

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

SARJANA TERAPAN TEKNIK INFORMATIKA
UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL
TAHUN 2022

HALAMAN PENGESAHAN



POS INDONESIA



LEMBAR PENGESAHAN Penelitian Internal

Judul : Pemodelan berbasis data untuk memprediksi gaji berdasarkan faktor-faktor spesifik dengan pendekatan Machine Learning

Ketua

Nama Lengkap : Syafrial Fachri Pane, ST., M.T.I., EBDP
Program Studi : D4 Teknik Informatika
NIDN : 0416048803
Nomor HP : 08112164882
Alamat Surel (e-mail) : syafrial.fachri@poltekpos.ac.id

Pendamping 1

Nama Lengkap : Amri Yanuar, ST., M.MOT
NIDN : 0412018603
Perguruan Tinggi : Politeknik Pos Indonesia
Lama Penelitian : 8 Bulan
Biaya Penelitian : Rp. 6.500.000 ,-

Mengetahui,
Ketua Program Studi



(Muhammad Yusril Helmi Setyawan,
S.Kom., M.Kom.)
NIK/NIP. 113.74.163

Bandung, 16-03-2022
Ketua

(Syafrial Fachri Pane, ST.,
M.T.I., EBDP)
NIK/NIP. 147.88.233

Menyetujui,
Ketua LPPM




(Sanjaya, S.T., M.T.)
NIK/NIP. 103.77.054

16-03-2022 16:50:04

1/1

LPPM Politeknik Pos Indonesia

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 memiliki 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 lainnya 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*

PRAKARTA

Puji syukur kami haturkan kehadiran Allah SWT. yang telah melimpahkan kekuatan dan kesabaran, karena berkat rahmat dan hidayah-Nya, sehingga penulisan laporan penelitian akhir dengan judul “Pemodelan Berbasis Data Untuk Memprediksi Gaji Berdasarkan Faktor-Faktor Spesifik Dengan Pendekatan *Machine Learning*” yang telah lama dipersiapkan dapat diselesaikan dengan baik. Penyusunan laporan ini mendapatkan banyak masukan dan motivasi dari berbagai pihak terutama rekan-rekan yang telah berkolaborasi dalam penulisan buku ini. Laporan ini bertujuan untuk memenuhi persyaratan pada penelitian internal. Maka dari itu, pada kesempatan ini kami ingin mengucapkan terima kasih yang sebesar-besarnya kepada seluruh pihak atas keihklasannya dalam melakukan proses penyusunan laporan ini hingga selesai sesuai rencana serta sesuai dengan kaedah-kaedah penulisan karya ilmiah. Penulis juga menghanturkan terima kasih kepada orang tua dan keluarga yang sangat kami cintai, yang selalu memberikan motivasi dan selalu berdoa atas setiap langkah kebenaran yang kami lakukan. Semoga laporan ini kelak bermanfaat untuk penulis maupun siapa saja yang membaca laporan ini.

Penulis menyadari bahwa dalam penyusunan laporan ini terdapat banyak kekurangan dan jauh dari sempurna serta menyadari bahwa laporan ini hanya merupakan sebagian kecil dari luasnya pengetahuan dalam algoritma regresi linier berganda. Oleh karena itu, segala bentuk masukan baik saran ataupun kritik yang berharga dari berbagai pihak untuk membangun kesempurnaan laporan ini sangat kami harapkan.

Terima kasih

Penulis

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
HALAMAN KETERLIBATAN MAHASISWA DALAM PENELITIAN.....	iii
ABSTRAK	iv
PRAKARTA	v
DAFTAR ISI.....	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	2
1.3 Rancangan Hipotesis Penelitian	3
1.4 Ruang Lingkup Penelitian	3
1.5 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 State-of-The-Art	5
2.2 Tinjauan Pustaka	7
2.3 Taksonomi <i>Literature Review</i>	16
2.4 <i>Machine Learning</i>	17
2.3.1 <i>Multivariate Linear Regression</i>	19
2.3.2 <i>Random Forest Classifier</i>	20
2.5 Statistika	22
2.5.1 <i>Simple Linear Regression</i>	22
2.5.2 <i>Quantile Regression</i>	23
2.6 Metode Evaluasi Model <i>Machine Learning</i>	24
2.6.1 <i>Root Mean Square Error (RMSE)</i>	25
2.6.2 <i>Mean Square Error (MSE)</i>	25
2.6.3 <i>Mean Absolute Error (MAE)</i>	25
BAB III TUJUAN DAN MANFAAT.....	27
3.1 Tujuan dan Manfaat Penelitian.....	27
3.1.1 Tujuan Penelitian	27

3.1.2	Manfaat Penelitian	28
BAB IV METODE PENELITIAN		29
4.1	Diagram Alur Metodologi Penelitian	29
4.1.1	Diagram Alur Utama.....	29
BAB V HASIL DAN LUARAN YANG DICAPAI.....		31
5.1	Kegiatan dan Hasil Pelaksanaan.....	31
5.2	Luaran Yang Dicapai.....	33
5.3	Hasil Penelitian.....	35
BAB VI KESIMPULAN DAN SARAN.....		56
5.1	Kesimpulan.....	56
5.2	Saran	56
DAFTAR PUSTAKA		57
LAMPIRAN-LAMPIRAN		62

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	7
Tabel 2. 2 Klasifikasi Model berdasarkan SoTA	16
Tabel 4. 1 Penjelasan Diagram Alur Metodologi Penelitian.....	30
Tabel 5. 1 Karakteristik Dataset yang Digunakan	32
Tabel 5. 2 Luaran dan Target Capaian	35
Tabel 5. 3 Dataset Data Training	36
Tabel 5. 4 Dataset Data Testing	36
Tabel 5. 5 Atribut Dataset Data Training.....	37
Tabel 5. 6 Atribut Dataset Data Testing	38

DAFTAR GAMBAR

Gambar 2. 1 Taksonomi Studi Literatur	16
Gambar 2. 2 Visualisasi Classification dan Regression.....	18
Gambar 4. 1 Gambar Diagram Alur Metodologi Penelitian	29
Gambar 5. 1 Heatmap Correlation Dataset Data Training	42
Gambar 5. 2 Heatmap Correlation Dataset Data Testing.....	43
Gambar 5. 3 Visualisasi Penyebaran Data Training	45
Gambar 5. 4 Visualisasi Penyebaran Data Testing	45
Gambar 5. 5 Evaluasi Model OLS	48
Gambar 5. 6 Nilai F-statistic dan Probabilitas	49
Gambar 5. 7 Nilai P-Values Dari Variabel Independen.....	50
Gambar 5. 8 Grafik Asumsi Linear.....	51
Gambar 5. 9 Grafik Distribusi Residual.....	52
Gambar 5. 10 Tabel VIF	53
Gambar 5. 11 Skor Durbin-Watson Dan Hasil Asumsi	54
Gambar 5. 12 Homoskedastisitas	54
Gambar 5. 13 Hasil Prediksi Gaji	55

DAFTAR LAMPIRAN

Lampiran 1 Susunan Organisasi Tim Peneliti dan Pembagian Tugas	62
Lampiran 2 Biodata Ketua dan Anggota Pengusul	63
Lampiran 3 Surat Pernyataan Bebas Plagiat Ketua Penelitian	73
Lampiran 4 Penggunaan Anggaran	74
Lampiran 5 Bukti Penerimaan Artikel Ilmiah (LOA) atau URL dan Screenshoot Halaman Jurnal yang Sudah Dipublikasi	75
Lampiran 6 Format Catatan Harian (Logbook)	76
Lampiran 7 Poster	78

BAB I

PENDAHULUAN

1.1 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 menggantikannya dengan pekerjaan dengan karakteristik baru [5]. Karakteristik baru pada pekerjaan juga membutuhkan kompetensi baru kepada para pekerja[6]. Tentunya perusahaan harus siap untuk saling bersaing dengan perusahaan yang lain[7]. Selanjutnya, perusahaan perlu memiliki keunggulan dan manajemen yang efektif untuk menghadapi persaingan tersebut[7]. Dengan demikian salah satu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja karyawannya[7]. Walaupun perusahaan tersebut memiliki teknologi yang canggih, namun tidak terdapat tenaga kerja didalamnya, perusahaan tidak akan dapat mencapai tujuannya[7].

Oleh karena itu, penentuan gaji yang tepat oleh perusahaan kepada karyawan adalah salah satu faktor yang berpengaruh secara internal terhadap kemajuan perusahaan. Selain itu, perusahaan juga harus bersedia mengeluarkan gaji bonus

bagi karyawannya yang telah bekerja dengan maksimal dan sesuai dengan apa yang dibutuhkan oleh sebuah perusahaan. Sangat disayangkan, perkembangan perusahaan saat ini belum memiliki suatu media keputusan untuk melakukan 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.

1.2 Identifikasi Masalah

Adapun identifikasi masalah pada penelitian ini sebagai berikut.

- a) Bagaimana menganalisa karakteristik dan koelasi data terkait gaji karyawan?
- b) Bagaimana membuat model prediksi gaji karyawan berdasarkan kualitas data dengan mempertimbangkan faktor-faktor spesifik?

- c) Bagaimana merancang *framework* yang dinamis untuk menampilkan hasil prediksi gaji?

1.3 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.

1.4 Ruang Lingkup Penelitian

Adapun ruang lingkup pada penelitian ini sebagai berikut.

- a) Dikarenakan kondisi pandemi, data yang digunakan dalam penelitian ini bersumber dari data kepegawaian disuatu perusahaan.
- b) Periode waktu 1 tahun.
- c) Bahasa pemograman menggunakan Python.
- d) *Software* yang digunakan Jupyter atau dan Google Colab.

1.5 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 yang diambil.

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

2.1 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 sehingga 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 untuk menganalisis dan menghitung risiko sistemik[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[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 [13]-[34]. Lebih lanjut, detail dari tiap referensi lainnya ditunjukkan pada tabel.

2.2 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	<i>Multiple Regression</i>	Analisis dan prediksi berdasarkan data survey dokter hewan dengan model <i>multiple regression</i> .	-	-	-	0,42
2.	Penentu kepuasan kerja ahli gizi di Yordania[14]	2022	Data kuesioner ahli gizi.	Statistika	<i>Multivariate Regression</i>	Analisis dan prediksi berdasarkan data kuesioner ahli gizi dengan model <i>multivariate regression</i> .	-	-	-	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	<i>Multivariate Regression</i>	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	<i>Logarithmic Regressions</i>	Analisis data dari survey komunitas Amerika menunjukkan bahwa besarnya perbedaan gaji guru meningkat dari waktu ke waktu.	-	-	-	-
7.	Analisis gaji dan bakat guru[19].	2021	Data <i>Schools and Staffing</i> Survey (SASS).	Statistika	<i>Quantile Regression</i>	Variasi gaji antara distrik sekolah yang berdekatan menunjukkan bahwa bakat guru berkorelasi positif dengan gaji guru.	-	-	-	0,90
8.	Prediksi gaji karyawan berdasarkan pengalaman bekerja[20].	2022	Data survey dari <i>Google Form</i> .	<i>Machine Learning</i>	Regresi Linier	Analisis dan prediksi gaji karyawan berdasarkan pengalaman lama bekerja.	-	-	2,051	-
9.	Analisis dan penerapan data mining	2018	Data rekap gaji karyawan tetap dan karyawan	<i>Machine Learning</i>	K-Means Clustering.	Analisis dan prediksi dalam menentukan gaji karyawan tetap dan kontrak pada PT.	-	-	-	-

	untuk menentukan gaji karyawan [21].		kontrak PT. Indomex Dwijaya Lestari.			Indomex Dwijaya Lestari dengan model K-Means Clustering.				
10.	Prediksi gaji dengan menggunakan kemampuan kecerdasan emosional [22]	2021	Data survey penelitian 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.	<i>Machine Learning</i>	Regresi Linear, Regresi Polinomial	Prediksi gaji karyawan dengan metode regresi linear dan regresi polinomial.	-	-	-	-
12.	Analisis empiris teknik regresi	2020	Dataset gaji pegawai dan harga rumah.	<i>Machine Learning</i>	<i>Multiple Linear Regression</i>	Prediksi gaji karyawan dan harga rumah dengan model <i>multiple linear regression</i> .	1704,3	29045 23,3	1410,9	0,92

	dengan harga rumah dan prediksi gaji[24].									
13.	Prediksi gaji dalam penerapan model regresi dalam <i>Data Mining</i> [25].	2015	Data gaji dan lama pengalaman bekerja.	<i>Machine Learning</i>	<i>Simple Linear Regression</i>	Prediksi gaji karyawan dari lama pengalaman bekerja dengan model <i>simple linear regression</i> .	-	-	-	-
14.	Analisis dan Prediksi Kepuasan Gaji dalam Organisasi Negeri dan Swasta[26].	2011	Data 250 karyawan staf-manajerial dan non-manajerial.	Statistika	<i>Simple Regression</i>	Prediksi kepuasan bekerja dengan metode statistika model <i>simple regression</i> .	-	-	-	0,915
15.	Perancangan sistem	2022	Data dari database	<i>Machine Learning</i> ,	<i>Linear regression</i> ,	Perancangan sistem prediksi kenaikan gaji berbasis	-	-	-	-

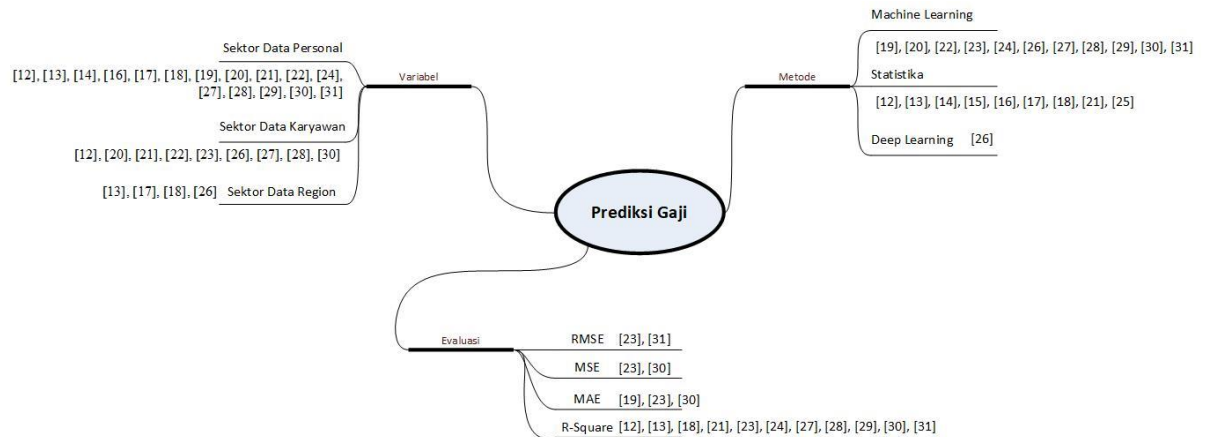
	berbasis <i>machine learning</i> untuk prediksi kenaikan gaji[27]		<i>Enterprise Resource Planning</i> (ERP)	<i>Deep Learning</i>	<i>artificial neural networks, random forest regression</i>	<i>machine learning</i> dengan menggunakan arsitektur <i>micro-services</i> .				
16.	Implementasi sistem prediksi gaji untuk meningkatkan motivasi siswa[28].	2016	13.541 data mahasiswa yang lulus dengan gajinya.	<i>Machine Learning</i>	<i>K-NN</i>	Prediksi gaji model <i>K-NN</i> memberikan akurasi terbaik sebesar 84,69% sedangkan	-	-	-	0,847
					<i>Naïve Bayes</i>		-	-	-	0,436
					<i>Decision trees</i>	<i>Multilayer Perceptron</i>	-	-	-	0,74
					<i>Multilayer perceptron</i>	memberikan akurasi terendah sebesar 38,08%.	-	-	-	0,381
					<i>SVM</i>		-	-	-	0,437
17.		2020			<i>Linear models</i>		-	-	-	0,586

	Prediksi gaji di pasar kerja TI [29].		Data <i>e-Recruitment</i> khusus untuk pekerjaan TI di Spanyol.	<i>Machine Learning</i>	<i>Logistic regression</i>	Prediksi kisaran gaji yang akurat dengan menggunakan model <i>random forest</i> secara umum lebih baik dengan menghasilkan akurasi 84%.	-	-	-	0,792
					<i>KNN</i>		-	-	-	0,591
					<i>Multi-layer perceptrons</i>		-	-	-	0,663
					<i>SVM</i>		-	-	-	0,836
					<i>Random forest</i>		-	-	-	0,840
					<i>Vote</i>		-	-	-	0,844
					<i>Vote3</i>		-	-	-	0,837
18.	Analisis prediktif gaji sumber daya manusia[30]	2022	Data ketenagakerjaan	<i>Machine Learning</i>	<i>Logistic Regression</i>	Prediksi gaji dengan model <i>Support Vector Machine</i> yang lebih akurat dibandingkan dengan <i>Logistic Regression</i> .	-	-	-	0,79
					<i>Support Vector Machine</i>		-	-	-	0,83
19.	Desain mesin prediksi baru	2018	Data kepegawaian	<i>Machine Learning</i>	<i>Decision tree classifier</i>	Memprediksi gaji yang sesuai untuk suatu pekerjaan dengan	-	389.64	6,04	0,844

	untuk memprediksi gaji[31]				<i>Random forest classifier</i>	metode <i>Decision tree classifier</i> dan <i>Random forest classifier</i> .	-	329.12	5,04	0,873
20.	Analisis prediktif untuk pendapatan alumni[32]	2022	Data survei studi dampak alumni Universitas Tecnológico de Monterrey	<i>Machine Learning</i>	<i>Quantile Regression</i>	Memprediksi pendapatan alumni untuk mendapatkan wawasan tentang prediktor terkuat dan kelas 'berpenghasilan tinggi' menggunakan metode <i>Quantile Regression (QR)</i>	50,431.45	-	-	0,44
					<i>Quantile Random Forest</i>		47,325.67	-	-	0,51
					<i>Quantile Gradient Boosting</i>		45,892.69	-	-	0,38
					<i>Linear Regression</i>		-	-	-	0,48
					<i>Random Forest Classifier</i>		-	-	-	0,50
					<i>Gradient Boosting Classifier</i>		-	-	-	0,53

21.	Komputasi cloud untuk fasilitas sinyal digital biomedis. [33]	2021	Data biomedis	<i>Framework Django</i>	Mmenyediakan layanan yang mampu menangani dan memproses data biomedis melalui <i>code-free interface</i> dengan <i>framework Django</i> .
22.	Aplikasi web prediksi diabetes [34]	2021	Data klinis penyakit diabetes	<i>Framework Django</i>	Model <i>machine learning</i> di- <i>deploy</i> menjadi aplikasi <i>web</i> dikembangkan untuk memprediksi diabetes yang sesuai.
23.	Deployment klasifikasi penyakit paru-paru [13]	2021	Dataset <i>x-ray</i> tubuh bagian atas untuk Covid-19, <i>Pneumonia</i> , dan <i>Normal</i> .	<i>Framework Django</i>	Klasifikasi penyakit paru-paru menggunakan model CNN di- <i>deploy</i> ke dalam <i>framework django</i> untuk menyediakan antarmuka pengguna yang lebih baik untuk memprediksi <i>output</i> .

2.3 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.	<i>Multivariate Linear Regression</i> [24]	92%
2.	<i>Simple Linear Regression</i> [26]	91,5%
3.	<i>Quantile Regression</i> [19]	90%
4.	<i>Random Forest Classifier</i> [31]	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 R^2 Koefisien Determinasi (R^2). 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.

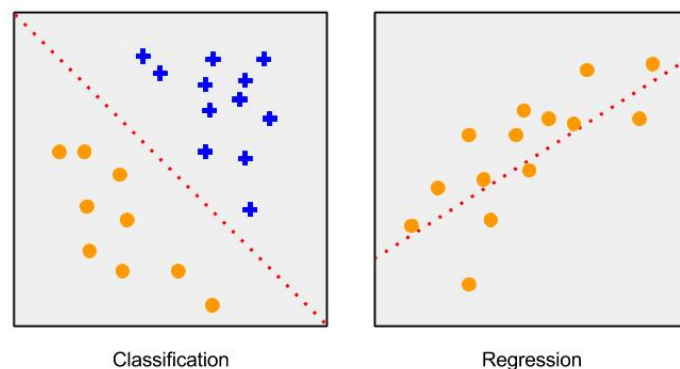
2.4 Machine Learning

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 dengan 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.



Gambar 2. 2 Visualisasi *Classification* dan *Regression*

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 :

2.3.1 *Multivariate 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 :

$$y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Keterangan :

y = Variabel tidak bebas (nilai yang diprediksikan)

X = Variabel bebas

a = Konstanta (nilai y apabila $X_1, X_2, \dots X_n = 0$)

b = Koefisien regresi (nilai peningkatan ataupun penurunan)

Dimana nilai $a, b_1, b_2, \dots b_n$ dapat dihitung dengan metode persamaan normal yaitu :

$$\sum y = a_n + b_1 \sum X_1 + b_2 \sum X_2 \quad (2)$$

$$\sum X_1 y = a \sum X_1 + b_1 \sum X_1 X_2 + b_2 \sum X_1 X_2 \quad (3)$$

$$\sum X_2 y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \quad (4)$$

Selain dihitung dengan persamaan normal di atas, nilai $1, b_1, b_2, \dots, b_n$ dapat dihitung dengan metode kuadran terkecil yaitu :

$$b_1 = \frac{(\sum X_2^2)(\sum X_1 y) - (\sum X_2 y)(\sum X_1 X_2)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} \quad (5)$$

$$b_2 = \frac{(\sum X_1^2)(\sum X_2 y) - (\sum X_1 y)(\sum X_1 X_2)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} \quad (6)$$

$$a = \frac{\sum y - (b_1 \sum X_1) - (b_2 \sum X_2)}{n} \quad (7)$$

Dimana :

$$\sum X_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n} \quad (8)$$

$$\sum X_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n} \quad (9)$$

$$\sum X_1 X_2 = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n} \quad (10)$$

$$\sum X_1 y = \sum X_1 y - \frac{(\sum X_1)(\sum y)}{n} \quad (11)$$

$$\sum X_2 y = \sum X_2 y - \frac{(\sum X_2)(\sum y)}{n} \quad (12)$$

$$\sum y^2 = \sum y^2 - \frac{(\sum y)^2}{n} \quad (13)$$

2.3.2 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 *tree* yang masing-masing *decision tree* memiliki kedalaman yang maksimal. *Random forest* adalah *classifier* yang berbentuk pohon dimana $\{h(x, \theta_k), k = 1, \dots\}$ dimana θ_k adalah *random vector* yang didistribusikan secara independent dan masing-masing *tree* pada sebuah unit kan memilih *class* yang paling populer input x . Berikut ini karakteristik akurasi pada *random forest*.

a. Memusatkan *random forest*

Terdapat *classifier* $h_1(x), h_2(x), \dots, h_k(x)$ dan dengan *training set* dari distribusi *random vector* Y, X . Berikut fungsi yang terbentuk.

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \quad (14)$$

Fungsi *error* yang digunakan..

$$PE^* = P_{X,Y}(mg(X, Y) < 0) \quad (15)$$

Hasil dari penggabungan fungsi.

$$P_{X,Y}(P_{\Theta}(h(X, \Theta) = Y) - \max_{j \neq Y} P_{\Theta}(h(X, \Theta) = j) < 0) \quad (16)$$

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

b. Kekuatan dan Korelasi

Fungsi yang dihasilkan adalah.

$$PE^* \leq \sum_i var(P_{\Theta}(h(X, \Theta) = Y) - P_{\Theta}(h(X, \Theta) = j))S_j^2 \quad (17)$$

Pada fungsi tersebut kekuatan tidak bergantung pada forest.

c. *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.

d. *Random Forest* menggunakan kombinasi input yang linier

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 linier F dihasilkan. Prosedur ini disebut *Forest-RC*[38].

2.5 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 sering digunakan pada masalah ilmiah, industri, atau sosial. 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 :

2.5.1 *Simple Linear Regression*

Simple linear regression adalah sebuah metode pendekatan untuk pemodelan hubungan antara satu variabel dependen dan satu variabel independen. Dalam regresi, variabel independen menerangkan variabel dependennya. Dalam analisis regresi sederhana, hubungan antar variabel bersifat linier, 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 linier sederhana sebagai berikut :

$$Y = a + bx \quad (18)$$

Keterangan :

Y = Variabel Dependen

a = Konstanta

b = 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 :

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \quad (19)$$

$$a = \frac{\sum Y - b \sum X}{n} \quad (20)$$

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

- a. Pembuatan dataset.
- b. Pembentukan model regresi linier.

Langkah pembentukan model sebagai berikut :

1. Langkah 1 : Hitung X^2 , Y^2 , XY dan total dari masing-masingnya.
2. Langkah 2 : Hitung a dengan menggunakan persamaan (2) dan b menggunakan persamaan (3).
3. Langkah 3 : Buat model persamaan *simple linear regression*.
4. Langkah 4 : Lakukan prediksi atau peramalan terhadap variabel faktor penyebab atau variabel akibat[40].
- 5.

2.5.2 *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 analisis 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 :

- a. Untuk memperkirakan median, atau kuantil 0,25 atau kuantil apapun.
- b. Asumsi kunci regresi linier tidak terpenuhi
- c. *Outlier* dalam data

- d. Residu tidak normal
- e. Peningkatan varians kesalahan dengan peningkatan variabel hasil

Berikut adalah model *linear regression* :

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (21)$$

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

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))^2 \quad (22)$$

Persamaan model *Quantile Regression* untuk kuantil ke- τ adalah.

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_p(\tau)x_{ip} \quad (23)$$

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

$$MAD = \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - (\beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_p(\tau)x_{ip})) \quad (24)$$

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

$$\rho_\tau(u) = \tau \max(u, 0) + (1 - \tau) \max(-u, 0) \quad (25)$$

2.6 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 :

2.6.1 *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] :

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad (26)$$

Keterangan :

Y' = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

2.6.2 *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].

$$MSE = \sum \frac{(Y' - Y)^2}{n} \quad (27)$$

Keterangan :

Y' = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

2.6.3 *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] :

$$MAE = \sum \frac{|Y' - Y|}{n} \quad (28)$$

Keterangan :

Y' = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

BAB III

TUJUAN DAN MANFAAT

3.1 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 penyampaian 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.

3.1.1 Tujuan Penelitian

Adapun tujuan pada penelitian ini sebagai berikut.

- a) Menganalisa korelasi data gaji karyawan berdasarkan faktor-faktor spesifik.
- b) 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.
- c) Menggunakan *framework* Django untuk menyajikan hasil prediksi gaji karyawan.

3.1.2 Manfaat Penelitian

Adapun manfaat pada penelitian ini sebagai berikut.

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

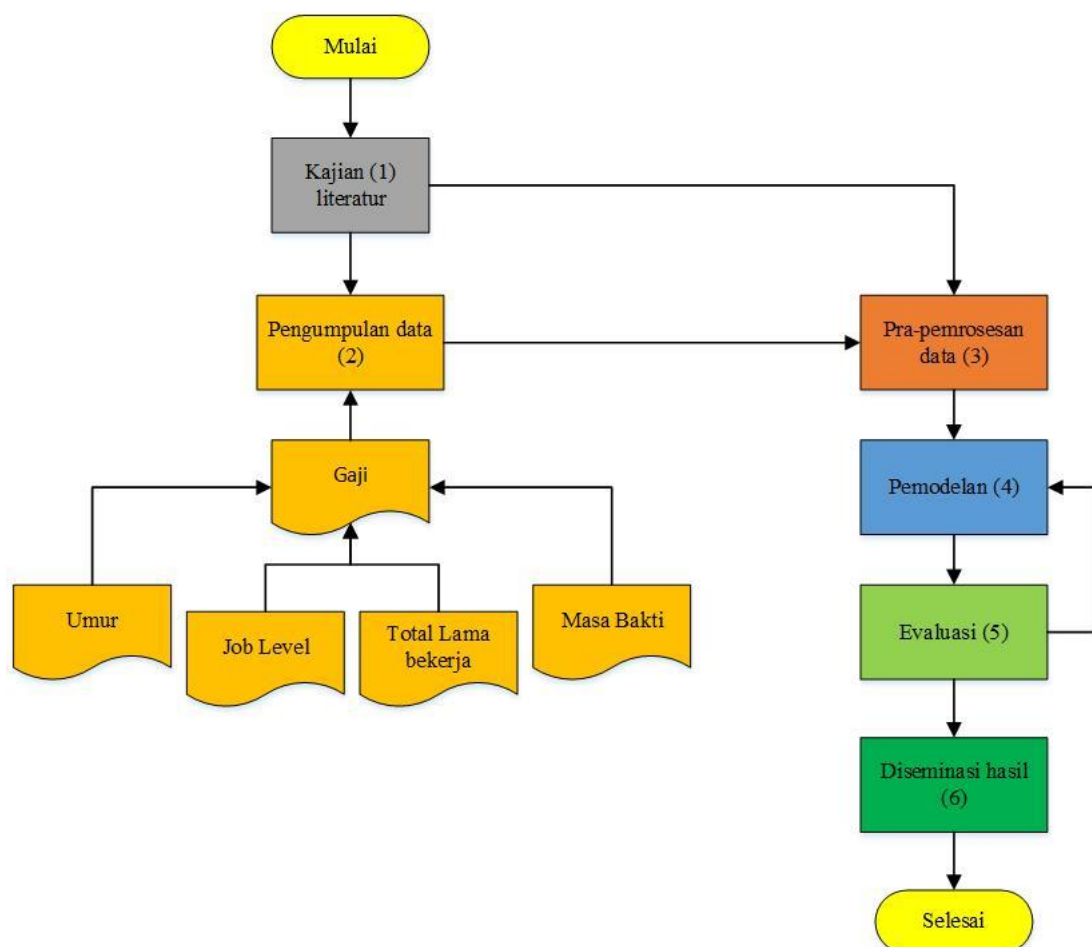
BAB IV

METODE PENELITIAN

4.1 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 tertentu. Metode penelitian tersebut bisa berupa diagram alur penelitian.

4.1.1 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 <i>machine learning</i> serta <i>framework</i> <i>django</i> .
2.	Studi literatur	→	2. <i>Mind map</i> prediksi gaji terhadap faktor-faktor spesifik berdasarkan data dan pada metode <i>machine learning</i> serta <i>framework</i> <i>django</i> ..
3.	Pengumpulan data	→	3. Data mentah dari berbagai faktor (umur, <i>job level</i> ,lama pengalaman bekerja,masa bakti)
3.	Pre-pemrosesan data	→	4. <i>Pre-processed</i> data yang sudah siap untuk pemodelan dengan tahapan pembersihan, penanganan nilai yang hilang dan transformasi.
4.	Pemodelan	→	5. Model Multivariat Linier Regresi digunakan untuk memprediksi gaji berdasarkan data dari setiap faktor-faktor spesifik.
5.	Evaluasi	→	6. Performansi model
6.	Diseminasi hasil	→	7. Artikel yang diterbitkan dalam jurnal nasional terakreditasi SINTA 3, HAKI dan Buku

BAB V

HASIL DAN LUARAN YANG DICAPAI

5.1 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.

5.1.1 Pelaksanaan Penelitian

Kegiatan penelitian yang dilakukan telah berjalan selama 6 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. Berdasarkan 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%.

5.1.2 Dataset yang Digunakan

Tabel 5. 1 Karakteristik Dataset yang Digunakan

Kode	Variabel	Jenis Variabel	Tipe Data	Sumber Data
X1	Umur	Data Personal	Numerik	Kaggle
X2	Gaji		Numerik	Kaggle
X3	Lama Pengalaman Bekerja		Numerik	Kaggle
X4	Job Level	Data Karyawan	Numerik	Kaggle
X5	Masa Bakti		Numerik	Kaggle

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.

5.1.3 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 perusahaan.

5.1.4 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.

5.1.5 Penyusunan Buku ISBN

Penyusunan buku ISBN dilakukan oleh anggota tim ketika masalah dari penelitian telah mendapatkan hasil yang dicapai. Penyusunan buku dilakukan selama 6 bulan. Buku yang disusun dari BAB I hingga BAB VI. Buku yang disusun terdiri cover, kata sambutan, kata pengantar dan terdiri dari 225 halaman terdaftar di Perpustakaan Nasional Republik Indonesia dan ber-ISBN.

5.1.6 Pembuatan Poster

Pembuatan poster dilakukan oleh anggota tim ketika masalah dari penelitian telah diatasi dan hasil penelitian telah mencapai hasil yang diharapkan. Poster yang dibuat berjumlah satu lembar dengan ukuran A2 secara vertical.

5.1.7 Pelaksanaan Praktikum Pada Mata Kuliah Database

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

5.2 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 jurnal nasional SINTA S3. Jurnal penelitian ini telah dihasilkan dalam bentuk draf. Draf tersebut belum sepenuhnya selesai karena belum direview lebih lanjut oleh anggota tim maupun diskusi bersama.

2. 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.

3. 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.

4. Poster penelitian

Poster penelitian adalah luaran yang ditargetkan untuk penelitian. Poster sebagai salah satu luaran dari pelaksanaan program penelitian ini telah dibuat dalam bentuk gambar desain grafis, dimana pembuatan poster ini disusun dari latar belakang, metode, hasil utama penelitian, kesimpulan, dan referensi. Penyusunan tersebut juga belum dikoreksi lebih lanjut karena belum melakukan diskusi lebih lanjut. Capaian luaran ini sesuai dengan target yaitu berupa pembuatan poster penelitian.

5. 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. 2 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 referensi 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.

5.3 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 menunjukkan 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. 3 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. 4 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. 5 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. 6 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())

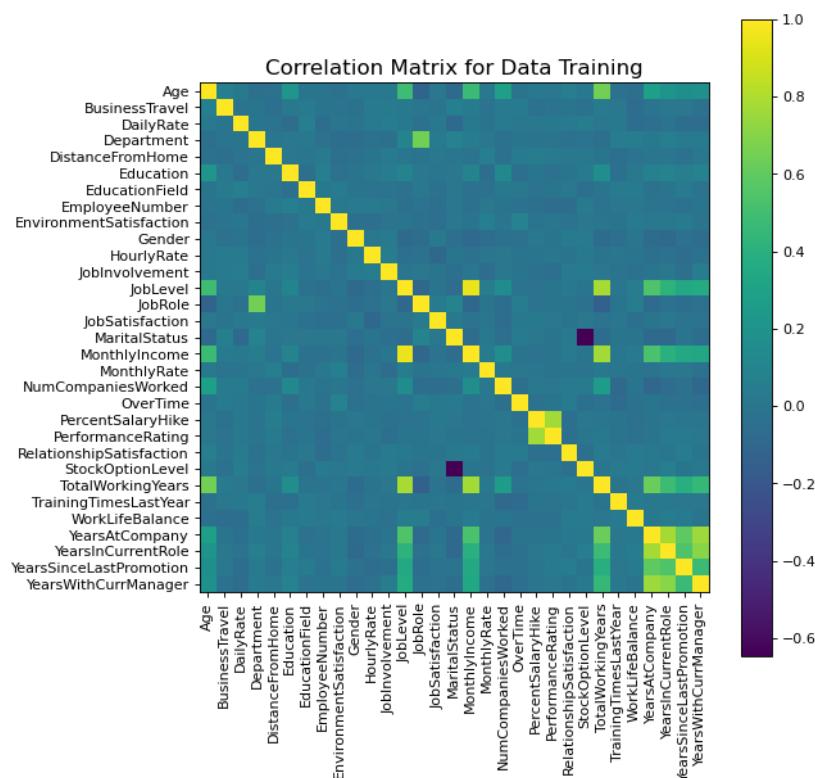
#Mengambil kolom DailyRate untuk diisi dengan mean-nya
DailyRate = df_train['DailyRate']
DailyRate.describe()
df_train.DailyRate =
df_train.DailyRate.fillna(value=df_train.DailyRate.mean())

```

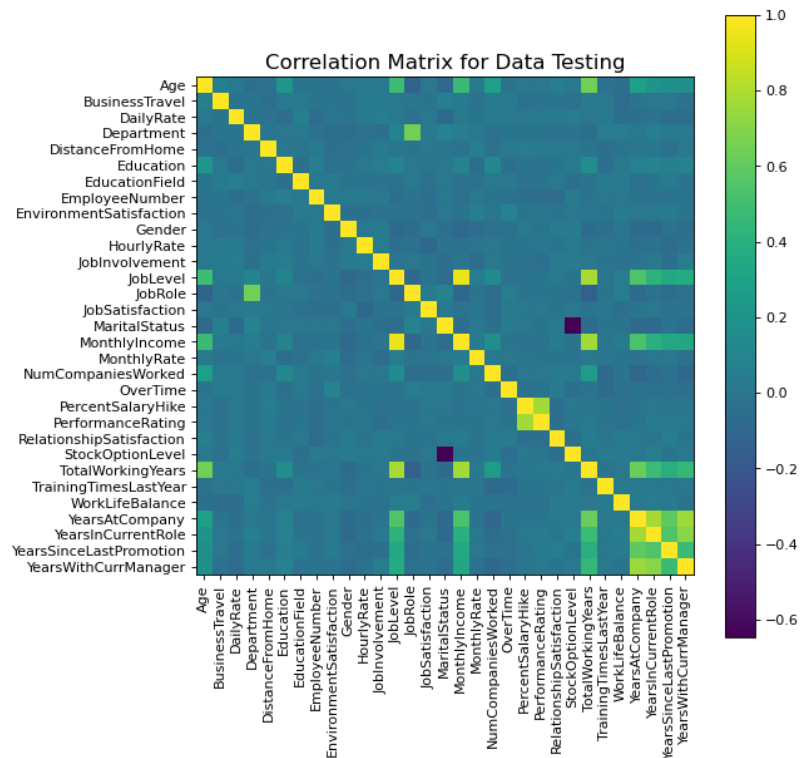
```
#Mengambil kolom DistanceFromHome untuk diisi dengan mean-nya
DistanceFromHome = df_train['DistanceFromHome']
DistanceFromHome.describe()

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.

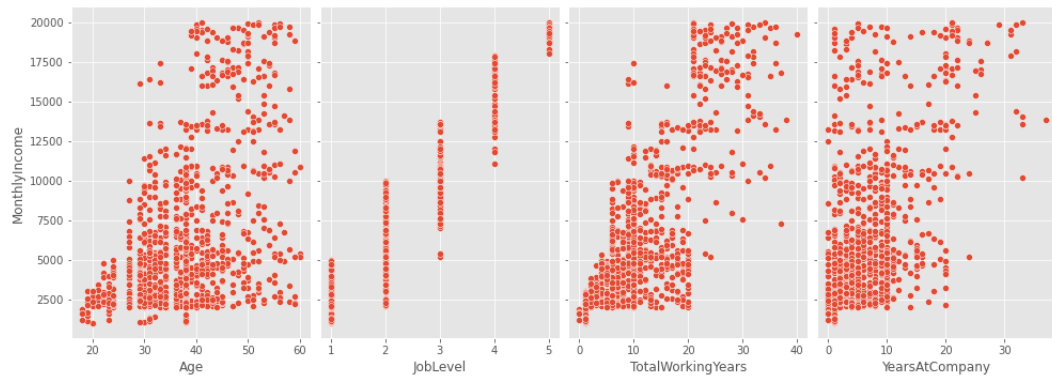
```

# 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')

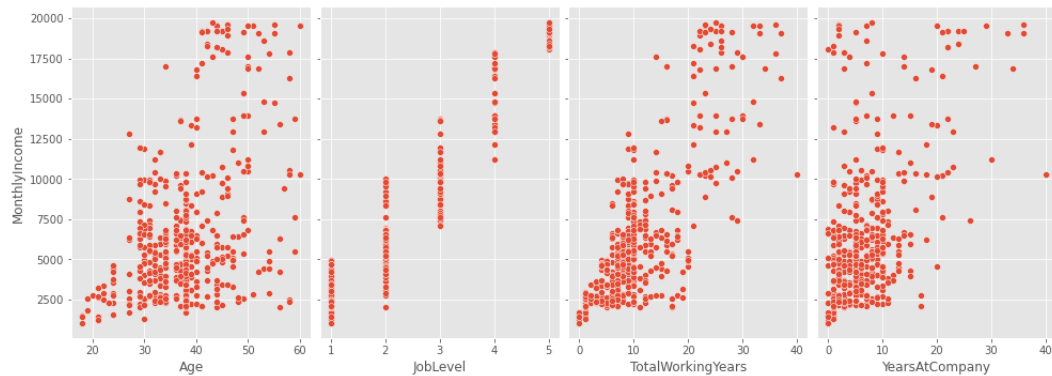
x_test = df_test_clean[['Age', 'JobLevel', 'TotalWorkingYears',
'YearsAtCompany']]
y_test = df_test_clean[['MonthlyIncome']]
df_test_clean.to_csv('employee_attrition_test_clean.csv')

```

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.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \beta_n X_n$$

β_0 adalah nilai *intersept* dari persamaan linear, dan β_1 , β_2 , β_3 sampai dengan β_n adalah konstanta dari variabel independen. Berdasarkan nilai koefisien

variabel independen dan *Intersept* didapat, maka persamaan regresi linear multivariabel sebagai berikut :

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

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 :

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

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 dipengaruhi oleh faktor Age dan YearsAtCompany sebesar 0,909 atau 90,9%. Nilai sisa dari

koefisien determinasi adalah 0.091 atau 9,1% dipengaruhi oleh faktor lain yang tidak diketahui.

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())
```

```

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

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

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 probabilitas 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 uji-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)
```

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

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 $F_s > P\text{-value}$, 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 :

- a. H_0 = Variabel independen secara simultan bukan penjelas yg signifikan terhadap variabel dependen (Model tidak cocok).
- b. H_1 = Variabel independen secara simultan merupakan penjelas yang signifikan terhadap variabel dependen (Model cocok).

Berdasarkan gambar 5.6, dapat diketahui bahwa $F_s > P\text{-value}$, yang artinya hipotesa yang dapat diambil adalah terima H_1 dan tolak H_0 . Dapat dikatakan, variabel Independen (Age, JobLevel, TotalWorkingYears, YearsAtCompany) dan MonthlyIncome berpengaruh signifikan terhadap permintaan. Pada taraf signifikansi 5% (0,05), H_0 ditolak karena nilai probabilitasnya yaitu 0,00 yang berarti dibawah dari 5%. Maka dapat disimpulkan, model yang dipakai cocok.

```
print(olsmod.pvalues)
```

const	1.419659e-13
Age	4.643777e-01
JobLevel	0.000000e+00
TotalWorkingYears	6.771641e-05
YearsAtCompany	3.136608e-01
dtype: float64	

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 :

- H_0 = Variabel independen tidak berpengaruh signifikan
- H_1 = Variabel independen berpengaruh signifikan.
- $\alpha = 0,05$ (Taraf signifikansi)

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

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

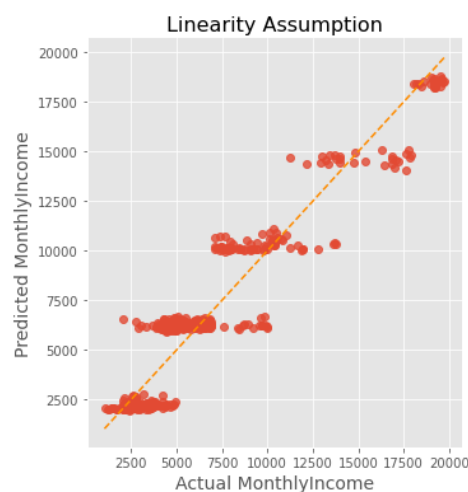
Berdasarkan hipotesa di atas, dapat disimpulkan variabel independen JobLevel dan TotalWorkingYears adalah variabel yang tidak mempengaruhi

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.

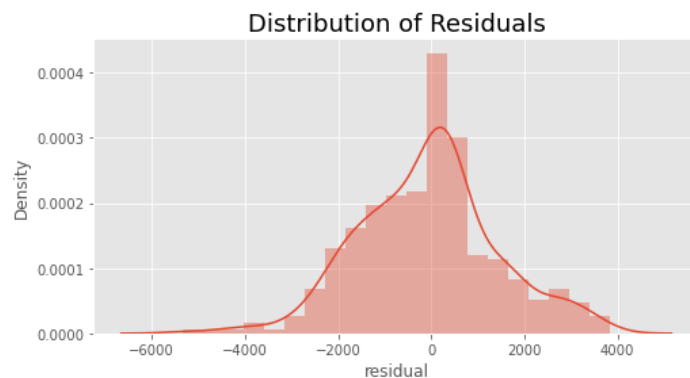
Setelah itu, dilakukan pengujian asumsi untuk model yang dibuat. Uji asumsi pertama yang dilakukan adalah uji linearitas. Yaitu pengujian untuk memeriksa apakah terdapat hubungan linear antara variabel independen dengan variabel dependen. Uji linearitas dilakukan dengan menggunakan plot pencar sehingga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

- a. H_0 = Residual terdistribusi normal.
- b. H_1 = 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 tolak H_0 terima H_1 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:

- Jika nilai VIF < 10 atau nilai Tolerance > 0,01, maka dinyatakan tidak terjadi multikolinearitas.
- Jika nilai VIF > 10 atau nilai Tolerance < 0,01, maka dinyatakan terjadi multikolinearitas.
- 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.

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

Gambar 5. 10 Tabel VIF

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

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

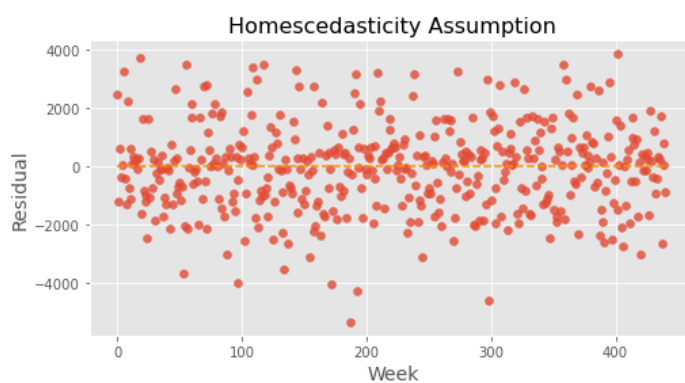
```
Durbin-Watson: 2.160636228778726
Little to no autocorrelation

Assumption satisfied
```

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 terdapat 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)
```

Gaji yang terprediksi pada pegawai perbulan yang bekerja sepanjang tahun tersebut adalah [[2173.12332303]]

Gambar 5. 13 Hasil Prediksi Gaji

BAB VI

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan dapat disimpulkan beberapa poin yaitu :

1. Berdasarkan hasil pada model OLS didapatkan nilai akurasi sebesar 0,909. Akurasi tersebut merupakan nilai akurasi yang baik, sehingga dapat dikatakan model *machine learning* dapat berperforma baik untuk memprediksi gaji.
2. Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa MonthlyIncome dipengaruhi oleh faktor independen (Age, JobLevel, TotalWorkingYears, YearsAtCompany) sebesar 0,909 atau 90,9%.
3. Visualisasi data dari hasil model prediksi gaji karyawan dapat digunakan menjadi bentuk aplikasi berbasis *web base* dengan menggunakan *framework* Django. Dengan aplikasi tersebut, admin dapat melakukan prediksi gaji karyawan dengan mudah dan dengan cepat.

5.2 Saran

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

1. Diharapkan model yang dibangun terhindar dari *outlier* dan *over fitting*
2. Diharapkan model yang dibangun dapat memberikan rekomendasi variabel apa saja yang dominan dan optimum untuk memprediksi gaji dengan menggabungkan *machine learning* dan algoritma optimasi
3. Perlu pengkajian secara komprehensif untuk mempelajari variabel sebelum pemodelan yang disebut *hyper parameter tuning* yang bertujuan mendapatkan variabel yang baik untuk pemodelan

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. Elsayhori, 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. Bresó, 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 Prediction Using Regression Technique,” *International Conference On Industry Interactive Innovations In Science, Engineering And Technology*, 2020, doi: <https://dx.doi.org/10.2139/ssrn.3526707>.
- [24] U. Bansal, A. Narang, A. Sachdeva, I. Kashyap, and S. P. Panda, “Empirical analysis of regression techniques by house price and salary prediction,” in *IOP Conference Series: Materials Science and Engineering*, Jan. 2021, vol. 1022, no. 1. doi: 10.1088/1757-899X/1022/1/012110.
- [25] S. Gupta, “A Regression Modeling Technique on Data Mining,” *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 Increase,” *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. Brilliant, 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

A. 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

B. 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	<i>Cumlaude</i>	<i>Cumlaude</i>	<i>Cumlaude</i>

C. 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

D. 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	Politeknik Pos Indonesia	2019

	Otak Dengan Metode Bayesian Learning dan Logistic Regression		
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	2019

		Terindex DOAI	
17	Collaboration Fmadm And K-Means Clustering To Determine The Activity Proposal In Operational Management Activity	Jurnal Emitter – 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

E. 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)

A. 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
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

B. 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

C. 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

D. 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

E. 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

A. 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

B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

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

C. 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)

A. 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

B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

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

C. 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)

A. 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
7.	Nomor Telepon/HP	085239807970

B. Kegiatan Kemahasiswaan Yang Sedang/Pernah Diikuti

No	Jenis Kegiatan	Status Dalam Kegiatan	Waktu dan Tempat
1.	Himatif	Tidak Aktif	ULBI
2.			
3.			

C. 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 Bebas Plagiat 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,



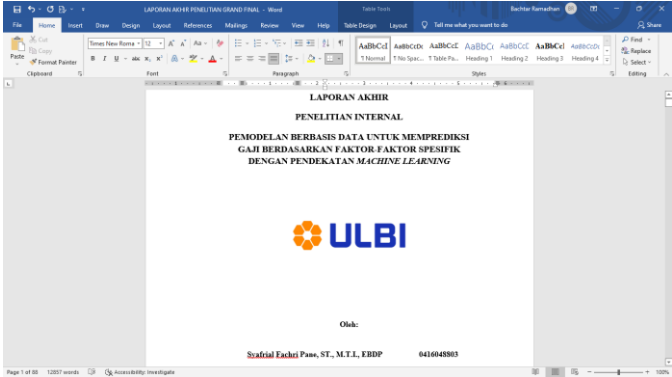
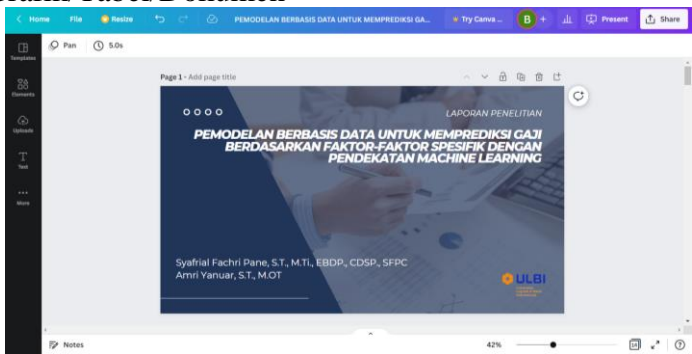
(Syafrial Fachri Pane.,S.T.,M.T.I.,EBDP)
NIK. 117.88.233

Lampiran 4 Penggunaan Anggaran

1. Gaji dan Upah				
Jabatan	Gaji & Upah / Jam	Waktu	Minggu	TOTAL
	(Rp)	(Jam/Minggu)		(Rp)
Ketua Peneliti	6500	6 Jam	16	832.000
Anggota 1 (Dosen)	4500	4 Jam	16	576.000
Anggota 2 (Mahasiswa)	4500	4 Jam	16	350.000
Anggota 3 (Mahasiswa)	4500	4 Jam	16	350.000
Anggota 4 (Mahasiswa)	4500	4 Jam	16	350.000
SUB TOTAL (Rp)				2.458.000
2. Bahan Habis Pakai				
Material	Justifikasi	Kuantitas	Harga Satuan	TOTAL (Rp)
	Pemakaian		(Rp)	
Kertas A4	Untuk Print Dokumen	2 Rim	50.000	100.000
Tinta Printer	Untuk Keperluan Print	1 buah	200.000	200.000
Konsumsi	Rapat/Diskusi 8 Bulan	8 Bulan	100.000	800.000
ATK	Administrasi	1 Pack	250.000	250.000
Sewa Zoom (100 Partisipan)	Meeting Online 8 Bulan	1 Paket	250.000	250.000
Subsidi Kuta Pulsa	Komunikasi	8 Bulan	50.000	2.000.000
SUB TOTAL (Rp)				3.600.000
3. Perjalanan				
Material	Justifikasi	Kuantitas	Harga Satuan	TOTAL (Rp)
	Pemakaian		(Rp)	
SUB TOTAL (Rp)				-
4. Lain - Lain				
Material	Justifikasi	Kuantitas	Harga Satuan	TOTAL (Rp)
	Pemakaian		(Rp)	
Administrasi, publikasi, laporan	Publikasi dan Laporan	1 kali	500.000	500.000
SUB TOTAL (Rp)				500.000
TOTAL DANA YANG DIBUTUHKAN (Rp)				6.558.000

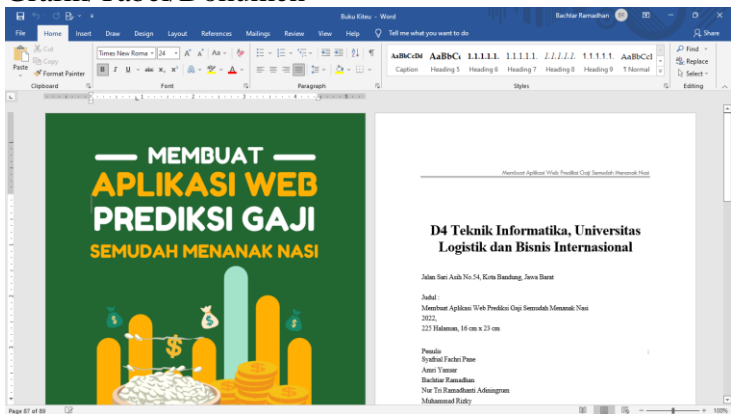
Lampiran 5 Bukti Penerimaan Artikel Ilmiah (LOA) atau URL dan Screenshoot Halaman Jurnal yang Sudah Dipublikasi

Lampiran 6 Format Catatan Harian (Logbook)

No	Tanggal	Kegiatan
1		<p>Catatan : Pembuatan Laporan Akhir Penelitian.</p> <p>Dokumen Pendukung :</p> <ol style="list-style-type: none"> 1. Foto 2. Grafik/Tabel/Dokumen  <ol style="list-style-type: none"> 3. dst
2		<p>Catatan : Pembuatan Power Point untuk Presentasi.</p> <p>Dokumen Pendukung :</p> <ol style="list-style-type: none"> 1. Foto 2. Grafik/Tabel/Dokumen  <p>dst</p>
3		<p>Catatan : Pembuatan Jurnal Ilmiah ter-akreditasi SINTA 3.</p> <p>Dokumen Pendukung :</p> <ol style="list-style-type: none"> 1. Foto 2. Grafik/Tabel/Dokumen <p>dst</p>
4		<p>Catatan : Pembuatan Buk uber-ISBN.</p>

Dokumen Pendukung :

- 1. Foto
- 2. Grafik/Tabel/Dokumen



dst

Lampiran 7 Poster

Pemodelan Berbasis Data Untuk Memprediksi Gaji Berdasarkan Faktor-Faktor Spesifik Dengan Pendekatan Machine Learning

Sekolah Vokasi, Sarjana Terapan Teknik Informatika, Universitas Logistik dan Bisnis Internasional

Syafrial Fakhri Pane, S.T., M.Ti., EBDP., CDSP., SFPC. (0416048803) , Amri Yanuar., S.T., M.OT. (0412018603)

Keterlibatan Mahasiswa : Bachtiar Ramadhan (1204077), Nur Tri Ramadhanti Adiningrum (1204061), Muhammad Rizky (1194021)



Abstrak

Persebaran gaji yang sesuai merupakan salah satu faktor penting untuk mendorong kinerja karyawan. Sayangnya, perkembangan teknologi saat ini belum memiliki media pengabdian perusahaan untuk memprediksi gaji karyawan berdasarkan kualitas data. Dalam memprediksi gaji tidak banyak faktor yang diperimbangkan, sehingga penelitian ini bertujuan untuk memprediksi gaji berdasarkan faktor-faktor tertentu. Penelitian ini dilakukan dengan menggunakan jumlah faktor yang lebih banyak dibandingkan penelitian lainnya. Faktor yang diambil antara lain faktor independen berupa Umur, Joblevel, Lama Pengalaman Bekerja, Masa Bakti dan faktor dependen berupa Pendapatan Bulanan. Teknik analisis data yang digunakan adalah analisis regresi linear multivariat. Model berbasis data uji pada tahap validasi model yang menghasilkan nilai akurasi sebesar 90,9%. Sehingga dapat diimpikan bahwa model dapat bekerja dengan baik. Selanjutnya penelitian ini menghasilkan visualisasi data menggunakan framework Django.

Kata Kunci – Prediksi Gaji, Regresi Linear Multivariat, Faktor Spesifik, Web Base.

Identifikasi Masalah

- Bagaimana mengolah karakteristik dan koefisien data terkait gaji karyawan?
- Bagaimana membuat model prediktor berdasarkan kualitas data dengan mempertimbangkan faktor-faktor spesifik?
- Bagaimana merancang framework yang dinamis untuk menampilkan visualisasi prediksi gaji?

Tujuan

Penelitian ini bertujuan untuk mengaitkan pengaruh faktor-faktor spesifik terhadap prediksi gaji karyawan dan bagaimana cara menampilkan visualisasi prediksi gaji secara real time.

Pendahuluan

Ergonomi akan mendukung berbagai kegiatan diberbagai bidang seperti pada bidang teknologi, ekonomi, sosial, dan politik[1]. Perubahan karakteristik pekerjaan adalah salah satu faktor penyebab terwujud dari dataregresi residual industri 4.0[2]. Karakteristik pekerjaan yang berubah akan mereduksi pekerjaan yang telah ada[3]. Tentunya perusahaan harus siap untuk saling bersaing dengan perusahaan yang lain[4]. Perusahaan perlu memiliki keunggulan dan manajemen yang efektif untuk menghadapi persaingan tersebut[5]. Salah satu aspek yang berpengaruh besar terhadap kemajuan sebuah perusahaan adalah kinerja karyawannya[6]. Oleh karena itu, pemberian gaji adalah salah satu faktor yang berpengaruh secara internal terhadap kemajuan perusahaan. Namun, perusahaan saat ini belum memiliki suatu media layanan untuk melakukan prediksi gaji karyawan berdasarkan kualitas data. Metodologi Karatecihik dataset yang digunakan terdiri dari parameter berdasarkan faktor-faktor spesifik. Selanjutnya faktor-faktor tersebut akan diuji validasi dan korelasinya dengan machine learning faktor-faktor tersebut akan diambil berdasarkan pedoman interpretasi koefisien korelasi[7]. Metode yang digunakan pada machine learning yaitu regresi linear. Hasil prediksi gaji karyawan divisualisasikan secara real time pada web base dengan framework Django.

Hasil Penelitian

Berdasarkan model yang diambil dari model OLS didapatkan nilai akurasi sebesar 0,909. Model prediksi yang dirancang dengan menggunakan machine learning dengan pendekatan regresi. Prediksi gaji karyawan menggunakan faktor-faktor spesifik yaitu umur, level, Lama Pengalaman Bekerja, dan Masa Bakti. Berdasarkan uji validasi, nilai akurasi 0,909 menunjukkan bahwa ModelMachine dapat menghasilkan prediksi gaji karyawan dengan akurasi sebesar 0,909 atau 90,9%.

Kesimpulan Penelitian

Berdasarkan hasil pada model OLS didapatkan nilai akurasi sebesar 0,909. Akurasi tersebut merupakan nilai akurasi yang baik, sehingga dapat dikatakan model machine learning dapat berperforma baik untuk memprediksi gaji. Berdasarkan uji validasi, nilai akurasi 0,909 menunjukkan bahwa uji dipengaruh oleh faktor independen (Umur, Masa Bakti) sebesar 0,909 atau 90,9%. Visualisasi data dari hasil model prediksi gaji karyawan dapat digunakan menjadi bentuk aplikasi berbasis web base dengan menggunakan framework Django. Dengan aplikasi tersebut, admin dapat melakukan prediksi gaji karyawan dengan mudah dan dengan cepat.

Daftar Pustaka

- [1] Prasetya B And Triyanti U, "Revisi Industri 4.0 Dan Tantangan Perubahan Sosial", *Journal Of Proceedings Series*, No. 5, Pp. 22-27, Nov. 2018. [2] A. A. Shahrudin and N. Riana, "Industrial Revolution 4.0 and Education", *International Journal of Academic Research in Business and Social Sciences*, vol. 8, no. 5, pp. 316-318, Oct. 2018. [3] S. Rengasach, "Industry 4.0: New Challenges And Opportunities For The Labour Market", *Foreign and STI Governance*, vol. 11, no. 4, pp. 6-9, 2017. doi: <http://dx.doi.org/10.17212/2500-2501.2017.4.6.8> [4] Y. Adhewane Da, Yash and Anjan, "Implementation Model Beyond Linear Regression Untuk Prediksi Gaji Berdasarkan Pengalaman Lama Bekerja", *Journal Education and Technology*, vol. 1, no. 2, pp. 56-70, Dec. 2019.[5] Tawana H.E., Sumpster Peter, Muzika R., "Pengaruh Professionalisme Berpaga Pegawai Terhadap Tingkat Kepuasan Pelayanan Pada Kantor PT. Papan Cahang Smanan", *Jurnal Administrasi Publik*, vol. 3, no. 01, pp. 1-9 2017.

Metodologi Penelitian



Gambar 1. Metodologi Penelitian

No	Langkah	Detail Langkah
1	Koleksi data parameter	1.1. Kumpulkan data gaji karyawan yang telah ada di perusahaan.
2	Studi literatur	2.1. Riset yang berkaitan dengan faktor-faktor yang mempengaruhi gaji karyawan.
3	Pengumpulan data	3.1. Dapatkan data gaji karyawan yang telah ada di perusahaan.
4	Analisis data	4.1. Lakukan analisis data untuk mengetahui faktor-faktor yang mempengaruhi gaji karyawan.
5	Pemodelan	5.1. Buatlah model prediksi gaji karyawan berdasarkan faktor-faktor yang telah ditentukan.
6	Validasi model	6.1. Lakukan validasi model untuk mengetahui akurasi model prediksi gaji karyawan.

Gambar 2. Indikator Capaian Tahapan Penelitian

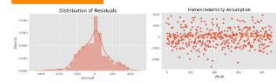
Implementasi



Gambar 3. Implementasi Framework

Berikut gambar 7 adalah implementasi web base dengan framework Django yang diintegrasikan dengan model machine learning untuk memprediksi gaji berdasarkan faktor-faktor spesifik.

Uji Validitas



Gambar 3. Uji Asumsi Normalitas

Dapat dilihat bahwa pada gambar 2 yang merupakan grafik untuk melihat normalitas dengan metode

Anderson-Darling. P-value yang didapat dari hasil perhitungan

sebesar 0,00032261. Angka tersebut berada di bawah nilai threshold yang

ditentukan yaitu 0,05, yang berarti Residual terdistribusi secara tidak

normal.

Gambar 3. Uji Asumsi Homoskedastisitas

Dari grafik scatterplot (gambar 3), terlihat titik-titik residual

menyebar secara acak, serta

tersebar baik di atas maupun di

bawah angka 0 (nol) pada sumbu

Y. Oleh karena itu, dapat diambil

kesimpulan bahwa tidak terdapat

gejala heteroskedastisitas pada

model regresi yang digunakan.

Gambar 5. Uji Asumsi Linearitas

Berdasarkan gambar 5, plot sebar menunjukkan

titik yang tersebar merata di sekitar garis

diagonal, sehingga dapat disimpulkan bahwa ada

hubungan linear antara variabel independen dan

dependen.

Gambar 6. Uji Asumsi Multikolinearitas

Berdasarkan gambar 5, dapat dilihat nilai

variabel umur, level, lama pengalaman

bekerja, masa bakti memiliki nilai kurang dari

10 sehingga dengan menggunakan tingkat

signifikansi sebesar 0,05 dapat disimpulkan

bahwa pada data tersebut tidak terdapat

multikolinearitas pada variabel-variabel

prediktor.

Keluaran Penelitian

Luaran yang dicapai dengan target capaian pada

penelitian sebagai berikut :

- Jurnal Ilmiah Nasional terakreditasi SINTA 3
- Buku ber-ISBN
- HAKI (Hak Atas Kekayaan Intelektual)

