

**PENERAPAN INTELEJEN BISNIS DENGAN VISUALISASI
DATA GAJI DAN ALGORITMA REGRESI LINEAR**



Disusun Oleh:

N a m a : Haryadi Tri Nugroho
NIM : 20523097

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2024**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**PENERAPAN INTELEJEN BISNIS DENGAN VISUALISASI
DATA GAJI DAN ALGORITMA REGRESI LINEAR**

TUGAS AKHIR



Pembimbing I,


(Dr. Syarif Hidayat, S.KOM., M.IT.)

HALAMAN PENGESAHAN DOSEN PENGUJI

PENERAPAN INTELEJEN BISNIS DENGAN VISUALISASI
DATA GAJI DAN ALGORITMA REGRESI LINEAR

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk
memperoleh gelar Sarjana Komputer dari Program Studi Informatika
di Fakultas Teknologi Industri Universitas Islam Indonesia
Yogyakarta, 22 Juli 2024

Tim Penguji

Dr. Syarif Hidayat, S.Kom., M.I.T.

Anggota 1

Rahardian Kurniawan, S.Kom., M.Kom.

Anggota 2

Zainudin Zuhri, S.T., M.I.T.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Thomas Hatta Fudholi, S.T., M.Eng., Ph.D)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Haryadi Tri Nugroho

NIM : 20523097

Tugas akhir dengan judul:

**PENERAPAN INTELEJEN BISNIS DENGAN VISUALISASI
DATA GAJI DAN ALGORITMA REGRESI LINEAR**

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila dikemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung resiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 5 Juli 2024



(Haryadi Tri Nugroho)

HALAMAN PERSEMBAHAN

Alhamdulillah hirobbil alamin.

Segala puji syukur kepada Allah SWT, yang telah melimpahkan segala kesempatan, ridho, rahmat, dan hidayah kepada saya sehingga dapat menuntut ilmu, menambah pengalaman dan dapat menyelesaikan tugas akhir ini dengan waktu yang telah ditentukan. Shalawat serta salam penulis curahkan kepada Nabi Agung Muhammad SAW, yang telah memberikan Kelancaran dan Kegigihan dalam Proses Tugas Akhir ini. Dengan mengucap hamdallah, tugas akhir ini penulis persembahkan kepada:

1. Kepada orang tua tercinta, Bapak Suyadi dan Ibu Siti Kamsiyah yang telah sabar dalam mendidik anaknya hingga saat ini dan selalu memberikan dukungan, motivasi serta doa yang terbaik kepada saya.
2. kepada Bapak dan Ibu Dosen Fakultas Teknik Industri Program Studi Informatika Universitas Islam Indonesia. Terima kasih atas ilmu, bimbingan, serta pengalaman yang sudah diberikan kepada saya pribadi sepanjang menempuh pendidikan di program studi ini. Saya merasa sangat beruntung dan terhormat telah menjadi bagian dari keluarga besar Fakultas Teknik Industri Program Studi Informatika Universitas Islam Indonesia yang telah memberikan pendidikan serta pendidikan yang sangat berguna untuk saya pribadi.
3. Kakak saya tersayang, Diah Ayu Susanti dan Nining Setyoningsih yang selalu memberikan motivasi, dukungan dan doa.
4. Bapak Dr. Syarif Hidayat, S.KOM., M.IT. yang telah memberikan dukungan dan bimbingan dn sabar hingga tugas akhir ini selesai.
5. Teruntuk sahabat-sahabat tersayang yang selalu sabar mendengar keluh kesah dan selalu memberikan semangat serta doa.
6. Semua teman-teman yang telah membantu dan memberi dukungan hingga tugas akhir ini selesai.

HALAMAN MOTO

Man Jadda Wa jadda

“Barang siapa yang bersungguh-sungguh pasti akan mendapatkan hasil”

Tidak ada satu pun perjuangan yang tidak melelahkan. “Dan berikanlah berita gembira kepada orang-orang yang sabar, yaitu yang ketika ditimpa musibah mereka mengucapkan: sungguh kita semua ini milik Allah dan sungguh kepada Nya lah kita kembali”.

(QS Al-Baqarah 155-156)

Pantang dalam menyerah, pantang dalam berpatah arang. Tidak ada kata gagal untuk orang yang enggan berhasil. “Dan janganlah kamu berputus asa dari rahmat Allah. Sesungguhnya tiada berputus dari rahmat Allah melainkan orang-orang yang kufur”.

(QS Yusuf 87)

Orang yang kuat tidak memamerkan harapannya. Orang yang kuat adalah mereka yang terus berusaha untuk mewujudkan harapannya. “Dan hanya kepada Tuhanmulah hendaknya kamu berharap.”

(QS Al-Insyirah 8)

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Segala puji hanya bagi Allah SWT, Tuhan semesta alam, yang telah memberikan rahmat, hidayah, dan karunia-Nya kepada kita semua. Shalawat serta salam senantiasa kita haturkan kepada Nabi Muhammad SAW, yang merupakan teladan sempurna bagi umat manusia dalam segala aspek kehidupan.

Dalam kesempatan ini, penulis dengan rendah hati ingin mempersembahkan Tugas Akhir ini berjudul “ Penerapan Intelegen Bisnis Dengan Visualisasi Data Gaji dan Algoritma Regresi Linear ”. sebagai hasil dari perjalanan panjang dalam menuntut ilmu. Dalam menulis skripsi ini, penulis tidak lupa untuk memohon pertolongan serta ridha dari Allah SWT.

Laporan ini di susun untuk memenuhi persyaratan tugas Akhir di Fakultas Teknologi Industri Jurusan Informatika Universitas Islam Indonesia. Penulis menyusun laporan ini tidak lepas dari bimbingan dan arahan berbagai pihak. Penulis mengucapkan rasa hormat dan terima kasih kepada semua pihak yang telah membantu. Adapun pihak-pihak yang terkait diantaranya sebagai berikut:

1. Bapak Fathul Wahid, S.T, M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia yang telah memberikan kesempatan kepada saya untuk menyelesaikan Pendidikan program sarjana di kampus tersayang
2. Bapak Hari Purnomo, Prof., Dr., Ir., M.T., selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Ketua Jurusan Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia
4. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Ketua Program Studi Informatika Program Sarjana, Fakultas Teknologi Industri, Universitas Islam Indonesia.
5. Bapak Dr. Syarif Hidayat, S.KOM., M.IT. selaku Dosen Pembimbing yang telah bersedia membimbing dan mengarahkan penyusunan laporan ini.
6. Kedua orang tua tercinta yang selalu mendukung baik dan memberikan doa selama kegiatan penelitian berlangsung.
7. Kakak yang selalu membantu dan mendukung dalam proses menyusun laporan ini.
8. Kepada teman-teman seperjuangan di Kontrakan Oyee : Rian Tri wahyudi, Zunanda Ibrahim , Ma'mun uddin. Yang selalu ada ketika saya sakit, sedih dan

senang. Terima kasih selalu meramaikan kontrakan , menghibur, dan membantu saya

penulis menyadari banyak kekurangan dalam menyusun laporan tugas akhir ini, untuk itu terima kasih atas kritikan dan saran yang membangun. Semoga tugas ini dapat bermanfaat.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 5 Januari 2024

A handwritten signature in black ink, appearing to read 'adi', with a horizontal line underneath it.

(Haryadi Tri Nugroho)

SARI

Untuk memahami dinamika lapangan kerja ini, penting untuk mengeksplorasi berapa lama Data Scientist biasanya bekerja dalam bidang ini dan dalam sektor apa mereka lebih cenderung berkecimpung. Informasi ini menjadi sangat penting bagi perusahaan yang ingin merekrut Data Scientists untuk menentukan overpaid dan underpaid gaji karyawan mereka. Dengan adanya kegiatan ini, manajer dapat dengan lebih jelas mengetahui kisaran gaji karyawan berdasarkan pengalaman kerja dan bidang spesialisasi mereka. Ini memberikan manajer alat yang lebih tepat untuk menilai kompensasi yang sesuai dengan tingkat pengalaman dan keahlian yang dimiliki oleh karyawan. Penelitian ini bertujuan untuk berkontribusi pada pemahaman prediksi gaji dan memberikan informasi berharga bagi pemilik bisnis untuk membuat keputusan yang tepat. Hasilnya menunjukkan bahwa regresi linier secara akurat dan membentuk model regresi linier dengan hasil R-square sebesar 0.263429. Proses perhitungan prediksi gaji dilakukan dengan mempertimbangkan faktor-faktor yang memengaruhi, dan hasilnya kemudian disajikan secara visual menggunakan Power BI untuk memberikan pemilik bisnis informasi yang lebih interaktif.

Kata Kunci: Gaji, Power BI, Prediksi, Visualisasi

GLOSARIUM

<i>Business intelligence (BI)</i>	Suatu proses ekstraksi data operasional organisasi atau perusahaan, lalu dikumpulkan ke dalam sebuah data warehouse
<i>Dashboard</i>	Sebuah sistem informasi yang menyajikan informasi mengenai indikator utama dari aktivitas organisasi secara sekilas
<i>Dataset</i>	Kumpulan data yang diatur dalam format yang terstruktur seperti tabel atau file, dan berisi dari berbagai sumber
Extract, Transform, and Load (ETL)	Alat integrasi data dengan melibatkan penggalan dari sumber data luar (sistem operasional), mengubah data sesuai dengan kebutuhan bisnis dan akhirnya data tersebut dipindahkan ke gudang data
Grafik	Kelompok representasi data yang matematis digunakan sebagai alat bantu maupun sarana bagi disiplin ilmu untuk mengungkapkan atau memvisualisasikan pernyataan verbal yang kompleks
<i>Machine learning</i>	Penerapan kecerdasan buatan yang memberikan sistem kemampuan belajar secara otomatis dari sekumpulan data untuk melakukan tugas tertentu tanpa diprogram secara eksplisit
Microsoft Power BI	Kumpulan layanan, aplikasi, dan konektor perangkat lunak yang bekerja sama untuk mengubah sumber data yang tidak saling terkait menjadi wawasan yang koheren, mendalam secara visual, dan interaktif
Regresi linier	Metode statistik yang memungkinkan peneliti untuk memahami hubungan antara dua atau lebih variabel dalam suatu konteks
Visualisasi data	Koneksi antara data dan teknologi komputer untuk mengubah data menjadi bentuk visual atau sonik

DAFTAR ISI

HALAMAN PENGESAHAN DOSEN PEMBIMBING	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN.....	v
HALAMAN MOTO.....	vi
KATA PENGANTAR.....	vii
SARI.....	ix
GLOSARIUM.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah	6
1.4 Manfaat Penelitian	6
1.5 Tujuan Penelitian	6
1.6 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Penelitian Terdahulu	8
2.2 <i>Business Intelligence</i> (BI).....	14
2.2.1 Manfaat <i>Business Intelligence</i>	16
2.2.2 Arsitektur Pada <i>Business Intelligence</i>	17
2.3 Extract, Transform, and Load (ETL)	18
2.4 Visualisasi Data	19
2.5 <i>Business Intelligence Dashboard</i>	20
2.6 Jenis Grafik Untuk Visualisasi Data	20
2.7 <i>Machine Learning</i>	23
2.8 Algoritma <i>Linear Regression</i>	24
BAB III METODOLOGI.....	27
3.1 Alat Penelitian.....	27
3.2 Pengumpulan Data.....	27

	xii
3.2.1 Sumber Data.....	27
3.2.2 Teknik Pengumpulan Data.....	28
3.3 Data yang Dianalisis	28
3.4 Informasi yang Dibutuhkan	29
3.5 Metode Penelitian	29
3.5.1 Dataset	30
3.5.2 Pengolahan Data	30
3.5.3 Data Warehouse.....	33
3.5.4 Visualisasi Data.....	33
BAB IV HASIL DAN PEMBAHASAN	35
4.1 Implementasi <i>Business Intelligence</i>	35
4.2 Visualisasi Dashboard.....	44
4.3 Regresi Linear.....	48
BAB V KESIMPULAN DAN SARAN	54
5.1 Kesimpulan	54
5.2 Saran	54
DAFTAR PUSTAKA	55
LAMPIRAN.....	60

DAFTAR TABEL

Tabel 4.1 Pengukuran Metriks Kesalahan.....	52
---	----

DAFTAR GAMBAR

Gambar 2.1 Arsitektur <i>Business Intelligence</i>	17
Gambar 3.1 Data <i>Salaries</i>	27
Gambar 3.2 Diagram Metode Penelitian	30
Gambar 3.3 Data Pelatihan.....	30
Gambar 3.4 Tampilan <i>Get Data</i> Power BI	31
Gambar 3.5 Tampilan Tabel Data Yang Digunakan	31
Gambar 4.1 Tampilan Awal <i>Microsoft Power BI</i>	35
Gambar 4.2 Tampilan <i>Microsoft Power BI</i>	35
Gambar 4.3 Tampilan <i>Microsoft Power BI</i>	36
Gambar 4.4 Memasukkan Data ke Aplikasi.....	37
Gambar 4.5 Proses Menentukan Data	37
Gambar 4.6 Menu <i>Transform data</i>	38
Gambar 4.7 Menu <i>Power Query Editor</i>	38
Gambar 4.8 Hubungan Antar Tabel	39
Gambar 4.9 Menu <i>Visualizations Build Visual</i>	40
Gambar 4.10 Tampilan Awal Grafik.....	40
Gambar 4.11 Tampilan Grafik Pada Microsoft Power BI	41
Gambar 4.12 Operasi Agregate	42
Gambar 4.13 Visualisasi Data Operasi Agregat.....	42
Gambar 4.14 Visualisasi Data Operasi <i>Slicer</i>	43
Gambar 4.15 Visualisasi Data Operasi <i>Drill Down</i>	43
Gambar 4.16 Menu Pivot Tabel	43
Gambar 4.17 Proses Pemilihan Pivot Tabel	44
Gambar 4.18 Hasil Pivot Tabel	44
Gambar 4.19 Tampilan <i>Dashboard</i>	45
Gambar 4.20 Grafik Total <i>Salary</i> Berdasarkan <i>Employment type</i>	45
Gambar 4.21 Grafik Total <i>Salary</i> Berdasarkan <i>Company location</i>	46
Gambar 4.22 Grafik Total <i>Salary</i> Berdasarkan <i>Job title</i>	47
Gambar 4.23 Grafik Total <i>Salary</i> Berdasarkan <i>Company size</i>	47
Gambar 4.24 Komparasi Total <i>Salary</i> Prediksi dengan Total <i>Salary</i> Aktual	48
Gambar 4.25 <i>Library</i> yang Digunakan.....	48
Gambar 4.26 Pengecekan Nilai <i>Missing Value</i>	49

Gambar 4.27 Matriks Korelasi	50
Gambar 4.25 Potongan <i>Script Library</i> Linear Regresi di Python	50
Gambar 4.29 Grafik Salary in USD Prediksi dan Salary in USD sebenarnya	51

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital yang terus meningkat telah mengubah secara signifikan cara orang berinteraksi, bekerja, dan berkomunikasi dalam masyarakat. Hal ini telah membawa dampak besar pada berbagai aspek kehidupan sosial (Asnawi, 2022). Peningkatan perkembangan teknologi komunikasi yang pesat telah membawa perubahan mendasar dalam cara orang berinteraksi dan mengakses informasi. Kemunculan berbagai alat komunikasi baru, seperti telepon genggam, internet, dan komputer, telah membawa dampak signifikan dalam percepatan dan perluasan akses terhadap informasi. Telepon genggam, misalnya, telah mengubah cara orang berkomunikasi dengan orang lain, memungkinkan orang untuk berbicara, mengirim pesan, dan bahkan mengakses internet di hampir setiap tempat dengan hanya menggunakan ponsel pintar. Internet, di sisi lain, telah membuka dunia informasi yang tak terbatas di ujung jari (Sundari & Urip, 2021).

Pada dasarnya, teknologi ini seharusnya menjadi sarana yang mendukung perkembangan pemikiran manusia dan memperkuat hubungan sosial dan profesional. Teknologi harus berperan dalam memfasilitasi pertukaran ide, meningkatkan kolaborasi, serta memberikan akses mudah ke sumber daya pengetahuan. Di ranah sosial, teknologi harus berperan dalam menghubungkan individu, sehingga mereka dapat menjalin hubungan yang lebih kuat, terlepas dari jarak geografis. Sementara dalam konteks profesional, teknologi harus membantu dalam meningkatkan efisiensi kerja, mendukung inovasi, dan membuka peluang baru untuk pertumbuhan ekonomi (Wahyudi & Sukmasari, 2014). Di tengah pertumbuhan teknologi yang sangat pesat ini, masyarakat semakin bergantung pada teknologi digital, terutama internet, sebagai alat utama untuk berkomunikasi dan berinteraksi. Meskipun terdapat manfaat yang nyata, masyarakat tidak bisa mengabaikan potensi dampak negatifnya perubahan dalam dinamika sosial yang terkait dengan penggunaan media teknologi digital (Anjar, Ritonga, Toni, & Rahmayani, 2021). Hal ini mengingatkan akan perlunya keseimbangan dalam memanfaatkan teknologi ini, dengan mempertimbangkan konsekuensi sosialnya dan upaya untuk meminimalkan dampak negatif yang mungkin timbul seiring dengan ketergantungan pada teknologi digital (Setiawan, 2018).

Salah transformasi digital yang dipacu oleh kemajuan teknologi informasi dan komputasi, salah satu aspek penting yang mendapatkan dampak positif adalah peningkatan

kemampuan komunikasi manusia (Rahmadani, 2023). Hal ini menjadi sangat krusial, terutama dalam pekerjaan yang melibatkan *Data Science*. *Data Science*, sebagai bidang yang berfokus pada pengumpulan, analisis, dan interpretasi data, tidak terlepas dari peran teknologi dalam transformasi digital. Kemajuan teknologi informasi dan komputasi telah menciptakan ledakan data yang kompleks, menciptakan tantangan dan peluang yang harus dihadapi. Dalam konteks ini, kemampuan komunikasi manusia menjadi semakin esensial, karena kolaborasi dan pertukaran informasi memainkan peran sentral dalam mengatasi kompleksitas data, mengidentifikasi pola, dan menghasilkan wawasan berharga (Dhar, 2013). Peningkatan kemampuan komunikasi tidak hanya memungkinkan para profesional *Data Science* untuk berbagi pengetahuan dan temuan mereka secara efektif, tetapi juga mempromosikan kerja sama lintas disiplin yang sangat diperlukan untuk memecahkan masalah-masalah kompleks yang melibatkan data. Dengan demikian, aspek ini menjadi landasan yang kuat untuk kesuksesan dalam lapangan *Data Science* di era digital ini data (Firmansyah, Saepuloh, & Dede, 2022).

Dengan Pekerjaan bidang *Data Science* telah menjadi salah satu yang paling dicari dan penting dalam era digital ini. *Data Scientists*, dengan keterampilan analisis data dan keahlian dalam menggali wawasan dari data, menjadi elemen kunci dalam pengambilan keputusan bisnis yang efektif. Selain itu, lama bekerja dalam pekerjaan *Data Science* juga memainkan peran penting. *Data Scientists* yang memiliki pengalaman bertahun-tahun dalam industri ini memiliki pemahaman yang mendalam tentang tren, tantangan, dan perkembangan dalam dunia data. Sehingga, mereka dapat memberikan kontribusi yang berharga dalam menghadapi perubahan cepat dalam teknologi dan memastikan bahwa strategi data perusahaan tetap relevan dan efisien (Hasudungan, 2017).

Karakteristik pekerjaan yang berubah akan mendisrupsi pekerjaan yang telah ada dan menggantikannya dengan pekerjaan dengan karakteristik baru. Karakteristik baru pada pekerjaan juga membutuhkan kompetensi baru pada para pekerja. Sekarang ini, dalam dunia bisnis, setiap perusahaan harus bersiap untuk saling bersaing. Manajemen yang efektif dan keunggulan perlu dimiliki oleh setiap perusahaan untuk menyambut persaingan itu. Dalam hal ini, salah satu aspek yang berpengaruh besar terhadap kemajuan dan keberhasilan sebuah perusahaan adalah kinerja dari karyawannya. Faktor yang berpengaruh secara internal dapat terlihat dari penentuan gaji yang tepat oleh sebuah perusahaan kepada karyawan. Salah satu masalah yang dihadapi adalah bagaimana mengetahui gaji yang akan dikeluarkan berdasarkan lama bekerja dan faktor lainnya (Tuah & Anyan, 2020).

Salah satu pekerjaan *Data Science*, yang sangat dipengaruhi oleh kemajuan teknologi. Pekerjaan *Data Science* telah menjadi pusat perhatian dalam era digital ini. Untuk memahami dinamika lapangan kerja ini, penting untuk mengeksplorasi berapa lama *Data Scientist* biasanya bekerja dalam bidang ini dan dalam sektor apa mereka lebih cenderung berkecimpung. Informasi ini menjadi sangat penting bagi perusahaan yang ingin merekrut *Data Scientists* atau menentukan rata-rata gaji karyawan mereka. Dengan pemahaman yang lebih mendalam tentang lama kerja dan spesialisasi dalam bidang *Data Science*, perusahaan dapat mengembangkan strategi yang lebih efektif dan efisien dalam menarik bakat-bakat yang dibutuhkan untuk kemajuan dan inovasi perusahaan mereka. Dengan demikian, penelitian ini bertujuan untuk memberikan pemahaman yang lebih baik tentang aspek-aspek kunci yang terkait dengan pekerjaan *Data Science*, yang pada gilirannya diharapkan dapat mendukung perkembangan dan kesuksesan perusahaan dalam menghadapi tantangan teknologi modern (Nasution, Sitompul, & Nababan, 2020).

Dengan adanya kegiatan ini, manajer dapat dengan lebih jelas mengetahui kisaran gaji karyawan berdasarkan pengalaman kerja dan bidang spesialisasi mereka. Ini memberikan manajer alat yang lebih tepat untuk menilai kompensasi yang sesuai dengan tingkat pengalaman dan keahlian yang dimiliki oleh karyawan. Selain itu, juga memungkinkan manajer untuk menempatkan karyawan pada posisi yang sesuai dengan bidangnya, sehingga mereka dapat memberikan kontribusi terbaik mereka dalam pekerjaan yang mereka kuasai. Hal ini pada akhirnya akan meningkatkan efisiensi dan efektivitas perusahaan dalam memanfaatkan bakat dan pengetahuan karyawannya, yang merupakan langkah penting dalam mencapai kemajuan dan kesuksesan organisasi (Prabowo & Sari, 2019).

Penerapan ini menampilkan tren pekerjaan berdasarkan gaji dan pengalaman, pemilik usaha dapat mengidentifikasi pola yang berharga dalam dunia kerja. Data ini membantu perusahaan untuk merencanakan kompensasi yang lebih adil, yang sesuai dengan tingkat pengalaman karyawan dan bidang spesialisasi mereka. Selain itu, memahami tren ini juga memberikan wawasan yang berharga bagi para profesional yang ingin memilih jalur karier yang paling sesuai dengan tujuan finansial mereka. Dengan demikian, menampilkan tren pekerjaan berdasarkan gaji dan pengalaman adalah langkah yang signifikan dalam mencapai efisiensi dan keadilan dalam dunia kerja serta dalam membantu individu merencanakan karier mereka dengan bijak (Hasudungan, 2017).

Data set ini berasal dari Kaggle.com dengan judul "Data Science Salaries 2023." Kaggle merupakan *platform online* yang terkenal dalam dunia *data science* dan analitika,

yang menyediakan berbagai data set dan tantangan untuk para profesional dan penggemar *data science*. *Data set* "Data Science Salaries 2023" merupakan salah satu sumber informasi penting yang digunakan oleh para ilmuwan data dan analis untuk memahami tren gaji di industri *data science* pada tahun 2023. *Platform* ini memfasilitasi akses ke berbagai data yang relevan dan memungkinkan analisis mendalam tentang perbandingan industri (Das, Barik, & Mukherjee, 2020).

Penerapan analisis ini untuk mengidentifikasi trend dalam kompensasi berdasarkan tiga faktor utama: pekerjaan, lamanya jabatan, dan bidang yang ditekuni. Analisis ini dapat memberikan manfaat signifikan kepada manajer dan pemangku kepentingan dalam pengambilan keputusan terkait gaji dan kompensasi. Dengan pemahaman yang lebih baik tentang tren gaji, manajer dapat mengatur kompensasi karyawan sesuai dengan peran mereka dalam organisasi, masa kerja yang mereka miliki, serta sektor atau bidang pekerjaan yang mereka geluti. Ini memungkinkan manajer untuk memperhatikan faktor-faktor yang relevan dan memutuskan kebijakan kompensasi yang lebih tepat, yang pada akhirnya dapat meningkatkan motivasi, retensi karyawan, dan kinerja (Mahaputra, 2022).

Dalam analisis data, regresi linier merupakan salah satu algoritma paling dasar dan sering digunakan dalam *machine learning*, terutama dalam masalah yang melibatkan prediksi nilai kontinu. Ini adalah metode statistik yang memungkinkan untuk memahami hubungan antara dua atau lebih variabel dalam suatu konteks. Dalam banyak kasus, regresi linier digunakan untuk mengukur dan menganalisis hubungan antara variabel dependen dan variabel independen. Regresi linier memungkinkan untuk mengeksplorasi sejauh mana perubahan dalam variabel independen mempengaruhi variabel dependen, dan dengan demikian, memungkinkan untuk membuat prediksi yang lebih baik dan informasi yang lebih berharga dalam berbagai disiplin ilmu, seperti ilmu sosial, ekonomi dengan menggunakan *linear regression* dengan menampilkan variabel terikat pada data kasus, regresi linier menjadi alat yang sangat relevan ilmu (Khan, et al., 2023). Analisis regresi linier memberikan wawasan yang mendalam tentang bagaimana variabel independen berdampak pada variabel terikat dalam kerangka data yang konkretnya. Dengan menggunakan alat ini, peneliti dapat mengidentifikasi hubungan sebab-akibat dan menggambarkan pola-pola yang mungkin tersembunyi dalam data. Hal ini memungkinkan pengambilan keputusan yang lebih cerdas dan akurat (Maulud & Abdulazeez, 2020). Algoritma regresi linier dipilih karena kesederhanaan dan kemudahan dalam interpretasi (mudah dipahami) serta mempunyai kecocokan dengan masalah prediksi. Algoritma regresi linear sangat cocok untuk masalah

prediksi yang tujuan utamanya adalah memprediksi nilai kontinu, seperti gaji berdasarkan variabel-variabel lain (misalnya, pengalaman kerja, pendidikan, posisi, dan lainnya).

Dalam era perkembangan teknologi dan analisis data yang pesat, beragam pekerjaan di bidang data seperti *Data Engineer*, *Data Scientist*, *Data Analyst*, *Machine Learning Engineer*, dan *Analytics Engineer* menawarkan peluang beragam. Dalam upaya memahami dinamika pekerjaan ini, algoritma regresi linier digunakan untuk mengidentifikasi *job title* yang tersedia, mengestimasi lama pekerjaan yang diperlukan, serta apakah pekerjaan tersebut lebih cenderung dilakukan secara *online* atau *offline*. Selain itu, algoritma ini juga membantu dalam menganalisis dan meramalkan gaji yang berkaitan dengan peran-peran ini, mencakup gaji tertinggi, gaji rata-rata, dan gaji terendah (Hope, 2020).

Setelah hasil analisis regresi linier diperoleh menggunakan bantuan bahasa pemrograman python sebagai *library machine learning*, penting untuk mengkomunikasikan temuan tersebut secara efektif. Dalam hal ini, aplikasi *Microsoft Power BI* menjadi alat yang sangat berguna. Dengan *Power BI*, para peneliti dapat mengimpor data hasil analisis regresi linier dan dengan mudah membuat visualisasi yang informatif. Visualisasi ini dapat berupa grafik garis atau grafik sebaran yang memperlihatkan hubungan antara variabel dependen dan independen, dengan garis regresi linier (Darman, 2018).

Selanjutnya divisualisasikan menggunakan *Power BI* juga memungkinkan pembuatan laporan interaktif yang dapat diakses oleh berbagai pemangku kepentingan. Laporan ini dapat menyediakan filter dan opsi interaktif yang memungkinkan para pengguna untuk menjelajahi data dengan lebih mendalam, serta memahami dampak berbagai faktor terhadap gaji dan pekerjaan di berbagai tingkatan. Dengan demikian, *Power BI* memberikan solusi yang lengkap dalam mengintegrasikan analisis regresi linier dengan visualisasi data yang informatif dan pelaporan yang memudahkan berbagai hasil penelitian secara lebih dinamis (Nurfarisi, 2022). Pada analisis data dengan regresi linier dan visualisasi menggunakan *Power BI* adalah pendekatan yang kuat untuk memahami dinamika pekerjaan di era digital. Memahami hubungan antara variabel pekerjaan, faktor lama pekerjaan, metode kerja, dan kompensasi adalah langkah penting dalam mengelola sumber daya manusia dan pengambilan keputusan (Zikra A. A., 2023).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas maka rumusan masalah pada penelitian ini Karyawan data scientist mengalami kesulitan dalam menentukan gaji yang tepat, yang menyebabkan terjadinya *overpaid* dan *underpaid*.

1.3 Batasan Masalah

Batasan pada penelitian ini adalah:

- a. Data yang diambil dari dataset Kaggle dengan judul Data Science Salaries.
- b. Data yang digunakan dalam penelitian data historis salaries tahun 2020-2023.
- c. Algoritma yang digunakan dalam penelitian ini adalah *linear regression*.
- d. Aplikasi visualisasi menggunakan Power BI.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari dilakukannya penelitian ini yaitu:

- a. Dapat menampilkan visualisasi data gaji
- b. Dapat membantu dalam menganalisis dan meramalkan gaji berdasarkan variabel yang tersedia.
- c. Membantu pekerjaan untuk memahami gajinya berdasarkan pengalaman di bidangnya.

1.5 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, maka tujuan dari penelitian ini adalah memanfaatkan aplikasi *business intelligence* berdasarkan data yang diperoleh untuk mengetahui faktor apa saja yang mempengaruhi gaji Overpaid dan Underpaid

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan laporan penelitian ini terdiri dari lima bab yang diuraikan seperti berikut:

BAB I : PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini membahas konsep, teori, dan studi terdahulu yang relevan dengan topik penelitian. Landasan teori membantu pembacanya memahami relevansi dan urgensi topik penelitian dengan membandingkan hasil penelitian terdahulu, teori yang telah ada, dan temuan terkini tentang bidang tersebut untuk memberikan dasar perumusan kerangka penelitian.

BAB III : METODOLOGI

Bab ini membahas tahapan penelitian yang akan dilakukan dengan memaparkan pengambilan data, pemrosesan data, implementasi algoritma regresi linier, dan visualisasi data.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas tentang hasil pengimplementasian metode serta pembahasannya. Hasil pengolahan data yang dibahas akan disesuaikan dengan tujuan penelitian untuk mendapatkan kesimpulan.

BAB V : KESIMPULAN DAN SARAN

Bab ini adalah bab terakhir yang membahas kesimpulan dan saran dari penelitian yang telah dilakukan.

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Pada bagian ini penulis akan memaparkan hasil dari beberapa penelitian yang telah ada dengan topik yang relevan. Pemaparan hasil penelitian terdahulu merupakan hal penting dan dapat digunakan sebagai referensi. Berikut ini merupakan beberapa penelitian terdahulu yang penulis jadikan referensi

Penelitian pertama dilakukan oleh Maesaroh, et al (2022) berjudul “Efektivitas Implementasi Manajemen *Business Intelligence* pada Industri 4.0”. Pada penelitian ini didapatkan, BI merupakan tantangan teknologi informasi Industri 4.0 dan menjadi isu manajemen yang sangat penting. Secara khusus, pentingnya mengembangkan keterampilan pengambilan keputusan yang baik secara analitis, yang tercermin dalam perangkat lunak dan sistem komputer, telah diakui. Industri 4.0 telah menjadikan lingkungan bisnis semakin kompleks. Perusahaan memerlukan inovasi dan teknologi canggih untuk merespons pasar yang dinamis ini dengan cepat. Dalam konteks ini, alat teknologi seperti intelijen bisnis (BI) diperlukan baik untuk pemrosesan informasi maupun pengambilan keputusan di tingkat perusahaan. Untuk menjalankan seluruh sistem BI, ada berbagai jenis kekuatan dan pengguna biasa. Pengguna yang kuat adalah pakar BI yang mengambil data diperlukan untuk menghasilkan laporan yang digunakan sebagai landasan untuk pengambilan keputusan (Elena & University, n.d.)

Penelitian kedua dilakukan oleh Husni dan Mukhlash (2014) berjudul “Implementasi *Business Intelligence* Pada Manajemen Report Bank XYZ”. Penelitian ini berfokus pada bagaimana BI memudahkan pengelolaan data besar dan mendukung pengambilan keputusan. Jurnal ini untuk pembuatan *prototype* data *warehouse* dan penerapan analisis multidimensi pada OLAP serta penggunaan SAS® *Enterprise Guide* untuk model multidimensi terkait *funding* dan *lending*. Hasilnya menunjukkan bahwa analisis multidimensi dengan data *warehouse* signifikan dalam memudahkan analisis data, mempercepat pengambilan keputusan, dan meningkatkan efisiensi dalam memantau pertumbuhan bisnis. Penelitian ini menggarisbawahi pentingnya BI dalam mengelola data besar di industri perbankan untuk pertumbuhan dan efisiensi bisnis.

Penelitian ketiga dilakukan oleh Hasan (2019) berjudul “Implementasi Sistem Business Intelligence Untuk Data Penelitian di Perguruan Tinggi” membahas Sistem BI ini dirancang

untuk mengatasi masalah kurangnya integrasi dan penyimpanan data yang tidak terstruktur di Perguruan Tinggi XYZ, dengan membangun data *warehouse* dan menerapkan analisis OLAP serta dashboard interaktif. Dengan sistem ini, perguruan tinggi dapat mengumpulkan, menganalisis, dan melaporkan data penelitian dengan lebih efektif, mendukung pengambilan keputusan dan evaluasi kinerja dosen. Jurnal ini menekankan manfaat penggunaan BI berbasis *open source* seperti Pentaho Data Integration untuk mengoptimalkan manajemen data penelitian, mempermudah pembuatan laporan, serta memberikan wawasan strategis melalui dashboard interaktif kepada pimpinan perguruan tinggi.

Penelitian keempat dilakukan oleh Akbar, et al (2018) berjudul “Implementasi Business Intelligence Menentukan Daerah Rawan Gempa Bumi di Indonesia dengan Fitur Geolokasi” bertujuan untuk menganalisis data gempa bumi dari Badan Nasional Penanggulangan Bencana (BNPB) Indonesia. Aplikasi ini memungkinkan visualisasi data secara spasial pada peta berdasarkan geolokasi, memberikan gambaran akurat tentang daerah rawan gempa bumi. Tujuan utama penelitian ini adalah menyediakan informasi penting untuk rencana pembangunan yang lebih aman di daerah rawan bencana. Jurnal ini menekankan betapa pentingnya penggunaan teknologi BI dalam mengelola dan menganalisis data besar, terutama untuk mendukung keputusan strategis dalam mitigasi bencana alam, seperti gempa bumi, di Indonesia.

Penelitian kelima dilakukan oleh Priyatna (2019) berjudul “Implementasi Sistem Penunjang Keputusan Menggunakan *Business Intelligence* Untuk UMKM Di Gunung Putri Kab. Bogor” bertujuan untuk penunjang keputusan (SPK) dengan Business Intelligence (BI) untuk mendukung Usaha Mikro, Kecil, dan Menengah (UMKM) di Gunung Putri, Kabupaten Bogor. Penelitian ini mengutilisasi alat Microsoft seperti *SQL Server Integration Services* (SSIS) dan *SQL Server Reporting Services* (SSRS) untuk mengolah data dan membantu UMKM memahami faktor-faktor yang mempengaruhi kinerja organisasi. Dengan menganalisis tren dan strategi, serta membandingkan hasil dari data sistem lama dan baru menggunakan software SPSS, penelitian ini bertujuan memberikan wawasan yang berharga untuk pengambilan keputusan dan peningkatan keunggulan kompetitif UMKM di wilayah tersebut.

Penelitian keenam dilakukan oleh Riyanda dan Suyanto (2020) berjudul “Implementasi *Business Intelligence* Pada Analisis Perkembangan Hasil Pertanian Provinsi Sumatera Selatan” bertujuan untuk mengatasi masalah pengelolaan data pertanian yang besar dan terfragmentasi dengan menciptakan sebuah *dashboard Business Intelligence*. *Dashboard* ini

memvisualisasikan data penting seperti persentase subsektor pertanian, tren produksi, dan produksi di berbagai wilayah, memudahkan analisis dan meningkatkan efisiensi dalam evaluasi hasil pertanian di provinsi tersebut. Kesimpulan dari penelitian ini menunjukkan bahwa penerapan teknologi *Business Intelligence* secara signifikan memperbaiki proses analisis data pertanian, khususnya di lingkungan dengan sumber daya dan infrastruktur teknologi terbatas.

Penelitian ketujuh dilakukan oleh Siska dan Putri (2021) berjudul “Implementasi *Business Intelligence* untuk Menganalisis perbandingan data kasus COVID-19 di Jawa Barat sebelum PSBB dan setelah PSBB” bertujuan untuk mengkaji penggunaan teknologi *Business Intelligence* (BI) dalam memetakan dan menganalisis data COVID-19 di Jawa Barat. Penelitian ini mendeskripsikan kasus COVID-19, termasuk Orang Dalam Pantauan (ODP), Pasien Dalam Pantauan (PDP), dan Pasien Positif, serta menggambarkan grafik terkonfirmasi sembuh dan meninggal. Studi ini melibatkan data historis selama 14 hari pada tiga periode berbeda: sebelum, selama, dan setelah PSBB. Hasilnya menunjukkan peningkatan kasus positif dan kematian selama PSBB, dengan tingkat kesembuhan yang relatif rendah. Penelitian ini memberikan wawasan penting mengenai dinamika COVID-19 di Jawa Barat dan menyoroti pentingnya teknologi BI dalam pengambilan keputusan dan kebijakan kesehatan masyarakat.

Penelitian kedelapan dilakukan oleh Putra et al (2023) berjudul “Pelatihan Power BI: Meningkatkan Kinerja Bisnis dengan Analisis Data dan Visualisasi yang Optimal.”. Penelitian ini terfokuskan pada Kolaborasi antara institusi pendidikan INSTIKI dan perusahaan konsultan Gama *Counseling & Preparing* bertujuan untuk meningkatkan pengetahuan dan kemampuan peserta dalam memanfaatkan *Power BI* untuk analisis informasi bisnis yang efisien dan pengambilan keputusan yang tepat. Pelatihan dilakukan dengan berbagai metode, termasuk penyampaian materi, diskusi, sesi praktik, dan kuesioner. Hasilnya menunjukkan kepuasan peserta secara keseluruhan, dengan umpan balik positif terhadap materi yang disampaikan. Studi ini merekomendasikan pelatihan lebih lanjut dengan fokus pada fitur-fitur yang belum dieksplorasi untuk mengoptimalkan potensi Power BI sebagai alat yang komprehensif untuk meningkatkan kinerja bisnis. Artikel ini mencakup penjelasan rinci tentang program pelatihan, termasuk topik yang dibahas dan metode yang digunakan, serta menyertakan dokumentasi visual dari sesi pelatihan dan hasil kuesioner pasca pelatihan.

Penelitian kesembilan dilakukan oleh Zikra (2022) berjudul "Implementasi *Business Intelligence* pada ACC Absensi Menggunakan Aplikasi Power BI." Penelitian ini bertujuan untuk mencatat data kehadiran karyawan, termasuk kedatangan, keterlambatan, dan kepulangan mereka. Informasi presensi ini memiliki manfaat dalam meningkatkan pengelolaan sumber daya manusia perusahaan, seperti pengaturan jadwal dan perhitungan waktu kerja karyawan. Untuk mencapai tujuan ini, Zikra membangun sebuah model intelijen bisnis menggunakan Power BI sebagai platform untuk membuat dashboard visualisasi. Proses pembuatan dashboard ini melibatkan langkah-langkah ETL (*Extract, Transform, Load*), yang dapat meningkatkan efisiensi penghitungan kinerja karyawan dengan kemampuan untuk melakukan pembaruan secara berkala.

Penelitian kesepuluh dilakukan oleh Nabilla dan Hidayat (2021) berjudul "Pengembangan *Business Intelligence* Pada Sistem Informasi Distributor", penelitian ini bertujuan untuk memantau pendapatan dan profit mereka sesuai dengan periode waktu yang mereka tentukan. Melalui pengembangan *Business Intelligence* (BI) dalam sistem ini, disoroti bahwa dengan mengetahui varian masker yang menghasilkan profit tertinggi dan terendah, distributor memiliki kesempatan untuk merencanakan langkah-langkah dan strategi dalam pengambilan keputusan pemasaran. Selain itu, distributor masker juga dapat memperoleh pemahaman tentang reseller yang memiliki penjualan tertinggi, yang dapat dijalin menjadi mitra bisnis potensial di masa mendatang.

Penelitian kesebelas dilakukan oleh Edhya dan Susilowati (2022) berjudul "*Business Intelligence Data Marketing* Menggunakan Metode Kimball dan ETL Dengan Power BI" penelitian ini untuk peningkatan signifikan dalam pendapatan distributor. Selain itu, penelitian ini juga mengungkapkan informasi penting mengenai jenis produk yang memiliki penjualan tertinggi, serta produk mana yang memiliki penjualan yang relatif lebih rendah. Selanjutnya, terdapat visualisasi data dalam bentuk peta yang menunjukkan jumlah pembelian tertinggi di berbagai daerah, memberikan gambaran yang jelas tentang sebaran penjualan produk. Penggunaan platform Power BI dalam proses visualisasi data tidak hanya membuatnya lebih menarik tetapi juga memudahkan pemahaman data bagi pembaca yang beragam.

Penelitian kedua belas dilakukan Kamasi dan Nangoy (2023) oleh berjudul "Pengembangan *Business Intelligence* untuk Pendukung Pengambilan Keputusan dalam Mengurangi Kemacetan di Kota Tomohon". Pada penelitian ini menyatakan bahwa penyebab utama kemacetan lalu lintas adalah ketidakseimbangan antara pertumbuhan jumlah kendaraan

dan kapasitas infrastruktur transportasi yang ada kemudian menjelaskan bahwa BI adalah kerangka kerja konseptual yang mendukung pengambilan keputusan bisnis dengan menggabungkan arsitektur, basis data atau data warehouse, alat analisis, dan aplikasi. Artikel ini juga menjelaskan metode penelitian tindakan yang digunakan untuk mengembangkan sistem BI, yang meliputi empat tahap: mendiagnosis, perencanaan tindakan, pengambilan tindakan, dan evaluasi. Artikel ini memberikan penjelasan rinci tentang setiap tahap dan hasil penelitian.

Penelitian ketiga belas dilakukan Rahmawati (2022) oleh berjudul “ Implementasi Power BI untuk Dashboard Analis Hasil Rating dan Review Produk Elektronik di Lazada.com“ bertujuan untuk pengembangan *dashboard* analisis menggunakan Power BI untuk memantau dan menganalisis hasil rating dan *review* produk elektronik di Lazada.com. Dengan menggunakan data dari 203.787 transaksi barang elektronik di Lazada pada tahun 2019, penelitian ini menghasilkan dashboard interaktif yang memungkinkan pemahaman yang lebih baik tentang tren konsumen dan perilaku pembelian. Analisis menunjukkan dominasi *review* positif dalam transaksi, memberikan wawasan berharga bagi penjual dan pembeli di *platform e-commerce* tersebut. Melalui teknik pengolahan data dan visualisasi yang canggih, penelitian ini menawarkan kontribusi signifikan dalam bidang analisis data *e-commerce* dan pembuatan keputusan bisnis.

Penelitian keempat belas dilakukan oleh Putri (2022) berjudul “Aplikasi Model Prediksi Gaji Profesi Data Berbasis Website Menggunakan Flask Dan Metode Decision Tree” bertujuan untuk memprediksi gaji profesi data yaitu *Data Engineer*, *Data Analyst*, dan *Data Scientist* dengan menggunakan metode terbaik yang diperoleh dari 3 metode yaitu *Random Forest*, *Support Vector Regression*, dan *Decision Tree*, serta menyajikannya dalam bentuk *website* sehingga mudah digunakan. Pengambilan data dilakukan dengan cara *web scraping* pada tanggal 4 November 2021 untuk data LinkedIn dan tanggal 6 Desember 2021 untuk data Jobstreet. Variabel-variabel yang berhubungan cukup kuat terhadap “gaji” adalah “nama profesi data”, “lokasi perusahaan”, “tingkat pekerjaan”, “lama pengalaman”, “ukuran perusahaan”, dan “industri”. Metode terbaik yang diperoleh dalam penelitian ini adalah *Decision Tree* menggunakan SMOTE dengan jumlah data per kategori “nama profesi data” adalah $n=87$. Parameter yang digunakan pada metode tersebut merupakan parameter terbaik yang diperoleh dari proses *hyperparameter tuning* dengan hasil evaluasi adalah R^2 sebesar 0,669, RMSE sebesar 147,817, serta MAPE sebesar 14,1%.

Penelitian kelima belas dilakukan oleh Saputra et al (2022) berjudul "Prediksi Gaji Berdasarkan Pengalaman Bekerja Menggunakan Metode Regresi Linear" menyatakan bahwa industri tidak bisa dipisahkan dari adanya sumber daya manusia (SDM). Walaupun industri memiliki teknologi yang maju dan juga modern, namun berhasilnya suatu perusahaan tak lepas dari jasa para sumber daya manusia yang unggul. Dengan begitu perlu bagi perusahaan untuk memperhatikan para pekerjanya. Salah satu usaha untuk meningkatkan mutu SDM yaitu dengan pemberian gaji berdasarkan pengalaman kerja. Ketika seseorang yang sudah lama bekerja di suatu perusahaan maka gajinya akan semakin naik. Penelitian ini ditujukan guna menganalisis prediksi gaji karyawan berdasarkan lama tahun bekerja. Dalam penelitian ini faktor pengujianya menggunakan variable (X) sebagai faktor pemicu terhadap variable (Y) konsekuensi. Metode yang digunakan dalam riset ini yaitu menggunakan metode Regresi linier. Kemudian kami menggunakan survei kuesioner kepada 30 responden sebagai metode pengambilan data. Hasil yang didapatkan yaitu tingkat error MAE 2.05 yang mana sangat bagus.

Penelitian keenam belas dilakukan Daniswara, Triwulandari, & Habyba (2022) dengan judul "Perancangan Gaji Karyawan Menggunakan Metode Point System Di PT. Mitra Karya Teknik Indonesia" menyakan bahwa penggajian karyawan dilakukan sama rata untuk setiap jabatan dan hal ini menimbulkan ketidakpuasan karyawan terhadap kompensasi yang diterima. Tujuan penelitian ini adalah memberikan usulan kepada perusahaan rancangan gaji karyawan. Berdasarkan metode *Point System* terdapat tiga tahapan utama yaitu analisa jabatan, evaluasi jabatan, dan penentuan besaran gaji. Metode *Point System* digunakan dalam penentuan bobot setiap indikator kinerja yang akan digunakan dalam penggajian. Analisa jabatan dilakukan pada lima manajer berbeda divisi yang dirancang ulang sistem penggajiannya. Keterampilan, tanggung jawab, kondisi kerja, dan usaha adalah empat faktor utama yang digunakan dalam evaluasi jabatan. Hasil perhitungan ulang gaji menghasilkan keseluruhan gaji yang lebih tinggi untuk masing-masing jabatan jika dibandingkan Upah Minimum Regional (UMR) dan diharapkan dapat meningkatkan loyalitas karyawan.

Penelitian ketujuh belas dilakukan oleh Hidayat (2022) dengan judul "Implementasi Machine Learning Prediksi Gaji Pekerjaan Perusahaan Rintisan Menggunakan Algoritma Random Forest Dan Framework Flask (Studi Kasus: Pekerjaan Pada Website Techinasia.Com)" menyatakan bahwa Indonesia merupakan negara yang memiliki perkembangan teknologi yang sangat pesat karena memiliki 2193 perusahaan rintisan pada

tahun 2019, terbanyak kelima di dunia. Meningkatnya kebutuhan pekerja di sektor teknologi membuat persoalan lain yaitu menentukan besaran gaji yang sesuai bagi pekerja maupun perusahaan sesuai dengan beban dan tanggung jawab pekerjaannya sehingga penulis membuat suatu penelitian yang dapat membantu dalam memprediksi besaran gaji di perusahaan rintisan sesuai dengan kriteria dan jenis pekerjaan di Indonesia seperti nama pekerjaan, ukuran perusahaan, tipe industri, pengalaman kerja, dan besaran gaji. Untuk memprediksi besaran gaji maka digunakan suatu metode *Machine Learning* yaitu algoritma *Random Forest* dengan *parameter* tuning menggunakan *Search Grid CV* dengan hasil *RMSE* dari data testing sebesar 3.243.109. Model yang telah dibuat selanjutnya di *deploy* dengan *framework flask* di *python* dan di implementasikan ke *website* dengan *Heroku*.

Berdasarkan uraian penelitian terdahulu di atas maka peneliti akan melakukan penelitian mengenai penerapan intelegen bisnis dengan visualisasi data gaji dan algoritma *regresi linear*. Pada penelitian ini visualisasi *dashboard* menggunakan Power BI dan bahasa pemrograman *python* untuk *machine learning* serta metode regresi linear. Metode regresi linear dipilih karena mudah dimengerti dan tidak mensyaratkan data harus dalam bentuk tertentu. Regresi linear mencoba untuk memodelkan hubungan antara dua variabel dengan mencocokkan persamaan linier dengan data yang diamati. Satu variabel dianggap sebagai variabel penjelas, dan yang lainnya dianggap sebagai variabel dependen. Dengan begitu maka metode Regresi Linear dapat mengetahui besarnya pengaruh secara kuantitatif setiap variabel bebas apabila pengaruh variabel dianggap konstan dan dapat menganalisis dengan menggunakan beberapa variabel bebas sehingga hasil yang didapat lebih akurat (Susanto, 2014).

2.2 Business Intelligence (BI)

Business intelligence (BI) adalah suatu cara untuk mengumpulkan, menyimpan, mengorganisasikan, membentuk ulang, meringkas information serta menyediakan informasi, baik berupa information aktivitas bisnis inside perusahaan, maupun information aktivitas bisnis eksternal perusahaan termasuk aktivitas bisnis para pesaing yang mudah diakses serta dianalisis untuk berbagai kegiatan manajemen (Akbar, et al., 2017). Kimball dan Caserta mendefinisikan *Business intelligence* (BI) sebagai suatu proses ekstraksi information operasional organisasi atau perusahaan, lalu dikumpulkan ke dalam sebuah information stockroom. Information stockroom dirancang untuk mendukung sebuah proses lanjutan untuk mendapatkan informasi berharga dengan menggunakan teknik data *mining*. Pendekatan ini melibatkan beberapa teknik dan alat, termasuk analisis bisnis mendalam, data *mining*,

visualisasi data kreatif, alat pemrosesan data canggih, dan infrastruktur data yang efisien. Semua ini bertujuan untuk membantu organisasi lebih memahami data, memandu pengambilan keputusan bisnis yang cerdas, dan meningkatkan kinerja dan daya saing di pasar yang kompetitif untuk mendukung intelijen bisnis dalam pengambilan keputusan bisnis dengan menggabungkan arsitektur, basis data atau gudang data, alat analisis, dan aplikasi. Ini digunakan untuk mengumpulkan, menyimpan, menganalisis, dan memberikan akses ke data, memungkinkan pengguna dari perusahaan atau organisasi untuk membuat keputusan yang lebih baik dan akurat. *Business Intelligence* juga digunakan untuk memprediksi dan melihat dampak keputusan yang diambil oleh organisasi.

Business Intelligence memiliki peran utama dalam memberikan metrik kinerja dan perbandingan standar dalam dunia bisnis yang senantiasa berubah. Dalam lingkungan bisnis yang dinamis, memiliki akses tepat waktu terhadap informasi mengenai kinerja perusahaan menjadi hal yang krusial. Melalui pemanfaatan *Business Intelligence*, para pemimpin bisnis mampu mengawasi perkembangan organisasi mereka terhadap target-target bisnis yang telah ditetapkan. Data yang disediakan tidak hanya mencakup aspek finansial, tetapi juga mengevaluasi performa operasional dan strategis perusahaan.

Keputusan bisnis yang didasarkan pada pemahaman mendalam terhadap data ini mampu membimbing organisasi menuju arah yang benar, meningkatkan daya saing, dan memberikan nilai tambah kepada pelanggan. *Business Intelligence* memungkinkan perusahaan untuk mengidentifikasi tren pasar, merespons perubahan lingkungan bisnis dengan cepat, dan mengoptimalkan operasi mereka. Dengan demikian, peran *Business Intelligence* bukan hanya untuk memantau kinerja saat ini, tetapi juga untuk mempersiapkan organisasi menghadapi tantangan masa depan. Intelijen bisnis membuka peluang bagi organisasi untuk memahami dengan lebih baik pola gaji. Di era data yang melimpah, alat-alat *Business Intelligence* membantu mengurai pola-pola yang tersembunyi di balik kumpulan angka dan fakta. Dengan pemahaman yang lebih mendalam tentang preferensi dan perilaku, bisnis dapat merancang strategi pemasaran yang lebih efektif dan personal. Analisis ini tidak hanya membatasi diri pada mengetahui produk atau layanan yang paling diminati, tetapi juga melibatkan pemahaman konteks di balik keputusan pemberian gaji.

Hasil analisis data yang diperoleh berupa kunci-kunci pengetahuan bisnis yang bermanfaat dalam peningkatan kinerja organisasi. Visualisasi dari BI dapat disajikan dalam bentuk *dashboard* untuk mendapatkan informasi detail. Tampilan visualisasi di dalam *dashboard* bisa berupa grafik, *pie*, *custom*, *drill down* dan lain sebagainya. *Dashboard*

berfungsi sebagai pemantau untuk mewujudkan strategi yang tepat dan berkualitas bagi *top level* manajemen khususnya yang terkait dengan pengambilan keputusan (Darman, 2018).

2.2.1 Manfaat *Business Intelligence*

Adapun manfaat yang dapat diperoleh dari mengimplementasikan *Business intelligence* adalah sebagai berikut (Sulistiyorini, 2010):

- a. Meningkatkan nilai data dan informasi organisasi.

Melalui implementasi *Business Intelligence* (BI), semua data dan informasi dapat disatukan sehingga membentuk dasar pengambilan keputusan yang komprehensif. Informasi yang sebelumnya terisolasi dan tidak tercakup sebagai faktor dalam pengambilan keputusan dapat dengan mudah dihubungkan dan digabungkan menggunakan BI. Selain itu, data dan informasi yang dihasilkan menjadi lebih mudah diakses dan dipahami oleh pengguna (informasi yang ramah pengguna).

- b. Memudahkan pemantauan kinerja organisasi.

Dalam mengevaluasi kinerja suatu organisasi, sering kali digunakan ukuran yang dikenal sebagai *Key Performance Indicator* (KPI). KPI tidak selalu diukur dalam satuan uang, melainkan dapat juga berdasarkan faktor seperti kecepatan pelaksanaan suatu layanan. *Business Intelligence* (BI) dapat dengan mudah, cepat, dan akurat menampilkan pencapaian KPI suatu organisasi. Dengan demikian, pihak yang terlibat dalam pengambilan keputusan akan lebih mudah menentukan langkah-langkah antisipatif yang diperlukan.

- c. Meningkatkan nilai investasi TI yang sudah ada.

Business Intelligence (BI) tidak perlu atau tidak harus mengubah atau menggantikan sistem informasi yang sudah ada sebelumnya. Sebaliknya, BI berperan sebagai tambahan layanan pada sistem-sistem tersebut, memungkinkan data dan informasi yang sudah ada menghasilkan informasi yang lebih komprehensif dan bermanfaat.

- d. Menciptakan pegawai yang memiliki akses informasi yang baik (*well-informed workers*).

Dalam menjalankan tugasnya, semua tingkatan dalam suatu organisasi, mulai dari pegawai atau bawahan hingga pimpinan, selalu memiliki keterkaitan dan/atau kebutuhan akses terhadap data dan informasi. *Business Intelligence* (BI) mempermudah akses data dan informasi bagi seluruh tingkatan pegawai, sehingga membantu dalam proses pengambilan keputusan. Dengan kondisi ini tercapai, pelaksanaan misi dan

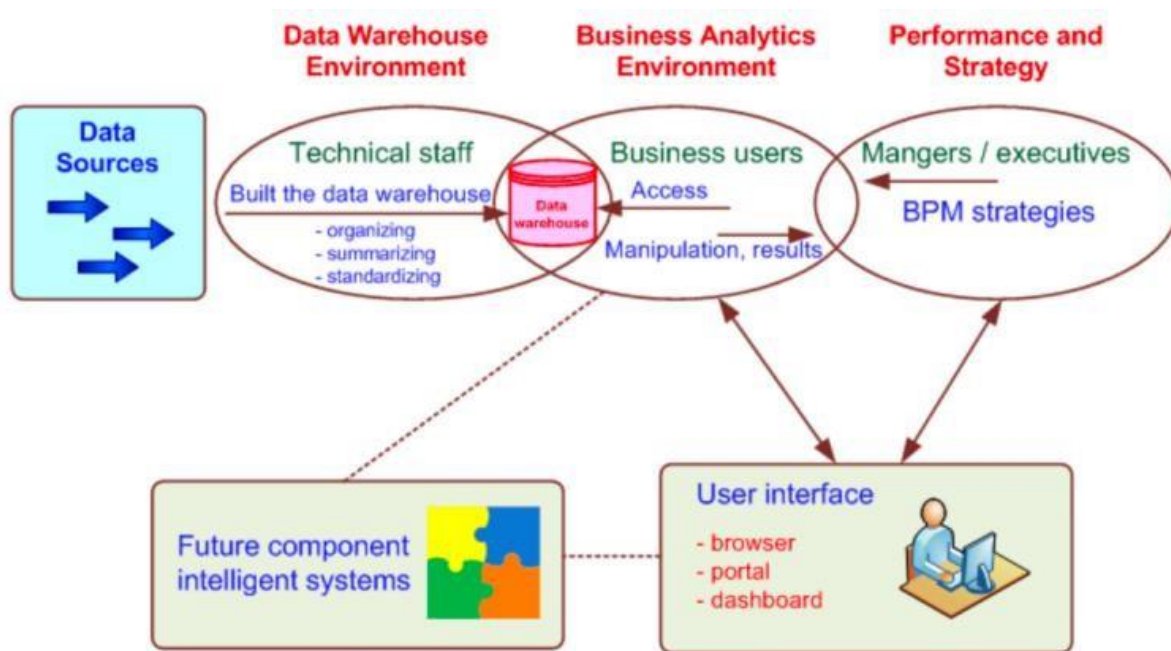
strategi organisasi dapat lebih mudah terwujud dan tingkat pencapaian dapat lebih efektif dipantau.

e. Meningkatkan efisiensi biaya.

Business Intelligence (BI) dapat meningkatkan efisiensi karena mempermudah individu dalam menjalankan tugasnya, menghemat waktu, dan mudah dalam pemanfaatannya. Proses pencarian data dan perolehan informasi yang dibutuhkan menjadi lebih cepat, dan cara mendapatkannya tidak memerlukan pengetahuan yang rumit atau pelatihan khusus. Dengan demikian, pelatihan yang biasanya dilakukan dengan biaya yang cukup besar dapat dihemat secara signifikan.

2.2.2 Arsitektur Pada *Business Intelligence*

Arsitektur pada *Business intelligence* memiliki empat komponen utama yaitu sebagai berikut (Silvana, Akbar, & Derisma, 2017) seperti ditunjukkan pada Gambar 2.1.



Gambar 2.1 Arsitektur *Business Intelligence*

a. *Data Warehouse*

Data warehouse merupakan sebuah landasan *Business Intelligence* baik data yang berjumlah maupun kecil. Sebelumnya *data warehouse* digunakan untuk menyimpan data yang terorganisir dan diringkas dengan tujuan untuk memudahkan pengguna dalam melihat dan memanfaatkannya. Namun sekarang penggunaan *data warehouse* telah menjadi sistem penyimpanan data real-time sehingga visualisasi yang ditampilkan berupa *data real-time*.

b. *Business Analytic*

Business analytics merupakan alat bantu untuk memudahkan pengguna dalam transformasi data, hasil data yang diseleksi akan disimpan ke dalam *warehouse* untuk dianalisis.

c. *Business Performance Management (BPM)*

Business Performance Management (BPM), digunakan untuk memantau dan menganalisis kinerja.

d. *User Interface (Dashboard)*

User Interface (Dashboard) menyediakan tampilan grafis atau gambar dari pengukuran performa perusahaan.

2.3 Extract, Transform, and Load (ETL)

ETL merupakan alat integrasi data dengan melibatkan penggalian dari sumber data luar (sistem operasional), mengubah data sesuai dengan kebutuhan bisnis dan akhirnya data tersebut dipindahkan ke gudang data. Selain menyeleksi data ETL dapat menyatukan kembali data-data yang relevan sebagai informasi yang berguna di gudang data (*warehouse*) (Oslan & Kristanto, 2019) sedangkan menurut Caserta (2004) ETL atau *extract, transformation, and loading* adalah suatu proses dasar dari data *warehouse*. Tolak ukur sistem ETL yang baik yaitu mampu mengekstrak data dari berbagai sumber serta dengan menerapkan standar dengan kualitas dan konsistensi yang sama sehingga data yang diperoleh dapat dianalisis dengan baik. Berikut merupakan penjelasan dari proses *extract, transformation, loading* (Hilmansyah & Maesaroh, 2022):

a. *Extraction*

Ekstraksi merupakan proses awal dari ETL pada proses ini dilakukan pemilihan dan pengambilan data dari satu sumber atau lebih. Langkah awal dari proses ini mengambil database dari sistem operasional perusahaan kemudian data tersebut dilakukan proses ekstraksi dari sumber aslinya dan dipindahkan ke data *warehouse* secara otomatis. Informasi mengenai data yang masuk ke dalam data *warehouse* didapatkan dari internal sistem (*legacy system*) maupun eksternal organisasi.

b. *Transform*

Transformasi merupakan aturan dalam proses ekstraksi data, pada proses ini terdapat aturan standar transformasi data yang melibatkan pengkodean data, perhitungan, penggabungan, pemisahan, dan pembuatan *Surrogate Key* atau *Primary Key*. *Output* dari transformasi adalah data bersih, konsisten dan siap dilakukan analisis.

c. *Loading*

Loading merupakan tahapan diantara data yang sudah bersih kemudian masuk pada target akhir. Proses *loading* memiliki jangka waktu yang disesuaikan kebutuhan organisasi. Beberapa data *warehouse* memiliki proses dalam jangka mingguan yang mana data dapat bertambah setiap minggunya untuk mengisi keseluruhan informasi yang ada secara kumulatif data diubah sementara data lain dapat menambahkan data baru dalam bentuk yang historikal, misalnya setiap jam data dapat bertambah.

2.4 Visualisasi Data

"Istilah 'Visualisasi data' dapat diartikan dalam beberapa cara, dengan sebagian besar definisi berfokus pada hubungan antara data dan teknologi komputer untuk mengubah data menjadi bentuk visual atau sonik. Proses dasar visualisasi data, dalam bentuk apa pun, dapat diubah menjadi gambar grafis. Ketika pengguna membaca atau melihat gambar grafis, gambar tersebut diinterpretasikan melalui sistem kognitif manusia untuk memperoleh wawasan atau pemahaman mengenai informasi yang bermanfaat. Visualisasi data dapat menggambarkan relasi dan pola antara variabel yang terdapat dalam data. Para pengambil keputusan dapat melihat koneksi antara data yang bersifat multidimensi. Tujuan utama dari visualisasi data adalah untuk menyampaikan informasi secara jelas dan efektif melalui representasi grafis. Visualisasi yang efektif biasanya terfokus, memberikan jawaban yang jelas, dan tidak terlalu terperinci" (Madyatmadja, Nuramalia, Kusumawati, Jamil, & Kusumawardhana, 2021). Secara umum, tujuan dari dilakukan visualisasi data adalah sebagai berikut:

- a. Meningkatkan pemahaman tentang suatu program, konteks, dan sejarahnya.
- b. Membantu pengumpulan data.
- c. Melakukan analisis terhadap beberapa bentuk data.
- d. Berkomunikasi dengan *stakeholder*.

2.5 *Business Intelligence Dashboard*

Dashboard merupakan suatu sistem informasi yang menyajikan informasi mengenai indikator utama dari aktivitas suatu organisasi dengan cara yang cepat dan ringkas. Biasanya, *dashboard* menciptakan model yang digunakan untuk mengembangkan alat pemantauan dan evaluasi kinerja sesuai kebutuhan suatu perusahaan. Proses pemantauan dan evaluasi kinerja dilakukan secara berkelanjutan oleh perusahaan untuk memastikan bahwa proses bisnis yang dijalankannya dapat mencapai tujuan yang telah ditetapkan melalui penerapan strategi pengelolaan yang tepat (Hariyanti, Werdiningsih, & Surendro, 2011). Visualisasi atau *dashboard* yang baik memiliki tahapan dan alasan dasar yang kuat dalam membuatnya (Satrio & Suryani, 2017):

- a. Identifikasi masalah
- b. Analisis dan perancangan
- c. Perancangan informasi *dashboard* / visualisasi
- d. Perancangan *prototype dashboard*

Tolok ukur dalam visualisasi data sama halnya dengan komunikasi berhasil atau tidaknya komunikasi tersebut ditentukan oleh bagaimana cara pembicara dalam menyampaikan informasi kepada lawan komunikasi. Tentunya visualisasi yang baik mampu memberikan informasi yang jelas dan tidak terlalu detail (Syaripul & Bachtiar, 2016).

2.6 **Jenis Grafik Untuk Visualisasi Data**

Grafik adalah kelompok representasi data yang matematis digunakan sebagai alat bantu maupun sarana bagi disiplin ilmu untuk mengungkapkan atau memvisualisasikan pernyataan verbal yang kompleks. Grafik dapat divisualisasikan dalam bentuk dua atau lebih variabel (Bunawan, Setiawan, Rusli, & Nahadi, 2015). Berikut merupakan jenis-jenis grafik:

- a. Grafik Batang

Grafik ini menampilkan kolom vertikal dengan menunjukkan perbandingan numerik antara kategori dan jumlah kolom tidak boleh terlalu besar, keunggulan dari grafik ini yaitu dapat memakai ketinggian kolom untuk menunjukkan perbedaan data karena mata manusia sangat sensitif terhadap perbedaan ketinggian, keterbatasan menggunakan grafik ini yaitu hanya cocok digunakan dengan data set ukuran kecil sampai medium.

- b. Grafik Balok

Grafik balok atau grafik bar hampir sama dengan grafik kolom, namun grafik ini memungkinkan penggunaan bar yang banyak.

c. Grafik Garis

Grafik garis merupakan alat visual yang berguna untuk menampilkan perubahan data dalam suatu rentang waktu atau kategori teratur. Grafik ini cenderung menunjukkan tren dan perubahan data seiring berjalannya waktu. Namun, penting untuk dicatat bahwa penggunaan grafik garis sebaiknya memperhatikan beberapa pertimbangan. Standar umum menyatakan bahwa jumlah data di dalam grafik garis sebaiknya melebihi dua agar dapat memberikan ukuran yang memadai. Selain itu, ketika jumlah *polyline* (garis) melebihi 5 pada satu grafik, grafik garis mungkin tidak cocok untuk data yang memiliki banyak klasifikasi atau kategori sehingga pemilihan jenis grafik yang sesuai perlu dipertimbangkan tergantung pada karakteristik dan jumlah data yang hendak disajikan.

d. Grafik Area

Grafik area yang dibentuk berdasarkan grafik garis, memiliki keunikan dengan mengisi area antara *polyline* dan sumbu (*axis*) dalam grafik garis menggunakan warna. Pengisian warna tersebut memiliki tujuan untuk memberikan klarifikasi tambahan terhadap informasi yang disajikan dalam grafik. Penggunaan warna pada grafik area dirancang untuk memperjelas hubungan antara serangkaian *polyline* yang berbeda. Pentingnya tingkat transparansi yang tepat pada warna-warna yang digunakan dalam grafik area adalah agar pengguna dapat dengan mudah mengamati dan memahami hubungan tumpang tindih antara berbagai seri data yang terdapat dalam grafik.

e. Grafik Pie

Grafik ini digunakan secara luas pada berbagai bidang untuk mempresentasikan proporsi dari setiap klasifikasi yang berbeda dan untuk membandingkan berbagai klasifikasi berdasarkan busurnya. Klasifikasi yang banyak akan cocok digunakan untuk grafik *pie* dengan perbedaan ukuran *pie* tiap klasifikasi menjadi tidak jelas. Grafik *Pie* dapat dibuat menjadi grafik *pie multi-layer* yang menunjukkan proporsi tiap klasifikasi dan diwaktu yang sama menampilkan relasi antar klasifikasi.

f. Scatter Plot

Menampilkan dua variabel dalam bentuk titik pada sistem koordinat persegi panjang. Posisi titik ditentukan oleh nilai variabel dengan mengamati distribusi data titik, pengguna dapat menyimpulkan korelasi antara variabel. Pembuatan scatter plot memerlukan banyak data, jika tidak korelasi menjadi tidak jelas.

g. Grafik Gelembung

Grafik gelembung adalah grafik dengan banyaknya variasi yang merupakan varian dari *scatter plot*. Perbedaannya ialah nilai variabel diwakili oleh sumbu X dan Y dan area disetiap balon mewakili data ketiga. Yang perlu diperhatikan pada grafik ini ukuran gelembung terbatas dan terlalu banyak gelembung akan membuat grafik sulit dibaca.

h. *Gauge*

Gauge dalam visualisasi data adalah sejenis grafik yang mematerialisasi skala mewakili metrik, pointer mewakili dimensi, dan sudut pointer menunjukkan nilai. Grafik ini dapat menampilkan progress atau situasi sebenarnya dari sebuah indikator. *Gauge* cocok digunakan dalam membandingkan interval.

i. Grafik Radar

Grafik radar digunakan untuk membandingkan beberapa variabel yang memiliki kuantitas tertentu, dengan tujuan memahami variabel mana yang memiliki nilai yang serupa atau terdapat nilai yang ekstrem. Grafik radar memungkinkan pengamatan terhadap variabel-variabel dalam satu set data untuk menentukan nilai yang lebih tinggi atau lebih rendah. Keunggulan dari grafik radar terletak pada kemampuannya untuk menunjukkan kinerja kerja, serta memiliki kolom yang dapat ditumpuk untuk membandingkan klasifikasi dan seri secara dua arah. Dengan demikian, grafik radar memungkinkan pengguna untuk melihat proporsi yang ada dalam suatu konteks tertentu.

j. Diagram Bingkai

Diagram bingkai adalah cara visual untuk menampilkan hirarki atau urutan dalam bentuk struktur pohon yang secara jelas menunjukkan hubungan hirarkis.

k. Diagram Pohon Persegi

Diagram pohon persegi panjang cocok dipakai untuk menampilkan data dengan hubungan hirarkis, secara visual dapat mencerminkan perbandingan antara tingkat yang sama. Jika dibandingkan dengan diagram pohon tradisional, diagram pohon persegi panjang membuat penggunaan ruang lebih efisien yang memiliki fungsi untuk menampilkan proporsi. Diagram pohon persegi panjang juga cocok untuk menampilkan hirarki dengan hubungan yang berat. Jika proporsi tidak perlu ditampilkan, diagram bingkai mungkin lebih jelas.

l. Grafik Corong

Grafik corong menunjukkan proporsi dari setiap tingkat dan secara visual mencerminkan ukuran masing-masing modul. Grafik ini cocok untuk membandingkan peringkat, di

saat yang sama grafik corong cocok digunakan untuk perbandingan. Kami mengatur beberapa diagram corong secara horizontal dan hasilnya perbedaan data nampak dengan sangat jelas.

m. Grafik *Word Cloud*

Word cloud adalah representasi visual dari data bentuk teks, tampilannya adalah grafik berwarna seperti awan yang terdiri dari banyak kosakata. Grafik ini dapat digunakan untuk menampilkan data bentuk teks dalam jumlah besar dan membantu pengguna teks yang paling menonjol dengan cepat. Grafik ini memerlukan jumlah data yang besar dan tingkat perbedaan data yang relatif besar, jika tidak, efeknya menjadi tidak jelas. Grafik ini tidak cocok dipakai untuk analisis yang akurat.

n. Grafik *Gantt*

Grafik *gantt* menampilkan waktu setiap tahap secara visual, progress yang actual dan perbandingan sesuai kebutuhan. Pengguna dapat dengan mudah memahami progress pada bidang organisasi.

o. Peta

Peta dapat dikelompokkan menjadi 3 jenis: peta daerah, peta titik, dan peta aliran.

p. *HeatMap*

HeatMap digunakan untuk menunjukkan bobot setiap titik di area geografis. Peta ini umumnya digunakan sebagai latar belakang ketebalan warna pada peta ini menunjukkan kepadatan yang terjadi pada beberapa wilayah.

2.7 *Machine Learning*

Machine learning, atau pembelajaran mesin, adalah penerapan kecerdasan buatan yang memberikan kemampuan sistem untuk belajar secara otomatis dari data yang ada, dengan tujuan melakukan tugas tertentu tanpa perlu diprogram secara eksplisit (Batta, 2020). Definisi lain oleh Rebala, Ravi, dan Churiwala (2019) menyebutkan bahwa *machine learning* merupakan bidang ilmu komputer yang mempelajari algoritma dan teknik untuk menghasilkan solusi otomatis pada masalah kompleks yang sulit diatasi dengan menggunakan metode pemrograman konvensional. Prosesnya melibatkan pencarian pola khusus pada setiap data atau dataset, dengan hasil berupa aturan-aturan yang dapat digunakan. Dalam pembelajaran *machine learning*, terdapat skenario-skenario seperti:

a. *Supervised Learning*

Penggunaan skenario *supervised learning* melibatkan penggunaan data pembelajaran yang sudah memiliki label. Dalam konteks ini, sistem belajar dari data yang telah diberi label untuk membuat prediksi pada data lainnya.

b. *Unsupervised Learning*

Pada skenario *unsupervised learning*, pembelajaran dilakukan dengan menggunakan data pembelajaran yang tidak memiliki label. Sistem kemudian mencoba mengelompokkan atau mengidentifikasi pola karakteristik dalam data tersebut tanpa adanya informasi label sebelumnya.

c. *Reinforcement learning*

Pada skenario *reinforcement learning*, fase pembelajaran dan tes saling dicampur. Proses ini melibatkan pembelajar yang secara aktif berinteraksi dengan lingkungan untuk mengumpulkan informasi. Pembelajar melakukan aksi, dan setiap aksi tersebut menghasilkan balasan atau konsekuensi. Dengan demikian, pembelajar belajar dari respons atau balasan yang diterima dan terus beradaptasi untuk meningkatkan kinerjanya dalam lingkungan tersebut.

2.8 Algoritma *Linear Regression*

Regresi linier adalah suatu metode perhitungan statistik yang digunakan untuk menilai dan memahami hubungan antara dua variabel. Metode ini membantu dalam menentukan sejauh apa variabel satu memengaruhi variabel lainnya. Dengan melakukan analisis regresi linier, peneliti dapat membuat peramalan nilai dari satu variabel berdasarkan nilai variabel lainnya dengan tingkat akurasi yang lebih baik (Katemba & Djoh, 2017). Dalam analisis regresi dikenal dua jenis variabel yaitu:

- a. Variabel tak bebas atau prediksi yang dilambangkan dengan Y merupakan variabel yang keadaannya dipengaruhi oleh keadaan variabel lain.
- b. Variabel bebas atau predictor yang dilambangkan dengan X merupakan variabel bebas yang keadaannya tidak dipengaruhi variabel lain.

Dalam penelitian ini, menggunakan regresi linier sederhana. Persamaan regresi linier sederhana (*simple analysis regeresi*), Regresi linier sederhana hanya memiliki satu variabel bebas X yang dihubungkan dengan satu variabel tidak bebas Y. Bentuk umum dari persamaan regresi linier sederhana adalah:

$$Y = b_0 + b_1X \quad (2.1)$$

Keterangan:

Y = variabel prediksi atau tak bebas (*dependent variable*)

X = variabel prediktor atau bebas (*independent variable*)

b₀ = konstanta (*intercept*)

b₁ = parameter koefisien regresi variabel bebas

Rumus untuk menentukan b₀ dan b₁ adalah

$$b_1 = \frac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum y^2)}$$

$$b_0 = y - b_1X \quad (2.2)$$

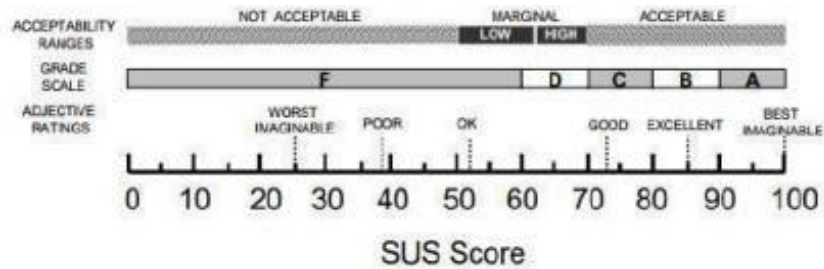
2.9 Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Python bisa dibilang bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat *source code* mudah dibaca. Python juga memiliki *library* yang lengkap sehingga memungkinkan *programmer* untuk membuat aplikasi yang mutakhir dengan menggunakan *source code* yang tampak sederhana (Parhusip, 2020).

2.10 System Usability Scale (SUS)

System Usability Scale (SUS) adalah kuesioner praktis yang digunakan untuk mendapatkan hasil terhadap pengujian *usability* sistem berdasarkan sudut pandang pribadi setiap pengguna (Sidik, 2018). Kuesioner tersebut memiliki 5 poin skala Likert “Sangat tidak setuju (STS)”, “Tidak setuju (TS)”, “Netral (N)”, “Setuju (S)”, dan “Sangat setuju (ST)”. Dengan menggunakan kuesioner SUS dapat mengukur sejauh mana pengguna merasa puas terhadap kinerja dan pengalaman menggunakan sistem yang telah dikembangkan. *System Usability Scale* (SUS) menghasilkan satu nomor mewakili ukuran gabungan dari kegunaan keseluruhan dari Sistem yang dipelajari. Perhatikan bahwa skor untuk setiap item yang tidak bermakna pada mereka sendiri. Untuk menghitung skor SUS, *sum* pertama kontribusi skor dari setiap item. Setiap item kontribusi skor akan berkisar dari 0 sampai 4. Untuk item

1,3,5,7, dan 9 kontribusi skor adalah skala posisi dikurangi 1. Untuk item 2,4,6,8 dan 10, kontribusi adalah 5 minus posisi skala. Kalikan jumlah nilai sebesar 2,5 untuk mendapatkan nilai keseluruhan SU (Manik, 2020). Skor SUS memiliki berbagai 0 sampai 100 seperti ditunjukkan pada Gambar 2.2.



Gambar 2.2 Skor SUS

BAB III METODOLOGI

3.1 Alat Penelitian

Pada penelitian ini menggunakan beberapa alat untuk memudahkan proses penelitian. Berikut merupakan beberapa alat penelitian yang ada:

a. Microsoft Excel

Software Microsoft excel digunakan sebagai format data yang dibutuhkan oleh peneliti, dengan format excel tersebut data dapat dianalisis dengan microsoft power BI.

b. Microsoft Power BI

Software tersebut digunakan untuk menganalisis data yang diperoleh dengan luaran berupa *dashboard* atau visualisasi.

c. Python sebagai *library* bahasa pemrograman *machine learning*.

3.2 Pengumpulan Data

3.2.1 Sumber Data

Sumber data penelitian ini berupa kumpulan dataset. Dataset ini berasal dari Kaggle.com dengan judul "Data Science Salaries 2023." Kaggle merupakan *platform online* yang terkenal dalam dunia data *science* dan analitika, yang menyediakan berbagai data set dan tantangan untuk para profesional dan penggemar data *science*.

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
2023	SE,FT	Principal	Data Scientist	80000	EUR	85847	ES,100	ES,L		
2023	MI,CT	ML Engineer		30000	USD	30000	US,100	US,S		
2023	MI,CT	ML Engineer		25500	USD	25500	US,100	US,S		
2023	SE,FT	Data Scientist		175000	USD	175000	CA,100	CA,M		
2023	SE,FT	Data Scientist		120000	USD	120000	CA,100	CA,M		
2023	SE,FT	Applied Scientist		222200	USD	222200	US,0	US,L		
2023	SE,FT	Applied Scientist		136000	USD	136000	US,0	US,L		
2023	SE,FT	Data Scientist		219000	USD	219000	CA,0	CA,M		
2023	SE,FT	Data Scientist		141000	USD	141000	CA,0	CA,M		
2023	SE,FT	Data Scientist		147100	USD	147100	US,0	US,M		
2023	SE,FT	Data Scientist		90700	USD	90700	US,0	US,M		
2023	SE,FT	Data Analyst		130000	USD	130000	US,100	US,M		
2023	SE,FT	Data Analyst		100000	USD	100000	US,100	US,M		
2023	EN,FT	Applied Scientist		213660	USD	213660	US,0	US,L		
2023	EN,FT	Applied Scientist		130760	USD	130760	US,0	US,L		
2023	SE,FT	Data Modeler		147100	USD	147100	US,0	US,M		
2023	SE,FT	Data Modeler		90700	USD	90700	US,0	US,M		
2023	SE,FT	Data Scientist		170000	USD	170000	US,0	US,M		
2023	SE,FT	Data Scientist		150000	USD	150000	US,0	US,M		
2023	MI,FT	Data Analyst		150000	USD	150000	US,100	US,M		
2023	MI,FT	Data Analyst		110000	USD	110000	US,100	US,M		
2023	SE,FT	Research Engineer		275000	USD	275000	DE,0	DE,M		

Gambar 3.1 Data Salaries

Adapun data gaji di atas memiliki atribut:

- a. *Work year* (Tahun kerja)
- b. *Experience level* (Level pengalaman)
- c. *Employment type* (Tipe perusahaan)
- d. *Job title* (Nama pekerjaan)
- e. *Salary* (Gaji)
- f. *Salary in USD* (Gaji dalam satuan USD)
- g. *Salary currency* (Mata uang gaji)
- h. *Employee residence* (Kabupaten/kota pegawai)
- i. *Remote ratio*
- j. *Company location* (Lokasi perusahaan)
- k. *Company size* (Ukuran perusahaan)

3.2.2 Teknik Pengumpulan Data

Pada pengumpulan data ini dilakukan dengan metode berdasarkan kebutuhan penelitian. Berikut merupakan beberapa teknik pengumpulan data yang digunakan dalam penelitian ini:

- a. Studi Literatur

Dalam penelitian ini, studi literatur dilakukan dengan mengumpulkan informasi dari penelitian terkait. Hal ini bertujuan untuk melengkapi data yang telah ada dan memperkuat dasar penelitian.

- b. Observasi

Observasi merupakan metode pengumpulan data yang melibatkan pengamatan dan pencatatan sistematis terhadap kejadian atau objek yang menjadi fokus observasi. Dalam konteks penelitian ini, peneliti melakukan observasi dengan merujuk pada data yang tersedia di Kaggle.com dan juga mencari data di luar perusahaan yang relevan dengan kebutuhan penelitian.

3.3 Data yang Dianalisis

Pemilihan data-data yang digunakan berdasarkan hasil observasi dan studi literatur untuk menentukan wawasan informasi yang diperlukan. Data yang digunakan yang berasal dari Kaggle.com adalah berupa data set *salaries* dari berbagai perusahaan di dunia dari tahun 2020-2023. Ini dilakukan untuk mengetahui wawasan yang ada dari keterkaitan pengaruh

terhadap atribut lain seperti lama kerja, tipe perusahaan, jenis pekerjaan dan lainnya. Selain itu juga digunakan jurnal atau artikel terkait yang berasal dari luar perusahaan sebagai referensi.

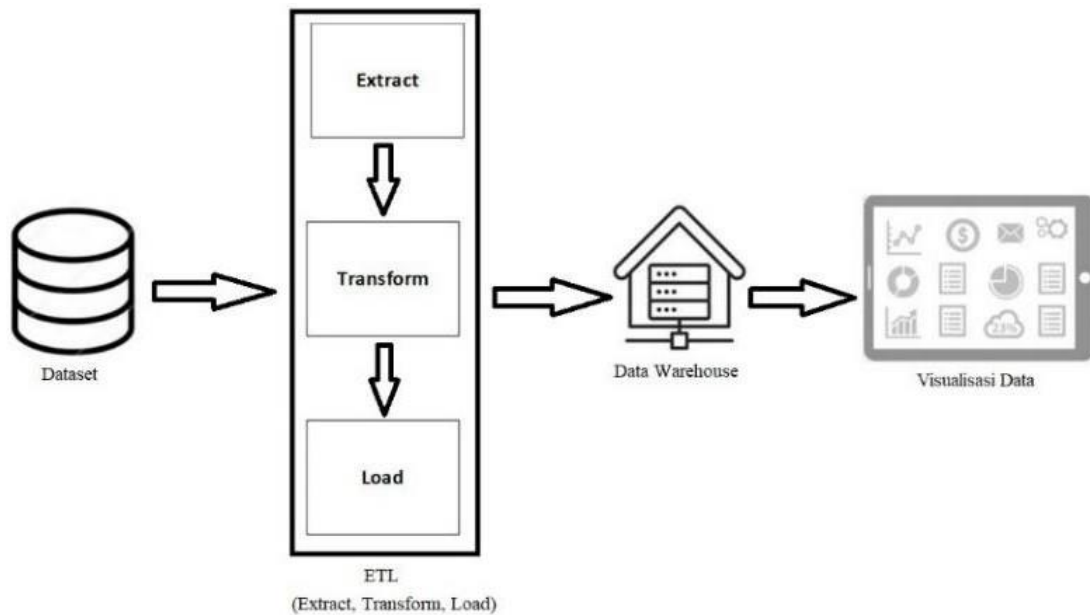
3.4 Informasi yang Dibutuhkan

Dalam rangka mengetahui informasi yang diperlukan, penelitian ini mengimplementasikan observasi dan studi literatur. Observasi dilakukan untuk memahami aspek-aspek yang terkait dengan pemberian gaji, sementara studi literatur dilakukan untuk menganalisis kebutuhan informasi perusahaan, terutama faktor-faktor yang mempengaruhi besarnya pemberian gaji. Melalui kedua metode tersebut, peneliti dapat mengidentifikasi kebutuhan informasi yang relevan diantaranya:

1. Kebutuhan informasi tentang jumlah gaji tahun 2020-2023.
2. Kebutuhan informasi tentang jumlah gaji berdasarkan *work year*
3. Kebutuhan informasi tentang jumlah gaji berdasarkan *experience level*
4. Kebutuhan informasi tentang jumlah gaji berdasarkan *employment type*
5. Kebutuhan informasi tentang jumlah gaji berdasarkan *job title*
6. Kebutuhan informasi tentang jumlah gaji berdasarkan *employee residence*
7. Kebutuhan informasi tentang jumlah gaji berdasarkan *company location*
8. Kebutuhan informasi tentang jumlah gaji berdasarkan *company size*.
9. Kebutuhan informasi tentang prediksi jumlah gaji yang akan datang.

3.5 Metode Penelitian

Diagram alir penelitian “Implementasi Intelegen Bisnis Dengan Visualisasi Data Gaji dan Algoritma Linear Regresion”, ditunjukkan melalui diagram pada Gambar 3.2.



Gambar 3.2 Diagram Metode Penelitian

3.5.1 Dataset

Dataset ini berasal dari Kaggle.com dengan judul "Data Science Salaries 2023." Kaggle merupakan platform online yang terkenal dalam dunia data science dan analitika, yang menyediakan berbagai data set dan tantangan untuk para profesional dan penggemar data science. Dataset "Data Science Salaries 2023" merupakan salah satu sumber informasi penting yang digunakan oleh para ilmuwan data dan analis untuk memahami tren gaji di industri data science pada tahun 2023. Platform ini memfasilitasi akses ke berbagai data yang relevan dan memungkinkan analisis mendalam tentang perbandingan industri (Das, Barik, & Mukherjee, 2020). Data pelatihan memiliki 3.755 baris dan 11 kolom, yang mencakup periode waktu dari tahun 2020 hingga 2023. Atribut dari dataset ini mencakup *work year*, *experience level*, *employment type*, *job title*, *salary*, *salary in USD*, *salary currency*, *employee residence*, *remote ratio*, *company location* dan *company size*.

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	compa
0	2023	SE	FT Principal Data Scientist	80000	EUR	85847	ES	100	ES	
1	2023	MI	CT ML Engineer	30000	USD	30000	US	100	US	
2	2023	MI	CT ML Engineer	25500	USD	25500	US	100	US	

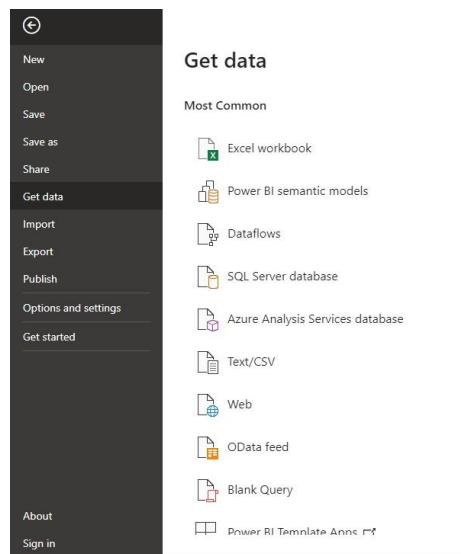
Gambar 3.3 Data Pelatihan

3.5.2 Pengolahan Data

Tahap ini dilakukan analisis data yang diperoleh selama penelitian yang berasal dari Kaggle.com berupa dataset Data Science Salaries 2023.

a. *Extract*

Proses ekstraksi data merupakan proses awal yang dilakukan pada saat melakukan pengembangan *dashboard*. Ekstraksi data merupakan proses yang dilakukan untuk mengambil data dari data *source* yang diinginkan, seperti *database*, *file*, *cloud* dan juga bisa diekstraksi menggunakan *script*. Power BI menyediakan fitur bernama *get data*, fitur ini dapat digunakan untuk melakukan berbagai macam ekstraksi data dengan berbagai cara seperti pada Gambar 3.4.



Gambar 3.4 Tampilan *Get Data* Power BI

Proses ekstraksi pada penelitian ini dilakukan dengan menggunakan *excel workbook*. Setelah data yang ingin digunakan sudah sesuai dengan yang ingin digunakan, proses ekstraksi dilakukan pada Power BI dengan memasukkan *script* ke dalam editor. Setelah proses ekstraksi data selesai maka data tersebut akan tertampil di Power BI seperti pada Gambar 3.5.

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location
2023	SE	FT	Data Engineer	253200	USD	253200	US	0	US
2023	SE	FT	Data Engineer	90700	USD	90700	US	0	US
2023	SE	FT	Data Engineer	270703	USD	270703	US	0	US
2023	SE	FT	Data Engineer	221484	USD	221484	US	0	US
2023	SE	FT	Data Engineer	238000	USD	238000	US	0	US
2023	SE	FT	Data Engineer	176000	USD	176000	US	0	US
2023	SE	FT	Data Engineer	115000	USD	115000	US	0	US
2023	SE	FT	Data Engineer	81500	USD	81500	US	0	US
2023	SE	FT	Data Engineer	185000	USD	185000	US	0	US
2023	SE	FT	Data Engineer	140000	USD	140000	US	0	US
2023	SE	FT	Data Engineer	165000	USD	165000	US	0	US
2023	SE	FT	Data Engineer	132300	USD	132300	US	0	US
2023	SE	FT	Data Engineer	179170	USD	179170	US	0	US
2023	SE	FT	Data Engineer	94300	USD	94300	US	0	US
2023	SE	FT	Data Engineer	247300	USD	247300	US	0	US
2023	SE	FT	Data Engineer	133800	USD	133800	US	0	US
2023	SE	FT	Data Engineer	185900	USD	185900	US	0	US
2023	SE	FT	Data Engineer	129300	USD	129300	US	0	US
2023	SE	FT	Data Engineer	252000	USD	252000	US	0	US
2023	SE	FT	Data Engineer	129000	USD	129000	US	0	US
2023	SE	FT	Data Engineer	145000	USD	145000	US	0	US
2023	SE	FT	Data Engineer	115000	USD	115000	US	0	US

Gambar 3.5 Tampilan Tabel Data Yang Digunakan

b. *Transform*

Setelah proses ekstraksi telah selesai, maka data tersebut harus dipindahkan pada sistem perantara atau sistem target agar bisa segera diproses lebih lanjut. Selanjutnya proses ini dinamakan dengan transformasi. Proses ini akan membantu kamu membuat gudang data terstruktur. Proses transformasi ini merupakan pembersihan dan mempersiapkan agregasi untuk analisis. Contoh umum dari transformasi adalah mengubah data menjadi tipe numerik. Hal ini dimaksudkan untuk memastikan bahwa semua input yang diberikan kepada model menjadi data numerik, sehingga membuat sistem semakin mudah dioperasikan. Data non-numerik dihapus selama proses ini, untuk memastikan bahwa model tersebut dapat menerima data masukan yang kompatibel dengan bentuk yang diharapkan. Proses ini sangat penting karena nantinya akan membantu memastikan data yang akan diolah sepenuhnya siap dan kompatibel. Proses transformasi terbagi menjadi beberapa proses, diantaranya pemberian, standarisasi, deduplikasi, verifikasi, pengurutan dan tugas lainnya. Proses transformasi dilakukan dengan menggunakan salah satu *tools* pada Power BI, yaitu *power query*. Dengan menggunakan *power query*, proses transformasi dapat dilakukan dengan menggunakan aplikasi yang sama sehingga memudahkan untuk melakukan pengolahan data.

Setelah dataset bersih dan terstruktur dengan baik, maka proses selanjutnya adalah menerapkan algoritma regresi linier. Tujuannya adalah untuk mengeksplorasi dan memahami hubungan antara lama pengalaman kerja dan besaran gaji yang diterima oleh para profesional di bidang data. Analisis ini membantu kami dalam mengidentifikasi pola dan tren yang relevan, serta faktor-faktor penting yang mempengaruhi tingkat gaji di industri ini. Dalam langkah selanjutnya, menggunakan Python, kami mengembangkan itemset yang mengkategorikan data berdasarkan durasi pengalaman kerja dan gaji. Pendekatan ini memungkinkan kami untuk melakukan analisis lebih lanjut dan mendalam terkait bagaimana pengalaman berpengaruh terhadap struktur gaji. Itemset ini membantu dalam menyajikan gambaran yang lebih jelas dan terstruktur tentang distribusi gaji di berbagai tingkat pengalaman. Melalui analisis ini, kami berupaya memberikan wawasan yang berguna bagi para profesional di bidang data science, termasuk bagaimana perkembangan karir dan kenaikan gaji dapat berkorelasi dengan peningkatan pengalaman kerja mereka.

c. *Load*

Proses "*load*" merujuk pada langkah untuk memuat data yang telah mengalami transformasi sebelumnya ke dalam data *warehouse Power BI*. Setelah data melalui proses transformasi di *Power Query*, langkah selanjutnya adalah melakukan proses "*load*" dengan menerapkan proses "*apply*" pada *Power Query*. Data yang telah melalui proses ini akan secara otomatis disimpan dalam data *warehouse Power BI*. Selanjutnya, data ini dapat digunakan untuk membuat visualisasi dalam bentuk *dashboard* pada *Power BI*.

3.5.3 Data Warehouse

Pada penelitian ini *Data Warehouse* yang digunakan oleh peneliti yaitu Microsoft Excel. Data yang disimpan pada Microsoft excel yaitu data historis data salaries tahun 2020- 2023.

3.5.4 Visualisasi Data

Setelah melalui proses ETL (*Extract, Transform, Load*), langkah berikutnya yang krusial dalam manajemen data adalah melakukan visualisasi data. Pada proses visualisasi ini juga dilakukan analisis data menggunakan algoritma regresi linear untuk memprediksi gaji di masa mendatang menggunakan bantuan Bahasa pemrograman *python* sebagai *library machine learning*. Hasil dari proses analisis tersebut kemudian ditampilkan pada proses visualisasi data. Visualisasi data memanfaatkan elemen visual seperti grafik, diagram, dan peta untuk menggambarkan informasi yang terkandung dalam dataset. Hal ini bertujuan untuk menyajikan data yang pada awalnya mungkin terasa monoton dalam bentuk tabel angka menjadi representasi visual yang lebih menarik dan informatif. Proses visualisasi membuka pintu bagi pembaca untuk lebih mudah memahami informasi, mendapatkan wawasan, dan meresapi cerita yang tersembunyi dalam data. Keberhasilan visualisasi data terletak pada kemampuannya membantu pembaca dalam melihat, berinteraksi, dan memahami data dengan lebih baik. Visualisasi, baik yang sederhana maupun kompleks, memiliki daya tarik untuk membawa semua pembaca ke arah pemahaman yang seragam, tanpa memandang tingkat keahlian mereka dalam membaca data. Pemilihan jenis grafik atau diagram yang tepat memainkan peran penting dalam merancang visualisasi yang efektif. Berbagai jenis visualisasi data, seperti diagram garis, batang, *pie*, *hash tree*, dan *network*, dapat digunakan untuk menyampaikan informasi dengan cara yang paling efektif. Setiap jenis

visualisasi memiliki keunggulannya sendiri dalam memfasilitasi pembaca untuk menginterpretasikan suatu dataset. Oleh karena itu, penting untuk memilih jenis visualisasi yang sesuai dengan karakteristik data yang akan disajikan agar pesan yang ingin disampaikan dapat tersampaikan secara optimal.

3.5.5 Pengujian

Untuk dapat mengetahui apakah desain *dashboard* informasi yang dibuat sesuai dengan kebutuhan perusahaan dan dapat digunakan dengan baik, maka perlu dilakukan pengujian validasi desain. Adapun metode yang digunakan *System Usability Scale* (SUS). Dalam uji ini digunakan kuesioner yang berisikan pertanyaan yang bertujuan untuk mengukur kebergunaan desain *dashboard* informasi yang telah dibuat. Kuesioner ini akan diisi oleh pakar atau pemegang jabatan dari perusahaan. *Draft* pertanyaan untuk kuisisioner ditunjukkan pada Tabel 3.1.

Tabel 3.1 Draft Kuisisioner

No	Pernyataan
1	Tulisan teks, warna dan visualisasi data (diagram dan chart) yang digunakan pada <i>dashboard</i> mudah dan jelas dipahami
2	Saya merasa kesulitan dalam memahami isi dan konten informasi yang disajikan
3	Anda dengan mudah dan cepat menerima informasi secara detail dan juga spesifik
4	Saya membutuhkan bantuan orang lain untuk menggunakan <i>dashboard</i> ini
5	Saya merasa mudah menggunakan <i>dashboard</i> pada aplikasi <i>Microsoft Power BI</i>
6	Saya merasa navigasi <i>dashboard</i> sulit untuk diingat dan digunakan
7	Saya merasa nyaman dalam menggunakan <i>dashboard</i> pada aplikasi <i>Microsoft Power BI</i>
8	Saya merasa <i>dashboard</i> ini sulit untuk digunakan
9	Paduan warna dan tata letak konten nyaman untuk dilihat
10	Saya perlu membiasakan diri terlebih dahulu sebelum menggunakan <i>dashboard</i> ini

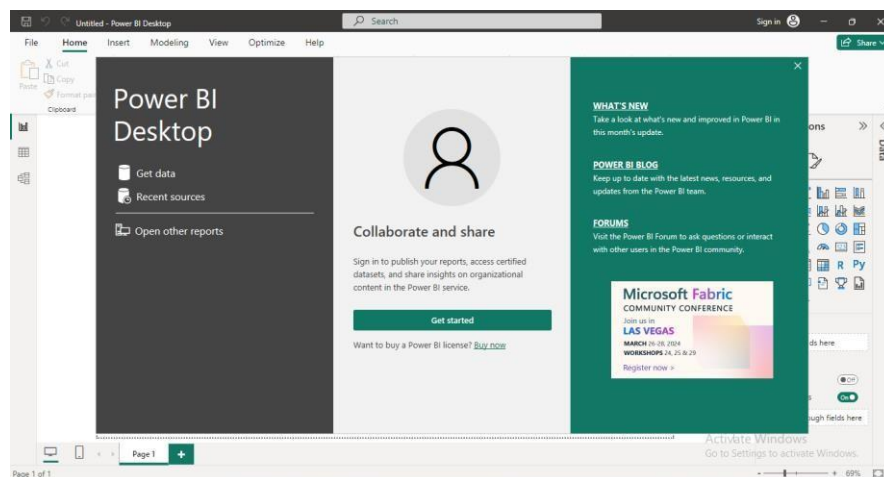
BAB IV

HASIL DAN PEMBAHASAN

4.1 Implementasi *Business Intelligence*

Pada sub bab ini akan menjelaskan proses *importing* data menggunakan *Software Microsoft Power BI*. Berikut ini merupakan langkah-langkah dalam melakukan pengolahan data menggunakan *Microsoft Power BI*.

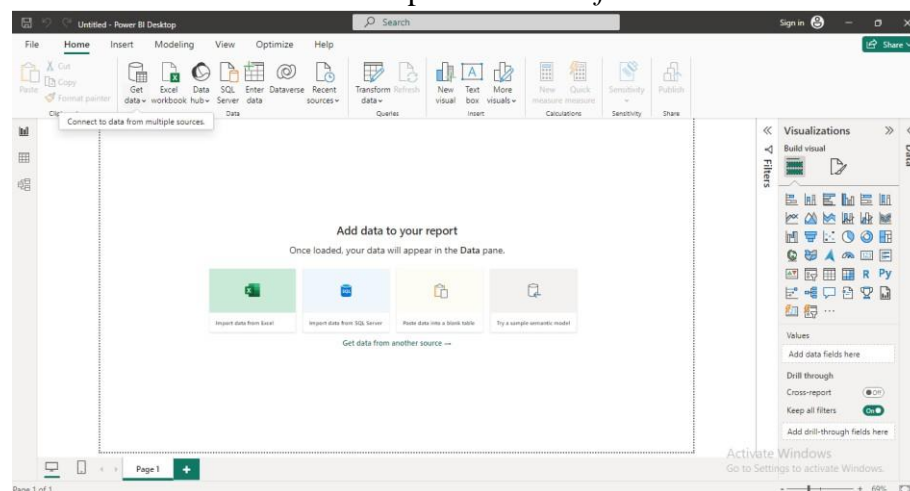
- Membuka *Microsoft Power BI* dengan *double* klik pada aplikasi hingga muncul tampilan seperti Gambar 4.1.



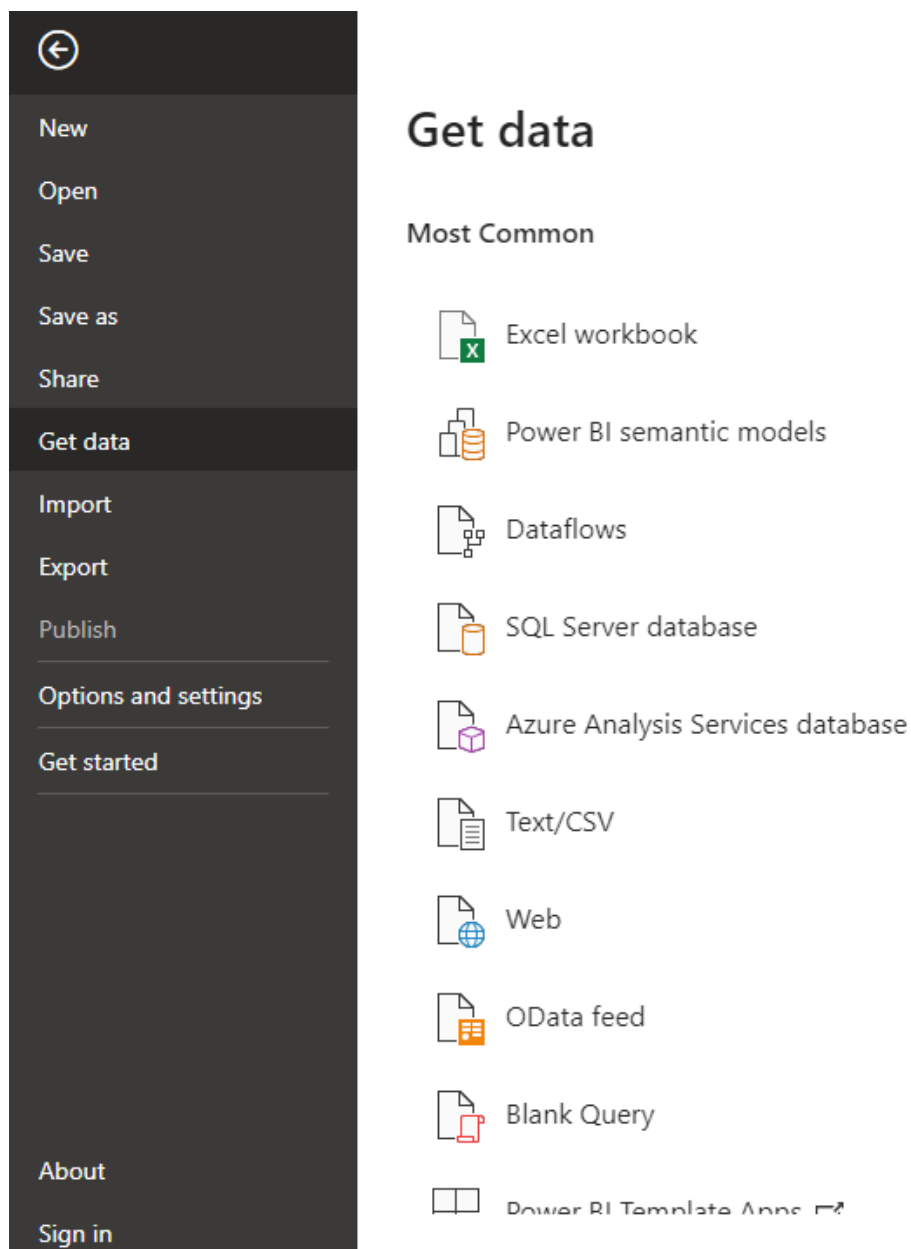
Gambar 4.1 Tampilan Awal *Microsoft Power BI*

- Lalu akan muncul tampilan menu, kemudian klik *Get Data* pada beranda untuk memasukan data pada aplikasi *Microsoft Power BI* seperti ditunjukkan pada Gambar 4.2.

Gambar 4.2 Tampilan *Microsoft Power BI*

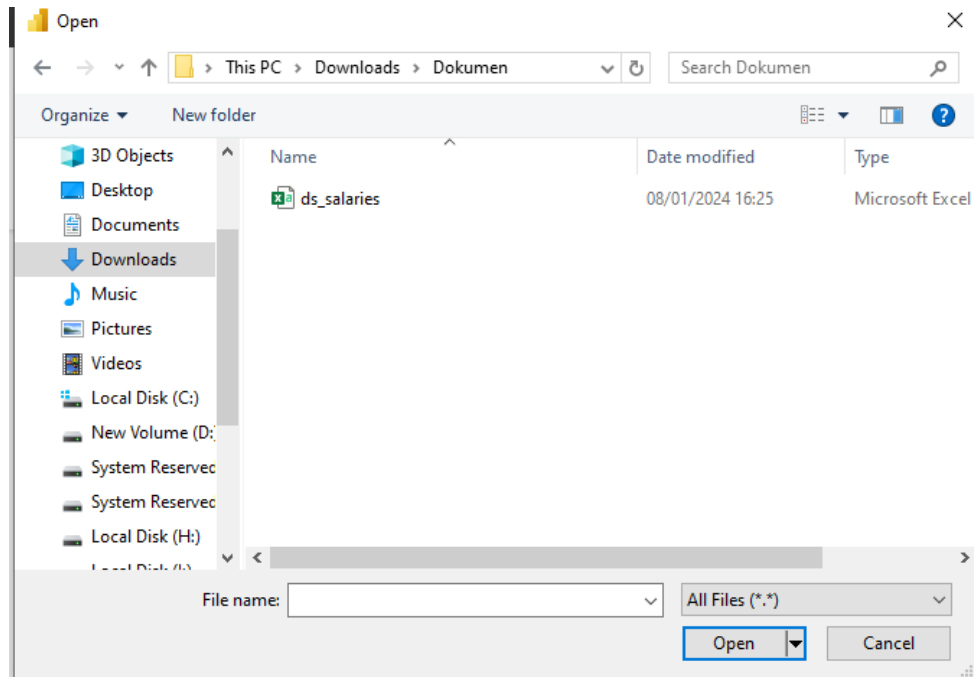


- c. Pada menu *Get Data* pilih *Excel Workbook* untuk menyesuaikan format yang digunakan pada data yang diolah lalu klik *Connect* seperti ditunjukkan pada Gambar 4.3.



Gambar 4.3 Tampilan *Microsoft Power BI*

- d. Memilih data yang akan dimasukkan pada aplikasi, klik data yang ingin dimasukkan lalu klik *Open* seperti ditunjukkan pada Gambar 4.4.



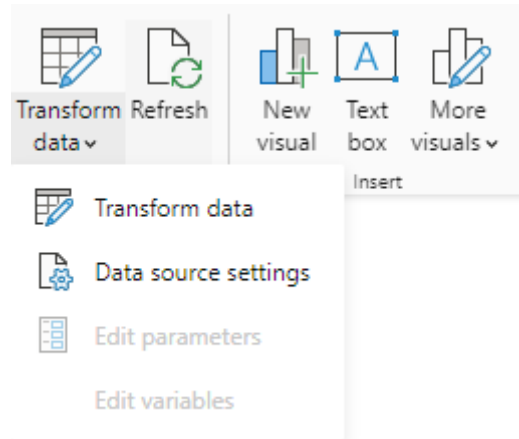
Gambar 4.4 Memasukkan Data ke Aplikasi

- e. Selanjutnya menentukan data yang akan dimasukkan ke dalam aplikasi dengan mengklik ceklis pada kolom kemudian klik *load* seperti ditunjukkan pada Gambar 4.5.

work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence
2023	SE	FT	Principal Data Scientist	80000	EUR	85847	ES
2023	MI	CT	ML Engineer	30000	USD	30000	US
2023	MI	CT	ML Engineer	25500	USD	25500	US
2023	SE	FT	Data Scientist	175000	USD	175000	CA
2023	SE	FT	Data Scientist	120000	USD	120000	CA
2023	SE	FT	Applied Scientist	222200	USD	222200	US
2023	SE	FT	Applied Scientist	136000	USD	136000	US
2023	SE	FT	Data Scientist	219000	USD	219000	CA
2023	SE	FT	Data Scientist	141000	USD	141000	CA
2023	SE	FT	Data Scientist	147100	USD	147100	US
2023	SE	FT	Data Scientist	90700	USD	90700	US
2023	SE	FT	Data Analyst	130000	USD	130000	US
2023	SE	FT	Data Analyst	100000	USD	100000	US
2023	EN	FT	Applied Scientist	213660	USD	213660	US
2023	EN	FT	Applied Scientist	130760	USD	130760	US
2023	SE	FT	Data Modeler	147100	USD	147100	US
2023	SE	FT	Data Modeler	90700	USD	90700	US
2023	SE	FT	Data Scientist	170000	USD	170000	US
2023	SE	FT	Data Scientist	150000	USD	150000	US
2023	MI	FT	Data Analyst	150000	USD	150000	US

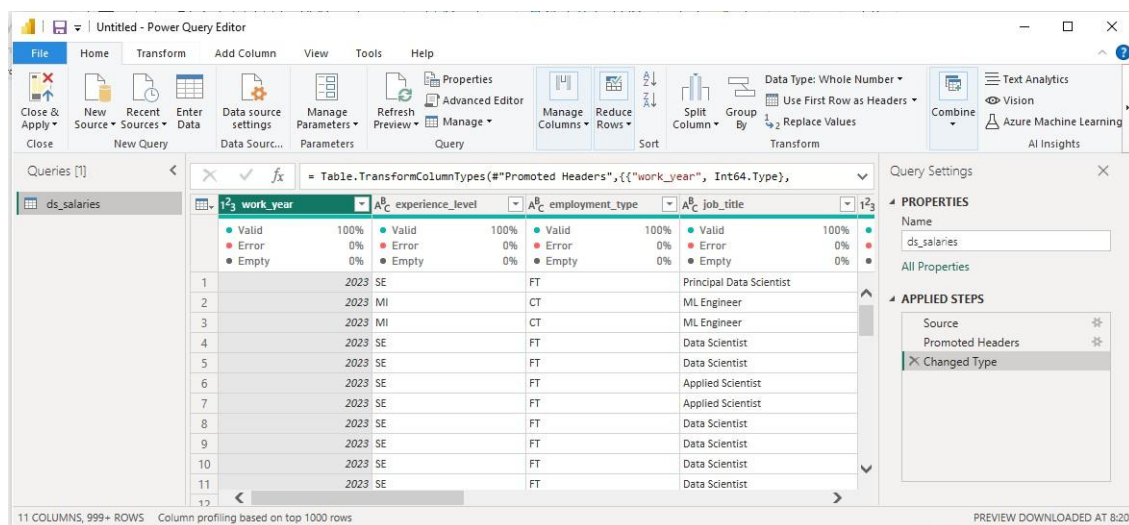
Gambar 4.5 Proses Menentukan Data

- f. Setelah memasukkan data pada aplikasi, langkah selanjutnya yaitu melakukan pengecekan kembali pada aplikasi untuk menghindari *error* pada hasil visualisasi, dengan cara mengklik *Transform Data*, dan mengklik lagi *Transform data*.



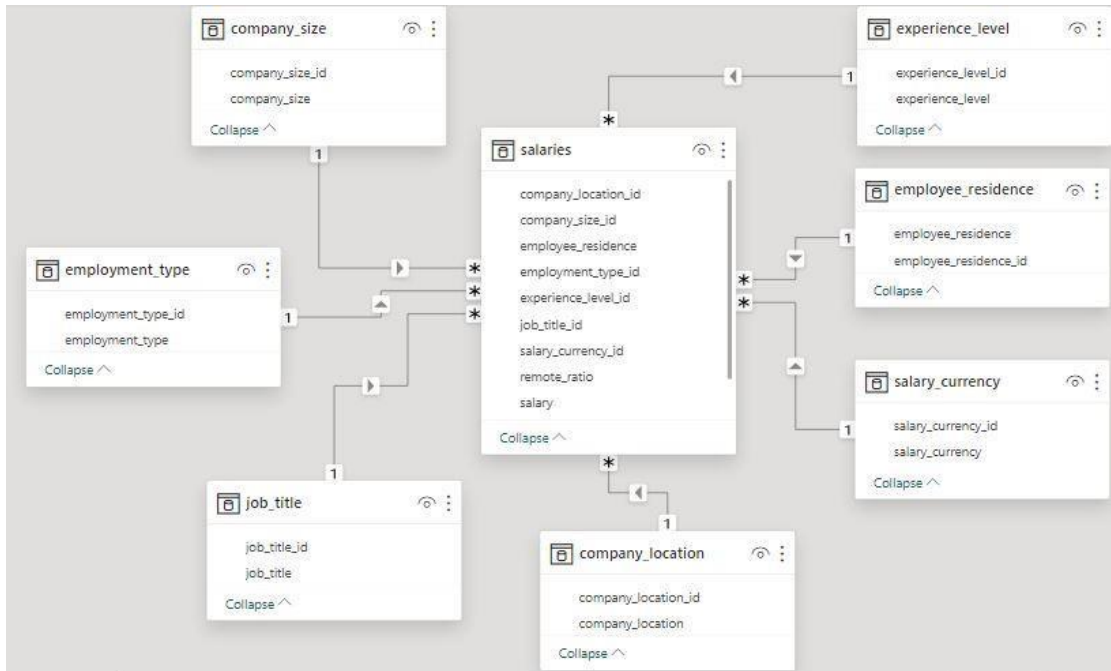
Gambar 4.6 Menu *Transform data*

- g. Setelah selesai dilakukan pengecekan klik *cancel* dengan cara klik simbol X pada bagian atas kanan menu *Power Query Editor*.



Gambar 4.7 Menu *Power Query Editor*

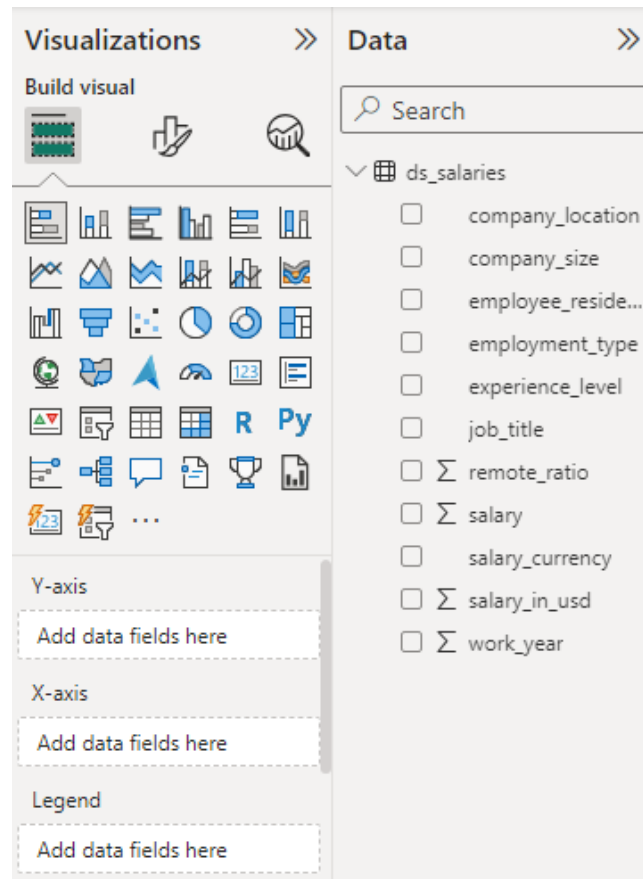
Setelah proses transformasi data selesai maka data perlu diberi relasi antar tabel terlebih dahulu sebelum digunakan. Pada Power BI terdapat menu *Manage Relationship* yang digunakan untuk mengatur *data modeling* yang akan diolah. Hal ini akan membantu dalam melakukan visualisasi data nantinya. Hubungan antar tabel ditunjukkan pada Gambar 4.8.



Gambar 4.8 Data Modeling

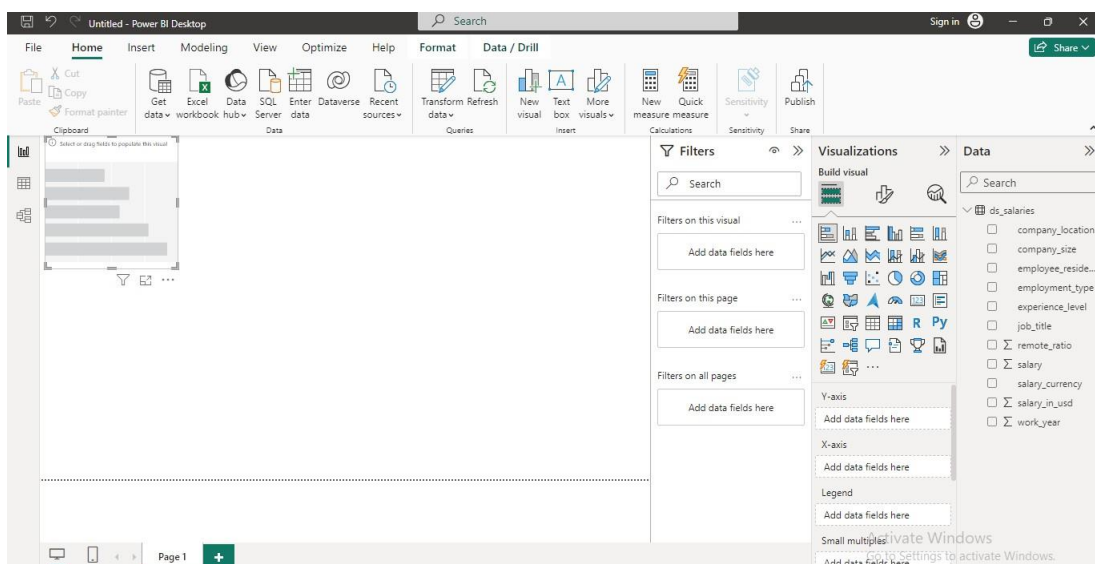
Model ini adalah *skema bintang* khas yang dapat dilihat dari gudang data. Model ini menyerupai bintang. Bagian tengah bintang adalah tabel Fakta. Tabel di sekitarnya disebut tabel Dimensi, yang berhubungan dengan tabel Fakta. Tabel Fakta berisi informasi numerik tentang *salaries*, seperti jumlah *salary* dan tahun kerja. Tabel dimensi terdiri dari tabel *company_size*, *employment_type*, *job_title*, *company_location*, *salary_currency*, *employee_residence* dan *experience_level*. Relasi yang digunakan dalam tabel ini adalah relasi 1-n yang berarti satu data boleh memiliki lebih dari satu data lainnya.

- h. Setelah proses *modeling data* selesai maka dapat dilakukan visualisasi *dashboard* dengan cara memilih grafik pada menu *Visualizations Build Visual* lalu memilih kategori data pada menu *fields* yang akan digunakan untuk dianalisis menggunakan aplikasi dengan cara mengklik pada kategori data hingga muncul tanda ceklis.



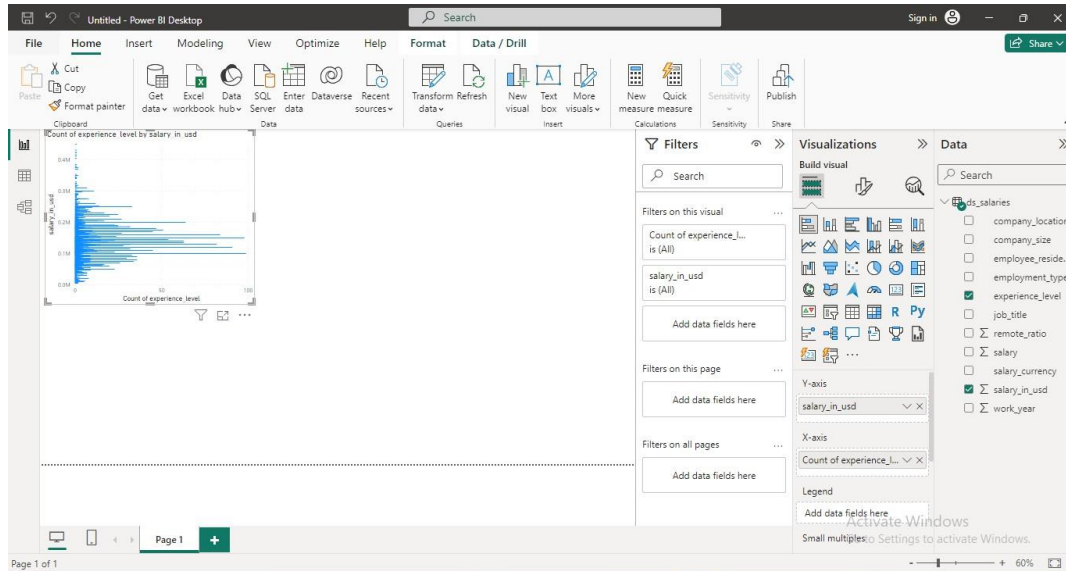
Gambar 4.9 Menu *Visualizations Build Visual*

Setelah ditentukan grafik dan dipilih kategori data maka secara otomatis data akan diubah dalam bentuk visualisasi seperti pada Gambar 4.10.



Gambar 4.10 Tampilan Awal Grafik

Langkah terakhir yaitu mengatur ukuran tampilan visualisasi dengan cara klik pada visualisasi kemudian sesuaikan dengan klik pada setiap sisi visualisasi seperti pada Gambar 4.11 Tampilan Pada Microsoft Power BI.

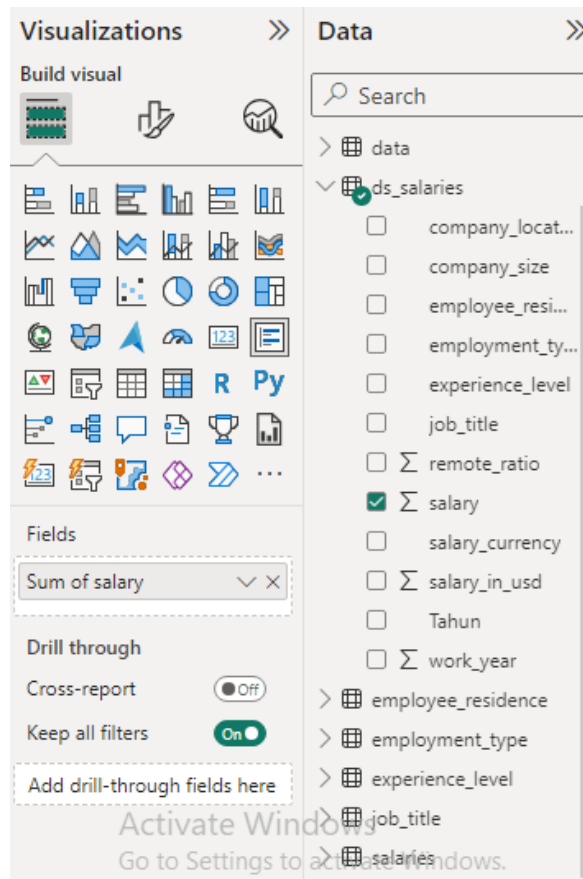


Gambar 4.11 Tampilan Grafik Pada Microsoft Power BI

Pada bagian bagian hasil visualisasi data terdapat beberapa operasi *dimensional data modeling* seperti:

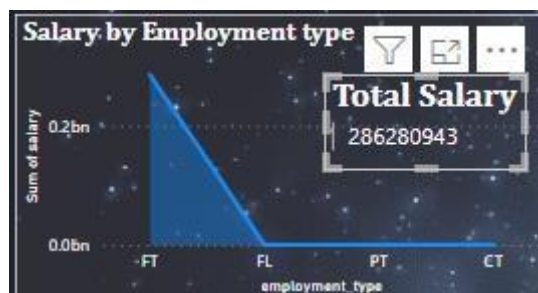
a. Agregat

Operasi agregat menampilkan hasil perhitungan matematis yang dapat berupa penjumlahan, rata-rata, nilai maksimal, nilai minimal atau lainnya. Misalkan menjumlahkan total gaji berdasarkan *employment type* dengan cara memilih model visualisasi data yang diinginkan dan dilanjutkan dengan memilih kolom *salary*. Namun sebelumnya tipe data kolom *salary* harus berupa data numeric agar dapat dilakukan penjumlahan dengan fungsi SUM seperti yang ditunjukkan pada Gambar 4.12.



Gambar 4.12 Operasi Agregate

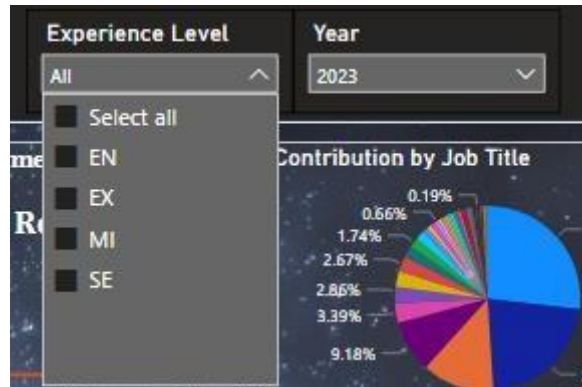
Hasil dari operasi agregat di atas akan menampilkan jumlah total data *salary* seperti yang ditunjukkan pada Gambar 4.13.



Gambar 4.13 Visualisasi Data Operasi Agregat

b. *Slice*

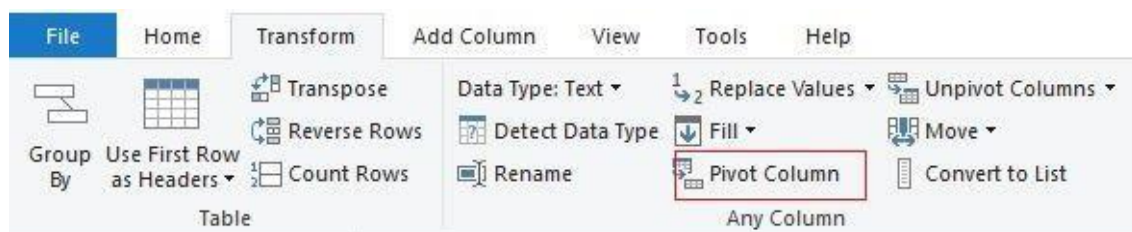
Operasi *slice* digunakan untuk memotong/memfilter data berdasarkan kategori tertentu. Pada data gaji ini, operasi *slice* digunakan untuk menampilkan data berdasarkan *experience level* dan *work year* (tahun kerja) seperti yang ditunjukkan pada Gambar 4.14.

Gambar 4.14 Visualisasi Data Operasi *Slicer*c. *Drill down*

Operasi *drill down* digunakan untuk menampilkan detail data secara spesifik seperti 4 negara (*company location*) yang memiliki pengeluaran gaji paling banyak seperti yang ditunjukkan pada Gambar 4.15.

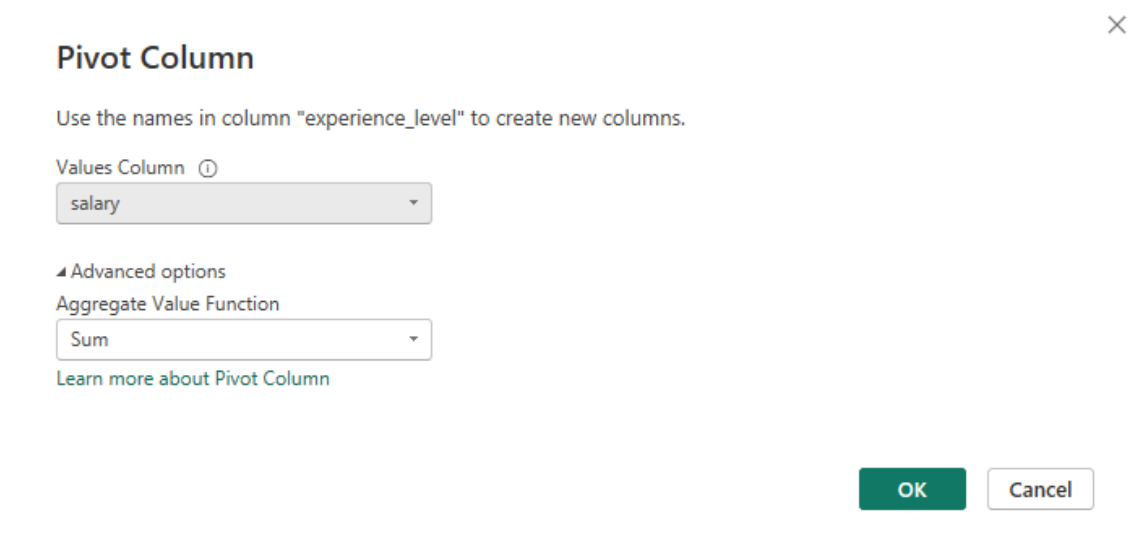
Gambar 4.15 Visualisasi Data Operasi *Drill Down*d. *Pivot*

Di Power Query dapat membuat tabel yang berisi nilai agregat untuk setiap nilai unik dalam kolom. Power Query mengelompokkan setiap nilai unik, melakukan perhitungan agregat untuk setiap nilai, dan mempivot kolom ke dalam tabel baru dengan cara memilih ikon pivot tabel pada Power Query. Pada tab Transformasi di grup kolom apa pun, pilih kolom pivot seperti ditunjukkan pada Gambar 4.16.

Gambar 4.16 Menu *Pivot* Tabel

Dalam kotak dialog Kolom pivot, di daftar kolom nilai, pilih *salary*. Secara *default*, *Power Query* akan mencoba melakukan jumlah sebagai agregasi, tetapi bisa memilih

opsi tingkat lanjut untuk melihat agregasi lain yang tersedia seperti SUM seperti ditunjukkan pada Gambar 4.17.



Gambar 4.17 Proses Pemilihan Pivot Tabel

Hasil dari pivot tabel ditampilkan seperti pada Gambar 4.17.

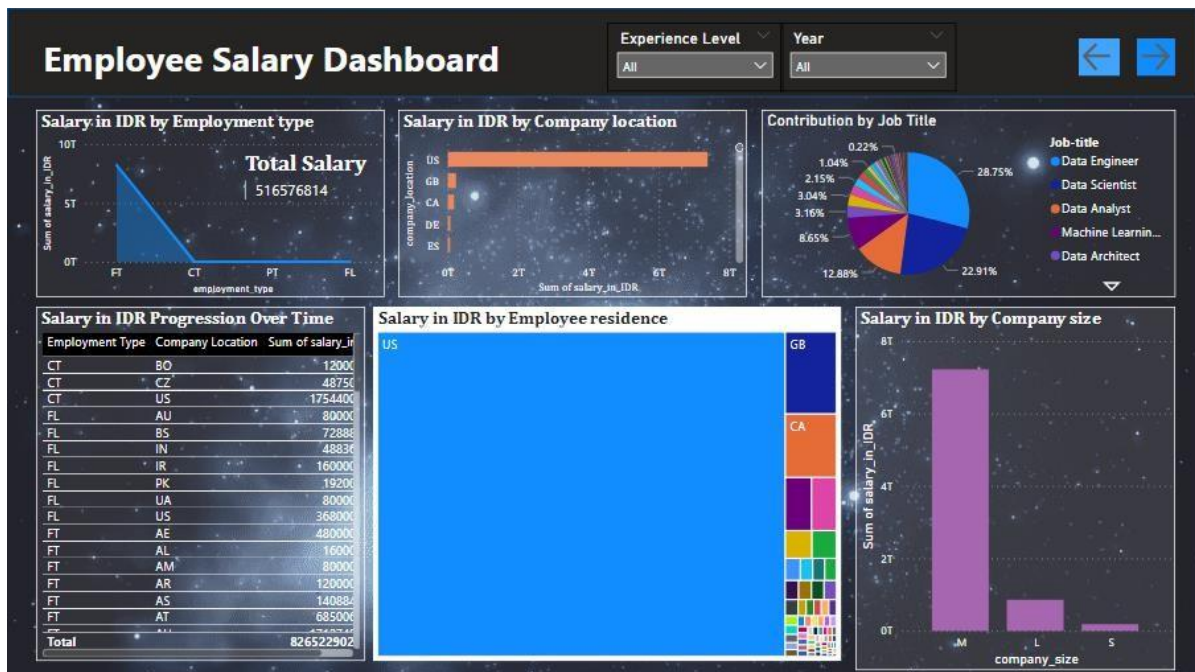
	A _C company_location	A _C company_size	1 ₂ 2023	1 ₂ 2022	1 ₂ 2020	1 ₂ 2021
			Valid 100% Error 0% Empty 0%	Valid 40% Error 0% Empty 60%	Valid 46% Error 0% Empty 54%	Valid 6% Error 0% Empty 94%
1	CZ	M		null	30469	null
2	US	L		null	null	100000
3	US	S		null	50000	null
4	IR	M		null	100000	null
5	AS	M		null	null	null
6	CR	S		null	50000	null
7	MK	S		null	6304	null
8	DE	M		64385	null	null
9	SE	M		80000	null	null
10	BA	S		120000	null	null
11	DE	L		214618	null	null
12	IN	L		null	300000	null
13	AU	M		null	40000	null
14	AU	L		70000	null	null
15	ES	M		null	31520	null
16	US	M		null	50000	null
17	CA	L		null	200000	null
18	DK	S		null	null	45896
19	AS	S		null	null	null
20	US	L		null	130000	null
21						

Gambar 4.18 Hasil Pivot Tabel

4.2 Visualisasi Dashboard

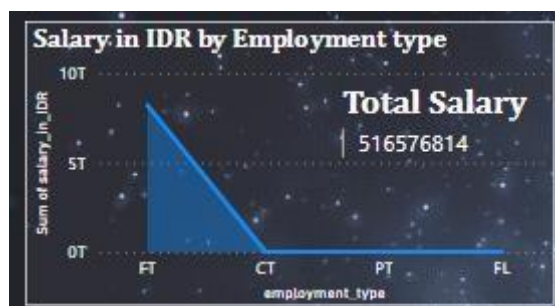
Setelah data diolah menggunakan *Microsoft Power BI* maka dihasilkan *dashboard* yang berisikan informasi dari data-data yang digunakan dalam bentuk visual. Bentuk tampilan

dashboard yang dibangun pada penelitian ini terdiri dari 3 halaman. Bentuk visualisasi dashboard yang terbentuk dalam penelitian ini ditunjukkan dalam Gambar 4.19.



Gambar 4.19 Tampilan *Dashboard*

Dashboard yang telah dibuat tersebut digunakan untuk menampilkan beberapa visualisasi yang berisi informasi, yaitu data aktual gaji (2020-2023) pemilihan *experience level* dan *work year*, *employment type*, *company location*, *job title*, *employee residence* dan *company size* yang dapat dimaksimalkan untuk menampilkan informasi yang lebih detail dan interaktif. Grafik pertama menunjukkan total *salary* berdasarkan *employment type* yang ditampilkan dalam bentuk grafik garis. *Employment type* tersebut adalah FT (*Full Time*), FL (*Freelance*), CT (*Contract*) dan PT (*Part Time*). Selain itu juga ditampilkan juga total *salary* yang ada seperti pada Gambar 4.20.



Gambar 4.20 Grafik Total *Salary* Berdasarkan *Employment type*

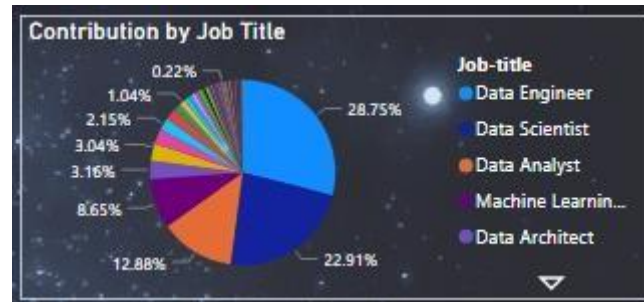
Grafik selanjutnya menggambarkan total *salary* berdasarkan *company location* yang digambarkan dalam bentuk grafik batang yang disusun secara horizontal. Informasi ini dapat digunakan sebagai dasar untuk mengambil sebuah strategi penetapan *salary* atau alokasi

sumber daya di berbagai lokasi. Dengan demikian, para pelaku bisnis atau analis dapat mengambil kebijakan yang lebih tepat dan akurat berdasarkan pemahaman yang didapat dari hasil visualisasi data. Dapat dilihat, untuk grafik yang menunjukkan rata-rata *salary* di area atau lokasi US masih lebih tinggi dibandingkan area lain. sebagian besar kantor pusat perusahaan dan tempat tinggal karyawan yang terwakili dalam kumpulan data adalah di US. Hal ini bisa jadi disebabkan oleh perbedaan geografis dan tingkat pendapatan, faktor spesifik wilayah seperti regulasi yang mengaturnya, atau karena lokasi US sendiri memiliki jumlah perusahaan yang tinggi dibandingkan dengan wilayah lain seperti pada Gambar 4.21.



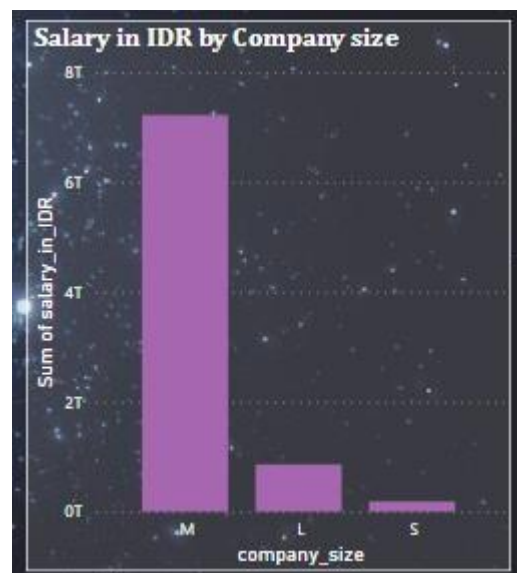
Gambar 4.21 Grafik Total *Salary* Berdasarkan *Company location*

Diagram *pie* adalah alat visual yang efektif untuk menganalisis perbedaan untuk setiap *job title* dengan *salary*. Warna yang berbeda pada setiap *pie* membuatnya lebih mudah dipahami, dan label yang jelas memberikan informasi yang mudah dicerna. Melalui analisis grafik, dimungkinkan untuk melihat pola atau tren perbedaan untuk setiap *job title*, dengan fokus pada *job title* yang menunjukkan perbedaan yang signifikan. Sebagai contoh, terlihat *job title* data *engineer* lebih mendominasi daripada *job title* lain seperti *data science*, *machine learning* dan lainnya. Hal ini dikarenakan data *engineer* merupakan pekerjaan yang paling banyak dalam kontribusi total *salary* yang dikeluarkan oleh perusahaan. Tingginya total *salary* pada data *engineer* dapat disebabkan oleh tingginya *salary* yang dikeluarkan oleh perusahaan maupun banyaknya pekerjaan tersebut. Grafik ini memberikan wawasan yang mendalam tentang sejauh mana model prediksi berhasil atau tidak untuk setiap *job title*, sehingga memudahkan identifikasi peningkatan atau penyempurnaan model seperti pada Gambar 4.22.



Gambar 4.22 Grafik Total *Salary* Berdasarkan *Job title*

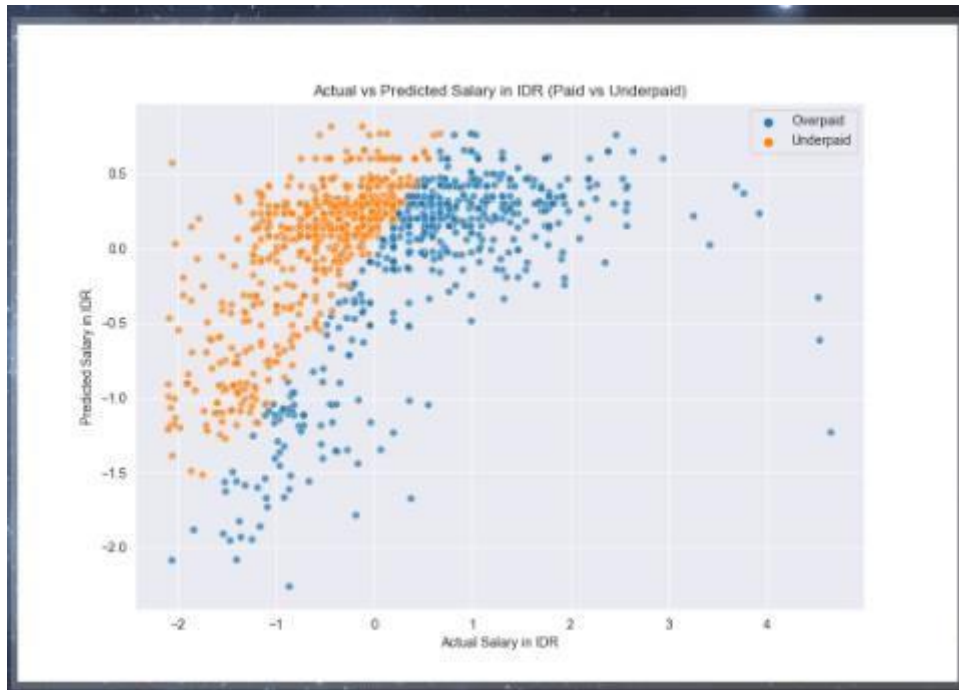
Grafik selanjutnya menggambarkan total *salary* berdasarkan *company size* yang digambarkan dalam bentuk grafik batang yang disusun secara vertikal. Informasi ini dapat digunakan untuk mengetahui pengaruh ukuran perusahaan terhadap besaran gaji yang dikeluarkan oleh perusahaan. Ukuran perusahaan merupakan gambaran untuk mengetahui besar kecilnya kekayaan yang dimiliki oleh suatu perusahaan. Ukuran perusahaan terbagi dari tiga golongan yang terdiri dari perusahaan *Large*, *Medium* dan *Small*. Kebanyakan perusahaan yang lebih besar yang memiliki total aset besar cenderung memiliki kemampuan untuk memberikan kompensasi kepada pihak eksekutifnya. Rata-rata gaji di perusahaan besar dan menengah telah meningkat secara signifikan dari tahun 2020-2023 sementara perusahaan kecil menunjukkan tingkat gaji rata-rata yang cukup stabil pada periode yang sama. Besaran total *salary* juga dipengaruhi seberapa banyak *company size* dan *salary* yang harus dikeluarkan oleh perusahaan.



Gambar 4.23 Grafik Total *Salary* Berdasarkan *Company size*

Dashboard prediksi *salary* menunjukkan perbandingan antara *salary* prediksi dengan *salary* aktual. Hasil visualisasi ini dapat memberikan informasi yang berharga bagi para

stakeholder, diantaranya para pemilik perusahaan serta dapat mengetahui variabel yang dapat mempengaruhi *salary*. Menganalisis tren ini dapat digunakan sebagai bahan pertimbangan dalam perencanaan bisnis dan juga menentukan strategi pemberian gaji bagi pemilik bisnis selanjutnya.



Gambar 4.24 Komparasi Total *Salary* Prediksi dengan Total *Salary* Aktual

4.3 Regresi Linear

Untuk pemrosesan analisis regresi linear dengan menggunakan bahasa pemrograman Python, digunakan beragam *library* seperti Numpy dan Pandas sebagai alat analisis data, serta Matplotlib dan Seaborn sebagai visualisasi data. Scikit-Learn dipilih untuk melakukan *machine learning*. Perintah untuk memanggil *library* yang dipakai ditunjukkan pada Gambar 4.25.

```

Python script editor
6 # Paste or type your script code here:
7 import numpy as np
8 import pandas as pd
9 import matplotlib.pyplot as plt
10 import seaborn as sns
11 sns.set_style('darkgrid')
12
13 import os
14 for dirname, _, filenames in os.walk('/kaggle/input'):
15     for filename in filenames:
16         print(os.path.join(dirname, filename))
17

```

Gambar 4.25 *Library* yang Digunakan

Proses pembuatan model diawali dengan mengunduh data di website Kaggle.com, suatu website yang menyediakan berbagai macam dataset dalam format file Microsoft Excel. File dataset yang sudah diunduh akan ekstrak terlebih dahulu kemudian dilakukan *preprocessing* data menggunakan python. Bahasa python digunakan karena memiliki karakteristik seperti memiliki distribusi Python yang kaya akan pustaka dan modul dan memiliki sintaks yang jelas dan mudah dipelajari. Selain melakukan preprocessing, juga dilakukan proses menghilangkan nilai *missing value* (NAN) serta mengecek tipe data yang digunakan sudah benar atau belum. Jika terdapat tipe data yang tidak sesuai maka harus diubah terlebih dahulu. Sejumlah langkah tersebut dilakukan bertujuan agar model mampu membaca data dengan kualitas tinggi dan juga memaksimalkan hasil dan akurasi yang diperoleh. *Output* data yang sudah melalui tahap *preprocessing* dan transformasi akan dilanjutkan ke tahap selanjutnya yaitu pelabelan data dan penanganan data *outlier*.

```

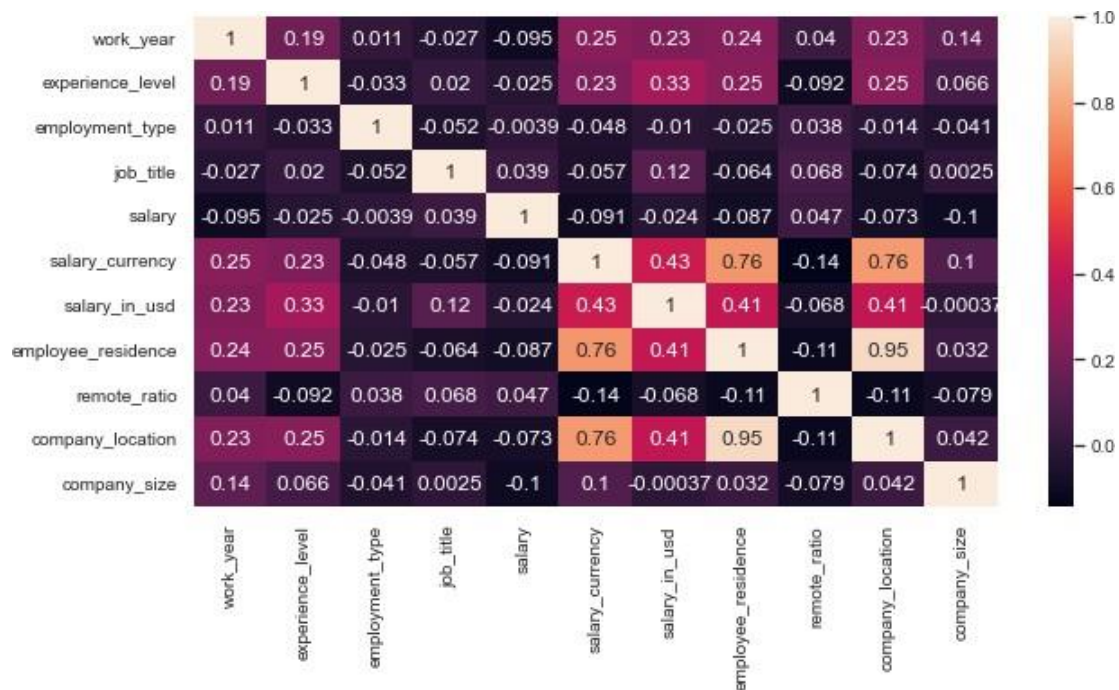
work_year          0
experience_level   0
employment_type    0
job_title          0
salary             0
salary_currency    0
salary_in_usd      0
employee_residence 0
remote_ratio       0
company_location   0
company_size       0
dtype: int64

```

Gambar 4.26 Pengecekan Nilai *Missing Value*

Identifikasi *outlier* dan pengamatan berpengaruh pada model regresi didasarkan pada asumsi bahwa model regresi yang diperoleh sudah tepat. Hal ini berarti model regresi yang telah dipilih telah cukup menggambarkan hubungan antara variabel respon dan variabel independen. Jika model regresi telah ditentukan, sebagian besar data seharusnya mendekati garis regresi atau *hyperplane*. Titik-titik data yang berada jauh dari garis regresi atau *hyperplane* mungkin bukan titik-titik data ideal bagi model yang dipilih dan dapat diidentifikasi sebagai *outlier*. Dalam konteks ini, matriks korelasi digunakan untuk mengevaluasi sejauh mana variabel-variabel lain berkaitan dengan variabel dependen. Korelasi yang tinggi dapat menunjukkan hubungan yang kuat, meskipun tidak selalu berarti sebab akibat. Hasil ini menjadi dasar kuat untuk pengambilan keputusan dalam analisis

regresi, menjamin bahwa model yang dikembangkan dapat memberikan hasil prediksi yang andal.



Gambar 4.27 Matriks Korelasi

Setelah proses matrik korelasi selesai maka dilanjutkan dengan proses analisis regresi linear. Modul yang digunakan untuk melakukan regresi linear yaitu Sklearn di Python seperti pada Gambar 4.25.

```

107 from sklearn.linear_model import LinearRegression
108
109 model = LinearRegression()
110
111 #Fitting the model to trainig data
112 model.fit(X_train, y_train)
113
114 from sklearn.metrics import mean_squared_error, r2_score
115
116 # Predicting test data using model
117 y_pred = model.predict(X_test)
118

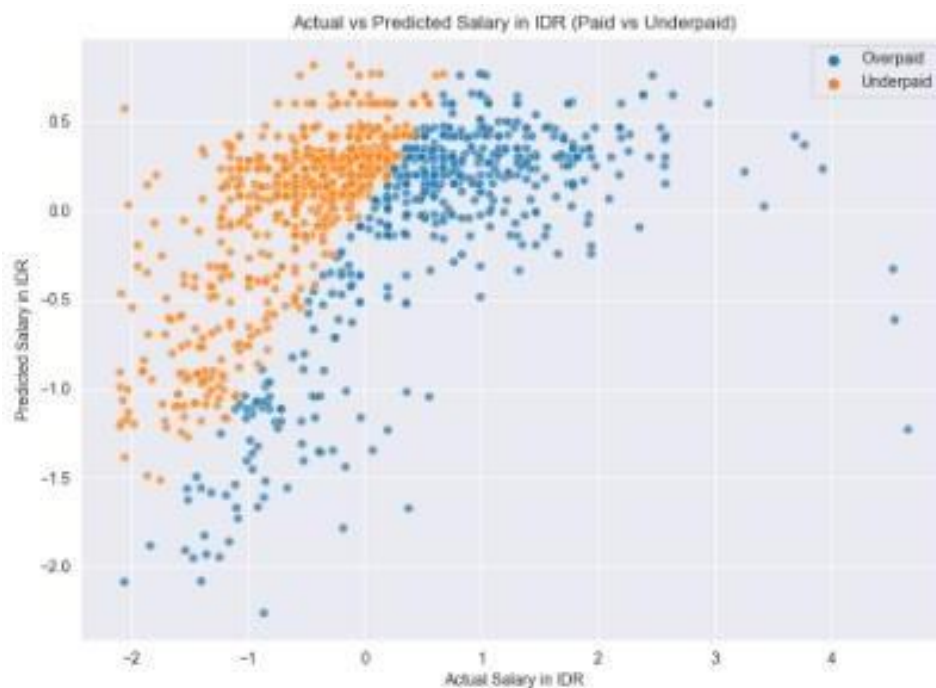
```

Gambar 4.28 Potongan Script Library Linear Regresi di Python

Pengujian dalam penelitian ini adalah dengan melakukan perbandingan pengaruh variabel bebas terhadap variabel terikat dengan mengacu korelasi antara tiap-tiap variabel bebas (x) dan variabel terikat (y). Pengujian dengan melakukan pengukuran tingkat ketepatan pada model multiple linear regression. Untuk melakukan pengujian data secara menyeluruh,

metode perhitungan akurasi sendiri akan diberlakukan. Penerapan metode melakukan perhitungan nilai total jumlah prediksi yang benar pada data memanfaatkan linear regression.

Variabel yang digunakan dalam penelitian ini adalah *work year*, *experience level*, *employment type*, *job title*, *salary currency*, *employee residence*, *remote ratio*, *company location* dan *company size* sebagai variabel penentu. Oleh karena itu, model regresi linier pada penelitian ini bisa dinyatakan sebagai persamaan 1:1 dimana y merupakan variabel respon, yaitu salary in USD. Adapun b_0 sebagai intercept, b_1 sebagai slope dari variabel x_1 yaitu *work year*, b_2 sebagai slope dari variabel x_2 yaitu *experience level*, b_3 sebagai slope dari variabel x_3 yaitu *experience type*, b_4 sebagai slope dari variabel x_4 yaitu *job title*, b_5 sebagai slope dari variabel x_5 yaitu *employee residence*, dan b_6 sebagai slope dari variabel x_6 yaitu Remote ratio, b_7 sebagai slope dari variabel x_7 yaitu *Company location* dan b_8 sebagai slope dari variabel x_8 yaitu *Company size*. Nilai b_0 (intercept), $b_1 - b_8$ (slope) dihasilkan melalui perhitungan kuadrat terkecil dengan memanggil fungsi model intercept dari library regresi linier dari modul Sklearn di Python. Seperti yang tampak pada nilai *intercept* dan nilai *slope* pada Gambar 4.29.



Gambar 4.29 Grafik Salary in USD Prediksi dan Salary in USD sebenarnya

Gambar 4.30 menampilkan grafik komparasi dari dataset yang diprediksi dan actual serta nilai *overpaid* dan *underpaid*. Grafik yang berwarna orange menunjukkan nilai *underpaid* dan grafik yang berwarna biru menunjukkan nilai *overpaid*. Nilai *underpaid* didapatkan dengan cara mencari selisih antara nilai aktual dan nilai prediksi. Jika selisih tersebut bernilai positif maka akan dianggap *overpaid*, jika selisih 0 maka dianggap *paid correctly* dan jika selisih bernilai negatif maka akan dianggap sebagai *underpaid* seperti ditunjukkan pada Gambar 4.30.

	Actual Salary	Predicted Salary	Difference	Status
2148	0.482648	0.060523	0.422125	Overpaid
1044	0.672585	0.281300	0.391285	Overpaid
3321	0.101981	0.315454	-0.213473	Underpaid
439	1.342321	0.248179	1.094142	Overpaid
3519	1.466038	-0.193616	1.659654	Overpaid
...
3666	-1.823385	-0.950607	-0.872778	Underpaid
1151	0.990204	0.299566	0.690638	Overpaid
942	0.416031	0.599840	-0.183809	Underpaid
3607	-0.357992	-0.529847	0.171856	Overpaid
1237	0.254248	0.232838	0.021410	Overpaid

Gambar 4.30 Perhitungan Nilai *Overpaid* dan *Underpaid*

Dari gambar ini, terlihat bahwa nilai prediksi yang dihasilkan oleh model kurang lebih sama dengan nilai aktual, tetapi nilai prediksi yang dihasilkan bisa saja sama atau lebih rendah dari nilai aktual. Keakuratan algoritma ini memberikan model yang lebih baik dengan akurasi yang lebih tinggi, dan nilai prediksi yang lebih mendekati nilai aktual. Setelah model regresi linear didapatkan maka akan lebih mudah dalam melakukan estimasi gaji. Langkah terakhir dalam model regresi linear adalah mengevaluasi model yang didapatkan. Tahap ini menentukan keakuratan dari model. Ukuran-ukuran yang digunakan untuk mengevaluasi hasil prediksi suatu model adalah ukuran R squared, *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Square Error* (RMSE).

Tabel 4.1 Pengukuran Metriks Kesalahan

MAE	MSE	RMSE	R-squared
0.648600	0.734666	0.857127	0.263429

Dari Tabel 4.1, didapatkan nilai MAE sebesar 0,648600. Nilai MAE menghitung rata-rata selisih absolut antara nilai aktual dengan nilai prediksi. Semakin rendah nilai MAE, semakin tinggi kemampuan model untuk memprediksi nilai. Selanjutnya, terdapat nilai MSE sebesar 0,734666. Nilai MSE ini mengukur rata-rata kesalahan kuadrat antara nilai aktual dan nilai prediksi. Nilai MSE yang rendah atau mendekati nol menunjukkan bahwa hasil prediksi

sesuai dengan data aktual. Kemudian didapatkan nilai RMSE sebesar 0,857127. Nilai RMSE berguna untuk merepresentasikan tingkat kesalahan pada data model yang digunakan. Semakin rendah nilai RMSE, maka semakin tinggi tingkat akurasi sistem. Terakhir, diketahui bahwa nilai R-squared yang dihasilkan oleh model adalah 0,263429. Semakin dekat nilai R- *squared* dengan 1, semakin akurat model yang dihasilkan. Peningkatan model juga dapat dilakukan dengan mengimplementasikan metode *machine learning* lainnya untuk mendapatkan akurasi model yang lebih akurat.

4.4 Pengujian

Teknik yang digunakan pada tahapan pengujian ini adalah *System Usability Scale* (SUS) untuk mengukur tingkat kepuasan pengguna terhadap *dashboard* yang sudah dibuat. Pengujian dilakukan dengan cara meminta responden untuk menggunakan *dashboard* yang sudah dibuat dan kemudian mengisi kuisioner yang sudah disediakan. Pengujian ini terdiri dari 2 responden. Jumlah pernyataan dalam kuisioner ini adalah 10 (Q1, Q2 sampai Q10). Hasil pengujian ditunjukkan pada Tabel 4.2.

Tabel 4.2 Hasil Pengujian SUS

Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Responden 1	5	2	4	2	4	1	5	3	4	2
Responden 2	4	2	4	1	4	2	5	2	4	3

Perhitungan SUS dilakukan dengan cara untuk pertanyaan bernomor ganjil seperti nomor 1,3,5,7, dan 9 kontribusi skor adalah skala posisi dikurangi 1. Untuk pertanyaan bernomor genap seperti nomor 2,4,6,8 dan 10, kontribusi adalah 5 minus posisi skala. Kalikan jumlah nilai sebesar 2,5 untuk mendapatkan nilai keseluruhan sehingga didapatkan hasil seperti Tabel 4.3.

Tabel 4.3 Hasil Akhir Perhitungan Pengujian SUS

No	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Jumlah	Nilai
Responden 1	4	3	3	3	3	4	4	2	3	3	32	80
Responden 2	3	3	3	4	3	3	4	3	3	2	31	77,5
												78,75

Berdasarkan perhitungan menggunakan pengujian SUS didapatkan hasil akhir sebesar 78,75 sehingga *dashboard* yang dibuat dapat dianggap baik sekali (*excellent*).

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil penelitian ini, dapat disimpulkan bahwa:

- a. Metode regresi linier memiliki kemampuan untuk memprediksi *salary*. R-squared yang dihasilkan oleh model adalah 0,263429. Semakin dekat nilai R-squared dengan 1, semakin akurat model yang dihasilkan. Namun demikian, masih terdapat ruang untuk meningkatkan akurasi dengan menambah jumlah data *training* yang digunakan, mengganti parameter yang dipakai, atau dengan menerapkan arsitektur yang lebih kompleks. Penelitian ini diharapkan dapat memberikan informasi yang berharga bagi perusahaan untuk mengambil keputusan yang tepat.
- b. Penelitian telah berhasil membuat data *warehousing* berdasarkan rancangan dimensional *data modeling* yang diusulkan.

5.2 Saran

Berdasarkan hasil analisis dan pembahasan, beberapa saran dapat diajukan yaitu:

- a. Peningkatan Model: Disarankan untuk mengeksplorasi dan mengimplementasikan metode *machine learning* lainnya, seperti regresi polinomial, *random forest*, atau *neural networks*, untuk membandingkan kinerja dan meningkatkan akurasi prediksi gaji.
- b. Penambahan Data: Mengumpulkan lebih banyak data dan memastikan data yang digunakan berkualitas tinggi dapat meningkatkan model prediksi. Data tambahan seperti informasi demografi, kinerja karyawan, dan kondisi pasar kerja juga dapat dipertimbangkan.

DAFTAR PUSTAKA

- Akbar, R., Darman, R., Marizka, Namora, J., & Ardewati, N. (2018). Implementasi Business Intelligence Menentukan Daerah Rawan Gempa Bumi di Indonesia dengan Fitur Geolokasi. *Jurnal Edukasi & Penelitian Informatika*, 30-35.
- Akbar, R., Soniawan, A., Dinur, R., Adrian, J., Azim, R., & Zikri, A. (2017). Implementasi Business Intelligence untuk Menganalisis Data Persalinan Anak di Klinik Ani Padang dengan Menggunakan Aplikasi Tableau Public. *JOIN*, 20-24.
- Anjar, A., Ritonga, M. K., Toni, & Rahmayani. (2021). Dampak Positif dan Negatif Perkembangan Teknologi Komunikasi Terhadap Mahasiswa PPKn FKIP Labuhanbatu. *CIVITAS (JURNAL PEMBELAJARAN DAN ILMU CIVIC)*, 41-44.
- Asnawi, A. (2022). Kesiapan Indonesia Membangun Ekonomi Digital di Era Revolusi Industri 4.0. Syntax Literate. *Jurnal Ilmiah Indonesia*, 398–413.
- Batta, M. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research*, 381-399.
- Bunawan, W., Setiawan, A., Rusli, A., & Nahadi. (2015). Penilaian Pemahaman Representasi Grafik Materi Optika Geometri Menggunakan Tes Diagnostik. *Cakrawala Pendidikan*, 257-267.
- Caserta, R. K. (2004). *The Data Warehouse ETL Toolkit*. New Jersey: In Wiley Publishing.
- Daniswara, D. M., Triwulandari, & Habyba, A. N. (2022). Perancangan Gaji Karyawan Menggunakan Metode Point System Di PT. Mitra Karya Tehnik Indonesia. *JIEOM (Journal of Industrial Engineering and Operation Management)*, 157-166.
- Darman, R. (2018). Analisis Data Tanaman Padi di Indonesia Analisis Data Tanaman Padi. *JURNAL INOVTEK POLBENG- SERI INFORMATIKA*, 155-163.
- Darman, R. (2018). Analisis Visualisasi dan Pemetaan Data Tanaman Padi di Indonesia Menggunakan Microsoft Power BI. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 156-162.
- Das, S., Barik, R., & Mukherjee, A. (2020). Salary Prediction Using Regression Techniques. *SSRN Scholarly Paper*, 1-5.
- Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM*, 64-73.
- Edhya, B. F., & Susilowati, M. (2022). Business Intelligence Data Marketing Menggunakan Metode Kimball dan ETL Dengan Power BI. *Jurnal Teknologi, Informasi dan Industri*, 87-97.

- Firmansyah, D., Saepuloh, D., & Dede. (2022). Daya Saing: Literasi Digital dan Transformasi Digital. *Journal of Finance and Business Digital (JFBD)*, 237-250.
- Hariyanti, E., Werdiningsih, I., & Surendro, K. (2011). Model Pengembangan Dashboard Untuk Monitoring dan Evaluasi Kinerja Perguruan Tinggi. *Jurnal Ilmiah Teknologi Informasi*, 13-20.
- Hasan, F. N. (2019). Implementasi Sistem Business Intelligence Untuk Data Penelitian di Perguruan Tinggi. *Prosiding Seminar Nasional Teknoka*, 1-10.
- Hasudungan, L. (2017). Pengaruh Faktor Pendidikan, Umur, dan Pengalaman Kerja Terhadap Kinerja Aparatur Sipil Negara (ASN) Pada Dinas Pekerjaan Umum Penata Ruang, Perumahan dan Kawasan Permukiman Kabupaten Kapuas Kalimantan Tengah. *Jurnal Ilmiah Ekonomi Bisnis*, 301 - 310.
- Hidayat, M. (2022). *Implementasi Machine Learning Prediksi Gaji Pekerjaan Perusahaan Rintisan Menggunakan Algoritma Random Forest Dan Framework Flask (Studi Kasus : Pekerjaan Pada Website Techinasia.Com)*. Yogyakarta: Universitas Islam Indonesia.
- Hilmansyah, L., & Maesaroh, S. (2022). Implementasi Datawarehouse Dalam Menentukan Tabel Fakta Melalui Proses Etl Di Alwy Minimart. *Jurnal Ilmiah Sains, Teknologi Dan Rekayasa*, 1-10.
- Hope, T. M. (2020). Chapter 4—Linear regression. In A. Mechelli & S. Vieira (Eds.), *Machine Learning*. Academic Press, 67-81.
- Husni, Z. N., & Mukhlash, I. (2014). Implementasi Business Intelligence Pada Manajemen Report Bank XYZ. *Jurnal Sains Dan Seni ITS*, 16-21.
- Kamasi, N. V., & Nangoy, S. M. (2023). Pengembangan Business Intelligence untuk Pendukung Pengambilan Keputusan dalam Mengurangi Kemacetan di Kota Tomohon. *Jurnal Informatika Dan Teknologi Informasi*, 15-21.
- Khan, S. P., Wahyudin, Ayuningtyas, S. M., Rohmah, W., Vindari, Z. I., & Azzahra, A. G. (2023). Analisa Perbandingan Nilai Akurasi Exponential Smoothing dan Linier Regresion pada Peramalan Permintaan Part Joint Brake Rod KTM. *Jurnal Serambi Engineering*, 15-21.
- Madyatmadja, E. D., Nuramalia, A. N., Kusumawati, L., Jamil, S. P., & Kusumawardhana, W. (2021). Data Visualization Of Internet Usage In The JABODETABEK Area. *INFOTECH: JOURNAL OF TECHNOLOGY INFORMATION*, 55-62.

- Maesaroh, S., Lubis, R. R., Husna, L. N., Widyaningsih, R., Susilawati, R., & Yasmin, P. M. (2022). Efektivitas Implementasi Manajemen Business Intelligence pada Industri 4.0. *ADI Bisnis Digital Interdisiplin Jurnal*, 1-9.
- Mahaputra, M. R. (2022). Effect of Salary and Work Environment on Productivity (Study of Human Resource Management Literature). *International Journal of Advanced Multidisciplinary*, 153-162.
- Manik, V. (2020). *Evaluasi Usability pada Aplikasi Mobile ACC.ONE menggunakan System Usability Scale (SUS) dan Usability Testing*. Yogyakarta: Universitas Atma Jaya.
- Maulud, D. H., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends* , 140-147.
- Nabilla, N. U., & Hidayat, S. (2021). Pengembangan Business Intelligence Pada Sistem Informasi Distributor. *AUTOMATA*, 1-5.
- Nasution, M. K., Sitompul, O. S., & Nababan, E. B. (2020). Data Science. *Journal of Physics: Conference Series*, 1-11.
- Nurfarisi, R. (2022). Visualisasi Data Covid-19 Klinik MariSehat Menggunakan Microsoft Power BI. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 146-149.
- Oslan, Y., & Kristanto, H. (2019). Proses ETL (Extract Transformation Loading) Data Warehouse Untuk Peningkatan Kinerja Biodata Dalam Menyajikan Profil Mahasiswa Dari Dimensi Asal Sekolah Studi Kasus: Biodata Mahasiswa UKDW. *Research Fair Unisri* , 528-535.
- Parhusip, H. A. (2020). *Pemrograman Python Untuk Penanganan Big Data*. Yogyakarta: Andi Offset.
- Prabowo, D., & Sari, B. W. (2019). Fuzzy Tsukamoto dan Mamdani Untuk Penentuan Bonus Gaji Pegawai PT. Indonesia IT. *INFOS Journal*, 25-31.
- Priyatna, A. (2019). Implementasi Sistem Penunjang Keputusan Menggunakan Business Intelligence Untuk UMKM Di Gunung Putri Kab. Bogor. *Jurnal Khatulistiwa Informatika*, 7-12.
- Putra, I. P., Nirwana, N. A., Aristana, I. G., Prayana, I. D., Pratiwi, N. A., & Desmayani, N. M. (2023). Pelatihan Power BI: Meningkatkan Kinerja Bisnis dengan Analisis Data dan Visualisasi yang Optimal. *Jurnal Widya Laksmi: Jurnal*, 77-80.

- Putri, W. L. (2022). *Aplikasi Model Prediksi Gaji Profesi Data Berbasis Website Menggunakan Flask Dan Metode Decision Tree*. Yogyakarta: Universitas Islam Indonesia.
- Rahmadani, D. A. (2023). Pemanfaatan Kemajuan Teknologi Informasi Terhadap Perkembangan Akuntansi. *Jurnal Ilmu Data*, 1-16.
- Rahmawati, E. (2022). Implementasi Power BI untuk Dashboard Analisis Hasil Rating dan Review Produk Elektronik di Lazada.com. *Jurnal SPIRIT*, 6-15.
- Rebala, G., Ravi, A., & Churiwala, S. (2019). *Machine Learning Definition and Basics*. Springer, 1-17.
- Riyanda, M. D., & Suyanto. (2020). Implementasi Business Intelligence Pada Analisis Perkembangan Hasil Pertanian Provinsi Sumatera Selatan. *Journal of Computer and*, 174-184.
- Saputra, M. A., Prasetyo, N., Zulfikar, I., Rijanandi, T., & Adhinata, F. D. (2022). Prediksi Gaji Berdasarkan Pengalaman Bekerja Menggunakan Metode Regresi Linear. *Journal of Dinda (Data Science, Information Technology, and Data Analytics)*, 58-63.
- Satrio, P., & Suryani, E. (2017). Penerapan Model Sistem Dinamik Untuk Melakukan Pemeliharaan Operasional Aset Unit Transmisi Dan Visualisasi Luaran Model Dengan Menggunakan Dashboard (Studi Kasus: PT.Pln (Persero) App Semarang). *Jurnal Teknik ITS*, 327-332.
- Setiawan, D. (2018). Dampak Perkembangan Teknologi Informasi dan Komunikasi Terhadap Budaya. *JURNAL SIMBOLIKA: Research and Learning in Communication Study*, 62-72.
- Sidik, A. (2018). Penggunaan System Usability Scale (SUS) Sebagai Evaluasi Website Berita Mobile. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 83-90.
- Silvana, M., Akbar, R., & Derisma. (2017). Pengembangan Model Business Intelligence Manajemen Rumah Sakit untuk Peningkatan Mutu Pelayanan. *JEPIN*, 124-133.
- Siska, & Putri, D. S. (2021). Implementasi Business Intelligence untuk Menganalisis perbandingan data kasus COVID-19 di Jawa Barat sebelum PSBB dan setelah PSBB. *Jurnal Ilmiah Edutic*, 94-104.
- Sulistyorini, P. (2010). Business Intelligence Dan Manfaatnya Bagi Organisasi. *Jurnal Ilmiah IC Tech*, 74-81.

- Sundari, S., & Urip, C. R. (2021). Kapabilitas Membangun Jaringan dengan Pemasok untuk Meningkatkan Kinerja Operasional Pada Toko Aksesoris Telepon Genggam Di Kabupaten Banyumas. *Eksis: Jurnal Ilmiah Ekonomi dan Bisnis*, 84-95.
- Susanto, H. (2014). Data Mining untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu. *Jurnal Pendidikan Vokasi*, 222-231.
- Syaripul, N. A., & Bachtiar, A. M. (2016). Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi DKI Jakarta: Topik Ekonomi dan Keuangan Daerah. *Jurnal Sistem Informasi*, 82-89.
- Tuah, Y. A., & Anyan. (2020). Implementasi Model Regresi Linear Sederhana Untuk Prediksi Gaji Berdasarkan Pengalaman Lama Bekerja. *Jurnal Education and Technology*, 56-70.
- Wahyudi, H. S., & Sukmasari, M. P. (2014). Teknologi dan Kehidupan Masyarakat. *Jurnal Analisa Sosiologi*, 13-24.
- Zikra, A. A. (2022). *Implementasi Business Intelligence pada ACC Absensi Menggunakan Aplikasi Power BI*. Yogyakarta: Universitas Islam Indonesia.
- Zikra, A. A. (2023). *Perancangan Dashboard Accabsensi Dan Opcent Menggunakan Power BI Di Astra Credit Companies*. Yogyakarta: Universitas Islam Indonesia.

LAMPIRAN