**ELECTRICITY PRICE PREDICTION**

PHASE – 3



Team members:

1.DIVYA.K(411621104008)

2.S.ANNIEMERCY(411621104003)

3.E.BARKAVI(411621104006)

4.S.JEEVARAJI(411621104017)

5.S.LEKHA SHREE(411621104028)

INTRODUCTION

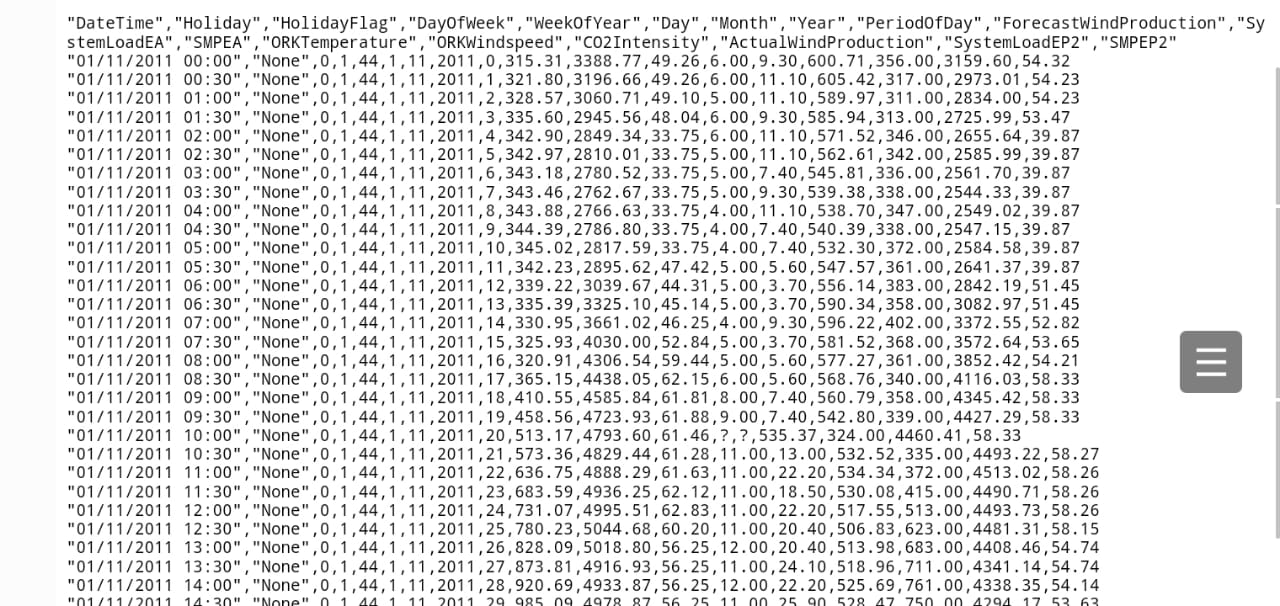
Electricity price prediction is the process of forecasting the future cost of electricity in a specific region or market. It is a crucial task for energy companies, consumers, and policymakers because it helps in making informed decisions related to energy consumption, production, and investment. Electricity price prediction plays a vital role in optimizing energy consumption, managing electricity generation cost of electricity in a specific region or market. It is a crucial task for energy companies, consumers, and policymakers because it helps in making informed decisions related to energy consumption, production, and investment.

DATASET

The dataset used for electricity price prediction typically consists of historical data related to electricity prices and various relevant factors that can influence those prices. Some of the key components of the dataset are Price data , Demand data, Supply data, weather data, Market data, Calendar data, Economic data.

The dataset we used here is:

<http://raw.githubusercontent.com/amankharwal/website-data/master/electricity.csv>



**1.Importing libraries and loading dataset:**

import pandas as p import numpy as np import seaborn as sns import matplotlib.pyplot as plt data = pd.read\_csv("http://raw.githubusercontent.com/amankharwal/website-data/master/electricity.csv") print(data.head())

Output:

DateTime Holiday HolidayFlag DayOfWeek WeekOfYear Day Month \

0 01/11/2011 00:00 NaN 0 1 44 1 11

1 01/11/2011 00:30 NaN 0 1 44 1 11

2 01/11/2011 01:00 NaN 0 1 44 1 11

3 01/11/2011 01:30 NaN 0 1 44 1 11

4 01/11/2011 02:00 NaN 0 1 44 1 11

Year PeriodOfDay ForecastWindProduction SystemLoadEA SMPEA \

0 2011 0 315.31 3388.77 49.26

1 2011 1 321.80 3196.66 49.26

2 2011 2 328.57 3060.71 49.10

3 2011 3 335.60 2945.56 48.04

4 2011 4 342.90 2849.34 33.75

ORKTemperature ORKWindspeed CO2Intensity ActualWindProduction SystemLoadEP2 \

0 6.00 9.30 600.71 356.00 3159.60

1 6.00 11.10 605.42 317.00 2973.01

2 5.00 11.10 589.97 311.00 2834.00

3 6.00 9.30 585.94 313.00 2725.99

4 6.00 11.10 571.52 346.00 2655.64

SMPEP2

0 54.32

1 54.23

2 54.23

3 53.47

4 39.87

**2. Exploratory Data Analysis(EDA):**

2.1 data.info()

data = data.dropna()

Output:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 38014 entries, 0 to 38013

Data columns (total 18 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 DateTime 38014 non-null object

1 Holiday 1536 non-null object

2 HolidayFlag 38014 non-null int64

3 DayOfWeek 38014 non-null int64

4 WeekOfYear 38014 non-null int64

5 Day 38014 non-null int64

6 Month 38014 non-null int64

7 Year 38014 non-null int64

8 PeriodOfDay 38014 non-null int64

9 ForecastWindProduction 38014 non-null object

10 SystemLoadEA 38014 non-null object

11 SMPEA 38014 non-null object

12 ORKTemperature 38014 non-null object

13 ORKWindspeed 38014 non-null object

14 CO2Intensity 38014 non-null object

15 ActualWindProduction 38014 non-null object

16 SystemLoadEP2 38014 non-null object

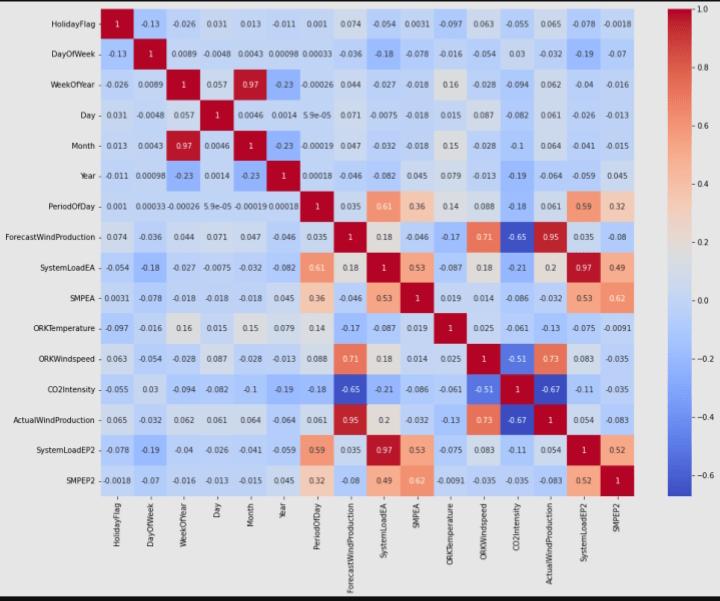
17 SMPEP2 38014 non-null object

dtypes: int64(7), object(11)

memory usage: 5.2+ MB

2.2

correlations = data.corr(method='pearson plt.figure(figsize=(16, 12)) sns.heatmap(correlations, cmap="coolwarm", annpt=True) plt.show()



**3.Prediction model:**

3.1 x=data[["Day","Month","ForecastWindProduction","SystemLoadEA","SMPEA","ORKTemperature","ORKWindspeed","CO2Intensity","ActualWindProduction","SystemLoadEP2"]] y = data["SMPEP2"] from sklearn.model\_selection import train\_test\_split xtrain,xtest,ytrain,ytest = train\_test\_split(x,y,test\_size = 0.2,random\_state=42)

3.2

from sklearn.ensemble import RandomForestRegressor model = RandomForestRegressor() model.fit(xtrain,ytrain)

Output:

RandomForestRegressor()

**4.Features:**

features = np.array([[10, 12, 54.10, 4241.05, 49.56, 9.0, 14.8, 491.32, 54.0, 4426.84]]) model.predict(features)

Output:

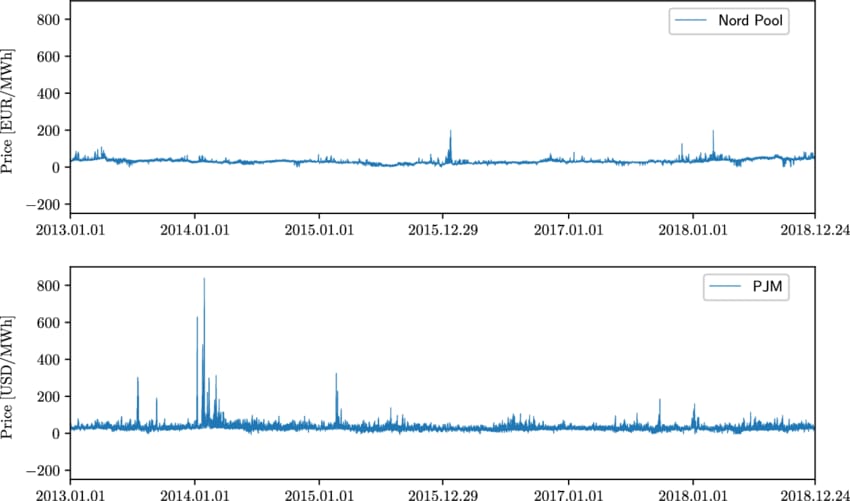
array([100.9588])

DIFFERENT ANALYSIS

1.Time series analysis:

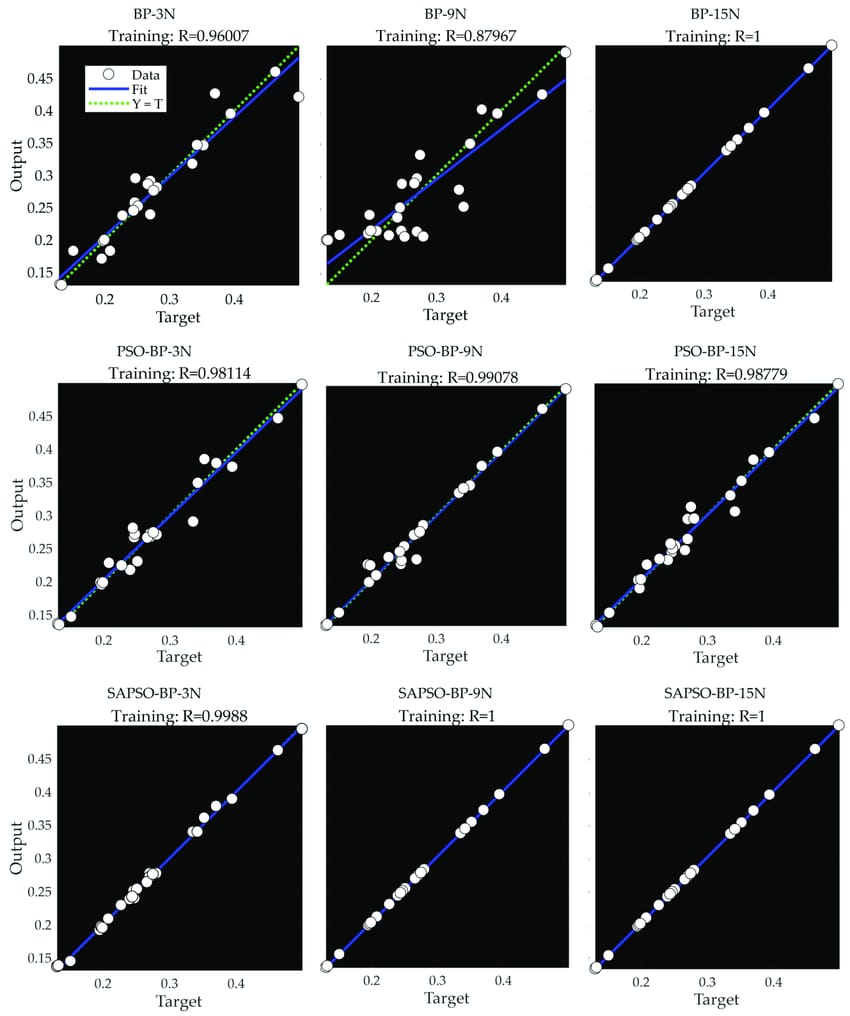
ARIMA (AutoRegressive Integrated Moving Average): This is a classic method for time series forecasting. It decomposes historical data into components like trend, seasonality, and noise, making it easier to make predictions.

Seasonal Decomposition of Time Series (STL): STL is a more advanced technique that can handle data with irregular seasonality and trends.

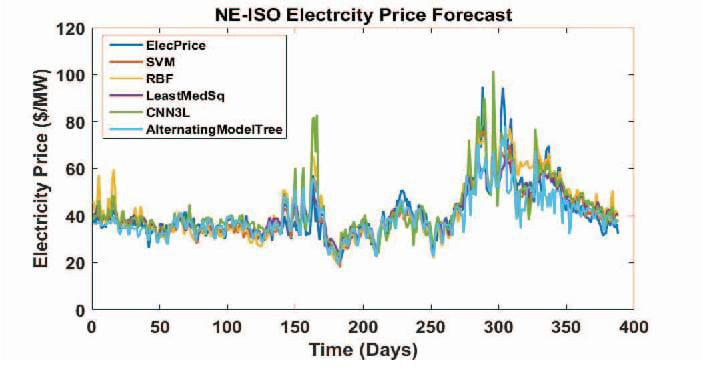


2. Machine Learning Models:

Regression Models: Linear regression, polynomial regression, or other regression techniques can be used to predict electricity prices by considering various features like historical prices, weather data, and demand.



Support Vector Machines (SVM): SVM can be applied for regression to predict electricity prices, especially when dealing with nonlinear relationships.



Random Forests and Gradient Boosting: These ensemble methods are powerful for capturing complex patterns in the data. They can handle various features and are robust to overfitting.

CONCLUSION

In Phase 3, we have uploaded the dataset and performed Exploratory Data Analysis and various analysis.