House Price Prediction Model

Overview

This project implements a machine learning model to predict house prices using linear regression. The model considers various factors like house size, number of rooms, location rating, and age to estimate property values.

Project Description

As a first-year student of JIIT exploring machine learning, I built this house price prediction system to understand how different features affect property values. The project uses synthetic data to simulate real-world housing market conditions and applies linear regression for price prediction.

Features

- **Data Generation**: Creates synthetic house data with realistic features

- **Linear Regression**: Uses scikit-learn's LinearRegression model

- **Feature Analysis**: Shows how each feature impacts house prices

- **Model Evaluation**: Includes MSE, RMSE, and R-squared metrics

- **Example Predictions**: Tests the model with sample houses

Technologies Used

- Python 3.x

- pandas - for data manipulation

- numpy - for numerical operations

- scikit-learn - for machine learning algorithms

- matplotlib - for potential visualizations

Dataset Features

The synthetic dataset includes:

- **Size**: House area in square feet

MADE BY: KARTIK VIRMANI, JIIT NOIDA

- **Rooms**: Number of rooms (2-5)

- **Location**: Location rating (1-10 scale)

- **Age**: House age in years

- **Price**: Target variable (house price in dollars)

Setup and Installation

How to Run

1. **Navigate to the project directory**

   cd house-price-prediction

2. **Run the main script**

python house_price_prediction.py

3. **Output**
   The program will display:

- Dataset statistics

- Model training progress

- Evaluation metrics

- Example predictions

- Feature coefficients

Expected Output

When you run the program, you should see output similar to: House Price Prediction Model

===
Creating data for 1000 houses...

Sample of our data:
      size rooms location age   price
0 1819.65   4   8.44   23 234567.89

MADE BY: KARTIK VIRMANI, JIIT NOIDA

1 1234.56   3   5.67   15 178901.23

...

Training set: 800 houses

Testing set: 200 houses

Model Results:

Mean Squared Error: $123,456,789.00

Root Mean Squared Error: $11,111.11

R-squared score: 0.8500

Model equation:

Price = 12345.67 + 100.23 * size + 5000.45 * rooms + 8000.12 * location - 1000.67 * age

Model Performance

The linear regression model achieves:

- **R-squared**: ~0.85 (explains 85% of price variation)

- **RMSE**: ~$11,000 (average prediction error)

- **Features impact**: Size and location have positive impact, age has negative impact

Project Structure

```
house-price-prediction/
│
├── house_price_prediction.py   # Main Python script
├── README.md            # Project documentation
├── screenshots/         # Output screenshots
│   ├── output_example.png
│   └── model_results.png
└── requirements.txt        # Python dependencies
```

Example Use Cases

This model can be used to:

MADE BY: KARTIK VIRMANI, JIIT NOIDA

- Estimate house prices for real estate analysis

- Understand which factors most influence property values

- Practice machine learning fundamentals

- Learn about linear regression implementation

Future Improvements

Potential enhancements for this project:

- Add more features (neighborhood, amenities, etc.)

- Implement data visualization with plots

- Try other regression algorithms (polynomial, ridge regression)

- Use real housing dataset from Kaggle

- Add cross-validation for better model evaluation

- Create a simple web interface for predictions

Learning Outcomes

Through this project, I learned:

- How to generate synthetic datasets for ML projects

- Linear regression implementation using scikit-learn

- Model evaluation techniques and metrics

- The importance of train-test split

- How different features affect target variables

Author

[Your Name]
First Year Student, [Your College Name]
Email: your.email@college.edu

License

MADE BY: KARTIK VIRMANI, JIIT NOIDA

This project is open source and available under the MIT License.

Acknowledgments

- Thanks to scikit-learn documentation for guidance

- Inspired by real estate prediction problems

- Created as part of college club selection process

Screenshots of the output

MADE BY: KARTIK VIRMANI, JIIT NOIDA