Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May

OPEN ACCESS

https://doi.org/10.35870/jimik.v5i2.657

Implementasi Intelejen Bisnis dengan Visualisasi Data Gaji dan Algoritma Linear Regresion

Haryadi Tri Nugroho [1], Syarif Hidayat [2*]

[1,2*] Program Studi Sistem Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia.

Email: 20523097@students.uii.ac.id [1], syarif@uii.ac.id [2*]

Histori Artikel:
Dikirim 27 Januari 2024; *Diterima dalam bentuk revisi* 15 Februari 2024; *Diterima* 15 Maret 2024; *Diterbitkan* 10 Mei 2024.

Abstrak

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 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. Pada masa sekarang, konsep Business Intelligence dapat diterapkan oleh semua sektor industri, selama industri tersebut sistem database untuk mengembangkan bisnisnya. Business Intelligence memproses dan menganalisa data mentah dalam jumlah yang besar untuk kemudian ditampilkan dalam sebuah laporan bisnis dengan visual yang interaktif yaitu dashboard. Penelitian ini bertujuan untuk berkontribusi pada pemahaman prediksi rata-rata 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.

Abstract

To understand these employment dynamics, it is important to explore how long Data Scientists typically work in this field and what sectors they are more likely to be involved in. This information is very important for companies looking to recruit Data Scientists or determine the average salary of their employees. With this activity, managers can more clearly know the employee's salary range based on their work experience and area of specialization. This gives managers a more precise tool for assessing compensation appropriate to the level of experience and expertise possessed by employees. Nowadays, the concept of Business Intelligence can be applied by all industrial sectors, as long as the industry has a database system to develop its business. Business Intelligence processes and analyzes large amounts of raw data and then displays it in a business report with interactive visuals, namely a dashboard. This research aims to contribute to the understanding of average salary predictions and provide valuable information for business owners to make informed decisions. The results show that linear regression is accurate and forms a linear regression model with an R-square result of 0.263429. The salary prediction calculation process is carried out by considering influencing factors, and the results are then presented visually using Power BI to provide business owners with more interactive information.

Keyword: Salary; Power BI; Predictions; Visualization.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

1. Pendahuluan

Peningkatan perkembangan teknologi komunikasi yang pesat telah membawa perubahan mendasar dalam cara kita berinteraksi dan mengakses informasi (Wahyudi & Sukmasari, 2014). Pada dasarnya, teknologi ini seharusnya menjadi sarana yang mendukung perkembangan pemikiran manusia dan memperkuat hubungan sosial dan profesional. 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 (Anjar *et al.*, 2021). 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, kita tidak bisa mengabaikan potensi dampak negatifnya perubahan dalam dinamika sosial yang terkait dengan penggunaan media teknologi digital (Firmansyah *et al.*, 2022).

Transformasi digital 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 (Syamsu & Widodo, 2021). Kemajuan teknologi informasi dan komputasi telah menciptakan ledakan data yang kompleks, menciptakan tantangan dan peluang yang harus dihadapi (Dhar, 2013). Keberadaan Data Scientist saat ini sangat diperlukan untuk memecahkan berbagai permasalahan dengan tantangan besar untuk mendukung keberadaan industri 4.0 dengan sasaran di berbagai sektor industri (Kusuma & Hidayat, 2024). Keberadaan Data Science dan Data Scientist mampu mentransformasi data yang berkembangan saat ini akibat revolusi industri 4.0, sehingga diperlukan sebuah penanganan yang cepat dan dapat dikontrol oleh pihak industri (Syamsu & Widodo, 2021). Transformasi data di industri 4.0 bertujuan untuk memberikan sebuah informasi yang sangat penting dan digunakan oleh perusahaan, untuk menentukan strategi segmentasi pasar, pengembangan suatu produk, merancang keputusan bisnis, dan memberikan data yang sesuai kebutuhan informasi yang dimiliki oleh perusahaan dan bahkan ke arah pengelompokan konsumen, dan lain sebagainya (Syamsu & Widodo, 2021).

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 (Rahmadani, 2023). Dengan pekerjaan bidang Data Science telah menjadi salah satu yang paling dicari dan penting dalam era digital ini (Syamsu & Widodo, 2021). Data Scientists, dengan keterampilan analisis data dan keahlian dalam menggali wawasan dari data, menjadi elemen kunci dalam pengambilan keputusan bisnis yang efektif (Kusuma & Hidayat, 2024). 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).

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 (Hasudungan, 2017). 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 (Hasudungan, 2017). 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

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

dan kesuksesan perusahaan dalam menghadapi tantangan teknologi modern (Nasution *et al.*, 2020).

Dengan adanya kegiatan ini, manajer dapat dengan lebih jelas mengetahui kisaran gaji karyawan berdasarkan pengalaman kerja dan bidang spesialisasi mereka (Prabowo & Sari, 2019). Ini memberikan manajer alat yang lebih tepat untuk menilai kompensasi yang sesuai dengan tingkat pengalaman dan keahlian yang dimiliki oleh karyawan (Prabowo & Sari, 2019). 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 (Prabowo & Sari, 2019). Penerapan ini menampilkan tren pekerjaan berdasarkan gaji dan pengalaman, kita dapat mengidentifikasi pola yang berharga dalam dunia kerja (Mahaputra, 2022). Data ini membantu perusahaan untuk merencanakan kompensasi yang lebih adil, yang sesuai dengan tingkat pengalaman karyawan dan bidang spesialisasi mereka (Mahaputra, 2022). 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 (Mahaputra, 2022).

Penerapan analisis ini untuk mengidentifikasi trend dalam kompensasi berdasarkan tiga faktor utama: pekerjaan, lamanya jabatan, dan bidang yang ditekuni (Khan *et al.*, 2023). Analisis ini dapat memberikan manfaat signifikan kepada manajer dan pemangku kepentingan dalam pengambilan keputusan terkait gaji dan kompensasi (Khan *et al.*, 2023). 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 (Khan *et al.*, 2023).

Dalam analisis data, regresi linier adalah alat yang sangat berharga (Hope, 2020). Ini adalah metode statistik yang memungkinkan kita untuk memahami hubungan antara dua atau lebih variabel dalam suatu konteks (Hope, 2020). Dalam banyak kasus, regresi linier digunakan untuk mengukur dan menganalisis hubungan antara variabel dependen dan variabel independen (Hope, 2020). Regresi linier memungkinkan kita untuk mengeksplorasi sejauh mana perubahan dalam variabel independen mempengaruhi variabel dependen, dan dengan demikian, memungkinkan kita untuk membuat prediksi yang lebih baik dan informasi yang lebih berharga dalam berbagai disiplin ilmu (Hope, 2020).

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 (Hope, 2020). 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 (Hope, 2020). 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).

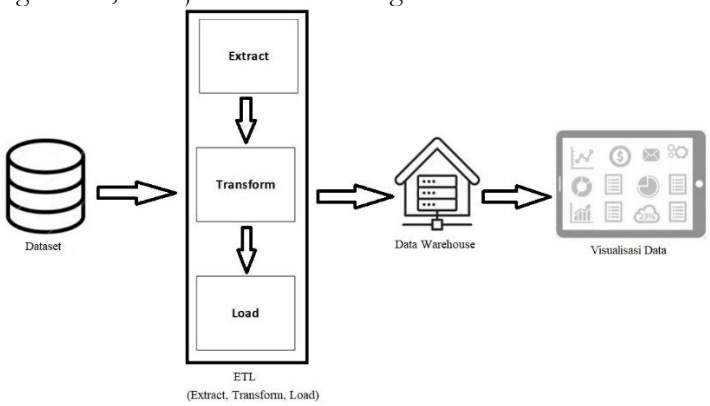
Setelah hasil analisis regresi linier diperoleh, penting untuk mengkomunikasikan temuan tersebut secara efektif (Darman, 2018). Dalam hal ini, aplikasi Microsoft Power BI menjadi alat yang sangat berguna (Darman, 2018). Dengan Power BI, para peneliti dapat mengimpor data hasil analisis regresi linier dan dengan mudah membuat visualisasi yang informatif (Darman, 2018). Visualisasi ini dapat berupa grafik garis atau grafik sebaran yang memperlihatkan hubungan antara variabel dependen dan independen, dengan garis regresi linier (Darman, 2018). Selanjutkan divisualisasikan menggunakan Power BI juga memungkinkan pembuatan laporan interaktif yang dapat diakses oleh berbagai pemangku kepentingan (Darman, 2018). 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 (Darman, 2018). Dengan demikian, Power BI memberikan solusi yang lengkap dalam mengintegrasikan analisis regresi linier dengan visualisasi data yang informatif dan pelaporan yang memudahkan berbagi hasil penelitian secara lebih dinamis (Darman, 2018). Pada analisis data dengan regresi linier dan visualisasi menggunakan Power BI adalah pendekatan yang kuat untuk memahami dinamika pekerjaan di era digital (Darman, 2018). Memahami hubungan antara variabel pekerjaan, faktor lama pekerjaan,

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

metode kerja, dan kompensasi adalah langkah penting dalam mengelola sumber daya manusia dan pengambilan keputusan (Darman, 2018).

2. Metode Penelitian

Diagram alir penelitian "Implementasi Intelejen Bisnis Dengan Visualisasi Data Gaji dan Algoritma Linear Regresion", ditunjukkan melalui diagram berikut ini.



Gambar 1. Metode Penelitian

2.1 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. Pemilihan penggunaan Kaggle dataset pada penelitian ini karena *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*.

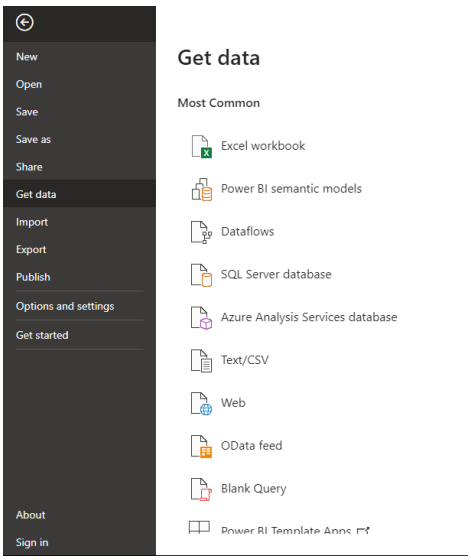
	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	compa
0	2023	SE	FT	Principal Data Scientist	80000	EUR	85847	ES	100	ES	
1	2023	MI	CT	ML Engineer	30000	USD	30000	US	100	US	
2	2023	MI	CT	ML Engineer	25500	USD	25500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

Gambar 2. Data Pelatihan

2.2 Extract

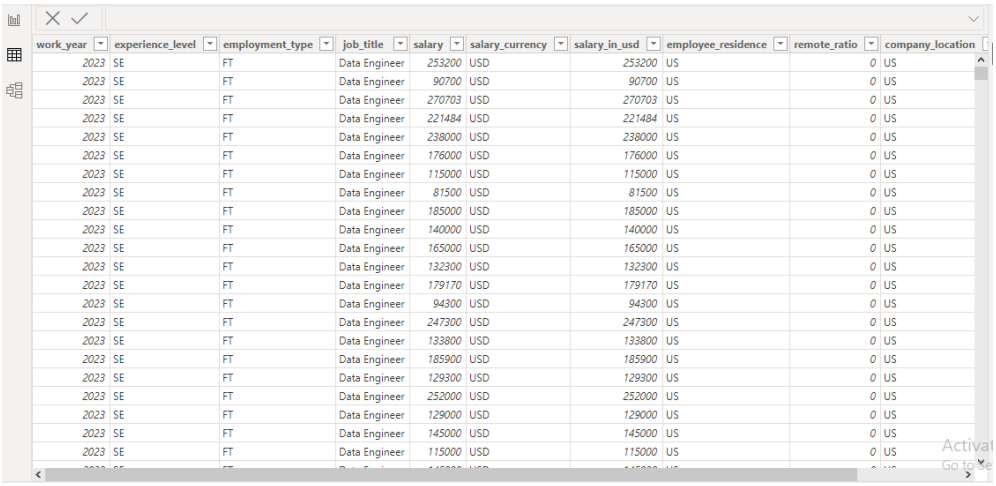
Proses ekstraksi data merupakan proses awal yang dilakukan pada saat melakukan pengembangan *dashboard*. Ekstraksi data merupakan proses yang dilakukan untuk mengambil data dari data *source* yang diinginkan, seperti *database, file, cloud* dan juga bisa diekstraksi menggunakan *script*. Power BI menyediakan fitur bernama *get* data, fitur ini dapat digunakan untuk melakukan berbagai macam ekstraksi data dengan berbagai cara seperti pada Gambar 3.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657



Gambar 3. 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 4.



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 4. Tampilan Tabel Data Yang Digunakan

2.3 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,

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

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. Algoritma regresi linear dipilih dalam penelitian ini karena kemudahan penggunaannya, dapat mengidentifikasi sekuat apa pengaruh yang diberikan oleh variabel independent terhadap variabel lainnya (dependen) serta dapat digunakan untuk memprediksi tren masa yang akan datang. 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.

2.4 Load

Load merupakan proses untuk memuat data yang telah dilakukan transformasi data pada proses sebelumnya ke dalam data *warehouse* Power BI. Data yang telah selesai dilakukan proses transformasi pada *power query*, selanjutnya dilakukan proses load ke data *warehouse* dengan cara melakukan proses *apply* pada *power query* dan data akan otomatis tersimpan pada data *warehouse* Power BI dan selanjutnya dapat dilakukan proses pembuatan dashboard visualisasi.

2.5 Data Warehouse

Data *warehouse* merupakan tempat penyimpanan data yang telah dilakukan pembersihan dandipetakan pada proses ETL. Pada penelitian ini gudang data yang dipakai bersumber dari Microsoft Excel. Data *warehouse* memiliki bentuk skema yang seperti bintang, atau biasa disebut Star Schema. Terdapat satu tabel fakta (fact table) yang menjadi pusat dari keseluruhan data dengan memiliki beberapa tabel dimensi (*dimension tables*) yang terhubung ke *fact table*. Data yang disimpan pada Microsoft excel yaitu data historis gaji tahun 2020-2023. Semua data tersebut digabungkan dalam satu tempat penyimpanan dengan format CSV dan dikelompokan sesuai kategori.

2.6 Visualisasi Data

Setelah melalui proses ETL (*Extract, Transform, Load*), langkah berikutnya yang krusial dalam manajemen data adalah melakukan 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.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

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. Hasil dan Pembahasan

3.1 Analisis Regresi Linear

Untuk pemrosesan analisis 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.

```
# Paste or type your script code here:
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

Gambar 5. Potongan *Script Library* yang Digunakan

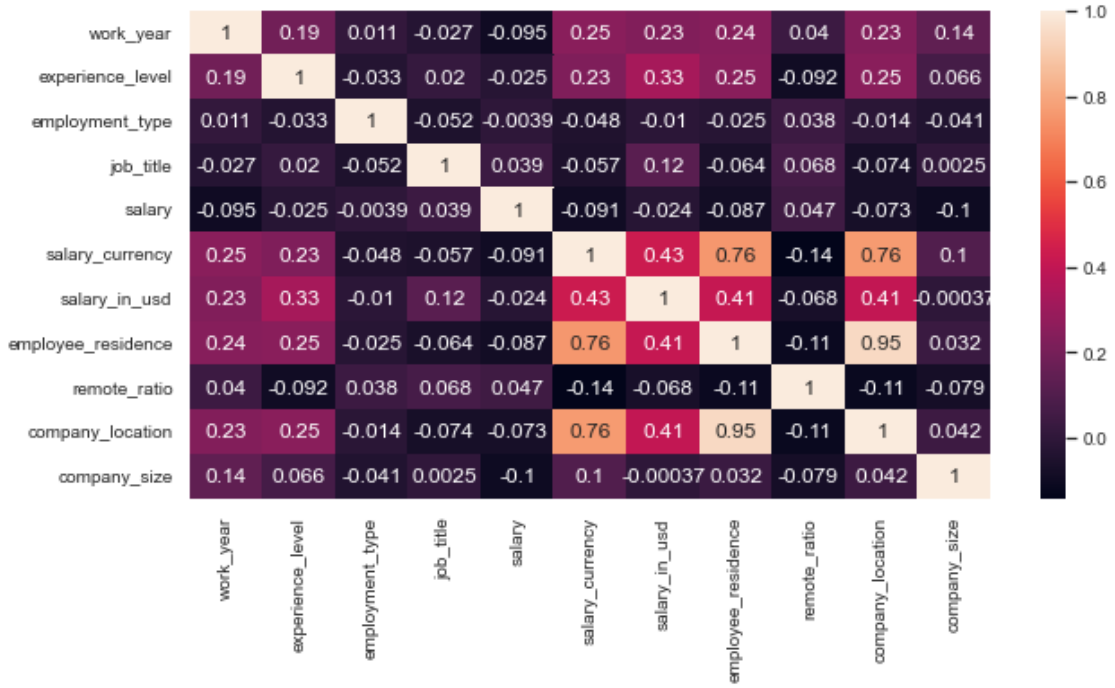
Proses pembuatan model diawali dengan mengunduh data di website Kaggle.com, suatu website yang menyediakan berbagai macam dataset dalam format file Microsoft Excel. File dataset yang sudah diunduh akan esktraksi terlebih dahulu kemudian dilakukan *preprocessing* data menggunakan python. Bahasa python digunakan karena memiliki karaketeristik 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 6. Pengecekan Nilai *Missing Value*

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

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 diidentikasikan 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 7. 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 8.

```
from sklearn.linear_model import LinearRegression

model = LinearRegression()

#Fitting the model to trainig data
model.fit(X_train, y_train)

from sklearn.metrics import mean_squared_error, r2_score

# Predicting test data using model
y_pred = model.predict(X_test)
```

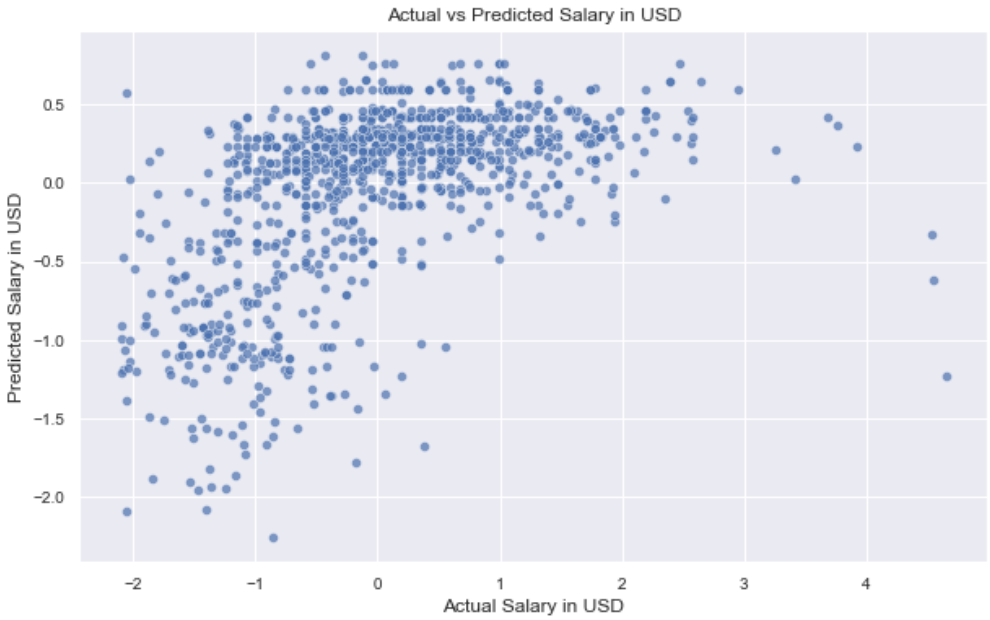
Gambar 8. Potongan *Script Library* Linear Regresi di Python

Pengujian dalam penelitian ini adalah dengan melakukan perbandingan pengaruh variabelibebas terhadap variabeliterikat dengan mengacu korelasiiantara tiap-tiap variabel bebas (x) dan variabeliterikat (y). Pengujianidengan melakukan pengukuran tingkat ketepatan pada model multiple

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

linear regresson. Pengujian ModeliLinear regression. Untuk melakukan pengujian dataisecara menyeluruh, metode perhitungan akurasi sendiri akan diberlakukan. Penerapan metode melakukan perhitungan nilai totalojumlah prediksi yang benar pada data memanfaatkan olinear 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 b0 sebagai intercept, b1 sebagai slope dari variabel x1 yaitu *work year*, b2 sebagai slope dari variabel x2 yaitu *experience level*, b3 sebagai slope dari variabel x3 yaitu *experience type*, b4 sebagai slope dari variabel x4 yaitu *job title*, b5 sebagai slope dari variabel x5 yaitu *employee residence*, dan b6 sebagai slope dari variabel x6 yaitu Remote ratio, b7 sebagai slope dari variabel x7 yaitu *Company location* dan b8 sebagai slope dari variabel x8 yaitu *Company size*. Nilai b0(intercept), b1 – b8 (slope) dihasilkan melalui perhitungan kuadrat terkecil dengan memanggil fungsi model intercept dari *library* regresi linier dari modul Sklearn di Python. Seperti yang tampak pada gambar 9 nilai intercept dan nilai slope pada Gambar 9.



Gambar 9. Grafik Salary in USD Prediksi dan Salary in USD sebenarnya

Gambar 9 menampilkan grafik komparasi dari dataset yang diprediksi dan aktual. 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 1. Pengukuran metriks kesalahan

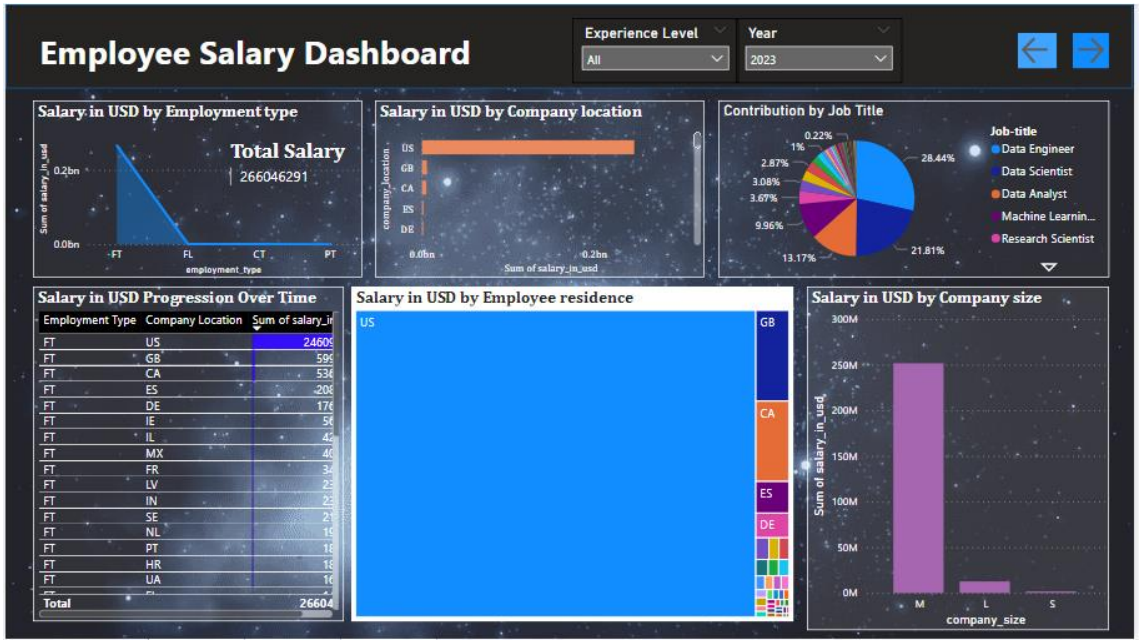
MAE	MSE	RMSE	R-squared
0.648600	0.734666	0.857127	0.263429

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

Dari Tabel 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. R-squared menggambarkan koefisien determinasi yang dapat menunjukkan sejauh mana kontribusi variabel bebas dalam model regresi mampu menjelaskan variasi dari variabel terikatnya. Uji koefisien determinasi (R-squared) dilakukan untuk menentukan dan memprediksi seberapa besar atau penting kontribusi pengaruh yang diberikan oleh variabel independen secara bersama – sama terhadap variabel dependen. Semakin dekat nilai R-squared dengan 1, semakin akurat model yang dihasilkan.

3.2 Visualisasi Dashboard

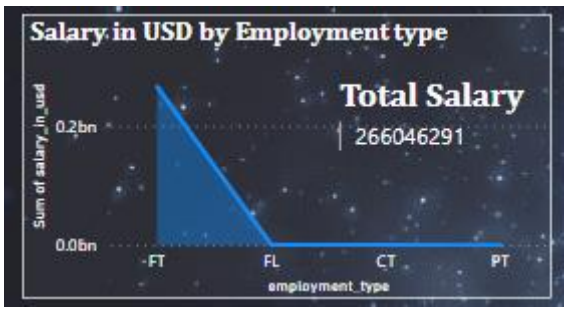
Bentuk visualisasi dashboard yang terbentuk dalam penelitian ini ditunjukkan dalam Gambar 10.



Gambar 10. 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*, *employement type, company location, job title, employee residence* dan *company size* yang dapat dimaksimalkan untuk menampilkan informasi yang lebih detail dan interaktif. Setiap grafik menunjukkan hubungan antara variabel dengan besarnya gaji (*salary in USD*). Terdapat 6 macam model visualisasi data yang berbeda untuk memberikan kemudahan dalam pembacaan data oleh pengguna. Terdapat pula menu filter *experience level* dan tahun jika pengguna ingin melihat data berdasarkan salah satu atau kedua variabel tersebut. 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 11.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657



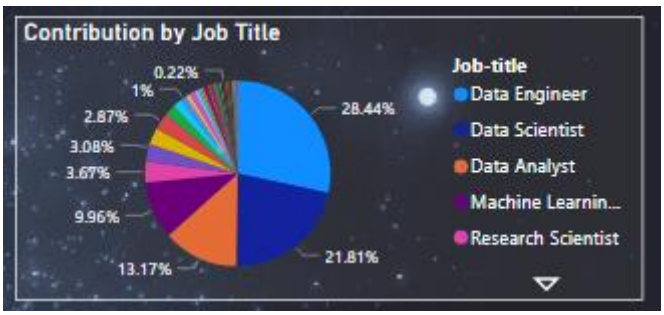
Gambar 11. Grafik Total *Salary* Berdasarkan *Employment type*

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 12.

Gambar 12. Grafik Total *Salary* Berdasarkan *Company location*

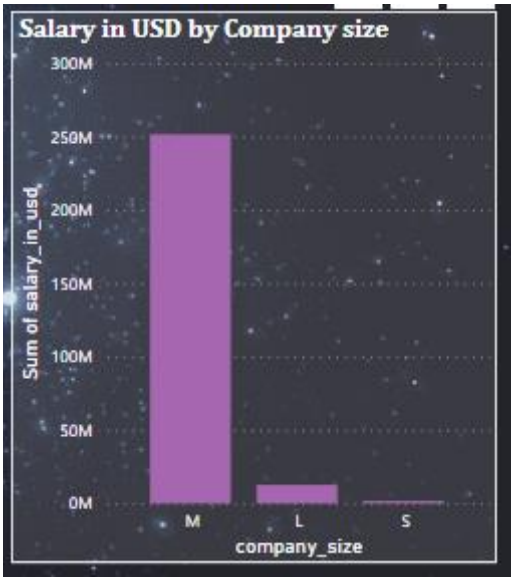
Diagram *pie* adalah alat visual yang efektif untuk menganalisis perbedaan untuk setiap *job title* dengan *salary*. Warna yang berbeda pada setiap pie membuatnya lebih mudah dipahami, dan label yang jelas memberikan informasi yang mudah dicerna. Melalui analisis grafik, dimungkinkan untuk melihat pola 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 laiinya. 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 13.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657



Gambar 13. Grafik Total *Salary* Berdasarkan *Job title*

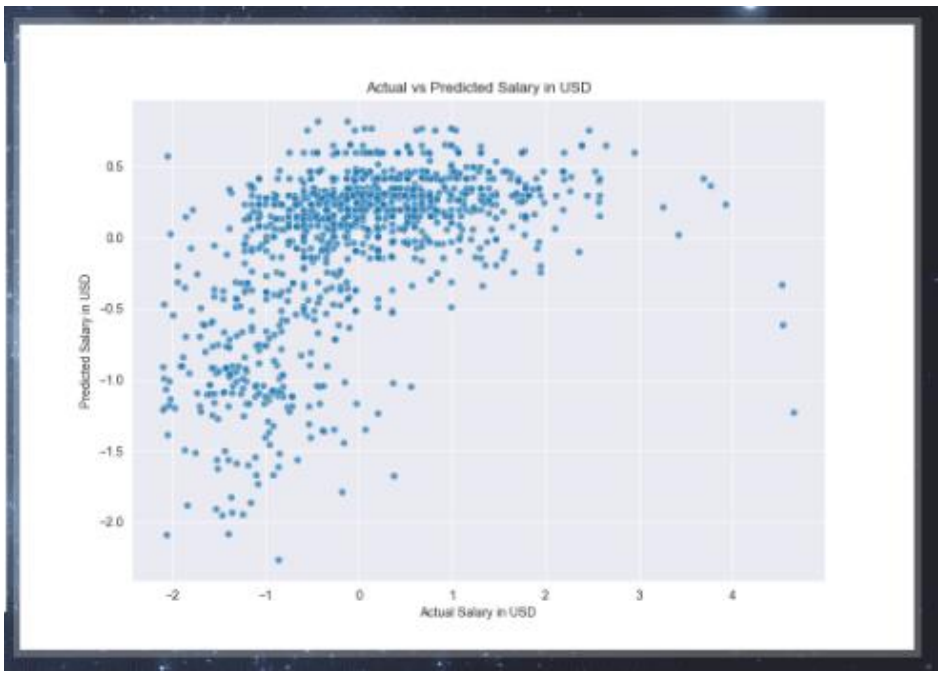
Grafik selanjutnya menggambarkan total *salary* berdasarkan *company size* yang digambarkan dalam bentuk grafik batang yang disusun secara vertikal. Informasi ini dapat digunakan untuk mengetahui pengaruh ukuran perusaahn terhadap besaran gaji yang dikeluarkan oleh perusahaan. Ukuran perusahaan merupakan 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 14. 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 mengethaui 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.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657



Gambar 15. Komparasi Total *Salary* Prediksi dengan Total *Salary* Aktual.

4. Kesimpulan

Dari hasil penelitian ini, dapat disimpulkan bahwa 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. Selama proses pengembangan visualisasi *dashboard* pada penelitian ini masih terdapat beberapa kekurangan sehingga pada penelitian selanjutnya diharapkan model regresi linear yang telah didapat sebelumnya dapat diperbarui dengan melakukan lagi pada proses *training* dengan menambahkan data *training* untuk meningkatkan akurasi dari model regresi linear yang telah didapatkan.

5. Daftar Pustaka

Anjar, A., Ritonga, M. K., & Toni, T. (2021). DAMPAK POSITIF DAN NEGATIF PERKEMBANGAN TEKNOLOGI KOMUNIKASI TERHADAP MAHASISWA PPKn FKIP LABUHANBATU. *CIVITAS (JURNAL PEMBELAJARAN DAN ILMU CIVIC)*, *7*(2), 41-44. DOI: https://doi.org/10.36987/civitas.v7i2.3535.

Das, S., Barik, R., & Mukherjee, A. (2020). Salary prediction using regression techniques. *Proceedings of Industry Interactive Innovations in Science, Engineering & Technology (I3SET2K19)*. DOI: https://dx.doi.org/10.2139/ssrn.3526707

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

Darman, R. (2018). Analisis Visualisasi dan Pemetaan Data Tanaman Padi di Indonesia Menggunakan Microsoft Power BI. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, *4*(2), 156-162. DOI: http://dx.doi.org/10.24014/rmsi.v4i2.5271.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64-73.

Firmansyah, D., & Saepuloh, D. (2022). Daya Saing: Literasi Digital dan Transformasi Digital. *Journal of Finance and Business Digital*, *1*(3), 237-250. DOI: https://doi.org/10.55927/jfbd.v1i3.1348.

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*, *3*(3), 301 - 310.

Hope, T. M. (2020). *Chapter 4—Linear regression*. In A. Mechelli & S. Vieira (Eds.), Machine Learning. Academic Press, 67-81.

Khan, S. P., Wahyudin, W., 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 KTMY. *Jurnal Serambi Engineering*, *8*(1), 15-21. DOI: https://doi.org/10.32672/jse.v8i1.5523.

Kusuma, M. D. H., & Hidayat, S. (2024). Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(2), 1097-1110. DOI: https://doi.org/10.35870/jimik.v5i2.629.

Mahaputra, M. R. (2022). Effect of Salary and Work Environment on Productivity (Study of Human Resource Management Literature). *International Journal of Advanced Multidisciplinary*, *1*(2), 153-162. DOI: https://doi.org/10.38035/ijam.v1i2.73.

Nasution, M. K., Sitompul, O. S., & Nababan, E. B. (2020). Data Science. *Journal of Physics: Conference Series*, 1-11.

Nurfarisi, R. (2022, June). Visualisasi Data Covid-19 Klinik MariSehat Menggunakan Microsoft Power BI. In *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis* (pp. 146-149).

Prabowo, D., & Sari, B. W. (2019). Fuzzy Tsukamoto Dan Mamdani Untuk Penentuan Bonus Gaji Pegawai PT. Indonesia IT. *INFOS Journal-Information System Journal*, *2*(1), 25-31.

Rahmadani, D. A. (2023). PEMANFAATAN KEMAJUAN TEKNOLOGI INFORMASI TERHADAP PERKEMBANGAN AKUNTANSI. *Jurnal Ilmu Data*, *3*(2).

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 2 (2024) | May
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i2.657

Syamsu, M., & Widodo, W. (2021). Peran Data Science dan Data Scientist Untuk Mentransformasi Data Dalam Industri 4.0. *Jurnal Teknologi Informasi (JUTECH)*, *2*(1), 27-36. DOI: https://doi.org/10.32546/jutech.v2i1.1540.

Wahyudi, H. S., & Sukmasari, M. P. (2018). Teknologi dan kehidupan masyarakat. *Jurnal Analisa Sosiologi*, *3*(1), 13-24.

Zikra, A. A. (2022). Perancangan Dashboard Accabsensi Dan Opcent Menggunakan Power Bi Di Astra Credit Companies.