INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Predicting Gemstone Prices Using Attribute Analysis

A Project Proposal by

Mohamed Azhar Mohamed Yoosuf

Supervised by

Murshid Hassen

Submitted in partial fulfilment of the requirements for the BEng in Software Engineering degree at the University of Westminster.

AUGUST 2024

ABSTRACT

The valuation and pricing of gemstones are challenging, traditional appraisal methods since they are heavily dependent on expert opinions. This project provides one such solution in developing a gemstone price prediction system that can automatically standardize the process of valuing a gemstone using machine learning techniques. It considers attributes of gemstone such as carat, clarity, color, cut, and hardness. The hardness of the gemstone is included as an important factor, because it mainly influences the value. Measured hardness value can vary due to gemstone treatments, Orientation and internal cracks. The aim of the project is to provide a transparent, data-driven system for consumers and jewelers, allowing for more consistent and reliable pricing.

Past research papers informed the choice of features such as carat, clarity, color, cut, and hardness highlighted as critical to determine the price of a gemstone using machine learning techniques. Attribute selection has been done to accurately predict based on data from public gemstone datasets and expert surveys. In this respect, three of the major models involved in price prediction are SVM, LR and Random Forest, trained by these attributes. The developed prediction system will take any given inputs of these attributes by the user and provide proper, data-driven price estimates. This system provides real value for its users by offering an objective, standardized solution for price determination.

Initial results of the prototype indicated that Random Forest has a higher level of accuracy and lower error rates compared with Linear Regression and SVM, thus showing the best performance in predicting the prices of gemstones. Though modest overall, these findings confirm the effectiveness of the proposed model. Further work in feature engineering and optimization may result in even better predictive accuracy.

Subject Descriptions: Computing methodologies → Machine learning → Machine learning approaches → Classification and regression trees

Keywords: Support Vector Machines, Machine Learning, Linear Regression, Convolutional Neural Network and Random Forest.