UNDERGRADUATE THESIS

PREDICTING MOVIE POPULARITY IN INDONESIA BASED ON METADATA USING GRADIENT BOOSTING

APRINIA SALSABILA ROIQOH
NPM 22081010166

THESIS ADVISORS
Dr. Rizky Parlika, S.Kom., M.Kom.
Dr. Firza Prima Aditiawan, S.Kom, M.T.I.

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

APPROVAL SHEET

PREDICTING MOVIE POPULARITY IN INDONESIA BASED ON METADATA USING GRADIENT BOOSTING

By :
APRINIA SALSABILA ROIQOH
NPM. 22081010166

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 May 12, 2026:

Approved,

Dr. Rizky Parlika, S.Kom, M.Kom NIP. 19840518 202121 1 003	.....	**(Advisor I)**
Dr. Firza Prima Aditiawan, S.Kom., MTI NIP. 19860523 202121 1 003	.....	**(Advisor II)**
Made Hanindia Prami Swari, S.Kom, M.Cs. NIP. 19890205 201803 2 001	.....	**(Head Assessor)**
Budi Mukhamad Mulyo, S.Kom., M.T. NIP. 19891118 202406 1 003	.....	**(Assessor I)**

Acknowledge by,

Dean of the Faculty of Computer Science

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

APPROVAL SHEET

PREDICTING MOVIE POPULARITY IN INDONESIA BASED ON METADATA USING GRADIENT BOOSTING

By :
APRINIA SALSABILA ROIQOH
NPM. 22081010166

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	:	Aprinia Salsabila Roiqoh
NPM	:	22081010166
Degree Program	:	Bachelor (S1)
Study Program	:	Informatic
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 education 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, May 25, 2026
Declarant,

APRINIA SALSABILA ROIQOH
NPM. 22081010166

ABSTRACT

Student Name / NPM	:	Aprinia Salsabila Roiqoh / 22081010166
Thesis Title	:	Predicting Movie Popularity In Indonesia Based On Metadata using Gradient Boosting
Advisor	:	1. Dr. Rizky Parlika, S.Kom., M.Kom. 2. Dr. Firza Prima Aditiawan, S.Kom, M.T.I,

The film industry in Indonesia has experienced significant growth; however, the success of a film in attracting audiences remains difficult to predict accurately. This study aims to develop a model for predicting the number of moviegoers in Indonesia based on pre-release metadata using gradient boosting algorithms, namely XGBoost, LightGBM, and CatBoost. The dataset was collected from Cinepoint and TMDb, consisting of 3,464 initial records, which were reduced to 2,595 after the preprocessing stage. The preprocessing steps included data cleaning, selective handling of missing values, logarithmic transformation of the target variable, and feature engineering using a Bayesian smoothing approach. The models were trained using two data split scenarios (80:20 and 70:30), and hyperparameter optimization was performed using Random Search and Bayesian Optimization (Optuna). Model performance was evaluated using RMSE, MAE, MAPE, and R^2 metrics. The results show that the best model was achieved by CatBoost with Random Search under the 80:20 data split scenario, yielding an R^2 value of 0.8729, MAE of 0.5538, RMSE of 0.7698, and MAPE 5,02%. These results indicate that CatBoost provides the most accurate and stable prediction performance compared to XGBoost and LightGBM. Furthermore, hyperparameter tuning was proven to improve model performance in predicting movie audience numbers. Feature importance and SHAP analysis reveal that the main actors, directors, and genres are the most influential features in the prediction results. This indicates that pre-release metadata plays a significant role in determining movie popularity in Indonesia.

Keywords: Movie, Popularity, Metadata, Gradient Boosting, RMSE

ACKNOWLEDGEMENTS

Praise be to Allah SWT for all His blessings, guidance, and grace bestowed upon the author, allowing this undergraduate thesis titled **"Predicting Movie Popularity in Indonesia Based on Metadata Using Gradient Boosting"** to be completed successfully.

The author would like to express the sincerest gratitude to Dr. Rizky Parlika, S.Kom., M.Kom. and Dr. Firza Prima Aditiawan, S.Kom., M.T.I., as the undergraduate thesis advisors, who have generously dedicated their time to provide guidance, invaluable advice, and motivation. The author has also received immense support from various parties, both morally, spiritually, and materially. Therefore, the author would like to extend appreciation to:

1. **Prof. Dr. Ir. Novirina Hendrasarie, M.T.**, as the Dean of the Faculty of Computer Science at Universitas Pembangunan Nasional "Veteran" Jawa Timur, for providing the facilities, opportunities, and a conducive learning environment that enabled me to pursue my studies optimally. Her leadership has been a great inspiration for students to continuously grow.
2. **Dr. Intan Yuniar Purbasari, S.Kom. MSc.**, as the Head of the Informatics Study Program, Faculty of Computer Science at Universitas Pembangunan Nasional "Veteran" Jawa Timur, for her guidance, academic support, and excellent management of the study program, ensuring that the academic process ran smoothly up to the final project presentation stage.
3. **First Advisor, Dr. Rizky Parlika, S.Kom., M.Kom.**, who with utmost patience and meticulousness provided guidance, constructive feedback, and direction throughout the undergraduate thesis writing process. His dedication has helped me understand many concepts and complete this research with a clearer focus.
4. **Second Advisor, Dr. Firza Prima Aditiawan, S.Kom., M.T.I.**, who generously spared his time, shared valuable insights, and provided tremendous guidance during the development of this undergraduate thesis. His motivation and suggestions were instrumental in refining my research.

5. **My beloved Parents**, who have always been the greatest source of strength, prayers, and motivation in my life Thank you for all the sacrifices, unwavering moral and material support, and endless love from the very beginning until the completion of this undergraduate thesis. Without your prayers and encouragement, I would not have been able to complete this academic journey successfully.
6. **My Sister**, who has always provided support, motivation, and advice during the writing of this undergraduate thesis. Thank you for the time and support given, which enabled me to complete this work on time.
7. **My closest friends**, who have constantly offered support, encouragement, and a place to share joys and sorrows throughout this undergraduate thesis journey. Thank you for the togetherness, cooperation, prayers, and laughter that made this journey feel lighter and more meaningful. Your presence has been an essential part of this process.
8. **Finally, a gentle note of gratitude to the author of this work, myself**. Thank you for holding on, for pushing through the exhaustion, and for refusing to give up despite the many hurdles in your way. You survived, and you finished. I hope the wisdom gathered during this long journey remains a guiding light for me, proving useful and meaningful for the rest of my life.

The author realizes that this undergraduate thesis is far from perfect and may contain shortcomings. Therefore, constructive criticism and suggestions from all parties are highly welcomed for the improvement of this work. Finally, despite all the limitations, the author hopes that this report will be beneficial to all readers in general and to the author in particular.

.

Surabaya, May 25th 2026

Author

TABEL OF CONTENT

LIST OF FIGURE

LIST OF TABLE