PHASE -2

Predicting House Prices using Machine Learning

NAME: Abhi.S

nmID : au952821106301

INTRODUCTION

Predicting house prices using machine learning is a valuable application of artificial intelligence in the real estate industry. By leveraging various data sources and algorithms, machine learning models can estimate the prices of residential properties accurately. This process involves collecting features such as square footage, location, number of bedrooms and bathrooms, and historical sales data. These features are then used to train a predictive model, which can be based on regression techniques like linear regression or more advanced models like gradient boosting or neural networks. The trained model can provide insights into property valuations, assist homebuyers in making informed decisions, and help real estate professionals in pricing listings effectively. This approach can significantly impact the housing market and facilitate more data-driven transactions.

The innovation in predicting house prices using machine learning lies in the utilization of advanced techniques and data sources to enhance accuracy, efficiency, and the overall user experience. Some key innovations include:

1.Advanced Algorithms:

Utilizing state-of-the-art machine learning algorithms such as XGBoost, Random Forest, or neural networks to capture complex patterns in housing data.

2.Feature Engineering:

Developing novel features or using alternative data sources like geospatial data, crime rates, school quality, and neighborhood sentiment analysis to improve model accuracy.

3.Real-time Predictions:

Providing real-time or near-real-time price predictions based on the most recent market data and updates, allowing users to make timely decisions.

3.Interpretability:

Developing models that are not just accurate but also interpretable, so users can understand why a prediction was made, which can be crucial in the real estate decision-making process.

4.Customization:

Allowing users to personalize their models by specifying their preferences and priorities, such as prioritizing proximity to specific amenities or schools.

5.Mobile Apps and Augmented Reality:

Developing mobile applications and augmented reality tools that enable users to point their smartphones at a property and instantly receive its estimated price and related information.

6.Blockchain Integration:

Using blockchain technology to provide transparent and tamper-proof records of property data and transactions, enhancing trust in the prediction process.

7.Sustainability Metrics:

Including environmental and energy efficiency data in predictions to cater to the growing interest in sustainable and energy-efficient homes.

Fun with Real Estate Data

-Use Rmarkdown to learn advanced regression techniques like random forests and XGBoost.

XGBoost with Parameter Tuning

-Implement LASSO regression to avoid multicollinearity.

-Includes linear regression, random forest, and XGBoost models as well.

Ensemble Modeling: Stack Model Example

-Use "ensembling" to combine the predictions of several models.

-Includes GBM (gradient boosting machine), XGBoost, ranger, and neural net using the caret package.

A Clear Example of Overfitting

-Learn about the dreaded consequences of overfitting data.

-Comprehensive Data Exploration with Python.

-Understand how variables are distributed and how they interact.

-Apply different transformations before training machine learning models.

House Prices EDA

-Learn to use visualization techniques to study missing data and distributions.

-Includes correlation heatmaps, pairplots, and t-SNE to help inform appropriate inputs to a linear model.

A Study on Regression Applied to the Ames Dataset

-Demonstrate effective tactics for feature engineering.

-Explore linear regression with different regularization methods including ridge, LASSO, and ElasticNet using scikit-learn.

Regularized Linear Models

-Build a basic linear model.

-Try more advanced algorithms including XGBoost and neural nets using Keras.

Data source:

Dataset Link: https://www.kaggle.com/datasets/vedavyasv/usa-housing

Program:

<u>Input</u>

House Price Prediction

Importing Dependencies

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_absolute_error,mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
import xgboost as xg
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
```

/opt/conda/lib/python3.10/site-packages/scipy/init.py:146: User Warning: A NumPy version>=1.16.5 and <1.23.0 is required for this version of SciPy (detected version 1.23.5

warnings.warn(f"A NumPy version>={np_minversion} and <{np_maxversion}"

Loading Dataset

dataset = pd.read_csv("E:/USA_Housing.csv')

Predicting Prices

In:

Prediction model lr.predict(X_test_scaly

Evaluation of Predicted Data

In:

pit plot(np.arange(len(Y_test)), Ytest, label Actual Trend)

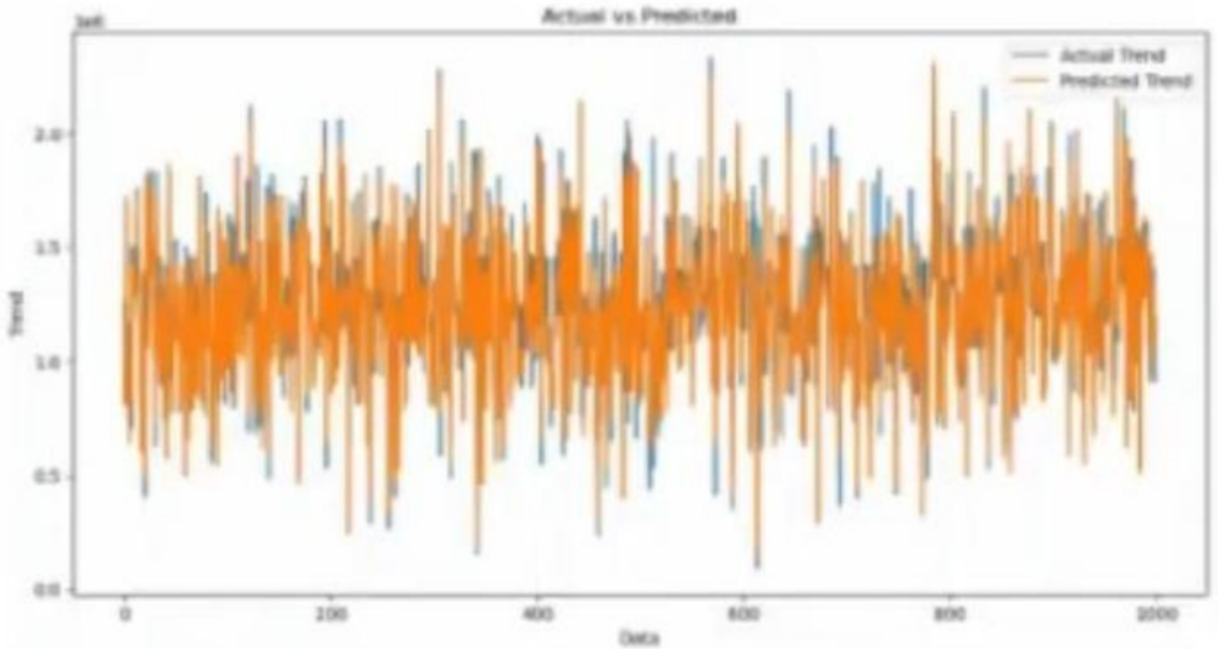
pit plotinp.arange(len(Ytest)), Prediction, label Predicted Trend)

pitylabel("Trand)

plt legend()

plttile Actual va Predicted)

output:



Support Vector Regressor

Input:

model_wr=SVR()

Out[S]:

Predicting Prices

Input:

Prediction2 = model_w.predict(X_test_scal)

Evaluation of Predicted Data

Input:

pit plot np.arange(len(Y_test)), Ytest, label Actual Trend)

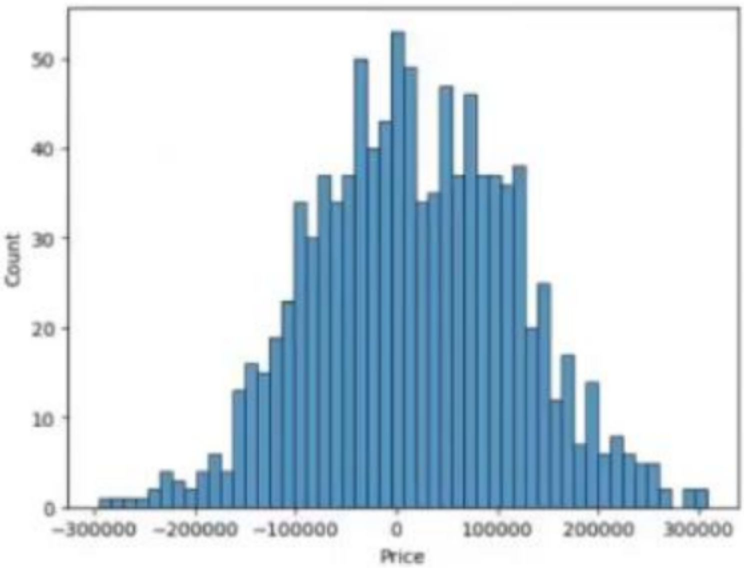
plt.plot(np.arange(len(Y_test)), Prediction2, label/Predicted Trend)

pitylabel(rond)

plt legend()

pilt title('Actual vs Predicted")

output:



In:

print2_score(Y_test, Prediction2))

printimean_absolute_error(Y_test, Prediction2))

print(mean squared_emory_test, Prediction2))

out:

-0.0006222175925689744

286137.81086908665

128209033251.4034

In:

print2_scoreY_test, Pradiction2))

print(mean absolute_error(Y_test, Prodiction2)) printman_squared_emor(Y_test, Prediction2))

Out:

-0.0006222175925689744

286137.81086908665

128209033251.4034

In:

print(r2_score(Y_test, Prediction2))

print(mean_absolute_error(Y_test, Prediction2)) print(mean_squared_error(Y_test, Prediction2))

Out:

-0.0006222175925689744

286137.81086908665

128209033251.4034

Conclusion and Future Work (Phase 2):

Project Conclusion:

In the Phase 2 conclusion, we will summarize the key findings and insights from the advanced regression techniques. We will reiterate the impact of these techniques on improving the accuracy and robustness of house price predictions.

Future Work:

We will discuss potential avenues for future work, such as incorporating additional data sources (e.g., real-time economic indicators), exploring deep learning models for prediction, or expanding the project into a web application with more features and interactivity.