“Project Summary: Property Price Prediction Capstone Project”

The Property Price Prediction Capstone Project aimed to develop a robust machine learning model for predicting property prices using various features from a dataset. The primary objective was to leverage data science techniques to forecast property values accurately, assisting stakeholders like real estate agents, buyers, and sellers in making informed decisions. The project followed a structured approach involving data loading, cleaning, exploratory data analysis (EDA), feature engineering, model training, evaluation, and results reporting.

1.Data Loading and Initial Exploration: The project began with the import of essential Python libraries such as pandas, numpy, seaborn, matplotlib, and sklearn. The dataset, Property\_data (1).csv, containing columns such as PropertyID, PropertyClass, PropertyZone, PropertyFrontage, PropertySize, Street, Alley, PropertyShape, Elevation, Amenities, and the target variable PropPrice, was loaded using pandas. Initial inspection of the dataset was carried out using .head() and .info() methods, providing an overview of the first few rows and data types.

2.Data Cleaning: The data cleaning process was crucial to ensure the accuracy and quality of the model. Numerical columns with missing values were filled using the median, while categorical columns were imputed with the most frequent values. This approach addressed missing data and reduced bias in the dataset. The dataset was checked for any remaining missing values to confirm the completeness of the data after imputation.

3. Exploratory Data Analysis (EDA): EDA was conducted to understand the underlying patterns and relationships within the data. Summary statistics were generated using data.describe() to provide insights into the distribution and central tendencies of the numerical features. A correlation matrix was plotted to visualize the relationships between different variables and identify strong correlations. Additionally, the distribution of the target variable, PropPrice, was analyzed using a histogram to understand its skewness and distribution characteristics.

4. Feature Engineering: To enhance the predictive power of the model, categorical features were encoded using LabelEncoder, transforming text labels into numerical values suitable for machine learning algorithms. A new feature, PropertyClass, was created as a proxy for property age by subtracting the sale year from the property price. This feature engineering aimed to introduce additional relevant information to the model.

5. Feature Selection and Scaling: Relevant features were selected for model training, including PropertyClass, PropertySize, PropertyFrontage, and PropPrice. Feature scaling was applied using StandardScaler to normalize the feature values and ensure that they contributed equally to the model training process.

6. Model Training and Evaluation: The dataset was split into training and testing sets using train\_test\_split. A Random Forest Regressor was chosen for its robustness and ability to handle complex relationships in the data. The model was trained on the training set and evaluated on the test set. Performance metrics such as Mean Squared Error (MSE) and R² Score were calculated to assess the model's accuracy and predictive capability. The model achieved a satisfactory R² score, indicating a good fit between predicted and actual property prices.

7. Reporting and Saving Results: The trained model was saved using joblib for future use and deployment. Predictions were made on the test set, and the results, including actual and predicted values, were saved to a CSV file for further analysis. The model’s performance metrics, along with other insights, were documented to ensure comprehensive reporting of the project's outcomes.

Overall, the project demonstrated the application of machine learning techniques in predicting property prices, providing valuable insights into the factors influencing property values.

This summary captures the project's scope, methodologies, and outcomes in detail, providing a clear overview of the work accomplished and its implications…