

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



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 : Universitas Logistik dan Bisnis Internasional

Lama Penelitian : 8 Bulan

Biaya Penelitian : Rp. 8.500.000 ,-

Mengetahui,
Ketua Program Studi



(Roni Andarsyah, ST., M.Kom.)

UNIK/NIP. 115.88.193

Logistik & Bisnis
Internasional

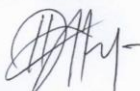
Bandung, 10-10-2022
Ketua



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

NIK/NIP. 117.88.233

Menyetujui,
Direktur RPIKK





(Dr. Prety Diawati, S.Sos., MM.)
NIK/NIP. 113.75.108

10-10-2022 17:25:49

1/1

Direktorat RPIKK Universitas Logistik
dan Bisnis Internasional

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	

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	21
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>	24
2.6.2 <i>Mean Square Error (MSE)</i>	25
2.6.3 <i>Mean Absolute Error (MAE)</i>	25
2.6.4 <i>OLS Regression (Uji Asumsi)</i>	25
BAB III TUJUAN DAN MANFAAT.....	30
3.1 Tujuan dan Manfaat Penelitian	30

3.1.1	Tujuan Penelitian	30
3.1.2	Manfaat Penelitian	31
BAB IV METODE PENELITIAN		32
4.1	Diagram Alur Metodologi Penelitian	32
4.1.1	Diagram Alur Utama.....	32
BAB V HASIL DAN LUARAN YANG DICAPAI.....		34
5.1	Kegiatan dan Hasil Pelaksanaan.....	34
5.2	Luaran Yang Dicapai.....	36
5.3	Hasil Penelitian.....	38
BAB VI KESIMPULAN DAN SARAN.....		54
5.1	Kesimpulan.....	54
5.2	Saran	54
DAFTAR PUSTAKA		55
LAMPIRAN-LAMPIRAN		60

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	7
Tabel 2. 2 Klasifikasi Model berdasarkan SoTA	16
Tabel 4. 1 Penjelasan Diagram Alur Metodologi Penelitian.....	33
Tabel 5. 1 Karakteristik Dataset yang Digunakan	35
Tabel 5. 2 Luaran dan Target Capaian	38
Tabel 5. 3 Kumpulan Data Staf.....	39
Tabel 5. 4 Dataset Data Training	39
Tabel 5. 5 Dataset Data Testing	39
Tabel 5. 6 Atribut Dataset Data Training.....	40
Tabel 5. 7 Atribut Dataset Data Testing	41
Tabel 5. 8 Data Yang Sudah Ter-Encoder	43
Tabel 5. 9 Data Yang Sudah Terisi Dengan Mean.....	43
Tabel 5. 10 Data Dengan Variabel Yang Digunakan.....	45
Tabel 5. 11 Variabel Independen Data.....	46
Tabel 5. 12 Variabel Dependen Data	46
Tabel 5. 13 Evaluasi Model OLS	49
Tabel 5. 14 Nilai P-Values Dari Variabel Independen	50
Tabel 5. 15 Tabel VIF	52

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	32
Gambar 5. 1 Heatmap Correlation Dataset Data Training	44
Gambar 5. 2 Heatmap Correlation Dataset Data Testing.....	45
Gambar 5. 3 Visualisasi Penyebaran Data Training	47
Gambar 5. 4 Visualisasi Penyebaran Data Testing	47
Gambar 5. 5 Grafik Asumsi Linear.....	51
Gambar 5. 6 Grafik Distribusi Residual.....	51
Gambar 5. 7 Homoskedastisitas.....	52

DAFTAR LAMPIRAN

Lampiran 1 Susunan Organisasi Tim Peneliti dan Pembagian Tugas	60
Lampiran 2 Biodata Ketua dan Anggota Pengusul	61
Lampiran 3 Surat Pernyataan Bebas Plagiat Ketua Penelitian	70
Lampiran 4 Penggunaan Anggaran	71
Lampiran 5 Bukti Penerimaan Artikel Ilmiah (LOA) atau URL dan Screenshoot Halaman Jurnal yang Sudah Dipublikasi	72
Lampiran 6 Format Catatan Harian (Logbook)	73
Lampiran 7 Poster	75

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Kemajuan ilmu pengetahuan dan teknologi maju lebih cepat dari sebelumnya sepanjang revolusi industri. Revolusi Industri 4.0 mulai terwujud berkat internet of things yang berfungsi sebagai pusat komunikasi serta mobilitas manusia dan mesin.[1]. Akibatnya, teknologi digital dan internet bergabung dengan industri tradisional, dengan tujuan untuk meningkatkan produksi, efisiensi, dan layanan pelanggan[2]. Era revolusi ini akan mendisrupsi berbagai kegiatan diberbagai bidang seperti pada bidang teknologi, ekonomi, sosial, dan politik[1]. Kita berada di fase awal revolusi yang secara fundamental akan mengubah cara kita berkomunikasi, hidup, dan bekerja[3].

Salah satu dampak nyata dari masuknya revolusi industri 4.0 adalah perubahan karakteristik ketenagakerjaan[4]. Pekerjaan yang ada akan terganggu dan digantikan oleh pekerjaan dengan fitur baru ketika karakteristik pekerjaan berubah[5]. Sebagai akibat dari tuntutan pekerjaan yang meningkat, karyawan juga membutuhkan kemampuan baru[6]. Secara alami, bisnis harus siap bersaing dengan orang lain[7]. Selain itu, untuk dapat bersaing, bisnis perlu memiliki keunggulan dan manajemen yang efisien[7]. Dengan demikian, salah satu faktor yang secara signifikan mempengaruhi perkembangan dan keberhasilan perusahaan adalah pekerjaan personelnnya[7]. Walaupun perusahaan tersebut memiliki teknologi yang canggih, namun tidak terdapat tenaga kerja didalamnya, perusahaan tidak akan dapat mencapai tujuannya[7].

Oleh karena itu, pengaturan kompensasi yang layak bagi karyawan merupakan salah satu elemen yang memiliki dampak internal terhadap perkembangan bisnis. Selain itu, bisnis harus siap untuk membayar kompensasi bonus kepada staf yang berkinerja baik dan sesuai dengan persyaratan bisnis. 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. Tujuan luaran dari penelitian ini adalah menghasilkan jurnal nasional SINTA 3, HAKI, dan buku ber-ISBN yang dimanfaatkan sebagai sarana pembelajaran dan praktikum pada mata kuliah Database pada Program Studi Diploma 4 Teknik Informatika Universitas Logistik dan Bisnis Internasional.

1.2 Identifikasi Masalah

Masalah dalam penelitian ini diidentifikasi 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

Berikut ini menjelaskan ruang lingkup penelitian.

- a) Data yang digunakan dalam penelitian ini berasal dari data karyawan di suatu perusahaan karena kondisi pandemi.
- b) Periode waktu 1 tahun.
- c) Bahasa pemrograman menggunakan Python.
- d) *Software* yang digunakan Jupyter atau dan Google Colab.

1.5 Sistematika Penulisan

Penyusunan laporan yang cermat menjadi lima bab ini didasarkan pada sejarah dan rumusan masalah yang diangkat di atas, khususnya :

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 menjelaskan bagaimana memecahkan masalah saat ini dan keuntungan dari penelitian yang dilakukan.

BAB IV METODOLOGI PENELITIAN

Bab ini memberikan gambaran tentang flowchart metodologi penelitian serta tahapan-tahapan flowchart penelitian yang harus diselesaikan agar penelitian dapat diselesaikan dan menghasilkan hasil yang diinginkan.

BAB V HASIL DAN LUARAN YANG DICAPAI

Bab ini memberikan penjelasan tentang temuan dan kesimpulan yang diambil dari penelitian yang dilakukan.

BAB II

TINJAUAN PUSTAKA

2.1 State-of-The-Art

Revolusi industri keempat melihat kemajuan yang signifikan dalam ilmu pengetahuan. Kita mungkin melihat pergeseran cara kita hidup, bekerja, dan berinteraksi satu sama lain sebagai akibat dari Revolusi Industri 4.0 saat ini[3]. Transformasi lapangan kerja merupakan salah satu efek nyata dari masuknya Revolusi Industri 4.0[4]. Ciri-ciri pekerjaan baru akan mengganggu ciri-ciri pekerjaan lama[5]. Perusahaan tentunya harus memiliki keunggulan dan manajemen yang efisien jika ingin bersaing dengan bisnis lain[7]. Akibatnya, salah satu aspek internal yang mungkin berdampak adalah bagaimana perusahaan memutuskan untuk membayar personelnnya. 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 pembelajaran mesin semakin banyak digunakan di berbagai sektor untuk memodelkan atau mengantisipasi hal-hal seperti gaji dengan menghasilkan prediksi dan mengekstrak informasi dari data[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 untuk merekomendasikan studi tambahan bagi akademisi. Penelitian ini menggunakan teknik regresi linier multivariat untuk melakukan proses pemilihan data berdasarkan kriteria yang dipilih, dengan penekanan pada kebijakan dalam menentukan pemilihan remunerasi karyawan di suatu perusahaan berdasarkan faktor-faktor tertentu. 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 mempelajari banyak referensi tentang kemajuan teknologi pada revolusi industri 4.0 yang dapat mengubah karakteristik kerja serta pemodelan prediksi pembelajaran mesin[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	-	-	-	-

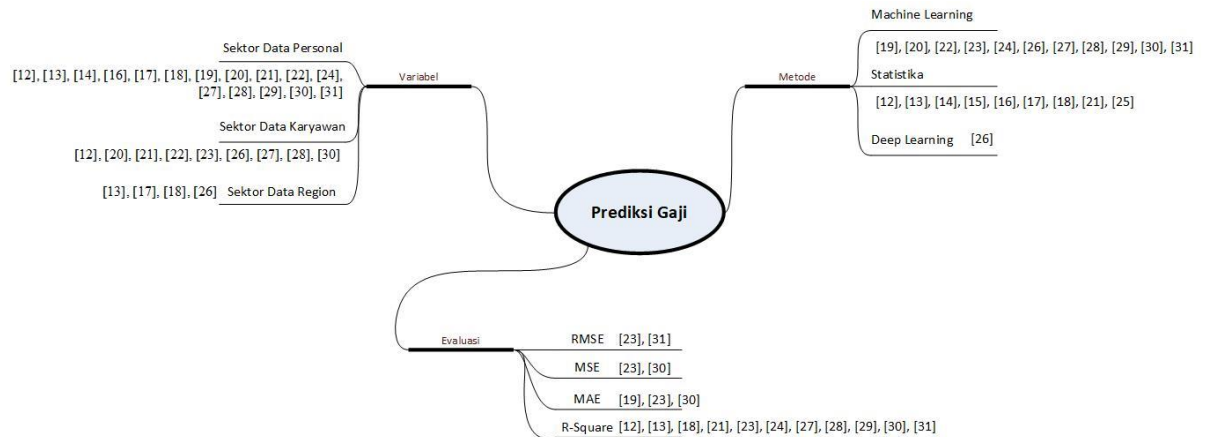
	berbasis <i>machine learning</i> untuk prediksi kenaikan gaji[27]		<i>Enterprise Resource Planning</i> (ERP)	<i>Deep Learning</i>	<i>artificial neural networks, random forest regression</i>	<i>machine learning</i> dengan menggunakan arsitektur <i>micro-services</i> .				
16.	Implementasi sistem prediksi 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.

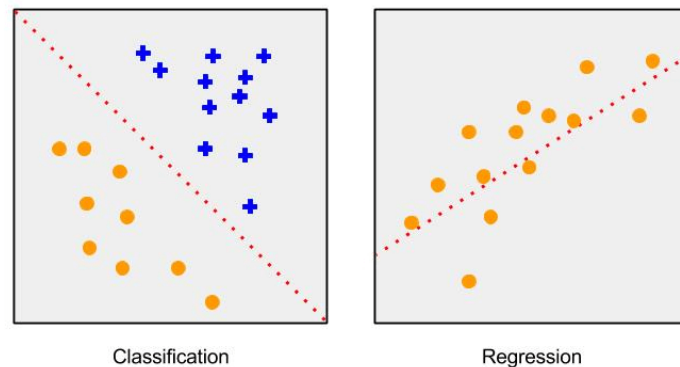
2.4 Machine Learning

Tujuan dari *machine learning* adalah untuk menciptakan sistem yang dapat belajar sendiri tanpa perlu terus menerus dilatih oleh manusia. *Machine Learning* adalah subbidang kecerdasan buatan yang berfokus pada pembelajaran dari data (*learn from data*). Sebelum menggunakannya untuk mengevaluasi hasil keluaran terbaik, pembelajaran mesin harus diberikan data yang dapat diandalkan untuk digunakan sebagai bahan pembelajaran.[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. Dalam *supervised learning*, ada dua variabel: variabel independen, juga dikenal sebagai variabel X,

dan variabel dependen, sering dikenal sebagai variabel Y. $Y = f$ adalah rumus yang digunakan untuk memetakan variabel X dan Y dalam banyak kasus (X). Untuk meramalkan variabel Y ketika menerima data masukan baru, fungsi pemetaan (f) diperkirakan menggunakan rumus metode pembelajaran terawasi ini (variabel X)..



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

Supervised learning dibagi menjadi dua jenis, yaitu :

1. Klasifikasi (*Classification*) : mengklasifikasikan *train data* ke dalam kategori tertentu dengan benar menggunakan *supervised learning*. Jenis ini dapat mengidentifikasi entitas tertentu dalam data dan membuat kesimpulan tentang bagaimana hal-hal itu harus dijelaskan atau diberi label..
2. Regresi (*Regression*) : memahami hubungan antara variabel dependen dan variabel independen melalui *supervised learning*. Biasanya, prediksi dibuat menggunakan bentuk regresi *supervised learning* ini.

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.

Berikut model *machine learning* yang digunakan dalam penelitian ini:

2.3.1 *Multivariate Linear Regression*

Pendekatan regresi yang dikenal sebagai regresi linier multivariat menggambarkan hubungan antara variabel respon (variabel dependen) dan variabel yang berdampak pada beberapa prediktor (variabel independen)[37]. Regresi linier adalah metode terbaik untuk digunakan ketika hasil, kelas, atau atribut numerik dan semua atribut numerik.

Pendekatan statistik utama adalah yang satu ini. Tujuannya adalah menggunakan rumus berikut untuk merepresentasikan kelas sebagai kombinasi linier kualitas dengan bobot yang telah ditentukan sebelumnya :

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

Keterangan :

- y = Variabel dependen
- X = Variabel independen
- a = Konstanta
- b = Koefisien regresi

Dimana $a, b_1, b_2, \dots b_n$ dihitung dengan metode persamaan sebagai berikut:

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

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

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

Nilai $1, b_1, b_2, \dots b_n$ juga dapat ditentukan dengan menggunakan pendekatan kuadran terkecil selain persamaan normal sebelumnya, 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 :

$$\sum X_1^2 = \sum X_1^2 - \frac{(\sum X_2)^2}{n} \quad (8)$$

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

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

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

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

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

2.3.2 Random Forest Classifier

Salah satu strategi *Decision Tree* adalah pendekatan *Random Forest*. Setiap pohon yang menguntungkan diintegrasikan ke dalam satu model dengan menggunakan hutan acak. Setiap pohon keputusan dalam hutan acak memiliki kedalaman maksimum dan bergantung pada nilai vektor acak dengan distribusi yang sama pada semua pohon. *Random forest* adalah pengklasifikasi yang mengambil bentuk pohon dimana setiap *tree* dalam suatu unit memilih input kelas yang paling umum θk dan dimana $\{h(x, \theta k), k = 1, \dots\}$ adalah vektor acak yang didistribusikan secara independen. Karakteristik akurasi *random forest* adalah sebagai berikut.

a. Pusat *random forest*

Satu set pelatihan untuk pengklasifikasi $h_1(x), h_2(x), \dots, h_k(x)$ berasal dari distribusi vektor acak dari variabel Y, X . Fungsi selanjutnya dibuat.

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

Fungsi untuk kesalahan.

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

Hasil dari menggabungkan banyak fungsi.

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

Temuan ini menjelaskan mengapa menambahkan *tree* ke *random forest* mencegahnya dari *overfitting* dan sebagai gantinya menghasilkan nilai yang hanya akurat untuk kesalahan.

b. Korelasi dan kekuatan

Hasilnya adalah fungsi.

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

Kekuatan tidak tergantung pada hutan dalam fungsi ini.

c. *Random Forest* memilih input secara acak

Untuk pemilihan fitur secara acak, bagging digunakan. Penggantian setiap training set diambil dari *training set* pertama. Set pelatihan kemudian ditanam dengan pohon menggunakan pemilihan fitur acak. Untuk meningkatkan akurasi saat fitur acak digunakan, bagging awalnya digunakan untuk tujuan ini. Kesalahan generalisasi pohon gabungan (PE^*) dan perkiraan kekuatan dan korelasinya dihitung menggunakan hasil bagging kedua. Sejumlah terbatas variabel *input* umum dipilih secara acak pada setiap *node* untuk membuat *random forest* paling dasar dengan karakteristik acak. membentuk pohon dengan ukuran sebesar mungkin menggunakan pendekatan CART.

d. *Input* yang digunakan oleh *random forest*

Jika ada beberapa input (M, F), mengambil proporsi dalam M akan menghasilkan peningkatan kekuatan tetapi korelasi tinggi. Definisi fitur tambahan menggunakan kombinasi acak linier dari sejumlah variabel input menghasilkan metode yang berbeda. Variabel L , yang mewakili jumlah variabel, adalah karakteristik. Koefisien dengan bilangan acak $[-1, 1]$ ditambahkan dengan variabel L , yang dipilih secara acak. Ini menghasilkan kombinasi linier F . Proses ini disebut sebagai *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, industry, atau social. Statistik berkaitan dengan berbagai metode, salah satunya yaitu regresi. Regresi adalah metode statistik untuk mencari tahu hubungan antara satu atau lebih variabel independen dan dependen. Metode ini juga dapat digunakan untuk menilai kemajuan hubungan antara variabel dan proyeksi masa depan[39]. Model statistik yang digunakan dalam penelitian ini adalah sebagai berikut :

2.5.1 *Simple Linear Regression*

Regresi linier sederhana adalah teknik paling sederhana untuk memodelkan hubungan antara satu variabel terikat dan satu variabel bebas. Regresi adalah proses menggunakan variabel independen untuk menjelaskan variabel dependen. Karena hubungan antarvariabel dalam analisis regresi sederhana bersifat linier, maka perubahan variabel X seringkali bersamaan dengan perubahan variabel Y. Namun, pada hubungan non-linier, perubahan variabel X tidak secara proporsional diikuti oleh perubahan variabel Y. Berikut ini adalah model sederhana untuk analisis regresi linier :

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

Keterangan :

Y = Variabel Dependen

a = Konstanta

b = Koefisien Regresi

Regresi linier sederhana memiliki tiga bagian: a adalah intersep, b adalah kemiringan, dan x adalah indeks waktu. Untuk menentukan nilai a dan b , gunakan persamaan :

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

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

Berikut adalah tahapan pendekatan yang disarankan berdasarkan regresi linier dasar :

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

Berikut ini adalah langkah-langkah untuk membuat model :

1. Hitung X^2 , Y^2 , XY dan jumlah masing-masing.
2. Hitung a dan b masing-masing menggunakan persamaan (2) dan (3).
3. Buat model persamaan dasar persamaan regresi linear.
4. Memprediksi atau meramalkan variabel-variabel yang mewakili faktor-faktor penyebab atau variabel-variabel akibat[40].

2.5.2 *Quantile Regression*

Kuantil adalah titik distribusi yang mewakili peringkat nilai dalam distribusi. 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*

Efektivitas model pembelajaran mesin dalam peramalan dievaluasi. Dengan memeriksa nilai kesalahan yang ditemukan dalam model prediksi, model ini dinilai. *Root Mean Square Error* (RMSE), *Mean Square Error* (MSE), dan *Mean Absolute Error* merupakan tiga faktor perhitungan untuk melihat nilai *error* (MAE). Spesifiknya ditunjukkan di bawah ini :

2.6.1 *Root Mean Square Error* (RMSE)

Teknik pengukuran yang disebut *Root Mean Square Error* (RMSE) memperkirakan nilai yang diamati dengan membandingkannya dengan nilai yang diantisipasi dari suatu model. Pendekatan estimasi akar dengan mean square error (RMSE) kurang dari 0,5 dianggap lebih akurat. Rumus RMSE ditunjukkan di bawah ini[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)

Kesalahan kuadrat rata-rata adalah perbedaan antara nilai aktual dan nilai yang diproyeksikan dalam rumus kuadrat rata-rata (MSE). Teknik MSE biasanya digunakan untuk menentukan nilai kesalahan peramalan yang diantisipasi. Temuan peramalan konsisten dengan data nyata dan dapat diterapkan untuk memprediksi perhitungan untuk periode mendatang ketika nilai MSE rendah atau mendekati nol[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)

Model evaluasi yang disebut *mean absolute error* (MAE) menampilkan kesalahan rata-rata, atau perbedaan antara nilai aktual dan nilai yang diantisipasi. Semakin akurat model dalam membuat prediksi, semakin rendah angka MAE yang harus dimiliki. Rumus MAE terlihat seperti ini[43] :

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

Keterangan :

Y' = Nilai Prediksi

Y = Nilai Sebenarnya

n = Jumlah Data

2.6.4 OLS Regression (Uji Asumsi)

Regresi adalah metode penelitian yang kuat yang dapat memeriksa banyak variabel sekaligus dan memberikan jawaban atas masalah penelitian yang menantang. Pada dasarnya, jika hasilnya cocok dengan Ordinary Least Square, dapat

dikatakan yakin terhadap OLS. Asumsi model regresi dapat diperiksa dengan menggunakan analisis Uji Asumsi. OLS mengandaikan bahwa variabel memiliki hubungan linier satu sama lain. OLS bukanlah teknik yang baik untuk analisis penelitian jika hubungannya tidak linier. OLS sering digunakan untuk memperkirakan berbagai parameter hubungan fungsional[44][45].

2.6.4.1 Uji Linearitas

Salah satu uji hipotesis tradisional adalah uji linieritas, yang digunakan untuk memastikan apakah distribusi data antara variabel X dan Y linier. Sangat penting untuk memahami bagaimana linearitas hubungan X dan Y mempengaruhi tingkat validitas model regresi. Ini dapat ditemukan dengan memeriksa distribusi data pada sumber diagonal grafik. Model regresi normal dan sesuai untuk digunakan dalam memprediksi variabel bebas dan sebaliknya jika menyebar dan mengikuti garis diagonal[46].

2.6.4.2 Uji Normalitas

Uji normalitas digunakan untuk menentukan apakah suatu kumpulan data atau distribusi data variabel terdistribusi normal atau tidak. Pengujian ini akan menunjukkan apakah nilai residual berdistribusi normal atau tidak. Regresi dengan nilai residual yang terdistribusi secara teratur merupakan model regresi yang baik. Ini dapat ditemukan dengan menggunakan distribusi data grafik pada sumber diagonal sebagai dasar untuk penilaian[47]. Dimungkinkan untuk mengevaluasi apakah residu terdistribusi normal dengan memasukkannya ke dalam histogram, memeriksa nilai-p dan uji normalitas Anderson-Darling, dan membandingkannya dengan batas 0,05. Nilai p dihitung menggunakan fungsi `ad() normal` dari `statsmodel`. Jika nilai p yang dihasilkan melebihi cutoff, residu dapat dianggap terdistribusi secara teratur. Jika nilai p yang dihasilkan jatuh di bawah cutoff, residu mungkin dianggap tidak terdistribusi normal.

Dapat diketahui hipotesa sebagai berikut :

- a. H_0 = Residual terdistribusi normal.
- b. H_1 = Residual terdistribusi secara tidak normal.

2.6.4.3 Uji-t

Untuk memastikan bagaimana setiap variabel independen mempengaruhi variabel dependennya sendiri, digunakan uji-t, yang juga dikenal sebagai uji parsial. Nilai p, seperti uji-F, menunjukkan kemungkinan mengamati hasil ekstrem yang serupa dengan yang diprediksi oleh model. Dengan menerapkan properti pvalues pada model, uji-t tambahan dapat memperoleh nilai-p untuk semua variabel[46].

Hipotesa yang dapat diambil adalah :

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

2.6.4.4 Koefisien Determinas (*R-Square*)

Pengukuran yang digunakan untuk menentukan seberapa besar kontribusi variabel X terhadap hasil Y adalah analisis koefisien determinasi (*R-Square*). Analisis ini digunakan untuk menghitung proporsi pengaruh variabel independen terhadap variabel dependen secara bersamaan[46].

2.6.4.5 Uji F-(ANOVA)

Uji F, juga dikenal sebagai uji Model atau uji Anova. Ini adalah tes untuk menentukan dampak gabungan dari semua faktor independen pada variabel dependen. Uji-F memungkinkan seseorang untuk mengevaluasi penerapan pemodelan dengan menghitung probabilitas pengamatan statistik-F yang setidaknya setinggi nilai yang dicapai oleh model bawaan. Properti .fvalues dan .pvalues dari model yang dikembangkan dapat diakses untuk mendapatkan statistik F dan probabilitasnya, seperti halnya skor R^2 .

Fs adalah hasil akhir dari analisis ANOVA. Nilai p dikontraskan dengan nilai Fs ini, juga dikenal sebagai F yang dihitung dalam uji hipotesis. Dapat disimpulkan bahwa baik variabel bebas maupun variabel terikat memiliki pengaruh yang signifikan terhadap permintaan secara bersamaan jika Fs melebihi nilai P. Tabel ANOVA tersebut di atas memberikan kesimpulan sebagai berikut :

- a. H_0 = Variabel independen tidak dapat sepenuhnya menjelaskan variabel dependen (Model tidak cocok).
- b. H_1 = Variabel dependen dijelaskan secara signifikan oleh variabel independen secara bersamaan (Model fit)[46].

2.6.4.6 Uji Multikolinearitas

Ketika variabel independen dalam model regresi memiliki hubungan linier yang sempurna atau hampir sempurna, ini disebut sebagai multikolinearitas. Jika terdapat fungsi linier yang sempurna pada beberapa atau semua variabel bebas dalam fungsi linier tersebut, maka model regresi dikatakan memiliki multikolinearitas. Uji multikolinearitas dilakukan untuk mengetahui apakah variabel independen dalam model regresi saling berhubungan. Apakah semua atau sebagian variabel yang digunakan untuk mengkarakterisasi model regresi linier sudah sempurna ditentukan dengan uji multikolinearitas. Pengujian ini dapat dihitung dengan melihat nilai toleransi dan nilai *variance inflation factor* (VIF). Menghitung faktor inflasi varians, atau VIF, adalah cara pengujian dilakukan. Jika nilainya terpusat, gunakan VIF (*Variance Inflation Factor*). Toleransi model regresi dan parameter *variance inflation factor* (VIF) dapat diuji. Variabel berikut diperhitungkan ketika uji multikolinearitas memberikan penilaian.:

- a. Jika angka VIF kurang dari 10 atau nilai toleransi lebih besar dari 0,01 dikatakan tidak ada multikolinearitas.
- b. Jika nilai VIF atau nilai toleransi lebih dari 10 atau kurang dari 0,01, multikolinearitas dinyatakan[46]

2.6.4.7 Uji Autokorelasi

Ketika residual dari periode t dan periode sebelumnya memiliki korelasi dalam model regresi, ini dikenal sebagai autokorelasi ($t-1$). Model regresi bebas autokorelasi adalah model yang baik. Uji autokorelasi dapat dilakukan dengan pengujian Durbin Watson. Model stats yang dibangun dalam fungsi `durbin watson()` langkah ini akan digunakan untuk menghitung skor Durbin-Watson, yang kemudian akan dievaluasi berdasarkan kriteria berikut :

- a. Jika skor Durbin-Watson kurang dari 1,5 dan terdapat autokorelasi positif, maka asumsi tersebut tidak benar.
- b. Jika skor Durbin-Watson antara 1,5 dan 2,5, maka tidak ada autokorelasi dan kondisi terpenuhi.
- c. Jika skor Durbin-Watson lebih besar dari 2,5, maka terjadi autokorelasi negatif, maka anggapan ini tidak benar[46][48].

2.6.4.8 Uji Heteroskedastisitas

Uji homoskedastisitas digunakan untuk menguji kesalahan dalam model statistik untuk menentukan apakah faktor lain berdampak pada varians atau variasi kesalahan. Ketika memvisualisasikan residu, homoskedastisitas dapat ditemukan dengan memeriksa apakah varians tampak seragam. Apabila terdapat ketidaksamaan varians dari residual untuk semua data dalam model regresi, dikatakan telah terjadi heteroskedastisitas[46].

BAB III

TUJUAN DAN MANFAAT

3.1 Tujuan dan Manfaat Penelitian

Tujuan dan sasaran yang ingin dicapai melalui penelitian disebutkan secara tepat dalam uraian tujuan penelitian. Beckingham (1971) menegaskan bahwa tujuan penelitian adalah penjelasan mengapa penelitian itu dilakukan. Tujuan suatu penelitian dapat mendefinisikan suatu gagasan dan memperjelas suatu masalah atau solusi, menunjukkan jenis penyelidikan yang harus dilakukan. Tujuan penelitian menentukan jalannya penelitian dan apa yang harus dicapai.

Dampak dari pencapaian tujuan penelitian adalah keuntungannya. Sugiyono (2011) mengklaim bahwa keunggulan penelitian adalah solusi atas pertanyaan yang diajukan oleh tujuan penelitian dan disajikan dalam temuan penelitian untuk membangun basis pengetahuan untuk memahami, menyelesaikan, dan meramalkan masalah yang telah terbentuk dalam topik penelitian. .

Berikut ini akan dijelaskan tujuan dan manfaat penelitian sehubungan dengan hal tersebut.

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

Kelebihan dari penelitian ini adalah 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*.

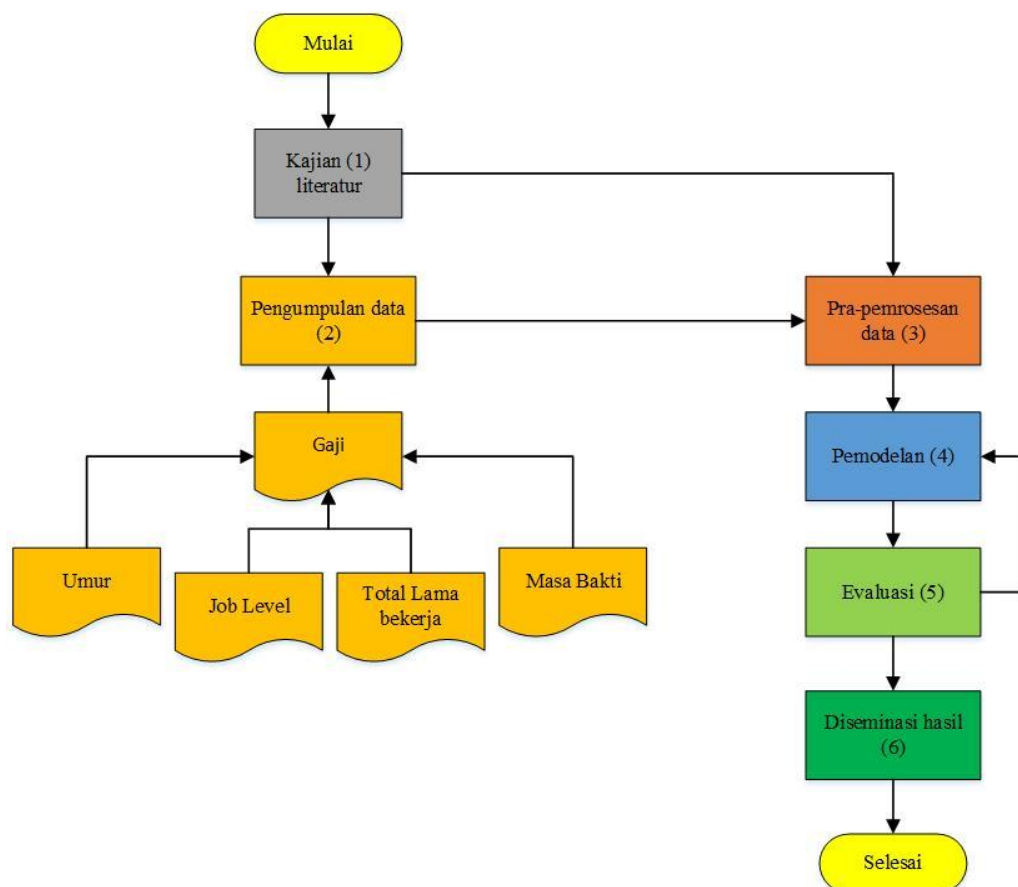
BAB IV

METODE PENELITIAN

4.1 Diagram Alur Metodologi Penelitian

Pada hakikatnya, teknik penelitian adalah suatu pendekatan metodis dalam pengumpulan data dengan tujuan dan keunggulan yang telah ditentukan. Untuk melakukan ini, kita membutuhkan taktik yang terhubung dengan hasil yang diinginkan. Teknik penelitian menurut Sugiyono (2017:2) pada dasarnya adalah proses ilmiah yang digunakan untuk mengumpulkan data untuk tujuan dan aplikasi tertentu. Sebuah diagram alur penelitian dapat digunakan sebagai metodologi 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

Tindakan yang diambil dan hasil dari pelaksanaan program hibah penelitian internal ini dimaksudkan untuk menghasilkan efek yang diinginkan. Berikut ini adalah deskripsi dari tindakan dan hasil tersebut.

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

Seperti yang telah dikemukakan di latar belakang, tujuan dari tindakan yang dilakukan dan hasil yang dicapai sebagai konsekuensi dari pelaksanaan program hibah penelitian internal adalah untuk menghasilkan hasil yang diinginkan. Keberhasilan keluaran dari program ini dapat diringkas sebagai berikut dengan mempertimbangkan tujuan keluaran saat ini.

1. Karya ilmiah yang dimuat di jurnal nasional

Penelitian ini akan memiliki artikel ilmiah di jurnal nasional yang diterbitkan untuk target audiens jurnal nasional SINTA S3. Versi draft jurnal penelitian ini telah dibuat. Karena belum ada tambahan anggota tim review atau menjadi bahan diskusi kelompok, draftnya belum selesai.

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. Produksi ini, yaitu penerbitan buku, sesuai dengan tujuan.

3. HAKI

HAKI pada penelitian 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. Karena kurangnya lebih banyak percakapan, tata letak belum disesuaikan lebih lanjut juga. Hasil ini sesuai dengan tujuan, yaitu menghasilkan 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.

Seperti yang dikatakan sebelumnya, tindakan yang telah diambil. Temuan-temuan yang diperoleh sehubungan dengan tujuan pencapaian dapat ditampilkan dalam format tabel di bawah ini.

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

Tabel di atas menjelaskan mengapa output dari 4 (empat) kategori output tersebut belum sepenuhnya mencapai tingkat tujuan. Tentunya, capaian tersebut akan terus dikembangkan dengan terus melakukan koreksi dan diskusi untuk mencapai target yang telah ditetapkan.

5.3 Hasil Penelitian

Fokus dari beberapa studi tentang perkiraan kompensasi karyawan, peneliti berpendapat bahwa ada hubungan positif dan substansial antara gaji dan masa kerja pada kinerja karyawan[49]. 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*. Model yang disarankan adalah regresi linier multivariat atau model regresi linier berganda

(MLR). Data penelitian ini berasal dari kumpulan data staf yang dibagi menjadi kumpulan pelatihan dan tes, dan diperoleh melalui Kaggle.

Tabel 5. 3 Kumpulan Data Staf

Age	Attrition	BusinessTravel	...	YearsAtCurrManager
50	No	Travel_Rarely	...	3
36	No	Travel_Rarely	...	1
21	Yes	Travel_Rarely	...	0
50	No	Travel_Frequently	...	7
...
50	No	Travel_Rarely	...	0

Hasil model prediksi dengan pendekatan regresi linier berganda diperoleh dengan mengolah data tidak berlabel menggunakan teknik *unsupervised learning*. Berikut adalah pemaparan data yang digunakan.

Tabel 5. 4 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. 5 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

Tabel 5.6 dan 5.7 memberikan karakteristik berikut dari data tersebut.

Tabel 5. 6 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. 7 Atribut Dataset Data Testing

Atribut	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.

Tabel 5. 8 Data Yang Sudah Ter-Encoder

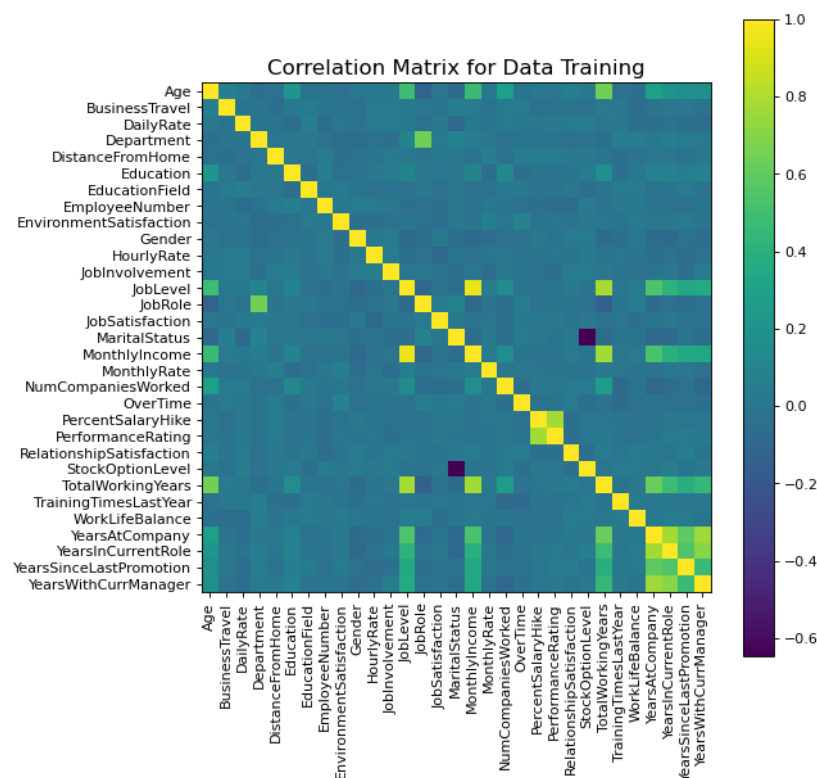
Age	Attrition	BusinessTravel	...	YearsAtCurrManager
50,0	0	2	...	3
36,0	0	2	...	1
21,0	1	2	...	0
50,0	0	1	...	7
...
50,0	0	2	...	0

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

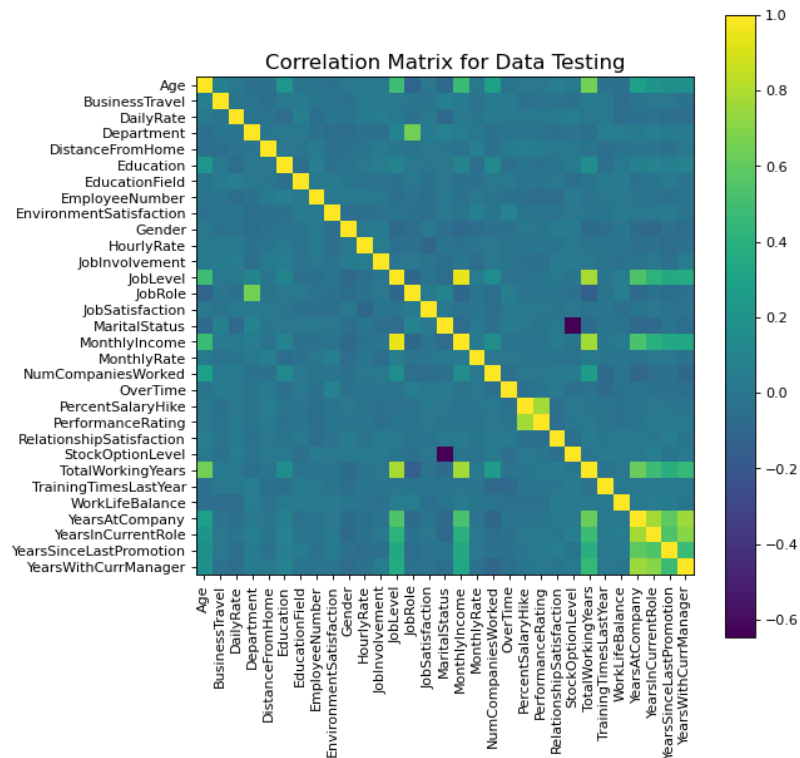
Tabel 5. 9 Data Yang Sudah Terisi Dengan Mean

Age	Attrition	BusinessTravel	DailyRate	...	YearsAtCurrManager
50,0	0	2	1126,0	...	3
36,0	0	2	216,0	...	1
21,0	1	2	337,0	...	0
50,0	0	1	1246,0	...	7
...
50,0	0	2	264,0	...	0

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 pada gambar 5.1 dan korelasi 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 lain memiliki hubungan yang lemah dengan 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.

Tabel 5. 10 Data Dengan Variabel Yang Digunakan

Age	JobLevel	MonthlyIncome	TotalWorkingYears	YearsAtCompany
50,0	4	17399	32	5
36,0	2	4941	7	3
21,0	1	2679	1	1

50,0	5	18200	32	32
...
50,0	5	19331	27	1

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

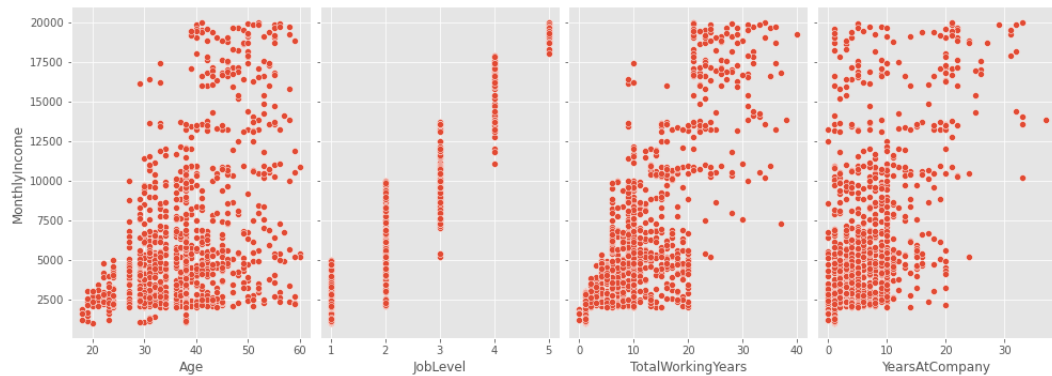
Tabel 5. 11 Variabel Independen Data

Age	JobLevel	TotalWorkingYears	YearsAtCompany
50,0	4	32	5
36,0	2	7	3
21,0	1	1	1
50,0	5	32	32
...
50,0	5	27	1

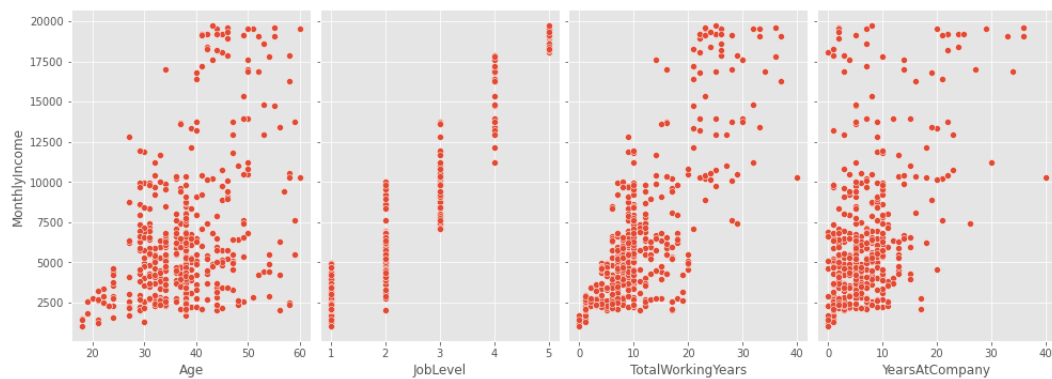
Tabel 5. 12 Variabel Dependen Data

No	MonthlyIncome
0	17399
1	4941
2	2679
3	18200
...	...
1029	19331

Distribusi data dari data pelatihan dan data pengujian kemudian ditampilkan menggunakan Seaborn. Gambar 5.3 untuk data pelatihan dan Gambar 5.4 untuk data pengujian keduanya menampilkan visualisasi.



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. 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. Intercept 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 dikurangi sebesar 1728 dari gaji sebelumnya per tahunnya. Proses selanjutnya adalah melakukan prediksi *data testing* menggunakan model *machine learning*. Kemudian buat kolom baru yang bernama MonthlyIncome Prediction yang berisikan nilai prediksi. Berikut persamaan umum dari model linear regresi multivariable.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \dots + \beta_nX_n$$

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 terikat yaitu MonthlyIncome

X1 = Variabel bebas pertama yaitu Age

X2 = Variabel bebas kedua yaitu JobLevel

X3 = Variabel bebas ketiga yaitu TotalWorkingYears

X4 = Variabel bebas 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.

Berdasarkan uji R-Square yang dibuat didapat nilai akurasi sebesar 0.90 atau 90%. Maka, MonthlyIncome dipengaruhi oleh faktor Age dan YearsAtCompany sebesar 0,909 atau 90,9%. Nilai residual koefisien determinasi atau 9,1% adalah 0,091, dan dipengaruhi oleh variabel yang tidak teridentifikasi lebih lanjut.

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. Berikut adalah tabel *summary* dari model OLS.

Tabel 5. 13 Evaluasi Model OLS

OLS Regression Results						
Dep. Variable	:	MonthlyIncome	R-Squared	:	0,909	
Model	:	OLS	Adj. R-Squared	:	0,909	
Method	:	Least Squares	F-Statistic	:	2571.	
Date	:	Mon, 10 Jan 2022	Prob (F-Statistic)	:	0,00	
Time	:	10:58:11	Log-Likelihood	:	-8944,9	
No.Observations	:	1029	AIC	:	1,790 x 104	
Df Residuals	:	1024	BIC	:	1,792 x 104	
Df Model	:	4				
Covariance Type	:	nonrobust				
Variable	Coef	Std. Err	t	P > t	[0,025	0,975]
const	-1728,52	230,58	-7,46	0,00	-2180,99	-1276,04
Age	-5,05	6,90	-0,73	0,46	-18,60	8,49
JobLevel	3871,75	65,63	58,98	0,00	3742,95	4000,54
TotalWorkingYears	46,94	11,73	4,01	0,00	23,91	69,96
YearsAtCompany	-9,84	9,76	-1,01	0,31	-29,01	9,32
Omnibus	:	12,798	Durbin-Watson	:	2,096	
Prob (Omnibus)	:	0,002	Jarque-Bera (JB)	:	15,262	
Skew	:	0,182	Prob (JB)	:	0,000485	
Kurtosis	:	3,472	Cond. No.	:	213	

Berdasarkan Tabel 5.13, 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. Jika kemungkinan F-statistik adalah 0,00

atau di bawah (di bawah 0,05), model tersebut signifikan dalam memprediksi variabel dependen (MonthlyIncome).

Berdasarkan uji-F didapat nilai F_s sebesar 2570,622 dan P-value sebesar 0,00. Dari hasil tersebut 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 tidak diperhitungkan karena nilai probabilitasnya adalah 0,00 yang lebih kecil dari 5%. Akibatnya, dapat diklaim bahwa model yang dipilih sesuai.

Tabel 5. 14 Nilai P-Values Dari Variabel Independen

No	Variabel	Nilai Uji-t
1	Age	$2,643777 \times 10^{-1}$
2	JobLevel	0
3	TotalWorkingYears	$6,771641 \times 10^{-5}$
4	YearsAtCompany	$3,136608 \times 10^{-1}$

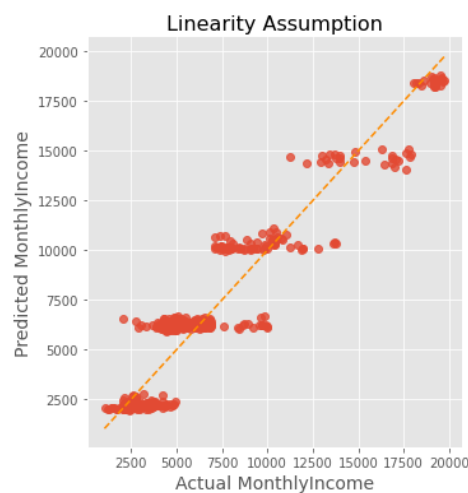
Berdasarkan uji-t pada Tabel 5.14, dapat diambil hipotesa sebagai berikut :

- 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 .

Variabel independen JobLevel dan TotalWorkingYears merupakan faktor yang tidak mempengaruhi variabel dependen, sesuai dengan hipotesis di atas. Sedangkan variabel terikat dipengaruhi oleh variabel bebas Age dan

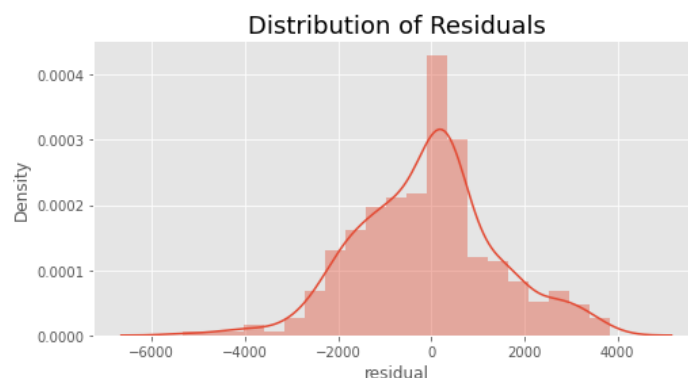
YearsAtCompany. Selanjutnya adalah menambahkan satu kolom baru dengan nama MonthlyIncome Prediction yang berisikan dengan hasil prediksi. Kemudian dibuat juga kolom residual untuk menyimpan nilai residualnya.

Uji asumsi pertama adalah uji linieritas. Tes ini menetapkan apakah hubungan antara variabel independen dan dependen adalah linier. Sebuah plot pencar digunakan untuk melakukan uji linieritas sehingga perbedaan antara nilai yang diantisipasi dan nilai yang sebenarnya dapat diperhatikan.



Gambar 5. 5 Grafik Asumsi Linear

Uji normalitas dilakukan setelah uji linieritas. Berikut adalah grafik uji normalitas di bawah ini :



Gambar 5. 6 Grafik Distribusi Residual

Berdasarkan Gambar 5.6, nilai p yang ditentukan oleh pendekatan Anderson-Darling adalah 0,00032261. Fakta bahwa angka ini lebih kecil dari titik

potong 0,05 yang ditentukan menunjukkan bahwa H_0 menerima H_1 , atau bahwa residu tidak terdistribusi normal. Oleh karena itu, dapat dikatakan bahwa praduga normal adalah benar.

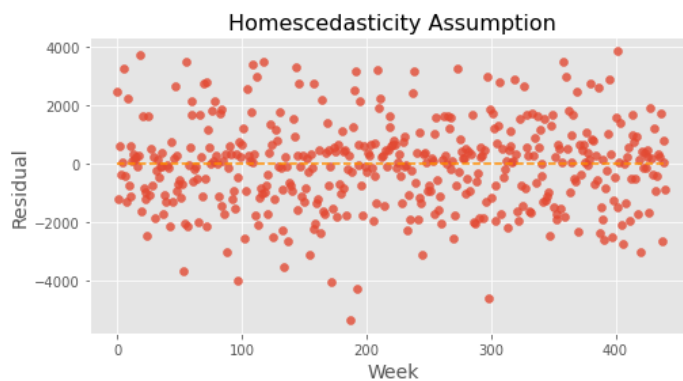
Uji selanjutnya adalah uji multikolinearitas. Berikut adalah tabel multikolinearitas variabel-variabel independent :

Tabel 5. 15 Tabel VIF

No	Variabel	VIF
1	Intercept	28,655370
2	Age	1,690786
3	JobLevel	2,489052
4	TotalWorkingYears	4,140803
5	YearsAtCompany	1,739893

Nilai Age, JobLevel, TotalWorkingYears, dan YearsAtCompany semuanya kurang dari 10, maka dengan menggunakan taraf signifikansi 0,05 dapat ditarik kesimpulan dari Tabel 5.15 di atas bahwa data tersebut tidak terdapat multikolinearitas.

Berdasarkan uji autokorelasi 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.



Gambar 5. 7 Homoskedastisitas

Gambar 5.12 menunjukkan bagaimana titik-titik ditempatkan secara acak di sepanjang sumbu Y, di atas dan di bawah nilai 0 (nol). Dapat dikatakan bahwa tidak terdapat indikasi heteroskedastisitas dalam model regresi yang digunakan.

BAB VI

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Beberapa kesimpulan yang dapat diambil dari hasil penelitian dan pembahasan yang telah dilakukan, antara lain sebagai berikut :

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 berikut dapat diberikan kepada peneliti yang akan melakukan dan melanjutkan penelitian ini :

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,” *J Biomed Inform*, 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,” *J Hydrol (Amst)*, 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,” *Inf Sci (N Y)*, 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,” *J Multivar Anal*, 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," *Top Companion Anim Med*, 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. Elshoryi, A. Alathamneh, I. Mahmoud, and F. Hammad, "Association of salary and intention to stay with the job satisfaction of the dietitians in Jordan: A cross-sectional study," *Health Policy Open*, vol. 3, Dec. 2022, doi: 10.1016/j.hpopen.2021.100058.
- [15] D. Webbink and J. Hartog, "Can students predict starting salaries? Yes!," *Econ Educ Rev*, vol. 23, no. 2, pp. 103–113, 2004, doi: 10.1016/S0272-7757(03)00080-3.
- [16] S. Gosh, K. Rascati, A. Shah, and P. Peebles, "PHP88 - Predictors of Annual Salary for Health Economics, Outcomes Research, and Market Access Professionals," *Value in Health*, vol. 21, p. S101, 2018, doi: <https://doi.org/10.1016/j.jval.2018.04.682>.
- [17] R. Marrero-Rodríguez, S. Morini-Marrero, and J. M. Ramos-Henriquez, "Tourism jobs in demand: Where the best contracts and high salaries go at online offers," *Tour Manag Perspect*, vol. 35, Jul. 2020, doi: 10.1016/j.tmp.2020.100721.
- [18] M. L. Blackburn, "Are U.S. teacher salaries competitive? Accounting for geography and the retransformation bias in logarithmic regressions," *Econ Educ Rev*, vol. 84, Oct. 2021, doi: 10.1016/j.econedurev.2021.102169.
- [19] G. A. Gilpin, "Teacher salaries and teacher aptitude: An analysis using quantile regressions," *Econ Educ Rev*, vol. 31, no. 3, pp. 15–29, Jun. 2012, doi: 10.1016/j.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," *Int J Environ Res Public Health*, vol. 18, no. 3, pp. 1–10, Feb. 2021, doi: 10.3390/ijerph18031322.
- [23] Sayan Das, Rupashri Barik, and Ayush Mukherjee, "Salary Prediction Using Regression Technique," *International Conference On Industry Interactive Innovations In Science, Engineering And Technology*, 2020, doi: <https://dx.doi.org/10.2139/ssrn.3526707>.
- [24] U. Bansal, A. Narang, A. Sachdeva, I. Kashyap, and S. P. Panda, "Empirical analysis of regression techniques by house price and salary prediction," in *IOP Conference Series: Materials Science and Engineering*, Jan. 2021, vol. 1022, no. 1. doi: 10.1088/1757-899X/1022/1/012110.
- [25] S. Gupta, "A Regression Modeling Technique on Data Mining," *Int J Comput Appl*, vol. 116, no. 9, pp. 975–8887, 2015, [Online]. Available: <http://www.nag.co.uk/stats/GDGE>
- [26] J. Prakash Sharma and N. Bajpai, "Salary Satisfaction as an Antecedent of Job Satisfaction: Development of a Regression Model to Determine the Linearity between Salary Satisfaction and Job Satisfaction in a Public and a Private Organization," *European Journal of Social Sciences*, vol. 18, no. 3, 2011.
- [27] Y. Gormez, H. Arslan, S. Sari, and M. Danis, "SALDA-ML: Machine Learning Based System Design to Predict Salary Increase," *Advances in Artificial Intelligence Research*, vol. 2, no. 1, pp. 15–19, Jan. 2022, doi: 10.54569/aair.1029836.
- [28] K. Pornthep and S. Pokpong, "Implement Of Salary Prediction System To Improve Student Motivation Using Data Mining Technique," *International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, 2016.
- [29] I. Martín, A. Mariello, R. Battiti, and J. Alberto Hernández, "Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study," *International Journal of Computational Intelligence Systems*, vol. 11, pp. 1192–1209, 2018, doi: <http://dx.doi.org/10.2991/ijcis.11.1.90>.
- [30] R. Voleti and B. Jana, "Predictive Analysis of HR Salary using Machine Learning Techniques," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 1, 2022, [Online]. Available: www.ijert.org
- [31] D. Sananda, H. Airiddha, and D. Kousik, "Design of a novel Prediction Engine for predicting suitable salary for a job," in *Fourth International*

Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 275–279. doi: <https://doi.org/10.1109/ICRCICN.2018.8718711>.

- [32] D. A. Gomez-Cravioto, R. E. Diaz-Ramos, N. Hernandez-Gress, J. L. Preciado, and H. G. Ceballos, “Supervised machine learning predictive analytics for alumni income,” *J Big Data*, vol. 9, no. 1, pp. 1–31, Dec. 2022, doi: 10.1186/s40537-022-00559-6.
- [33] M. R. Jennings *et al.*, “Code-free cloud computing service to facilitate rapid biomedical digital signal processing and algorithm development,” *Comput Methods Programs Biomed*, vol. 211, Nov. 2021, doi: 10.1016/j.cmpb.2021.106398.
- [34] N. Ahmed *et al.*, “Machine learning based diabetes prediction and development of smart web application,” *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229–241, Jun. 2021, doi: 10.1016/j.ijcce.2021.12.001.
- [35] 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 STEEM*, 2019, vol. 1, no. 1, pp. 1–19. Accessed: Aug. 14, 2022. [Online]. Available: <http://seminar.uad.ac.id/index.php/STEEM/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] M. Kuncoro, *Metode Kuantitatif: Teori Dan Aplikasi Untuk Bisnis Dan Ekonomi*. Yogyakarta, 2001. Accessed: Oct. 14, 2022. [Online]. Available: http://library.fip.uny.ac.id/opac/index.php?p=show_detail&id=1721
- [45] F. Pavelescu, “Features Of The Ordinary Least Square (OLS) Method Implications For The Estimation Methodology,” *Romanian Journal Of Economic Forecasting*, vol. 1, no. 2, pp. 85–101, 2004, Accessed: Oct. 14, 2022. [Online]. Available: https://econpapers.repec.org/article/rjrromjef/v_3a1_3ay_3a2004_3ai_3a2_3ap_3a85-101.htm
- [46] G. Mardiatmoko, “Pentingnya Uji Asumsi Klasik Pada Analisis Regresi Linier Berganda (Studi Kasus Penyusunan Persamaan Allometrik Kenari Muda [*Canarium Indicum L.*]),” *Jurnal Ilmu Matematika dan Terapan*, vol. 4, no. 3, pp. 333–342, 2020, doi: 10.30598/barekengvol14iss3pp333-342.
- [47] G. Imam, *Aplikasi Analisis Multivariat Dengan Program IBM SPSS 23*, vol. 8. Semarang, 2016. Accessed: Oct. 14, 2022. [Online]. Available: <https://onesearch.id/Record/IOS2863.JATEN000000000218217>
- [48] C. Trihendradi, *Langkah Mudah Menguasai Analisis Statistik Menggunakan SPSS 21*. Bandung, 2013. Accessed: Oct. 14, 2022. [Online]. Available: <https://onesearch.id/Record/IOS13258.ai:slims-83102>
- [49] 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)

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,



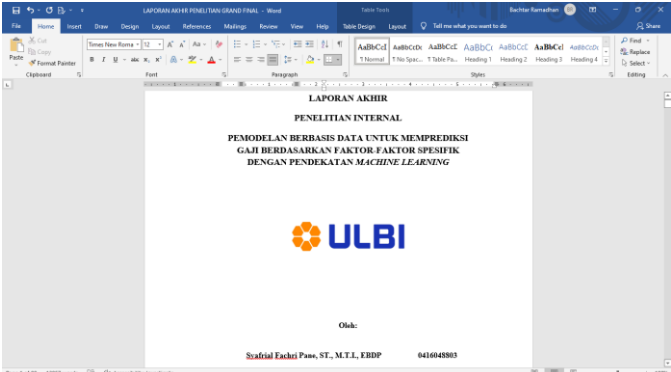

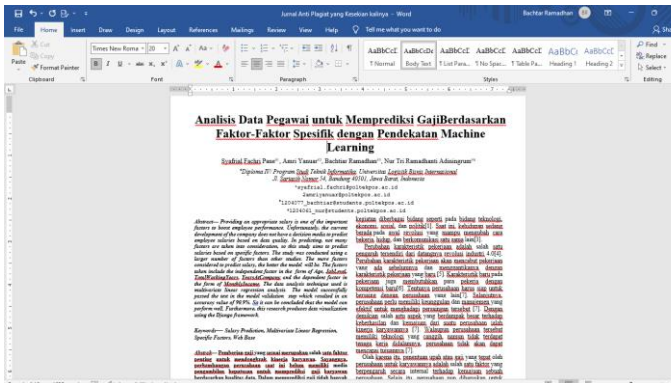
(Syafrial Fachri Pane.,S.T.,M.T.I.,EBDP)
NIK. 117.88.233

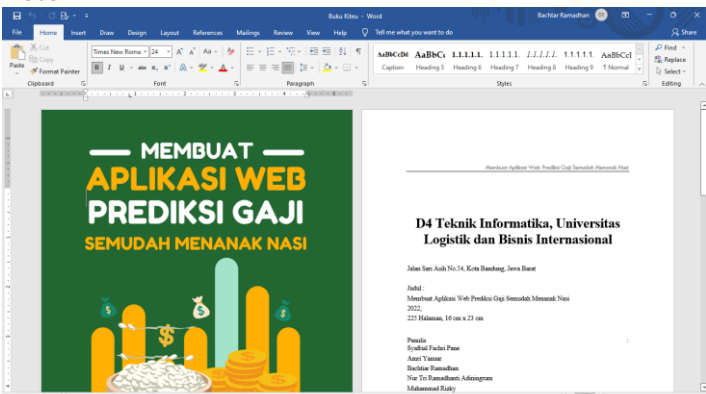
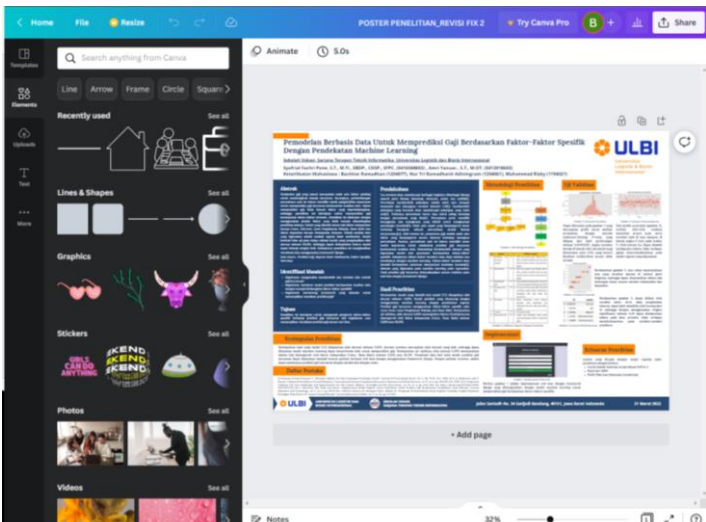
Lampiran 4 Penggunaan Anggaran

1. Gaji dan Upah				
Jabatan	Gaji & Upah/Jam	Waktu	Minggu	TOTAL (Rp)
	(Rp)	(Jam/Minggu)		
Ketua Peneliti	6500	6 Jam	16	832000
Anggota 1 (Dosen)	4500	4 Jam	16	576000
Anggota 2 (Dosen)	4500	4 Jam	16	350000
Anggota 3 (Dosen)	4500	4 Jam	16	350000
SUB TOTAL (Rp)				2108000
2. Bahan Habis Pakai				
Material	Justifikasi	Kuantitas	Harga Satuan	TOTAL (Rp)
	Pemakaian		(Rp)	
Kertas A4	Untuk Print Dokumen	5 Rim	50000	250000
Tinta Printer	Untuk Keperluan Print	4 Buah	200000	800000
Konsumsi	Rapat/Diskusi 8 Bulan	8 Bulan	200000	1600000
ATK	Administrasi	2 Pack	250000	500000
Sewa Zoom (100 Partisipan)	Meeting Online 8 Bulan	1 Paket	250000	250000
Subsidi Kuota Pula	Komunikasi	8 Bulan	50000	2550000
SUB TOTAL (Rp)				5950000
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	500000	500000
SUB TOTAL (Rp)				500000
TOTAL DANA YANG DIBUTUHKAN (Rp)				8558000

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>Hari Rabu, 17 Agustus 2022</p>	<p>Catatan : Pembuatan Laporan Akhir Penelitian.</p> <p>Dokumen Pendukung :</p> <ol style="list-style-type: none"> 1. Foto 
2	<p>Hari Sabtu, 6 Agustus 2022</p>	<p>Catatan : Pembuatan Power Point untuk Presentasi.</p> <p>Dokumen Pendukung :</p> <ol style="list-style-type: none"> 1. Foto 
3	<p>Hari Kamis, 1 September 2022</p>	<p>Catatan : Pembuatan Jurnal Ilmiah ter-akreditasi SINTA 3.</p> <p>Dokumen Pendukung :</p> <ol style="list-style-type: none"> 1. Foto 

<p>4</p> <p>Hari Selasa, 9 Agustus 2022</p>		<p>Catatan : Pembuatan Buku ber-ISBN.</p> <p>Dokumen Pendukung :</p> <p>1. Foto</p> 
<p>5</p> <p>Hari Sabtu, 20 Agustus 2022</p>		<p>Catatan : Pembuatan Poster Penelitian</p> <p>Dokumen Pendukung :</p> <p>1. Foto</p> 

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 Fachri 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

Pendapatan gaji yang sesuai merupakan salah satu faktor penting untuk mendorong kinerja karyawan. Sayangnya, perkembangan perusahaan saat ini belum memiliki media pengambilan keputusan untuk memprediksi gaji karyawan berdasarkan analisis data. Dalam memprediksi gaji, tidak banyak faktor yang diperhatikan, 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, level, lama pengalaman bekerja, masa haki dan faktor dependen berupa Pendapatan Bulanan. Teknik analisis data yang digunakan adalah analisis regresi linier multivariat. Model berhasil jika uji pada tahap validasi model yang menghasilkan nilai akurasi sebesar 90,9%. Sehingga dapat disimpulkan bahwa model dapat bekerja dengan baik. Selanjutnya penelitian ini menghasilkan visualisasi data menggunakan framework Django.

Kata Kunci: Prediksi Gaji, Regresi Linier Multivariat, Faktor Spesifik, Web User.

Identifikasi Masalah

- Bagaimana mengonversi karakteristik dan korelasi data terkait gaji karyawan?
- Bagaimana membuat model prediksi berdasarkan kualitas data dengan mempertimbangkan faktor-faktor spesifik?
- Bagaimana merancang framework yang dinamis untuk menampilkan visualisasi prediksi gaji secara real time.

Tujuan

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

Pendahuluan

Era revolusi akan mendorong berbagai kegiatan diberbagai bidang seperti pada bidang teknologi, ekonomi, sosial, dan politik[1]. Perubahan karakteristik pekerjaan adalah salah satu dampak terwujud dari datangnya revolusi industri 4.0[2]. Karakteristik pekerjaan yang berubah akan menimbulkan pekerjaan yang tidak ada[3]. Tentunya perusahaan harus siap untuk saling bersaing dengan perusahaan yang lain[4]. Perusahaan perlu mencari keunggulan dan manajemen yang efektif untuk menghadapi persaingan tersebut[5]. Salah satu aspek yang berpengaruh besar terhadap kinerja sebuah perusahaan adalah berupa 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 keputusan untuk melakukan prediksi gaji karyawan berdasarkan kualitas data. Metodologi Karakteristik dataset yang digunakan terdiri dari parameter berdasarkan faktor-faktor spesifik. Selanjutnya faktor-faktor tersebut akan diuji validitas dan korelasinya dengan machine learning. Faktor-faktor tersebut akan diambil berdasarkan pedoman interpretasi koefisien korelasi[7]. Metode yang digunakan pada machine learning yaitu regresi. Hasil prediksi gaji karyawan divisualisasikan secara real time pada web user 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 Haki. Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa Model Machine Learning dapat menghasilkan nilai faktor independen (Umur, Masa Haki) sebesar 0,909 atau 90,9%.

Metodologi Penelitian



Gambar 1. Metodologi Penelitian

No		Deskripsi	
1	Identifikasi masalah	1. Bagaimana mengonversi karakteristik dan korelasi data terkait gaji karyawan?	1. Bagaimana membuat model prediksi berdasarkan kualitas data dengan mempertimbangkan faktor-faktor spesifik?
2	Tujuan	2. Bagaimana merancang framework yang dinamis untuk menampilkan visualisasi prediksi gaji secara real time.	2. Bagaimana merancang framework yang dinamis untuk menampilkan visualisasi prediksi gaji secara real time.
3	Pendahuluan	3. Bagaimana mengonversi karakteristik dan korelasi data terkait gaji karyawan?	3. Bagaimana membuat model prediksi berdasarkan kualitas data dengan mempertimbangkan faktor-faktor spesifik?
4	Hasil Penelitian	4. 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 Haki. Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa Model Machine Learning dapat menghasilkan nilai faktor independen (Umur, Masa Haki) sebesar 0,909 atau 90,9%.	4. 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 Haki. Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa Model Machine Learning dapat menghasilkan nilai faktor independen (Umur, Masa Haki) sebesar 0,909 atau 90,9%.

Gambar 2. Indikator Capaian Tahapan Penelitian

Implementasi



Gambar 3. Implementasi Framework

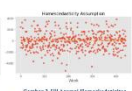
Berikut gambar 7 adalah implementasi web base dengan framework django yang diintegrasikan dengan model machine learning untuk memprediksi gaji berdasarkan faktor-faktor spesifik.

Uji Validitas



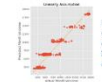
Gambar 4. Uji Asumsi Normalitas

Dapat dilihat pada gambar 2 yang merupakan grafik untuk melihat terdistribusi titik-titik residual 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 5. Uji Asumsi Homoskedastisitas

Dari grafik scatterplot (gambar 5), 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 disimpulkan bahwa tidak terdapat gejala heteroskedastisitas pada model regresi yang digunakan.



Gambar 6. Uji Asumsi Multikolinearitas

Berdasarkan gambar 5, plot sebar menunjukkan sisa yang tersebar merata di sekitar garis diagonal, sehingga dapat disimpulkan bahwa ada hubungan linear antara variabel independen dan dependen.

Berdasarkan gambar 5, dapat dilihat nilai variabel umur, level, lama pengalaman bekerja, masa haki 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.

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 berperforansi baik untuk memprediksi gaji. Berdasarkan uji validitas, nilai akurasi 0,909 menunjukkan bahwa Gaji dipengaruhi oleh faktor independen (Umur, Masa Haki) 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] Prasetyo R and Triyandani I. "Resensi Industri 4.0 dan Tantangan Perubahan Sosial", *Journal Of Proceedings Series*, No. 5, pp. 22-27, Nov. 2018. [2] A. A. Shabroun and K. Hamis. "Industrial Revolution 4.0 and Education", *International Journal of Academic Research in Business and Social Sciences*, vol. 4, no. 8, pp. 314-319, Oct. 2018. [3] S. Kerguelen. "Industry 4.0: New Challenge and Opportunities for The Labour Market", *Technology and Society Governance*, vol. 1, no. 4, pp. 4-6, 2017. <http://dx.doi.org/10.1016/j.tsg.2017.04.001>. [4] Y. Ahsanov A. Taib and A. Anas. "Implementasi Model Regresi Linier Sederhana Untuk Prediksi Gaji Berdasarkan Pengalaman Lama Bekerja", *Journal Education and Technology*, vol. 1, no. 2, pp. 56-70, 2019. [5] Tawfik A. S. Rongga Pratomo, Muband R. "Pengaruh Produktivitas Kerja Pegawai Terhadap Tingkat Kepuasan Pelayanan Pada Kantor PT. Taspen Cabang Manado", *Jurnal Administrasi Publik*, vol. 3, no. 46, pp. 1-9, 2017.



Jalan Sariasih No. 54 Sarjadi Bandung, 40151, Jawa Barat Indonesia

31 Maret 2022

