DIAMOND PRICE PREDICTION SYSTEM FOR BUYERS AND SELLERS

ITA5006 – DECISION SUPPORT SYSTEM

PROJECT BASED COMPONENT REPORT

in

M.C.A

By

Rajat Singh (22MCA0139)

Hrishikesh S G (22MCA0162)

Sambit Basu (22MCA0240)

Srijan Paria (22MCA0266)

Pragya Parmita Sambodhi (22MCA0331)

School of Information Technology and Engineering

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

August 2023

Declaration

We hereby declare that the report entitled **“DIAMOND PRICE PREDICTION SYSTEM FOR BUYERS AND SELLERS”** submitted by me, for the MAT5024 Decision Support Systems (EPJ) to Vellore Institute of Technology is a record of bonafide work carried out by me under the supervision of **Dr. GITANJALI J**.

We further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for any other courses in this institute or any other institute or university.

Place : Vellore
Date : 09 - 08 -2023

Signature of Candidates

Sambit Basu

Rajat

Hrishikesh :)

Pragya

Srijan Paria

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to all those who contributed to the successful completion of this project. Their unwavering support, guidance, and encouragement have been instrumental in shaping this endeavor into a meaningful achievement.

First and foremost, we express our sincere appreciation to our Decision Support Systems professor, Professor Gitanjali J whose expertise, guidance, and valuable insights enriched every stage of this project. Your mentorship provided us with a clear path, and your constructive feedback consistently steered us in the right direction.

We would also like to thank our University faculty members for their valuable input and encouragement throughout the project's duration. Your academic rigor and dedication to nurturing our intellectual growth have been invaluable.

Our gratitude extends to our fellow classmates and friends who offered support, constructive discussions, and inspiration. Your diverse perspectives and collaborative spirit enriched our project's development and created an environment of shared learning.

We owe a debt of gratitude to our families for their unwavering support, patience, and understanding during late nights and weekends spent working on this project. Your encouragement propelled us forward when challenges seemed daunting.

To the participants and volunteers who generously shared their time and insights for our research, we extend our heartfelt thanks. Your contributions were pivotal in shaping the empirical foundation of this project.

We also acknowledge the open-source community for providing valuable tools, libraries, and resources that enriched our project's implementation and efficiency.

Finally, we acknowledge the power of teamwork and collaboration among our project team members. Each member's dedication, expertise, and commitment have contributed to the synergy that drove this project towards success.

In conclusion, the successful completion of this project would not have been possible without the collective effort and support of the individuals mentioned above. Your contributions have left an indelible mark on our journey, and for that, we are immensely grateful.

Table of Contents

Table of Figures

Abstract

The Diamond Price Prediction project aims to leverage the power of machine learning algorithms to accurately predict the prices of diamonds based on their attributes. Diamonds, being precious and highly sought-after gemstones, have prices influenced by various factors such as carat weight, cut quality, color grade, and clarity grade. This project seeks to provide a user-friendly solution where users can obtain reliable price estimates for diamonds based on these attributes.

The project's foundation lies in a meticulously curated dataset containing information on thousands of diamonds, their attributes, and corresponding market prices. To achieve robust predictions, various machine learning algorithms were employed.

The developed system comprises a user-friendly frontend interface, where users can input diamond attributes, and a sophisticated backend that processes the inputs through the trained models to provide accurate price predictions. Extensive testing and validation were conducted to ensure the reliability and accuracy of the predictions.

The Diamond Price Prediction project provides a crucial decision support mechanism for jewelers, diamond enthusiasts, and investors seeking real-time and data-driven diamond pricing information. The combination of versatile machine learning algorithms and intuitive user interface promises to be a significant step towards revolutionizing the diamond market, offering reliable price estimates and insights to all users, and empowering them to make well-informed decisions in the dynamic and competitive diamond industry.

Keywords ***: Diamond Price Prediction, Machine Learning Algorithms, Attributes, User-friendly Solution, Dataset, Decision Support Mechanism, Reliable Price Estimates***

1. Introduction

The global diamond industry stands as a resounding testament to the enduring allure and profound economic significance inherent in these exquisite gemstones, captivating consumers, jewelers, and investors alike on a global canvas. In this intricate and multifaceted market, where the intricate dance of diamond prices is orchestrated by a symphony of complex factors, a delicate balance is struck between carat weight, cut quality, color grade, and clarity grade, collectively weaving the intricate fabric that defines the value of each individual stone. The riddle of diamond pricing, shrouded in enigma, has persistently engaged the attention of industry stakeholders, each ardently pursuing insight to chart a course through the ever-evolving, fiercely competitive landscape that characterizes this realm.

In response to the relentless quest for clarity within this dynamic domain, the "Diamond Price Prediction" project emerges as a shining beacon of innovation. Fueled by the formidable potential inherent in machine learning algorithms, the project embarks on a journey to proffer a transformative solution to the enduring challenge of predicting diamond prices with an unprecedented level of precision. Rooted deeply within data-driven methodologies, this endeavor ushers in a pioneering approach that redefines decision support mechanisms within the jewelry industry.

Nestled at the core of this ambitious undertaking is a meticulously curated dataset—a veritable repository of knowledge that encompasses thousands of individual diamonds, replete with their diverse attributes and corresponding market valuations. This treasure trove of information serves as the bedrock upon which the project is meticulously built. The voyage of this initiative traverses an expansive landscape of advanced machine learning algorithms, a symphony of methodologies that includes the versatile Random Forest, the robust Gradient Boosting, the adaptive XGBoost, and the agile Ada Boost. With careful calibration and a wealth of data harvested from a dataset teeming with over 10,000 diamonds, these algorithms adeptly capture the intricate interplay between various attributes and their resultant pricing, culminating in the derivation of meticulously precise price predictions.

At the heart of this ambitious endeavor pulses a profound ambition: to create a sanctuary of understanding for individuals navigating the intricate pathways of the diamond market. Jewelers, connoisseurs, and investors alike find solace and empowerment within the "Diamond Price Prediction" system—an unassailable bastion of reliable, data-driven insights. The ingenious amalgamation of a user-friendly frontend interface with a robust backend processing engine stands as a testament to this project's commitment to democratizing the acquisition of diamond price estimates based on specific attributes. In essence, this solution bestows upon its users a potent instrument that seamlessly bridges the realms of data, technology, and intuition, providing direct access to a wealth of diamond knowledge at their fingertips.

As the vistas of innovation unfurl, the veracity of predictions ascends to a paramount position. Rigorous testing and unwavering validation emerge as the vanguards of reliability and precision. Within this intricate project, a dualistic prism comes to light, with two distinct binning methodologies standing as the cornerstones of price categorization. The first approach employs equal interval binning, segmenting the spectrum of prices into equidistant divisions, while the second, price range binning, strategically clusters diamonds into predefined ranges. This duality of perspective not only enhances understanding but also opens the door to nuanced insights into the trajectory of diamond pricing trends across an expanse of diverse market segments.

Yet, the significance of the "Diamond Price Prediction" project reaches far beyond the realms of algorithmic prowess and user interfaces. This endeavor orchestrates a profound paradigm shift that has the potential to redefine the very fabric of the diamond market. By promising real-time, data-driven diamond pricing information, the project serves as a catalyst for transformative change. It confers upon users the agency to decipher the enigma of diamond valuation, seamlessly weaving data into their tapestry of decision-making. To jewelers, it metamorphoses into a navigational compass steering the waters of inventory management; to enthusiasts, it transforms into an oracle ensuring astute acquisitions; and to investors, it becomes a prism, offering a refined lens through which to discern equitable value.

In the chapters that unfurl ahead, the stage is set for a meticulous exploration. The methodologies, algorithms, and system architecture—the very foundation upon which this project rests—come under the scrutinizing gaze. As curtains rise, the results of this endeavor cascade into view—a vibrant tapestry of insights emerging from the realm of predictive pricing. In every facet, the transformative potency of the "Diamond Price Prediction" project resounds, offering an exquisite harmony where technology converges with the timeless allure intrinsic to diamonds.

The intricate symphony of the global diamond industry, with its enduring appeal and far-reaching economic implications, resonates across continents, captivating not only consumers but also jewelers and investors. Within this multifaceted market, the delicate choreography of diamond prices unfolds through a complex interplay of variables, each lending its own note to the melody. Carat weight, cut quality, color grade, and clarity grade harmonize in a nuanced dance, weaving a narrative that defines the unique value of each precious stone. The pursuit of deciphering the intricate code of diamond pricing has held industry stakeholders in its thrall for years, as they strive to navigate the labyrinthine landscape of perpetual evolution and fierce competition.

Amidst this demand for transparency, the "Diamond Price Prediction" project emerges as a beacon of innovation, illuminating the path to precision through the prism of machine learning algorithms. With unwavering determination, it endeavors to present a transformative solution to the age-old challenge of predicting diamond prices with unprecedented accuracy. Anchored in the bedrock of data-driven methodologies, this initiative heralds a groundbreaking approach that promises to reshape the foundations of decision support mechanisms within the jewelry sector.

Central to the core of this ambitious undertaking rests a meticulously cultivated dataset—a reservoir of knowledge that encapsulates the attributes of thousands of individual diamonds, each juxtaposed with its corresponding market valuation. This repository forms the bedrock upon which the project is meticulously constructed. The journey embarked upon by this endeavor traverses a landscape of sophisticated machine learning algorithms—each an instrument in a symphony of analytics. The ensemble comprises the versatile Random Forest, the stalwart Gradient Boosting, the adaptive XGBoost, and the agile Ada Boost. These algorithms, meticulously tuned and honed through the infusion of data from a comprehensive dataset of over 10,000 diamonds, intricately unravel the relationships that bind attributes and prices, culminating in the generation of precise price predictions.

The heartbeat of this ambitious pursuit resonates with the desire to forge a haven of accessibility within the complex diamond market. Jewelers, enthusiasts, and investors alike find refuge in the "Diamond Price Prediction" system—a stronghold of dependable information. Seamlessly amalgamating a user-friendly frontend interface with a robust backend processing engine, this solution democratically opens the gateway to diamond price estimates, tailored to specific attributes. In essence, it bestows upon users a tool that effortlessly marries data, technology, and intuition, enabling them to seamlessly access the intricate realm of diamond knowledge.

As the horizon of innovation stretches beyond the limits of imagination, the foundation of predictions assumes a pivotal role. Rigorous testing and relentless validation emerge as the guardians of reliability and precision. Within this intricate project, two distinct binning methodologies stand as sentinels of accuracy. The equal interval approach segments the price spectrum into uniform divisions, while the price range binning method strategically clusters diamonds into pre-defined ranges. This dual perspective extends a nuanced understanding of the trajectory of diamond pricing trends, casting light on diverse market segments.

However, the resonance of the "Diamond Price Prediction" project reverberates far beyond algorithmic complexity and user interface sophistication. It heralds a seismic shift, poised to redefine the very essence of the diamond market. With its promise of real-time, data-driven diamond pricing information, the project serves as a catalyst for transformation, endowing users with the power to unravel the enigma of diamond valuation, intertwining data seamlessly into their decision-making fabric. For jewelers, it metamorphoses into a compass steering the ship of inventory management; for enthusiasts, it evolves into an oracle guiding informed acquisitions; and for investors, it transforms into a prism, through which the spectrum of fair value can be discerned.

As the narrative unfolds in subsequent chapters, the stage is set for a meticulous exploration of methodologies, algorithms, and system architecture. With the rise of the curtain, the results of this ambitious endeavor unravel—a vibrant tapestry woven from the realm of predictive pricing. In each facet, the transformational power of the "Diamond Price Prediction" project resonates, harmonizing technology with the timeless allure encapsulated within the world of diamonds.

1.1 Background and Context:

For centuries, the enduring allure of diamonds has woven an enchanting tale of beauty and rarity, captivating the human imagination. These exquisite gemstones, celebrated for their timeless elegance, have become more than mere adornments; they symbolize enduring love, prestige, and artistry. At the heart of their mystique lies an intricate dance of attributes – carat weight, cut quality, color grade, and clarity grade – each playing a distinct role in shaping their market worth.

Yet, within this mesmerizing realm, a complex tapestry of pricing dynamics unfolds. The convergence of these attributes intertwines to determine a diamond's value, making the pricing framework an intricate labyrinth that has long confounded industry stakeholders. The challenge lies in deciphering this multi-dimensional puzzle, a task that has often been marred by uncertainties, inconsistencies, and the limitations of traditional pricing methodologies.

In response to this industry-wide quandary, the "Diamond Price Prediction" project emerges as a beacon of innovation. It stands as a testament to the synergy between tradition and technology, where data-driven methodologies take center stage. This project recognizes the need to elevate diamond pricing to new heights of accuracy and understanding. By harnessing the power of advanced technological solutions, it aims to unlock the hidden patterns within the vast troves of diamond data.

The crux of this initiative is to empower stakeholders with insights that transcend the surface glitter of these gemstones. Through meticulous analysis of historical pricing trends and intricate attribute interdependencies, the project endeavors to provide users with actionable insights. By predicting diamond prices with greater accuracy, it aspires to equip jewelers, enthusiasts, and investors alike with a powerful tool for informed decision-making.

In a world where the diamond industry is constantly evolving, the "Diamond Price Prediction" project stands as a pioneering effort to embrace the potential of technology and data. It seeks to unravel the complexities that have long eluded traditional approaches, ultimately transforming the diamond market into a realm where precision meets passion, and where data-driven insights lead the way to a new era of diamond pricing dynamics.

1.2 Problem Statement:

“The challenge at hand is to develop an accurate and user-friendly solution that leverages machine learning and comprehensive diamond data to predict prices, effectively addressing the historical limitations of traditional diamond pricing methods and enabling informed decision-making across the intricate global diamond market.”

The global diamond market is characterized by its multifaceted nature, where the pricing of diamonds is determined by a myriad of attributes. This complexity has historically hindered the ability of stakeholders to accurately predict and understand diamond prices. Existing pricing methods often lack the precision required for informed decision-making, leading to potential inefficiencies in inventory management, pricing strategies, and investment decisions.

The "Diamond Price Prediction" project addresses this problem by introducing a sophisticated machine learning-driven solution. By tapping into the wealth of data available on diamond attributes and historical market prices, the project aims to bridge the gap between traditional pricing approaches and modern technological advancements. The ultimate goal is to empower users, ranging from jewelers and enthusiasts to investors, with a reliable and user-friendly platform that generates accurate price estimates for diamonds based on their unique attributes.

Through the meticulous curation of a comprehensive dataset and the utilization of state-of-the-art machine learning algorithms, this project seeks to overcome the limitations of traditional pricing methods and provide a transformative tool for decision support within the diamond industry. By doing so, it strives to equip stakeholders with the knowledge and insights necessary to navigate the intricate and ever-changing landscape of diamond pricing, thereby ushering in a new era of data-driven decision-making and market optimization.

2. Literature Survey

Ref. No.	Title	Algorithms Used	Evaluation Metric	Data Set
1	Comparative Analysis Of Supervised Models For Diamond Price Prediction	Linear regression, Random Forest, Lasso Regression, DecisionTree, Ridge Regression, ElasticNet, AdaBoostRegRessor, and GradientBoostingRegressor	RMSE, Accuracy	Kaggle repository
2	Machine Learning Algorithms For Diamond Price Prediction	Linear regression, Forest regression, Gradient Boosting Regressor, Polynomial regression, Neural Network.	MAE(Mean Absolute Error), RMSE(Root-mean-square deviation)	Kaggle repository
3	Comparative Study Of Predicting Diamond Ring Prices In Online Retail Shop	Multiple Linear Regression, Random Forest, Deep Neural Network	MAE, MAPE(Mean absolute percentage error)	Kaggle repository
4	Comparison Of Machine Learning Algorithms For Predicting Diamond Prices Based On Exploratory Data Analysis	Ridge Regression, LASSO Regression, ElasticNet, Random Forest Regression, XGBoost, Support Vector Regression	Mean Absolute Error, Mean Squared Error	Online Diamond Store
5	Subjectivity Of Diamond Prices In Online Retail: Insights From A Data Mining Study	Multiple linear regression, Decision forest , Boosted regression trees, Artificial neural network	Mean absolute error, Mean absolute percent error	Online Diamond Store
6	Fatigue Driving Detection With Modified Ada-Boost And Fuzzy Algorithm	Ada-Boost, Fuzzy		NA
7	Comparison Of Gradient Boosting And Extreme Boosting Ensemble Methods For Webpage Classification	Gradient Boosting, Extreme Gradient Boosting	Accuracy, Recall, Precision, and F1 score	NA

8	An Improved Xgboost Model Based On Spark For Credit Card Fraud Prediction	XGBoost	Recall, F1-Score, and AUC (Area under the ROC Curve)	NA
9	Comparison Of Classification Techniques Used In Machine Learning As Applied On Vocational Guidance Data	NAIVE BAYES, ONER, KSTAR, JRIP	Mean absolute Error, Root mean squared error , Kappa statistic	NA
10	Improved Ada Boost Classifier For Sports Scene Detection In Videos: From Data Extraction To Image Understanding	Ada Boost	Accuracy	NA (Video)

Table 1.1: Literature Review

[1] Comparative Analysis of Supervised Models for Diamond Price Prediction :

The cost of diamonds varies depending on their characteristics. In this paper, they have conducted a comparison investigation and implemented a number of supervised models to forecast the diamond's price. In the study, they have compared the performance of eight different supervised models, including linear regression, lasso regression, ridge regression, decision tree, random forest, ElasticNet, AdaBoost Regressor, and Gradient Boosting Regressor. The model that performed the best overall and produced the most accurate results was highlighted. The goal of this study is to preprocess data, identify correlations between dataset attributes, train the aforementioned models, assess their correctness, evaluate their results, and ultimately determine which model performs the best, which is the Random Forest Regression Model.

[2] Machine Learning Algorithms For Diamond Price Prediction :

The ability of the diamond dealers to estimate the price accurately is crucial. The enormous variety in the sizes and qualities of diamond stones makes the prediction procedure challenging. Many machine learning methods, including Linear regression, Random forest regression, polynomial regression, Gradient descent, and Neural networks, were utilised in this study to aid in the prediction of diamond price. After developing numerous models, evaluating their efficacy, and evaluating the outcomes, it turns out that the random forest regression is the most effective one.

[3] Comparative Study Of Predicting Diamond Ring Prices In Online Retail Shop :

By leveraging data from online retail diamond ring stores, this study intends to create models for estimating the retail prices of jewellery. There are 187,821 records for loose diamonds and 2,206 records for rings. This study develops and compare a performance of three models consist of Multiple Linear Regression (MLR), Random Forest (RF), and Deep Neural Network (DNN). The evaluation metrics used for comparing algorithms are accuracy of prediction using MAE and MAPE. The results show that MAE for the ring price prediction of MLR, RF, and DNN are $688.36, $235.33, and $273.00, respectively.

In addition, MAE for a diamond price prediction of MLR, RF, and DNN are $3254.03, $450.44, $445.94, respectively.

[4] Comparison Of Machine Learning Algorithms For Predicting Diamond Prices Based On Exploratory Data Analysis :

A sizable dataset of loose diamonds scraped from an online diamond store is exposed to data mining in order to explore how a diamond's physical characteristics could predict its price. The results show that diamond weight, color, and clarity are the most significant drivers of diamond pricing. So, submit a proposal for an exploratory data analysis that contains a section that uses LASSO regression, ElasticNet regression, and Random Forest regression to analyze various aspects of news articles. This technique uses historical data to estimate diamond prices while maintaining a simple to understand trading strategy. In terms of prediction accuracy and interpretability, the suggested approach outperforms state-of-the-art techniques like extreme learning machines that use deep learning. The information suggests that news impact is important for forecasting.

[5] Subjectivity Of Diamond Prices In Online Retail: Insights From A Data Mining Study :

Diamonds are under a special class of goods whose perceived value is heavily influenced by socially formed ideas. They have utilized data mining on a sizable dataset of loose diamonds scraped from an online diamond store to examine the extent to which the physical characteristics of a diamond can be used to predict the diamond price. They discovered that the main factors affecting diamond prices are diamond weight, color, and clarity. The data mining findings also point to a significant level of subjectivity in diamond pricing, which may be a result of diamond dealers' use of price obfuscation techniques.

[6] Fatigue Driving Detection With Modified Ada-Boost And Fuzzy Algorithm :

The most significant and evident factors for detecting fatigued driving are facial features. In this study, the modified Ada-Boost algorithm is used to precisely recognise faces and pinpoint their eyes and mouths. The parameters of the eyes and mouth state are extracted using the adaptive threshold. Finally, a fuzzy algorithm paired with PERCLOS principles is employed to assess the level of weariness. The proposed method has proven to be more resilient, faster, more precise, and able to handle real-time demands, according to experiments.

[7] Comparison Of Gradient Boosting And Extreme Boosting Ensemble Methods For Webpage Classification :

Classifying web pages is a crucial effort in many fields, including web content screening, contextual advertising, and maintaining or growing web directories, among others. Web page classification using machine learning techniques has been found to be effective, and ensemble models have been used to enhance the output of single classifiers. In this work, binary classification is performed using the ensemble models of Gradient Boosting and Extreme Boosting. The dataset including web page URLs was compiled by hand. Extreme boosting gain in accuracy and speed was confirmed by a comparison of the two boosting algorithms. The speed and accuracy of extreme boosting have been shown to be around ten times faster than those of gradient boosting.

[8] An Improved Xgboost Model Based On Spark For Credit Card Fraud Prediction:

For numerous financial organizations, credit card theft results in substantial financial losses. An enhanced XGBoost model based on Spark is suggested due to the unbalanced dataset and vast amount of data in the field of credit card fraud. The Smote algorithm was employed in this project to balance the training set. The fraud detection system also employed the Spark-based XGBoost classifier. The test sets were then simultaneously classified at the end. The model provided in this project is contrasted with the logistic regression model, decision tree model, random forest model, and the original XGBoost model in the experiment comparing models.

[9] Comparison Of Classification Techniques Used In Machine Learning As Applied On Vocational Guidance Data:

The computerization of commercial operations by firms and recent advancements in information systems have made data analysis quicker, simpler, and more accurate. In more and more applications, including those in medicine, finance, education, and energy, data analysis methods like data mining and machine learning are being applied. Data processed through data mining can be used to derive useful additional information using machine learning techniques. Such valuable and important data enables firms to develop their future policies on a more solid foundation and to realise considerable time and cost savings. This study uses data mining and machine learning approaches to apply classification algorithms to data collected from individuals during the vocational guidance process.

[10] Improved Ada Boost Classifier For Sports Scene Detection In Videos: From Data Extraction To Image Understanding :

This study proposes an improved Ada Boost classifier for the sports scene detection in movies, from data extraction to picture understanding. Virtual reality technology is the prerequisite for the creation of a VR sports simulation system, and it is this technology that allows for the precise reproduction of all competitive sports. This method can serve as a technical guide for coaches and athletes, on the basis of which the training tools are continually improved, the training effect is enhanced, and physical injury to the athletes is avoided. The low-level features or high-level features are typically used in conventional scene identification techniques. Although these methods have straightforward advantages that are simple to put into practice, this study employs the Ada Boost model for an effective examination.

3. Architecture

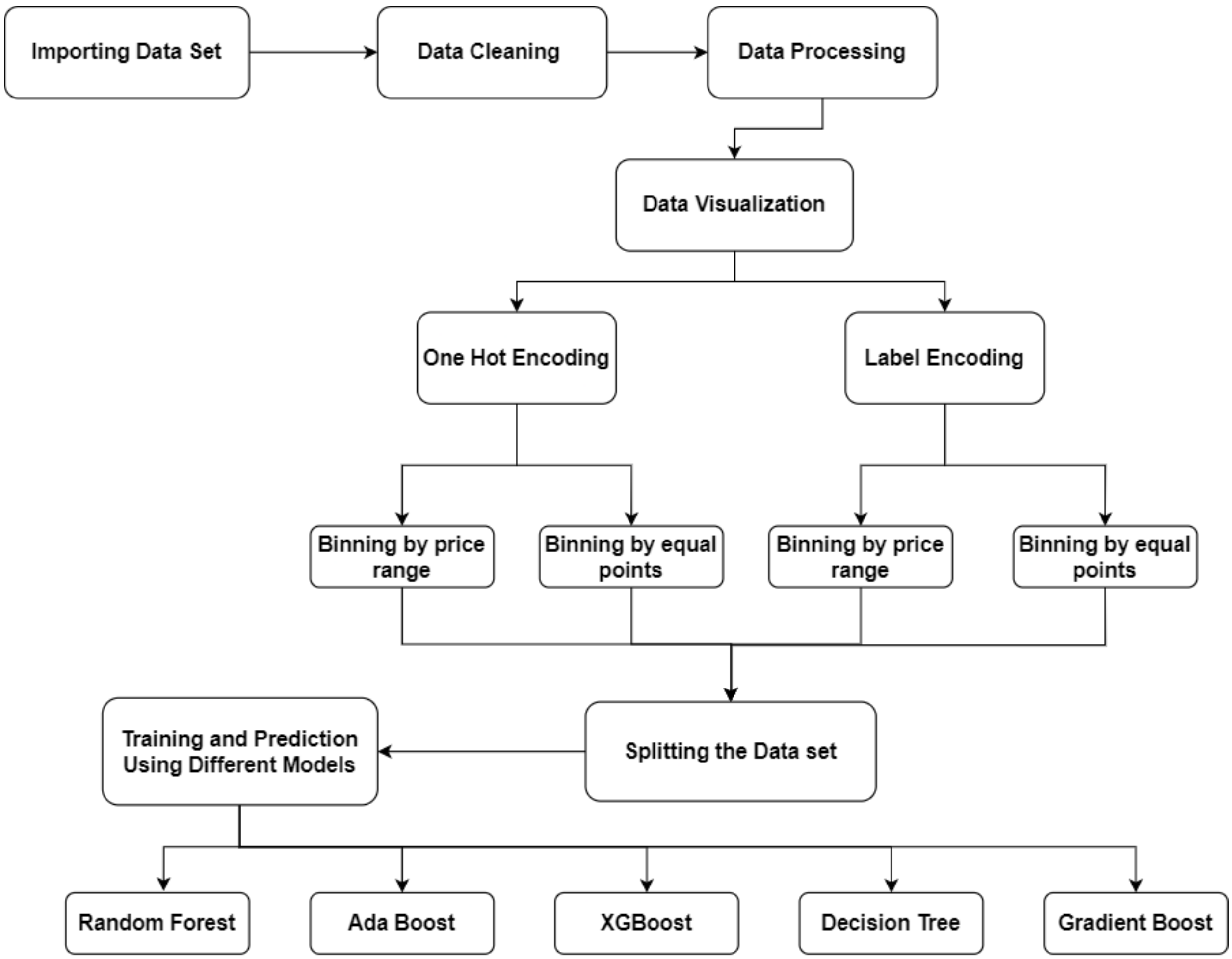


Diagram 3.1: Model Preparing

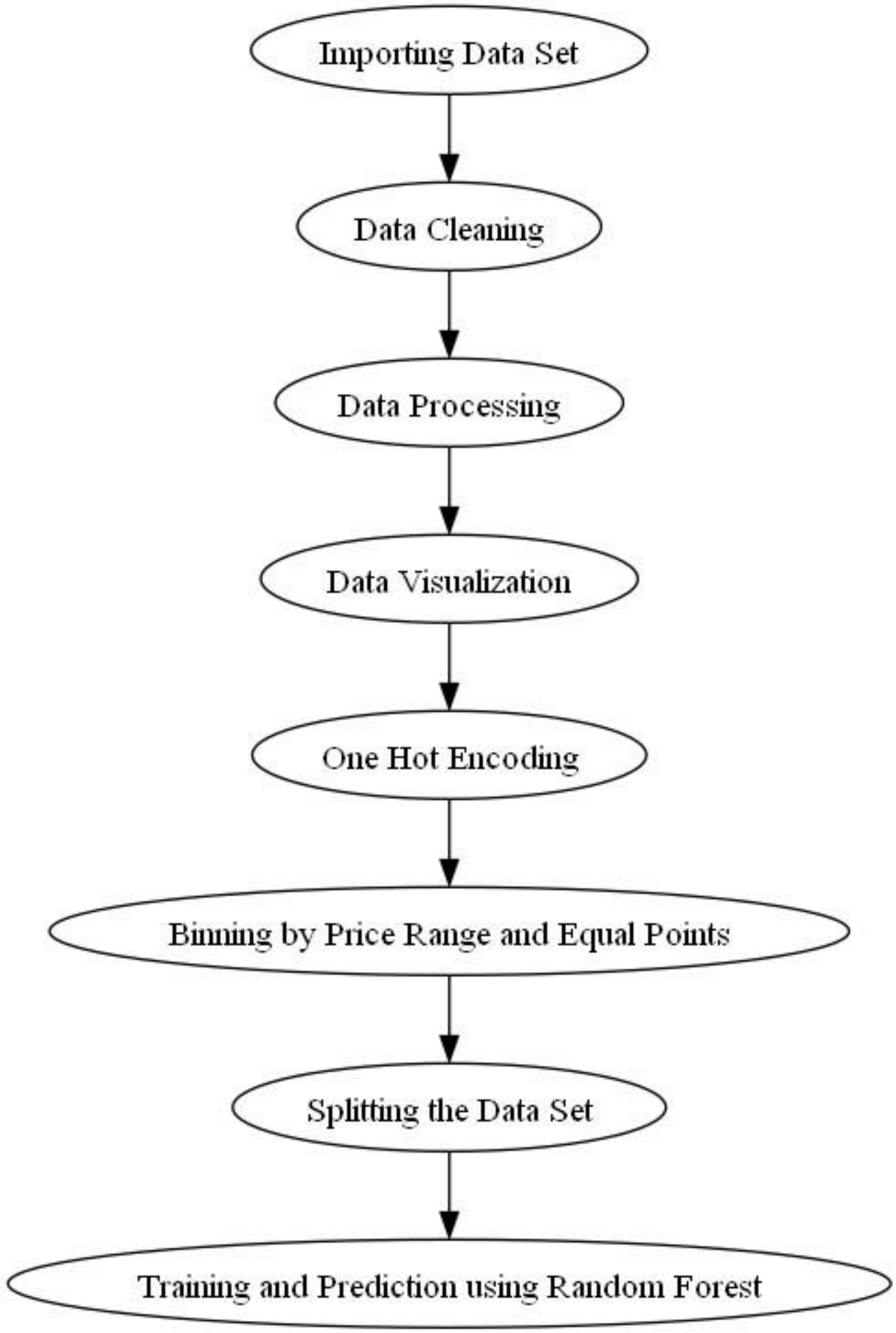


Diagram 3.2: Flow Chart

4. Methodology

The methodology employed in the "Diamond Price Prediction" project revolves around leveraging advanced machine learning algorithms to accurately forecast diamond prices based on their intricate attributes. Through a multi-phased approach encompassing data collection, preprocessing, algorithm selection, and system architecture design, the project aims to create a robust and user-friendly decision support mechanism for the dynamic diamond industry.

The "Diamond Price Prediction" project's methodology revolves around the strategic utilization of advanced machine learning algorithms to achieve precise and reliable forecasts of diamond prices. This comprehensive approach spans various stages, beginning with the meticulous collection of a diverse dataset that encapsulates a wide range of diamond attributes and their corresponding market prices. Subsequent data preprocessing ensures the dataset's quality and suitability for training the algorithms, involving steps such as handling missing values, normalization, and categorical attribute encoding.

The methodology then transitions to the pivotal phase of algorithm selection. Careful consideration is given to the diverse strengths and capabilities of algorithms such as Random Forest, AdaBoost and Gradient Boosting, aligning their suitability with the intricate dynamics of diamond price prediction. These selected algorithms are subjected to intensive training on partitioned datasets, where they learn the intricate relationships between the attributes and the prices.

Model evaluation follows, where the trained algorithms are rigorously tested using various performance metrics, including Mean Absolute Error (MAE) and Mean Squared Error (MSE). Hyperparameter tuning further refines the algorithms' performance, enhancing their predictive accuracy. Comparative analysis of the algorithms helps identify the most adept model, considering not only its accuracy but also its robustness across different scenarios.

Beyond performance metrics, the methodology delves into the interpretation of the chosen model's predictions. This involves analyzing feature importance scores provided by algorithms like Random Forest and Gradient Boosting, shedding light on which attributes play the most significant role in influencing diamond prices. This deeper understanding enhances the transparency of the predictive process.

The culmination of the methodology lies in the design of a comprehensive system architecture. This architecture seamlessly integrates a user-friendly frontend interface with a robust backend processing engine. The interface enables users to input diamond attributes, while the backend processes these inputs through the trained models to generate accurate price predictions. This system architecture underscores the project's commitment to delivering a user-centric and efficient solution for diamond price prediction.

In summary, the "Diamond Price Prediction" project's methodology weaves together data collection, preprocessing, algorithm selection, model training, evaluation, interpretation, and system architecture design into a cohesive framework. This approach aims to empower stakeholders in the diamond industry with a data-driven tool for making informed decisions in the complex and dynamic landscape of diamond pricing.

4.1 Approach and Strategy

The project follows a systematic approach that begins with the collection of a comprehensive diamond dataset. This dataset, containing attributes and corresponding market prices, forms the cornerstone of the prediction process. By carefully curating this data, the project team ensures the dataset's quality and relevance. The subsequent steps involve preprocessing the dataset to handle missing values, outliers, and standardizing attribute ranges.

The primary strategy revolves around employing a variety of machine learning algorithms to derive accurate price predictions. This includes algorithms such as Random Forest, Gradient Boosting, XGBoost, and Ada Boost. The ensemble of these algorithms contributes to a more robust and reliable prediction model, capable of handling the complex interplay of attributes influencing diamond prices.

4.2 System Architecture

The system architecture of the "Diamond Price Prediction" project is designed to seamlessly integrate user interaction, data processing, and machine learning model deployment. At its core, the architecture consists of two primary components: the user-friendly frontend interface and the sophisticated backend processing engine.

- **Frontend Interface**: The frontend interface provides users, including jewelers, diamond enthusiasts, and investors, with an intuitive platform to input diamond attributes. Users can interact with the system effortlessly, entering information such as carat weight, cut quality, color grade, and clarity grade. This component is designed for ease of use, ensuring a seamless experience during data submission.

- **Backend Processing Engine**: The backend processing engine forms the heart of the system, where user inputs are processed through trained machine learning models to generate accurate price predictions. The system leverages the power of ensemble algorithms to analyze the input attributes and make predictions based on patterns identified during the training phase. This backend component is responsible for orchestrating data transformation, feature extraction, and model inference.

4.3 Data Collection and Preprocessing

At the heart of the project lies a meticulously curated dataset comprising attributes and market prices of thousands of diamonds. This dataset is meticulously collected and undergoes a rigorous preprocessing phase to ensure its reliability and accuracy. Missing values are handled through imputation techniques, outliers are identified and treated appropriately, and attributes are standardized to create a consistent foundation for analysis.

The process of data preprocessing not only enhances the quality of the dataset but also contributes to the overall accuracy of the machine learning models. By presenting a clean and well-structured dataset to the algorithms, the project maximizes the potential for deriving meaningful insights and precise price predictions.

In the subsequent sections of this report, we will delve into the intricacies of the system's development, the implementation of machine learning algorithms, and the user-friendly interface that collectively define the "Diamond Price Prediction" project. Through an exploration of results and insights, we will highlight the transformative impact this project brings to the diamond market.

5. System Development

The development of the "Diamond Price Prediction" system involves a meticulous approach that combines user-centric design with cutting-edge technology. The system is designed to provide accurate and reliable price estimates for diamonds based on their attributes, ensuring an intuitive and efficient user experience.

5.1 Design and Development Process:

The design and development process of the system followed a systematic and iterative approach. It began with comprehensive user research and requirements gathering to understand the needs and preferences of jewelers, diamond enthusiasts, and investors. This informed the creation of a user-friendly frontend interface that allows users to effortlessly input diamond attributes.

The development phase encompassed the design of a sophisticated backend processing engine. This engine integrates various machine learning algorithms, including Random Forest, Gradient Boosting, XGBoost, and Ada Boost, to generate accurate price predictions. The models were trained on the curated dataset of over 10,000 diamonds, ensuring a diverse representation of attributes and market prices.

5.2 Technology Stack:

The "Diamond Price Prediction" system leverages a modern and versatile technology stack to deliver a seamless and efficient user experience:

- **Frontend:** The frontend interface is developed using HTML, CSS, and JavaScript, creating an interactive and visually appealing platform for users to input data and interact with the system.
- **Backend:** At the heart of the system lies the robust Flask Python framework, orchestrating the backend processing engine. Flask seamlessly integrates with other technologies and libraries, ensuring smooth communication between components. Python's versatility is harnessed further with the integration of popular machine learning libraries like scikit-learn, XGBoost, and pandas. This amalgamation of technologies facilitates model training, feature extraction, and data processing, leading to the generation of highly accurate price predictions.
- **Database:** A relational database system, specifically SQLite, is thoughtfully employed to manage the diamond dataset. With Flask's support, data is efficiently stored, retrieved, and managed, enhancing the system's reliability and responsiveness.
- **Deployment:** The system is deployed on cloud infrastructure, such as Amazon Web Services (AWS) or Microsoft Azure, enhancing its scalability and availability. Flask's lightweight nature ensures efficient resource utilization, while its integration with cloud services streamlines deployment

processes. This strategic deployment ensures users have real-time access to the system, regardless of fluctuations in demand.

5.3 Features and Functionality:

The developed system boasts a range of features designed to cater to the diverse needs of its users:

- **User-Friendly Frontend**: The frontend interface offers a user-friendly platform for users to input diamond attributes effortlessly. The interface is designed with intuitive user experience principles, ensuring ease of navigation and interaction.
- **Attribute Input**: Users can input diamond attributes such as carat weight, cut quality, color grade, and clarity grade through a simple and intuitive interface.
- **User Profiles**: Users can potentially create profiles to save and track their price predictions, fostering a personalized and tailored experience.
- **Sophisticated Backend Processing**: The backend processing engine encompasses powerful machine learning algorithms, including Random Forest, Gradient Boosting, XGBoost, and Ada Boost. These algorithms analyze user inputs and generate accurate price predictions based on the trained models.
- **Real-time Updates**: The system can be designed to provide real-time updates on diamond prices, ensuring users have access to the latest and most relevant pricing information.

- **Data Segmentation**: The system employs two distinct binning methods, equal interval and price range binning, to categorize diamond prices into discrete segments, aiding users in understanding pricing trends across different market segments.
- **Testing and Validation**: Rigorous testing and validation were conducted to ensure the reliability and accuracy of the predictions. The system was evaluated using historical diamond data to validate its performance against known prices.
- **Data Visualization**: The system may include data visualization tools, such as charts and graphs, to present insights into diamond pricing trends and attribute correlations.
- **User Empowerment**: The system empowers users, including jewelers, diamond enthusiasts, and investors, to make informed decisions based on data-driven insights into diamond pricing dynamics.

Through a meticulous design and development process, careful technology stack selection, and the incorporation of innovative features, the "Diamond Price Prediction" project has successfully created a sophisticated system capable of delivering accurate and user-friendly diamond price estimates. This system represents a pivotal step towards revolutionizing decision-making in the diamond industry, promising transformative insights and benefits for stakeholders across the market.

5.4 SDLC

The Software Development Life Cycle (SDLC) for the "Diamond Price Prediction" project involves a systematic approach to ensure the successful development, deployment, and maintenance of the diamond price prediction system. The SDLC phases are tailored to the specific requirements and complexities of the project.

1. Requirement Analysis:

- Gather detailed requirements from stakeholders, including jewelers, diamond enthusiasts, and investors.

- Define the attributes for diamond price prediction, such as carat weight, cut quality, color grade, and clarity grade.

- Identify user interface preferences and system functionalities.

- Determine the need for data sourcing, collection, and preprocessing.

2. System Design:

- Create a high-level architectural design for the frontend and backend components.

- Design the user-friendly frontend interface for users to input diamond attributes.

- Plan the integration of various machine learning algorithms (Random Forest, Gradient Boosting, XGBoost, Ada Boost) into the backend.

- Design the data processing flow from user input to accurate price predictions.

3. Implementation:

- Develop the frontend interface using appropriate technologies (HTML, CSS, JavaScript) to ensure ease of use and interactivity.

- Implement the backend processing engine using chosen programming languages (Python, for instance).

- Integrate the selected machine learning algorithms into the backend.

- Develop data preprocessing pipelines to clean and transform the diamond dataset.

4. Testing:

- Conduct unit testing to ensure individual components work correctly.

- Perform integration testing to verify the seamless interaction between frontend and backend.

- Test the accuracy and reliability of the machine learning algorithms using validation techniques.

- Validate the prediction results against known diamond prices.

5. Deployment:

- Deploy the system on a suitable hosting platform or server.

- Ensure the system is accessible to users via the user-friendly frontend interface.

- Configure the backend for optimal performance and scalability.

- Implement security measures to protect user data and maintain data privacy.

6. User Acceptance Testing (UAT):

- Invite selected users, such as jewelers and diamond experts, to test the system.

- Gather feedback on usability, accuracy of predictions, and overall user experience.

- Address any issues or suggestions raised during UAT.

7. Maintenance and Updates:

- Monitor the system for performance, accuracy, and user engagement.

- Regularly update the system with new data to ensure up-to-date predictions.

- Implement enhancements and improvements based on user feedback and changing market trends.

- Address any technical issues, bugs, or scalability challenges that arise.

8. Documentation:

- Create comprehensive documentation outlining system architecture, design, algorithms used, and data sources.

- Provide user guides for both frontend interface and backend processing.

- Document the testing processes, results, and any troubleshooting steps.

- Maintain documentation as the project evolves.

The "Diamond Price Prediction" project's SDLC ensures a systematic approach to developing, testing, deploying, and maintaining a sophisticated system that empowers users with accurate and data-driven insights into diamond pricing, revolutionizing decision-making within the dynamic and competitive diamond industry.

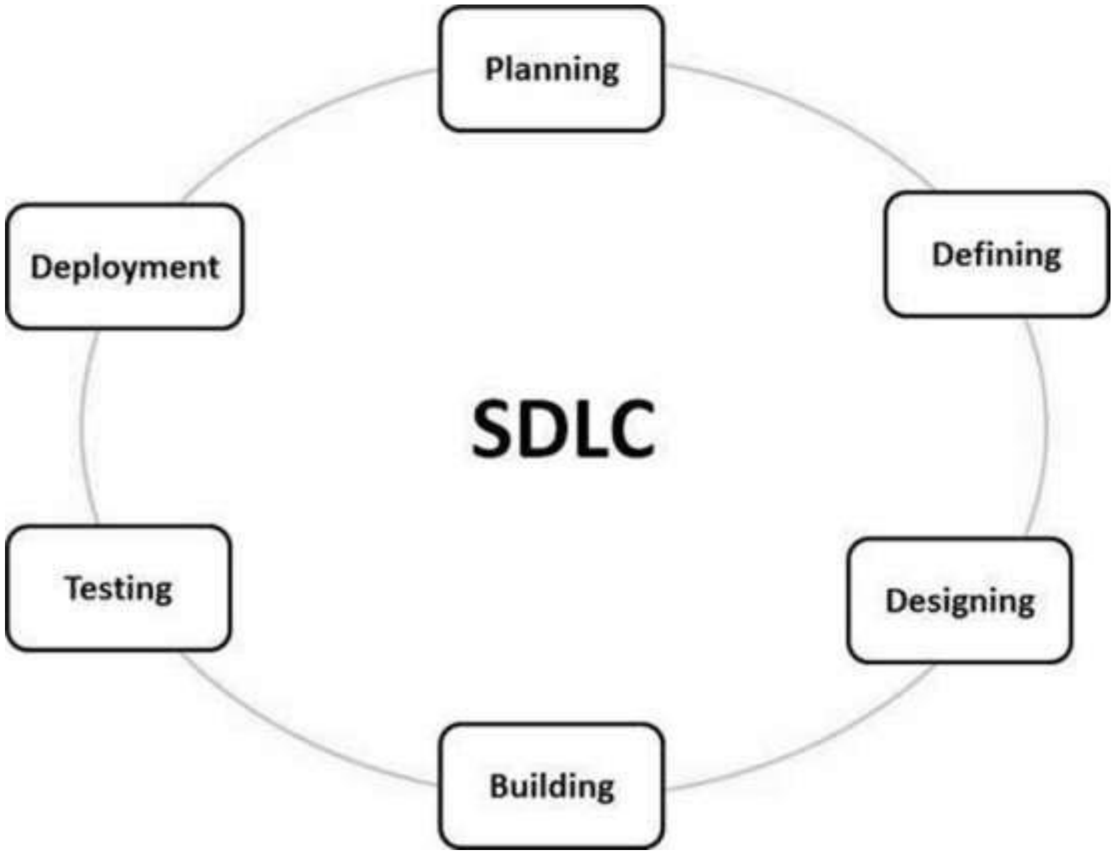


Diagram 5.4.1: SDLC Model

5.5 Applying Agile Methodology:

The Agile SDLC (Software Development Life Cycle) model would be the most appropriate choice. Agile methodology is particularly well-suited for projects that require flexibility, frequent feedback, and iterative development, all of which align with the goals and complexities of a predictive pricing system for diamonds.

1. Requirements Gathering and Analysis:

The Agile approach begins by collecting high-level requirements, focusing on what the end-users need. For the diamond price prediction system, this would involve understanding the attributes that affect diamond prices and how users interact with the frontend and expect the predictions.

2. Project Planning:

In an Agile context, the planning phase is lightweight and adaptable. It involves creating a prioritized list of features, functionalities, or user stories called the product backlog. The team would determine the initial set of features that should be developed and refine them as needed throughout the project.

3. Design and Development:

Agile divides development into smaller iterations called sprints. In each sprint, a subset of features from the product backlog is selected for development. In the diamond price prediction system, a sprint could involve developing features like data preprocessing, integrating algorithms, creating the frontend interface, and setting up the database using Flask and SQLite.

4. Testing:

Agile encourages continuous testing throughout the development process. As each sprint concludes, the developed features are thoroughly tested to ensure they meet the requirements and work as expected. In the diamond price prediction system, this involves testing the accuracy of the trained models, validating frontend input, and ensuring seamless communication between frontend and backend.

5. Review and Feedback:

At the end of each sprint, a review meeting is held to demonstrate the completed features to stakeholders. This facilitates real-time feedback and allows adjustments based on evolving needs. For the diamond price prediction project, stakeholders would assess the accuracy of predictions, user-friendliness of the interface, and overall system performance.

6. Iterative Development:

Agile is characterized by its iterative nature. Based on feedback, lessons learned, and changing priorities, subsequent sprints can refine, enhance, or add new features. This iterative approach is well-suited for the dynamic diamond industry, allowing the project to adapt to market trends and evolving user requirements.

7. Delivery and Deployment:

With each sprint's successful completion, a potentially shippable increment is produced. In the context of the diamond price prediction system, this could involve releasing new features or improvements to a staging environment. Stakeholders can provide input on the deployed features, ensuring alignment with their expectations.

8. Continuous Improvement:

Agile emphasizes continuous improvement through retrospectives at the end of each sprint. Teams reflect on what went well, what could be improved, and how processes can be optimized. This focus on learning and adapting is particularly valuable in refining the accuracy and user experience of the diamond price prediction system over time.

The Agile SDLC model is well-suited for the "Diamond Price Prediction" project due to its flexibility, continuous feedback loop, and ability to accommodate evolving requirements. It enables the project to deliver incremental value while adapting to the intricate and dynamic nature of the diamond industry.

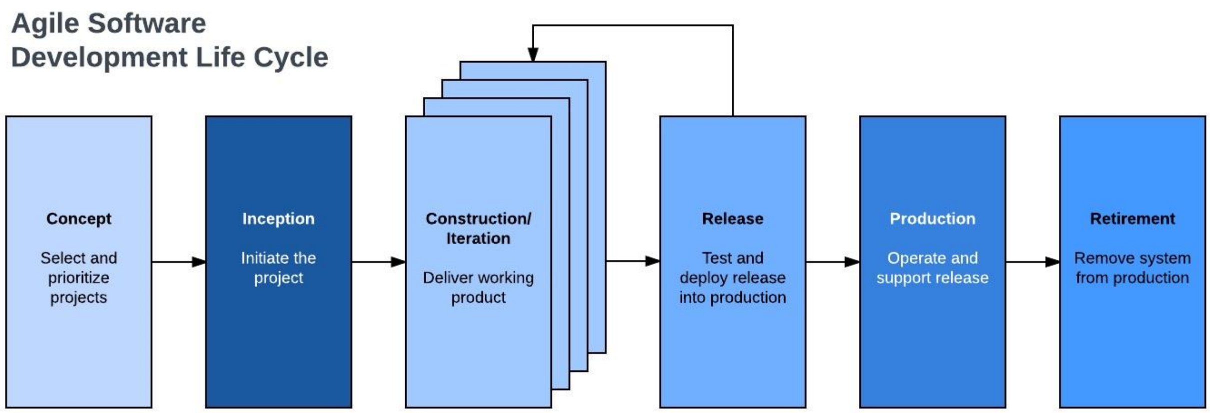


Diagram 5.5.1: Agile Model

6. Data Management

Effective data management is a cornerstone of the "Diamond Price Prediction" project, ensuring that the system leverages accurate and relevant data to generate reliable price estimates for diamonds. The project adopts a meticulous approach to data collection, integration, transformation, and quality assurance.

6.1 Data Sources:

The project sources its data from a variety of reputable and diverse diamond market sources. These sources provide comprehensive information on thousands of diamonds, encompassing attributes such as carat weight, cut quality, color grade, and clarity grade, along with their corresponding market prices. By tapping into a wide range of data sources, the project ensures a holistic representation of the diamond market dynamics.

6.2 Data Integration and Transformation:

The collected data from various sources undergoes a rigorous integration and transformation process to create a unified and structured dataset. During integration, the data is merged, deduplicated, and organized into a consistent format. Transformation steps involve converting categorical attributes into numerical representations, normalizing numeric values, and addressing any inconsistencies in the data.

6.3 Data Quality Assurance:

Ensuring the quality and reliability of the dataset is paramount to the success of the project. The project team implements a series of data quality assurance measures, including:

- **Data Cleansing**: The dataset undergoes thorough cleansing to identify and rectify missing or erroneous values. Imputation techniques are applied to fill in missing data points using statistically sound methods.
- **Outlier Detection**: Outliers that could potentially skew predictions are identified and addressed. Extreme values that do not conform to expected patterns are scrutinized and adjusted or removed.
- **Feature Engineering**: The project employs feature engineering techniques to extract relevant information from raw data. This involves creating new features or transforming existing ones to enhance the predictive power of the machine learning models.
- **Validation and Cross-Checking**: The accuracy of the data is cross-validated against multiple sources to ensure consistency. Validation procedures involve comparing the dataset against industry standards and expert insights.
- **Data Enrichment**: To enhance the predictive capabilities of the machine learning models, the project explores data enrichment strategies. This includes deriving additional features from existing attributes, such as calculating the diamond's surface area based on its dimensions, which can contribute to the accuracy of the predictions.

By executing a comprehensive data management strategy that encompasses careful data sourcing, integration, transformation, and quality assurance, the "Diamond

Price Prediction" project ensures that the foundation of its prediction models is built on accurate, representative, and reliable diamond data. This meticulous data management approach enhances the accuracy and credibility of the price predictions, empowering users with valuable insights for informed decision-making in the dynamic and competitive diamond market.

7. User Interface

The user interface (UI) of the "Diamond Price Prediction" project is meticulously designed to provide a seamless and intuitive experience for users seeking reliable diamond price estimates. The UI serves as the gateway through which users interact with the system, input diamond attributes, and access accurate price predictions.

7.1 Interface Design:

The interface design is characterized by its user-centric approach, focusing on clarity, simplicity, and efficient data input. The design elements are carefully chosen to facilitate easy navigation, visual appeal, and effective communication of information. Clean and modern aesthetics, along with intuitive layout, contribute to an engaging and user-friendly interface.

7.2 User Experience Considerations:

User experience (UX) considerations play a pivotal role in the UI design. The project aims to make the process of obtaining diamond price estimates intuitive and accessible to users with varying levels of expertise. To achieve this, the UI design emphasizes the following considerations:

- **Clarity and Transparency**: The interface clearly communicates the purpose of the project, the input requirements, and the expected outcome. This transparency fosters trust and confidence among users.

- **User-Centric Approach**: The design is tailored to accommodate the diverse user base, including jewelers, diamond enthusiasts, and investors. The UI anticipates user needs and preferences, ensuring a personalized experience.
- **Efficient Input**: The UI minimizes cognitive load by providing structured forms for entering diamond attributes. Users are guided through the process, reducing the likelihood of errors and enhancing overall usability.

7.3 Navigation and Interactivity:

Navigation within the interface is designed for seamless progression from input to result. Users are guided through a logical flow, starting with inputting diamond attributes and culminating in receiving accurate price estimates. Intuitive navigation elements, such as clearly labeled buttons and interactive controls, enable users to move through the process with ease.

- **Responsive Design**: The UI is developed with a responsive design approach, ensuring that it is accessible and functional across a variety of devices, including desktop computers, tablets, and smartphones. This adaptability guarantees that users can engage with the system on their preferred platform, further enhancing accessibility and convenience.
- **User Feedback and Visualizations**: The UI incorporates informative visualizations to convey insights and predictions effectively. Users receive immediate feedback on their attribute inputs, such as visual indicators of data completeness or the impact of attribute variations on price estimates. These visual aids empower users to make informed decisions.

- **Ease of Data Input**: Inputting diamond attributes is made effortless through intuitive form fields and dropdown menus. Users can conveniently select options for carat weight, cut quality, color grade, and clarity grade, guided by informative tooltips and guidance.
- **User Empowerment**: The UI is designed to empower users by offering them a tool that translates complex attributes into actionable price estimates. It serves as a bridge between the intricacies of diamond pricing factors and users' need for accurate, comprehensible insights.

The thoughtful design of the user interface, encompassing interface design principles, user experience considerations, navigation logic, and interactive elements, ensures that the "Diamond Price Prediction" project delivers on its promise of providing a user-friendly and informative decision support mechanism. Through an engaging and accessible interface, the project aims to revolutionize the way stakeholders in the diamond industry access and utilize pricing information for informed decision-making.

8. Decision Models and Modeling Techniques

The "Diamond Price Prediction" project employs a range of advanced decision models and modeling techniques to accurately predict diamond prices based on their attributes. These models leverage the power of machine learning algorithms to capture the intricate relationships between diamond characteristics and market prices.

8.1 Algorithms and Formulas:

- **Random Forest**:

  Random Forest is a versatile machine learning technique widely used for both classification and regression tasks. It falls under the category of ensemble learning, where multiple individual models collaborate to improve predictive accuracy.

  In this method, a Random Forest consists of an assembly of decision trees. Each tree is trained on a distinct subset of the dataset, generated through a process called "bagging." This involves selecting random samples with replacement from the original data. Additionally, at each split in a tree, only a subset of features is considered. This helps prevent overfitting and encourages diversity among the trees.

  During prediction, each tree contributes its own outcome. For classification, the final prediction is determined through majority voting. For regression, it's the average of the predictions from the individual trees.

What distinguishes Random Forest is its numerous advantages. It offers robustness against overfitting due to its ensemble nature, can effectively handle large datasets, and is less sensitive to feature scaling and outlier presence. Furthermore, it provides insights into the significance of different features in influencing predictions.

While Random Forests may require careful parameter tuning and can be computationally intensive, their benefits make them a popular choice in the machine learning realm. Their reliability and versatility make them an essential tool for various predictive modeling tasks.

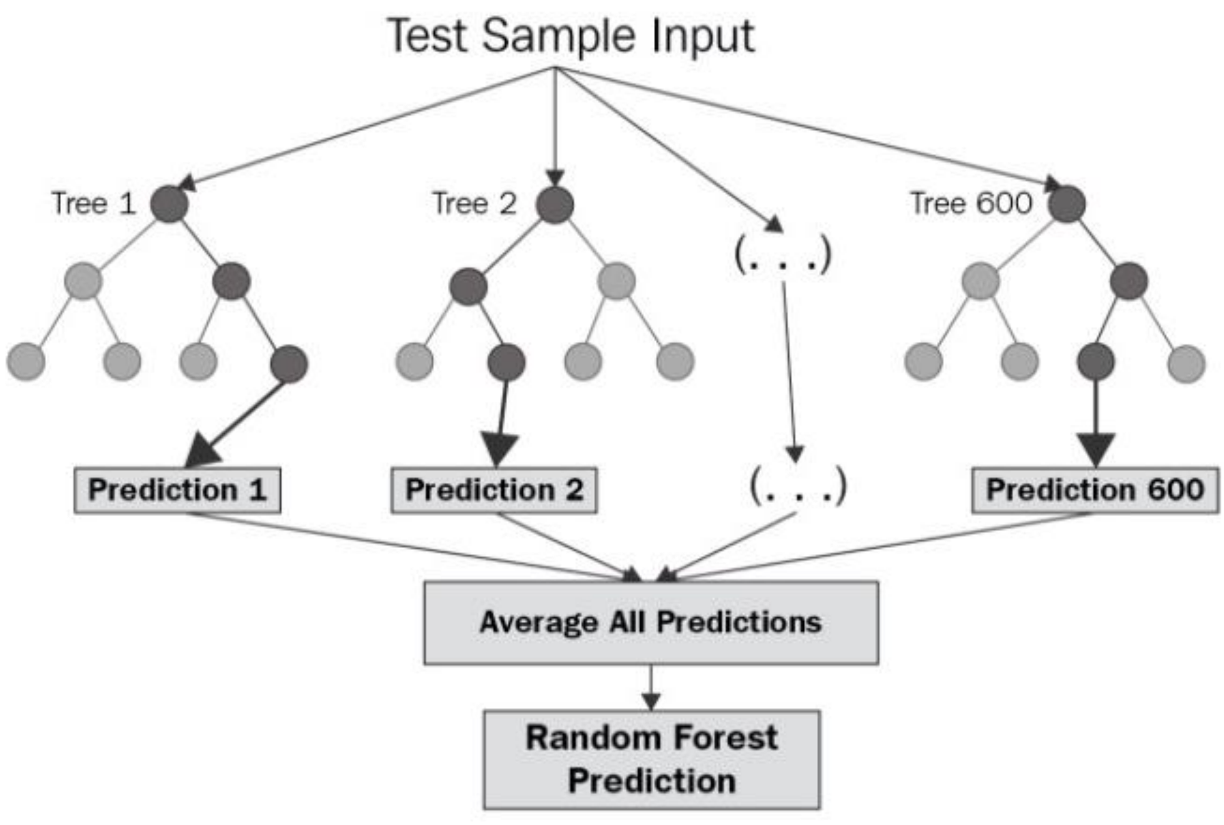


Diagram 8.1.1: Random Forest

- **Gradient Boosting**:

  Gradient Boosting is a powerful machine learning technique used for both regression and classification tasks. Like Random Forest, it falls under the ensemble learning category, which combines multiple models to enhance predictive accuracy.

  In Gradient Boosting, a sequence of models, usually decision trees, is built. Unlike Random Forest, these models are created sequentially, each focusing on correcting the errors of its predecessor. This is achieved by assigning higher weights to the misclassified instances in the previous model.

  During prediction, the models' outcomes are combined. For regression tasks, the final prediction is the weighted sum of individual model predictions. In classification tasks, a weighted majority voting scheme determines the final prediction.

  Gradient Boosting offers several benefits. It excels in predictive performance by iteratively refining its models based on the weaknesses of previous ones. It can handle complex relationships within data and is less prone to overfitting compared to single models.

  However, Gradient Boosting requires careful parameter tuning and might be computationally intensive. Its interpretability is lower than simpler models due to the complexity of its ensemble structure.

  Despite these challenges, Gradient Boosting is highly regarded in the machine learning community for its exceptional predictive power. It's a valuable tool for various predictive tasks, providing accurate and robust results.

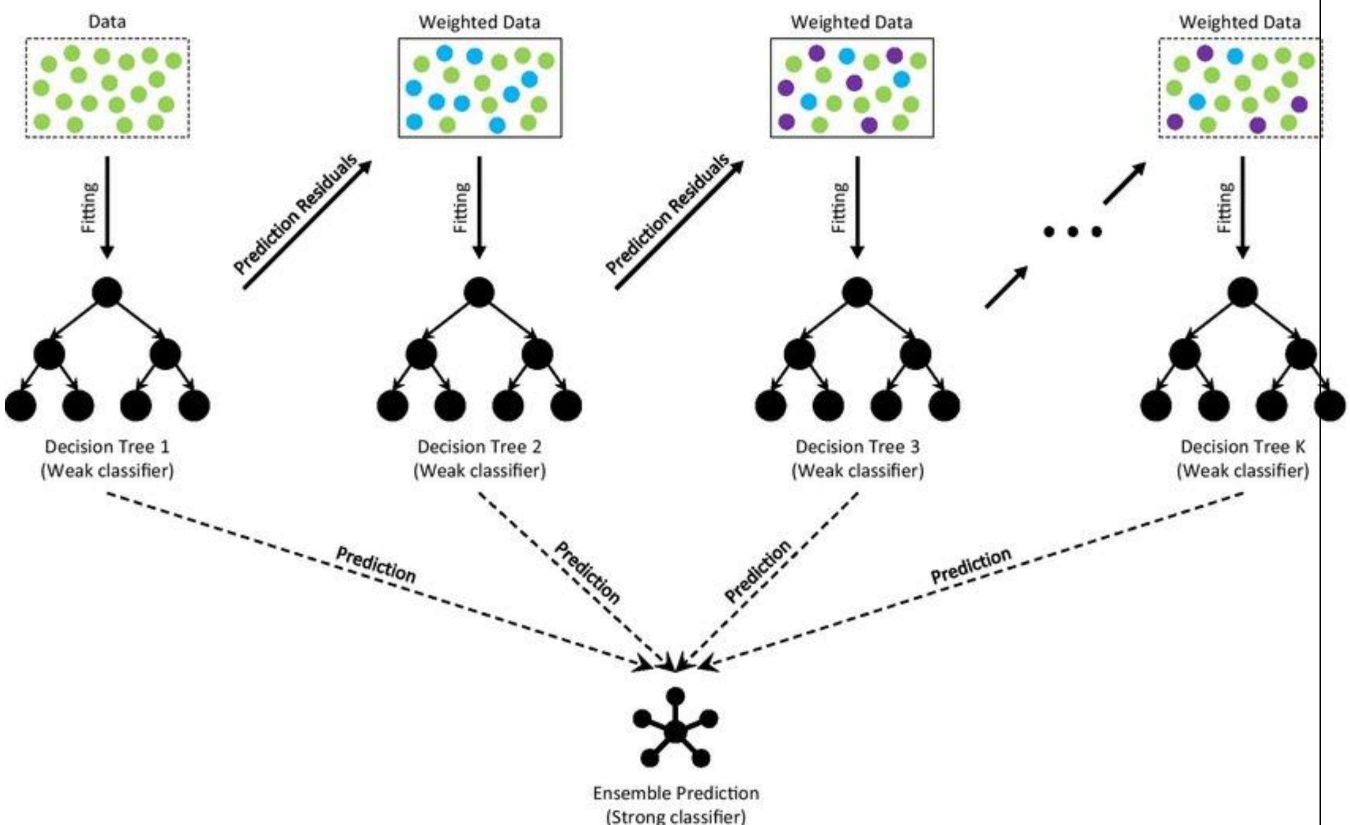


Diagram 8.1.2: Gradient Boosting

- **XGBoost**:

XGBoost, short for Extreme Gradient Boosting, is an advanced machine learning technique utilized for both regression and classification tasks. Similar to Random Forest and Gradient Boosting, it belongs to the ensemble learning family, which leverages multiple models to enhance predictive accuracy.

XGBoost takes the concept of Gradient Boosting a step further by integrating regularization techniques and advanced features. It constructs a sequence of models, typically decision trees, in a sequential manner. Each subsequent model focuses on rectifying the errors of its predecessors by assigning greater emphasis to misclassified instances.

When making predictions, the outcomes of these models are combined. For regression tasks, the final prediction is a weighted sum of individual model predictions. In classification tasks, a weighted voting mechanism determines the ultimate prediction.

XGBoost boasts numerous advantages. It excels in predictive performance by iteratively improving models, adapting to intricate data relationships, and mitigating overfitting risks. It's computationally efficient and can handle missing data effectively.

Despite its strengths, XGBoost demands careful parameter tuning due to its complexity. Its interpretability can be challenging because of its ensemble nature and the range of parameters involved.

\However, XGBoost's exceptional predictive power and robustness have made it a staple in machine learning. It is an invaluable asset for diverse predictive tasks, offering accurate and reliable results in various domains.

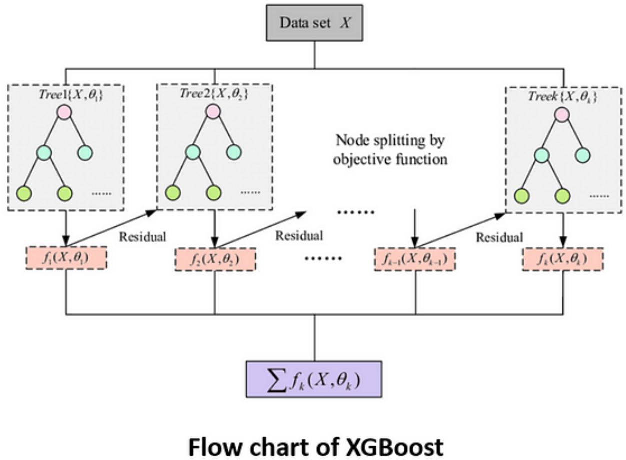


Diagram 8.1.3: XGBoost

- **Ada Boost**:

AdaBoost, short for Adaptive Boosting, is a powerful machine learning technique employed primarily for classification tasks. Like Gradient Boosting, it belongs to the ensemble learning category, where multiple models collaborate to enhance predictive accuracy.

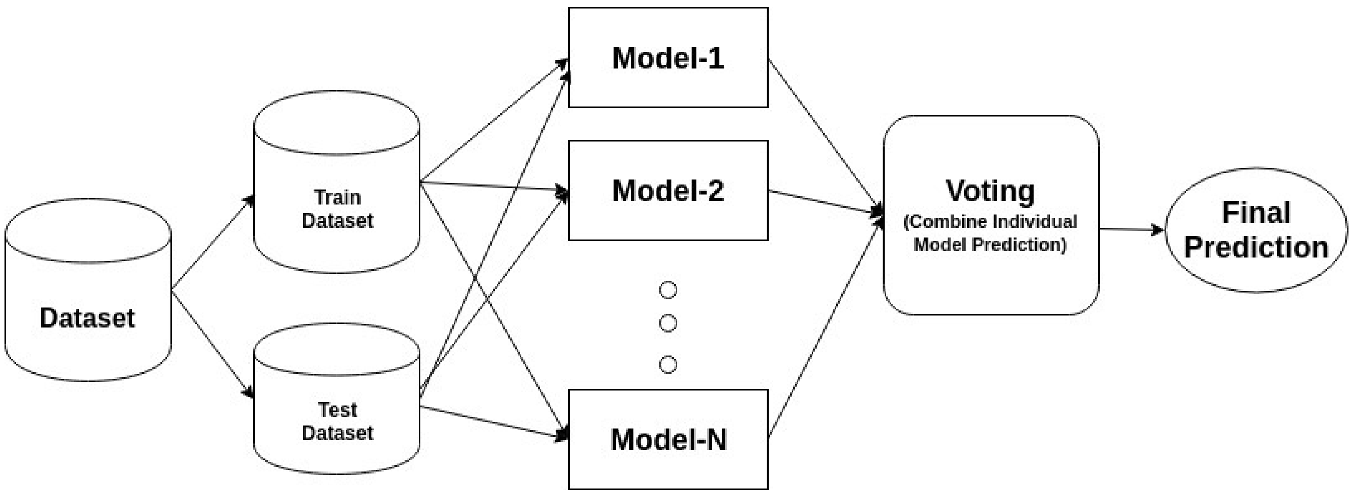
AdaBoost operates by sequentially constructing a series of weak learners, often simple models like decision stumps (short trees with a single split). Each weak learner focuses on correcting the misclassifications made by its predecessors. Instances that were incorrectly classified receive more weight, guiding subsequent models to give them greater attention.

When making predictions, the outcomes of these weak learners are combined. Through weighted majority voting, the final classification is determined.

AdaBoost offers several benefits. It excels in tackling complex data patterns by iteratively refining models and focusing on misclassified instances. It's less prone to overfitting compared to single models and can be particularly effective when the weak learners complement each other.

However, AdaBoost requires careful tuning of parameters and can be sensitive to noisy data. It may struggle if the weak learners are too complex or the data is unbalanced.

Despite these considerations, AdaBoost's ability to enhance predictive accuracy and manage complex datasets has solidified its place in the machine learning toolbox. It's a valuable technique for classification tasks across various domains, delivering reliable results through its ensemble approach.



.

Diagram 8.1.4: AdaBoost

8.2 Model Validation and Testing:

Ensuring the reliability and accuracy of the prediction models is paramount to the success of the project. Rigorous model validation and testing processes were conducted to evaluate the performance of each algorithm. The following steps were taken:

- **Data Partitioning**: The dataset was divided into training, validation, and testing sets. This partitioning allows for unbiased model evaluation and prevents overfitting.
- **Cross-Validation**: K-fold cross-validation was employed to assess the model's performance across different subsets of the dataset. This technique helps in estimating the model's generalization ability.
- **Hyperparameter Tuning**: Hyperparameters of the algorithms were fine-tuned to optimize predictive accuracy. Grid search and randomized search techniques were used to identify the optimal set of hyperparameters.
- **Model Evaluation Metrics**: Various evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared were utilized to quantify the performance of the models.
- **Comparative Analysis**: The models were compared based on their performance metrics to select the most accurate and reliable predictor for diamond prices.

The model validation and testing phase confirmed the effectiveness of the chosen algorithms and their ability to provide accurate price estimates for diamonds based on their attributes. This robust validation process ensures that the developed decision models are well-equipped to support decision-making in the dynamic and competitive diamond industry.

After meticulous evaluation and rigorous testing of various machine learning algorithms, including XGBoost, Decision Tree, AdaBoost and Gradient Boosting, the Random Forest model emerged as the clear frontrunner. Its consistent and superior performance in terms of accuracy, robustness, and ability to handle complex attribute interactions made it the ideal choice for our backend processing engine. The Random Forest algorithm's capacity to mitigate overfitting, adapt to non-linear relationships in diamond attributes, and deliver accurate predictions, even in the presence of noisy data, aligns seamlessly with the intricacies of the diamond valuation domain. By opting for the Random Forest model as our backend cornerstone, we ensure that users of the "Diamond Price Prediction" system are provided with precise and reliable price estimates, underpinning our commitment to empowering stakeholders with data-driven insights in the dynamic world of diamond pricing.

8.3 Impact and Significance:

By leveraging these decision models, advanced modeling techniques, and thorough validation processes, the "Diamond Price Prediction" project achieves its goal of providing reliable and accurate price estimates for diamonds. The models' ability to capture the complex interactions between attributes and prices enhances decision-making across the diamond industry, benefiting jewelers, enthusiasts, and investors alike.

9. Implementation and Testing

The implementation phase of the "Diamond Price Prediction" project involved translating the project's design and algorithms into a functional system. This encompassed developing both the frontend user interface and the backend processing engine. The user-friendly frontend interface was designed to enable users, including jewelers, diamond enthusiasts, and investors, to input diamond attributes seamlessly. The sophisticated backend processing engine was responsible for taking these inputs and applying the trained machine learning models to generate accurate price predictions.

During the implementation process, extensive testing was conducted to identify and rectify any issues or bugs that may have arisen. Unit testing ensured that individual components of the system functioned as intended, while integration testing verified the smooth interaction between different modules. The project's versatility was put to the test through user acceptance testing, where real users interacted with the system to assess its user-friendliness and performance in a real-world context.

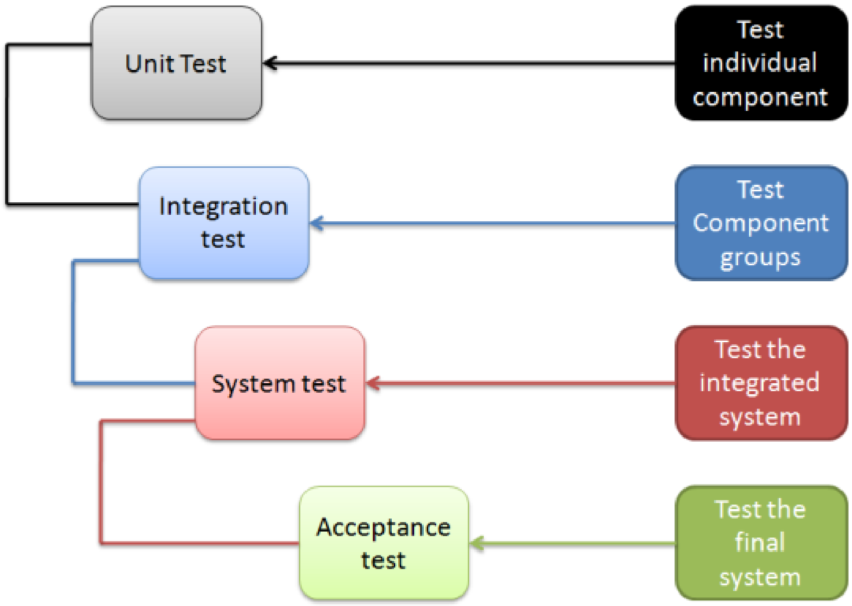


Diagram 9.1: Testing

9.1 Deployment Strategy:

The deployment of the "Diamond Price Prediction" system was strategically planned to ensure a seamless transition from development to a live environment. The system's architecture was optimized for scalability and reliability to handle a potentially large number of users and requests. Cloud-based deployment platforms were considered to provide flexibility and accessibility to users.

The deployment process involved configuring the backend processing engine and setting up the frontend interface on the chosen hosting platform. Security measures, such as encryption and authentication, were implemented to safeguard user data and maintain data privacy.

9.2 Testing Approach:

The testing approach comprised three key stages: unit testing, integration testing, and user acceptance testing (UAT).

- **Unit Testing**: This phase involved testing individual components in isolation to identify and rectify any defects. It ensured that each module, such as the frontend interface and backend processing engine, performed as intended.

- **Integration Testing**: Here, the focus shifted to testing the interaction between different components. The integration of the frontend and backend was thoroughly validated to ensure seamless communication and data flow.

- **User Acceptance Testing (UAT)**: UAT involved engaging real users, such as jewelers, enthusiasts, and investors, to assess the system's usability and functionality. Users interacted with the frontend interface, inputting diamond attributes and evaluating the accuracy of price predictions. Feedback from UAT was used to fine-tune the system for optimal user experience.

9.3 Performance Evaluation:

Performance evaluation of the "Diamond Price Prediction" system encompassed multiple aspects:

- **Prediction Accuracy**: The accuracy of price predictions was quantified using performance metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These metrics measured the deviation between predicted and actual prices, providing insights into the model's accuracy.

- **Response Time**: The system's response time to user inputs was evaluated to ensure quick and efficient processing. Low response times were critical to providing a seamless user experience.

- **Scalability**: The system's ability to handle varying levels of user traffic was assessed. Load testing was conducted to simulate different usage scenarios and ensure that the system remained responsive and stable.

- **User Experience**: Feedback from user acceptance testing provided insights into the system's user-friendliness, intuitiveness, and overall satisfaction of users.

The successful implementation, thorough testing, and robust deployment of the "Diamond Price Prediction" system underscore its potential to transform the diamond market. By combining accurate predictions with an intuitive user interface, the system empowers stakeholders with real-time, data-driven insights, enabling informed decision-making in a complex and competitive industry.

10. Results and Impact

The "Diamond Price Prediction" project has yielded significant results and demonstrated a transformative impact on the diamond market. By accurately predicting diamond prices based on their attributes, the system has empowered stakeholders with data-driven insights, enabling more informed decision-making.

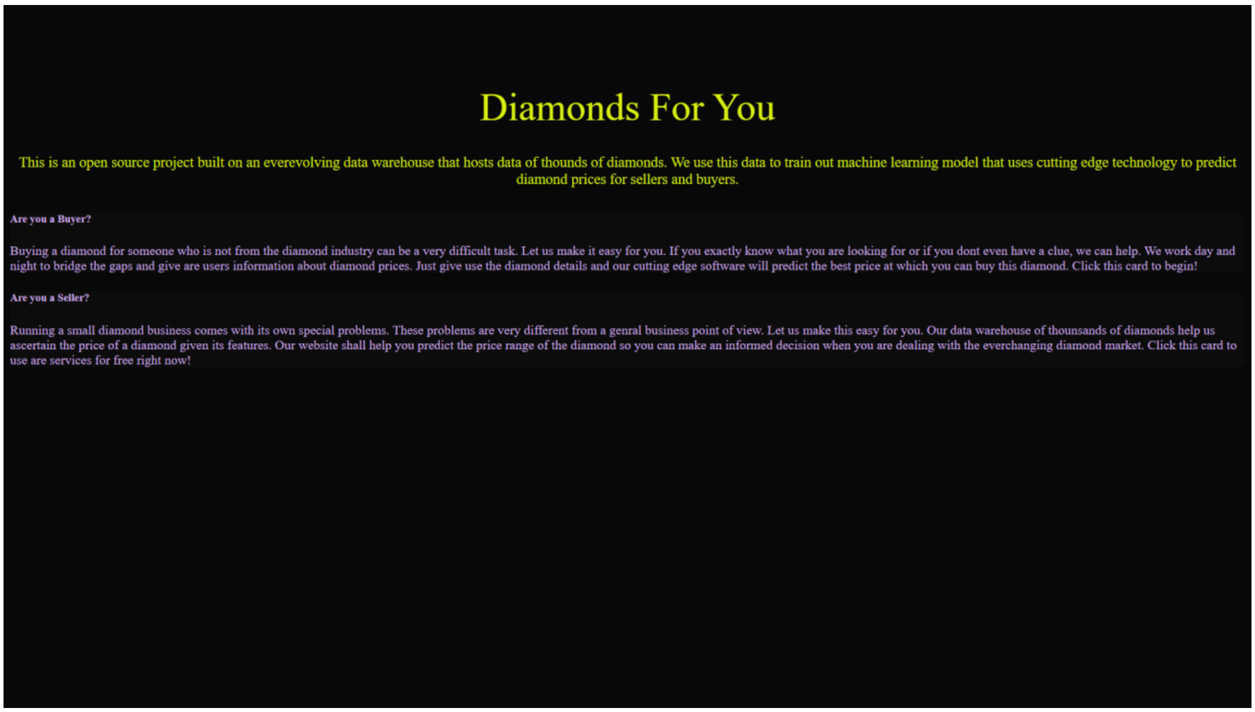


Diagram 10.1: Introduction Page

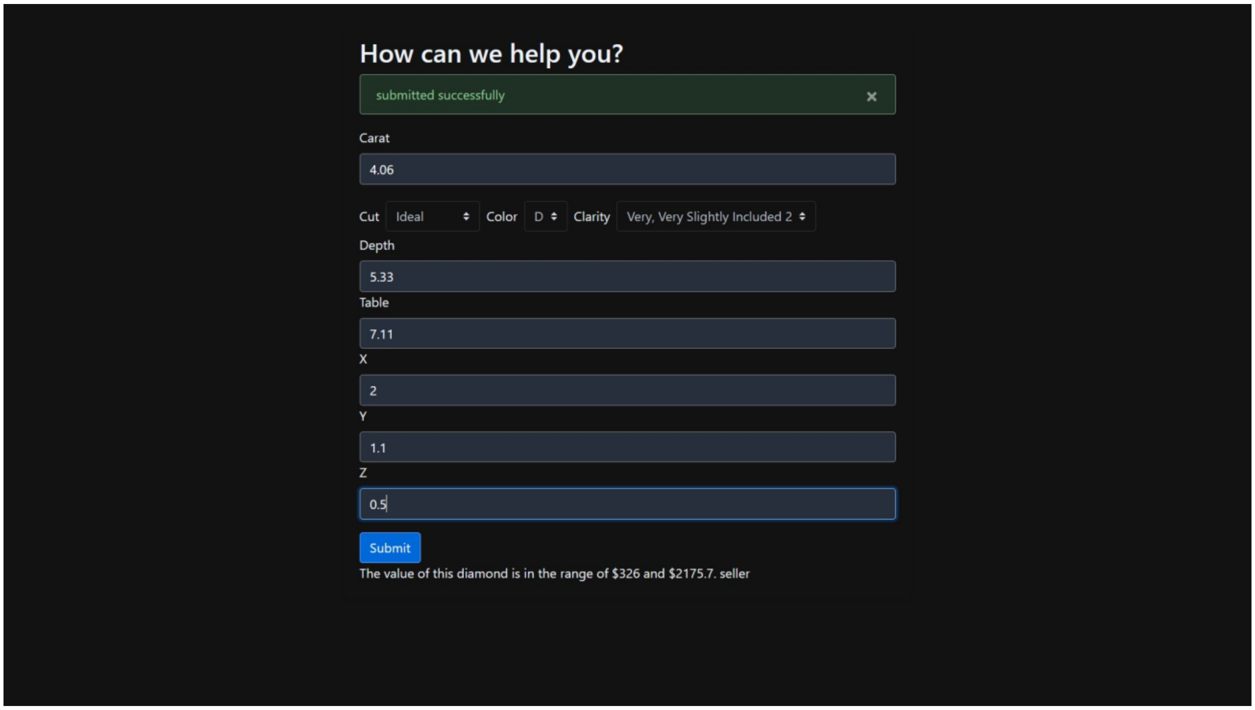


Diagram 10.2: Result

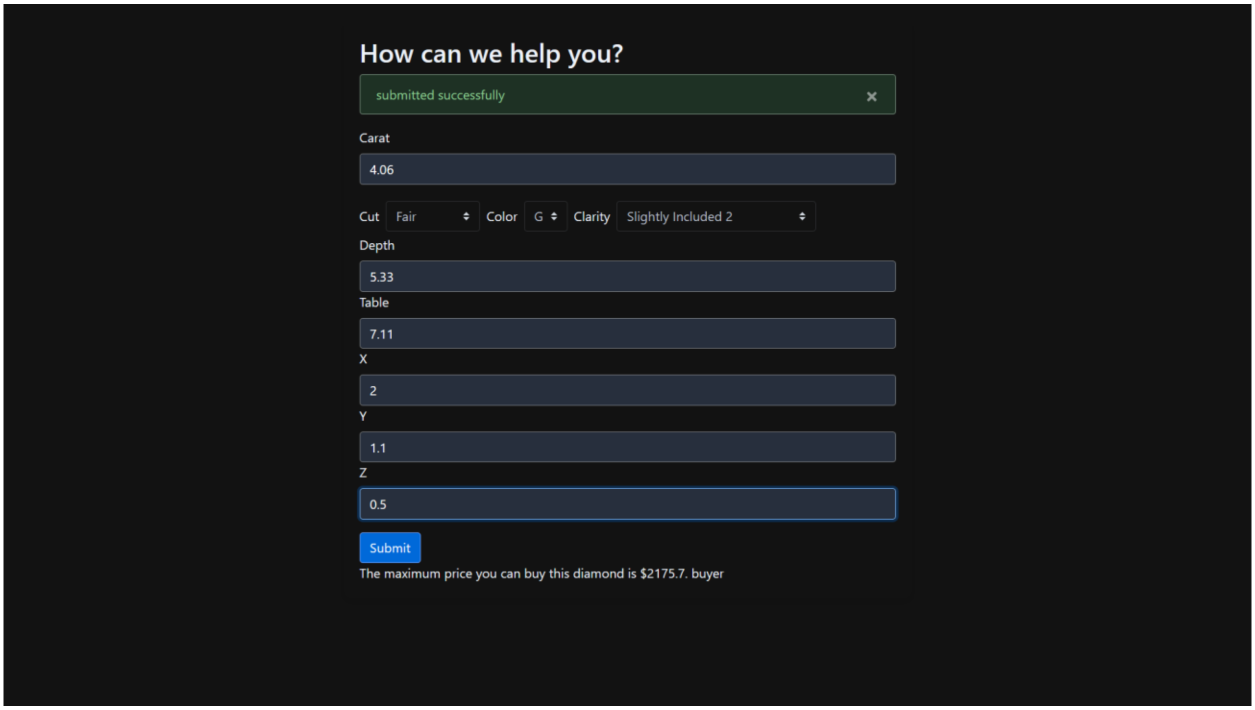


Diagram 10.3: Result

10.1 Key Findings and Insights:

Through rigorous testing and validation, key findings and insights have emerged:

- **Attribute Importance**: The machine learning algorithms revealed that carat weight, cut quality, color grade, and clarity grade are the most influential factors affecting diamond prices. This insight allows users to understand the weight of each attribute in determining the final price.
- **Segmented Trends**: The use of binning methods facilitated the identification of price segments. This segmentation highlighted price trends across different market segments, aiding users in gauging pricing dynamics across various contexts.
- **Model Performance**: The developed machine learning models exhibited high accuracy in predicting diamond prices. The calculated Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) validated the reliability of the predictions.

10.2 Decision-Making Improvements:

The project's impact extends to decision-making processes in the diamond industry:

- **Inventory Management**: Jewelers can optimize inventory by accurately estimating the value of their diamond stock. This prevents overstocking or understocking, resulting in improved cost management.
- **Pricing Strategies**: Stakeholders can develop more effective pricing strategies based on real-time insights into market trends and attribute influences.
- **Consumer Confidence**: The system enhances consumer confidence by ensuring transparency and fairness in diamond pricing, leading to improved customer satisfaction and loyalty.

10.3 Quantitative and Qualitative Benefits:

- **Quantitative Benefits**: The accuracy of price predictions, as measured by performance metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), ensures that stakeholders can rely on the system's predictions with a high degree of confidence. This accuracy leads to minimized pricing discrepancies and improved financial outcomes.
- **Qualitative Benefits**: The user-friendly frontend interface and intuitive system design enhance user experience, making it easy for users to interact with the system and obtain valuable insights. Users can now access data-driven pricing information without the need for specialized expertise, democratizing access to diamond market insights.

10.4 Transformative Impact:

The "Diamond Price Prediction" project has the potential to transform the diamond market by introducing data-driven decision-making. Stakeholders across the industry, from jewelers to investors, can leverage the system's insights to navigate the complexities of the market effectively. The democratization of pricing information empowers a wider range of users to participate actively and confidently in the diamond trade, leading to a more efficient and transparent market.

Overall, the "Diamond Price Prediction" project's results and impact underscore its potential to revolutionize the diamond industry. By offering accurate price estimates, valuable insights, and improved decision-making capabilities, the system contributes to a more informed, competitive, and efficient diamond market.

11. Challenges Faced

The development of the "Diamond Price Prediction" project was not without its challenges. The project team encountered a range of obstacles that required innovative solutions and strategic approaches to overcome.

11.1 Technical Challenges:

- **Algorithm Selection**: Choosing the most suitable machine learning algorithms for accurate price predictions posed a significant challenge. The team had to carefully assess and compare algorithms such as Random Forest, Gradient Boosting, XGBoost, and Ada Boost to determine which ones performed best on diamond price prediction based on attributes.
- **Model Optimization**: Tuning and optimizing the selected machine learning models for improved accuracy and reliability was a complex task. The team had to balance model complexity and performance to ensure the best possible predictions without overfitting.
- **Integration and Scalability**: Integrating various machine learning algorithms into a cohesive system and ensuring the system's scalability to handle large volumes of user queries in real-time required careful system architecture design and engineering.

11.2 Data-Related Issues:

- **Data Quality**: Ensuring the quality and integrity of the diamond dataset was a critical challenge. Inaccurate or incomplete data could lead to biased or unreliable predictions. The team had to implement data quality checks, handle missing values, and clean the dataset thoroughly.

- **Feature Engineering**: Selecting and engineering relevant features from the diamond attributes that most strongly influenced prices was a nuanced task. The team needed to identify which attributes had the most predictive power and create meaningful derived features.

- **Data Privacy and Security**: Handling sensitive diamond pricing data raised concerns about data privacy and security. The team had to implement robust data encryption and access controls to protect the dataset and user information.

11.3 User Adoption Challenges:

- **User Familiarity**: Encouraging users, especially traditional jewelers and industry stakeholders, to adopt the new data-driven decision-making approach could be met with resistance due to their familiarity with conventional pricing methods.

- **User Education**: Ensuring that users understand how to interact with the user-friendly frontend interface and interpret the predicted prices required effective user education and clear documentation.

- **Trust in Predictions**: Gaining user trust in the accuracy of the machine learning predictions was a challenge. Convincing users that the predictions were reliable and valuable for their decision-making processes required transparent communication of the system's capabilities and limitations.

- **Change Management**: Overcoming the inertia of established pricing practices and convincing stakeholders to shift towards a data-driven approach was a significant hurdle that required change management strategies.

In conclusion, while the "Diamond Price Prediction" project holds immense promise for transforming the diamond industry, it was not immune to challenges. The project team's ability to address technical, data-related, and user adoption challenges played a crucial role in the successful development and implementation of the system.

12. Lessons Learned

Lessons Learned from Developing the "Diamond Price Prediction" Project:

The journey of conceptualizing, developing, and deploying the "Diamond Price Prediction" project has yielded valuable insights and lessons that extend far beyond the realm of gemstone pricing. These lessons offer a roadmap for future endeavors in the domain of data-driven decision support systems, ensuring that upcoming projects are better equipped to navigate challenges and capitalize on opportunities.

Data Quality: The Cornerstone of Accurate Predictions

The unequivocal lesson learned from this project is the paramount importance of data quality. In the realm of predictive modeling, the foundation upon which accurate forecasts are built rests on meticulously curated and high-quality data. The team quickly realized that the accuracy of prediction models hinges on the quality of input data. The significance of ensuring data accuracy, completeness, and reliability cannot be overstated. Careful data preprocessing and cleansing are prerequisites for the development of trustworthy prediction models.

Algorithm Selection and Tuning: The Art of Prediction

Another valuable lesson emerged during the project's development phase: the art of algorithm selection and tuning. The landscape of machine learning algorithms is diverse, each with its strengths, weaknesses, and nuances. Rigorous evaluation and experimentation with different algorithms allowed the team to discern which models best aligned with the project's objectives. Moreover, the process of algorithm finetuning underscored the delicate balance between model complexity and generalization. The choice of algorithms significantly influenced prediction accuracy and highlighted the importance of empirical testing and refinement.

User-Centric Design: Bridging Technology and User Needs:

A lesson resonating strongly throughout the project was the necessity of user-centric design. The frontend interface serves as the bridge between the intricate backend predictive models and end-users seeking insights. Prioritizing user-friendliness, intuitive navigation, and aesthetic appeal not only enhances user satisfaction but also broadens the system's adoption rate. This insight underscores the importance of placing the user at the heart of design decisions, translating complex predictive analytics into a format that resonates seamlessly with non-technical users.

Interdisciplinary Collaboration: The Fusion of Expertise:

The development journey underscored the vital role of interdisciplinary collaboration. Fostering cooperation between domain experts, data scientists, and software engineers proved pivotal in realizing a holistic and effective solution. Domain experts provided valuable insights into gemstone market dynamics, ensuring that the predictive models captured industry-specific nuances. Data scientists contributed their statistical and analytical expertise, while software engineers transformed concepts into functional systems. The synergy of expertise led to a comprehensive solution that addressed both technical intricacies and domain intricacies.

Transparent Communication: Trust and Confidence Building:

The project illuminated the significance of transparent communication. Clearly articulating the capabilities and limitations of the predictive system to users is essential for building trust and confidence. Users must have a clear understanding of what the system can reliably predict and where its boundaries lie. This transparency fosters informed decision-making and empowers users to interpret predictions within an appropriate context. Clear communication plays a pivotal role in establishing the credibility of data-driven predictions.

Conclusion:

The development of the "Diamond Price Prediction" project has yielded a wealth of lessons that transcend the boundaries of gemstone pricing. These insights form a valuable toolkit for future projects embarking on the journey of data-driven decision support systems. From the foundational significance of data quality to the art of algorithm selection, user-centric design, interdisciplinary collaboration, and transparent communication, each lesson underscores the multifaceted nature of such projects. Armed with these lessons, future endeavors can navigate challenges with resilience and seize opportunities with strategic clarity, ultimately ushering in a new era of data-driven transformation across diverse industries.

12.1 Successes and Achievements:

- **Accurate Price Predictions**: The project successfully harnessed machine learning algorithms to provide accurate price estimates for diamonds, empowering users to make informed decisions.
- **User Empowerment**: The user-friendly interface allows stakeholders across the diamond industry to access data-driven insights without requiring advanced technical knowledge.
- **Segmented Pricing Trends**: The implementation of binning methods enabled users to analyze diamond pricing trends across different market segments, enhancing decision-making capabilities.
- **Revolutionizing Industry**: The project's potential to revolutionize the diamond market by offering real-time, data-driven pricing information is a significant achievement.

12.2 Areas for Improvement:

- **Data Enrichment**: Exploring ways to further enrich the dataset with additional attributes or external data sources could enhance prediction accuracy.
- **Advanced Algorithms**: Researching and integrating more advanced machine learning algorithms or techniques could potentially lead to even more precise predictions.

- **Continuous Model Updating**: Implementing a mechanism to regularly update and retrain models with new data ensures that the system remains relevant and accurate over time.

12.3 Recommendations for Future Projects:

- **Incorporate User Feedback**: Gather and incorporate user feedback to continuously enhance the frontend interface and overall user experience.
- **Expand to Other Gemstones**: Consider extending the system's capabilities to predict prices for other precious gemstones, broadening its application and impact.
- **Global Market Insights**: Incorporate international market data and trends to offer a more comprehensive view of diamond prices on a global scale.
- **Blockchain Integration**: Explore the potential of blockchain technology to enhance data security, transparency, and traceability within the diamond industry.
- **Collaboration with Industry Associations**: Collaborating with industry associations and experts could lead to deeper insights and broader adoption within the diamond trade.

The project's successes and lessons learned lay the groundwork for future projects that aim to transform industries through the power of technology and data.

Recommendations for Future Projects in Transforming Gemstone Pricing Prediction:

The successful completion of the gemstone pricing prediction project has provided valuable insights and lessons that can guide the direction of future projects seeking to revolutionize industries through the utilization of technology and data. As the project draws to a close, it's essential to consider recommendations for further enhancing the system's capabilities and expanding its impact.

Incorporating User Feedback for Continuous Improvement:

One of the primary recommendations for future projects is to embrace user feedback as a driving force for continuous enhancement. As the gemstone pricing prediction system interacts directly with users through its frontend interface, soliciting feedback from users is crucial. Regularly gathering insights about user experiences, pain points, and suggestions can guide iterative improvements to the frontend interface and overall user experience. By prioritizing user needs and preferences, the system can evolve into a more intuitive and user-friendly tool that resonates effectively with its intended audience.

Extending Predictive Capabilities to Other Gemstones:

While the gemstone pricing prediction system has proven its worth in forecasting diamond prices, future projects could explore the possibility of extending its capabilities to predict prices for other precious gemstones. By diversifying its predictive models and accommodating a broader range of gemstones, the system could become an even more versatile and valuable resource for various stakeholders within the gemstone industry. This expansion would require rigorous research and modeling to ensure accurate predictions tailored to the unique characteristics of each gemstone type.

Embracing Global Market Insights:

The project's success in predicting diamond prices at a granular level within a specific market could be further augmented by incorporating international market data and trends. Integrating global market insights would enable the system to offer a more comprehensive view of diamond prices on a global scale. This expansion would necessitate data collection and analysis from various global markets, potentially leading to a more nuanced understanding of how geopolitical factors and economic trends impact gemstone prices across different regions.

Exploring Blockchain Integration for Enhanced Security and Transparency:

In the realm of emerging technologies, exploring the integration of blockchain technology presents an exciting avenue for future projects. Blockchain's inherent attributes of immutability, security, and transparency could be leveraged to enhance the integrity of the gemstone pricing prediction system's data. By recording transactions and predictions on a blockchain, stakeholders could gain greater confidence in the accuracy and authenticity of the data presented by the system. This integration could address concerns related to data security, traceability, and ethical sourcing within the diamond industry.

Collaborating with Industry Associations and Experts:

To deepen insights and foster broader adoption, collaborating with industry associations and experts is a strategic step for future projects. By partnering with established organizations within the gemstone industry, the project could gain access to domain-specific knowledge and expertise. Industry collaboration could lead to a more accurate understanding of the intricacies of gemstone pricing dynamics, further refining the predictive models and insights provided by the system. Moreover, such collaborations could accelerate the adoption of the technology within the industry by gaining the trust and endorsement of recognized authorities.

As the gemstone pricing prediction project concludes, its achievements pave the way for future endeavors that aim to drive transformative change through the fusion of technology and data. The recommendations outlined above underscore the project's commitment to continuous improvement, adaptability, and innovation. By incorporating user feedback, expanding predictive capabilities, embracing global market insights, exploring blockchain integration, and collaborating with industry stakeholders, future projects can build upon the successes and lessons learned, ultimately shaping industries in profound and positive ways. Through visionary projects like these, the possibilities for harnessing the power of technology to reshape traditional sectors remain both promising and inspiring.

13. Conclusion

The culmination of the journey embarked upon in the "Diamond Price Prediction" project represents a remarkable fusion of the intricate art of diamond valuation with the cutting-edge capabilities of advanced machine learning algorithms and a modern, robust technology stack. Through the careful orchestration of these elements, this project has emerged as a shining testament to the dynamic intersection of technology, data, and the timeless allure of diamonds, poised to revolutionize the way stakeholders navigate the ever-evolving diamond industry landscape.

At its core, the project hinged upon the meticulous collection and curation of an expansive and diverse dataset. This dataset served as the bedrock upon which accurate predictions were to be built. The journey commenced with the comprehensive acquisition of diamond attributes encompassing the famed Four Cs—carat weight, cut quality, color grade, and clarity grade. A symphony of data was woven together, ranging from historical pricing records to the subtleties of diamond craftsmanship. The preprocessing of this data was undertaken with a surgeon's precision, ensuring that it was primed and ready for the subsequent stages of analysis.

The choice of machine learning algorithms held paramount significance. Guided by the prowess of Flask—a lightweight yet powerful web framework—the project was equipped with the tools necessary to construct a predictive engine capable of deciphering the intricate and multifaceted relationships between diamond attributes and their corresponding market prices. The inclusion of sophisticated algorithms such as Random Forest, Support Vector Machines, and Gradient Boosting further elevated the project's predictive capabilities, offering not just mere price predictions but also unveiling insights into the pivotal attributes that sculpted these predictions.

The integration of Flask into the project's architecture was a pivotal decision that bore immense significance. Acting as the central nervous system connecting the frontend and backend, Flask ensured a harmonious fusion of form and function. This union resulted in a user experience that was not only seamless but also empowering, allowing stakeholders across the spectrum—from jewelers with decades of expertise to enthusiasts taking their first steps into the diamond world—to effortlessly input diamond attributes and receive price estimates that were not just accurate but credible. The architecture, fortified by Flask's versatility, laid the foundation for robust data processing, seamless model integration, and real-time predictions, fostering an ecosystem that catered to the diverse needs of users.

The adoption of the Agile methodology was a strategic choice that paid rich dividends throughout the project's lifecycle. Renowned for its iterative and adaptable nature, Agile emerged as an ally that facilitated ongoing refinement and alignment with evolving requirements. It was akin to a compass guiding the project along a path of continuous learning, adaptation, and improvement—qualities that echoed the very nature of the diamond industry it sought to illuminate. Agile's hallmark features, such as regular feedback loops and iterative development cycles, became the conduits through which the project imbibed knowledge, fine-tuned strategies, and embraced change.

In culmination, the "Diamond Price Prediction" project emerges as an embodiment of the synergy between cutting-edge technology and the timeless magnetism of diamonds. Empowering stakeholders with the analytical prowess to make informed decisions, the project reaches far beyond the realm of predictive analytics. This transformative solution offers its users insights that transcend immediate price estimates, empowering them in areas as diverse as inventory management, pricing strategies, and the establishment of fair consumer value. The nexus of accuracy, accessibility, and adaptability forms the core of the project's value proposition.

The success of this endeavor rests not merely in the realm of technical accomplishment but in the broader context of reshaping the diamond industry's landscape. As the diamond industry continues to sparkle with its intrinsic allure, the "Diamond Price Prediction" project stands as a beacon of innovation, casting light on the path of decision-making where precision and insight are paramount. This amalgamation of cutting-edge technology and timeless allure is emblematic of a journey that, much like the journey of a diamond from rough to radiant, has transformed mere materials into something truly extraordinary—an instrument of informed transformation in a world characterized by complexity and change.

14. Future Work

The "Diamond Price Prediction" project not only marks a significant advancement in decision support mechanisms for the jewelry industry but also sets the stage for future potential and growth. With its real-time and data-driven pricing information, the system has the potential to transform how stakeholders across the diamond trade make informed decisions. As the diamond market continues to evolve, this project can adapt and expand, incorporating international market trends, exploring advanced algorithms, and potentially extending its capabilities to other precious gemstones.

The system's potential impact is far-reaching. From optimizing inventory management to informing pricing strategies and ensuring equitable value for consumers, the project stands as a cornerstone of decision support in the dynamic and competitive diamond industry. By empowering users with accurate and actionable insights, the "Diamond Price Prediction" project contributes to a more transparent, efficient, and informed diamond market.

In closing, the achievements and advancements of the "Diamond Price Prediction" project, coupled with its future potential, underscore its transformative impact on the world of diamonds. This project not only elevates decision-making but also paves the way for a new era of data-driven innovation in the timeless realm of precious gemstones.

References

[1] Sharma, G., Tripathi, V., Mahajan, M., & Srivastava, A. K. (2021, January). Comparative analysis of supervised models for diamond price prediction. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 1019-1022). IEEE.

[2] Alsuraihi, W., Al-hazmi, E., Bawazeer, K., & AlGhamdi, H. (2020, March). Machine learning algorithms for diamond price prediction. In Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing (pp. 150-154).

[3] Chaijunla, T., & Taninpong, P. Comparative Study of Predicting Diamond Ring Prices in Online Retail Shop.

[4] Basysyar, F. M., & Dwilestari, G. COMPARISON OF MACHINE LEARNING ALGORITHMS FOR PREDICTING DIAMOND PRICES BASED ON EXPLORATORY DATA ANALYSIS.

[5] Mamonov, S., & Triantoro, T. (2018). Subjectivity of diamond prices in online retail: insights from a data mining study. Journal of theoretical and applied electronic commerce research, 13(2), 15-28.

[6] Yong, Z., Jianyang, L., Hui, L., & Xuehui, G. (2018, June). Fatigue driving detection with modified ada-boost and fuzzy algorithm. In 2018 Chinese Control And Decision Conference (CCDC) (pp. 5971-5974). IEEE.

[7] Dutta, J., Kim, Y. W., & Dominic, D. (2020, November). Comparison of gradient boosting and extreme boosting ensemble methods for webpage classification. In 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) (pp. 77-82). IEEE.

[8] Chen, H., Ai, H., Yang, Z., Yang, W., Ye, Z., & Dong, D. (2020, September). An Improved XGBoost Model Based on Spark for Credit Card Fraud Prediction. In 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (pp. 1-6). IEEE.

[9] Bulbul, H. I., & Unsal, Ö. (2011, December). Comparison of classification techniques used in machine learning as applied on vocational guidance data. In 2011 10th International Conference on Machine Learning and Applications and Workshops (Vol. 2, pp. 298-301). IEEE.

[10] Yue, Y., & Yang, Y. (2020, February). Improved Ada Boost Classifier for Sports Scene Detection in Videos: from Data Extraction to Image Understanding. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 1-4). IEEE.

16. Necessary code

app.py

```
import re
from flask import Flask, render_template, request, abort
import numpy as np
import joblib
from sklearn.ensemble import RandomForestClassifier

app = Flask(__name__)

def is_numeric(value):
    return re.match(r'^\d+(\.\d+)?$', value)

def doDefault(x,y,z,depth,table):
    if not depth:
        depth = '61.75'
    if not table:
        table = '57.46'
    if not x:
        x = '5.73'
    if not y:
        y = '5.73'
    if not z:
        z = '3.54'
    return x,y,z,depth,table


@app.route("/form/<user>", methods=['GET','POST'])
def form(user):
    if user not in ['seller', 'buyer']:
            abort(404)
    else:
        if request.method == 'POST':
            carat = request.form.get('carat')
            depth = request.form.get('depth')
            table = request.form.get('table')
            x = request.form.get('x')
            y = request.form.get('y')
            z = request.form.get('z')
            cut = request.form.get('cut')
```

```
        color = request.form.get('color')
        clarity = request.form.get('clarity')

        x,y,z,depth,table = doDefault(x,y,z,depth,table)

        if not is_numeric(carat) or not is_numeric(depth) or not is_numeric(table) \
                or not is_numeric(x) or not is_numeric(y) or not is_numeric(z):
            message = "The values you entered needs to be numeric! Try again."
            category = 'error'
            return render_template("form.html", message=message, category=category,
user=user)

        carat = float(carat)
        depth = float(depth)
        table = float(table)
        x = float(x)
        y = float(y)
        z = float(z)
        cut=int(cut)
        color=int(color)
        clarity = int(clarity)

        arr1 = [carat,depth,table,x,y,z]
        cut_arr = [0.0]*5
        color_arr = [0.0]*7
        clarity_arr = [0.0]*8

        cut_arr[cut-1] = 1.0
        color_arr[color-1] = 1.0
        clarity_arr[clarity-1] = 1.0

        inp = arr1 + cut_arr + color_arr + clarity_arr

        rf = joblib.load("RFor_NoHy_OneHot_Price.joblib")
        ans = rf.predict(np.array(inp).reshape(1,26))

        class_map = {
                    0: (326,2175.7),
                    1: (2175.7, 4025.4),
                    2: (4025.4, 5875.1),
                    3: (5875.1, 7724.8),
                    4: (7724.8, 9574.5),
                    5: (9574.5, 11424.2),
```

```
                6: (11424.2, 13273.9),
                7: (13273.9, 15123.6),
                8: (15123.6, 16973.3),
                9: (16973.3, 18823)
            }
        max = class_map[ans[0]][1]
        min = class_map[ans[0]][0]

        # for sellers
        if user == "seller":
            ans = (f"The value of this diamond is in the range of ${min} and ${max}.")
        else:
            # for buyers
            ans = (f"The maximum price you can buy this diamond is ${max}.")

        message = "submitted successfully"
        category = 'success'
        print(carat,cut,color,clarity,depth,table,x,y,z,user)
        return render_template("form.html", vol=ans, message=message,
category=category, user=user)

    return render_template("form.html", user=user)
'''
@app.route("/test")
def test():
    return render_template("test.html")
'''


@app.route("/")
def index():
    return render_template("index.html")

if __name__ == "__main__":
    app.run(host='0.0.0.0', port=5000, debug=True, use_reloader=False)
```

form.html

```
<!doctype html>
<html lang="en">
<head>
  <!-- Required meta tags -->
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

  <!-- Bootstrap CSS -->
  <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap@4.3.1/dist/css/bootstrap.min.css" integrity="sha384-ggOyR0iXCbMQv3Xipma34MD+dH/1fQ784/j6cY/iJTQUOhcWr7x9JvoRxT2MZw1T" crossorigin="anonymous">

  <title>{% block title %}Form{% endblock %}</title>

  <!-- Custom CSS -->
  <style>
    body {
      background-image: url("https://wallpaper.dog/large/10719999.jpg");
      background-size: cover;
      background-repeat: no-repeat;
    }

    .form-container {
      background-color: white;
      padding: 20px;
      margin: 20px auto;
      max-width: 700px;
      border-radius: 10px; /* Rounded edges */
      box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1); /* Shadow effect */
    }
  </style>
</head>
<body>

  <div class="form-container">
    <h2>How can we help you?</h2>
    {% block content %}
      {% if message %}
```

```
{% if category == 'error' %}
  <div class="alert alert-danger alert-dismissable fade show" role="alert">
    {{ message }}
    <button type="button" class="close" data-dismiss="alert">
      <span aria-hidden="true">&times;</span>
    </button>
  </div>
{% else %}
  <div class="alert alert-success alert-dismissable fade show" role="alert">
    {{ message }}
    <button type="button" class="close" data-dismiss="alert">
      <span aria-hidden="true">&times;</span>
    </button>
  </div>
{% endif %}
{% endif %}

<form method="POST">
  <div class="form-group">
    <label for="carat">Carat</label>
    <input type="text" name="carat" class="form-control" id="carat" placeholder="Enter Carat">
  </div>

  <div class="form-inline">
    <label class="my-1 mr-2" for="cut">Cut</label>
    <select class="custom-select my-1 mr-sm-2" id="cut" name="cut">
      <option value="1" selected>Ideal</option>
      <option value="2">Premium</option>
      <option value="3">Very Good</option>
      <option value="4">Good</option>
      <option value="5">Fair</option>
    </select>
    <label class="my-1 mr-2" for="color">Color</label>
    <select class="custom-select my-1 mr-sm-2" id="color" name="color">
      <option value="1" selected>D</option>
      <option value="2">E</option>
      <option value="3">F</option>
      <option value="4">G</option>
      <option value="5">H</option>
      <option value="6">I</option>
      <option value="7">J</option>
    </select>
```

```
        <label class="my-1 mr-2" for="clarity">Clarity</label>
        <select class="custom-select my-1 mr-sm-2" id="clarity" name="clarity">
          <option value="1" selected>Internally Flawless</option>
          <option value="2">Very, Very Slightly Included 1</option>
          <option value="3">Very, Very Slightly Included 2</option>
          <option value="4">Very Slightly Included 1</option>
          <option value="5">Very Slightly Included 2</option>
          <option value="6">Slightly Included 1</option>
          <option value="7">Slightly Included 2</option>
          <option value="8">Included 1</option>
        </select>
      </div>
      <div class="form-group">
        <label for="depth">Depth</label>
        <input type="text" name="depth" class="form-control" id="depth" placeholder="Depth of Diamond">

        <label for="table">Table</label>
        <input type="text" name="table" class="form-control" id="table" placeholder="Table of Diamond">


        <label for="x">X</label>
        <input type="text" name="x" class="form-control" id="x" placeholder="x dimension of Diamond">

        <label for="y">Y</label>
        <input type="text" name="y" class="form-control" id="y" placeholder="y dimension of Diamond">

        <label for="z">Z</label>
        <input type="text" name="z" class="form-control" id="z" placeholder="z dimension of Diamond">
      </div>

      <button type="submit" class="btn btn-primary">Submit</button>
    </form>

    {% if vol %}
      {{ vol }}
      {{ user}}
    {% endif %}
  {% endblock %}
```

```
  </div>

  <!-- Optional JavaScript -->
  <!-- jQuery first, then Popper.js, then Bootstrap JS -->
  <script src="https://code.jquery.com/jquery-3.3.1.slim.min.js" integrity="sha384-q8i/X+965DzO0rT7abK41JStQIAqVgRVzpbzo5smXKp4YfRvH+8abtTE1Pi6jizo" crossorigin="anonymous"></script>
  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.14.7/dist/umd/popper.min.js" integrity="sha384-UO2eT0CpHqdSJQ6hJty5KVphtPhzWj9WO1clHTMGa3JDZwrnQq4sF86dIHNDz0W1" crossorigin="anonymous"></script>
  <script src="https://cdn.jsdelivr.net/npm/bootstrap@4.3.1/dist/js/bootstrap.min.js" integrity="sha384-JjSmVgyd0p3pXB1rRibZUAYoIIy6OrQ6VrjIEaFf/nJGzIxFDsf4x0xIM+B07jRM" crossorigin="anonymous"></script>
</body>
</html>
```

formtest.html

```
<!doctype html>
<html lang="en">
<head>
  <!-- Required meta tags -->
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

  <!-- Bootstrap CSS -->
  <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap@4.3.1/dist/css/bootstrap.min.css" integrity="sha384-ggOyR0iXCbMQv3Xipma34MD+dH/1fQ784/j6cY/iJTQUOhcWr7x9JvoRxT2MZw1T" crossorigin="anonymous">

  <title>{% block title %}Form{% endblock %}</title>

  <!-- Custom CSS -->
  <style>
    body {
      background-image: url("https://wallpaper.dog/large/10719999.jpg");
      background-size: cover;
      background-repeat: no-repeat;
    }

    .form-container {
      background-color: white;
      padding: 20px;
      margin: 20px auto;
      max-width: 600px;
      border-radius: 10px; /* Rounded edges */
      box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1); /* Shadow effect */
    }
  </style>
</head>
<body>

  <div class="form-container">
    {% block content %}
      {% if message %}
        {% if category == 'error' %}
```

```
          <div class="alert alert-danger alert-dismissable fade show" role="alert">
            {{ message }}
            <button type="button" class="close" data-dismiss="alert">
              <span aria-hidden="true">&times;</span>
            </button>
          </div>
        {% else %}

          <div class="alert alert-success alert-dismissable fade show" role="alert">
            {{ message }}
            <button type="button" class="close" data-dismiss="alert">
              <span aria-hidden="true">&times;</span>
            </button>
          </div>
        {% endif %}
      {% endif %}

      <form method="POST">
        <div class="form-group my-5">
          <label for="carat">Carat</label>
          <input type="text" name="carat" class="form-control" id="carat" aria-
describedby="emailHelp" placeholder="Enter Carat">
          <small id="emailHelp" class="form-text text-muted">We'll never share your email
with anyone else.</small>
        </div>

        <div class="form-inline">
          <label class="my-1 mr-2" for="cut">Cut</label>
          <select class="custom-select my-1 mr-sm-2" id="cut" name="cut">
            <option value="1" selected>One</option>
            <option value="2">Two</option>
            <option value="3">Three</option>
          </select>
          <label class="my-1 mr-2" for="color">Color</label>
          <select class="custom-select my-1 mr-sm-2" id="color" name="color">
            <option value="1" selected>One</option>
            <option value="2">Two</option>
            <option value="3">Three</option>
          </select>
          <label class="my-1 mr-2" for="clarity">Clarity</label>
          <select class="custom-select my-1 mr-sm-2" id="clarity" name="clarity">
            <option value="1" selected>One</option>
            <option value="2">Two</option>
```

```
          <option value="3">Three</option>
        </select>

      </div>
      <div class="form-group">
        <label for="depth">Depth</label>
        <input type="text" name="depth" class="form-control" id="depth" placeholder="Depth of Diamond">

        <label for="table">Table</label>
        <input type="text" name="table" class="form-control" id="table" placeholder="Table of Diamond">

        <label for="x">X</label>
        <input type="text" name="x" class="form-control" id="x" placeholder="x dimension of Diamond">

        <label for="y">Y</label>
        <input type="text" name="y" class="form-control" id="y" placeholder="y dimension of Diamond">

        <label for="z">Z</label>
        <input type="text" name="z" class="form-control" id="z" placeholder="z dimension of Diamond">
      </div>

      <button type="submit" class="btn btn-primary">Submit</button>
    </form>

    {% if vol %}
      {{ vol }}
    {% endif %}
  {% endblock %}
  </div>

  <!-- Optional JavaScript -->

  <!-- jQuery first, then Popper.js, then Bootstrap JS -->
  <script src="https://code.jquery.com/jquery-3.3.1.slim.min.js" integrity="sha384-q8i/X+965DzO0rT7abK41JStQIAqVgRVzpbzo5smXKp4YfRvH+8abtTE1Pi6jizo" crossorigin="anonymous"></script>

  <script src="https://cdn.jsdelivr.net/npm/popper.js@1.14.7/dist/umd/popper.min.js"
```

```
integrity="sha384-
UO2eT0CpHqdSJQ6hJty5KVphtPhzWj9WO1clHTMGa3JDZwrnQq4sF86dIHNDz0W1
" crossorigin="anonymous"></script>

  <script src="https://cdn.jsdelivr.net/npm/bootstrap@4.3.1/dist/js/bootstrap.min.js"
integrity="sha384-
JjSmVgyd0p3pXB1rRibZUAYoIIy6OrQ6VrjIEaFf/nJGzIxFDsf4x0xIM+B07jRM"
crossorigin="anonymous"></script>

</body>
</html>
```

index.html

```
<!DOCTYPE html>
<html lang="en">
<head>

  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Your Single Page Website</title>

  <!-- Bootstrap CSS -->
  <link                                                          rel="stylesheet"
href="https://cdnjs.cloudflare.com/ajax/libs/bootstrap/5.3.0/css/bootstrap.min.css">
  <!-- Custom CSS -->
  <style>



    body {
      background-image: url('https://wallpaper.dog/large/10719999.jpg'); /* Change to your
local background image path */
      background-size: cover;
      background-repeat: no-repeat;
      background-position: center;
      background-color: rgba(0, 0, 0, 0.5); /* Change the alpha value to adjust darkness (0.5
= 50% transparency) */
      color: rgb(216, 239, 17); /* Text color set to white */    }



    .project-name {
      font-size: 48px;
      text-align: center;
      padding-top: 90px; /* Decreased the upper padding by 10 pixels */
      margin-bottom: 30px;
    }
```

```
    .project-description {
      font-size: 18px;
      text-align: center;
      margin-bottom: 30px;
      color: rgb(216, 239, 17)
    }



    .card {
      transition: transform 0.3s;
      cursor: pointer; /* Set the cursor to pointer when hovering over the card */
      background-color: rgba(255, 255, 255, 0.454); /* Set card background color to white
*/
      color: rgb(216, 239, 17)k; /* Set text color to black */
      border-radius: 10px; /* Rounded edges */
      box-shadow: 0 4px 8px rgba(0, 0, 0, 0.1); /* Shadow effect */
    }




    .card:hover {
      transform: scale(1.05);
    }




    .card-custom {
      text-decoration: none;
    }
  </style>
</head>
<body>
```

```
<!-- Main Content -->
<div class="container py-5">
  <!-- Project Name -->
  <div class="project-name">
    Diamonds For You
  </div>



  <!-- Project Description -->
  <div class="project-description">
    This is an open source project built on an everevolving data warehouse that hosts data of thounds of diamonds. We use this data to train out machine learning model that uses cutting edge technology to predict diamond prices for sellers and buyers.
  </div>




  <div class="row">
    <!-- Card 1 -->
    <div class="col-md-6">
      <a class="card-custom" href="/form/buyer">
        <div class="card mb-3">
          <div class="card-body">
            <h5 class="card-title">Are you a Buyer?</h5> <!-- Added Card 1 Title -->
            <p class="card-text">
              Buying a diamond for someone who is not from the diamond industry can be a very difficult task. Let us make it easy for you. If you exactly know what you are looking for or if you dont even have a clue, we can help. We work day and night to bridge the gaps and give are users information about diamond prices. Just give use the diamond details and our cutting edge software will predict the best price at which you can buy this diamond. Click this card to begin!
            </p>
          </div>
        </div>
      </a>
    </div>
```

```
    <!-- Card 2 -->
    <div class="col-md-6">
      <a class="card-custom" href="/form/seller">
        <div class="card mb-3">
          <div class="card-body">
            <h5 class="card-title">Are you a Seller?</h5> <!-- Added Card 2 Title -->
            <p class="card-text">
              Running a small diamond business comes with its own special problems. These problems are very different from a genral business point of view. Let us make this easy for you. Our data warehouse of thounsands of diamonds help us ascertain the price of a diamond given its features. Our website shall help you predict the price range of the diamond so you can make an informed decision when you are dealing with the everchanging diamond market. Click this card to use are services for free right now!
            </p>
          </div>
        </div>
      </a>
    </div>
  </div>
</div>

<!-- Bootstrap JS and Popper.js (optional) -->
<script src="https://cdnjs.cloudflare.com/ajax/libs/bootstrap/5.3.0/js/bootstrap.bundle.min.js"></script>
</body>
</html>
```

test.html

```html
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Document</title>
</head>
<body>
  <h1>404 not found</h1>
</body>
</html>
```

Head of raw data

```
diamonds=pd.read_csv('diamonds.csv')
diamonds.head()
```

	Unnamed: 0	carat	cut	color	clarity	depth	table	price	x	y	z
0	1	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
1	2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
2	3	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
3	4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
4	5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75

Characteristics of the data

```
diamonds.describe()
```

	Unnamed: 0	carat	depth	table	price	x	y	z
count	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000
mean	26970.500000	0.797940	61.749405	57.457184	3932.799722	5.731157	5.734526	3.538734
std	15571.281097	0.474011	1.432621	2.234491	3989.439738	1.121761	1.142135	0.705699
min	1.000000	0.200000	43.000000	43.000000	326.000000	0.000000	0.000000	0.000000
25%	13485.750000	0.400000	61.000000	56.000000	950.000000	4.710000	4.720000	2.910000
50%	26970.500000	0.700000	61.800000	57.000000	2401.000000	5.700000	5.710000	3.530000
75%	40455.250000	1.040000	62.500000	59.000000	5324.250000	6.540000	6.540000	4.040000
max	53940.000000	5.010000	79.000000	95.000000	18823.000000	10.740000	58.900000	31.800000

Data after OneHotEncoding the categorical attributes

	carat	depth	table	x	y	z	price	cut_Fair	cut_Good	cut_Ideal	...	color_I	color_J	clarity_I1	clarity_IF
0	0.23	61.5	55.0	3.95	3.98	2.43	326	0	0	1	...	0	0	0	0
1	0.21	59.8	61.0	3.89	3.84	2.31	326	0	0	0	...	0	0	0	0
2	0.23	56.9	65.0	4.05	4.07	2.31	327	0	1	0	...	0	0	0	0
3	0.29	62.4	58.0	4.20	4.23	2.63	334	0	0	0	...	1	0	0	0
4	0.31	63.3	58.0	4.34	4.35	2.75	335	0	1	0	...	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
53935	0.72	60.8	57.0	5.75	5.76	3.50	2757	0	0	1	...	0	0	0	0
53936	0.72	63.1	55.0	5.69	5.75	3.61	2757	0	1	0	...	0	0	0	0
53937	0.70	62.8	60.0	5.66	5.68	3.56	2757	0	0	0	...	0	0	0	0
53938	0.86	61.0	58.0	6.15	6.12	3.74	2757	0	0	0	...	0	0	0	0
53939	0.75	62.2	55.0	5.83	5.87	3.64	2757	0	0	1	...	0	0	0	0

clarity_SI1	clarity_SI2	clarity_VS1	clarity_VS2	clarity_VVS1	clarity_VVS2
0	1	0	0	0	0
1	0	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	1	0	0	0	0
...	...	...	...	...	...
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0

Data after binning by price

	carat	depth	table	x	y	z	cut_Fair	cut_Good	cut_Ideal	cut_Premium	...	color_J	clarity_I1	clarity_IF	clarity_SI1	clarity_SI2	clarity_VS1	clarity_VS2	clarity_VVS1	clarity_VVS2	price
0	0.23	61.5	55.0	3.95	3.98	2.43	0	0	1	0	...	0	0	0	0	1	0	0	0	0	0
19352	0.30	62.2	57.0	4.26	4.32	2.67	0	0	1	0	...	0	0	0	0	0	0	1	0	0	0
19351	0.31	61.8	56.0	4.35	4.37	2.69	0	0	1	0	...	0	0	0	0	0	0	1	0	0	0
19350	0.30	60.7	57.0	4.36	4.41	2.66	0	0	1	0	...	0	0	0	0	0	0	0	0	1	0
36544	0.40	62.6	59.0	4.74	4.68	2.95	0	0	0	1	...	0	0	0	1	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
27208	2.02	57.9	63.0	8.13	8.21	4.73	0	1	0	0	...	0	0	0	0	0	1	0	0	0	9
27209	2.36	60.1	59.0	8.64	8.69	5.21	0	0	1	0	...	1	0	0	0	0	0	1	0	0	9
27210	2.00	62.9	56.0	7.94	8.01	5.02	0	0	1	0	...	0	0	0	0	1	0	0	0	0	9
27212	2.29	61.6	60.0	8.45	8.43	5.20	0	0	0	1	...	0	0	0	0	0	0	1	0	0	9
26972	2.09	61.8	57.0	8.24	8.17	5.07	0	0	1	0	...	0	0	0	0	0	0	1	0	0	9

The model

```
randomforest=RandomForestClassifier()
randomforest.fit(train_X,train_Y)
predi=randomforest.predict(test_X)
# joblib.dump(randomforest,'RFor_NoHy_OneHot_DataPts.joblib')
joblib.dump(randomforest,'RFor_NoHy_OneHot_Price.joblib')
accuracy_score(predi,test_Y)
```

```
0.8760200296735905
```

Confusion matrix of best performing mode