



UNDEGRUATE THESIS

**GOLD PRICE PREDICTION ANALYSIS
USING SUPPORT VECTOR REGRESSION
AND RANDOM FOREST ALGORITHMS
BASED ON BPS DATA**

Hanin Fatma Soraya
NPM 22081010069

THESIS ADVISOR

Dr. Budi Nugroho, S.Kom., M.Kom
Dr. Firza Prima Aditiawan, S.Kom., M. TI.

**MINISTRY OF EDUCATION, CULTURE, RESEARCH, AND
TECHNOLOGY VETERAN NATIONAL DEVELOPMENT
UNIVERSITY OF EAST JAVA FACULTY OF COMPUTER
SCIENCE INFORMATICS STUDY PROGRAM
SURABAYA
2026**

APPROVAL SHEET

GOLD PRICE PREDICTION ANALYSIS USING SUPPORT VECTOR REGRESSION AND RANDOM FOREST ALGORITHMS BASED ON BPS DATA

By :
HANIN FATMA SORAYA
NPM. 22081010069

Has been defended before and accepted by the Thesis Examining Team of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on June 17, 2026 :

Approved,

Dr. Budi Nugroho, S.Kom., M.Kom.
NIP. 198009072021211005



(Advisor I)

Dr. Firza Prima Aditiawan, S.Kom., MTI.
NIP. 19860523 2021211 003



(Advisor II)

Dr. Rizky Parlika, S.Kom, M.Kom.
NIP. 19840518 2021211 003



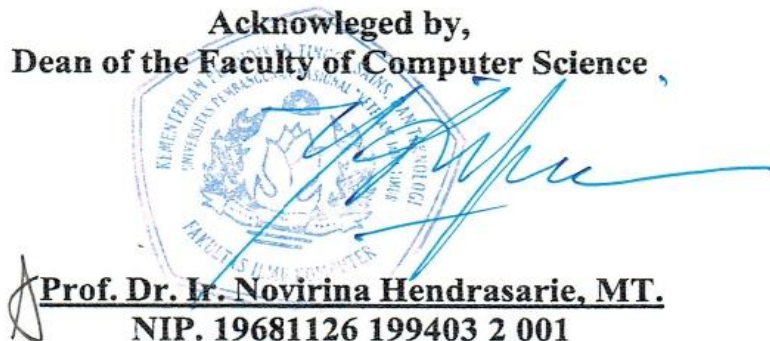
(Head Assessor)

Achmad Junaidi, S.Kom, M.Kom.
NIP. 197811102025211048



(Assessor I)

Acknowledged by,
Dean of the Faculty of Computer Science ,



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

APPROVAL SHEET

GOLD PRICE PREDICTION ANALYSIS USING SUPPORT VECTOR REGRESSION AND RANDOM FOREST ALGORITHMS BASED ON BPS DATA

By :

HANIN FATMA SORAYA

NPM. 22081010069

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 : Hanin Fatma Soraya
NPM : 22081010069
Degree Program : Bachelor (S1)
Study Program : Informatics
Faculty : Faculty of Computer Science

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher educational institution. Furthermore, it does not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

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



Surabaya, June 30, 2026
Declarant,



HANIN FATMA SORAYA

NPM.22081010069

ABSTRACT

Student Name / Number : Hanin Fatma Soraya / 22081010069

Thesis Title : Gold Price Prediction Analysis Using Support Vector Regression and Random Forest Based On BPS Data

Advisor : 1.Dr. Budi Nugroho, S.Kom, M.Kom.
2. Dr. Firza Prima Aditiawan, S.Kom., MTI

Fluctuations in gold prices, influenced by various global and domestic economic factors, make gold price prediction a complex time series problem. A multivariate time series approach was applied by comparing the Support Vector Regression (SVR), Random Forest (RF), and Naive Forecast baseline models using daily data from January 2018 to April 2026 with 3,015 observations. Predictor variables used include global gold prices, the USD/IDR exchange rate, inflation, and world crude oil prices. To capture temporal patterns, feature engineering was performed using lag, moving average, volatility, and day_index. Data was divided chronologically (time-based split) with a ratio of 80:20, while hyperparameter optimization used Bayesian Optimization. Model validation on training data was performed using TimeSeriesSplit to adjust for the characteristics of the time series data and minimize the risk of data leakage. Test results showed that Support Vector Regression (SVR) provided the best performance with a MAPE of 0.06%, an RMSE of 1,825.82, and an MAE of 1,353.09. The Naive Forecast model ranked second with a MAPE of 0.23%, while Random Forest produced the lowest performance with a MAPE of 28.87% due to extrapolation failure. These results indicate that the combination of multivariate data, temporal feature engineering, Bayesian Optimization, and the Support Vector Regression algorithm can produce more accurate gold price predictions, while Random Forest is less suitable for time series data with strong long-term upward trends.

Keywords: Gold Price Prediction, Support Vector Regression, RandomForest, Multivariate Time Series, Bayesian Optimization.

ACKNOWLEDGEMENTS

Praise be to God Almighty for all His grace, guidance, and gifts to the author, enabling the successful completion of this thesis entitled "**Gold Price Prediction Analysis Using Support Vector Regression and Random Forest Algorithms Based on BPS Data**" The author has also received substantial support from various parties, both moral, spiritual, and material. Therefore, the author would like to express his gratitude to:

1. To my parents, Mom and Dad, thank you for all your prayers, love, sacrifices, and never-ending support. I realize that every step I have taken is inseparable from the struggle, patience, and sincerity of Mom and Dad in accompanying me. In every process that is not always easy, Mom and Dad's faith is one of the biggest reasons for me to keep trying. I hope this achievement can be a small form of my gratitude for all the love and struggle that Mom and Dad have given me.
2. Prof. Dr. Ir. Novirina Hendrasarie, MT, Dean of the Faculty of Computer Science, Veteran National Development University, East Java, who has provided facilities, opportunities, and a conducive learning environment so that I can optimally pursue my studies. Her leadership provides much inspiration for students to continue to develop.
3. Ms. Intan Yuniar Purbasari, S.Kom. MSc. as the Head of the Informatics Study Program, Faculty of Computer Science, National Development University "Veteran" East Java, who has provided guidance, academic support, and good study program management so that the lecture process can run smoothly until the final assignment preparation stage.
4. My first advisor, Dr. Budi Nugroho, S.Kom., M.Kom., who patiently and meticulously provided direction, input, and guidance throughout the thesis writing process. His dedication in guiding me has helped me understand many things and complete this research with more direction.
5. Supervisor 2, Mr. Firza Prima Aditiawan, S.Kom., MTI, who has dedicated his time, provided invaluable insights, and guidance throughout the process of writing this thesis. His encouragement and suggestions have been

invaluable in helping me perfect my research.

6. To my sister, Yuvan Aqila. Thank you for always being a constant companion in my life, a source of encouragement on this journey, and a source of encouragement. Your presence has been a source of motivation in every step of this journey, including completing my thesis and continuing to strive to bring positive things that bring pride to my family.
7. To my friends during college, "Heavenly Angels", whom the author cannot mention one by one, who have always given encouragement, motivation, and accompanied the author since the first year of college.
8. Hanin's close friends, namely Aulia, Hanif, Silvia, who always provide support, prayers, and positive affirmations for all the activities that the author does and during the lecture process from the first semester to the preparation of the thesis.
9. The Difams family of Aulia, Inka, Karisa, and Nath as boarding house friends who became a second home for the author to support, strengthen, pray for, listen to each other's complaints, laugh together, so that every step feels lighter to go through.
10. The author's best friend of all time, Aulia Arifany Safitri, is not just a friend but a friend who has always been there in every phase of the author's life journey. Always a loyal listener and the main support who is able to accept the good and bad sides of the author without ever judging. Her presence makes the author strong and inspires the author during this lecture. The sincere support you give has been an encouragement in facing every academic and personal challenge. Thank you for the happy and sad moments we have gone through together. May God grant happiness and success in every step of your life.
11. Yudha Wahyu Devstiandi, who has been a sincere source of encouragement, emotional support, and a source of sharing throughout the process of compiling this thesis. Their presence and support, both in the form of motivation and understanding, have been invaluable in helping the author overcome various academic challenges and pressures, leading to the completion of this thesis.

12. Lastly, I would like to thank a girl who has been silently struggling non-stop, a simple woman with a small heart but with big dreams. Thank you to the researcher of this thesis, namely myself, Hanin Fatma Soraya. The first daughter and the hope of her parents. Thank you for being present in this world, for surviving this far, and continuing to walk through all the challenges the universe presents. Thank you for remaining brave enough to be yourself. I am proud of every small step you take, for all the achievements that others may not celebrate. Even though sometimes your expectations do not match what the universe gives, keep learning to accept and be grateful for whatever you get. Don't get tired of trying, be happy wherever you are. I pray that your small steps will always be strengthened, surrounded by good and great people, and your dreams will be answered one by one. Amen

The author realizes that the writing and compilation of this thesis are far from perfect due to limited knowledge and experience. Therefore, the author is very open to criticism and suggestions aimed at the thesis. The author hopes that this thesis will provide knowledge and benefits to readers, academics, and the author in particular.

Surabaya, June 10, 2026

Author

LIST OF CONTENTS

VALIDITY SHEET.....	ii
APPROVAL SHEET	iii
ORIGINALITY STATEMENT LETTER.....	iv
ABSTRACT	v
ABSTRACT.....	vi
FOREWORD	vii
LIST OF CONTENTS	x
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xiv
CHAPTER I INTRODUCTION	1
12.1. Background.....	1
12.2. Formulation of the problem.....	2
12.3. Research purposes	3
12.4. Scope of problem.....	3
12.5. Benefits of research	4
CHAPTER II LITERATURE REVIEW	5
2.1. Previous Research.....	5
2.2. Basic theory	10
2.2.1 Gold	10
2.2.2 PredictionPrice (Time Series)	10
2.2.3 Data Mining.....	12
2.2.4 Support Vector Regression (SVR)	13
2.2.5 Random Forest.....	15
2.2.6 NormalizationData (Min–Max Scaler).....	17
2.2.7 Bayesian Optimization.....	19
2.2.8 Overfitting and underfitting	21
2.2.9 Model Evaluation Matrix	22
2.2.10 Streamlit.....	25
CHAPTER III SYSTEM DESIGN AND IMPLEMENTATION	27
3.1. Research methods	27

3.2. Research Flow	28
3.3. Analysis Steps.....	31
3.4. System Design	54
3.1.1 System Architecture.....	54
3.1.2 System Workflow	55
CHAPTER IV RESULTS AND DISCUSSION.....	61
4.1. Data Collection	61
4.2. Data Pre-processing	63
4.2.1. Date Format Handling	63
4.2.2. Handling Missing Values	65
4.2.3. Exploratory Data Analysis.....	65
4.3. Feature Engineering.....	68
4.3.1. Lag Features	69
4.3.2. Moving Average (Rolling Mean) Feature	70
4.3.3. Volatility Features	71
4.3.4. Daily Return Feature	71
4.3.5. Time Index Feature (day_index).....	72
4.3.6. Time-Based Data Sharing.....	73
4.3.7. Definition of Features and Target Variables.....	75
4.3.8. Data Normalization.....	75
4.4. Data Modeling	76
4.4.1. TimeSeriesSplit Technique for Cross Validation	77
4.4.2. Baseline Model: Naive Forecast.....	77
4.4.3. Support Vector Regression (SVR) Model	78
4.4.4. Random Forest Model	79
4.4.5. Hyperparameter Optimization with Bayesian Optimization.....	79
4.4.6. Model Training Results	80
4.5. Model Evaluation.....	81
4.5.1. Evaluation Metrics Used	81
4.5.2. Evaluation Results of the Three Models.....	83
4.6. Gold Price Prediction.....	84
4.6.1. Prediction Methodology	84
4.6.2. Prediction Results on Test Data.....	84

4.6.3. Residual Analysis	87
4.7. Discussion of Research Results	88
4.7.1. SVR as the Best Model: Analysis and Explanation.....	89
4.7.2. Why Random Forests Fail: Extrapolation Failure	90
4.7.3. Naive Forecast and the Implications of High Autocorrelation	91
4.7.4. Analysis of Feature Importance and Its Meaning.....	92
4.7.5. Relationship between Results and Research Objectives and Research Gaps	94
4.8. Implementation of Streamlit-Based Prediction System.....	95
CHAPTER V CLOSING	99
5.1. Conclusion	99
5.2. Suggestion.....	100
BIBLIOGRAPHY.....	102
ATTACHMENT.....	109

LIST OF TABLES

Table 2.1 Previous Research	10
Table 2.2 Prediction Evaluation Criteria Based on MAPE Value	29
Table 3.1 Gold Dataset	39
Table 3.2 Examples of Handling Missing Values in Gold Price Data	41
Table 3. 3 Examples of Feature Engineering in Gold Price Data	45
Table 3.4 Example of Gold Price Data Normalization Process	48
Table 4.1 Description of Gold Price Dataset Variables	66
Table 4.2 Results of Missing Values Examination on the Original Dataset	69
Table 4.3 Descriptive Statistics of Gold Price IDR/gram (2018-2026)	70
Table 4.4 Augmented Dickey-Fuller (ADF) Test Results	71
Table 4.5 Summary of All Features Used in the Study	75
Table 4.6 Results of Training and Testing Data Division	77
Table 4. 7 Best Hyperparameters of SVR Results of Bayesian Optimization	83
Table 4. 8 Random Forest Hyperparameters Bayesian Optimization Results	83
Table 4.9 Examples of Manual Calculation of Evaluation Metrics	85
Table 4. 10 Results of Model Performance Comparison on Test Data	85
Table 4. 11 Sample SVR Prediction Results at Several Test Points	88
Table 4. 12 Feature Importance of Random Forest (Top 15 Features)	93

LIST OF FIGURES

Figure 2.1 Stages of the SVR 20 Algorithm	
Figure 2. 2 Random Forest Workflow	22
Figure 3.1 Research Analysis Steps	37
Figure 3.2 Flowchart of Gold Prediction System	61
Figure 3.3 Model Testing	63
Figure 4.1 Load Dataset and Date Format	68
Figure 4.2 Output Code Load Dataset and Date Format	68
Figure 4.3 Trend of IDR/gram Gold Prices for the Period January 2018 - April 2026	71
Figure 4.4 Lag Features	73
Figure 4.5 Moving Average Features	74
Figure 4.6 Daily Volatility and Return Features	75
Figure 4.7 Output Results of Splitting Data	78
Figure 4.8 Normalization Visualization	80
Figure 4.9 TimeSeriesSplit Structure (n_splits=5)	80
Figure 4. 10 Convergence of Bayesian Optimization	81
Figure 4.11 Comparison of Actual and Predicted SVR Values	87
Figure 4.12 Comparison of Three Algorithms	88
Figure 4.13 Distribution of SVR Prediction Residuals on Test Data	89
Figure 4. 14 Feature Importance Random Forest (Top 15 Features)	93
Figure 4.15 User Interface View of Streamlit Based System	97
Figure 4.16 Visualization of Modeling Dataset	98