



HOUSE PRICE PREDICTION USING GRADIENT BOOSTING REGRESSOR, XGBOOST REGRESSOR, AND LIGHTGBM

UNDERGRADUATE THESIS

Submitted as one of the requirements to obtain

Sarjana Komputer (S.Kom.)

By

MUHAMMAD SYAFIQ AYASI

001201900054

**FACULTY OF COMPUTING
INFORMATION TECHNOLOGY STUDY PROGRAM**

CIKARANG

SEPTEMBER, 2023

PANEL OF EXAMINER APPROVAL


The Panel of Examiners declare that the undergraduate thesis entitled "**HOUSE PRICE PREDICTION USING GRADIENT BOOSTING REGRESSOR, XGB REGRESSOR AND LIGHTGBM**" that was submitted by **Muhammad Syafiq Ayasi** majoring in IT from the Faculty of Computer Science was assessed and approved to have passed the Oral Examination on Thursday September 7, 2023.

Panel of Examiner



RUSDIANTO ROESTAM

Chair of Panel Examiner



ROSALINA

Examiner I

STATEMENT OF ORIGINALITY

In my capacity as an active student of President University and the author of the thesis/final project/business plan stated below:

Name : Muhammad Syafiq Ayasi

Student ID Number : 01201900054

Study Program : Information Technology

Faculty : Computing

I hereby declare that my thesis/final project/business plan entitled "**HOUSE PRICE PREDICTION USING GRADIENT BOOSTING REGRESSOR, XGBOOST REGRESSOR, AND LIGHTGBM**" is to the best of my knowledge and belief, an original piece of work based on sound academic principles. If there is any plagiarism detected in this thesis/final project/business plan, I am willing to be personally responsible for the consequences of these acts of plagiarism, and will accept the sanctions against these acts in accordance with the rules and policies of President University.

I also declare that this work either in whole or in part, has not been submitted to another university to obtain a degree.

Cikarang, August 21, 2023



(Muhammad Syafiq Ayasi)

**SCIENTIFIC PUBLICATION APPROVAL FOR ACADEMIC
INTEREST**

As an academic community member of the President's University, I, the undersigned:

Name : Muhammad Syafiq Ayasi

Student ID Number : 01201900054

Study program : Information Technology

for the purpose of development of science and technology, certify, and approve to give President university a non-exclusive royalty-free right upon my final report with the title:

**HOUSE PRICE PREDICTION USING GRADIENT BOOSTING
REGRESSOR, XGBOOST REGRESSOR, AND LIGHTGBM**

With this non-exclusive royalty-free right, President University is entitled to converse, to convert, to manage in a database, to maintain, and to publish my final report. There are to be done with the obligation from President University to mention my name as the copyright owner of my final report.

This statement I made in truth.

Cikarang, August 21, 2023



(Muhammad Syafiq Ayasi)

**ADVISOR APPROVAL FOR JOURNAL/INSTITUTION'S
REPOSITORY**

As an academic community member of the President's University, I, the undersigned:

Name : Genta Sahuri, S.Kom., M.Kom.

ID number : 20181000777

Study program: Information System

Faculty : Computing

Declare that following thesis:

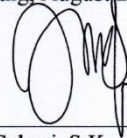
Title of thesis : House Price Prediction using Gradient Boosting Regressor,
XGBoost Regressor, and LightGBM

Thesis of author : Muhammad Syafiq Ayasi

Student ID Number : 01201900054

will be published in journal / institution's repository / proceeding / unpublish.

Cikarang, August 21, 2023



(Genta Sahuri, S.Kom., M.Kom.)

Final Draft - Muhammad Syafiq Ayasi

ORIGINALITY REPORT

20% SIMILARITY INDEX	10% INTERNET SOURCES	7% PUBLICATIONS	15% STUDENT PAPERS
--------------------------------	--------------------------------	---------------------------	------------------------------

PRIMARY SOURCES

1	Submitted to President University Student Paper	11%
2	towardsdatascience.com Internet Source	1%
3	Zisen Zhong, Ying Li, Zhouzhou Han, Zhenliu Yang. "Ship Target Detection Based on LightGBM Algorithm", 2020 International Conference on Computer Information and Big Data Applications (CIBDA), 2020 Publication	1%
4	as-proceeding.com Internet Source	1%
5	dspace.lib.ntua.gr Internet Source	1%
6	Nurfadhlina Mohd Sharef, Nurul Amelina Nasharuddin, Raihani Mohamed, Nabila Wardah Zamani, Mohd Hafeez Osman, Razali Yaakob. "Applications of Data Analytics and Machine Learning for Digital Twin-based Precision Biodiversity: A Review", 2022	<1%

or, choose a file to upload

CHOOSE FILE

Final Draf... Ayasi.pdf

Accepted file types: pdf, docx, doc, txt

I agree to the terms of service

Get Results

Your text is likely to be written entirely by a human

There is a 0% probability this text was entirely written by AI

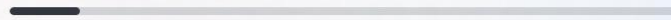


The nature of AI-generated content is changing constantly. As such, these results should not be used to punish students. While we build more robust models for GPTZero, we recommend that educators take these results as one of many pieces in a holistic assessment of student work. See our [FAQ](#) for more information.

GPTZero Model Version: [2023-09-14](#)

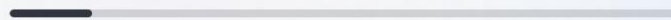
Stats

Average Perplexity Score: 105.024



A document's perplexity is a measurement of the randomness of the text

Burstiness Score: 122.945



A document's burstiness is a measurement of the variation in perplexity

Your sentence with the highest perplexity, "The Waterfall model is a software develop", has a perplexity of: 671

ABSTRACT

The intricate and essential topic of predicting housing prices within the real estate industry will influence customer preferences. Various machine learning algorithms have been applied in an effort to produce predictions that are more accurate. In this work, use and compare the performance of three well-known gradient boosting algorithms for predicting home prices: the Gradient Boosting Regressor (GBR), XGBoost, and LightGBM.

For the training and testing of the model, pertinent housing price information is gathered and compiled. Through rigors feature analysis, significant features for predicting home prices are found and chosen. Additionally, the same dataset was used to train the prediction model using GBR, XGBoost, and LightGBM three gradient boosting methods.

The experimental findings demonstrate that the three algorithms are capable of effectively resolving the issue of house price prediction. Each approach, nevertheless, has benefits and drawbacks in terms of model stability, accuracy, and speed. Based on pertinent evaluation measures, such as R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE), we compare the performance of the three algorithms.

In conclusion, picking the appropriate algorithm can enhance the precision and efficacy of house price projections. The findings of this study can serve as a useful manual for practitioners, researchers, and application developers as they select the optimal algorithm for more accurate house price projections.

ACKNOWLEDGMENT

First, I would like to thank Allah SWT for always giving me the health, strength, and ease that you have given me. Secondly, I would like to thank myself for finally being able to complete my final project. Thirdly, I would also like to thank those who have supported me during my final project support me during the completion of this final project:

1. Mr. Genta Sahuri S.Kom., M.Kom. was the person responsible for supervising my final project, who also helped me and guided me during the final project period.
2. Mr. Rila Mandala Ph.D. as Dean of Faculty Computing, Mrs. Cutifa Safitri, M.Sc, Ph.D. as Program Head of Information Technology, Mrs Indah as Secretary of the Faculty of Computing, and other lecturers who have provided a lot of knowledge during my studies at president university.
3. My beloved parents, for all their prayers and support
4. My friends, who always helped me study during college.

TABLE OF CONTENTS

ABSTRACT.....	1
DEDICATION.....	2
ACKNOWLEDGMENT	3
TABLE OF CONTENTS.....	4
LIST OF TABLES.....	8
CHAPTER I INTRODUCTION	12
1.1 Background	12
1.2 Problem Statement	13
1.3 Research Objective.....	13
1.4 Scope and Limitations.....	14
1.5 Methodology	14
1.6 Outline.....	16
CHAPTER II: LITERATURE STUDY.....	18
2.1 Data Mining.....	18
2.2 Machine Learning	18
2.3 Boosting	19
2.4 Gradient Boosting Regressor	20
2.5 XGBoost Algorithm	26
2.6 Light Gradient Boost Machine.....	27

2.7	Phyton.....	29
2.8	Mean Squared Error (MSE)	32
2.9	Root Mean Squared Error	32
2.10	Mean Absolute Error	33
2.11	R-Squared.....	34
2.12	Related Work	35
2.11.1	Car Price Prediction	35
2.11.2	Flight Price Prediction	36
2.13	Work Comparative Overview in Relation to Previous Work	36
CHAPTER III: SYSTEM ANALYSIS.....		38
3.1	System Overview	38
3.2	Function Analysis.....	38
3.3	Research Data.....	39
3.4	Hardware and Software Requirements.....	40
3.4.1	Necessary Hardware	40
3.4.2	Required Software	40
3.5	Data Mining Process and Algorithm	41
3.5.1	Load Dataset	43
3.5.2	Data Processing.....	43
3.4.3	Modelling Preparation	44
3.5.3	Train Models.....	44
3.6	Use Case Diagram.....	45
3.7	Use Case Narrative.....	46

3.7.1	Load Dataset	46
3.7.2	Data Processing.....	47
3.7.3	Split Data	48
3.7.4	Complete The Prediction Form.....	49
3.7.5	Graph Comparison	50
3.8	Swim Lane Diagram.....	51
3.8.1	Load Dataset	51
3.8.2	Cleaning the Dataset	52
3.8.3	Split Data	52
3.8.4	Complete the Prediction Form	54
CHAPTER IV: SYSTEM DESIGN.....		55
4.1	User Experience Design Prototype	55
4.4.1	Page of predictions.....	55
4.4.2	Front page	57
4.4.3	Graph.....	57
4.2	Formal Design	58
CHAPTER V: SYSTEM IMPLEMENTATION		60
5.1	Development of user interfaces.....	60
5.1.1	Front page	60
5.1.2	House Prediction Page	61
5.1.3	Garph.....	63
5.2	Processing Datasets	64
5.2.1	Load Dataset	64
5.2.2	Data Processing.....	65
5.2.3	Explaratory Data Analysis	67
5.2.4	Modelling Preparation	72
5.2.5	Train Model.....	75

5.2.6	Model Evaluation.....	76
5.3	Application Details.....	79
5.3.1	Import Library & Loaded Model	79
5.3.2	Front Page	79
5.3.3	Prediction Page	80
5.3.4	Garph Page.....	82
CHAPTER VI: SYSTEM TESTING		83
6.1	Testing Environment	83
6.2	Testing Scenario	83
6.2.1	User Features Testing.....	84
6.2.2	Prediction Form Testing.....	85
6.2.3	Predicton Form Algorihm Testing.....	85
CHAPTER VII: CONCLUSIONS & FUTURE WORK.....		88
7.1	Conclusion.....	88
7.2	Future Work.....	89
REFERENCES.....		90

LIST OF TABLES

Table 2.1 Comparison with Related Work	37
Table 3.1 List of Function.....	39
Table 3.2 Hardware Specification.....	40
Table 3.3 Software Specifications.....	40
Table 3.4 Load Datasets	47
Table 3.5 Cleaning Dataset	48
Table 3.6 Split Data.....	49
Table 3.7 Fill out the prediction form	49
Table 4.1 Prediction Page Function	56
Table 4.2 Hardware Requirements.....	59
Table 4.3 Software Requirements	59
Table 6.1 User Testing Scenario.....	84
Table 6.2 Testing Scenario Prediction.....	85

LIST OF FIGURES

Picture 1. 1 Waterfall Method	15
Picture 2. 1 Boosting Process Illustration	20
Picture 2. 2 Structure of Gradient Boosting Regressor Algorithm	25
Picture 2. 3 Structure of XGBoost Algorithm.....	27
Picture 2. 4 Structure of LightGBM Algorithm	27
Picture 3. 1 Flowchart of the System	42
Picture 3. 2 Use Case Diagram	46
Picture 3. 3 Swim Lane Dataset of Load Diagram	51
Picture 3. 4 Diagram of the Cleaning Datasets' Swim Lanes	52
Picture 3. 5 Split Data Swim Lane Diagram.....	53
Picture 3. 6 Swim Lane Diagram of Prediction Form To Complete.....	54
Picture 4. 1 Page of predictions	56
Picture 4. 2 Front Page.....	57
Picture 4. 3 Graph Page	58
Picture 5. 1 Front Page.....	61
Picture 5. 2 House Prediction Page.....	62
Picture 5. 3 House Prediction Page.....	62
Picture 5. 4 House Prediction Page.....	63
Picture 5. 5 Graph Page	64
Picture 5. 6 Import Library Code.....	64
Picture 5. 7 Load Dataset	65
Picture 5. 8 Check Data Type.....	65

Picture 5. 9 Check Data Dimensions	66
Picture 5. 10 Check Missing Values	66
Picture 5. 11 Smallest Price of House.....	67
Picture 5. 12 Largest Price of House.....	67
Picture 5. 13 Smallest Built Area.....	67
Picture 5. 14 Largest Built Area.....	68
Picture 5. 15 Smallest Land Area.....	68
Picture 5. 16 Smallest Land Area.....	68
Picture 5. 17 Number of Bedroom.....	69
Picture 5. 18 Bedroom Bar Chart.....	69
Picture 5. 19 Number of Bathroom.....	70
Picture 5. 20 Bathroom Bar Chart.....	70
Picture 5. 21 Number of Garage	71
Picture 5. 22 Garage Bar Chart.....	71
Picture 5. 23 Encoding Name	72
Picture 5. 24 Encoding Result.....	72
Picture 5. 25 Data Type After Encoding	73
Picture 5. 26 Splitting Features and Target	73
Picture 5. 27 Splitting Result Variable X	74
Picture 5. 28 Splitting Result Variable Y	74
Picture 5. 29 Split Datasets	75
Picture 5. 30 Train Model	75
Picture 5. 31 Gradient Boost score	76

Picture 5. 32 XGboost Score.....	77
Picture 5. 33 LightGBM Score	78
Picture 5. 34 Saving Model.....	79
Picture 5. 35 Import Library & Loaded Model in Web Application.....	79
Picture 5. 36 Front Page Code	80
Picture 5. 37 Prediction Page Code.....	81
Picture 5. 38 Graph Page Code	82