ELECTRICITY PRICES PREDICTION USING APPLIED DATA SCIENCE

Phase-4 submission document

Team members reg.no

S. SHYAM KUMAR 731221104033

P. RAVI KUMAR 731221104029

S. BABUSHANKAR 731221104005

E. DHAKSHINAMOORTHI 731221104009

Project Title: Electricity Prices Prediction

Phase 4: Development Part 2

Topic: Continue building the electricity prices prediction model by performing feature engineering, model training and evaluation.

**Introduction**

In the age of digital transformation, electricity is the only power source used by all applications. Therefore, a suitable mechanism is needed to assess the amount of power used for both home and industrial uses in our daily lives. Electricity price predictions give you information about how much you will have to pay as well as how much capacity you are using, how much is needed, and other relevant details.

This study work has designed a model to anticipate electricity by utilizing machine learning techniques, as a good analysis and prediction of electricity is necessary. Predicting the price of electricity is difficult since it depends on a variety of variables, including national wind, wind production, and natural causes. Therefore, a suitable mechanism is needed to assess the amount of power used for both home and industrial uses in our daily lives.

**Primary Price Determinants**

The interplay of supply and demand determines electricity prices. On the other hand, demand is somewhat price-dependent, while supply prices vary greatly depending on a number of factors. The short- to mid-term fundamental price determinants are outlined here in summary.

**• Demand**

In order to activate the energy sources that provide power, demand is crucial.

**• Vent**

Wind speed has a significant impact on the price of electricity because it generates a significant amount of the nation's electricity.

**• Condensation**

The quantity of snow and rain that hydro reservoirs retain can have a significant impact on the cost of energy.

**• Temperature**

Both the supply and demand for power are directly impacted by temperature, as is the other way around. Demand is impacted by the usage of electricity for heating. Given that temperature and wind are typically connected, temperature can have an impact on energy output. Temperature has an impact on hydropower as well; for instance, melting snow or ice can enhance output.

**• Prices of commodities**

The price of commodities has a significant impact on power pricing, as the overall energy consumption is heavily reliant on these commodities. As energy moves between zones, the prices of commodities also have an impact on other contraries, since energy systems become less reliant on these commodities.

**• Transmission capacity**

The division of various price ranges determined by the capacity of transmission between regions and national boundaries. The price in a particular area is influenced by the transmission capacity, which restricts the flow of electricity. The price in the two bidding areas would be the same if there were unlimited transmission capabilities.

**Model Training**

**Context: -**

Dataset containing the price of electricity for a data center in addition to factors that might affect the price.

**Dataset:-** <https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction>

**Column Descriptions:**

1. DateTime: String, defines date and time of sample
2. Holiday: String, gives name of holiday if day is a bank holiday
3. HolidayFlag: integer, 1 if day is a bank holiday, zero otherwise
4. DayOfWeek: integer (0–6), 0 Monday, day of week
5. WeekOfYear: integer, running week within year of this date
6. Day integer: day of the date
7. Month integer: month of the date
8. Year integer: year of the date
9. PeriodOfDay integer: denotes half hour period of day (0–47)
10. SystemLoadEA: the national load forecast for this period
11. SMPEA: the price forecast for this period
12. ORKTemperature: the actual temperature measured at Cork airport
13. ORKWindspeed: the actual windspeed measured at Cork airport
14. CO2Intensity: the actual CO2 intensity in (g/kWh) for the electricity produced
15. ActualWindProduction: the actual wind energy production for this period
16. SystemLoadEP2: the actual national system load for this period
17. SMPEP2: the actual price of this time period, the value to be forecasted

**Program Of Training :-**

import os

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import statsmodels.api as sm

import tensorflow as tf

import xgboost as xgb

import os

import warnings

from tensorflow.keras.layers import Dense, LSTM, Conv1D, MaxPooling1D, TimeDistributed, Flatten, Dropout, RepeatVector

from statsmodels.graphics.tsaplots import plot\_acf, plot\_pacf

from statsmodels.tsa.stattools import adfuller, kpss, ccf

from sklearn.metrics import mean\_squared\_error, r2\_score

from sklearn.preprocessing import LabelEncoder, StandardScaler, MinMaxScaler

from sklearn.decomposition import PCA

from sklearn.model\_selection import train\_test\_split

from math import sqrt

%matplotlib inline

for dirname, \_, filenames **in** os.walk('/kaggle/input'):

for filename **in** filenames:

print(os.path.join(dirname, filename))

**Importing the necessary libraries**:-

Import numpy as np

Import pandas as pd

From sklearn.model\_selection import train\_test\_split

From sklearn.metrics import mean\_squared\_error

From math import sqrt

Import keras

From keras.models import Sequential

From keras.layers import Dense

From sklearn.preprocessing import StandardScaler

**Reading the dataset:-**

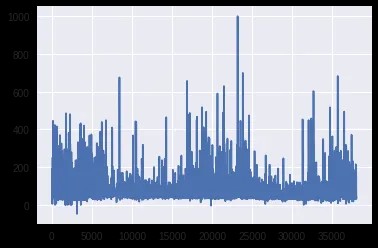
Df = pd.read\_csv(“/content/electricity\_prices.csv”, na\_values=[‘?’])

Df.head()

Df.shape

**Plotting the target feature**:-

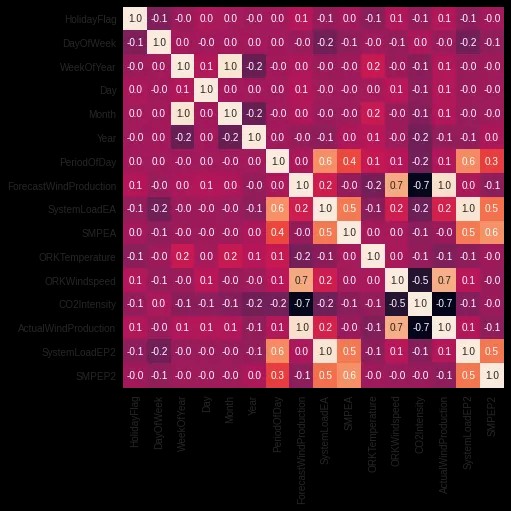
Plt.plot(“SMPEP2”, data=df)



**Correlation plot of Independent attributes:-**

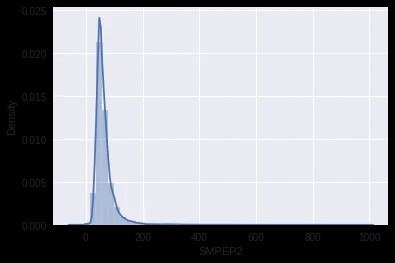
Plt.figure(figsize=(9,7))

Sns.heatmap(df.corr(), annot=True, square=True, fmt=’.1f’, cbar=False);



**Distribution plot of Target feature**:-

Sns.distplot(df[‘SMPEP2’])



**Splitting the independent features and target feature:-**

X = df[[‘ActualWindProduction’, ‘SystemLoadEP2’, ‘SMPEA’, ‘SystemLoadEA’, ‘ForecastWindProduction’,

‘DayOfWeek’, ‘Year’, ‘ORKWindspeed’, ‘CO2Intensity’, ‘PeriodOfDay’]]

Y = df[‘SMPEP2’]

**Compiling the model:-**

Model.compile(loss=’mse’, optimizer=’adam’, metrics=[‘mse’,’mae’])

## **Evaluating the model on test set:-**

**from** **sklearn.metrics** **import** mean\_absolute\_error,r2\_score

predictions = model.predict(X\_test)

print (f"MAE: **{**mean\_absolute\_error(y\_test, predictions)**}**")

print (f"R2\_score: **{**r2\_score(y\_test, predictions)**}**")

**Evaluation: -**

By using this analysis, we can predict electricity prices that is the actual price of this time period and forecast future business strategies.

**CONCLUSION**:

In conclusion, applied data science-based electricity price prediction is a useful instruction for resolving the issues with the energy market. Enhancing the precision of pricing projections through the application of sophisticated data analytics and machine learning methodologies can provide more informed choices for suppliers and customers alike. This strategy can help maximize energy use , cut expenses , and improve the energy sector’s sustainability and efficiency .