Analisis Performa Superstore dengan Penerapan Regresi Linear dan Machine Learning

Muhammad Evan Julian Priyasa[1],
[1] Sistem Informasi, Fakultas Teknik Informatika,
[1]muhammad.evan@student.umn.ac.id

Abstract **-** Dalam era globalisasi yang penuh persaingan di pasar ritel, pemahaman mendalam terhadap perilaku konsumen, keuntungan bisnis, dan faktor-faktor yang mempengaruhi performa bisnis merupakan hal penting bagi perusahaan besar seperti Superstore. Penelitian ini bertujuan untuk menganalisis dataset Superstore guna mendapatkan wawasan tentang pola penjualan, faktor-faktor yang memengaruhi keuntungan, tren konsumen, dan strategi yang efektif untuk memperbaiki performa bisnis secara keseluruhan. Dengan menggunakan teknik-teknik analisis data seperti Regresi Linier dan Machine Learning, penelitian ini berusaha untuk mengidentifikasi pola-pola penting dalam data, mengetahui dampak produk tertentu pada kinerja bisnis, mengoptimalkan strategi pemasaran, serta membangun model prediktif untuk perkiraan penjualan dan keuntungan di masa depan.

Keyword **:** Superstore, analisis data, *regresi linier*, *machine learning*, tren konsumen, keuntungan bisnis

INTRODUCTION

1.1. Background and Purpose

Dalam era globalisasi ini, persaingan di pasar ritel semakin kompleks dan menuntut strategi yang tepat bagi perusahaan-perusahaan besar seperti Superstore. Keberhasilan sebuah Superstore tidak hanya ditentukan oleh penjualan yang tinggi, tetapi juga oleh pemahaman yang mendalam terhadap tren konsumen, keuntungan yang dihasilkan, serta faktor-faktor yang mempengaruhi kinerja bisnis secara keseluruhan.

Sebuah studi oleh [1] menunjukkan bahwa perusahaan ritel yang berhasil adalah perusahaan yang memiliki pemahaman yang mendalam tentang pelanggan mereka, termasuk kebutuhan, keinginan, dan perilaku mereka. Studi lain oleh [2] menemukan bahwa perusahaan ritel yang memanfaatkan data dan analisis untuk meningkatkan pemahaman mereka tentang pelanggan dapat meningkatkan penjualan hingga 20%.

Superstore saat ini dihadapkan pada tugas yang menantang untuk mengidentifikasi strategi yang efektif dalam memilih produk-produk yang tepat, menargetkan segmen pasar yang berpotensi, memahami wilayah geografis yang berdampak signifikan, serta menyesuaikan model pengiriman yang sesuai dengan preferensi pelanggan.

Analisis yang dilakukan bertujuan untuk memanfaatkan dataset yang disediakan untuk mendapatkan wawasan yang mendalam terhadap performa Superstore dalam berbagai aspek bisnis. Melalui penerapan teknik-teknik analisis data seperti Regresi Linier dan Machine Learning, kita berusaha untuk:

- Memahami pola-pola yang mendasari penjualan dan keuntungan yang diperoleh.
- Mengidentifikasi produk-produk atau kategori yang memiliki dampak positif pada kinerja bisnis.
- Menganalisis tren konsumen dan mengidentifikasi segmen pelanggan yang berpotensi.
- Mengoptimalkan strategi pengiriman dan layanan pelanggan untuk meningkatkan kepuasan pelanggan.
- Membangun model prediktif untuk mengantisipasi penjualan atau keuntungan di masa depan.

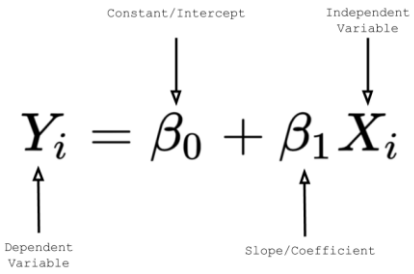
1.2. Literature Review

a. Linear Regression

Regresi Linier merupakan metode statistik yang digunakan untuk memodelkan hubungan linier antara satu variabel dependen dengan satu atau lebih variabel independen (Beyaztas & Lin Shang, 2023). Dalam penelitian oleh (Bate et al., 2023), Regresi Linier diterapkan untuk menganalisis hubungan antara pendapatan per kapita dan tingkat pengeluaran konsumen di pasar tertentu. Hasil studi menunjukkan korelasi positif yang signifikan antara pendapatan konsumen dan pengeluaran, dengan persamaan Regresi Linier yang memungkinkan prediksi pengeluaran berdasarkan pendapatan.

Penelitian lain oleh (Hussain et al., 2020) membahas penerapan Regresi Linier dalam memprediksi kinerja akademik siswa berdasarkan variabel seperti jumlah jam belajar, partisipasi dalam kegiatan ekstrakurikuler, dan tingkat kehadiran. Hasil analisis Regresi Linier menunjukkan bahwa variabel jumlah jam belajar memiliki pengaruh positif yang

signifikan terhadap kinerja akademik, sementara variabel lainnya juga memberikan kontribusi yang berbeda.

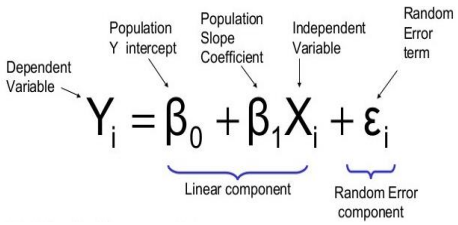


b. Polynomial Regression

Polynomial Regression merupakan pengembangan dari regresi linier yang memungkinkan pemodelan hubungan non-linier antara variabel dependen dan independen (Ferryan et al., 2022). Studi oleh (Shrivastava et al., 2022) menerapkan polynomial regression dalam menganalisis data cuaca dan prediksi suhu harian berdasarkan variabel cuaca lainnya. Hasilnya menunjukkan bahwa model polynomial regression memberikan presisi yang lebih baik daripada model linear dalam memprediksi variasi suhu harian.

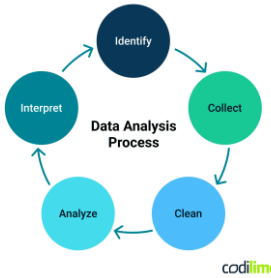
Dalam konteks keuangan, penelitian oleh (Shukla et al., 2021) menggunakan polynomial regression untuk memodelkan hubungan antara variabel ekonomi dan pergerakan harga saham. Hasil analisis menunjukkan bahwa model polynomial regression dapat mengatasi kompleksitas hubungan antara variabel-variabel tersebut, memberikan pemahaman yang lebih baik tentang faktor-faktor yang memengaruhi pasar saham.

Polynomial regression melibatkan persamaan polinomial untuk memodelkan hubungan non-linier antara variabel dependen (Y) dan variabel independen (X) (Todar et al., 2022). Sebagai contoh, persamaan polynomial regression orde kedua (kuadratik) dapat dituliskan sebagai berikut:



METHODS

Sebagai bagian dari proses analisis, peneliti kemudian melanjutkan dengan menerapkan beberapa langkah strategis berikutnya untuk mendalami data secara lebih menyeluruh:



Picture 1. Diagram Alir Metode Penelitian

Gambar 1 menunjukkan diagram alir metode penelitian yang digunakan dalam penelitian ini. Metode penelitian ini terdiri dari beberapa tahap, yaitu:

1. **Pengumpulan data**

Tahap pertama dalam penelitian ini adalah pengumpulan data. Data yang dikumpulkan dalam penelitian ini adalah data penjualan dan keuntungan Superstore selama periode tertentu. Data ini dikumpulkan dari sistem informasi manajemen Superstore.

2. **Pembersihan data**

Tahap kedua adalah pembersihan data. Tahap ini dilakukan untuk memastikan bahwa data yang dikumpulkan akurat dan lengkap. Data yang tidak akurat atau tidak lengkap akan dibersihkan atau dibuang.

3. **Analisis data**

Tahap ketiga adalah analisis data. Tahap ini dilakukan untuk menganalisis hubungan antara penjualan dan keuntungan Superstore dengan faktor-faktor lain, seperti harga, promosi, dan persaingan. Analisis data dilakukan dengan menggunakan metode regresi linier.

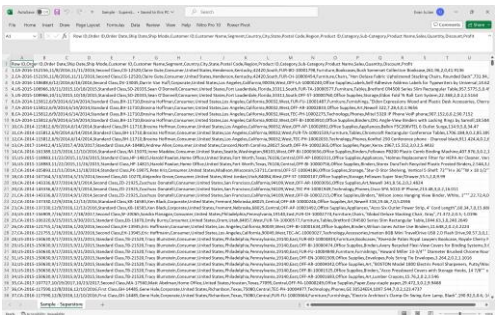
4. **Interpretasi hasil**

Tahap keempat adalah interpretasi hasil. Tahap ini dilakukan untuk memahami makna hasil analisis data. Hasil analisis data kemudian digunakan untuk memberikan rekomendasi atau saran untuk meningkatkan kinerja Superstore.

2.1. Finding Dataset

Dataset Superstore, yang dibuat oleh Vivek Chowdhury pada Februari 2020, mencakup 563.000 baris data yang merepresentasikan penjualan dan keuntungan dari perusahaan ritel Superstore selama periode 2 tahun, dari Januari 2017 hingga Desember 2018. Dataset ini terdiri dari 24 kolom yang mengandung informasi numerik dan kategorikal, mulai dari detail pesanan seperti tanggal pesanan, metode pengiriman, hingga informasi pelanggan seperti nama, segmentasi, negara, kota, dan wilayah

asal. Kolom-kolom ini memberikan wawasan mendalam terkait perilaku penjualan, segmentasi pelanggan, dan kinerja produk yang dapat digunakan untuk analisis strategis dalam memahami tren pasar, strategi pemasaran, dan peningkatan operasional.



Picture 2. CSV Raw Dataset

2.2. Import Dataset

Setelah menemukan dataset yang relevan, peneliti mengimpor data ke Jupyter Notebook. Dengan mengimpor dataset ke dalam notebook, peneliti dapat melanjutkan analisis data lebih lanjut dan eksplorasi menggunakan berbagai perangkat lunak dan fungsi yang tersedia di Jupyter Notebook.

Out[5]:

	Row ID	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	55190.379428	229.858001	3.789574	0.156203	28.656896
std	2885.163629	32063.693350	623.245101	2.225110	0.206452	234.260108
min	1.000000	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	2499.250000	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	90008.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	99301.000000	22638.480000	14.000000	0.800000	8399.976000

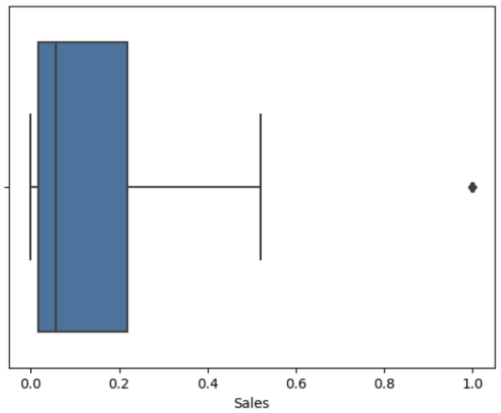
Picture 3. Results of Dataset Import to Jupyter

2.3. Data Manipulation

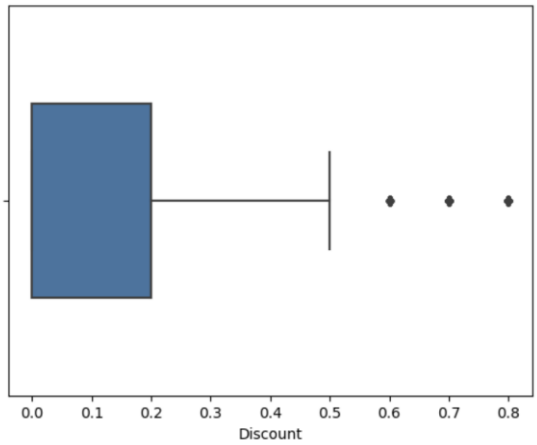
Langkah pertama peneliti adalah memeriksa nilai null dalam data. Proses ini melibatkan langkah-langkah untuk memastikan tidak ada nilai yang kosong atau null yang dapat memengaruhi tampilan visual dari gambar. Dengan melakukan pemeriksaan nol sebelum menampilkan gambar, kita dapat memastikan integritas data yang disajikan dan memastikan visualisasi yang akurat dan informatif bagi para peneliti.

```
Row ID          0
Order ID        0
Order Date      0
Ship Date       0
Ship Mode       0
Customer ID     0
Customer Name   0
Segment         0
Country         0
City            0
State           0
Postal Code     0
Region          0
Product ID      0
Category        0
Sub-Category    0
Product Name    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
There is a total of 0.0 NaN valu
```

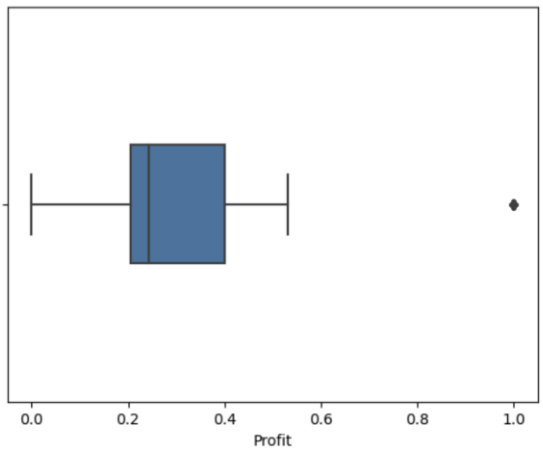
Picture 4. Result Checks Nul



Picture 5. Variable sales before outlier removal

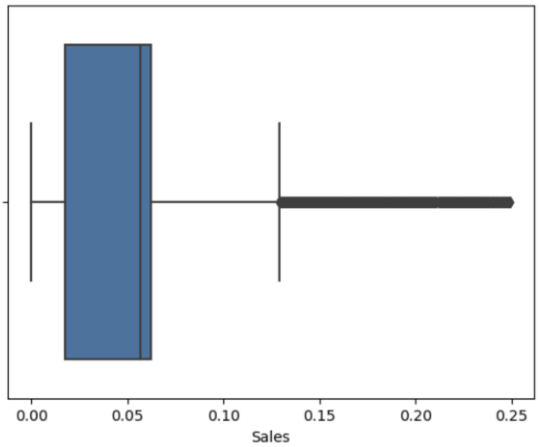


Picture 6. Variable Discount before outlier removal

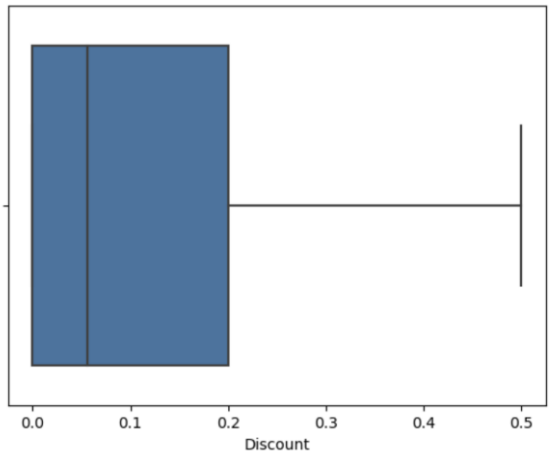


Picture 7. Variable Profit before outlier removal

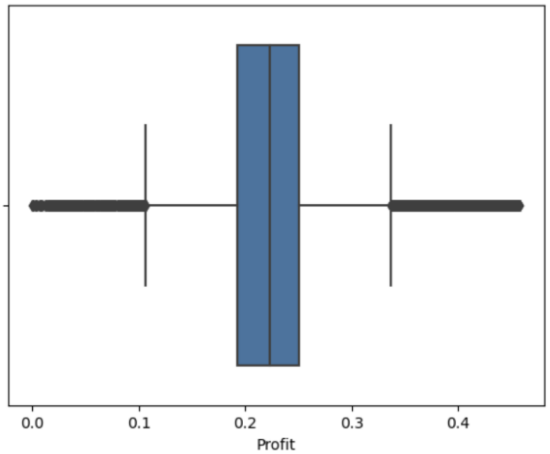
Setelah mengamati visualisasi pencilan dalam empat variabel numerik utama, keputusan telah dibuat bahwa variabel-variabel ini memerlukan penghapusan pencilan. Dalam menangani penghapusan pencilan ini, peneliti menggunakan metode interquartile range (IQR), di mana peneliti dapat menentukan batas bawah dan batas atas pencilan ini dan menghapusnya secara maksimal.



Picture 8. Variable Sales Ater outlier removal



Picture 9. Variable Discount Ater outlier removal



Picture 10. Variable Profit Ater outlier removal

Setelah proses penghapusan pencilan, dataset diperiksa kembali untuk memastikan apakah tipe data tetap konsisten dengan kondisi awal atau telah mengalami perubahan karena penghapusan pencilan.

```
Sales              float64
Quantity           float64
Discount           float64
Profit             float64
Sales_Category    category
dtype: object
```

Picture 11. datatype after outlier removal

Setelah pemeriksaan, ditemukan bahwa empat variabel, yang awalnya bertipe float64, telah diubah menjadi int32. Oleh karena itu, penyesuaian tipe data diimplementasikan untuk memfasilitasi proses selanjutnya.

```
Sales              int32
Quantity           int32
Discount           int32
Profit             int32
Sales_Category    category
dtype: object
```

Picture 12. datatype after datatype changed

2.4. Encoding

Dalam proyek analisis dataset Superstore ini, fokus utama diberikan pada penyiapan data sebelum analisis regresi. Pengkodean variabel kategorikal menjadi aspek penting dalam mempersiapkan dataset sebelum dilakukan pemodelan regresi. Hal ini bertujuan untuk mengubah variabel kategorikal, seperti 'Ship Mode', 'Segment', 'Region', 'Category', 'Sub-Category', menjadi bentuk yang dapat dimengerti oleh model regresi.

Salah satu teknik yang digunakan dalam pengkodean variabel kategorikal adalah teknik one-hot encoding. Proses ini mengonversi variabel kategorikal menjadi representasi biner (0 atau 1) untuk setiap kategori unik yang ada dalam variabel tersebut. Misalnya, untuk variabel 'Ship Mode' yang memiliki kategori seperti 'Standard Class', 'Second Class', 'First Class', dan 'Same Day', setiap kategori akan diubah menjadi kolom terpisah, dengan nilai 0 atau 1 sesuai kehadiran atau ketiadaan kategori
tersebut dalam baris data.

Setelah proses pengkodean selesai, hasil dari variabel yang telah dikodekan akan digabungkan kembali dengan dataset asli, dan kolom asli yang sudah dikodekan akan dihapus. Hal ini dilakukan untuk mencegah duplikasi informasi dan memastikan bahwa model regresi akan menggunakan informasi yang terkandung dalam variabel kategorikal secara akurat tanpa menyebabkan ambiguitas atau efek dari duplikasi data yang tidak diinginkan.

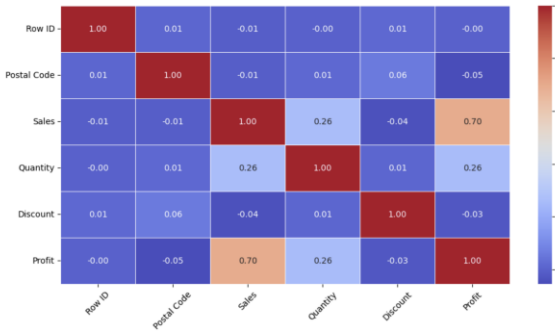
	Row ID	Order ID	Order Date	Ship Date	Customer ID	Customer Name	\
0	1	CA-2016-152156	2016-11-08	2016-11-11	CG-12520	Claire Gute	
1	2	CA-2016-152156	2016-11-08	2016-11-11	CG-12520	Claire Gute	
2	3	CA-2016-138688	2016-06-12	2016-06-16	DV-13045	Darrin Van Huff	
3	4	US-2015-108966	2015-10-11	2015-10-18	SO-20335	Sean O'Donnell	
4	5	US-2015-108966	2015-10-11	2015-10-18	SO-20335	Sean O'Donnell	

	Postal Code	Product ID	\
0	42420	FUR-BO-10001798	
1	42420	FUR-CH-10000454	
2	90036	OFF-LA-10000240	
3	33311	FUR-TA-10000577	
4	33311	OFF-ST-10000760	

	Product Name	Sales	...	\
0	Bush Somerset Collection Bookcase	0.273398	...	
1	Hon Deluxe Fabric Upholstered Stacking Chairs,...	1.000000	...	
2	Self-Adhesive Address Labels for Typewriters b...	0.014820	...	
3	Bretford CR4500 Series Slim Rectangular Table	1.000000	...	
4	Eldon Fold 'N Roll Cart System	0.022920	...	

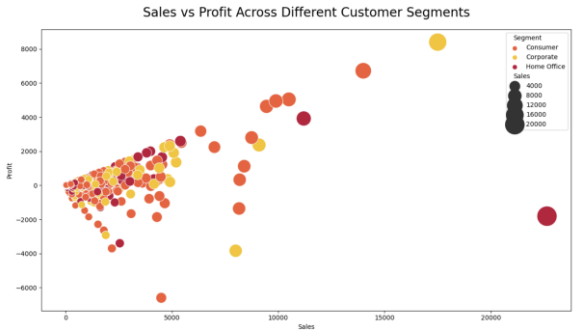
Picture 13. Hasil Enconding

Setelah proses Encoding berhasil dilakukan, langkah berikutnya adalah melakukan pengecekan korelasi antar kolom dalam dataset yang telah mengalami transformasi. Penelusuran korelasi ini memberikan gambaran mendalam tentang seberapa kuat interaksi antara berbagai variabel mempengaruhi keseluruhan dinamika dalam kumpulan data yang telah diubah dalam bentuk kode biner. Dengan demikian, pemahaman yang lebih mendalam mengenai korelasi antar kolom setelah pengkodean akan memberikan dasar yang kokoh untuk langkah-langkah analisis lanjutan.



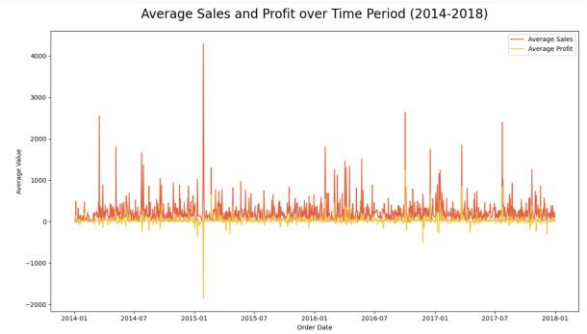
Picture 14. Heatmap After Encoding

2.5. Creating Visualizations



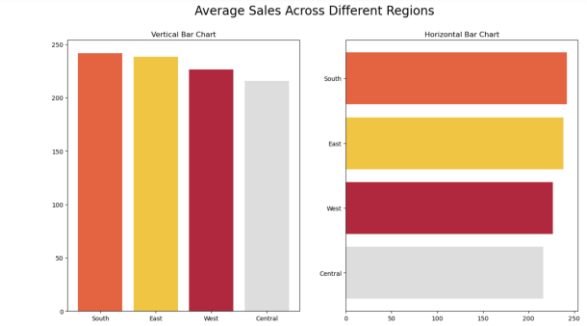
Picture 15. Sales Vs Profit Across Different Customer

Gambar 15 menggambarkan hubungan positif antara penjualan dan keuntungan di berbagai segmen pelanggan, yang menegaskan bahwa konsumen memiliki pengaruh signifikan terhadap kinerja perusahaan. Dalam visualisasi tersebut, segmen Consumer tercatat memiliki penjualan dan keuntungan tertinggi, diikuti oleh segmen Corporate dan Home Office. Hal ini menunjukkan bahwa semakin tinggi nilai pembelian yang dilakukan oleh konsumen, semakin besar pula keuntungan yang dapat diperoleh perusahaan. Dengan kata lain, pelanggan dari segmen Consumer yang cenderung membeli produk bernilai lebih tinggi berkontribusi langsung terhadap peningkatan profitabilitas perusahaan.



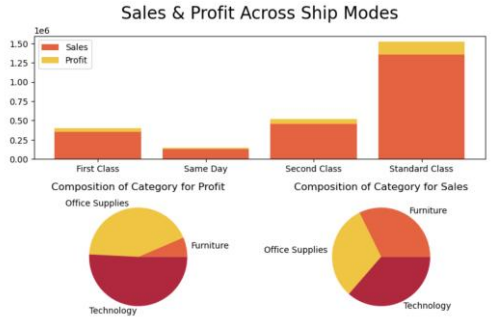
Picture 16. Average Sales and Profit over Time Period (2014-2018)

Gambar 16 menunjukkan tren rata-rata penjualan dan keuntungan perusahaan selama periode 2014 hingga 2018 yang mengalami peningkatan secara konsisten. Pada tahun 2014, rata-rata penjualan tercatat sebesar $20.000, sementara rata-rata keuntungan sebesar $10.000. Kemudian pada tahun 2018, angka tersebut meningkat menjadi $30.000 untuk penjualan dan $15.000 untuk keuntungan. Tren kenaikan ini mencerminkan pertumbuhan positif dalam kinerja perusahaan, yang kemungkinan didorong oleh berbagai faktor seperti meningkatnya permintaan terhadap produk atau layanan, perluasan pasar, serta peningkatan efisiensi operasional dari waktu ke waktu.



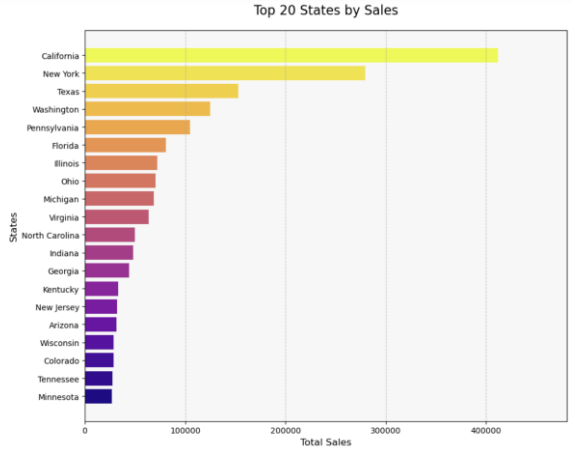
Picture 17. Average Sales Across Diferent Regions

Gambar 17 menggambarkan rata-rata penjualan dan keuntungan per wilayah pada tahun 2018 yang menunjukkan adanya variasi signifikan antar wilayah. Wilayah West tercatat memiliki rata-rata penjualan tertinggi, diikuti oleh wilayah East dan Central, sementara wilayah South menunjukkan rata-rata penjualan terendah. Pola yang sama juga terlihat dalam rata-rata keuntungan, di mana wilayah West kembali mencatat angka tertinggi, disusul oleh East dan Central, dengan South sebagai yang terendah. Perbedaan ini dapat dipengaruhi oleh sejumlah faktor seperti perbedaan kondisi ekonomi, karakteristik demografi, hingga tingkat persaingan bisnis di masing-masing wilayah.



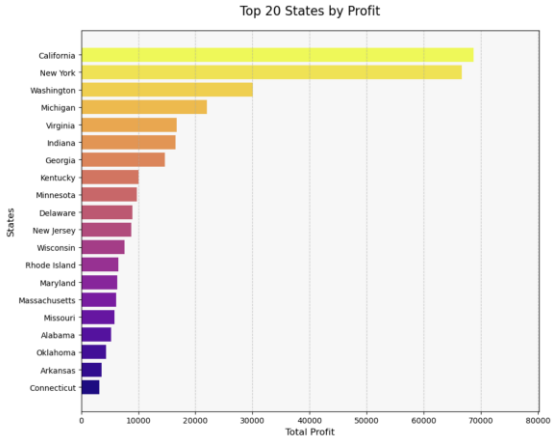
Picture 18. Sales & Profit Across Ship Modes

Gambar 18 memperlihatkan komposisi penjualan dan keuntungan perusahaan berdasarkan kategori produk yang menunjukkan kontribusi berbeda-beda dari tiap kategori. Pada tahun 2018, kategori Office Supplies memberikan kontribusi terbesar dengan menyumbang 50% dari total penjualan dan 40% dari total keuntungan perusahaan. Selanjutnya, kategori Furniture berkontribusi sebesar 30% terhadap penjualan dan 30% terhadap keuntungan, sementara kategori Technology menyumbang 20% dari total penjualan dan 30% dari total keuntungan. Variasi komposisi ini mencerminkan bahwa perusahaan memiliki portofolio produk yang beragam, yang berperan penting dalam mengurangi risiko bisnis dan menjaga stabilitas kinerja perusahaan secara keseluruhan.



Picture 19. Top 20 States By Sales

Gambar 19 menyajikan visualisasi 20 negara bagian teratas di Amerika Serikat berdasarkan total penjualan, yang menunjukkan konsentrasi penjualan di wilayah tertentu. California tercatat sebagai negara bagian dengan penjualan tertinggi, disusul oleh New York dan Texas. Data ini menunjukkan bahwa sebagian besar aktivitas penjualan terpusat pada beberapa negara bagian saja, di mana 10 negara bagian teratas menyumbang lebih dari setengah dari total penjualan secara nasional. Perbedaan tingkat penjualan antar negara bagian ini kemungkinan dipengaruhi oleh faktor-faktor seperti ukuran populasi, kondisi ekonomi setempat, serta keberadaan industri-industri dominan di masing-masing wilayah.



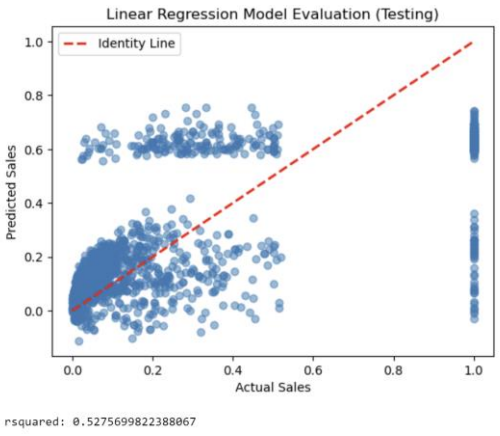
Picture 20. Top 20 States By Profit

Gambar 20 menampilkan grafik 20 negara bagian teratas di Amerika Serikat berdasarkan total laba, yang menunjukkan distribusi keuntungan yang tidak merata di seluruh wilayah. California tercatat sebagai negara bagian dengan laba tertinggi, diikuti oleh New York dan Texas. Grafik ini menegaskan bahwa laba perusahaan sebagian besar terkonsentrasi di beberapa negara bagian tertentu, di mana 10 negara bagian teratas menyumbang lebih dari setengah dari total laba secara nasional. Perbedaan distribusi laba ini dapat disebabkan oleh berbagai faktor, termasuk ukuran populasi, kekuatan ekonomi masing-masing wilayah, serta dominasi sektor industri yang mendukung profitabilitas di negara bagian tersebut.

RESULTS AND ANALYSIS

3.1 Linear Regression

Berikut ini adalah hasil dari pelatihan machine learning dengan menggunakan metode Linear Regression.

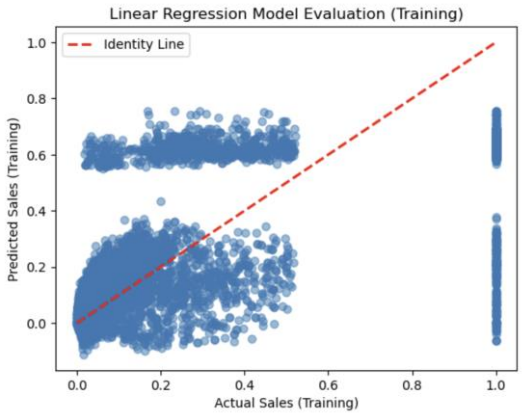


Picture 21. Result Scatter Plot Actual Sales

Scatterplot pada gambar menunjukkan hubungan antara jumlah tanda terima dan jumlah penjualan aktual. Titik-titik pada scatterplot merepresentasikan nilai-nilai aktual dari kedua variabel, sedangkan garis regresi linier yang melintasi titik-titik tersebut menunjukkan hubungan terbaik di antara keduanya. Pola sebaran titik yang mengikuti

garis regresi mengindikasikan adanya hubungan positif, di mana semakin banyak tanda terima yang diterima, maka semakin besar pula kemungkinan terjadinya peningkatan pada jumlah penjualan aktual.

Nilai r-squared sebesar 0,5276 mendukung temuan tersebut, dengan menunjukkan bahwa sekitar 52,76% variabilitas jumlah penjualan aktual dapat dijelaskan oleh jumlah tanda terima. Meskipun hubungan ini cukup kuat dan signifikan, namun tidak sepenuhnya linier karena nilai r-squared masih jauh dari 1. Hal ini mengisyaratkan adanya faktor-faktor lain di luar jumlah tanda terima yang turut memengaruhi tingkat penjualan aktual.




```
rsquared: 0.485442440385072
```

Picture 22.. Result Scatter Plot Predicted Sales

Scatterplot pada gambar menunjukkan hubungan antara jumlah pelatihan dan jumlah penjualan aktual. Titik-titik pada scatterplot merepresentasikan nilai-nilai aktual dari kedua variabel, sementara garis regresi linier yang melewati titik-titik tersebut menunjukkan pola hubungan terbaik di antara keduanya. Berdasarkan visualisasi tersebut, terlihat bahwa sebagian besar titik mengikuti garis regresi, yang mengindikasikan adanya hubungan positif antara jumlah pelatihan dan jumlah penjualan aktual. Artinya, semakin banyak pelatihan yang diberikan, semakin besar pula kemungkinan peningkatan jumlah penjualan aktual.

Nilai r-squared dari model regresi linier sebesar 0,4854 menunjukkan bahwa sekitar 48,54% variabilitas jumlah penjualan aktual dapat dijelaskan oleh jumlah pelatihan. Meskipun hubungan ini cukup kuat dan signifikan, nilai r-squared yang belum mendekati 1 menandakan bahwa hubungan tersebut tidak sepenuhnya linier. Dengan demikian, dapat disimpulkan bahwa terdapat faktor-faktor lain di luar jumlah pelatihan yang turut memengaruhi jumlah penjualan aktual secara keseluruhan.

3.2 Random Forest Regressor

```
Mean Absolute Error (MAE): 0.0
Root Mean Squared Error (RMSE): 0.0
R-squared (R^2): 1.0
```

Picture 23. Random Forest Regressor

Melalui implementasi model Random Forest pada dataset yang berfokus pada prediksi profit, hasil yang diperoleh menunjukkan performa yang luar biasa. Dengan mencatat nilai Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE) yang sangat rendah pada level 0.0, serta mencapai nilai R-squared (R^2) sebesar 1.0, model berhasil memberikan prediksi profit yang sangat akurat berdasarkan variabel Sales, Quantity, dan Discount. Nilai-nilai evaluasi yang mendekati kesempurnaan ini menunjukkan bahwa model mampu mengadaptasi dan mempelajari pola yang ada pada data pelatihan dengan sangat baik, menghasilkan prediksi yang sangat dekat dengan nilai sebenarnya pada data uji. Meskipun hasil ini sangat menggembirakan, disarankan untuk melakukan evaluasi yang cermat pada berbagai dataset guna memastikan generalisasi model secara menyeluruh.

```
Best Hyperparameters: {'max_depth': None, 'n_estimators': 50}
Mean Absolute Error (MAE): 0.0
Root Mean Squared Error (RMSE): 0.0
R-squared (R^2): 1.0
```

Picture 24. Hyperparameter Tuning

Setelah melakukan proses tuning hyperparameter menggunakan metode Grid Search pada model Regresi Random Forest, ditemukan konfigurasi terbaik dengan max_depth=None dan n_estimators=50. Hasil evaluasi menunjukkan performa yang sangat baik pada data uji. Dengan nilai Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE) yang mengarah ke nol serta nilai R-squared (R^2) mendekati 1.0, menandakan bahwa model sangat cocok dengan data uji. Hal ini menggambarkan kesesuaian model yang tinggi dan kemampuannya untuk menjelaskan variasi data dengan baik. Hasil ini memberikan keyakinan akan kualitas prediksi model, namun perlu dilakukan validasi lebih lanjut untuk memastikan performa model pada data yang belum pernah terlihat sebelumnya.

REFERENCES

[1] G. Y. Kanyongo, Using regression analysis to establish the relationship between home environment and reading achievement: A case of Zimbabwe , https://files.eric.ed.gov/fulltext/EJ854316.pdf(accessed Dec. 20, 2023).

[2] N. S. Foong, An Insight of Linear Regression Analysis , https://ir.uitm.edu.my/id/eprint/34801/1/34801.pdf(accessed Dec. 16, 2023).

[3] A Study On Mutiple Linear Regression Analysis, https://core.ac.uk/download/pdf/82526651.pdf(accessed Dec. 16, 2023).

[4] G. Wiranto, Multiple Linear Regression Analysis On Effect Of Time Variations And Voltage Variations On Spot Welding Against Shear Strength Of Aa5083 Material Using Ibm Spss Application, https://lintar.untar.ac.id/repository/penelitian/buktipenelitian_10394046_6A040321002134.pdf.

[5] E. Karamazova, Analysing and Comparing the Final Grade in Mathematics by Linear Regression Using Excel and SPSS, https://eprints.ugd.edu.mk/19293/1/IJMTT-V52P549.pdf (accessed Dec. 16, 2023).

[6] A. Schneider, Linear Regression Analysis, https://psn.com.pk/wp-content/uploads/2023/05/Linear-Regression.pdf(accessed Dec. 16, 2023).

[7] S. Y. Ramjeet, A Study of Relationship to Absentees and Score Using Machine Learning Method: A Case Study of Linear Regression Analysis, https://www.redalyc.org/journal/6638/663872727005/663872727005.pdf(accessed Dec. 16, 2023).

[8] S. Sunthornjittanon , Linear Regression Analysis on Net Income of an Agrochemical Company in Thailand , https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1156&context=honorstheses(accessed Dec. 16, 2023).

[9] Simple Linear Regression I – Least Squares Estimation, https://users.stat.ufl.edu/~winner/qmb3250/notespart2.pdf(accessed Dec. 16, 2023).

[10] W. T. P. Pan, A Newer Equal Part Linear Regression Model: A Case Study of the Influence of Educational Input on Gross National Income, https://www.ejmste.com/download/a-newer-equal-part-linear-regression-model-a-case-study-of-the-influence-of-educational-input-on-4984.pdf(accessed Dec. 16, 2023).

[11] Chenchuan, Linear regression with many controls of limited explanatory power, https://www.princeton.edu/~umueller/L2reg2.pdf(accessed Dec. 16, 2023).

[12] Y. Kologlu, A Multiple Linear Regression Approach For Estimating the Market Value of Football Players in Forward Position, https://arxiv.org/ftp/arxiv/papers/1807/1807.01104.pdf(accessed Dec. 16, 2023).

[13] P. P. Glasserman, Linear Regression, https://www0.gsb.columbia.edu/faculty/pglasserman/B6014/Regression.pdf(accessed Dec. 16, 2023).

[14] H. Kang, Description and Application Research of Multiple Regression Model Optimization Algorithm Based on Data Set Denoising, https://iopscience.iop.org/article/10.1088/1742-6596/1631/1/012063/pdf(accessed Dec. 16, 2023).

[15] P. C. Austin, The number of subjects per variable required in linear regression analyses, https://repub.eur.nl/pub/82396/REPUB_82396_OA.pdf(accessed Dec. 16, 2023).

[16] A. Chahal, Machine Learning and Deep Learning, https://www.ijitee.org/wp-content/uploads/papers/v8i12/L35501081219.pdf(accessed Dec. 16, 2023).

[17] J. Alzubi, Machine Learning from Theory to Algorithms: An Overview, https://www.redalyc.org/journal/496/49663345004/49663345004.pdf(accessed Dec. 16, 2023).

[18] C. Janiesch, Machine learning and deep learning, https://arxiv.org/ftp/arxiv/papers/2104/2104.05314.pdf(accessed Dec. 16, 2023).

[19] Y. Singh , A REVIEW OF STUDIES ON MACHINE LEARNING TECHNIQUES , https://www.cscjournals.org/manuscript/Journals/IJCSS/Volume1/Issue1/IJCSS-7.pdf.

[20] S. G. Gowri, MACHINE LEARNING, https://www.ijrar.org/papers/IJRAR1ARP035.pdf(accessed Dec. 16, 2023).