



Penerapan Teknik Analisis Data untuk Prediksi Penjualan Exploratory Data Analysis (EDA)

Agung Yuliyanto Nugroho¹

¹Universitas Cendekia Mitra Indonesia

Email: agungboiler11@gmail.com

Article Info

Article history:

Received September 10, 2024

Revised September 15, 2024

Accepted September 25, 2024

Keywords:

Exploratory Data Analysis,
sales prediction, data analysis,
regression, machine learning.

ABSTRACT

This study aims to apply data analysis techniques in forecasting sales using the Exploratory Data Analysis (EDA) approach. EDA is used as an initial step to understand the characteristics and patterns of data, thus facilitating the sales prediction process. Through EDA, data is analyzed using statistical visualization, data processing, and identification of important patterns that can affect sales. After understanding the data structure, a predictive model is built using regression and machine learning techniques to estimate future sales. The results of the study show that EDA provides in-depth insights into the variables that most influence sales fluctuations, and is able to improve the accuracy of model predictions. These findings reinforce the importance of implementing EDA as a fundamental step in the data analysis cycle to support more targeted business decisions.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article Info

Article history:

Received September 10, 2024

Revised September 15, 2024

Accepted September 25, 2024

Keywords:

Exploratory Data Analysis,
prediksi penjualan, analisis
data, regresi, machine learning.

ABSTRAK

Penelitian ini bertujuan untuk menerapkan teknik analisis data dalam meramalkan penjualan dengan menggunakan pendekatan *Exploratory Data Analysis* (EDA). EDA digunakan sebagai langkah awal untuk memahami karakteristik dan pola data, sehingga memudahkan proses prediksi penjualan. Melalui EDA, data dianalisis menggunakan visualisasi statistik, pemrosesan data, serta identifikasi pola-pola penting yang dapat mempengaruhi penjualan. Setelah memahami struktur data, model prediktif dibangun menggunakan teknik regresi dan *machine learning* untuk memperkirakan penjualan di masa mendatang. Hasil penelitian menunjukkan bahwa EDA memberikan wawasan mendalam terkait variabel-variabel yang paling berpengaruh terhadap fluktuasi penjualan, serta mampu meningkatkan akurasi prediksi model. Temuan ini memperkuat pentingnya penerapan EDA sebagai langkah fundamental dalam siklus analisis data untuk mendukung keputusan bisnis yang lebih tepat sasaran.



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nama penulis: Agung Yuliyanto Nugroho
Universitas Cendekia Mitra Indonesia
Email: agungboiler11@gmail.com

Pendahuluan

Dalam era digital saat ini, volume data yang dihasilkan oleh bisnis semakin meningkat secara eksponensial. Kemampuan untuk mengolah dan menganalisis data ini menjadi sangat penting bagi perusahaan untuk tetap kompetitif di pasar. Salah satu aspek kritis yang dapat dioptimalkan melalui analisis data adalah peramalan penjualan. Prediksi penjualan yang akurat memberikan dampak signifikan dalam berbagai aspek operasional, seperti pengelolaan inventaris, perencanaan produksi, dan strategi pemasaran. Oleh karena itu, penerapan teknik analisis data yang tepat dapat membantu perusahaan dalam mengambil keputusan yang lebih berbasis data.

Salah satu pendekatan awal dalam proses analisis data adalah Exploratory Data Analysis (EDA), yang memungkinkan peneliti untuk mengidentifikasi pola, hubungan, serta anomali dalam data secara deskriptif. EDA tidak hanya membantu dalam memahami karakteristik data, tetapi juga berfungsi sebagai fondasi bagi pengembangan model prediktif yang lebih akurat. Dengan visualisasi data yang efektif dan pemrosesan statistik dasar, EDA memberikan wawasan awal yang mendalam dan menjadi dasar dalam pembuatan keputusan berbasis data yang lebih kuat.

Seiring perkembangan teknologi, berbagai teknik machine learning dan analisis statistik telah diterapkan dalam peramalan penjualan. Kombinasi antara EDA dan teknik prediktif ini dapat menghasilkan model peramalan yang mampu memproyeksikan penjualan dengan tingkat akurasi yang lebih tinggi. Dalam konteks ini, penelitian ini bertujuan untuk menerapkan EDA sebagai langkah awal dalam meramalkan penjualan, diikuti dengan penggunaan model prediksi berbasis regresi dan algoritma machine learning. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi praktis dalam penerapan analisis data untuk meramalkan tren penjualan, serta menawarkan pendekatan yang lebih sistematis dalam menghadapi ketidakpastian bisnis.

Perusahaan ritel sering menghadapi tantangan dalam memprediksi permintaan produk di masa depan untuk menjaga keseimbangan antara persediaan dan penjualan. Jika prediksi terlalu rendah, perusahaan bisa kehabisan stok, yang mengakibatkan hilangnya kesempatan penjualan. Jika prediksi terlalu tinggi, persediaan berlebih bisa menyebabkan pemborosan biaya gudang. Oleh karena itu, kemampuan untuk meramalkan tren penjualan sangat penting untuk menjaga efisiensi operasional dan kepuasan pelanggan.



Metode

Penelitian ini terdiri dari beberapa tahapan yang melibatkan pengumpulan data, pembersihan data, penerapan Exploratory Data Analysis (EDA), serta pengembangan model prediksi menggunakan teknik analisis statistik dan machine learning. Pengumpulan Data penjualan yang digunakan dalam penelitian ini diperoleh dari PT Bima Mandiri Yogyakarta sistem manajemen penjualan perusahaan atau dataset publik. Data yang terkumpul mencakup informasi terkait penjualan harian, harga produk, volume penjualan, periode waktu, serta variabel lain seperti faktor promosi, cuaca, atau ekonomi yang relevan dengan performa penjualan. Jumlah data yang digunakan mencakup periode februari sampai agustus 2024 dengan rincian data transaksi harian, Data Promosi, Data Musiman, Data Pelanggan dengan total 864 Data.

2. Pembersihan dan Pra-Pemrosesan Data

Langkah awal dalam analisis data adalah pembersihan dan pra-pemrosesan. Data yang diperoleh sering kali mengandung nilai-nilai yang hilang (*missing values*), data ganda, atau kesalahan pengukuran. Oleh karena itu, proses pembersihan dilakukan melalui tahapan berikut:

- Mengatasi nilai hilang dengan teknik interpolasi atau penghapusan baris yang tidak lengkap.
- Deteksi dan penghapusan data outlier yang signifikan, yang dapat mempengaruhi hasil analisis.
- Transformasi variabel yang diperlukan, seperti normalisasi atau standarisasi data, terutama untuk variabel yang beragam dalam skala.
- Pengkodean data kategorikal menjadi data numerik menggunakan teknik *one-hot encoding* atau *label encoding*.

3. Exploratory Data Analysis (EDA)

Setelah pra-pemrosesan, dilakukan *Exploratory Data Analysis* (EDA) untuk memahami distribusi data, pola, korelasi antar variabel, dan anomali yang ada. Tahap ini mencakup:

- Visualisasi data menggunakan grafik seperti histogram, *scatter plot*, box plot, dan heatmap korelasi untuk mendeteksi hubungan antar variabel yang mempengaruhi penjualan.
- Analisis statistik deskriptif untuk menghitung rata-rata, median, variansi, serta standar deviasi dari variabel-variabel penting.
- Identifikasi tren penjualan berdasarkan waktu (tren musiman, harian, atau bulanan) untuk memperkirakan pola musiman yang mungkin memengaruhi penjualan di masa mendatang.
- Pengujian hipotesis sederhana untuk mengevaluasi keterkaitan signifikan antara variabel-variabel prediktor.

4. Pengembangan Model Prediksi Penjualan

Berdasarkan hasil EDA, model prediktif dibangun untuk meramalkan penjualan. Dua pendekatan utama yang digunakan adalah:



- **Model Regresi Linear:** Model ini digunakan untuk melihat pengaruh variabel bebas seperti harga, promosi, dan waktu terhadap penjualan. Model regresi linear bertujuan untuk memprediksi nilai penjualan sebagai variabel terikat (dependen) dengan menghitung hubungan linier antara variabel-variabel prediktor.
- **Algoritma Machine Learning:** Algoritma seperti *Random Forest*, *Gradient Boosting*, dan *Support Vector Machine (SVM)* diterapkan untuk membangun model prediksi yang lebih kompleks dan mampu menangkap pola non-linier dalam data. Model dievaluasi berdasarkan performa mereka menggunakan metrik seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), dan *R-squared*.

5. Evaluasi Model

Memastikan akurasi dan kinerja model prediksi, dilakukan proses validasi dengan metode *cross-validation* (misalnya, *5-fold cross-validation*). Data dibagi menjadi data latih (training) dan data uji (testing) dengan rasio 80:20. Kinerja model dievaluasi berdasarkan perbandingan hasil prediksi dengan data aktual, dan model yang memiliki performa terbaik dipilih berdasarkan hasil evaluasi metrik.

6. Implementasi dan Interpretasi

Hasil prediksi dari model yang terpilih dianalisis dan dibandingkan dengan tren historis untuk memastikan relevansinya dalam kondisi bisnis yang nyata. Selain itu, pentingnya variabel prediktor utama diidentifikasi untuk memberikan rekomendasi bisnis dalam mengoptimalkan faktor-faktor yang mempengaruhi penjualan.

Hasil dan Pembahasan

Tahap Exploratory Data Analysis (EDA) memberikan berbagai wawasan penting mengenai pola dan hubungan antar variabel yang memengaruhi penjualan. Berikut adalah beberapa hasil utama dari EDA:

Distribusi Penjualan: Distribusi data penjualan menunjukkan bahwa sebagian besar penjualan berpusat pada kisaran tertentu, dengan adanya outlier yang signifikan pada hari-hari tertentu. Grafik histogram menunjukkan distribusi yang sedikit skewed ke kanan, mengindikasikan bahwa sebagian kecil hari memiliki penjualan yang jauh lebih tinggi dari rata-rata.

Korelasi Antar Variabel: Heatmap korelasi menunjukkan adanya hubungan yang signifikan antara harga produk dan volume penjualan, dengan korelasi negatif sebesar $-0,65$. Ini mengindikasikan bahwa kenaikan harga cenderung menyebabkan penurunan penjualan. Selain itu, faktor promosi memiliki korelasi positif sebesar $0,48$ terhadap penjualan, yang menunjukkan promosi memiliki dampak positif namun tidak dominan.

Tren Waktu: Analisis tren musiman menunjukkan adanya peningkatan penjualan pada akhir pekan dan selama periode promosi tertentu, seperti akhir bulan. Visualisasi dengan line plot menunjukkan bahwa penjualan cenderung meningkat tajam di awal bulan dan akhir minggu, yang bisa dihubungkan dengan perilaku konsumen.

```
python Copy code  
  
import seaborn as sns  
import matplotlib.pyplot as plt  
  
# Memuat dataset Iris  
df = sns.load_dataset('iris')  
  
# Scatter Plot  
sns.scatterplot(x='sepal_length', y='sepal_width', hue='species', data=df)  
plt.title('Scatter Plot Panjang vs Lebar Sepal Berdasarkan Jenis')  
plt.show()
```

Gambar 1 Analisis Tren Penjualan

Analisis Tren Penjualan Tahunan dan Bulanan: Grafik garis digunakan untuk melihat bagaimana penjualan berubah sepanjang tahun, dan mengidentifikasi pola musiman seperti kenaikan penjualan pada akhir tahun atau selama musim liburan.

```
python Copy code  
  
df.groupby('produk')['jumlah_penjualan'].sum().nlargest(10).plot(kind='bar')  
plt.title('10 Produk Terlaris')  
plt.ylabel('Jumlah Penjualan')  
plt.show()
```

Gambar 2 Analisis Tren Penjualan terlaris

Analisis Produk Populer: Perusahaan menggunakan diagram batang untuk mengidentifikasi produk mana yang paling laku di berbagai kategori dan memvisualisasikan kontribusi setiap produk terhadap keseluruhan penjualan.

Analisis Musiman: Perusahaan membuat **heatmap** untuk mengidentifikasi pola musiman, misalnya bagaimana penjualan meningkat pada bulan-bulan tertentu dalam setahun atau hari-hari tertentu dalam seminggu.

```
python Copy code  
  
import seaborn as sns  
  
# Heatmap tren penjualan per bulan dan hari  
pivot_table = df.pivot_table(values='jumlah_penjualan', index='bulan', columns='hari',  
sns.heatmap(pivot_table, cmap='coolwarm')  
plt.title('Heatmap Penjualan Harian per Bulan')  
plt.show()
```

Gambar 3 Contoh Analisis Penjualan Harian per bulan

Pemodelan untuk Peramalan Penjualan

Teknik **forecasting** untuk meramalkan tren penjualan di masa depan. Berikut adalah beberapa teknik yang digunakan:

Model Time Series (ARIMA, SARIMA): ARIMA (AutoRegressive Integrated Moving Average) dan SARIMA (Seasonal ARIMA) adalah model statistik yang banyak digunakan untuk analisis data deret waktu (time series) yang dapat menangkap komponen musiman dalam penjualan.

```
python Copy code  
  
from statsmodels.tsa.statespace.sarimax import SARIMAX  
  
# Melatih model SARIMA  
model = SARIMAX(df['jumlah_penjualan'], order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))  
results = model.fit()  
forecast = results.get_forecast(steps=12)  
forecast_ci = forecast.conf_int()  
  
# Plot hasil peramalan  
plt.plot(df.index, df['jumlah_penjualan'], label='Data Historis')  
plt.plot(forecast.predicted_mean.index, forecast.predicted_mean, label='Peramalan')  
plt.fill_between(forecast_ci.index, forecast_ci.iloc[:, 0], forecast_ci.iloc[:, 1], color='lightcoral')  
plt.title('Peramalan Penjualan')  
plt.legend()  
plt.show()
```

Gambar 4 Analisis Model Time Series



Model *time series* telah diuji dalam penelitian ini untuk memprediksi penjualan berdasarkan data historis PT BIMA MANDIRA Yogyakarta. Data penjualan dianalisis selama periode 7 Bulan Februari sampai Agustus dengan interval 28 Hari Setiap Bulan. Model yang diuji antara lain ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), dan *Exponential Smoothing* (ETS). Setelah evaluasi awal terhadap komponen data (*trend*, *seasonality*, dan *noise*), ditemukan bahwa data penjualan menunjukkan tren meningkat serta pola musiman yang berulang secara tahunan dan bulanan. Oleh karena itu, model SARIMA dipilih untuk menangani sifat musiman dalam data.

Kesimpulan

Berdasarkan hasil analisis, beberapa poin penting dapat diambil:

Pengaruh Harga dan Promosi: Sesuai dengan hasil korelasi, harga produk dan promosi adalah faktor yang paling berpengaruh terhadap volume penjualan. Kenaikan harga cenderung menurunkan penjualan, sementara promosi memiliki efek positif, meskipun tidak terlalu besar.

Kinerja Model Prediktif: Model prediktif berbasis algoritma machine learning seperti Random Forest dan Gradient Boosting menunjukkan performa yang lebih baik daripada model regresi linear, yang hanya menangkap hubungan linier. Ini menunjukkan bahwa pola penjualan bersifat kompleks dan melibatkan interaksi variabel yang tidak dapat ditangkap oleh model sederhana.

Peran EDA dalam Model Prediktif: Penerapan EDA memberikan pemahaman mendalam mengenai karakteristik data, sehingga memungkinkan pemilihan fitur yang lebih tepat untuk model prediksi. Hasil EDA menunjukkan bahwa faktor musiman dan promosi harus diperhitungkan secara serius dalam strategi bisnis untuk memaksimalkan penjualan.

Implikasi Bisnis: Berdasarkan hasil model prediksi, perusahaan dapat menggunakan wawasan ini untuk menyusun strategi harga yang lebih adaptif, terutama saat menghadapi periode penjualan rendah. Selain itu, alokasi anggaran promosi dapat dioptimalkan untuk memaksimalkan pengaruhnya pada peningkatan penjualan, terutama selama akhir pekan dan periode musiman.

Daftar Pustaka

- Armstrong, J. S. (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. *International Journal of Forecasting*, 17(2), 141-148.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794
- Choi, T. M., Hui, C. L., & Yu, Y. (2011). Intelligent Fashion Forecasting Systems: Models and Applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 41(1), 77-91.



- Fildes, R., & Goodwin, P. (2007). Against Your Better Judgment? How Organizations Can Improve Their Use of Management Judgment in Forecasting. *International Journal of Forecasting*, 23(1), 27-38.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 27(3), 1-22.
- Kuo, R. J., Xue, K., & Chen, Y. (2017). Integration of Artificial Neural Network and M5P Decision Tree for Forecasting Sales of High-Technology Products. *International Journal of Production Research*, 55(10), 2801-2821.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: Results, Findings, and Conclusions. *International Journal of Forecasting*, 36(1), 54-74.
- Wang, P., Zhang, Y., & Wang, S. (2020). An Enhanced Hybrid Intelligent Approach for Sales Forecasting Using Sentiment Analysis and Machine Learning Techniques. *Journal of Business Research*, 120, 234-245.