataset x-ray tubuh bagian atas untuk Covid-19, Pneumonia, dan Normal. | *Framework Django* | | Klasifikasi penyakit paru-paru mengguanakn model CNN di-deploy ke dalam *framework django* untuk menyediakan antarmuka pengguna yang lebih baik untuk memprediksi output. | | | | |

## *Taksonomi Literature Review*



Gambar 2. 1 Taksonomi Studi Literatur

Penjelasan dari gambar 2.1 taksonomi studi literatur dari kumpulan penelitian-penelitian sebelumnya dan yang berkaitan dengan permasalahan dalam penelitian ini dijelaskan dari referensi [12 – 31] bahwa pendekatan *machine learning* dapat digunakan untuk memprediksi untuk digunakan dalam membuat model berdasarkan kualitas data tersebut. Adapun pendekatan *machine learning* yang memiliki akurasi terbaik yaitu:

Tabel 2.2 Klasifikasi Model berdasarkan SoTA

|  |  |  |
| --- | --- | --- |
| **No** | **Pendekatan** | **Nilai Akurasi** |
| 1. | *Mulivariate Linear Regression* [23] | 92% |
| 2. | *Simple Linear Regression* [25] | 91,5% |
| 3. | *Quantile Regression* [18] | 90% |
| 4. | *Random Forest Classifier* [30] | 87,3% |

Langkah selanjutnya yaitu membandingkan kinerja dari beberapa model *machine learning* dalam memprediksi gaji karyawan. Ada beberapa parameter yang dipilih sebagai pengukur kinerja machine learning, yaitu *Root Means Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Standard Error* (MSE) dan R2 Koefisien Determinasi (R2). Kemudian, ada beberapa model *machine learning* yang dirangkum dari referensi, yaitu *Multivariate Regression Modelling* dengan nilai akurasi 92%, *Simple Linear Regression* dengan nilai akurasi 91,5%, *Quantile Regression* dengan nilai akurasi 90%, dan *Random Forest Classifier* dengan akurasi 87,3%. Secara keseluruhan, model-model ini mampu mengidentifikasi parameter pembelajaran yang mempengaruhi perbedaan dalam memprediksi gaji karyawan. Namun berdasarkan model evaluasi yang digunakan sebagai parameter kinerja machine learning, *Multivariate Regression Modelling* menjadi salah satu pilihan yang tepat untuk memprediksi gaji karyawan.

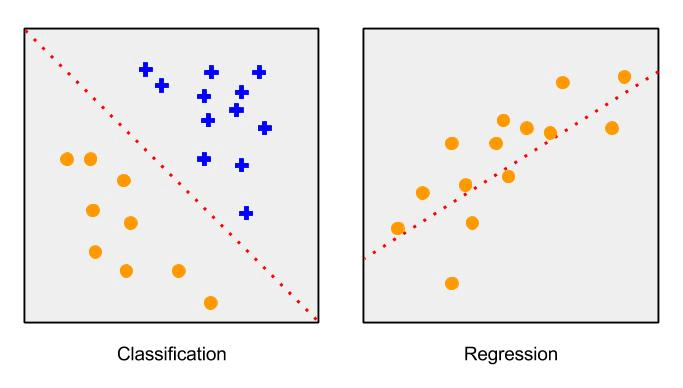
Selain itu, berdasarkan studi literatur dari kumpulan penelitian-penelitian sebelumnya, dimana permasalahan berupa visualisasi menggunakan *framework* django dapat dijelaskan dari referensi [32 – 34] bahwa dengan penggunaan bantuan *framework* django, kegiatan untuk melakukan keputusan dapat lebih mudah dilakukan tanpa harus menggunakan pengkodean, tetapi hanya dengan mengakses web-base yang mudah dimengerti.

## *Machine Learning*

*Machine Learning* atau pembelajaran mesin adalah cabang dari *artificial intelligence* yang fokus belajar dari data (*learn from data*), yaitu fokus pada pengembangan sistem yang mampu belajar secara “mandiri” tanpa harus berulang kali diprogram manusia. *Machine learning* membutuhkan data yang valid sebagai bahan belajar sebelum digunakan ketika testing untuk hasil *output* yang optimal[35]. Model *machine learning* pada umumnya dibagi menjadi tiga kategori, yaitu :

1. *Supervised Learning*
2. *Unsupervised Learning*
3. *Reinforcement Learning*

Pada penelitian ini, model *machine learning* yang digunakan adalah *supervised learning regression*. *Supervised learning* adalah metode *machine learning* yang membutuhkan pembelajaran fungsi yang sesuai denagn pasangan input nilai dengan output. *Supervised learning* mengekstrak pengetahuan dari *data training* berlabel dan setiap pasangan input dengan nilai berlabel[36]. *Supervised learning* membutuhkan data berlabel untuk membangun sebuah model. Ada dua variabel yang terdapat dalam *supervised learning*, yaitu variabel independen yang biasa disebut sebagai variabel X dan variabel dependen yang biasa disebut sebagai variabel Y. Pada umumnya, rumus pemetaan variabel X dan Y adalah Y = f(X). Rumus algoritma *supervised learning* ini digunakan untuk memperkirakan fungsi pemetaan (f) agar dapat memprediksi variabel Y ketika memiliki data input (variabel X) yang baru.



*Supervised learning* dibagi menjadi dua jenis, yaitu :

1. Klasifikasi (*Classification*) : menggunakan *supervised learning* untuk menetapkan *train data* ke dalam kategori tertentu secara akurat. Jenis ini dapat mengenali entitas tertentu dalam data dan mencoba untuk menarik beberapa kesimpulan bagaimana entitas tersebut harus diberi label atau didefinisikan.
2. Regresi (*Regression*) : menggunakan *supervised learning* untuk memahami hubungan antara variabel dependen dan variabel independen. Biasanya jenis *supervised learning* regresi ini digunakan untuk membuat proyeksi.

*Supervised learning* mendeteksi pola dalam *data training* dan menghasilkan fungsi yang dapat memprediksi pasangan *input* baru atau pengamatan yang tidak pernah terlihat. Algoritma tersebut dapat menggeneralisasikan fungsi untuk memprediksi secara akurat. Algoritma *supervised learning* menerapkan langkah-langkah sebagai berikut :

1. Mengumpulkan data dan sumber data yang relevan.
2. Memproses data dengan mengisi nilai-nilai yang miss, menormalkan data, dan menghapus data.
3. Menentukan jenis variabel target.
4. Memisahkan data (*train data* dan *test data*).
5. Melatih model *machine learning*.
6. Memprediksi.

Adapun model *machine learning* yang digunakan pada penelitian ini adalah sebagai berikut :

## *Mutivariate Linear Regression*

*Multivariate linear regression* adalah analisis regresi yang menjelaskan hubungan antara peubah respon (variabel dependen) dengan faktor-faktor yang mempengaruhi lebih dari satu prediktor (variabel independen)[37]. Ketika suatu hasil/keluaran, atau kelas berupa numerik, dan semua atribut adalah numerik, regresi linear adalah teknik yang tepat untuk menyelesaikan.

Ini adalah metode pokok di dalam ilmu statistik. Gunanya adalah untuk mengekspresikan kelas sebagai kombinasi linier dari atribut, dengan bobot yang telah ditentukan, dengan rumus sebagai berikut :

(1)

Keterangan :

= Variabel tidak bebas (nilai yang diprediksikan)

= Variabel bebas

= Konstanta (nilai apabila = 0)

= Koefisien regresi (nilai peningkatan ataupun penurunan)

Dimana nilai , , , … dapat dihitung dengan metode persamaan normal yaitu :

(2)

(3)

(4)

Selain dihitung dengan persamaan normal di atas, nilai 1, , , … dapat dihitung dengan metode kuadran terkecil yaitu :

(5)

(6)

(7)

Dimana :

(8)

(9)

(10)

(11)

(12)

(13)

## *Random Forest Classifier*

Metode *Random Forest* merupakan salah satu metode dalam *Decision Tree*. *Random forest* adalah kombinasi dari masing-masing *tree* yang baik kemudian dikombinasikan ke dalam satu model. *Random forest* bergantung pada sebuah nilai *vector random* dengan distribusi yang sama pada semua pohon yang masing-masing decision tree memiliki kedalaman yang maksimal. *Random forest* adalah *classifier* yang berbentuk pohon dimana dimana adalah *random vector* yang didistribusikan secara independent dan masing-masing *tree* pada sebuah unit kan memilih *class* yang paling popular input x. Berikut ini karakteristik akurasi pada *random forest*.

1. Memusatkan *random forest*

Terdapat *classifier* dan dengan *training set* dari distribusi *random vector* Berikut fungsi yang terbentuk.

(14)

Fungsi *error* yang digunakan..

(15)

Hasil dari penggabungan fungsi.

(16)

Pada hasil tersebut menjelaskan mengapa *random forest* tidak *overfit* saat *tree* ditambahkan, tetapi menghasilkan nilai yang terbatas pada *error*.

1. Kekuatan dan Korelasi

Fungsi yang dihasilkan adalah.

(17)

Pada fungsi tersebut kekuatan tidak bergantung pada forest.

1. *Random Forest* menggunakan seleksi input yang random

Bagging digunakan untuk pemilihan fitur secara random. Masing-masing *training set* diambil dengan penggantinya dari *training set* asli. Kemudian sebuah *tree* ditanam pada sebuah *training set* menggunakan seleksi fitur random. Ada dua alasan penggunaan *bagging* yaitu yang pertama penggunaan *bagging* untuk meningkatkan akurasi ketika fitur random digunakan. Yang kedua *bagging* digunakan untuk memberikan perkiraan dari kesalahan generalisasi (PE\*) dari gabungan *tree*, untuk memperkirakan kekuatan dan korelasi. *Random forest* yang paling sederhana dengan fitur random dibentuk dengan seleksi secara random, pada masing-masing *node*, sebuah grup kecil dari input variabel yang terbagi. Membentuk tree menggunakan metodologi CART ke ukuran yang maksimum.

1. *Random Forest* menggunakan kombinasi input yang linear

Misalkan terdapat beberapa input, M, F mengambil fraksi pada M yang kana memimpin dalam meningkatkan kekuatan tetapi pada korelasi yang tinggi. Pendekatan yang lain terbentuk dengan mendefinisikan lebih banyak fitur dengan mengambil kombinasi *random linear* dari sejumlah variabel input. Fitur tersebut variabel L yaitu jumlah dari variabel yang dikombinasikan. Variabel L secara random diseleksi dan ditambahkan bersama dengan koefisien yang memiliki nomor random [-1, 1]. Kombinasi linear F dihasilkan. Prosedur ini disebut *Forest-RC*[38].

## Statistika

Statistik adalah bidang studi yang berfokus pada pengumpulan, menafsirkan, dan mewakili set data numerik (angka). Statistik juga berkaitan dengan organisasi, analisis, interpretasi, dan penyajian data yang seting digunakan pada maslaah ilmiah, industry, atau social. Statistik berkaitan dengan berbagai metode, salah satunya yaitu regresi. Regresi adalah metode statistik yang dipakai untuk memperkirakan hubungan antara sebuah variabel terikat dan satu atau lebih variabel bebas. Metode ini juga bisa digunakan untuk menilai kekuatan hubungan antara variabel dengan perkiraan masa depan[39]. Adapun model statistika yang digunakan pada penelitian ini adalah sebagai berikut :

## Simple Linear Regression

Simple linear regression adalah sebuah metode pendekatan untuk pemodelan hubungan antara satu variabel dependen dan satu variabel independent. Dalam regresi, variabel independent menerangkan variabel dependennya. Dalam analisis regresi sederhana, hubungan antar variabel bersifat linear, dimana perubahan pada variabel X akan diikuti oleh perubahan variabel Y secara tetap. Sementara pada hubungan non-linear, perubahan variabel X tidak diikuti variabel Y secara proposional. Model analisis regresi linear sederhana sebagai berikut :

(18)

Keterangan :

= Variabel Dependen

= Konstanta

= Koefisien Regresi

Komponen pada simple linear regression ada tiga yaitu a sebagai intersep, b sebagai slope, dan x sebagai indeks waktu. Persamaan untuk mendapatkan nilai a dan b adalah :

(19)

(20)

Langkah-langkah metode yang diusulkan berdasarkan simple linear regression sebagai berikut :

1. Pembuatan dataset
2. Pembentukan model linear regresi

Langkah pembentukan model sebagai berikut :

1. Langkah 1 : Hitung , , dan total dari masing-masingnya.
2. Langkah 2 : Hitung a dengan menggunakan persamaan (2) dan b menggunakan persamaan (3).
3. Langkah 3 : Buatkan model persamaan simple linear regression.
4. Langkah 4 : Lakukan prediksi atau peramalan terhadap variabel faktor penyebab atau variabel akibat[40].

## Quantile Regression

Kuantil adalah titik dalam distribusi yang berhubungan dengan urutan peringkat nilai dalam distribusi tersebut. Metode quantile regression memungkinkan pemahaman hubungan antar variabel di luar rata-rata data, sehingga berguna dalam memahami hasil yang tidak terdistribusi secara normal dan yang memiliki hubungan non-linear dengan variabel prediktor. Quantile Regression memungkinkan analis untuk berasumsi bahwa variabel beroperasi sama di ekor atas distribusi seperti pada rata-rata dan untuk mengidentifikasi faktor-faktor yang merupakan penentu penting variabel. Quantile Regression digunakan ketika :

1. Untuk memperkirakan median, atau kuantil 0,25 atau kuantil apapun.
2. Asumsi kunci regresi linear tidak terpenuhi
3. Outlier dalam data
4. Residu tidak normal
5. Peningkatan varians kesalahan dengan peningkatan variabel hasil

Berikut adalah model linear regression :

(21)

Garis linear regression terbaik ditemukan dengan meminimalkan kesalahan kuadrat rata-rata.

(22)

Persamaan model Quantile Regression untuk kuantil ke- adalah.

(23)

Dimana p adalah jumlah variabel regressor, n adalah jumlah titik data. Garis regresi kuantil terbaik ditemukan dengan minimalkan dengan meminimalkan penyimpangan absolut median.

(24)

Di sini fungsi adalah fungsi cek yang memberikan bobot asimetris pada kesalahan tergantung pada kuantil dan tanda keseluruhan kesalahan. Secara matematis, mengambil bentuk[41] [] :

(25)

## Metode Evaluasi Model *Machine Learning*

Evaluasi model *machine learning* digunakan untuk menilai kualitas model tersebut dalam memprediksi. Model-model tersebut dievaluasi dengan cara melihat nilai error yang diperoleh pada model prediksi. Ada beberapa parameter perhitungan untuk melihat nilai error tersebut, yaitu *root mean square error* (RMSE), *mean square error* (MSE), dan *mean absolute error* (MAE). Adapun rinciannya sebagai berikut :

### *Root Mean Square Error* (RMSE)

*Root Mean Square Error* (RMSE) adalah metode pengukuran dengan mengukur perbedaan nilai dari prediksi sebuah model sebagai estimasi atas nilai yang diobservasi. Metode estimasi *root mean square error* (RMSE) lebih kecil dari 0,5 dikatakan lebih akurat. Berikut rumus dari RMSE[42] :

(26)

Keterangan :

Y’ = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

### *Mean Square Error* (MSE)

*Mean squared error* (MSE) adalah rata-rata kesalahan kuadrat diantara nilai actual dan nilai peramalan. Metode MSE secara umum digunakan untuk melihat estimasi berapa nilai kesalahan pada peramalan. Nilai MSE yang rendah atau mendekati nol menunjukkan bahwa hasil peramalan sesuai dengan data aktual dan dapat dijadikan untuk perhitungan peramalan di periode mendatang[42].

(27)

Keterangan :

Y’ = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

### *Mean Absolute Error* (MAE)

*Mean absolute error* (MAE) adalah model evaluasi yang menunjukkan kesalahan rata-rata yang error dari nilai sebenarnya dengan nilai prediksi. Jika nilai MAE lebih kecil dari 0,2 maka semakin baik model tersebut dalam melakukan prediksi. Rumus dari MAE sebagai berikut[43] :

(28)

Keterangan :

Y’ = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

# BAB III

# TUJUAN DAN MANFAAT

## Tujuan dan Manfaat Penelitian

Tujuan penelitian memuat uraian yang menyebutkan secara spesifik maksud dan tujuan yang ingin dicapai dari penelitian. Menurut Beckingham (1971), tujuan penelitian adalah ungkapan mengapa penelitian itu dilakukan. Tujuan dari suatu penelitian dapat menggambarkan suatu konsep dan menjelaskan suatu situasi atau solusi yang mengindikasikan jenis stui yang akan dilakukan. Tujuan penelitian menentukan arah dari suatu penelitian dan merinci apa saja yang ingin dicapai.

Manfaat penelitian merupakan dampak dari penyampaiannya tujuan. Menurut Sugiyono (2011), manfaat penelitian merupakan jawaban atas tujuan penelitian yang dibahas dalam hasil penelitian guna mendapatkan sistem pengetahuan dalam memahami, memecahkan, dan mengantisipasi masalah yang sudah ditrumuskan dalam topik penelitian.

Berdasarkan hal tersebut Adapun tujuan dan manfaat penelitian yang dilakukan yang akan dijelaskan sebagai berikut.

### Tujuan Penelitian

Adapun tujuan pada penelitian ini sebagai berikut.

1. Menganalisa korelasi data gaji karyawan berdasarkan faktor-faktor spesifik.
2. Menggunakan pendekatan *machine learning* yaitu model *multivariate linier regression* untuk pemodelan prediksi gaji karyawan berdasarkan parameter dari faktor-faktor spesifik seperti umur, *job level*, total lama bekerja, masa bakti.
3. Menggunakan *framework* django untuk menyajikan hasil prediksi gaji karyawan.

### Manfaat Penelitian

Adapun manfaat pada penelitian ini sebagai berikut.

1. Merekomendasikan model prediksi gaji karyawan berdasarkan kualitas dari faktor-faktor spesifik.
2. Membuat tampilan *framework* agar mudah digunakan untuk melakukan prediksi gaji karyawan secara *realtime*.

# BAB IV METODE PENELITIAN

## Diagram Alur Metodologi Penelitian

Metode penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan manfaat tertentu. Untuk mencapai tujuan tersebut, diperlukan suatu metode yang relevan dengan tujuan yang ingin dicapai. Sugiyono (2017:2) menyatakan bahwa metode penelitian pada dasarnya merupakan cara ilmiah yang digunakan untuk mendapatkan data dengan tujuan dan kegunaan tertenu. Metode penelitian tersebut bisa berupa diagram alur penelitian.

### Diagram Alur Utama



Gambar 4.1 Gambar Diagram Alur Metodologi Penelitian

Berdasarkan diagram alur metodologi penelitian diatas, terdapat indikator capaian sebagai berikut.

Tabel 4.1 Penjelasan Diagram Alur Metodologi Penelitian

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Tahapan** |  | **Indikator capaian** |
| 1. | Identifikasi dan perumusan masalah | 🡪 | 1. Mind map prediksi gaji terhadap faktor-faktor spesifik berdasarkan data dan pada metode *machine learning* serta *framework* django. |
| 2. | Studi literatur | 🡪 | 1. *Mind map* prediksi gaji terhadap faktor-faktor spesifik berdasarkan data dan pada metode *machine learning* serta *framework* django.. |
| 3. | Pengumpulan data | 🡪 | 1. Data mentah dari berbagai faktor (umur,*job level*,lama pengalaman bekerja,masa bakti) |
| 3. | Pre-pemrosesan data | 🡪 | 1. *Pre-processed* data yang sudah siap untuk pemodelan dengan tahapan pembersihan, penanganan nilai yang hilang dan transformasi. |
| 4. | Pemodelan | 🡪 | 1. Model Multivariat Linier Regresi digunakan untuk memprediksi gaji berdasarkan data dari setiap faktor-faktor spesifik. |
| 5. | Evaluasi | 🡪 | 1. Performansi model |
| 6. | Diseminasi hasil | 🡪 | 1. Artikel yang diterbitkan dalam jurnal nasional terakreditasi SINTA 3, HAKI dan Buku |

# BAB V

# HASIL DAN LUARAN YANG DICAPAI

## Kegiatan dan Hasil Pelaksanaan

Kegiatan dan hasil pelaksanaan program hibah penelitian internal ini ditujukan untuk mendapatkan luaran yang diharapkan. Kegiatan dan hasil tersebut dapat dipaparkan seperti berikut.

* + 1. **Pelaksanaan Penelitian**

Kegiatan penelitian yang dilakukan telah berjalan selama x bulan oleh anggota tim. Penelitian dilakukan secara bertahap dengan didasarkan metode penelitian yang telah dirancang sebelumnya. Pelaksanaan penelitian ini dilakukan dengan cara berdiskusi dan mencari solusi berdasarkan literatur yang sesuai dengan tema penelitian. Permasalahan-permasalahan yang didapatkan selama penelitian dapat diatasi dengan mengenali masalah untuk kemudian dilakukannya pencarian solusi secara berdiskusi berdasarkan tinjauan literatur.

Hasil penelitian yang telah dilaksanakan mampu mencapai hasil yang diinginkan. Dimana penelitian ini berhasil mengatasi masalah berupa bagaimana cara memprediksi gaji karyawan, yang dilakukan dengan menggunakan kode pemrograman bahasa Python.

Berdasarkan penelitian yang dilakukan, didapatkan hasil nilai akurasi sebesar 0,909. Akurasi tersebut merupakan nilai akurasi yang baik, sehingga dapat dikatakan model *machine learning* dapat berperforma baik untuk memprediksi gaji. Berdasakan uji validitas, nilai akurasi 0,909 menunjukkan bahwa gaji dipengaruhi oleh faktor independen (umur, masa bakti, dan lama pengalaman bekerja) sebesar 0,909 atau 90,9%. Nilai sisa dari akurasi tersebut adalah 0,091 atau 9,1% yang artinya gaji dipengaruhi oleh faktor lain yang tidak diketahui sebesar 9,1%.

* + 1. **Dataset yang Digunakan**

Tabel Karakteristik Dataset yang Digunakan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Kode** | **Variabel** | **Jenis Variabel** | **Tipe Data** | **Sumber Data** |
| X1 | Umur | Sektor Penggajian | Numerik | Hasil Kuesioner |
| X2 | Job Level | Numerik |
| X3 | Masa Bakti | Numerik |
| X4 | Lama Pengalaman Bekerja | Numerik |
| X5 | Gaji | Numerik |

Kumpulan dataset pada tabel di atas yang telah disiapkan memiliki implikasi untuk memprediksi gaji dengan melakukan eksplorasi pengaruh berdasarkan faktor-faktor spesifik diantaranya umur, *job level*, masa bakti, lama pengalaman bekerja.

* + 1. **Pembuatan Aplikasi Prediksi Gaji Pegawai**

Pembuatan Aplikasi Prediksi Gaji Pegawai dilakukan secara bersamaan ketika melakukan proses pembuatan kode program *machine learning* untuk memprediksi gaji pegawai serta penyusunan laporan penelitian.

Aplikasi telah dibuat oleh anggota tim dengan menggunakan bahasa pemrograman Python dengan *framework* Django. Aplikasi telah berhasil dijalankan sesuai dengan harapan yaitu melakukan prediksi gaji pegawai berdasarkan usia, level pekerjaan, tahun masa bakti, dan lama bekerja di perusahan.

* + 1. **Penyusunan Jurnal Ilmiah Nasional**

Penyusunan Jurnal Ilmiah Nasional dilakukan oleh anggota tim ketika masalah dari penelitian telah diatasi dan hasil penelitian telah mencapai hasil yang diharapkan. Dimana jurnal ilmiah akan dipublikasikan dan ditargetkan untuk mencapai jurnal nasional SINTA S3. Susunan jurnal tersebut terdiri dari judul, abstrak, pendahuluan, tinjauan pustaka, implementasi, kesimpulan dan saran, serta daftar pustaka.

* + 1. **Penyusunan Buku ISBN**

Penyusunan buku ISBN dilakukan oleh anggota tim ketika masalah dari penelitian telah mendapatkan hasil yang dicapai. Penyusunan buku dilakukan selama x bulan. Buku yang disusun dari BAB I hingga BAB VI. Buku yang disusun terdiri cover, kata sambutan, kata pengantar dan terdiri dari 250 halaman terdaftar di Perpustakaan Nasional Republik Indonesia dan ber-ISBN.

* + 1. **Pelaksanaan Praktikum Pada Mata Kuliah *Database***

Pelaksanaan pada praktikum mata kuliah *Database* menggunakan hasil penelitian ini yaitu buku berlisensi ISBN.

## Luaran yang Dicapai

Sebagaimana yang telah diungkapkan pada latar belakang, kegiatan-kegiatan yang telah dilaksanakan dan hasil diperoleh dari pelaksanaan program hibah penelitian internal, dimaksudkan untuk mendapatkan luaran yang ditargetkan. Sesuai dengan target luaran yang ada, adapun capaian luaran dari program ini yang dapat dipaparkan sebagai berikut.

1. Publikasi ilmiah pada jurnal nasional

Publikasi ilmiah pada jurnal nasional pada penelitian ini akan dipublikasikan untuk target junral nasional SINTA S3. Jurnal penelitian ini telah dihasilkan dalam bentuk draf. Draf tersebut belum sepenuhnya selesai karena belum diriview lebih lanjut oleh anggota tim maupun diskusi bersama.

1. Buku ISBN

Buku ISBN adalah luaran yang ditargetkan untuk publikasi. Buku ISBN sebagai salah satu luaran dari pelaksanaan program penelitian ini telah dibuat dalam bentuk draf, dimana penyusunan buku ini disusun sebanyak enam bab. Bab tersebut juga belum dikoreksi lebih lanjut karena belum melakukan diskusi lebih lanjut. Capaian luaran ini sesuai dengan target yaitu berupa publikasi buku.

1. Hak atas kekayaan intelektual

Hak atas kekayaan intelektual pada penelitian ini didasarkan publikasi buku yang telah dipaparkan sebelumnya. Pada saat ini belum adanya hak atas kekayaan intelektual karena penyusunan dan publikasi buku belum sepenuhnya dituntaskan. Target dari luaran ini adalah mendapatkan HAKI berdasarkan penyusunan buku dari hasil program penelitian.

1. Referensi praktikum pada matakuliah Database

Penelitian yang telah dilakukan akan dijadikan sebagai referensi praktikum pada matakuliah Database jurusan Diploma 4 Teknik Informatika. Capaian ini ditargetkan untuk terlaksananya praktikum pada matakuliah tersebut sebagai bahan ajar referensi.

Sebagaimana yang telah diungkapkan pada latar belakang, kegiatan-kegiatan yang telah dilaksanakan

Luaran yang dicapai dengan target capaiannya dapat disajikan dalam bentuk tabel berikut.

Tabel 5.1 Luaran dan Target Capaian

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Jenis Luaran** | **Target** | **Capaian** |
| 1. | Publikasi jurnal ilmiah nasional | Publikasi SINTA S3 | Draf |
| 2. | Buku ISBN | Publikasi Buku | Draf |
| 3. | Hak atas Kekayaan Intelektual | Mendapatkan HAKI | Belum ada |
| 4. | Referensi praktikum pada matakuliah Database | Terlaksananya praktikum berdasarkan refernsi penelitian | Belum ada |

Berdasarkan tabel diatas dapat dijelaskan bahwa dari 4 (empat) jenis luaran yang ada, capaian luaran belum sepenuhnya mencapai target. Tentunya, capaian tersebut akan terus dikembangkan dengan terus melakukan koreksi dan diskusi untuk mencapai target yang telah ditetapkan.

## Hasil Penelitian

Kajian tentang prediksi gaji pegawai telah banyak dilakukan, peneliti memprediksi terdapat pengaruh positif dan signifikan antara gaji dan masa kerja terhadap kinerja karyawan[44] []. Studi ini mengusulkan metode prediksi *Machine* *Learning* dengan menganalisis data penggajian yang dikumpulkan dengan metode angket (kuesioner). Penelitian menggunakan model *Linear* *Regression* sebagai algoritma *Machine* *Learning*. Hasil penelitian menunjukan nilai akurasi sebesar 96% atau 0.96, sehingga dikatakan bahwa model tersebut memiliki nilai yang baik.

Pada penelitian ini, peneliti melakukan hal sama dengan memprediksi gaji pegawai dengan menggunakan pendekatan *machine* *learning*. Adapun model yang diusulkan adalah model Regresi Linier Berganda atau *Multivariate* *Linear* *Regression* (MLR). Data yang digunakan pada penelitian ini adalah dataset kepegawaian yang terdiri atas data training dan data test, dimana data ini diperoleh dari Kaggle.

|  |
| --- |
| df\_train = pd.read\_csv('employee\_attrition\_train.csv')  df\_train  df\_test = pd.read\_csv('employee\_attrition\_test.csv')  df\_test |

Data bersifat tidak berlabel dan data diolah dengan teknik *Unsupervised Learning* untuk mendapatkan hasil model prediksi dengan menggunakan metode regresi linier berganda. Berikut adalah pemaparan data yang digunakan.

Tabel 5. 1 Dataset Data Training

|  |  |  |  |
| --- | --- | --- | --- |
| Age | Attrition | ... | YearsWithCurrManager |
| 50 | No |  | 3 |
| 36 | No |  | 1 |
| 21 | Yes |  | 0 |
| 50 | No |  | 7 |
| ... | ... | ... | ... |
| 41 | No |  | 2 |
| 22 | Yes |  | 0 |
| 29 | No |  | 3 |
| 50 | No |  | 0 |

Tabel 5. 2 Dataset Data Testing

|  |  |  |  |
| --- | --- | --- | --- |
| Age | BusinessTravel | ... | YearsWithCurrManager |
|  | Travel\_Rarely |  | 3 |
| 53 | Travel\_Rarely |  | 3 |
| 24 | Travel\_Rarely |  | 0 |
| 45 | Travel\_Rarely |  | 0 |
| ... | ... | ... | ... |
| 27 | Non-Travel |  | 4 |
|  | Travel\_Rarely |  | 2 |
| 39 | Travel\_Rarely |  | 4 |
|  | Travel\_Rarely |  | 0 |

Adapun Tabel 5.4 dan Tabel 5.5 menjelaskan tentang attribut data-data tersebut adalah sebagai berikut.

Tabel 5. 3 Atribut Dataset Data Training

|  |  |  |
| --- | --- | --- |
| Attribut | Keterangan | Tipe Data |
| Age | Usia pegawai | numerik |
| Attrition | Eliminasi pegawai | object |
| BusinessTravel | Perjalanan bisnis pegawai | object |
| DailyRate | Tarif harian pegawai | numerik |
| Department | Departemen pegawai | object |
| DistanceFromHome | Jarak perusahaan dari rumah pegawai | numerik |
| Education | Pendidikan pegawai | object |
| EducationField | Bidang pendidikan pegawai | numerik |
| EmployeeCount | Jumlah pegawai | numerik |
| EmployeeNumber | Nomor pegawai | numerik |
| EnvironmentSatisfaction | Kepuasan lingkungan pegawai | numerik |
| Gender | Gender pegawai | object |
| HourlyRate | Tarif per jam pegawai | numerik |
| JobInvolvement | Keterlibatan kerja pegawai | numerik |
| JobLevel | Tingkat kerja pegawai | numerik |
| JobRole | Peran pekerjaan pegawai | object |
| JobSatisfaction | Kepuasan kerja pegawai | numerik |
| MaritalStatus | Status perkawinan pegawai | object |
| MonthlyIncome | Penghasilan bulanan pegawai | numerik |
| MonthlyRate | Tarif bulanan pegawai | numerik |
| NumCompaniesWorked | Jumlah perusahaan yang dikerjakan pegawai | numerik |
| Over18 | Pegawai dengan usia lebih dari 18 tahun | object |
| OverTime | Lembur pegawai | object |
| PercentSalaryHike | Persen kenaikan gaji pegawai | numerik |
| PerformanceRating | Peringkat kinerja pegawai | numerik |
| RelationshipSatisfaction | Kepuasan hubungan pegawai | numerik |
| StandardHours | Jam standar pegawai | numerik |
| StockOptionLevel | Tingkat opsi saham pegawai | numerik |
| TotalWorkingYears | Jumlah tahun kerja pegawai | numerik |
| TrainingTimesLastYear | Waktu pelatihan tahun lalu pegawai | numerik |
| WorkLifeBalance | Keseimbangan kehidupan kerja pegawai | numerik |
| YearsAtCompany | Tahun di perusahaan pegawai | numerik |
| YearsInCurrentRole | Tahun berperan sekarang pegawai | numerik |
| YearsSinceLastPromotion | Tahun sejak promosi terakhir pegawai | numerik |
| YearsWithCurrManager | Tahun dengan manajer saat ini pegawai | numerik |

Tabel 5. 4 Atribut Dataset Data Testing

|  |  |  |
| --- | --- | --- |
| Attribut | Keterangan | Tipe Data |
| Age | Usia pegawai | numerik |
| BusinessTravel | Perjalanan bisnis pegawai | object |
| DailyRate | Tarif harian pegawai | numerik |
| Department | Departemen pegawai | object |
| DistanceFromHome | Jarak perusahaan dari rumah pegawai | numerik |
| Education | Pendidikan pegawai | object |
| EducationField | Bidang pendidikan pegawai | numerik |
| EmployeeCount | Jumlah pegawai | numerik |
| EmployeeNumber | Nomor pegawai | numerik |
| EnvironmentSatisfaction | Kepuasan lingkungan pegawai | numerik |
| Gender | Gender pegawai | object |
| HourlyRate | Tarif per jam pegawai | numerik |
| JobInvolvement | Keterlibatan kerja pegawai | numerik |
| JobLevel | Tingkat kerja pegawai | numerik |
| JobRole | Peran pekerjaan pegawai | object |
| JobSatisfaction | Kepuasan kerja pegawai | numerik |
| MaritalStatus | Status perkawinan pegawai | object |
| MonthlyIncome | Penghasilan bulanan pegawai | numerik |
| MonthlyRate | Tarif bulanan pegawai | numerik |
| NumCompaniesWorked | Jumlah perusahaan yang dikerjakan pegawai | numerik |
| Over18 | Pegawai dengan usia lebih dari 18 tahun | object |
| OverTime | Lembur pegawai | object |
| PercentSalaryHike | Persen kenaikan gaji pegawai | numerik |
| PerformanceRating | Peringkat kinerja pegawai | numerik |
| RelationshipSatisfaction | Kepuasan hubungan pegawai | numerik |
| StandardHours | Jam standar pegawai | numerik |
| StockOptionLevel | Tingkat opsi saham pegawai | numerik |
| TotalWorkingYears | Jumlah tahun kerja pegawai | numerik |
| TrainingTimesLastYear | Waktu pelatihan tahun lalu pegawai | numerik |
| WorkLifeBalance | Keseimbangan kehidupan kerja pegawai | numerik |
| YearsAtCompany | Tahun di perusahaan pegawai | numerik |
| YearsInCurrentRole | Tahun berperan sekarang pegawai | numerik |
| YearsSinceLastPromotion | Tahun sejak promosi terakhir pegawai | numerik |
| YearsWithCurrManager | Tahun dengan manajer saat ini pegawai | numerik |

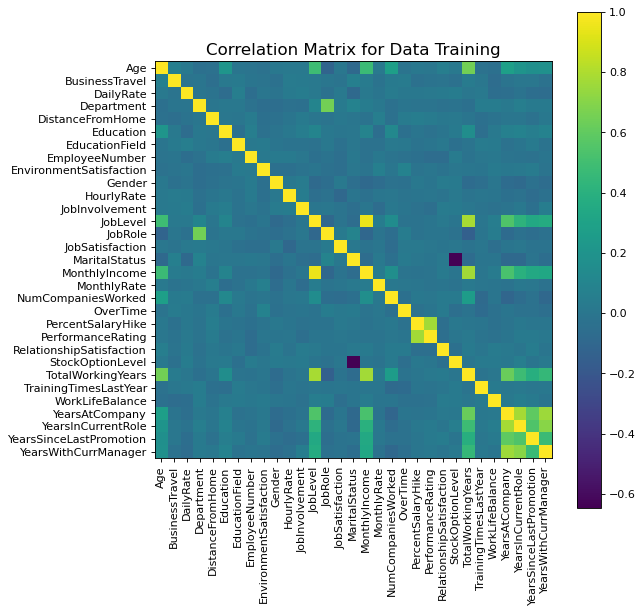
Proses selanjutnya adalah mengganti kolom yang memiliki tipe data object menjadi numerik menggunakan metode encoder.

|  |
| --- |
| # 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  # Encoder EducationField Variable  df\_train['EducationField'] = df\_train['EducationField'].astype('category')  # Assigning numerical values and storing in another column  df\_train['EducationField'] = df\_train['EducationField'].cat.codes  # Encoder Gender Variable  df\_train['Gender'] = df\_train['Gender'].astype('category')  # Assigning numerical values and storing in another column  df\_train['Gender'] = df\_train['Gender'].cat.codes  # Encoder JobRole Variable  df\_train['JobRole'] = df\_train['JobRole'].astype('category')  # Assigning numerical values and storing in another column  df\_train['JobRole'] = df\_train['JobRole'].cat.codes  # Encoder MaritalStatus Variable  df\_train['MaritalStatus'] = df\_train['MaritalStatus'].astype('category')  # Assigning numerical values and storing in another column  df\_train['MaritalStatus'] = df\_train['MaritalStatus'].cat.codes  # Encoder Over18 Variable  df\_train['Over18'] = df\_train['Over18'].astype('category')  # Assigning numerical values and storing in another column  df\_train['Over18'] = df\_train['Over18'].cat.codes  # Encoder OverTime Variable  df\_train['OverTime'] = df\_train['OverTime'].astype('category')  # Assigning numerical values and storing in another column  df\_train['OverTime'] = df\_train['OverTime'].cat.codes  df\_train |

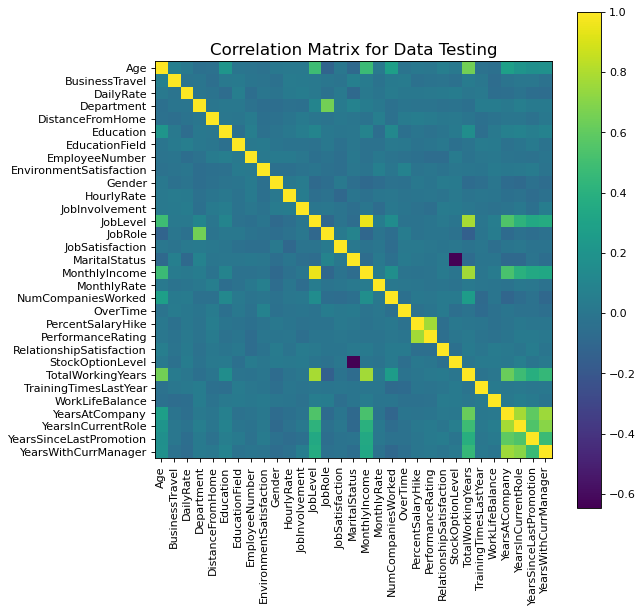
Proses selanjutnya adalah mengisi nilai yang hilang pada dataset tersebut dengan metode mengisi nilai yang hilang menggunakan mean.

|  |
| --- |
| #Cek Apakah Ada Data yang Kosong?  df\_train.isnull().values.any()  #Data kosong pada kolom  df\_train.isnull().sum()  #Mengambil kolom Age untuk diisi dengan mean-nya  Age = df\_train['Age']  Age.describe()  df\_train.Age = df\_train.Age.fillna(value=df\_train.Age.mean())  #Mengambil kolom DailyRate untuk diisi dengan mean-nya  DailyRate = df\_train['DailyRate']  DailyRate.describe()  df\_train.DailyRate = df\_train.DailyRate.fillna(value=df\_train.DailyRate.mean())  #Mengambil kolom DistanceFromHome untuk diisi dengan mean-nya  DistanceFromHome = df\_train['DistanceFromHome']  DistanceFromHome.describe()  df\_train.DistanceFromHome = df\_train.DistanceFromHome.fillna(value=df\_train.DistanceFromHome.mean()) |

Kemudian dilakukan langkah pemilihan atribut yang akan digunakan dengan menggunakan heatmap correlation untuk memilih fitur atau variabel independen yang berhubungan erat dengan variabel dependen model, yaitu “Monthly Income”. Pemilihan variabel independen dilihat berdasarkan nilai korelasi yang memiliki tingkat hubungan sedang, kuat, dan sangat kuat. Hasil korelasi dataset data training dapat dilihat pada gambar 5.1 dan dataset data testing pada gambar 5. 2.



Gambar 5. 1 Heatmap Correlation Dataset Data Training



Gambar 5. 2 Heatmap Correlation Dataset Data Testing

Berdasarkan nilai korelasi dataset data training pada gambar 5.1 dan korelasi dataset data testing pada gambar 5.2 di atas dapat diperhatikan bahwa variabel independen Age, JobLevel, TotalWorkingYears, dan YearsAtCompany memiliki hubungan kuat terhadap MonthlyIncome dengan rata-rata nilai akurasi 0,66. Sedangkan variabel lainnya memiliki hubungan yang lemah terhadap MonthlyIncome.

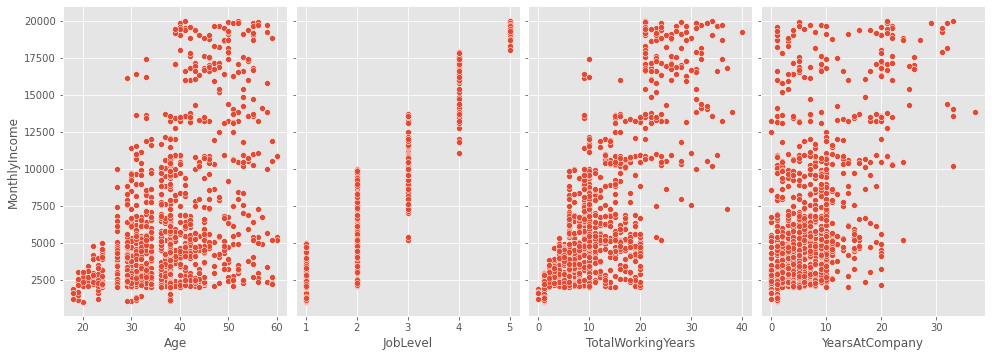
Selanjutnya dataset data training dan data testing dieliminasi berdasarkan pemilihan variabel yang digunakan berdasarkan nilai akurasi untuk melakukan prediksi. Variabel independent terdiri dari Age, JobLevel, TotalWorkingYears, dan YearsAtCompany. Sedangkan variabel dependen adalah MonthlyIncome.

|  |
| --- |
| # Eliminasi Variabel yang Tidak akan digunakan  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)  df\_test\_clean = df\_test.drop(['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) |

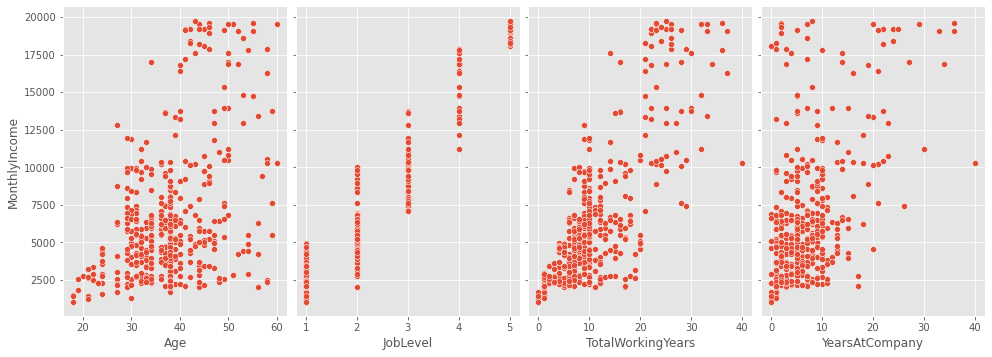
Proses selanjutnya adalah melakukan penetapan variabel independen (sumbu x) yaitu Age, JobLevel, TotalWorkingYears, dan YearsAtCompany dan variabel dependen (sumbu y) yaitu Monthly Income.

|  |
| --- |
| # Menentukan variabel X dan variabel Y  x\_train = df\_train\_clean[['Age', 'JobLevel', 'TotalWorkingYears', 'YearsAtCompany']]  y\_train = df\_train\_clean[['MonthlyIncome']]  df\_train\_clean.to\_csv('employee\_attrition\_train\_clean.csv')  x\_test = df\_test\_clean[['Age', 'JobLevel', 'TotalWorkingYears', 'YearsAtCompany']]  y\_test = df\_test\_clean[['MonthlyIncome']]  df\_test\_clean.to\_csv('employee\_attrition\_test\_clean.csv') |

Proses selanjutnya adalah melakukan visualisasi penyebaran data dari data training dan data testing menggunakan seaborn. Visualisasi dapat dilihat pada gambar 5.3 untuk data training dan gambar 5.4 untuk data testing.



Gambar 5. 3 Visualisasi Penyebaran Data Training



Gambar 5. 4 Visualisasi Penyebaran Data Testing

Selanjutnya, membuat variabel regressor yang menggunakan metode LinearRegression, kemudian membuat variabel persamaan menggunakan *method* regressor.fit dengan parameter x\_train dan y\_train.

|  |
| --- |
| regressor = LinearRegression()  persamaan = regressor.fit(x\_train, y\_train)  print(regressor.coef\_)  print(regressor.intercept\_) |

Berdasarkan kode tersebut, terdapat output yang dihasilkan yaitu sebagai berikut.

|  |
| --- |
| [[ -5.05425997 3871.75302822 46.94057998 -9.84604878]]  [-1728.52019705] |

Dari model tersebut didapat nilai koefisien dari variabel Independen. Nilai koefisien dari Age adalah -5,054 yang berarti tiap karyawan yang mengalami pertambahan satu tahun umur kerja, maka akan mengalami penurunan gaji sebesar 5,054. Nilai koefisien dari JobLevel adalah 3871,7530 yang berarti tiap karyawan yang mengalami pertambahan satu tingkat job level akan mengalami kenaikan gaji sebesar 3871,7530. Nilai koefisien TotalWorkingYears adalah 46,9405 yang berarti tiap karyawan yang mengalami pertambahans satu tahun pengalaman bekerja akan mengalami kenaikan kerja sebesar 46,9405. Nilai koefisien YearsAtCompany adalah -9,8460 yang berarti tiap karyawan yang mengalami pertambahan satu tahun akan mengalami penurunan gaji sebesar 9,8460.

Selanjutnya adalah mencari konstanta/intercept menggunakan regressor. Intecept merupakan sebuah koefisien dari sebuah persamaan model regresi linear. Untuk kasus ini, berarti untuk variabel X yang bernilai nol atau karyawan yang belum berpengalaman kerja, karyawan tersebut akan menerima gaji sebesar 26611 per tahunnya.

|  |
| --- |
| y\_pred = regressor.predict(x\_test) |

Proses selanjutnya adalah melakukan prediksi *data testing* menggunakan model *machine learning*. Kemudian buat kolom baru yang bernama MonthlyIncome Prediction yang berisikan nilai prediksi.

|  |
| --- |
| vis\_test['MonthlyIncome Prediction'] = y\_pred.tolist()  vis\_test |

Berikut persamaan umum dari model linear regresi multivariable.

Y = β0 + β1X1 + β2X2 + β3X3 + β4X4 +….+ βnXn

β0 adalah nilai *intersept* dari persamaan linear, dan β1, β2, β3 sampai dengan βn adalah konstanta dari variabel independen. Berdasarkan nilai koefisien variabel independen dan *Intersept* didapat, maka persamaan regresi linear multivariabel sebagai berikut :

Y = -1728 – 5,054X1 + 3871,7530X2 + 46,9405X3 – 9,8460X4

Keterangan:

Y = Variabel Dependen yaitu MonthlyIncome

X1 = Variabel Independen pertama yaitu Age

X2 = Variabel Independen kedua yaitu JobLevel

X3 = Variabel Independen ketiga yaitu TotalWorkingYears

X4 = Variabel Independen keempat yaitu YearsAtCompany

Maka dapat disimpulkan, persamaan regresi linear multivariabel sebagai berikut :

**MonthlyIncome = -1728 – 5,054(Age) + 3871,7530(JobLevel) + 46,9405(TotalWorkingYears) – 9,8460(YearsAtCompany)**

Pada tahap selanjutnya yaitu evaluasi data, dimana evaluasi yang dilakukan pertama adalah menilai akurasi model dengan menggunakan metode R Square.

|  |
| --- |
| from sklearn.metrics import r2\_score  r2 = r2\_score (y\_test, y\_pred)  print ("Nilai R2 adalah ", r2) |

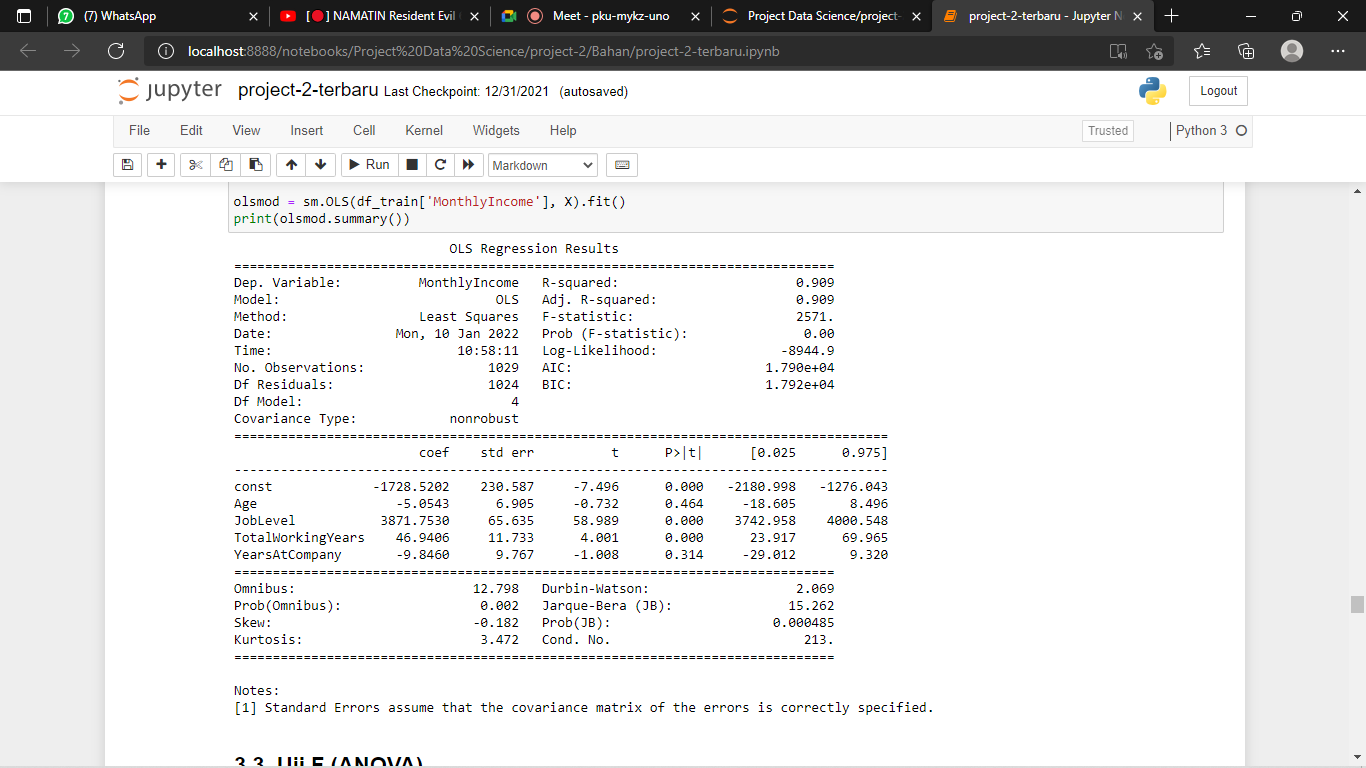
Adapun output nilai R square dari model yang digunakan adalah sebagai berikut.

|  |
| --- |
| Nilai R2 adalah 0.894582813656067 |

Berdasarkan nilai R square tersebut, dapat dijelaskan bahwa model yang dibuat memiliki nilai akurasi 0.90 atau 90%. Maka, MonthlyIncome dipengarui 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.

Proses evaluasi selanjutnya adalah menggunakan model OLS untuk mengevaluasi terhadap analisis model dan kinerja metode. Tahap pertama untuk membuat model OLS adalah membuat variabel x atau variabel independen yaitu Age, JobLevel, TotalWorkingYears dan YearsAtCompany.

|  |
| --- |
| 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()) |

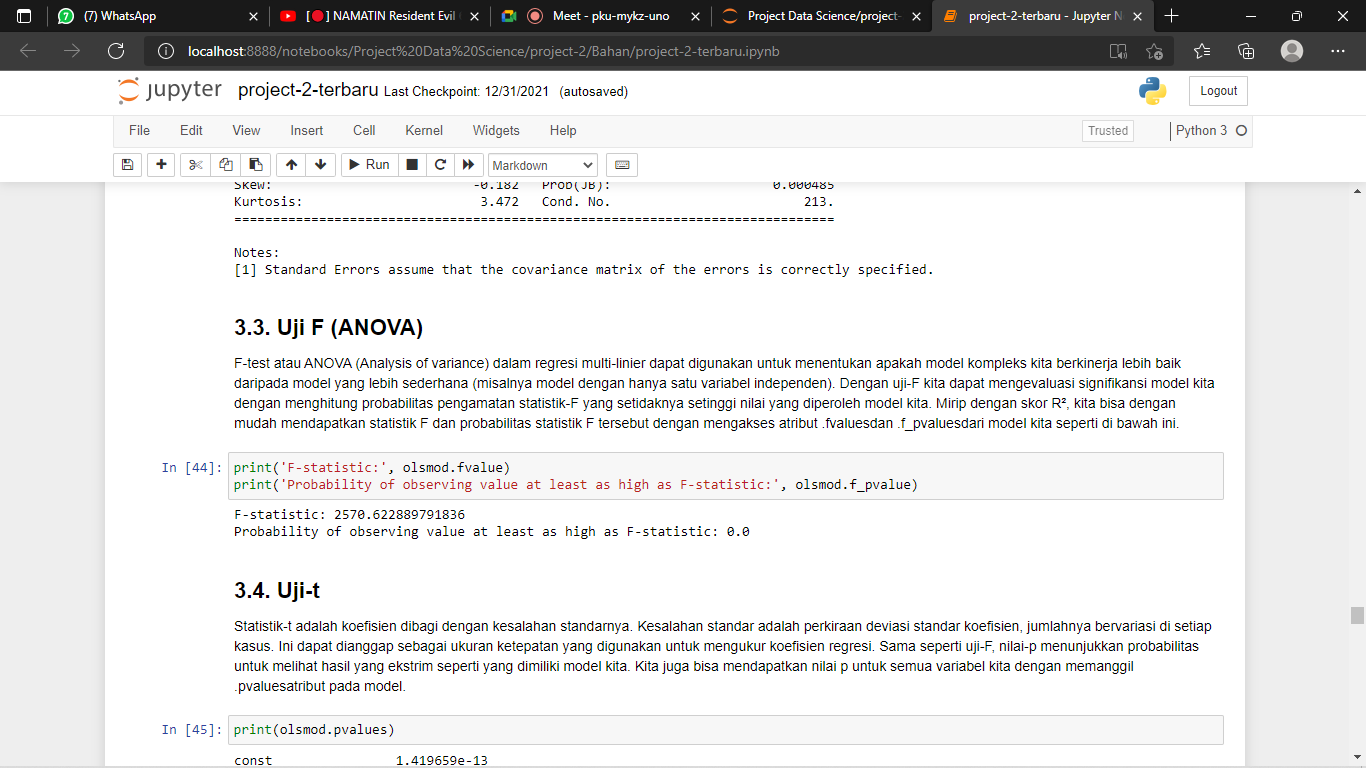


Gambar 5. 5 Evaluasi Model OLS

Berdasarkan Gambar 5.5, dapat dilihat hasil evaluasi model dan kinerja metode, dimana R-Square memperoleh nilai akurasi sebesar 0.909 atau 90%. F-statistic memperoleh nilai sebesar 2571. Kemudian untuk probablitas F-statistic memperoleh nilai 0,00 atau dibawah 0.05 hal ini menandakan bahwa model signifikan dalam melakukan prediksi variabel dependen (MonthlyIncome).

Langkah evaluasi selanjutnya ialah Uji F atau uji Model/Uji Anova yaitu uji untuk melihat bagaimana pengaruh semua variabel bebasnya secara bersama-sama terhadap variabel terikatnya. Dengan uij-F dapat dievaluasi signifikansi model yang dibuat dengan menghitung probabilitas pengamatan statistic-F yang setidaknya setinggi nilai yang diperoleh model yang dibuat. Mirip dengan skor R2, dapat dengan mudah mendapatkan statistic F dan probabilitas statistic F tersebut dengan mengakses atribut .fvalues dan .f\_pvalues dari model yang dibuat.

|  |
| --- |
| print('F-statistic:', olsmod.fvalue)  print('Probability of observing value at least as high as F-statistic:', olsmod.f\_pvalue) |



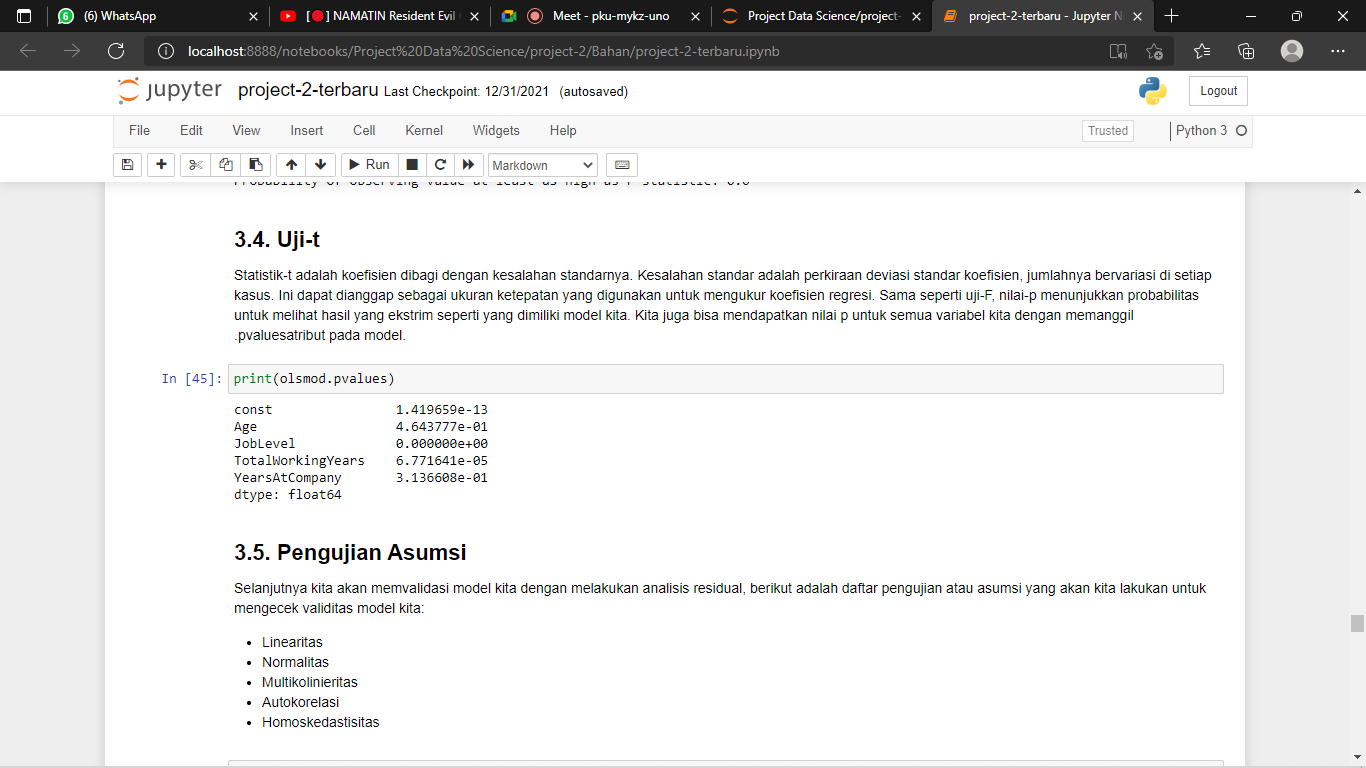
Gambar 5. 6 Nilai F-statistic dan Probabilitas

Fs adalah hasil akhir dari analisis ANOVA. Nilai Fs ini yang dikenal dengan F hitung dalam pengujian hipotesa dibandingkan dengan nilai p-value. Jika Fs > P-value, maka dapat dinyatakan bahwa secara simultan (bersama-sama) variabel dependen dan variabel independen berpengaruh signifikan terhadap permintaan. Hipotesa yang didapat dari tabel ANOVA di atas adalah :

1. H0 = Variabel independen secara simultan bukan penjelas yg signifikan terhadap variabel dependen (Model tidak cocok).
2. H1 = Variabel independen secara simultan merupakan penjelas yang siginifikan terhadap variabel dependen (Model cocok).

Berdasarkan gambar 5.6, 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.

|  |
| --- |
| print(olsmod.pvalues) |



Gambar 5. 7 Nilai P-Values Dari Variabel Independen

Uji-t dikenal dengan uji parsial adalah uji yang digunakan untuk menguji bagaimana pengaruh masing-masing variable bebasnya secara sendiri-sendiri terhadap variable terikatnya. Sama seperti uji-F, nilai-p menunjukkan probabilitas untuk melihat hasil yang ekstrim seperti yang dimiliki oleh model yang dibuat. Selain itu, uji-t juga bisa mendapatkan nilai p untuk seluruh variabel dengan memanggil .pvalues atribut pada model.

Hipotesa yang dapat diambil adalah :

1. H0 = Variabel independen tidak berpengaruh signifikan
2. H1 = Variabel independen berpengaruh signifikan.
3. α = 0,05 (Taraf signifikansi)

Berdasarkan uji-t pada gambar 5.7, dapat diambil hipotesa sebagai berikut :

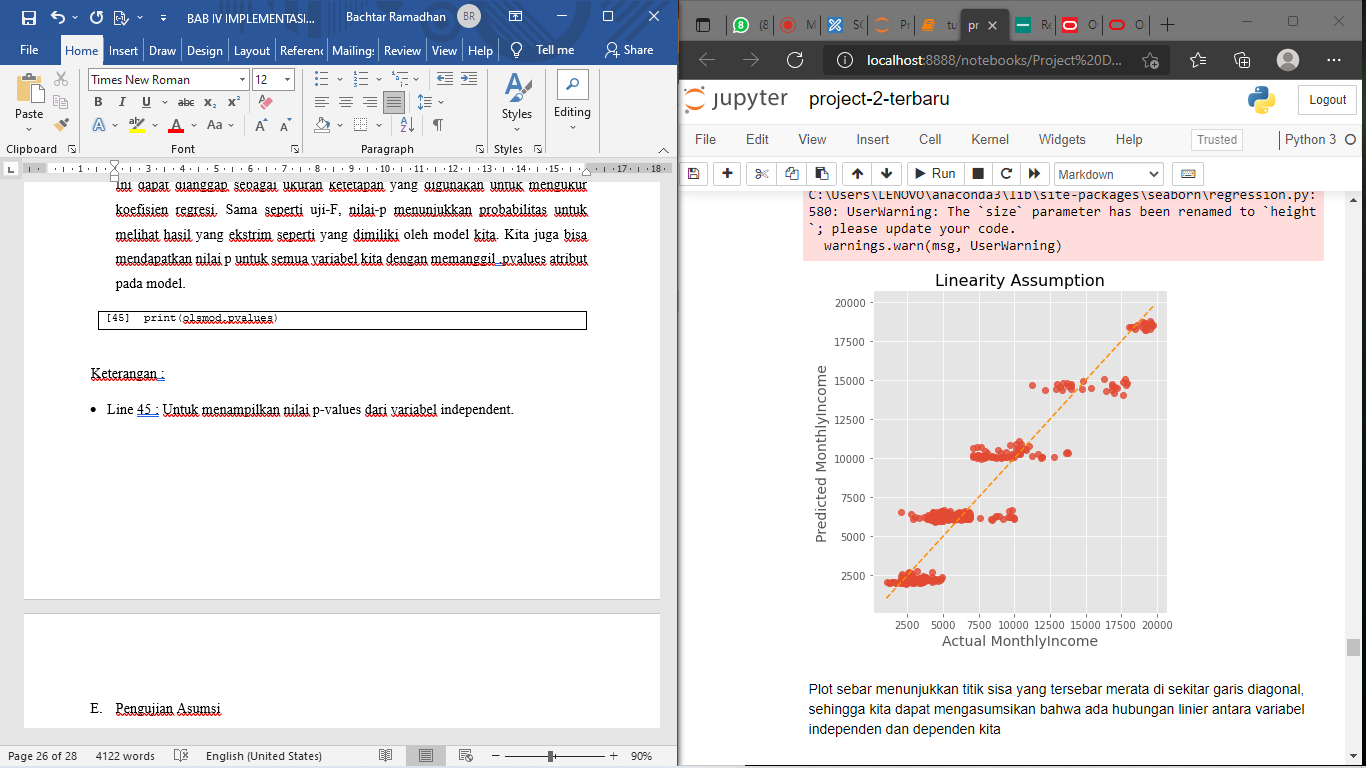
1. Nilai variabel X1 (Age) berada di atas taraf signifikansi yang berarti terima H1.
2. Nilai variabel X2 (JobLevel) berada di bawah taraf signifikansi yang berarti terima H0.
3. Nilai variabel X3 (TotalWorkingYears) di bawah taraf signifikansi yang berarti terima H0.
4. Nilai variabel X4 (YearsAtCompany) di atas taraf signifikansi yang berarti terima H1.

Berdasarkan hipotesa di atas, dapat disimpulan variabel independen JobLevel dan TotalWorkingYears adalah variabel yang tidak mempengaruhi variabel dependen. Sedangkan variabel independen Age dan YearsAtCompany adalah variabel yang mempengaruhi variabel dependen.

|  |
| --- |
| df\_test\_new2['MonthlyIncome Prediction'] = olsmod.predict(X)  df\_test\_new2['residual'] = olsmod.resid  df\_test\_new2 |

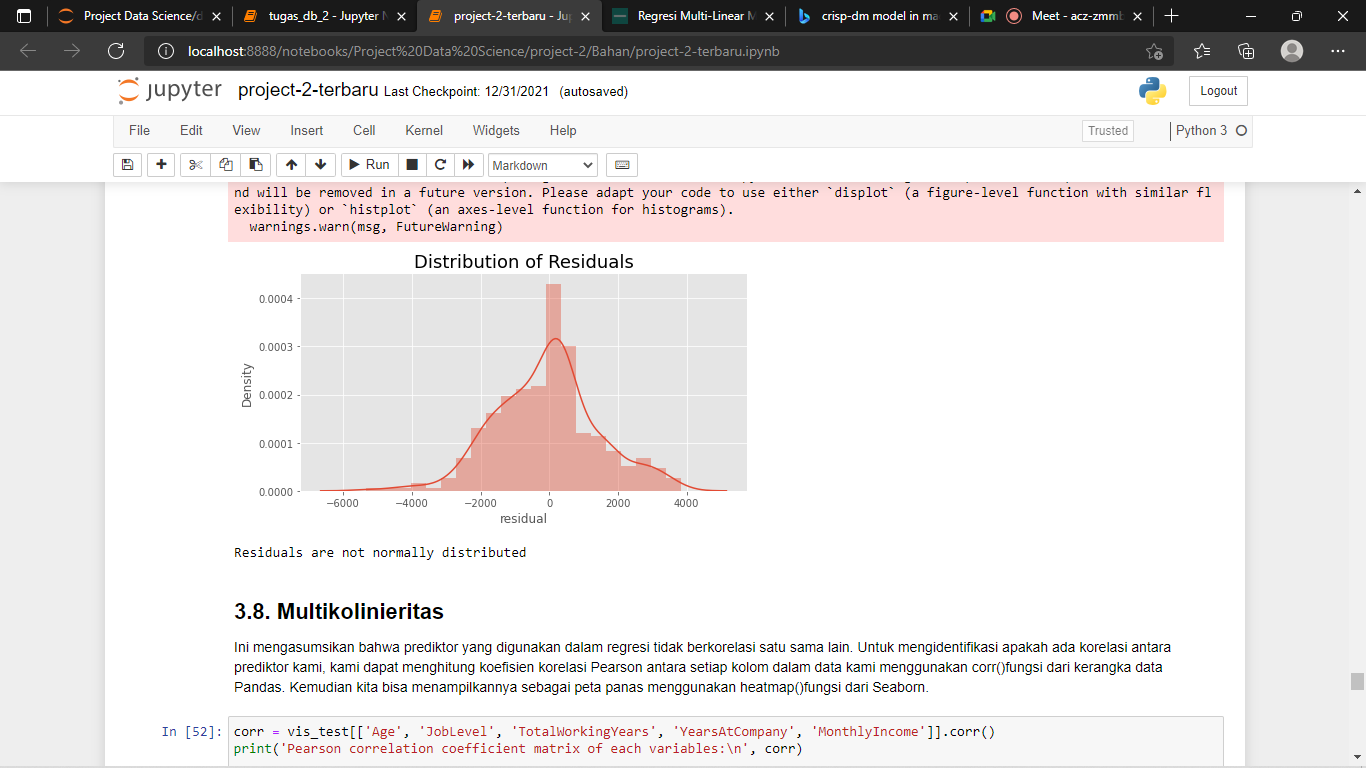
Selanjutnya adalah menambahkan satu kolom baru dengan nama MonthlyIncome Prediction yang berisikan dengan hasil prediksi. Kemudian dibuat juga kolom residual untuk menyimpan nilai residualnya.

Setelah itu, dilakukan pengujian asumsi untuk model yang dibuat. Uji asumsi pertama yang dilakukan adalah uji linearitas. Yaitu pengujian untuk memeriksa apakah terdapat hubungan linear antara variabel independen dengan variabel dependen. Uji linearitas dilakukan dengan menggunakan plot pencar sehingga dapat diketahui perbandingan nilai prediksi dengan nilai sebenarnya.



Gambar 5. 8 Grafik Asumsi Linear

Setelah uji linearitas, selanjutnya dilakukan uji normalitas. Uji Normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variable, apakah sebaran data tersebut berdistribusi normal ataukah tidak. Memeriksa normalitas residual dengan memplotnya ke dalam histogram dan melihat nilai p dan uji normalitas Anderson-Darling dengan menggunakan normal\_ad( ) fungsi dari statsmodel untuk menghitung p-value dan kemudian membandingkannya dengan threshold 0,05. Jika p-value yang diperoleh lebih tinggi dari threshold maka dapat diasumsikan bahwa residual terdistribusi normal. Jika p-value yang diperoleh lebih kecil dari threshold, maka dapat diasumsikan bahwa residual terdistribusi tidak normal.



Gambar 5. 9 Grafik Distribusi Residual

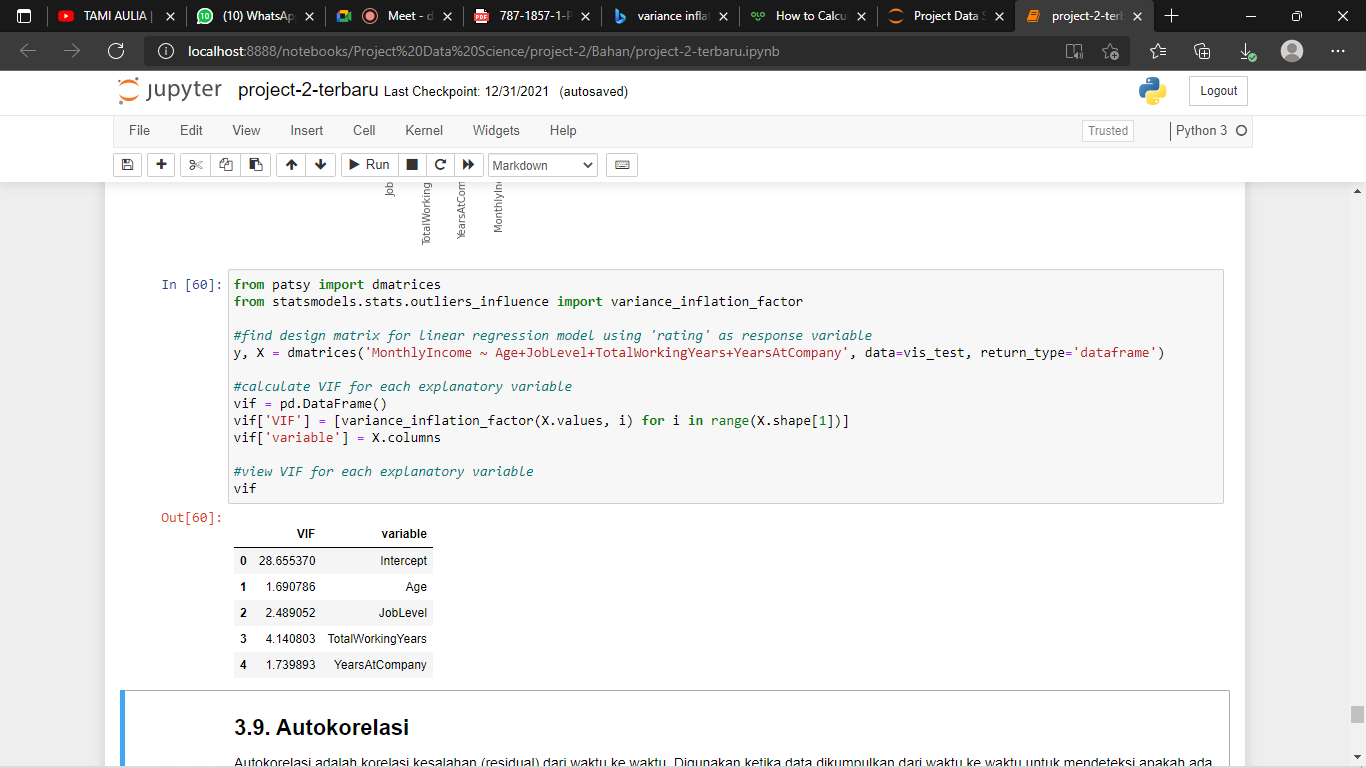
Dapat diketahui hipotesa sebagai berikut :

1. H0 = Residual terdistribusi normal.
2. H1 = Residual terdistribusi secara tidak normal.

Berdasarkan gambar 5.9, 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 tolah H0 terima H1 atau dapat dikatakan residual terdistribusi secara tidak normal. Sehingga disimpulkan asumsi normalitas terpenuhi.

Uji multikolinearitas dilakukan untuk menguji apakah terdapat korelasi antar variable bebas dalam model regresi. Multikolinearitas menguji apakah terdapat linier yang sempurna antara beberapa atau semua variable yang menjelaskan model regresi. Pengujian ini dapat diketahui dengan melihat nilai toleransi dan nilai variance inflation factor (VIF). Pengujian dilakukan dengan melihat nilai VIF atau variance inflation factors. Apabila nilai centered VIF (Variance Inflation Factor). Pengujian dapat dilakukan dengan melihat nilai Tolerance dan Variance Inflation Factor (VIF) pada model regresi. Kriteria pengambilan keputusan terkait uji multikolinearitas adalah sebagai berikut:

1. Jika nilai VIF < 10 atau nilai Tolerance > 0,01, maka dinyatakan tidak terjadi multikolinearitas.
2. Jika nilai VIF > 10 atau nilai Tolerance < 0,01, maka dinyatakan terjadi multikolinearitas.
3. Jika koefisien korelasi masing-masing variabel bebas > 0,8 maka terjadi multikolinearitas. Tetapi jika koefisien korelasi masing-masing variabel bebas < 0,8 maka tidak terjadi multikolinearitas.

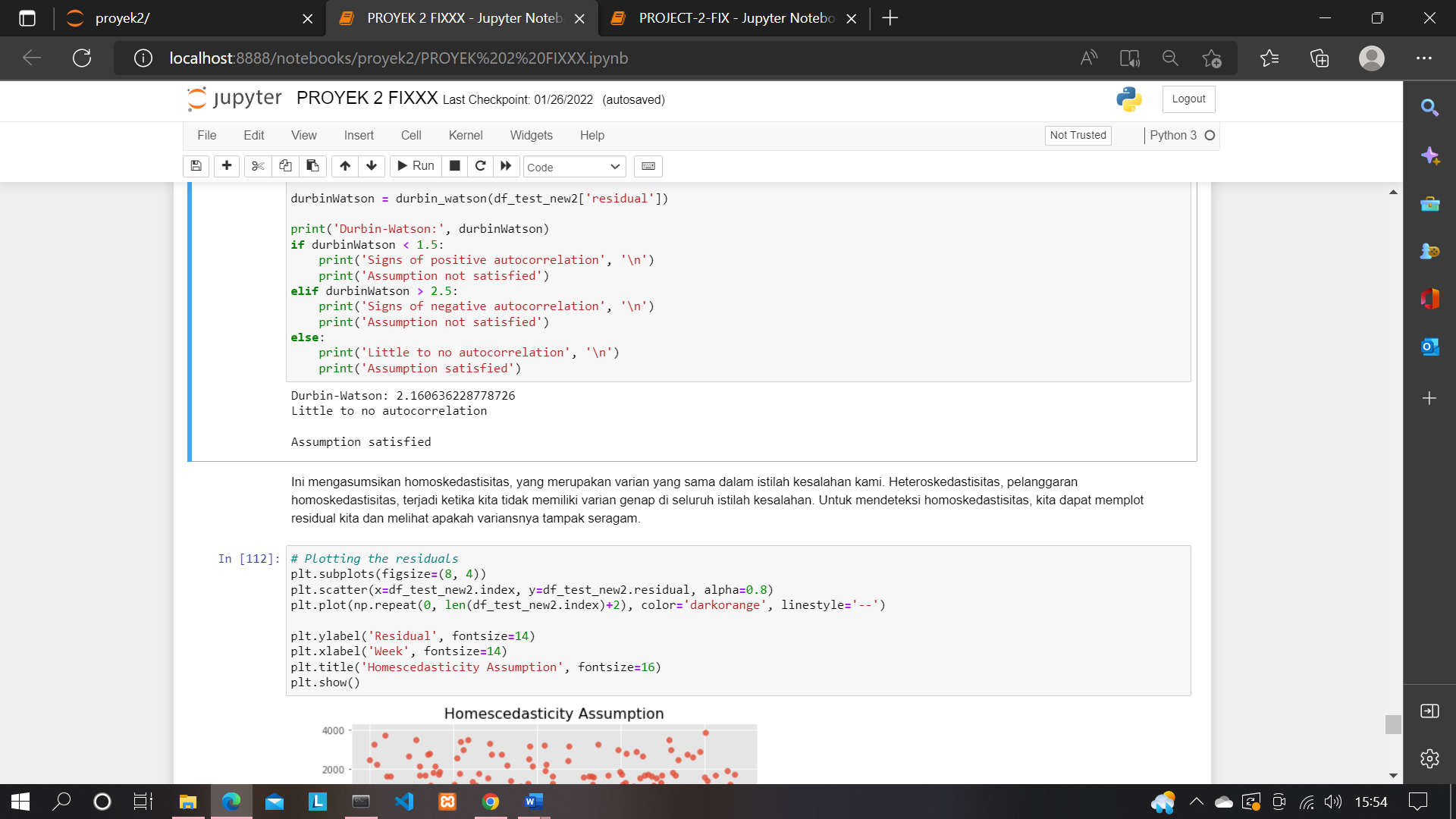


Gambar 5. 10 Tabel VIF

Berdasarkan gambar 5.10 di atas dapat dilihat nilai variabel Age, JobLevel, TotalWorkingYears, YearsAtCompany memiliki nilia kurang dari 10 sehingga dengan menggunakan tingkat signifikansi sebesar 0,05 dapat disimpulkan bahwa pada data tersebut tidak terdapat multikolinearitas pada variabel-variabel prediktor.

Uji asumsi selanjutnya adalah pengujian autokorelasi. Autokorelasi adalah korelasi kesalahan (residual) dari waktu ke waktu. Digunakan ketika data dikumpulkan dari waktu ke waktu untuk mendeteksi apakah ada autokorelasi. Autokorelasi dapat dideteksi dengan melakukan uij Durbin-Watson untuk menentukan apakah ada korelasi positif atau negative. Pada langkah ini akan dilakukan perhitungan skor Durbin-Watson menggunakan durbin\_watson( ) fungsi dari statsmodel yang dibuat, kemudian menilainya dengan kondisi sebagai berikut:

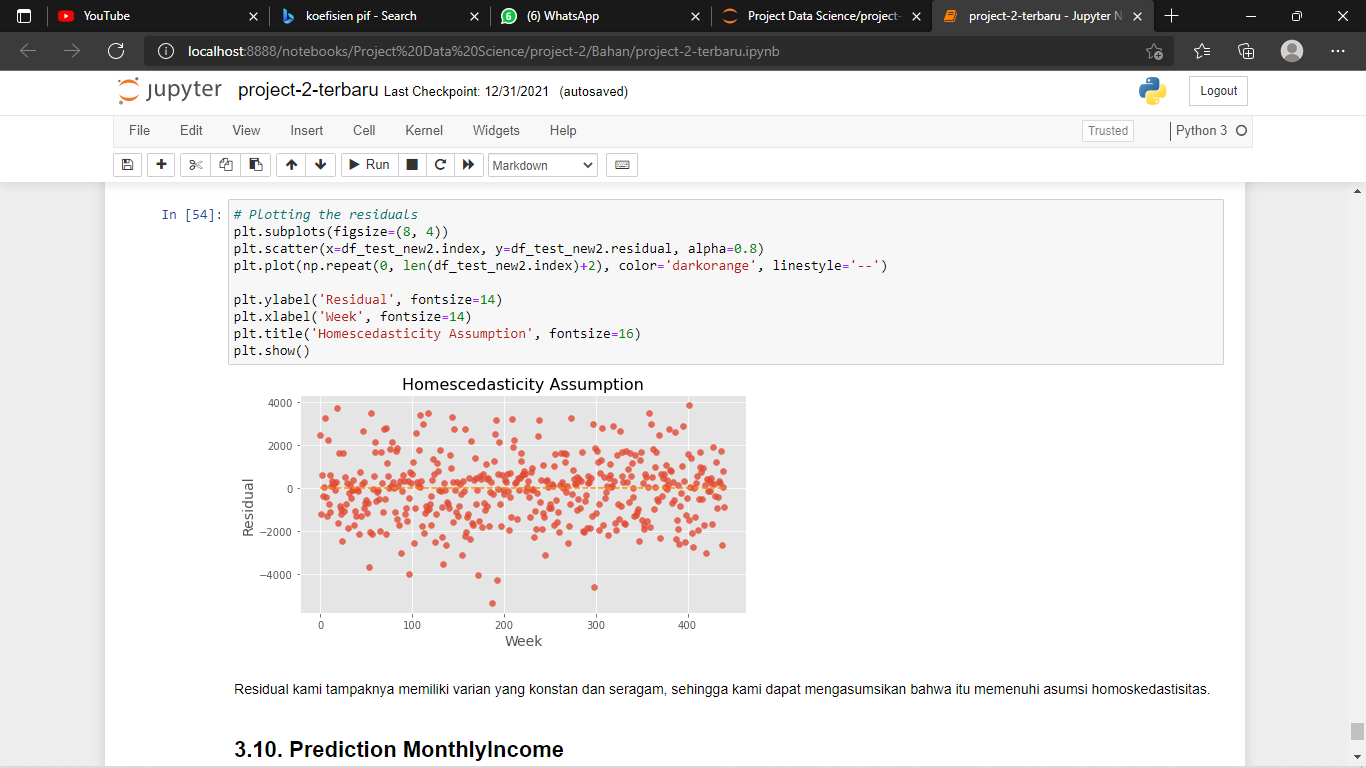
1. Jika skor Durbin-Watson kurang dari 1,5 maka terdapat autokorelasi positif dan asumsi tidak terpenuhi.
2. Jika skor Durbin-Watson antara 1,5 – 2,5 maka tidak ada autokorelasi dan asumsi puas.
3. Jika skor Durbin-Watson lebih dari 2,5 maka terdapat autokorelasi negatif dan asumsi tidak puas.



Gambar 5. 11 Skor Durbin-Watson Dan Hasil Asumsi

Berdasarkan gambar 5.11, didapat hasil perhitungan skor Durbin-Watson sebesar 2,160636228. Dapat diasumsikan bahwa terdapat sedikit atau tidak ada autokorelasi, sehingga asumsi yang didapatkan memiliki hasil puas.

Uji homoskedastisitas digunakan dalam menguji error dalam model statistik untuk melihat apakah varians atau keragaman dari error terpengaruh oleh faktor lain atau tidak. Untuk mendeteksi homoskedastisitas, dapat dilakukan berupa memplot residual dan melihat apakah variansnya tampak seragam.

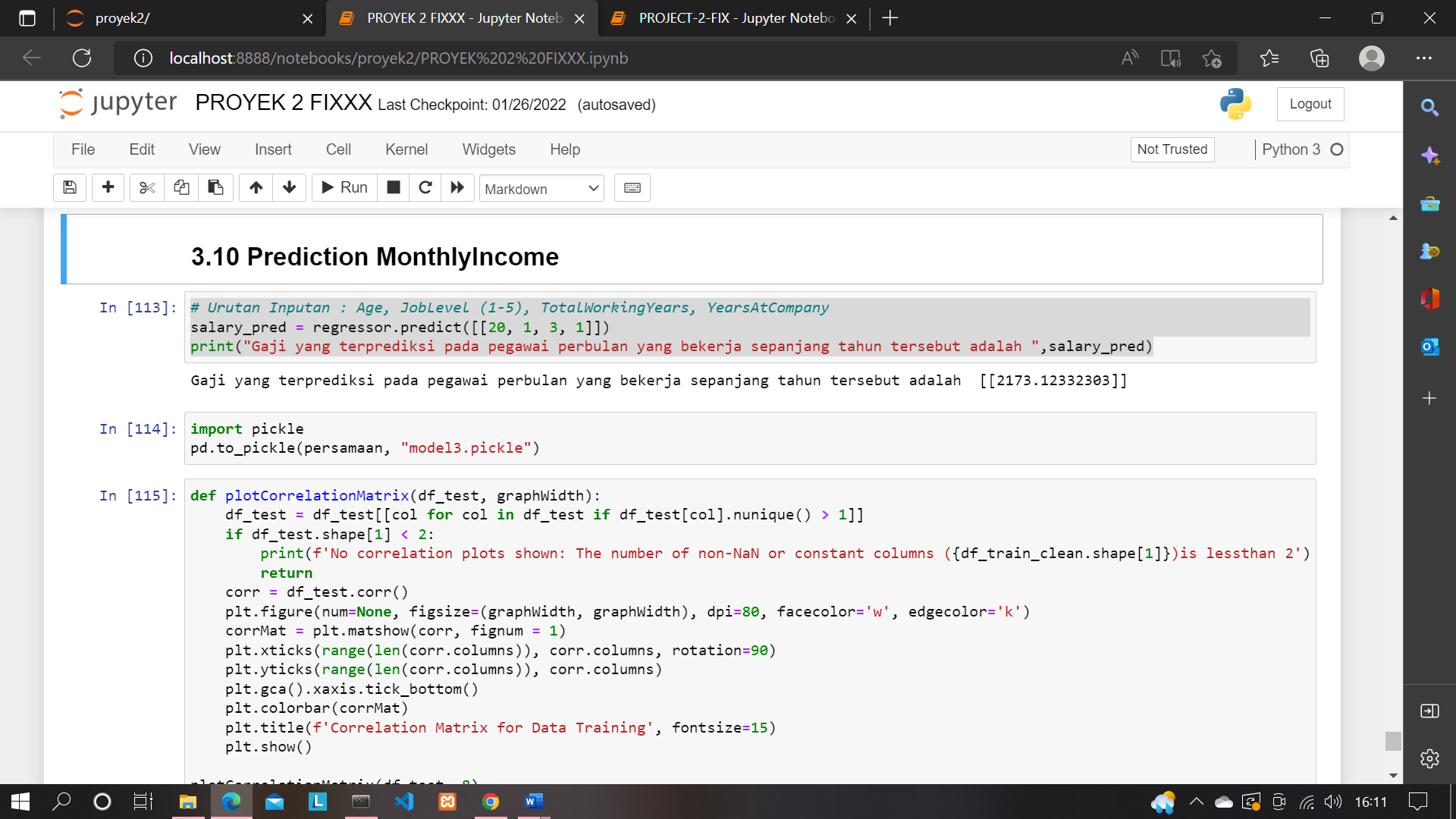


Gambar 5. 12 Homoskedastisitas

Dari gambar 5.12 berupa grafik scatterplot, terlihat titik-titik menyebar secara acak, serta tersebar baik di atas maupun di bawah angka 0 (nol) pada sumbu Y. Maka dapat diambil kesimpulan bahwa tidak terdaat gejala heteroskedastisitas pada model regresi yang digunakan.

Langkah terakhir setelah berbagai pemodelan dan pengujian adalah melakukan prediksi gaji dengan parameter Age = 20 tahun, JobLevel = 1, TotalWorkingYears = 3 tahun, dan YearsAtCompany = 1 tahun menggunakan metode pada variabel regressor, kemudian simpan di variabel MonthlyIncome. Menampilkan nilai dari variabel MonthlyIncome.

|  |
| --- |
| # Urutan Inputan : Age, JobLevel (1-5), TotalWorkingYears, YearsAtCompany  salary\_pred = regressor.predict([[20, 1, 3, 1]])  print("Gaji yang terprediksi pada pegawai perbulan yang bekerja sepanjang tahun tersebut adalah ",salary\_pred) |



Gambar 5. 13 Hasil Prediksi Gaji

# DAFTAR PUSTAKA

[1] B. Prasetyo and U. Trisyanti, “REVOLUSI INDUSTRI 4.0 DAN TANTANGAN PERUBAHAN SOSIAL.”

[2] H. Prasetyo and W. Sutopo, “Perkembangan Keilmuan Teknik Industri Menuju Era Industri 4.0,” Surakarta, May 2017.

[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, 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 *ACM International Conference Proceeding Series*, 2019, vol. Part F148155, pp. 244–253. doi: 10.1145/3326365.3326398.

[7] Y. Adrianova Eka Tuah, P. Studi Pendidikan Komputer, and S. Persada Khatulistiwa Sintang, “IMPLEMENTASI MODEL REGRESI LINEAR SEDERHANA UNTUK PREDIKSI GAJI BERDASARKAN PENGALAMAN LAMA BEKERJA,” 2020.

[8] G. Nicora, M. Rios, A. Abu-Hanna, and R. Bellazzi, “Evaluating pointwise reliability of machine learning prediction,” *Journal of Biomedical Informatics*, vol. 127, Mar. 2022, doi: 10.1016/j.jbi.2022.103996.

[9] Q. Ke and K. Zhang, “Interaction effects of rainfall and soil factors on runoff, erosion, and their predictions in different geographic regions,” *Journal of Hydrology*, vol. 605, Feb. 2022, doi: 10.1016/j.jhydrol.2021.127291.

[10] W. Xu, B. Wang, J. Liu, Y. Chen, P. Duan, and Z. Hong, “Toward practical privacy-preserving linear regression,” *Information Sciences*, vol. 596, pp. 119–136, Jun. 2022, doi: 10.1016/j.ins.2022.03.023.

[11] Y. Fujikoshi, “High-dimensional consistencies of KOO methods in multivariate regression model and discriminant analysis,” *Journal of Multivariate Analysis*, vol. 188, Mar. 2022, doi: 10.1016/j.jmva.2021.104860.

[12] 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,” *Topics in Companion Animal Medicine*, vol. 40, Aug. 2020, doi: 10.1016/j.tcam.2020.100428.

[13] 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.

[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!,” *Economics of Education Review*, 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,” *Tourism Management Perspectives*, 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,” *Economics of Education Review*, 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,” *Economics of Education Review*, 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,” *International Journal of Environmental Research and 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,” *International Journal of Computer Applications*, 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,” *Journal of 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,” *Computer Methods and Programs in Biomedicine*, 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] C. Imam, Sutrisno, A. S. Arief, H. Uswatun, and I. F. Yessica, *AI, MACHINE LEARNING  & DEEP LEARNING  (Teori & Implementasi)*. 2020. [Online]. Available: http://bit.ly/3piOnnU

[36] N. Al-Azzam and I. Shatnawi, “Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer,” *Annals of Medicine and Surgery*, vol. 62, pp. 53–64, Feb. 2021, doi: 10.1016/j.amsu.2020.12.043.

[37] Y. Herlambang Ngumar, “Aplikasi Metode Numerik Dan Matrik Dalam Perhitungan Koefisien-Koefisien Regresi Linier Multiple Untuk Peramalan,” *Konferensi Nasional Sistem dan Informatika*, pp. 157–168, 2008, Accessed: Aug. 14, 2022. [Online]. Available: https://yudiagusta.files.wordpress.com/2009/11/157-162-knsi08-029-aplikasi-metode-numerik-dan-matrik-dalam-perhitungan-koefisien-koefisien-regresi-linier-multiple-untuk-peramalan.pdf

[38] Y. Aditya, “Random Forest,” *Universitas Gadjah Mada Menara Ilmu Machine Learning*, Jul. 20, 2018.

[39] E. H. Briliant, M. Hasan, and S. Kurniawan, “Perbandingan Regresi Linier Berganda dan Regresi Buckley-James Pada Analisis Survival Data Tersensor Kanan,” in *PROCEEDINGS OF THE 1 st STEEEM*, 2019, vol. 1, no. 1, pp. 1–19. Accessed: Aug. 14, 2022. [Online]. Available: http://seminar.uad.ac.id/index.php/STEEEM/article/view/3349/721

[40] A. Afifah Muhartini *et al.*, “Analisis Peramalan Jumlah Penerimaan Mahasiswa Baru Dengan Menggunakan Metode Regresi Linear Sederhana,” *Jurnal Bayesian*, vol. 1, no. 1, pp. 17–23, 2021, [Online]. Available: http://bayesian.lppmbinabangsa.id/index.php/home

[41] Greatlearning Blog, “What is Quantile Regression? | Introduction to Quantile Regression,” *Greatlearning blog*, Jul. 16, 2020. https://www.mygreatlearning.com/blog/what-is-quantile-regression/ (accessed Aug. 14, 2022).

[42] Khoiri, “Pengertian dan Cara Menghitung Root Mean Square Error (RMSE),” *khoiri.com*, Dec. 23, 2020. https://www.khoiri.com/2020/12/cara-menghitung-root-mean-square-error-rmse.html (accessed Aug. 14, 2022).

[43] K. Abdul Muiz, “Cara Hitung RMSE , MSE, MAPE, dan MAE Dengan Excel,” *Pengalaman Edukasi*, Jan. 2021. https://www.pengalaman-edukasi.com/2021/01/cara-menghitung-rmse-root-mean-square.html (accessed Aug. 14, 2022).

[44] 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.

# LAMPIRAN-LAMPIRAN

Lampiran 1 Susunan Organisasi Tim Peneliti dan Pembagian Tugas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Nama/NIDN** | **Asal Prodi** | **Bidang Ilmu** | **Alokasi Waktu (jam/minggu)** | **Uraian Tugas** |
| 1. | Syafrial Fachri Pane 0416048803 | D4 Teknik Informatika | Machine Learning, Data Science, Big Data | 16 Minggu | Menentukan pendekatan machine learning yang digunakan |
| 2. | Amri yanuar 0412018603 | D4 Logistik Bisnis | Logistik | 16 Minggu | Roadmap tinjauan pustaka |
| 3. | Bachtiar Ramadhan | D4 Teknik Informatika | Machine Learning | 16 Minggu | Pemograman |
| 4. | Nur Tri Ramadhanti Adiningrum | D4 Teknik Informatika | Machine Learning | 16 Minggu | Analisi dan pengumpulan data |
| 5. | M. Rizky | D4 Teknik Informatika | Machine Learning | 16 Minggu | Data Engineer |

Lampiran 2 Biodata Ketua dan Anggota Pengusul

1. Identitas Diri Ketua Peneliti

|  |  |  |
| --- | --- | --- |
| 1 | Nama Diri | Syafrial Fachri Pane, S.T.,M.T.I.,EBDP |
| 2 | Jenis Kelamin | Laki – Laki |
| 3 | Program Studi | Teknik Informatika |
| 4 | NIDN/NUPN | 0416048803 |
| 5 | Tempat dan Tanggal Lahir | Medan, 16 April 1988 |
| 6 | Alamat Email | syafrial.fachri@poltekpos.ac.id |
| 7 | Nomor Telepon/HP | 085362383988 |

1. Riwayat Pendidikan

|  |  |  |  |
| --- | --- | --- | --- |
|  | **D-3** | **S-1** | **S-2** |
| Nama Perguruan Tinggi | Politeknik Pos Indonesia | Universitas Pasundan | Universitas Bina Nusantara |
| Bidang Ilmu & Tahun Lulus | Teknik Informatika (Lulus 2009) | Teknik Informatika (Lulus 2013) | Teknik Informatika (Lulus 2017) |
| **IPK** | **3.76** | **3.60** | **3.67** |
| **Penghargaan** | ***Cumloude*** | ***Cumloude*** | ***Cumloude*** |

1. Rekam Jejak Tri Dharma PT Pendidikan/Pengajaran

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Semester** | **Kode Mata Kuliah** | **Nama Mata Kuliah** |
| 1. | Ganjil 2013 | T4I322AG4 | Basis Data II/Database II |
| 2. | Genap 2013 | L3452S2 | Fundamental SAP |
| 3. | Genap 2013 | P3M222D3 | Basis Data |
| 4. | Genap 2013 | T4I222D4 | Basis Data I/Database I |
| 5. | Genap 2013 | T4I722AF4 | Sistem Pendukung Keputusan / DSS |
| 6. | Ganjil 2014 | T4I322AG4 | Basis Data II/Database II |
| 7. | Ganjil 2014 | T4T4I611C | Internship 1 |
| 8. | Ganjil 2014 | A4K733Y3 | Manajemen Accounting II berbasis SAP |
| 9. | Ganjil 2014 | D4L352C3 | Database + PRKT |
| 10. | Genap 2014 | T4I722AF2 | Sistem Pendukung Keputusan |
| 11. | Genap 2014 | T4I222D4 | Basis Data I |
| 12. | Ganjil 2015 | T4I162D2 | Matematika Diskrit |
| 13. | Ganjil 2015 | T4I322AG4 | Basis Data II/Database II |
| 14. | Ganjil 2015 | A4K733Y3 | Manajemen Accounting II berbasis SAP |
| 15. | Genap 2015 | M4P422D3 | Basis Data |
| 16. | Genap 2015 | T4I222D4 | Basis Data I |
| 17. | Ganjil 2016 | D4L352C3 | Database + Praktek |
| 18. | Ganjil 2016 | T4I322AG4 | Basis Data II/Database II |
| 19. | Genap 2016 | T4I222D4 | Basis Data I |
| 20. | Ganjil 2017 | T4I322AG4 | Basis Data II/Database II |
| 21. | Genap 2017 | T4I222D4 | Basis Data I |
| 22. | Ganjil 2018 | T4I322AG4 | Basis Data II/Database II |
| 23. | Genap 2018 | T4I222D4 | Basis Data I |

1. Penelitian

|  |  |  |  |
| --- | --- | --- | --- |
| No | Judul Penelitian | Penyandang Dana | Tahun |
| 1 | Perancangan Sistem Penerimaan Mahasiswa Baru Berbasis Website | Politeknik Pos Indonesia | 2014 |
| 2 | Analisis kinerja proses bisnis dengan pendekatan BPMN menggunakan Bizagi | Politeknik Pos Indonesia | 2015 |
| 3 | Mengevaluasi Pengelolaan Dan Perencanaan Investasi Teknologi Informasi dari Sumber Dana Hibah Pemerintah Untuk Perguruan Tinggi Swasta Menggunakan Cobit 5 | Politeknik Pos Indonesia | 2016 |
| 4 | Prototype RFID Conveyor Belt Pada Warehouse Management System Berbasis IoT | Politeknik Pos Indonesia | 2018 |
| 5 | Perancangan Aplikasi E-Recruitment Beasiswa Mahasiswa/I kurang mampu dengan penerapan Teknologi Geispatial Intelligence dan Webservice (OAUTH) menggunakan metode Electre Berbasis Webs | Politeknik Pos Indonesia | 2017 |
| 6 | PROFIT-WMS Prototype RFID Conveyor Belt pada Warehouse Management System Berbasis IoT | Politeknik Pos Indonesia | 2018 |
| 7 | Simulasi Auto Turn Sign Pengantar Pos Menggunakan Aktifitas Gelombang Otak Dengan Metode Bayesian Learning dan Logistic Regression | Politeknik Pos Indonesia | 2019 |
| 8 | Perancangan Simulasi Warehouse Management System (Wms) Berbasis Internet Of Things Pada Center Of Technology | Politeknik Pos Indonesia | 2019 |
| 9 | Straglog : Analisis Strategi Pengadaan Barang dan Jasa Menggunakan Algoritma Heuristic Miner | Politeknik Pos Indonesia | 2020 |
| 10 | Qualitative Evaluation of RFID Implementationon Warehouse Management System | Jurnal Telkomnika – Jilid 16 Terbitan 3 (International) Akreditasi Dikti A dan Terindex Scopus | 2018 |
| 11 | K Means Clustering and Meanshift Analysis for Grouping the Data of Coal Term in Puslitbang tekMIRA | Jurnal Telkomnika – Jilid 16 Terbitan 3 (International) Akreditasi Dikti A dan Terindex Scopus | 2018 |
| 12 | Sireuboh-klasifikasi data lokasi barang menggunakan region of interest (roi) dan algoritma ransac | Jurnal Nasional  Tekno Insentif LLDIKTI IV | 2018 |
| 13 | Implementation of web scraping on github task monitoring system | Jurnal Telkomnika – Jilid 17 Terbitan 1 (International) Akreditasi Dikti A dan Terindex Scopus | 2019 |
| 14 | Ontology Design of Family Planning Field Officer for Family Planning Agency Using  OWL and RDF | Jurnal Telkomnika – Jilid 17 Terbitan 1 (International) Akreditasi Dikti A dan Terindex Scopus | 2019 |
| 15 | RFID-based conveyor belt for improve warehouse operations | Jurnal Telkomnika – Jilid 17 Terbitan 2 (International) Akreditasi Dikti A dan Terindex Scopus | 2019 |
| 16 | Implementasi algoritma genetika untuk optimalisasi pelayanan kependudukan | Jurnal Nasional  Tekno Insentif LLDIKTI IV  Terindex DOAI | 2019 |
| 17 | Collaboration Fmadm And K-Means Clustering To Determine The Activity Proposal In Operational Management Activity | Jurnal Emiter – EMITTER International Journal of Engineering Technology Terindex Scopus | 2019 |
| 18 | Implementasi Algoritma Genetika Untuk Optimalisasi Pelayanan Kependudukan | Jurnal Tekno Insentif 13 (2), 36-43 | 2019 |
| 19 | MILA: Low-cost BCI framework for acquiring EEG data with IoT | Telkomnika 18 (2), 846-852 | 2020 |
| 20 | Ovmp: Operational sVehicle Management Application Using Extreme Programming (Xp) Method | Jurnal Tekno Insentif 14 (1), 9-16 | 2020 |
| 21 | Sistem Informasi Absensi Pegawai Menggunakan Metode RAD dan Metode LBS Pada Koordinat Absensi | Jurnal Media Informatika Budidarma 4 (1),  59-64 | 2020 |
| 22 | AMCF: A Novel Archive Modeling Based on Data Cluster and Filtering | Technomedia Journal 4 (2), 139-152 | 2020 |

1. Pengabdian Kepada Masyarakat

|  |  |  |  |
| --- | --- | --- | --- |
| No | Judul Pengabdian kepada Masyarakat | Penyandang Dana | Tahun |
| 1 | Pelatihan Penyusunan Proposal Penelitian Tindakan Kelas | SD Panorama Bandung | 2015 |
| 2 | Pelatihan Penyusunan Laporan Penelitian Tindakan Kelas | SD Panorama Bandung | 2016 |
| 3 | Pelatihan Publikasi Penelitian Tindakan Kelas | SD Panorama Bandung | 2017 |
| 4 | Pelatihan Pemanfaatan Google Map | Desa Wangunharja Lembang | 2019 |

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Internal Tahun 2022

Bandung, 14 Maret 2022

 Ketua Peneliti,

(Syafrial Fachri Pane,S.T.,M.T.I.,EBDP)

1. Identitas Diri Anggota Peneliti

|  |  |  |
| --- | --- | --- |
| 1 | Nama Lengkap (dengan gelar) | Amri Yanuar, ST.,M.MT |
| 2 | Jenis Kelamin | Laki-laki |
| 3 | Jabatan Fungsional Akademik | Asisten Ahli |
| 4 | NIK | 116.86.207 |
| 5 | NIDN | 0412018603 |
| 6 | Tempat dan Tanggal Lahir | Bandung, 12 Januari 1986 |
| 7 | E-mail | [amriyanuar@poltekpos.ac.id](mailto:amriyanuar@poltekpos.ac.id) |
| 8 | Nomor Telepon/HP | 081910027205 |
| 9 | Alamat Kantor | Jl. Sari Asih No. 54 Bandung |
| 10 | Nomor Telepon kantor |  |
| 11 | Lulusan yang telah dihasilkan | D4 = 9 Orang |
| 12 | Mata kuliah yang diampu | 1. Manajemen Persediaan |
| 2. Manajemen pergudangan |

1. Riwayat Pendidikan

|  |  |  |
| --- | --- | --- |
|  | S1 | S2 |
| Nama Perguruan Tinggi | Universitas Pasundan | Universiti Teknologi Malaysia |
| Bidang Ilmu | Teknik Industri | Management of Technology |
| Tahun Masuk/Tahun Lulus | 2004/2009 | 2011/2013 |
| Judul Skripsi/Tesis | Implementasi MFG/Pro di PT Pindad Persero | Implementation of Inventory Management in SMEs Metal Industries Bandung |
| Nama Pembimbing/Promotor | Putri Mety Zalynda, ST.,MT | Dr. Low Hock Heng |

1. Pengalaman Penelitian dalam 5 tahun Terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Tahun | Judul Penelitian | Pendanaan | |
| Sumber | Jml (Juta Rp) |
| 1 | 2015 | Penerapan Software ERP SAP di PT Purinusa Eka Persada | LPPM Politeknik Pos | 5.000.000 |
| 2 | 2017 | Perancangan Kebutuhan Persediaan Untuk Barang Penjualan Online (Studi Kasus: UKM Pelaku E-Commerce) | LPPM Politeknik Pos | 8.000.000 |
| 3 | 2018 | Formulasi Model Bisnis Surat Kabar Elektronik di Indonesia | PDP Ristekdikti | 14.500.000 |

1. Pengalaman Pengabdian Kepada Masyarakat dalam 5 tahun terakhir

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Tahun | Judul Pengabdian Kepada Masyarakat | Pendanaan | |
| Sumber | Jml (Juta Rp) |
| 1 | 2017 | Pelatihan Aplikasi Monitoring Distribusi Beras Berbasis SMS Gateway di Kecamatan Cikancung | LPPM Politeknik Pos | 7.000.000 |
| 2 | 2018 | Pelatihan Aplikasi Pos Pelayanan Keluarga Berencana dan Kesehatan Terpadu di RW 06 Kelurahan Rancaekek Kencana | LPPM Politeknik Pos | 8.000.000 |

1. Publikasi Artikel Ilmiah Dalam Jurnal dalam 5 tahun Terakhir

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Judul Artikel | Nama Jurnal | Volume/Nomor/Tahun |
| 1 | Penerapan Software ERP SAP PT Purinusa Eka Persada | Jurnal Logistik Bisnis | Vol. 6 No 2 2016 |
| 2 | Formulasi Bisnis Model Surat Kabar Elektronik di Pikiran Rakyat | Jurnal Competitive | Vol. 13 No 1 2018 |

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Internal Tahun 2022. Bandung,14 Maret 2022

Anggota Peneliti,

Amri Yanuar, ST.,M.MT

1. Identitas Diri Anggota Peneliti

|  |  |  |
| --- | --- | --- |
| 1 | Nama Diri | Bachtiar Ramadhan |
| 2 | Jenis Kelamin | Laki – Laki |
| 3 | Program Studi | DIV-Teknik Informatika |
| 4 | NIM | 1204077 |
| 5 | Tempat dan Tanggal Lahir | Pekanbaru, 21 Desember 2000 |
| 6 | Alamat Email | 1204077\_bachtiar@students.poltekpos.ac.id |
| 7 | Nomor Telepon/HP | 085213921331 |

1. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

|  |  |  |  |
| --- | --- | --- | --- |
| No | Jenis Kegiatan | Status Dalam Kegiatan | Waktu dan Tempat |
| 1 | RPPI | Aktif |  |
| 2 | LDK Commitment | Aktif |  |
| 3 | Himatif | Aktif |  |

1. Penghargaan Yang Pernah Diterima

|  |  |  |  |
| --- | --- | --- | --- |
| No | Jenis Penghargaan | Pihak Pemberi Penghargaan | Tahun |
| 1 |  |  |  |
| 2 |  |  |  |

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Internal Tahun 2022.

Bandung, 14 Maret 2022

 Anggota Peneliti,

(Bachtiar Ramadhan)

1. Identitas Diri Anggota Penelitian

|  |  |  |
| --- | --- | --- |
| 1 | Nama Diri | Nur Tri Ramadhanti Adiningrum |
| 2 | Jenis Kelamin | Perempuan |
| 3 | Program Studi | DIV-Teknik Informatika |
| 4 | NIM | 1204061 |
| 5 | Tempat dan Tanggal Lahir | Bandung,16 Desember 2001 |
| 6 | Alamat Email | 1204061\_nur@students.poltekpos.ac.id |
| 7 | Nomor Telepon/HP | 081221950983 |

1. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

|  |  |  |  |
| --- | --- | --- | --- |
| No | Jenis Kegiatan | Status Dalam Kegiatan | Waktu dan Tempat |
| 1 | Himatif | Aktif |  |
| 2 | Composer | Aktif |  |
| 3 | Popeys | Aktif |  |

1. Penghargaan Yang Pernah Diterima

|  |  |  |  |
| --- | --- | --- | --- |
| No | Jenis Penghargaan | Pihak Pemberi Penghargaan | Tahun |
| 1 |  |  |  |
| 2 |  |  |  |

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Internal Tahun 2022.

Bandung, 14 Maret 2022

Anggota Tim

(Nur Tri Ramadhanti Adiningrum)

1. Identitas Diri Anggota Penelitian

|  |  |  |
| --- | --- | --- |
| 1. | Nama Diri | M. Rizky |
| 2. | Jenis Kelamin | Laki – Laki |
| 3. | Program Studi | DIV-Teknik Informatika |
| 4. | NIM | 1194021 |
| 5. | Tempat dan Tanggal Lahir | Dompu, 17 April 2000 |
| 6. | Alamat Email | [mriski889@gmail.com](mailto:mriski889@gmail.com) |
| 7. | Nomor Telepon/HP | 085239807970 |

1. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

|  |  |  |  |
| --- | --- | --- | --- |
| No | Jenis Kegiatan | Status Dalam Kegiatan | Waktu dan Tempat |
| 1. |  |  |  |
| 2. |  |  |  |
| 3. |  |  |  |

1. Penghargaan Yang Pernah Diterima

|  |  |  |  |
| --- | --- | --- | --- |
| No | Jenis Penghargaan | Pihak Pemberi Penghargaan | Tahun |
| 1. |  |  |  |
| 2. |  |  |  |

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Internal Tahun 2022.

Bandung, 14 Maret 2022

Anggota Tim

(M. Rizky)

Lampiran 3 Surat Pernyataan Ketua Penelitian

**SURAT PERNYATAAN BEBAS PLAGIAT**

Saya yang bertandatangan di bawah ini:

|  |  |  |
| --- | --- | --- |
| Nama | : | Syafrial Fachri Pane.,S.T.,M.TI.,EBDP |
| NIDN | : | 0416048803 |
| Program Studi | : | D4 Teknik Informatika |

Dengan ini menyatakan bahwa judul laporan ini **pemodelan berbasis data untuk memprediksi gaji berdasarkan faktor-faktor spesifik dengan pendekatan machine learning** benar bebas dari plagiat, dan apabila pernyataan ini terbukti tidak benar maka saya bersedia menerima sanksi sesuai ketentuan yang berlaku.

Demikian surat pernyataan ini saya buat untuk dipergunakan sebagaimana mestinya.

Bandung, 14 Maret 2022

Yang membuat pernyataan,

**(Syafrial Fachri Pane.,S.T.,M.T.I.,EBDP)**

**NIK. 117.88.233**

Lampiran 4 Penggunaan Anggaran

Lampiran 5 Bukti Penerimaan Artikel Ilmiah (LOA) atau URL dan Screenshoot Halaman Jurnal yang Sudah Dipublikasi

Lampiran 6 Format Catatan Harian (Logbook)

|  |  |  |
| --- | --- | --- |
| No | Tanggal | Kegiatan |
| 1 |  | Catatan :  Dokumen Pendukung :   1. Foto 2. Grafik/Tabel/Dokumen 3. dst |
| 2 |  | Catatan :  Dokumen Pendukung :   1. Foto 2. Grafik/Tabel/Dokumen   dst |
| 3 |  | Catatan :  Dokumen Pendukung :   1. Foto 2. Grafik/Tabel/Dokumen   dst |
| 4 |  | Catatan :  Dokumen Pendukung :   1. Foto 2. Grafik/Tabel/Dokumen   dst |

Lampiran 7 Poster