



UNDERGRADUATE THESIS

FORECASTING INFLATION IN INDONESIA USING ARIMAX-LIGHTGBM MODEL WITH OPTUNA HYPERPARAMETER OPTIMIZATION BASED ON MACROECONOMIC INDICATORS

DIAN MAHARANI
NPM 22081010184

THESIS ADVISORS

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
Muhammad Muharrom Al Haromainy, S.Kom., M.Kom.

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026**

APPROVAL SHEET

FORECASTING INFLATION IN INDONESIA USING ARIMAX- LIGHTGBM MODEL WITH OPTUNA HYPERPARAMETER OPTIMIZATION BASED ON MACROECONOMIC INDICATORS

By:
DIAN MAHARANI
NPM. 22081010184

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on June 11, 2026:

Approved,

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400


.....

(Advisor I)

M. Muharrom Al Haromainy, S.Kom., M.Kom.
NIP. 19950601 202203 1 006


.....

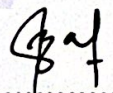
(Advisor II)

Fetty Tri Anggraeny, S.Kom. M.Kom.
NIP. 19820211 202121 2 005


.....

(Head Assessor)

Made Hanindia Prami Swari, S.Kom, M.Cs.
NIP. 19890205 201803 2 001


.....

(Assessor)

Acknowledged by,

Dean of the Faculty of Computer Science



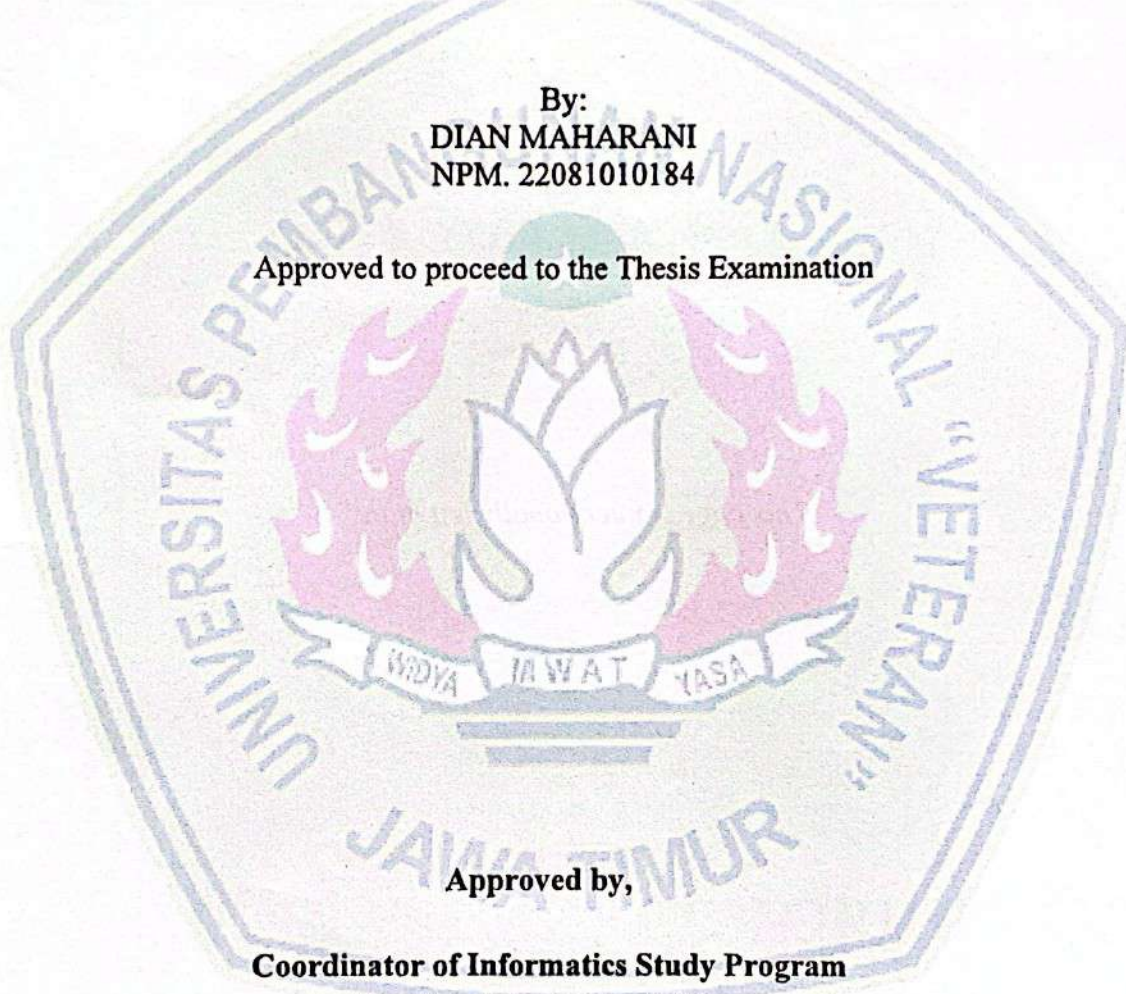
Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

APPROVAL SHEET

FORECASTING INFLATION IN INDONESIA USING THE ARIMAX- LIGHTGBM MODEL WITH OPTUNA HYPERPARAMETER OPTIMIZATION BASED ON MACROECONOMIC INDICATORS

By:
DIAN MAHARANI
NPM. 22081010184

Approved to proceed to the Thesis Examination



Approved by,

Coordinator of Informatics Study Program
Faculty of Computer Science

Dr. Intan Yuniar Purbasari, S.Kom. MSc.

NIP. 19800602 202521 2 029

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Dian Maharani
NPM : 22081010184
Degree Program : Bachelor (S1)
Study Program : Informatics
Faculty : Faculty of Computer Science

Hereby affirm that this undergraduate thesis is an original scholarly work and does not incorporate, either wholly or partially, any academic manuscript that has previously been submitted in fulfillment of the requirements for an academic degree at any institution of higher education. Furthermore, this thesis contains no ideas, arguments, findings, or intellectual contributions originating from other authors.

And I declare that this scientific manuscript is free from any vestige of plagiarism. Should any indication thereof be established in the future, I willingly submit to the sanctions imposed under the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.



Surabaya, June 29, 2026
Declarant,



DIAN MAHARANI
NPM. 22081010184

ABSTRACT

Student Name / NPM : Dian Maharani/22081010184
Thesis Title : Forecasting Inflation in Indonesia Using ARIMAX-LightGBM Model with Optuna Hyperparameter Optimization Based on Macroeconomic Indicators
Supervisor : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. M. Muharrom Al Haromainy, S.Kom., M.Kom.

Inflation is one of the key indicators reflecting a country's economic stability; therefore, inflation forecasting is crucial as a basis for monetary policy-making and economic planning. However, the characteristics of inflation data, which contain linear and nonlinear patterns, mean that a single method is often unable to produce optimal forecasts. This study aims to develop an inflation forecasting model for Indonesia using a hybrid ARIMAX-LightGBM approach with Optuna hyperparameter optimization based on macroeconomic indicators. The exogenous variables used include money supply (M1 and M2), exchange rate, interest rates, inflation expectations, administered prices, and volatile food prices.

The research stages include data cleaning, exploratory data analysis, data transformation, stationarity tests, ARIMAX model development, residual modeling using LightGBM based on lag features and statistics, hyperparameter optimization using Optuna, and evaluation using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), and coefficient of determination (R^2).

The results show that the hybrid ARIMAX-LightGBM model with Optuna optimization delivers the best performance compared to the ARIMAX model, the LightGBM model, and the hybrid model without optimization. On the test data, the model produced an RMSE of 0.055973, an MAE of 0.042383, an MBE of 0.008892, and an R^2 of 0.989878. Furthermore, optimization using Optuna successfully improved the hybrid model's performance by reducing the RMSE and MAE values compared to the model without optimization. Model analysis shows that information derived from the Administered Price and Volatile Food components contributed more significantly to improving inflation prediction capabilities than other macroeconomic variables during the study period.

Keywords: Indonesian inflation, ARIMAX, LightGBM, Optuna, time series forecasting, macroeconomic indicators, hybrid models.

ACKNOWLEDGEMENTS

Profound praise and gratitude are ascribed to Allah SWT, whose immeasurable mercy, benevolence, and divine favor have sustained the author throughout the completion of this undergraduate thesis titled “Forecasting Inflation in Indonesia Using the ARIMAX-LightGBM Model with Optuna Hyperparameter Optimization Based on Macroeconomic Indicators”. The completion of this thesis would not have been possible without the help, support, guidance, and prayers from many people. Henceforth, the author takes this opportunity to convey her profound appreciation to those whose support and encouragement have been invaluable:

1. Mrs. Prof. Dr. Ir. Novirina Hendrasarie, M.T., Dean of the Faculty of Computer Science at UPN “Veteran” Jawa Timur, for providing facilities and support throughout the author’s studies.
2. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., Coordinator of the Informatics Study Program, for the guidance and support throughout the author’s studies.
3. Dr. Eng. Ir. Anggraini Puspita Sari, S.T., M.T., as the First Advisor, who patiently devoted time, energy, and thought to providing guidance, direction, feedback, and motivation to the author throughout the process of writing this thesis.
4. M. Muharrom Al Haromainy, S.Kom., M.Kom., as the Second Advisor, who provided guidance, support, and constructive advice, enabling this thesis to be successfully completed.
5. All lecturers of the Informatics Study Program at UPN “Veteran” Jawa Timur, who have shared their knowledge, experience, and invaluable lessons throughout my academic studies, which have served as a foundation for the author in completing this thesis.
6. The author’s beloved parents, Father and Mother, as well as the entire family, who have always been a source of strength for the author. Thank you for never-ending prayers, sincere love, unconditional support, and the sacrifices that have enabled the author to reach this stage.
7. Dhenis Lailathul Syiva, the author’s best friend and closest companion who has always been there through every phase of college life. Thank you for being

someone with the author could share stories, complaints, joys, and worries. Thank you for the support, care, and companionship provided throughout this journey, which was not always easy.

8. All members of “Makasih Ibu Diah”: Naila Jinan Gaisani, Dhea Intan Sagita, I Gusti Ngurah Karunya Pratama, Radendha Muhammad Arthansa, Muhammad Adeva, and Muizzul Rizqon Gunawan. Thank you for being there to share stories, exchange ideas, support one another, and grow together. The presence of each member made this journey feel lighter, more enjoyable, and full of meaning. The various moments experienced together are precious memories for which the author will always be grateful.
9. All members of “D for Dudidudidamdamm”: Salsa Pramudhita Agustiardani, Caritta Elizabeth, and Dwitamara Septyana. Thank you for the togetherness, laughter, support, and shared experiences throughout our college years. The support of all members has enriched the author’s life and will remain a cherished part of the author’s memories.
10. To everyone who has offered support, assistance, or prayers throughout the completion of this thesis, whether directly or indirectly.

Although this thesis has several limitations, the author welcomes constructive suggestions for future improvement. It is expected that this study will contribute to future research in computer science and economic forecasting.

Surabaya, June 29, 2026

Author

TABLE OF CONTENTS

APPROVAL SHEET	ii
APPROVAL SHEET	iii
STATEMENT OF ORIGINALITY	iv
ABSTRACT	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	viii
LIST OF FIGURES	xii
LIST OF TABLES	xv
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Research Questions	4
1.3 Research Objectives	4
1.4 Research Benefits	4
1.5 Research Limitations	5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Related Studies	6
2.2 Inflation	8
2.3 Time Series	11
2.4 ARIMAX	12
2.4.1 ARIMAX Components	14
2.4.2 Augmented Dickey Fuller	16
2.4.3 ACF and PACF	17
2.4.4 AIC and BIC	19
2.4.5 Cross-Correlation Function	20

2.4.6	β Coefficient Interpretation	21
2.4.7	Significance Test	22
2.5	LightGBM	23
2.5.1	LightGBM Architecture	24
2.5.2	LightGBM for Time Series	26
2.6	Hyperparameter Optimization.....	28
2.7	Hybrid Model	30
2.8	Model Evaluation	32
CHAPTER 3	METHODOLOGY.....	34
3.1	Research Stages.....	34
3.2	Data Collection.....	35
3.3	Exploratory Data Analysis (EDA)	37
3.3.1	Basic Data Characteristics Analysis.....	38
3.3.2	Variable Correlation Analysis	38
3.4	Preprocessing Data.....	39
3.4.1	Parsing and Time Series Data Sorting.....	39
3.4.2	Missing Value Checking	40
3.4.3	Data Transformation	40
3.4.4	Lag Feature Formation	42
3.4.5	Splitting Data	43
3.5	ARIMAX Modeling	43
3.5.1	Stationarity Test and <i>Differencing</i> (d)	44
3.5.2	ARIMA Model Order Identification (p and q).....	45
3.5.3	Cross-Correlation Function Analysis	46
3.5.4	Determine the ARIMAX Model Structure	46
3.5.5	Best ARIMAX Model Selection	47

3.5.6	Prediction Using ARIMAX.....	47
3.6	Modeling LightGBM on Residuals.....	50
3.6.1	Creating Lag Features as LightGBM Inputs	50
3.6.2	Statistical Feature Creation	51
3.6.3	Prediction Using LightGBM.....	51
3.7	Hyperparameter Optimizing.....	53
3.8	ARIMAX-LightGBM Prediction.....	56
3.9	Model Evaluation	57
3.10	Analysis of Variable Influence on Forecasting Results.....	59
3.11	Testing Scenario	60
3.12	Wireframe Design	61
CHAPTER 4	RESULT AND DISCUSSION	63
4.1	Data Cleaning.....	63
4.2	Explanatory Data Analysis (EDA).....	66
4.2.1	Basic Data Characteristics.....	66
4.2.2	Relationships Between Variables	70
4.3	Data Preprocessing.....	74
4.3.1	Data Stationarity Test	75
4.3.2	Data Transformation	76
4.3.3	Lag Formation.....	78
4.3.4	Data Splitting	79
4.4	ARIMAX-LightGBM Modeling.....	81
4.4.1	ARIMA Order Identification.....	81
4.4.2	Cross-Correlation Function Analysis	83
4.4.3	ARIMAX Model Structure Identification.....	85
4.4.4	Selection of the Best ARIMAX Model	86

4.4.5	ARIMAX Residual Testing.....	88
4.4.6	Feature Engineering as Input LightGBM.....	91
4.4.7	Hyperparameter Optimization.....	92
4.4.8	ARIMAX-LightGBM Predictions	93
4.4.9	Model Evaluation.....	94
4.5	Test Results	95
4.5.1	Testing on Data Split.....	95
4.5.2	Testing on Combinations of Orders p and q.....	102
4.5.3	Testing on Residual Lag.....	110
4.5.4	Learning Rate Testing	119
4.5.5	Best Test Scenario Results	127
4.6	Model Evaluation.....	130
4.6.1	Models ARIMAX.....	131
4.6.2	Model LightGBM	132
4.6.3	Model ARIMAX-LightGBM	135
4.6.4	Model Evaluation Results	137
4.7	Analysis of the Influence of Variables on Prediction Results	138
4.7.1	Model Analysis Using Significant Variables.....	142
4.7.2	Model Analysis Using Macroeconomic Variables	144
4.8	Development of the Inflation Prediction Dashboard	147
CHAPTER 5	CONCLUSION	150
5.1	Conclusion	150
5.2	Suggestions	151
BIBLIOGRAPHY		152

LIST OF FIGURES

Figure 2.1 ARIMAX Model Flowchart	13
Figure 2.2 ACF plot of the data before and after differencing	18
Figure 2.3 PACF plot of the data before and after differencing	18
Figure 2.4 Differences Between Leaf-Wise and Level-Wise	24
Figure 2.5 LightGBM Architecture Diagram	26
Figure 2.6 Optuna Architecture	29
Figure 3.1 Research Workflow.....	34
Figure 3.2 ARIMAX Modeling Flowchart.....	44
Figure 3.3 LightGBM Modeling Workflow	54
Figure 3.4 System Wireframe	62
Figure 4.1 Number of duplicate data points.....	64
Figure 4.2 Results of the missing value check.....	64
Figure 4.3 Results of time series data format adjustments.....	65
Figure 4.4 Inflation Trend Graph	67
Figure 4.5 Inflation volatility chart	68
Figure 4.6 Graph of extreme inflation shocks.....	69
Figure 4.7 Heatmap (correlation matrix) between variables.....	71
Figure 4.8 Scatter Plot.....	73
Figure 4.9 Data transformation results.....	77
Figure 4.10 Data distribution results.....	80
Figure 4.11 ACF and PACF Plots	82
Figure 4.12 CCF plots for each variable	84
Figure 4.13 Residual distribution in the ARIMAX training data.....	89
Figure 4.14 Results of the Ljung-Box test on the residuals	90

Figure 4.15 70:20:10 split data test results.....	96
Figure 4.16 Comparison chart of split data test results 70:20:10.....	97
Figure 4.17 80:20:10 split data test results.....	98
Figure 4.18 Comparison chart of split data test results 80:20:10.....	99
Figure 4.19 90:5:5 split data test results	100
Figure 4.20 Comparison chart of split data test results 90:5:5.....	101
Figure 4.21 Results of testing order $p=0$ and $q=2$	103
Figure 4.22 Comparison chart of test results of order $p=0$ and $q=2$	104
Figure 4.23 Results of the $p=2$ and $q=1$ order tests	105
Figure 4.24 Comparison graph of test results for $p=2$ and $q=1$	106
Figure 4.25 Results of the $p=2$ and $q=2$ order tests	107
Figure 4.26 Results of comparison of test results of order $p=2$ and $q=2$	108
Figure 4.27 Test results Lag Residual 1, 2, 3	111
Figure 4.28 Comparison chart of test results Lag Residual 1, 2, 3	112
Figure 4.29 Test results Lag Residual 1, 2, 3, 4, 5	113
Figure 4.30 Comparison chart of test results Lag Residual 1, 2, 3, 4, 5	114
Figure 4.31 Test results Lag Residual 1, 2, 3, 4, 5, 6, 12	116
Figure 4.32 Comparative testing graph Lag Residual 1, 2, 3, 4, 5, 6, 12.....	117
Figure 4.33 The test results of the learning rate are 0.001-0.003.....	120
Figure 4.34 Comparison chart of learning rate test results 0.001-0.003	121
Figure 4.35 The test results of the learning rate are 0.003-0.007.....	122
Figure 4.36 Comparison chart of learning rate test results 0.003-0.007	123
Figure 4.37 Learning rate test results 0.007-0.01.....	124
Figure 4.38 Comparison chart of learning rate test results 0.007-0.01	125
Figure 4.39 Best test evaluation results.....	128
Figure 4.40 Predicted results on test data.....	129

Figure 4.41 Results of the evaluation of the ARIMAX model	131
Figure 4.42 Comparison chart of the results of the ARIMAX model.....	132
Figure 4.43 Results of the evaluation of the LightGBM model.....	133
Figure 4.44 LightGBM model results comparison chart	134
Figure 4.45 ARIMAX-LightGBM evaluation results	135
Figure 4.46 Comparison chart of the results of the ARIMAX-LightGBM model	136
Figure 4.47 Results of model evaluation with significant variables	143
Figure 4.48 Comparison graph of models with significant variables	143
Figure 4.49 Results of model evaluation with macroeconomic variables	146
Figure 4.50 Comparison graph of the model with macroeconomic variables	146
Figure 4.51 Model implementation view	148

LIST OF TABLES

Table 3.1 Dataset Used.....	37
Table 3.2 Example data without missing values in the dataset	40
Table 3.3 Example of an index difference transformation	41
Table 3.4 Example of transforming raw data into log-differencing	41
Table 3.5 Example of lag generation.....	42
Table 3.6 Example of lag formation after removing null values.....	43
Table 3.7 Example of data split.....	43
Table 3.8 Sample data for manual calculations.....	44
Table 3.9 ARIMAX Forecast Results and Residuals	50
Table 3.10 Example of lag feature creation in LightGBM.....	50
Table 3.11 Example of a combination of lag and rolling mean features.....	51
Table 3.12 Example of a summary of LightGBM prediction results	53
Table 3.13 Hyperparameters for optimization using Optuna	55
Table 3.14 Comparison of each model's predictions with actual inflation	57
Table 3.15 Test Scenarios.....	61
Table 4.1 Comparison of the percentage contributions of the linear and nonlinear components of the variables	74
Table 4.2 ADF test results before transformation for each variable	75
Table 4.3 Results of the ADF test after transformation and differencing	77
Table 4.4 Results of the CCF test for each variable from lags 0-6	85
Table 4.5 ARIMAX model selection results using AIC and BIC	87
Table 4.6 <i>Hyperparameter</i> Best on 70:20:10 data split testing.....	95
Table 4.7 Hyperparameter Best on 80:20:10 split data testing	98
Table 4.8 Hyperparameter Best at 90:5:5 split data testing	100
Table 4.9 Results of testing data split scenarios.....	101

Table 4.10 Hyperparameter best at testing order $p=0$ and $q=2$	103
Table 4.11 Hyperparameter best in testing order $p=2$ and $q=1$	105
Table 4.12 Hyperparameter best at testing order $p=2$ and $q=2$	107
Table 4.13 Performance comparison results in the p and q Order scenarios	109
Table 4.14 Hyperparameter Best on testing Lag Residual 1, 2, 3	111
Table 4.15 Hyperparameter Best on testing Lag Residual 1, 2, 3, 4, 5	113
Table 4.16 Hyperparameter Best Testing Lag Residual 1, 2, 3, 4, 5, 6, 12	116
Table 4.17 Comparison of test results <i>Lag Residual</i>	118
Table 4.18 <i>Hyperparameter</i> Best on the learning rate test 0.001-0.003	120
Table 4.19 <i>Hyperparameter</i> Best on the learning rate test 0.003-0.007	122
Table 4.20 <i>Hyperparameter</i> best on the learning rate test 0.007-0.01	124
Table 4.21 Comparison of test results <i>Learning Rate</i>	126
Table 4.22 Results of the best test scenarios	127
Table 4.23 Comparison of the results of the overall evaluation of the model	137
Table 4.24 The results of variable influence using coefficients and p-values.....	139
Table 4.25 Comparison of model evaluation results with significant variables and all variables	143
Table 4.26 The results of the influence of macroeconomic variables using coefficients and p-values	145
Table 4.27 Comparison of the results of the evaluation of all models with the macroeconomic variable model	147