REPORT TITLE
HOUSE PRICE PREDICTION USING ARTIFICIAL NEURAL NETWORKS

Introduction

In the realm of real estate, accurately predicting housing prices is a crucial task for various stakeholders including investors, buyers, real estate agents, and policymakers. The ability to forecast house values based on a multitude of features can significantly enhance decision-making processes, improve market analysis, and optimize investment strategies.

This project aims to develop a robust predictive model for housing prices using a comprehensive dataset encompassing various property characteristics and location-specific information. By leveraging machine learning techniques, particularly neural networks, we aim to understand the complex relationships between different features and housing prices, ultimately providing a reliable tool for price prediction. This model not only assists in price estimation but also offers valuable insights into the factors influencing housing market trends.

Project Overview

This project aims to develop a predictive system for housing prices using a machine learning model. By leveraging a comprehensive dataset that includes various features related to property characteristics and their locations, we aim to accurately predict the median house values. The predictive model is built using a neural network and includes preprocessing steps to ensure the data is appropriately scaled for optimal model performance.

Data Description

The dataset used in this project includes the following features:

- **Longitude and Latitude:** Geographic coordinates of the properties.
- **Housing Median Age:** The median age of houses in the area.
- **Total Rooms:** The total number of rooms in the houses.
- **Total Bedrooms:** The total number of bedrooms in the houses.
- **Population:** The population of the area.
- **Households:** The number of households in the area.
- **Median Income:** The median income of the households in the area.
- **Ocean Proximity:** The proximity of the property to the ocean, categorized as follows:
    - `<1H OCEAN` (within one hour of an ocean)
    - `INLAND` (non-coastal regions)
    - `NEAR OCEAN` (close to the ocean but not directly on the coast)
    - `NEAR BAY` (near a bay or inlet)
    - `ISLAND` (islands)

Data Preprocessing

The data preprocessing steps include:

- **Scaling:** Features are scaled using a `MinMaxScaler` to ensure they are within a similar range, which is crucial for the performance of neural networks.

- **Encoding:** Categorical features such as `ocean_proximity` are encoded numerically to be used as inputs for the model.

Model Architecture

Our neural network model is designed with the following architecture:

- **Input Layer:** Accepts the feature vector of shape (9,).
- **Hidden Layers:**
    - Dense layer with 1000 neurons and ReLU activation function.
    - Dropout layer with a rate of 0.2 to prevent overfitting.
    - Dense layer with 500 neurons and ReLU activation function.
    - Dropout layer with a rate of 0.2 to prevent overfitting.
    - Dense layer with 250 neurons and ReLU activation function.
    - Dense layer with 100 neurons and ReLU activation function.
- **Output Layer:** A single neuron with a linear activation function to predict the median house value.

Model Training

The model is compiled using the `RMSprop` optimizer and the `mean_squared_error` loss function, with `mean_absolute_error` as a metric. An early stopping mechanism is employed to monitor the validation loss and stop training when the model stops improving, restoring the best weights.

- **R^2 Score:** 0.7036
- **Mean Squared Error (MSE):** 4053679193.21
- **Mean Absolute Error (MAE):** 43947.08

Observations

- **Effect of Ocean Proximity:** Houses closer to the ocean tend to have higher median values. The largest proportion of the dataset falls under `<1H OCEAN`, indicating a significant influence of ocean proximity on house prices.
- **Bedrooms and House Value:** Properties with fewer bedrooms generally have lower median values, but there is considerable variability, with some properties having high values despite fewer bedrooms.
- **Population and House Value:** A slight trend shows that higher populations may correlate with higher house values, but this trend is not consistent, and there are outliers.

- **Households and House Value:** Similar to the population trend, areas with more households show a slight increase in house values, with significant variability.
- **Rooms and House Value:** More rooms generally correlate with higher house values, but the relationship is not strictly linear, with considerable spread in house values.
- **Income and House Value:** There is a positive correlation between median income and house values, with more variability in house prices in higher-income areas.
- **Age and House Value:** No clear trend is observed between housing age and value, indicating other factors may play a more significant role.

Predictive System

The predictive system takes the following input features to predict the median house value:

- Longitude
- Latitude
- Housing Median Age
- Total Rooms
- Total Bedrooms
- Population
- Households
- Median Income
- Ocean Proximity

EXAMPLE PREDICTION

```
longitude = -122.2300
latitude = 37.8800
housing_median_age = 41.0000
total_rooms = 880.0000
total_bedrooms = 129.0000
population = 322.0000
households = 126.0000
median_income = 8.3252
ocean_proximity = 3.0000

price = pred(longitude, latitude, housing_median_age, total_rooms, total_bedrooms, populat
print(price)  # Output: 400603.66
```

CONCLUSION

This project successfully demonstrates the application of neural networks in predicting housing prices. The model's performance, reflected by its R^2 score, MSE, and MAE, indicates a good fit for the data, providing valuable insights into the factors influencing house prices. Further improvements can be made by incorporating additional features, tuning the model, and using more advanced machine learning techniques.

THANK YOU