Final Project Data Science

Jihad Akbar

+62 813 3232 6785
https://www.linkedin.com/in/jihadakbr
jihadakbr@gmail.com

TABLE OF CONTENTS

01

Introduction

Jihad Akbar

linkedin.com/in/jihadakbr

"Transforming complexity into **clarity**, uncovering **insights**, and turning data into actionable strategies through **collaboration**."

Experience

Now–Oct 2024
Machine Learning Annotator
PT Nomura Research Institute Indonesia

Sep–Feb 2024
Data Annotator
CVAT.ai Corporation

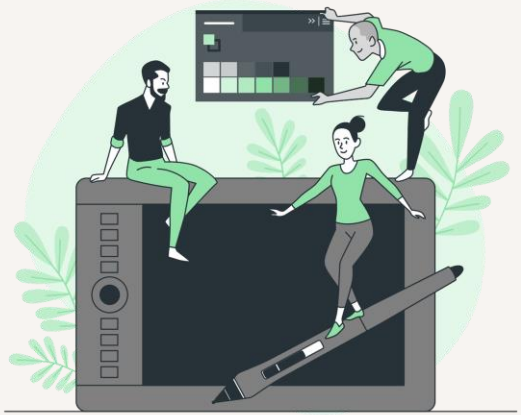
Education

Now–Oct 2024
Data Science Bootcamp
dibimbing.id

2020–2016
Bachelor of Science in Physics
Universitas Sebelas Maret

02

Previous projects



What I Have Learned in the Bootcamp

New York City Taxi Trip Duration (Oct-Dec 2024)

- Predicted NYC taxi trip durations using 2016 data by defining business objectives, preprocessing data, engineering features like distance and speed, and addressing NYC-specific location constraints.
- Built scalable machine learning workflows using reusable transformer components with OOP principles in scikit-learn pipelines, achieving an RMSLE score of 0.55 with a LightGBM Regressor.

github.com/jihadakbr/new-york-city-taxi-trip-duration

Credit Card Fraud Detection (Sep-Oct 2024)

- Developed a fraud detection model using European credit card transaction data, addressing class imbalance and optimizing fraud detection accuracy by prioritizing the precision-recall AUC metric and adjusting the decision threshold.
- Fine-tuned models with Grid Search CV, achieving a PR AUC of 0.812 and F2 score of 0.794 using the Extra Trees Classifier, with a decision threshold set at 0.570 to balance accuracy and minimize false negatives.

github.com/jihadakbr/credit-card-fraud-detection

FutureSight: Revolutionizing E-Commerce with Data-Driven Forecasting

03

Main Project

MAIN PROJECT OUTLINE

01 Project Background

02 Business Objective

03 Data Understanding

04 Data Preprocessing

05 Findings and Results

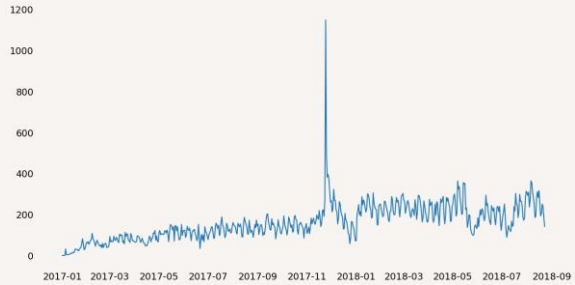
06 Recommendations

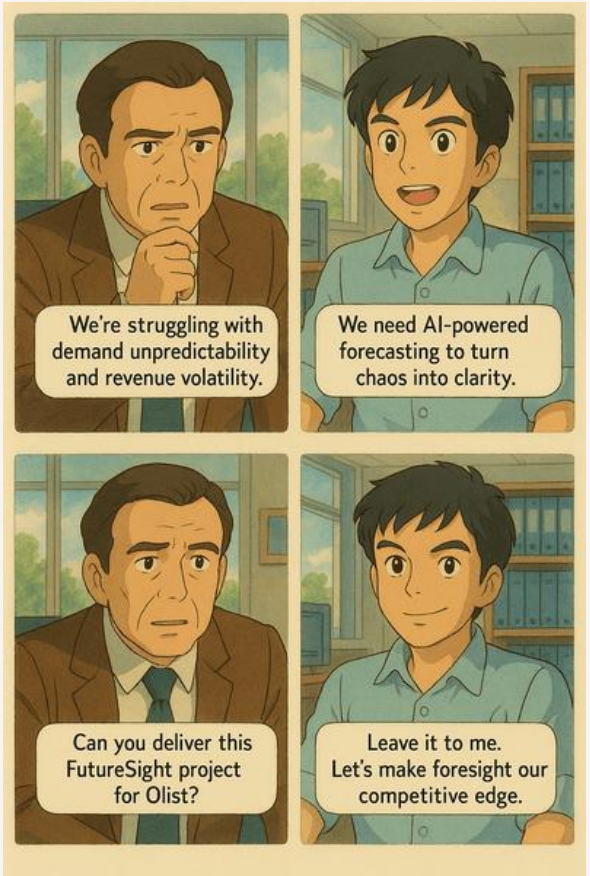
Project Background

Olist, one of Brazil's largest e-commerce platforms, faces critical challenges despite its growing sales:

- ⚠ Unpredictable Order Volumes: Sudden fluctuations make it challenging to manage resources effectively.
- ⚠ Revenue Volatility: Shifting revenue trends are hard to track, hindering strategic planning.

Daily Orders Over Time





Ghibli-style Data Scientist Illustration created by ChatGPT

Disclaimer: *The following story is fictional and created solely to illustrate the challenges addressed by this project.*

Business Objective

The Mission: Predict the Future, Optimize the Present

The primary goal of this project is to build a state-of-the-art time series forecasting system that accurately predicts key business metrics. This mission includes:

- 📈 Accurately forecast order volumes to optimize inventory and resource management, minimizing stockouts and overstocking.
- 📈 Forecast revenue trends to provide a clear financial roadmap, supporting strategic planning and long-term growth.

This project, **FutureSight**, will deliver an AI-driven forecasting engine that leverages advanced analytics to address these business-critical need.

**Powered by Olist's real transaction data, comprising over 100,000 orders across Brazil's diverse regions, collected from Kaggle.*

Data Understanding

Collected from Kaggle – Brazilian E-Commerce Public Dataset by Olist.

Approximately 100,000 orders from Oct 2016 to Sep 2018 across Brazil.

Includes order details, payments, customer and seller information, product data, shipping logistics, and customer reviews.

Data Preprocessing

Step 1

Combine Data

Merged related CSV files for daily order and revenue forecasting.

Step 2

Filter Relevant Data

Focused on delivered orders (order_status = 'delivered')

Step 3

Remove Anomalous Dates

Excluded dates with near-zero orders or revenue, keeping data from Jan 2017–Aug 2018).

Data Preprocessing

Step 4

Exclude Outlier Dates

Removed end-of-period dates with unusually low activity, likely due to in-transit orders.

Step 5

Convert Data Types

Reformatted ‘date’ to date type

Step 6

Address Outliers

Identified Black Friday 2017 as an outlier. Handled as an exogenous variable in the model.

Data Preprocessing

Step 7

Analyze Seasonality

Detected weekly and yearly patterns through ACF/PACF analysis.

Step 8

Assess Stationarity

Data approached non-stationarity (p-value ~0.5).

Step 9

Save Clean Data

Exported a clean CSV dataset for use in the dashboard.

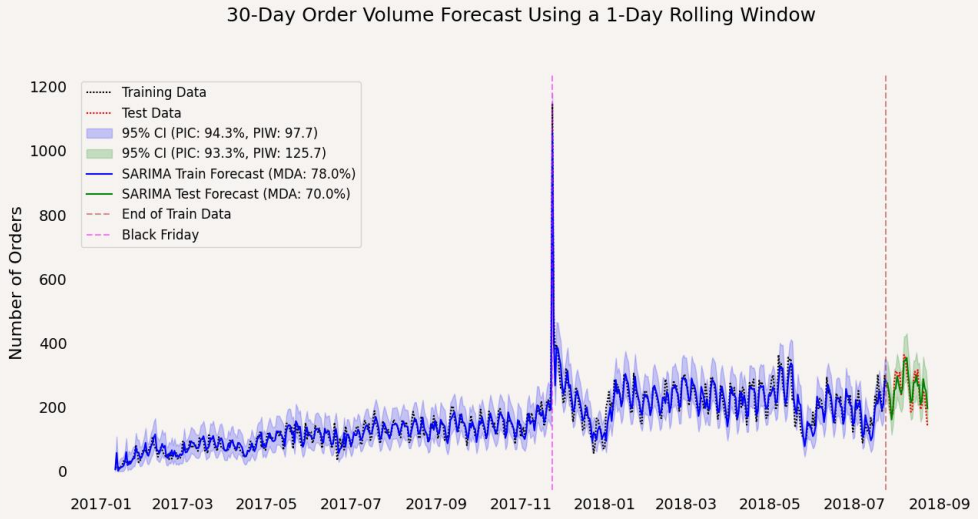
Data Preprocessing Result

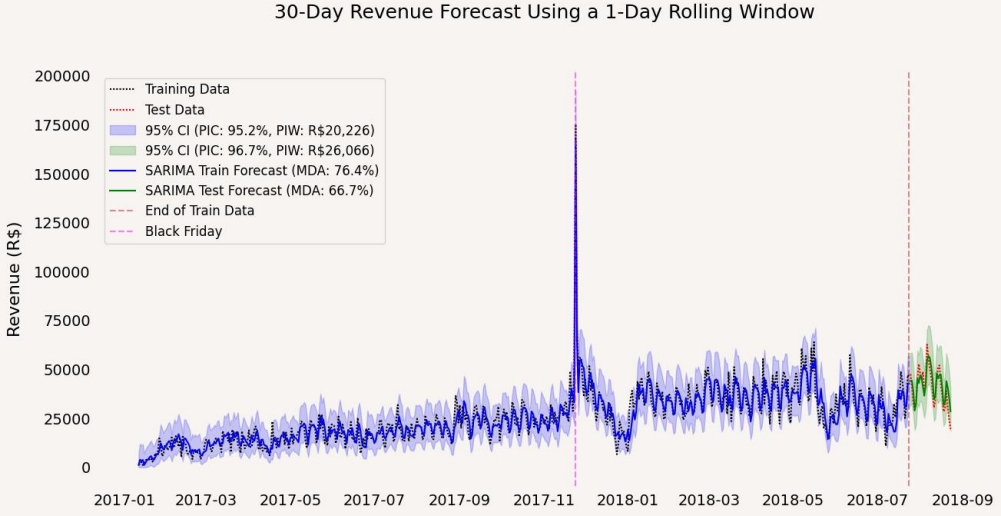
Findings and Results

Insight

- The actual and predicted daily orders and revenue exhibit a similar pattern throughout the year, with a notable peak around the Black Friday event (2017-11-24).
- The average prediction error (RMSE) is 30 for daily orders and R$5,500 for daily revenue.
- The prediction interval width (PIW) ranges approximately 126 orders (e.g., 100–226) and R$26,000 for daily revenue (e.g., R$10,000–36,000).
- The prediction coverage is 93% for orders and 97% for revenue, indicating that 93% of actual daily orders and 97% of actual daily revenue fall within their respective predicted ranges.
- 70% of trend predictions (up/down compared to the previous day) are correct for daily orders, while 67% of trend predictions are correct for daily revenue.

	Daily Order Volume	Daily Revenue
Avg. Prediction Error (RMSE)	30	R$5,458

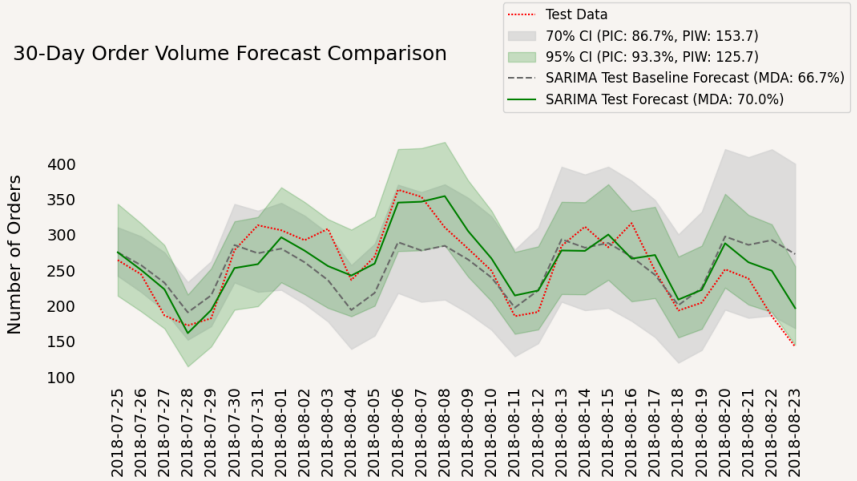


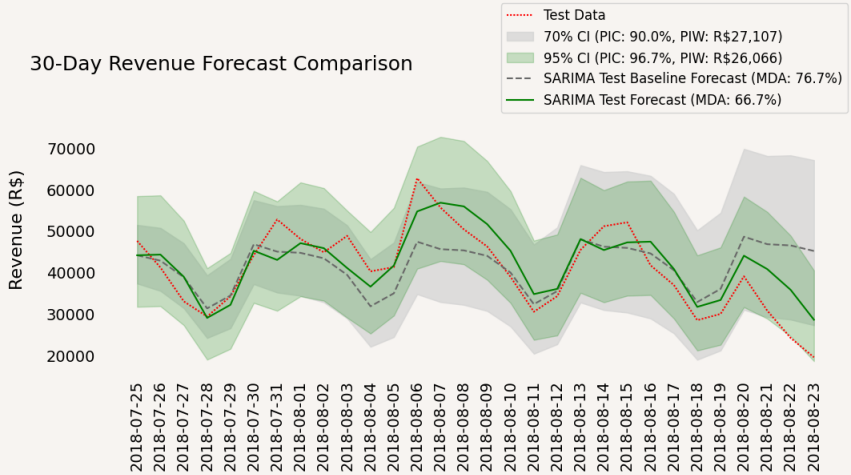


Findings and Results

Insight

- The baseline forecast provides a rough prediction for orders or revenue over the next 30 days, giving a broad overview of future trends. In contrast, the 1-day rolling window forecast leverages today's actual data to predict tomorrow's values, making it significantly more accurate.
- As a result, while the baseline forecast is less precise than the 1-day rolling window forecast, it serves as a useful tool for initial preparation and planning.

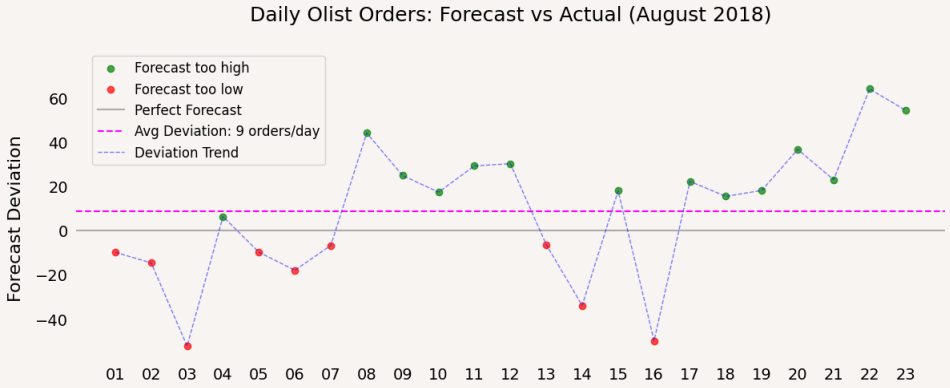


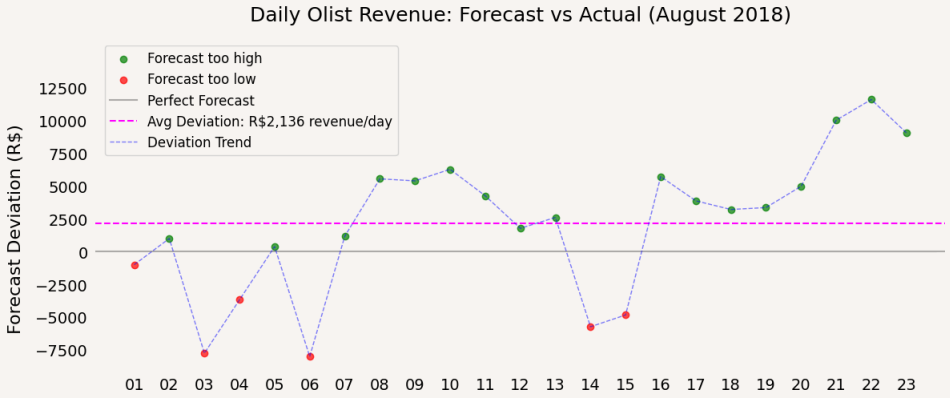


Findings and Results

Feature

- This monitoring dashboard helps identify whether the predictions are too high or too low compared to the actual data.
- The gray line represents the actual number of orders or revenue, serving as a benchmark for comparison.





Findings and Results

Daily Order Volume

Business Impact Analysis

- Average Daily Orders: 255
- Average Deviation: 30 orders (11.9% of average)

Error Cost Analysis (Local Currency R$)

- Under-prediction Cost: R$50 per order (lost profit)
- Over-prediction Cost: R$30 per order (excess inventory)
- Daily Average Cost: R$958
- Monthly Error Cost Impact: R$28,728

Error Cost Analysis (USD $)

- Under-prediction Cost: $9 per order
- Over-prediction Cost: $5 per order
- Daily Average Cost: $165
- Monthly Error Cost Impact: $4,945

Findings and Results

Daily Revenue

Business Impact Analysis

- Average Daily Revenue: R$40,787
- Average Deviation: R$5,458 revenue (13.4% of average) ($939 per day)

Error Cost Analysis (Local Currency R$)

- Under-prediction Cost: R$50 per R$100 error (lost profit)
- Over-prediction Cost: R$30 per R$100 error (excess inventory)
- Daily Average Cost: R$1,680
- Monthly Error Cost Impact: R$50,415

Error Cost Analysis (USD $)

- Under-prediction Cost: $9 per R$100 error
- Over-prediction Cost: $5 per R$100 error
- Daily Average Cost: $289
- Monthly Error Cost Impact: $8,677

Recommendations

- Use updated daily order forecasts to **adjust supply and inventory plans**, ensuring products are available without overstocking or understocking.
- **Base inventory and revenue strategies** on daily updates to forecasts, ensuring alignment with actual sales and preventing overproduction or lost opportunities.
- **Regularly monitor forecast accuracy** on the dashboard, adjusting orders and inventory levels when predictions are off to minimize excess or shortage costs.

Forecasting Dashboard

To monitor daily order volume and revenue trends, I have provided a dashboard for the relevant stakeholders to access.

Link to the Dashboard: Streamlit

📈 FutureSight Analytics

🏠 Home

📊 Dashboard

🛒 Order Volume

💰 Revenue Trend

📞 Contact

Made with ❤️ by Jihad Akbar

FutureSight: Revolutionizing E-Commerce with Data-Driven Forecasting

Welcome to FutureSight!

***Disclaimer**: The following story is fictional and created solely to illustrate the challenges addressed by this project.*

🎬 The Background: A Business on the Brink

Olist, one of Brazil's largest e-commerce platforms, faces critical challenges despite its growing sales:

⚠️ **Unpredictable Order Volumes**: Sudden fluctuations make it challenging to manage resources effectively.

⚠️ **Revenue Volatility**: Shifting revenue trends are hard to track, hindering strategic planning.

The CEO of Olist has issued a bold challenge:

FutureSight Analytics Dashboard

FutureSight Analytics

Home

Dashboard

Order Volume

Revenue Trend

Contact

Made with ❤ by Jihad Akbar

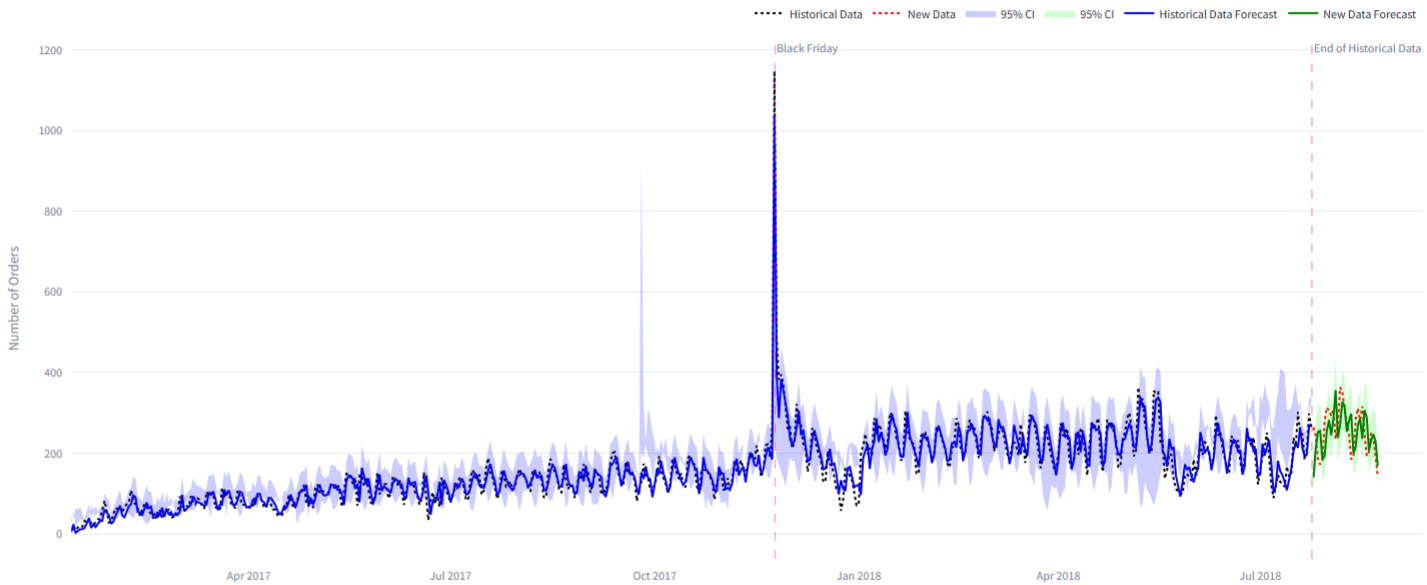
Next Day Orders

264

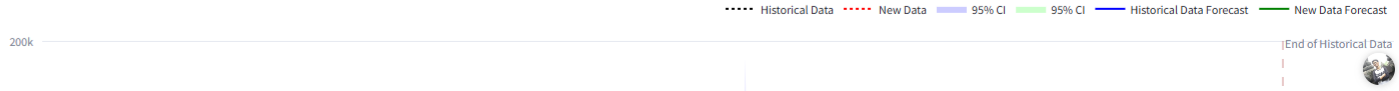
Next Day Revenue

$47,600

30-Day Order Volume Forecast Using a 1-Day Rolling Window



30-Day Revenue Forecast Using a 1-Day Rolling Window



FutureSight Analytics

Home
Dashboard
Order Volume
Revenue Trend
Contact

Made with ❤ by Jihad Akbar

Expected Order Volume Range

Next 30-day projections with possible variation

Date	Expected Orders	Worst Case	Best Case
2018-07-25	140	96	192
2018-07-26	211	157	272
2018-07-27	252	193	319
2018-07-28	255	196	322
2018-07-29	185	134	243
2018-07-30	192	141	251

At a Glance:

- Typical expected volume: **250 orders/day**
- Normal fluctuation range: **192–316**

Recommended Actions

Order Volume	Forecast Condition	Priority Action	Next Step
> 300	High Demand	Hire 20% more staff	Negotiate shipping contracts
100 - 200	Moderate Demand	Maintain current staffing	Optimize delivery routes
< 100	Low Demand	Freeze hiring	Audit retention metrics

Time Series Model Performance

Metric	Value	Interpretation
Avg. Prediction Error (RMSE)	30	On average, predictions are off by 30 orders/day.
Prediction Width (PIW)	126	Prediction ranges are 126 orders wide (e.g., 100–226).
Prediction Coverage (PIC)	93	93% of actual sales fall within the predicted range.
Trend Accuracy (MDA)	70	70% of trend predictions (up/down vs. yesterday) are correct.

Date	Expected Revenue	Worst Case	Best Case
2018-07-25	21,133	12,636	31,565
2018-07-26	25,490	16,132	36,752
2018-07-27	35,166	24,151	48,031
2018-07-28	45,923	33,341	60,309
2018-07-29	36,582	25,347	49,660
2018-07-30	34,686	23,746	47,476

At a Glance:

- Typical expected volume: **40,395 revenue/day**
- Normal fluctuation range: **28,665–53,955**

Recommended Actions

Revenue Range	Forecast Condition	Priority Action	Next Step
> R$50,000	High Revenue	Scale marketing campaigns (+20% budget)	Upsell premium products/services
R$30,000 - R$50,000	Moderate Revenue	Maintain current ad spend	Run targeted promotions
< $30,000	Low Revenue	Reduce non-essential costs	Analyze customer churn drivers

Time Series Model Performance

Metric	Value	Interpretation
Avg. Prediction Error (RMSE)	R$5,458	On average, predictions are off by R$5,458 revenue/day.
Prediction Width (PIW)	R$26,066	Prediction ranges are R$26,066 revenue wide (e.g., R$10,000–R$36,066).
Prediction Coverage (PIC)	97	97% of actual sales fall within the predicted range.
Trend Accuracy (MDA)	67	67% of trend predictions (up/down vs. yesterday) are correct.

📈 **FutureSight Analytics**

- 🏠 Home
- 📊 Dashboard
- 🛒 Order Volume
- 💰 Revenue Trend
- 📞 Contact

Contact Information

- Jihad Akbar
- Indonesia
- (+62) 8133 2326 785
- jihadakbr@gmail.com
- LinkedIn
- GitHub

Feel free to reach out!

04

Appendix

Link

Streamlit: time-series-forecasting-olist-e-commerce-jihadakbar.streamlit.app

Jupyter Notebook: G-Drive

THANK YOU

Do you have any questions?

Jihad Akbar

+62 813 3232 6785

https://www.linkedin.com/in/jihadakbr

jihadakbr@gmail.com