

**MACHINE LEARNING-BASED FORECASTING OF INCOMING CUSTOMER
ORDERS: A CASE STUDY ON HEAVY-DUTY FILTER PRODUCTS**

By
Taryono
22351042

MASTER'S DEGREE
in
INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

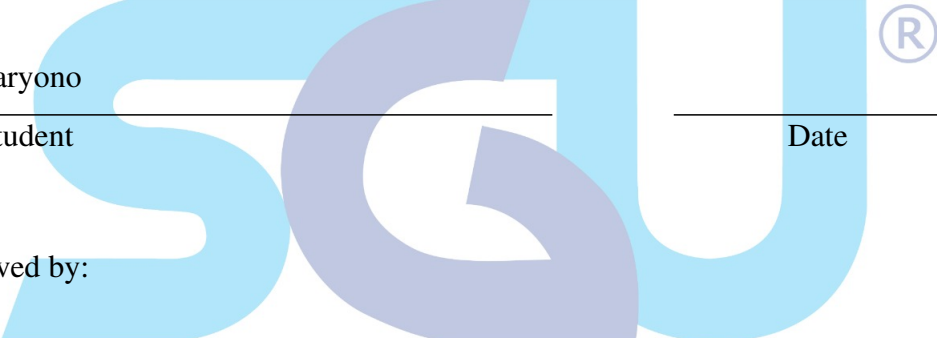
August 2025

STATEMENT BY THE AUTHOR

Name of Student : Taryono
Student ID : 22351042
Faculty : Engineering and Information Technology
Study Program : Information Technology
Date of Defense : 11-Aug-2025
Thesis Title : Machine Learning-Based Forecasting of Incoming Customer Orders:
A Case Study on Heavy-Duty Filter Products

I hereby declare that this submission is my own work and, to the best of my knowledge, contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgment is made in the thesis.

Taryono
Student _____ Date _____

Approved by: 

Dr. Tanika D. Sofianti ST MT

Thesis Advisor

Date

Dr. Ir. Eka Budiarto ST M.Sc

Thesis Co-Advisor

Date

Dr. Ir. Maulahikmah Galinium, S.Kom, M.Sc

Dean of Faculty of Engineering and
Information Technology

Date

ABSTRACT

MACHINE LEARNING-BASED FORECASTING OF INCOMING CUSTOMER ORDERS: A CASE STUDY ON HEAVY-DUTY FILTER PRODUCTS

By

Taryono

Dr. Tanika D. Sofianti ST MT, Advisor

Dr. Ir. Eka Budiarto ST M.Sc, Co-Advisor

SWISS GERMAN UNIVERSITY

Accurate demand forecasting plays a critical role in inventory management and production planning, particularly in the manufacturing of heavy-duty filter products, where customer order patterns are influenced by seasonal and external factors. This study proposes a machine learning-based forecasting framework to predict incoming customer orders using structured sales data.

The analysis focuses on the 2021–2025 period, during which demand behavior was observed to be more stable compared to previous years. The study evaluates the effects of feature engineering and feature selection on forecast accuracy, with special attention to time-based, seasonal, and external economic variables. Random Forest, LightGBM, and Linear Regression models were developed and their performance was assessed using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), MAPE, and R^2 score.

Experimental results show that the inclusion of seasonal and exogenous features significantly enhances model performance, particularly for item types with pronounced cyclical trends. Furthermore, the study conducts hypothesis testing to assess the statistical impact of engineered features and model types on forecast accuracy.

The findings demonstrate the applicability of machine learning in industrial forecasting scenarios, offering a replicable and scalable approach to data-driven decision making in

procurement and inventory planning.



Keywords: Demand Forecasting, Machine Learning, Feature Engineering, Feature Selection, Random Forest, LightGBM, Linear Regression, Heavy-Duty Filters, Forecast Accuracy



ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Dr. Tanika D. Sofianti ST MT and Dr. Ir. Eka Budiarto ST M.Sc as my thesis advisor for the time, support, advice and guidance given throughout this research project and the completion of this thesis report. It is because of their priceless contributions that this thesis report and the whole research project can reach this point.

I would like to thank my wife and daughter for the countless moral supports throughout my pursuit of knowledge in SGU. I cannot thank you enough for the time that has already been spent by you to write this thesis.

I would like to thank all my direct supervisors, my team at work, and my friends, especially Batch 33 MIT SGU, for their companionship, and the countless number of people who have helped me throughout this research project, either directly or indirectly.

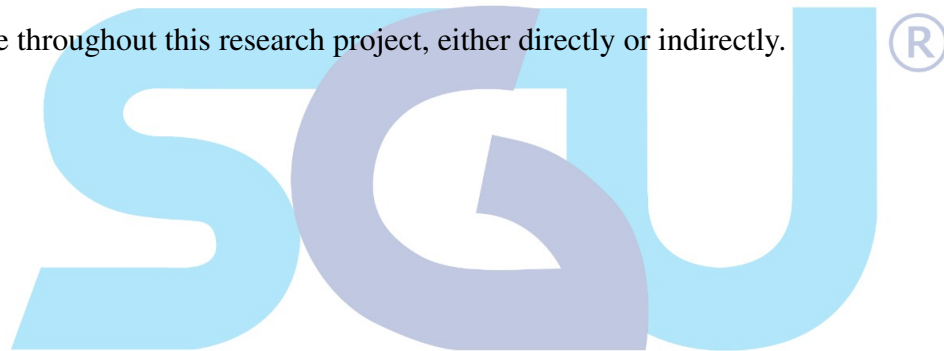


TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR	2
ABSTRACT	3
COPYRIGHT	5
ACKNOWLEDGEMENT	6
TABLE OF CONTENTS	7
LIST OF FIGURES	11
LIST OF TABLES	14
1. INTRODUCTION	16
1.1 Background	16
1.2 Research Problem	18
1.3 Research Objectives	19
1.4 Research Question	20
1.5 Hypothesis	20
1.6 Research Scope and Limitation	21
1.6.1 Scope of the research	21
1.6.2 Limitation of the Research	22
1.7 Significance of Study	23
1.8 Thesis Structure	25
2. LITERATURE REVIEW	27
2.1 Problem Description	27

2.2	Organizational and Problem Context	28
2.2.1	Functional Roles	30
2.2.2	Operational Challenges Across Functional Areas	31
2.2.3	Relationship to Forecasting Challenges	32
2.3	Characteristics of Inventory and Forecasting in Manufacturing Industries: Case of PT. Heavy Duty Vehicle Filter Manufacture	33
2.4	Feature Engineering and Feature Selection in Demand Forecasting	36
2.5	Conceptual Description	38
2.5.1	Machine Learning for Demand Forecasting	38
2.5.2	Demand Seasonality	38
2.5.3	Feature Engineering	38
2.5.4	Feature Selection	39
2.5.5	Forecast Evaluation	39
2.6	Forecasting Methods and Frameworks Used in Previous Studies	40
2.6.1	Research Gap	41
3.	RESEARCH METHODS	43
3.1	Framework Used Description	43
3.1.1	Problem Identification and Data Collection	44
3.1.2	Exploratory Data Analysis (EDA) and Demand Pattern Characterization	44
3.1.3	Model Development Framework (2021–2025)	44
3.1.3.1	Baseline Modeling	45
3.1.3.2	Naïve Baseline Model Parameters	45
3.1.3.3	Moving Average (MA) Baseline Parameters	45
3.1.3.4	Feature Engineering	46
3.1.3.5	Initial Model Training (All Features)	46
3.1.3.6	Feature Selection as the Feature Reduction Method	49
3.1.3.7	Comparison with Other Feature Selection Methods	52
3.1.3.8	Selected Model Training	54
3.1.4	Forecast Evaluation	54
3.1.5	Insight Generation and Success Criteria	54
4.	RESULTS AND DISCUSSION	56

4.1	Exploratory Data Analysis (EDA) Results	56
4.1.1	Demand Trends by Year	56
4.1.2	Demand Distribution by Item Type	57
4.1.3	Monthly and Yearly Seasonality Patterns	58
4.1.4	Macro-Economic Context: IHSG and Currency Trends	58
4.2	Overview of the Experiment Process	59
4.2.1	Summary of Dataset Used (2021–2025)	59
4.2.2	Reiteration of Modeling Phases: Baseline, Feature Engineering, and Feature Selection	61
4.3	Baseline Modeling	61
4.3.1	Performance Metrics Naïve Forecasting Model	62
4.3.2	Performance Metrics Moving Average Model	66
4.3.3	Baseline Selection for Model Evaluation	70
4.4	Baseline Model Performance Using Raw Features	71
4.5	Feature Engineering	72
4.5.1	Feature Engineering Framework	73
4.5.2	Baseline Selection after Feature Engineering	74
4.6	Feature Importance Analysis	75
4.6.1	Feature Importance in Linear Regression with Feature Engineering . . .	75
4.6.2	Feature Importance in Random Forest with Feature Engineering	77
4.7	Feature Selection and Reduction (FS)	79
4.7.1	Feature Importance Rankings of Random Forest	80
4.7.2	Feature Importance Rankings of Linear Regression	83
4.7.3	Results of Feature Removal Simulation	87
4.8	Selected Best Model Performance	89
4.8.1	Best Model per Item Type	90
4.9	Cross Model Comparison	90
4.9.1	Visual Forecast Comparison	92
4.10	Insight Generation	95
4.10.1	Top Contributing Factors	96
4.10.2	Forecast Behavior Insight	96
4.11	Limitations and Future Opportunities	97

4.11.1	Limitations of the Current Modeling	97
4.11.2	Suggestions for Improving External Data Integration	98
4.11.3	Potential for Multistep and Real-Time Forecasting	99
5.	CONCLUSIONS AND RECOMMENDATIONS	100
5.1	Conclusions	100
5.2	Recommendations	101
5.3	Implications	102
5.3.1	Academic Implications	102
5.3.2	Methodological Implications	102
5.3.3	Managerial / Practical Implications	103
5.3.4	Implications of Feature Importance for Business Understanding	103
5.4	Roadmap for Implementation and Managerial Implications	104
5.4.1	Implementation Roadmap	105
5.4.2	Feature Insights to Business Actions	107
5.4.3	Managerial Implications	108
5.4.4	Consulting and Scalability Opportunities	108
REFERENCES		119

LIST OF FIGURES

1.1	Fishbone Diagram – Root Causes of Demand Forecast Absence / Planning Disruption	17
2.1	Company Organizational Structure	29
2.2	Pain-Point Prioritization Matrix	31
3.1	Overall forecasting framework combining exploratory analysis (2018–2025), model development (2021–2025), and evaluation stages to predict incoming customer orders at SKU level.	43
3.2	Feature Reduction Method (Feature Take-Out Simulation) Flowchart	50
4.1	Monthly demand quantity trend across all SKUs (2021–2025).	56
4.2	Demand quantity distribution by item type (Top 4 + Others).	57
4.3	Monthly demand quantity with seasonal calendar flags (2021–2025).	58
4.4	Monthly demand quantity vs. USD exchange rate and IHSG index.	58
4.5	Bar chart of item type distribution within the dataset. Major product types (OF, FF, HF, SX) dominate the transaction records.	61
4.6	Actual demand for Item Type FF shows sharp fluctuations, while the naïve baseline lags behind and fails to capture turning points. This demonstrates the inability of the naïve method to represent volatile demand patterns.	62
4.7	Actual demand for Item Type HF exhibits sharp peaks and troughs, while the naïve baseline produces a smoother, lagging response. This lag causes the baseline to miss sudden turning points and underestimate demand volatility. . .	63
4.8	Actual and naïve baseline demand for Item Type OF both show a general downward trend, but the baseline tends to overestimate during declines and lags at turning points. While closer to actual demand than in other item types, the naïve method still fails to capture sharper shifts in the trend.	64

4.9	Actual demand for Item Type SX fluctuates sharply, while the naïve baseline provides a lagging and smoother approximation. This causes systematic underestimation during upswings and overestimation during downturns, showing the method's weakness in volatile demand conditions.	65
4.10	Actual demand for Item Type FF shows sharp peaks and troughs, while the moving average baseline smooths these fluctuations into a gradual trend. This results in underestimation during demand surges and overestimation during steep declines, limiting its usefulness in volatile demand conditions.	66
4.11	Actual demand for Item Type HF shows sharp volatility, while the moving average baseline smooths fluctuations into a gradual trend. This smoothing effect causes underestimation during spikes and overestimation during drops, limiting its suitability for volatile demand conditions.	67
4.12	Actual and moving average demand for Item Type OF both show a downward trend, but the MA baseline smooths sharper fluctuations into a gradual decline. This leads to overestimation during steep drops and a lagging response during partial recoveries.	68
4.13	Actual demand for Item Type SX is highly volatile, while the moving average baseline smooths fluctuations into a gradual line. This results in underestimation during peaks and overestimation during troughs, showing that the MA approach cannot capture irregular demand patterns.	69
4.14	Feature importance via Δ MAPE for Fuel Filter (FF) using Random Forest with feature engineering. Temporal features (Lag and MA) dominate the model, while macroeconomic and holiday variables show limited or negative contribution.	80
4.15	Feature importance via Δ MAPE for Hydraulic Filter (HF) using Random Forest with feature engineering. Moving Average and Lag are dominant features, while holiday and macroeconomic variables provide secondary contributions. .	81
4.16	Feature importance via Δ MAPE for Oil Filter (OF) using Random Forest with feature engineering. Temporal features (Lag and MA) dominate the model, while calendar and macroeconomic variables provide little or negative contribution.	82

4.17	Feature importance via Δ MAPE for Special Design Filter (SX) using Random Forest with feature engineering. Temporal features (MA and Lag) dominate, with Imlek providing some seasonal contribution, while fiscal, monthly, and currency features reduce accuracy.	83
4.18	Feature importance rankings for Fuel Filter (FF) using Linear Regression. Lag and Moving Average are the most influential predictors, while macroeconomic and calendar features contribute weakly.	84
4.19	Feature importance rankings for Hydraulic Filter (HF) using Linear Regression with feature engineering. Moving Average and Lag dominate the model, while calendar and macroeconomic variables show limited or negative contributions. .	85
4.20	Feature importance rankings for Oil Filter (OF) using Linear Regression with feature engineering. Moving Average and Lag dominate the model, while Imlek and Month also emerge as meaningful contributors.	86
4.21	Feature importance rankings for Special Design Filter (SX) using Linear Regression with feature engineering. Moving Average is the only dominant predictor, while most other features contribute little or negatively.	87
4.22	Actual demand for Item Type FF fluctuates sharply, while the baseline shows only a gradual rise and fall. The best model follows the actual trend more closely, capturing peaks and declines with greater responsiveness than the baseline.	92
4.23	Actual demand for Item Type HF is highly volatile, with sharp peaks and troughs. The best model follows these turning points more closely than the baseline, which smooths the series and fails to capture sudden shifts.	93
4.24	Actual demand for Item Type OF shows a general downward trend with mild fluctuations. The best model follows this trend more closely than the baseline, which smooths the series and tends to overestimate during sharper declines. . .	94
4.25	Actual demand for Item Type SX is highly volatile with sharp peaks and troughs. The best model tracks these fluctuations more closely than the baseline, which smooths the series and misses extreme variations.	95
5.1	Roadmap Machine Learning-Based Forecasting Implementation.	105

LIST OF TABLES

1.1	Summary of Research Contributions	24
2.1	Key Operational Challenges Across Functional Areas	31
2.2	Demand Forecasting Characteristics at PT. Heavy Duty Vehicle Filter Manufacturer	35
2.3	Summary of Feature Engineering and Selection in Forecasting Customer Orders	37
2.4	Comparison of Machine Learning-Based Forecasting Studies, Miguel Rodrigues (2024) and Feizabadi (2022).	41
3.1	Random Forest Parameters Used in This Study	47
3.2	LightGBM Parameters Used in This Study	48
3.3	Linear Regression Parameters Used in This Study	48
3.4	Comparison of Feature Selection Methods	53
4.1	Baseline Forecasting Results by Item Type	70
4.2	Performance of ML Models Using Raw Features (2021–2025)	71
4.3	Comparison of Model Performance: Raw Features vs Feature Engineering (FE). Bold values indicate the best-performing model per item type.	74
4.4	Feature Importance (Ablation) Analysis for Linear Regression with Feature Engineering	76
4.5	Feature Importance via Δ MAPE (Linear Regression + FE)	76
4.6	Feature Importance (Ablation) Analysis for Random Forest with Feature Engineering	78
4.7	Feature Importance via Δ MAPE (Random Forest + FE)	78
4.8	Results of Feature Removal Simulation: Best Models and Feature Sets per Item Type	88
4.9	Selected Best Model Performance per Item Type	89
4.10	Cross Model Comparison: Baseline vs. Selected Best Model (MAPE and R^2) .	91

5.1	Examples of SLA Targets for Data Governance	106
5.2	Translating Feature Importance into Business Actions	107

