Electricity Prices Prediction

TEAM MEMBERS

- 210821205007: Akash M
- 210821205043: Keerthivasan V
- 210821205048: Kumar P
- 210821205049: Lingeshwaran R

Phase 2 Submission Document

Project: Electric Price Prediction

Introduction:

Electric price prediction plays a vital role in our increasingly electrified world. It involves using data and models to anticipate fluctuations in electricity costs, enabling individuals, businesses, and policymakers to make informed decisions and adapt to changing market conditions. In this introduction, we'll delve into the methods, applications, and significance of electric price prediction.

Content for Project Phase 2 :

Consider exploring advanced machine learning techniques like random forest, gradient boosting and support vector machine .

Data Source:

A good data source for house price prediction using machine learning should be Accurate, Complete, Covering the geographic area of interest, Accessible.

Dataset Link: [**https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction**](https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	DateTime	Holiday	HolidayFla	DayOfWee	WeekOfYe	Day	Month	Year	PeriodOfD	ForecastW	SystemLoa	SMPEA	ORKTempe	ORKWinds	CO2Intens	ActualWin	SystemLoa	SMPEP2
2	########	None	0	1	44	1	11	2011	0	315.31	3388.77	49.26	6	9.3	600.71	356	3159.6	54.32
3	########	None	0	1	44	1	11	2011	1	321.8	3196.66	49.26	6	11.1	605.42	317	2973.01	54.23
4	########	None	0	1	44	1	11	2011	2	328.57	3060.71	49.1	5	11.1	589.97	311	2834	54.23
5	########	None	0	1	44	1	11	2011	3	335.6	2945.56	48.04	6	9.3	585.94	313	2725.99	53.47
6	########	None	0	1	44	1	11	2011	4	342.9	2849.34	33.75	6	11.1	571.52	346	2655.64	39.87
7	########	None	0	1	44	1	11	2011	5	342.97	2810.01	33.75	5	11.1	562.61	342	2585.99	39.87
8	########	None	0	1	44	1	11	2011	6	343.18	2780.52	33.75	5	7.4	545.81	336	2561.7	39.87
9	########	None	0	1	44	1	11	2011	7	343.46	2762.67	33.75	5	9.3	539.38	338	2544.33	39.87
10	########	None	0	1	44	1	11	2011	8	343.88	2766.63	33.75	4	11.1	538.7	347	2549.02	39.87
11	########	None	0	1	44	1	11	2011	9	344.39	2786.8	33.75	4	7.4	540.39	338	2547.15	39.87
12	########	None	0	1	44	1	11	2011	10	345.02	2817.59	33.75	4	7.4	532.3	372	2584.58	39.87
13	########	None	0	1	44	1	11	2011	11	342.23	2895.62	47.42	5	5.6	547.57	361	2641.37	39.87
14	########	None	0	1	44	1	11	2011	12	339.22	3039.67	44.31	5	3.7	556.14	383	2842.19	51.45
15	########	None	0	1	44	1	11	2011	13	335.39	3325.1	45.14	5	3.7	590.34	358	3082.97	51.45
16	########	None	0	1	44	1	11	2011	14	330.95	3661.02	46.25	4	9.3	596.22	402	3372.55	52.82
17	########	None	0	1	44	1	11	2011	15	325.93	4030	52.84	5	3.7	581.52	368	3572.64	53.65
18	########	None	0	1	44	1	11	2011	16	320.91	4306.54	59.44	5	5.6	577.27	361	3852.42	54.21
19	########	None	0	1	44	1	11	2011	17	365.15	4438.05	62.15	6	5.6	568.76	340	4116.03	58.33
20	########	None	0	1	44	1	11	2011	18	410.55	4585.84	61.81	8	7.4	560.79	358	4345.42	58.33
21	########	None	0	1	44	1	11	2011	19	458.56	4723.93	61.88	9	7.4	542.8	339	4427.29	58.33
22	########	None	0	1	44	1	11	2011	20	513.17	4793.6	61.46	?	?	535.37	324	4460.41	58.33
23	########	None	0	1	44	1	11	2011	21	573.36	4829.44	61.28	11	13	532.52	335	4493.22	58.27

Data Collection and Preprocessing:

 Importing the dataset: Obtain a comprehensive dataset containing relevant features such as square footage, number of bedrooms, location, amenities, etc.

 Data preprocessing: Clean the data by handling missing values, outliers, and categorical variables. Standardize or normalize numerical features.

Exploratory Data Analysis (EDA):

 Visualize and analyze the dataset to gain insights into the relationships between variables.

 Identify correlations and patterns that can inform feature selection and engineering.

 Present various data visualizations to gain insights into the dataset.

 Explore correlations between features and the target variable (electric prices).

 Discuss any significant findings from the EDA phase that inform feature selection.

Advanced machine learning model Techniques:

Machine learning models are widely used for electric price prediction due to their ability to handle complex, non-linear relationships and incorporate a variety of input features. Here are some common machine learning models for electric price prediction:

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It is effective in capturing non-linear relationships and handling a large number of features. Random Forest is often used for both short-term and long-term electric price forecasting.

Gradient Boosting: Gradient boosting algorithms like XG Boost and Light GBM are powerful for regression tasks. They build an ensemble of decision trees in a sequential manner, focusing on the errors of previous trees. These models are known for their high prediction accuracy.

Support Vector Machines (SVM): SVM is a versatile machine learning model that can be used for electric price prediction. SVM seeks to find a hyperplane that best separates the data points into different price categories.

Model Evaluation and Selection:

 Split the dataset into training and testing sets.

 Evaluate models using appropriate metrics (e.g., Mean Absolute Error, Mean Squared Error, R-squared) to assess their performance.

 Use cross-validation techniques to tune hyper parameters and ensure model stability.

 Compare the results with traditional linear regression models to highlight improvements.

 Select the best-performing model for further analysis.

Model Interpretability:

 Explain how to interpret feature importance from Gradient Boosting and XG Boost models.

 Discuss the insights gained from feature importance analysis and their relevance to electric price prediction.

 Interpret feature importance from ensemble models like Random Forest and Gradient Boosting to understand the factors influencing electric prices.

Deployment and Prediction:

 Deploy the chosen regression model to predict electric prices.

 Develop a user-friendly interface for users to input property features and receive price predictions.

Program:

Electric 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.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:
UserWarning: 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(': https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction')

Model 4 - Random Forest Regressor

In [1]: model_rf = RandomForestRegressor(n_estimators=50)

In [2]: model_rf.fit(X_train_scal, Y_train)

Out[2]:

```
RandomForestRegressor
RandomForestRegressor(n_estimators=50)
```

Predicting Prices

In [3]:Prediction4 = model_rf.predict(X_test_scal)

Evaluation of Predicted Data

In [4]:plt.figure(figsize=(12,6))

plt.plot(np.arange(len(Y_test)), Y_test, label='Actual Trend')

plt.plot(np.arange(len(Y_test)), Prediction4, label='Predicted Trend')

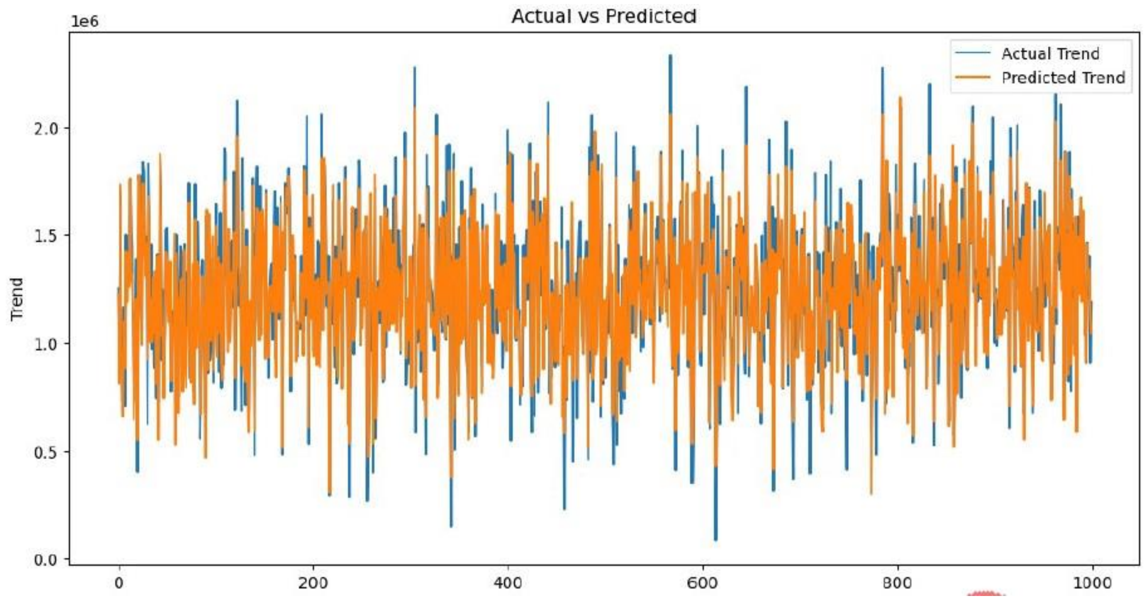
plt.xlabel('Data')

plt.ylabel('Trend')

plt.legend()

plt.title('Actual vs Predicted')

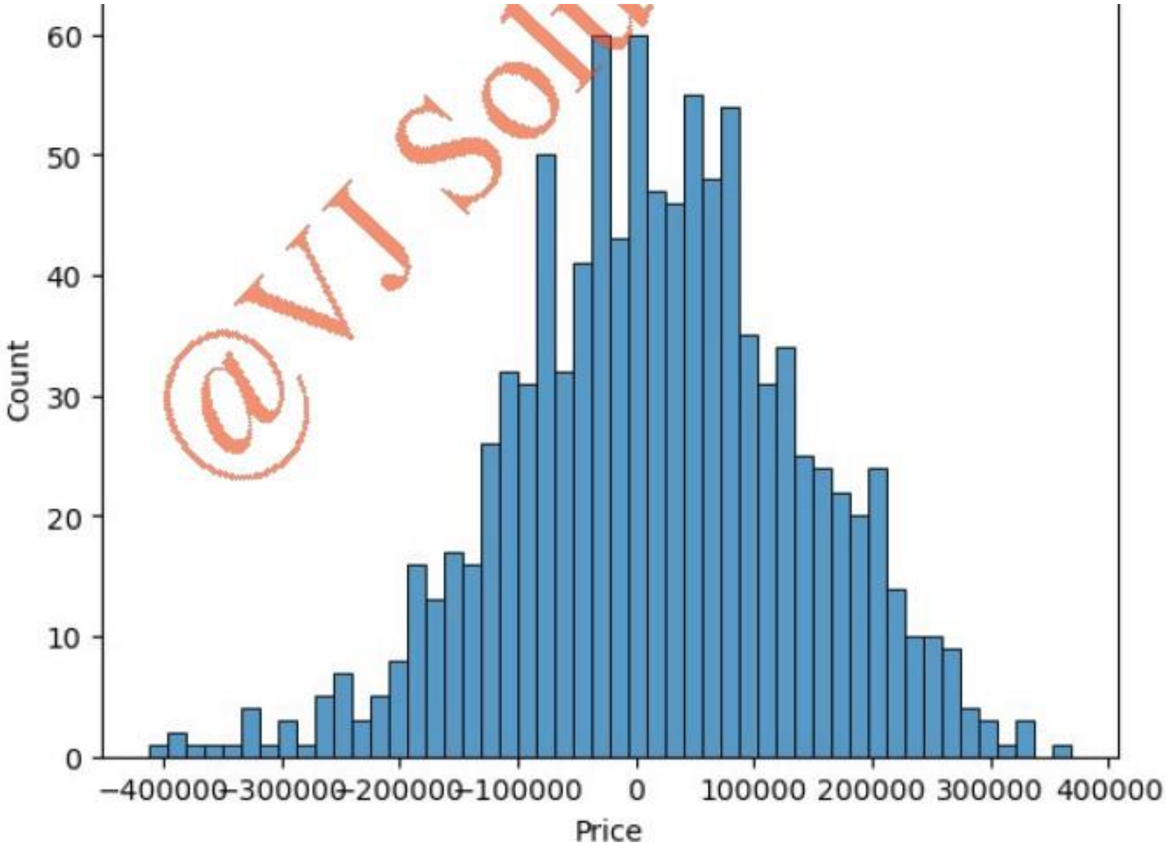
Out[4]:



Text(0.5, 1.0, 'Actual vs Predicted')

In [5]: sns.histplot((Y_test-Prediction4), bins=50)

Out[5]:



<Axes: xlabel='Price', ylabel='Count'>

In [6]: print(r2_score(Y_test, Prediction2))

print(mean_absolute_error(Y_test, Prediction2))

print(mean_squared_error(Y_test, Prediction2))

Out [6] :-0.0006222175925689744

286137.81086908665

128209033251.4034

Model 2 - Support Vector Regressor

In [7]: model_svr = SVR()

In [8]:model_svr.fit(X_train_scal, Y_train)

Out[8]:

Predicting Prices

In [9]:Prediction2 = model_svr.predict(X_test_scal)

Evaluation of Predicted Data

In [10]:plt.figure(figsize=(12,6))

plt.plot(np.arange(len(Y_test)), Y_test, label='Actual Trend')

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

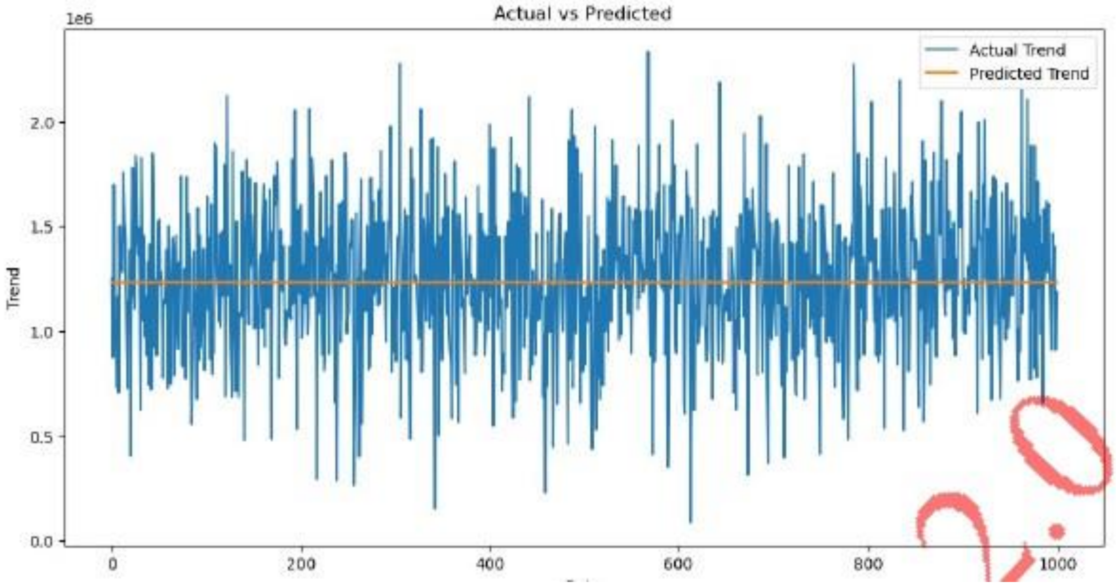
plt.xlabel('Data')

plt.ylabel('Trend')

plt.legend()

plt.title('Actual vs Predicted')

Out[10]:



Text(0.5, 1.0, 'Actual vs Predicted')

In [11]:sns.histplot((Y_test-Prediction2), bins=50)

<Axes: xlabel='Price', ylabel='Count'>

In [12]:print(r2_score(Y_test, Prediction2))

print(mean_absolute_error(Y_test, Prediction2))

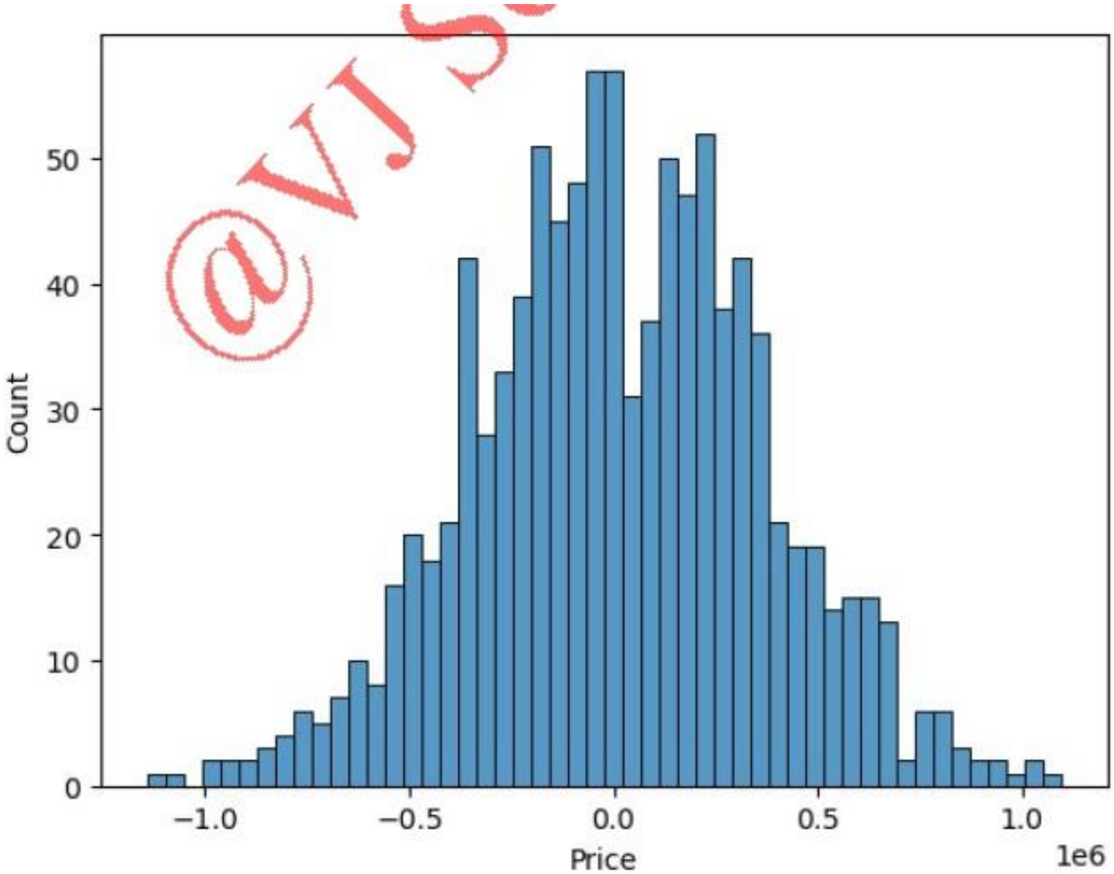
print(mean_squared_error(Y_test, Prediction2))

-0.0006222175925689744

286137.81086908665

128209033251.4034

Out[12]:



Conclusion:

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 Electric price predictions.