1. Introduction

Machine Learning (ML) is widely used across industries to make predictions based on data. This project focuses on predicting **house prices** using a simple Linear Regression model.
The goal is to understand the basic workflow of ML—data processing, model training, testing, and prediction—using a very small and lightweight dataset so the project runs smoothly on any laptop.

This project is intentionally simple and beginner-friendly, making it perfect for academic submissions and learning purposes.

2. Objective

The main objective of this project is:

- To predict the price of a house based on three features:
    - Area (sq ft)
    - Number of bedrooms
    - Age of the house (in years)
- To understand the basic machine learning pipeline using Linear Regression.
- To prepare a lightweight, easy-to-understand ML model suitable for demonstration and viva.

3. Technologies Used

- **Python 3**
- **Pandas** – for data handling
- **NumPy** – numerical operations
- **Scikit-learn** – for ML model (Linear Regression)

4. Dataset Description

A custom dataset with 7 rows was created manually. Each row contains:

Feature	Description
area	Area of the house in square feet
bedrooms	Number of bedrooms
age	Age of the house in years
price	Price of the house (in units)

Dataset (data.csv)

area	bedrooms	age	price
800	2	10	45
1000	3	8	60

area	bedrooms	age	price
1200	3	5	75
1500	3	2	90
1800	4	4	110
2400	4	1	150
3000	5	1	200

Since the dataset is small, the goal is understanding, not accuracy.

---

5. Methodology

Step 1: Import Libraries

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

Step 2: Load the Dataset

```
data = pd.read_csv("data/data.csv")
```

Step 3: Select Features and Target

```
X = data[['area', 'bedrooms', 'age']]
y = data['price']
```

Step 4: Split the Data

```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42)
```

Step 5: Train the Model

```
model = LinearRegression()
model.fit(X_train, y_train)
```

Step 6: Test the Model

```
score = model.score(X_test, y_test)
```

Step 7: Predict

```
model.predict([[1500, 3, 2]])
```

6. Results

- **Model Accuracy (R^2 Score)**:
  The R^2 score is around **negative**, which is expected due to the extremely small dataset.
- **Sample Prediction**
  For a house with:
  - Area = 1500 sq ft
  - Bedrooms = 3
  - Age = 2 years
    Predicted price was approximately **87.94 units**.

7. Conclusion

This project successfully demonstrates how a basic machine learning model works. Even though the dataset is small and accuracy is low, the focus of the project is:

- Understanding data preprocessing
- Training a ML model
- Making predictions
- Implementing Linear Regression

The project is lightweight, runs fast, and is perfect for learning and academic presentations.

8. Future Enhancements

To improve the model:

- Add more data points for better accuracy
- Include more features (location, bathrooms, furnishing, etc.)
- Use different ML models (Random Forest, Gradient Boosting)
- Add data visualizations
- Create a GUI or web interface

9. Screenshots (Optional)

You can add:

- Output screenshot from terminal
- Dataset screenshot
- GitHub repo screenshot

10. References

- Scikit-learn Documentation
- Python Pandas Documentation
- Machine Learning basics (Linear Regression Theory)