

UNIVERSITI TEKNOLOGI MARA

**DIAMOND PRICE PREDICTION
USING RANDOM FOREST ALGORITHM**

NUR AMIRAH BINTI MOHD AZMI

BACHELOR OF COMPUTER SCIENCE (Hons.)

JAN 2025

ACKNOWLEDGEMENT

Bismillahirrahmanirrahim and Alhamdulillah, praise and thanks to God because with His abundant grace, I was successful in completing this study within the given period. I want to extend my sincere gratitude to everyone who helped complete this study successfully, whether directly or indirectly. There have been many obstacles along the way, and I appreciate everyone's encouragement and support.

First of all, I would like to give a lot of thanks to my supervisor, Dr. Gloria Jennis Tan, for her guidance, support, and insightful feedback throughout the entire research process. Her expertise has played an important role in shaping the direction of this project. Not to be forgotten is Madam Ummu Fatimah Binti Mohd Bahrin, my CSP600 lecturer, for her commitment, efficient information delivery, supply of quality note sources, and above all her suggestions for improvement. The knowledge she imparted has helped me greatly in raising the caliber of my report writing. Also, my family and friends supported and understood me no matter what during the research process. I've found strength in your support, and I consider myself incredibly fortunate to have you in my life.

To conclude, I express my gratitude for the chance and confidence I have been given to carry out this research. I'm happy with how this study has advanced my abilities and understanding. Without the assistance of all those mentioned above, this accomplishment would not have been feasible. I appreciate you being a part of my educational experience.

ABSTRACT

This project addresses the challenge of accurately pricing diamonds by leveraging machine learning techniques. Diamonds are priced based on various attributes such as carat, cut, color, clarity, depth, and table, which exhibit complex interrelationships. Traditional methods struggle to model these complexities effectively, necessitating adoption of advanced algorithms to improve accuracy. The aim of this project is to develop a Diamond Price Prediction System using Random Forest, designed to accurately predict diamond prices based on attributes. Project seeks to compare performance of Random Forest with other regression models using key performance metrics. Some of the broad steps of methodology involve data preprocessing, by means of which handling of missing values, outliers, and inconsistencies for quality were developed. Development for a customized Random Forest-based model and a library-based one is performed. In both versions, feature selection was done along with hyperparameter tuning to have a better performance for both models. Comparisons among the MAE, RMSE, and R^2 on a custom-based, library-based model, along with other regression models, have been drawn on a comparative basis. The best balance was achieved at a 70:30 train-test split. Of the six regressors tried, Random Forest had the highest predictive accuracy. It outperformed those from Gradient Boosting and Decision Tree. On the contrary, SVR has the weakest performance among the six regressors. All in all, a library-based Random Forest model gives, consistently better accuracy compared to a custom-based one. That achieves lower MAE and RMSE which is 101.24 and 203.53. Besides the regression, the R^2 score, namely 99%. The research indicates that a custom Random Forest model surpasses standard implementations when properly optimized. Its deeper predictive accuracy arises from tuning hyperparameters to better suit the dataset. The model, however, is resource-consuming, has the requirement for pseudo-expertise in tuning parameters, and is therefore less accessible for a layperson. Future work should focus on improving the interpretability and extending the models to capture localized diamond pricing trends in real-life transactions.

TABLE OF CONTENTS

SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT.....	v
TABLE OF CONTENTS.....	vi
LIST OF FIGURES	x
LIST OF TABLES.....	xiii

CHAPTER 1: INTRODUCTION

1.1 Background of the Study	1
1.2 Problem Statement	2
1.3 Objective	3
1.4 Project Scope.....	3
1.5 Project Significance	6
1.6 Overview of Research Framework.....	7
1.6.1 Preliminary Phase.....	8
1.6.2 Design and Implementation Phase	9
1.6.3 Evaluation Phase	9
1.6.4 Gantt Chart	10
1.7 Summary	12

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction	13
2.1.1 Overview of Diamond Pricing.....	13
2.1.2 Machine Learning in Pricing Models	14
2.2 Prediction System.....	15
2.3 Diamond Pricing Determinants.....	20
2.3.1 Market Trends and Demand	21
2.3.2 Historical Pricing Data	23

2.4	Types of Prediction	24
2.5	Prediction System using Machine Learning	28
2.5.1	Implementation of Machine Learning Algorithms in Predicting Systems	32
2.5.2	Implementation of Regression Algorithm in Predicting System.....	35
2.6	Diamond Price Prediction System using Regression	38
2.7	Random Forest Algorithm.....	43
2.7.1	Implementation of Random Forest Algorithm in Various Problems	43
2.8	Similar Works	48
2.8.1	Performance Evaluation in Similar Works.....	55
2.9	Implication of Literature Review	56
2.10	Summary	58

CHAPTER 3: RESEARCH METHODOLOGY

3.1	Overview of Research Methodology	60
3.2	Preliminary Phase.....	63
3.2.2	Data Acquisition Phase.....	66
3.3	Design Phase	68
3.3.1	System Architecture	69
3.3.2	System Flowchart	72
3.3.3	Pseudocode of Selected Algorithm	73
3.3.4	User Interface Design	75
3.4	Implementation and Development Phase.....	76
3.5	Evaluation Phase	79
3.5.1	Classification and Prediction Accuracy.....	80
3.6	Gantt Chart.....	82
3.7	Conclusion.....	84

CHAPTER 4: RESULT AND DISCUSSION

4.1	Overview of Project System.....	85
4.2	Prototype Development.....	87
4.2.1	Data Understanding	87