**House Price Prediction: Final Project Report**

**1. Introduction**

**1.1 Project Overview**

This project aims to develop a machine learning model to predict house prices based on various features such as area, number of bedrooms, location, etc. Accurate house price prediction is crucial for real estate professionals, buyers, and sellers to make informed decisions.

**1.2 Dataset Description**

The dataset contains house-related features including:

* **Area**: Total square footage of the house
* **Bedrooms**: Number of bedrooms
* **Bathrooms**: Number of bathrooms
* **Location**: Categorical variable indicating the area of the house
* **Price**: Target variable representing the price of the house

**2. Data Pre-processing**

**2.1 Handling Missing Values**

* Missing values in numerical columns were replaced with the median value.
* Categorical missing values were replaced using the mode.

**3. Exploratory Data Analysis (EDA)**

**3.1 Data Visualization**

* **Distribution Plots**: House price distribution was visualized using histograms.
* **Correlation Heatmap**: Showed the relationships between features.

**4. Feature Engineering**

**4.1 Feature Selection**

* Using **Random Forest feature importance**, the most impactful features were selected: Area, Number of Bedrooms, and Location.

**4.2 Data Transformation**

* House prices were **log-transformed** to normalize the distribution and improve model performance.

**5. Model Training and Evaluation**

**5.1 Model Selection**

* The following models were trained and compared:
  + **Linear Regression**
  + **Decision Tree Regressor**
  + **Random Forest Regressor**

**5.2 Model Evaluation**

* The models were evaluated using:
  + **R² Score**
  + **Mean Absolute Error (MAE)**
  + **Mean Squared Error (MSE)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **R² Score** | **MAE** | **MSE** |
| Random Forest | **0.61** | **1022506** | **1964193399645** |

**6. Model Deployment**

**6.1 Creating a Flask API**

* The trained model was **saved using joblib**.
* A Flask API was created to receive JSON requests and return house price predictions.

**6.2 API Testing**

* The API was tested using **Python requests**, successfully returning predicted prices based on input features.

**7. Insights & Conclusion**

**7.1 Key Findings**

* **Data Distribution:** The distribution of house prices shows a right-skewed pattern.
* **Feature Importance:** Some features have stronger correlations with the target variable.
* **Model Performance:** Random Forest performed well with MAE and MSE scores indicating reasonable accuracy.
* **Deployment:** The model is successfully deployed using Flask and can be accessed via API for predictions.

## Conclusion

The project successfully implemented a **House Price Prediction** model with optimized pre-processing, feature engineering, model training, and deployment. Future improvements could involve tuning additional hyperparameters, trying different regression models, or using more data for better generalization.

**8. References**

* Dataset Source: [Kaggle Housing Dataset]