|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title** | Diamond Price Prediction | | |
| **Project Code** | **AIS301** | **Course Name** | **Machine Learning** |
| **Professor** | **Dr. Ghada Khoriba** | | |
| **TA** | **AlAmir Hassan** | **Mentor Name** | Text. |
| **Team Name** | **MNAY** | | |
| **Team Members** | **Nour Mohamed** | **Youssef Abbas** | **Maryam Walid** |
| **Abdelrahman Medhat** | Text. | Text. |
| **Problem Summary** | Scarcity plays a huge role in any commodity price, and when you think of scarcity, the first thing that comes to your mind is diamonds, as they are one of the rarest and most valuable substances that are produced naturally on this planet. However, unlike gold and silver, determining the price of a diamond could be very complex. The actual value of a diamond is determined by a gemologist (a person that specializes in identifying, grading, and appraising gemstones) after inspecting its various features. In his project, we aim to come up with the most efficient model for the price prediction of diamonds. This model should work based on previously trained and analyzed data about diamond type, carat, cut, color, clarity, depth, prices, etc. as each of these feature’s effects the diamond price differently depending on its weight. This model will be trained and updated until it reaches the highest accuracy possible compared to human error | | |
| **Methodology** | In our project, we are going to use some machine learning algorithms and models to help in predicting diamond prices with high accuracy. Our dataset contains prices and attributes for approximately 54,000 round-cut diamonds. it will consist of 10 features (carat, cut, color, clarity, depth, table, price, x, y, and z). Most variables are numeric in nature, but the variables cut, color, and clarity are ordered factor variables. for the price currency is $. And About the columns x,y, and z they are diamond measurements as (( x: length in mm, y: width in mm, z: depth in mm )). After understanding the domain of the project, we imported Diamond dataset using Pandas library, checked for NULL values and cleared redundant data, decoded numerical columns, virtualized data to find correlated attribute with price. To minimize outliers, we scaled the data. We used three different models, linear regression, decision tree regressor, random forest regressor. | | |
| **Achievements and Skills Gained** | 1. **Data Analysis** 2. **Visualization** 3. **Machine Learning Models** 4. **Analyzing dataset** 5. **Cleaning dataset** 6. **Exploring more libraries in Pandas** 7. **Compare between better models according to accuracy** | | |

|  |  |
| --- | --- |
| **Project Title** | Diamond Price Prediction |
| **Main Results** | A picture containing text  Description automatically generatedGraphical user interface, text, application  Description automatically generated A picture containing table  Description automatically generated |
| **Discussion and Conclusion** | We tried many models to achieve the best accurate model |
| **References** | https://www.kaggle.com/datasets/nancyalaswad90/diamonds-prices |
| **Future Work and Suggestions** | We are aiming to create a model to predict expensive gemstones type from images using CNN |
| **Group Photo** |  |