

# “Diamond Market Insights – A Predictive Modelling Framework for Pricing Analysis”

**Mr.P.Srinivasa\*1, Ch.Likhitha\*2, M.Shivani\*3, M.Jayadithya\*4,**

\*<sup>1</sup>Associate Professor, CSE , ACE Engineering College, Hyderabad, India.

\*<sup>2,3,4</sup> Students of Department CSE , ACE Engineering College, Hyderabad, India.

**ABSTRACT-**This study will create a full-scale framework for predictive modelling focused on extracting vital information regarding the market and improving price analysis for diamonds. The analysis will be based on an organized set of data comprising the globally accepted parameters for valuing diamonds, which are the 4 Cs (carat, cut, colour, and clarity), along with other characteristics such as diamond depth, table, and dimensions.

The study will be conducted by applying ML algorithms for regression analysis, such as Random Forest, Gradient Boosting (XGBoost), and, if needed, Deep Learning models. The goal of the research is to find the correlation between these characteristics and predict the retail price of a diamond.

One of the essential aspects of the process is to conduct proper data processing, feature engineering, and hyperparameters tuning. It will allow us to achieve accurate results that are not prone to errors in the future and have good RMSE and  $R^2$  scores. This study will become a helpful instrument in making decisions and creating strategies for jewellers, valuers, and financial experts in their work with diamonds.

## I. INTRODUCTION

The diamond market cannot be easily estimated due to limitations of straightforward pricing methods (i.e., the Four C's). This study aims to develop a reliable price prediction model using artificial intelligence to address the problem.

## II LITERATURE REVIEW

Modern research in the field of diamond price analysis using Supervised Machine Learning shows impressive results achieved by various ensemble regression methods, namely XGBoost and Random Forest. These models produce high-quality results in terms of the coefficient of determination ( $R^2$ ) due to effective use of the inherent physical properties of diamonds, called 4Cs (Carat, Cut, Color, Clarity). But along with high accuracy comes a set of limitations and shortcomings which make these models less useful in practical use: static feature approach (modeling based solely on physical attributes ignoring dynamics of gold rate, inflation, or Rapaport index); black box nature of models that prevents industry professionals from understanding the reasons behind the prediction and adopting the model into the process; inability of current databases to reflect changes that occurred during the recent explosive popularity of Lab-Grown Diamonds (LGDs). Our project strives to tackle these problems via several improvements, such as development of a hybrid model including both regression and time series analysis that would involve dynamics of diamond prices; application of explainable AI methods (SHAP values) to increase the interpretability of our models; creation of a segmented model to cover two types of diamonds, natural and lab-grown.

**2.11. Comparison Table: Literature Review on Daimond market insights.**

Author(s)	Titles	Focus	Keypoints	Limitations
Reddy et al. (2025): (2025):	"Diamond Price Prediction..."	Comparative ML Regression	XGBoost is the optimal ensemble model for high accuracy on the 4Cs data.	Static Modeling: Only uses intrinsic features; cannot forecast market-driven price changes over time.
Alani et al. (2024):	"Predictive modelling and benchmarking..."	Benchmarking & Hyperparameter Tuning	Established a new benchmark through rigorous hyperparameter tuning across multiple ML models.	Feature Isolation: Does not integrate macro-economic indicators (e.g., gold price) for true market insight.
Wyld et al. (2025):	"Integrating Machine Learning for Diamond Price..."	Dual-Task: Price + Classification	Proposed a unified model for both price prediction and Natural vs. Lab-Grown Diamond (LGD) classification.	Lacks Interpretability: The advanced ensemble models are black-box and fail to provide transparent explanations for the prediction.

**3. Research Gaps:**

**i.Scope of Limited Features:**

The existing algorithm mainly considers features like 4Cs and physical aspects. Nevertheless, additional influencing features like brand reputation, certification (GIA), trends in demand, and regional price discrepancies are not considered.

**2.No Inclusion of Market Dynamics:**

Price of diamonds depends on dynamic global market conditions. No economic metrics, currency conversion rates, and seasonal demand changes are considered.

**3.No Consideration for Images:**

Aspects of visual appearance like sparkle and brilliance have considerable influence on the final price, but they are not accounted for in the existing dataset. Implementation of image processing can improve accuracy.

**4.Generalization of Model:**

The existing model has been designed using a particular dataset. It may require modifications when applied to different datasets.

**4. CONCLUSION**

The analysis demonstrates the ability to predict the price of diamonds with high accuracy using ML algorithms such as random forest and XGBoost based on 4Cs and others factors. With appropriate pre-processing and configuration, the algorithms perform efficiently (RMSE,R2R2), offering an effective tool in pricing decisions in the diamond business.

## 5. REFERENCES

- 1.A. J. M. S. R. B. P. N. D. L. V. K. M. A. Reddy et al., "Diamond Price Prediction Based on Regression Algorithms: Analyzing Key Factors and Model Performance," *ResearchGate*, August 2025.A. J. M. S. R. B. P. N. D. L. V. K. M. A. Reddy et al., "Diamond Price Prediction Based on Regression Algorithms: Analyzing Key Factors and Model Performance," *ResearchGate*, August 2025.
2. David C. Wyld et al., "Integrating Machine Learning for Diamond Price Prediction and Distinguishing Natural Diamonds From Lab Grown," *Computer Science & Information Technology (CS & IT)*, 2025.David C. Wyld et al., "Integrating Machine Learning for Diamond Price Prediction and Distinguishing Natural Diamonds From Lab Grown," *Computer Science & Information Technology (CS & IT)*, 2025.
4. M. A. R. E. A. W. A. Alani et al., "Predictive modeling and benchmarking for diamond price estimation: integrating classification, regression, hyperparameter tuning and execution time analysis," *International Journal of System Assurance Engineering and Management*, October 2024.M. A. R. E. A. W. A. Alani et al., "Predictive modeling and benchmarking for diamond price estimation: integrating classification, regression, hyperparameter tuning and execution time analysis," *International Journal of System Assurance Engineering and Management*, October 2024.
- 5.H. G. K. S. W. M. T. L. S. Gunathilaka et al., "Diamond Price Prediction using Machine Learning Techniques," *IEEE Conference Publication*, September 2024.H. G. K. S. W. M. T. L. S. Gunathilaka et al., "Diamond Price Prediction using Machine Learning Techniques," *IEEE Conference Publication*, September 2024.
- 6.M. S. Sánchez, "Optimization of diamond price prediction strategies using machine learning techniques," *Journal of Economics and International Finance*, 2024.M. S. Sánchez, "Optimization of diamond price prediction strategies using machine learning techniques," *Journal of Economics and International Finance*, 2024.
- 6.Mihir et al., "Diamond Price Prediction using Machine Learning," *2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, 2021.Mihir et al., "Diamond Price Prediction using Machine Learning," *2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, 2021.