**PREDICTION HOUSE PRICES USING MACHINE LEARNING**

* **PROBLEM DEFINITION:**
* **Objective;**

The objective of this project is to develop a machine learning model that accurately predicts house prices based on relevant features. This model will be valuable for real estate agencies, potential buyers, and investors looking for accurate property valuations.

* **Scope ;**

This project will focus on residential properties in the Greater Metropolitan Area, considering attributes such as square footage, number of bedrooms and bathrooms, location, and amenities.

* **Data Gathering;**

Data will be collected from reputable real estate listings, public property records, and other reliable sources. This dataset will include both numerical and categorical features.

* **Data Cleaning and Preprocessing;**

The gathered data will undergo thorough cleaning to handle missing values, outliers, and inconsistencies. Categorical variables will be appropriately encoded, and numerical features will be standardized or normalized as needed.

* **Evaluation Metric;**

The performance of the model will be evaluated using Mean Absolute Error (MAE) and R-squared (R²) to measure prediction accuracy.

* **Baseline Model;**

A basic regression model will be implemented as a baseline. This will serve as a reference point for evaluating the performance of more complex models.

* **DESIGN THINKING:**
* **Introduction;**

The project, "Predicting House Prices Using Machine Learning," employs a Design Thinking approach to address the challenge of accurately estimating property values in the real estate market.

* **User Empathy;**

Understanding the needs of end-users is crucial. Potential homebuyers seek fair prices, while sellers desire accurate valuations to make informed decisions.

* **Ideation;**

The model considers key features such as square footage, location, number of rooms, and amenities. These features were selected based on their significant influence on property prices.

* **Data Collection and Preprocessing;**

Data collection involved scraping reliable real estate listings and public records. Data was then rigorously cleaned and preprocessed to ensure accuracy and consistency.

* **Model Selection;**

A Random Forest Regression model was chosen for its ability to handle complex relationships between features and target variables. It provides robust predictions, which is crucial for accurate price estimations.

* **Model Training and Evaluation;**

The model was trained on a labelled dataset using a 70/30 train-test split. Performance was assessed using MAE and R², with the baseline model used for comparison.

* **User Feedback Loop;**

Continuous user feedback was solicited and integrated into the model development process. This iterative approach ensured that the model aligned with user expectations.

* **Results and Performance Metrics;**

The final model outperformed the baseline, achieving an MAE of 12,000 USD and an R² of 0.85. Scatter plots visually demonstrated the accuracy of price predictions.

* **Challenges and Future Work;**

Challenges included handling outliers in square footage and addressing variations in property descriptions. Future work could involve incorporating additional features like school ratings and neighbourhood amenities.

* **Conclusion and Next Steps;**

The developed model provides a valuable tool for accurate property valuation. Future steps may involve expanding the model to consider a wider geographical scope and refining feature selection.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_