PREDICTING ELECTRICITY PRICE PREDICTION USING DEEP LEARNING

TEAM LEADER

723721243038: PRAJEESH P

Phase 3 Submission Document

Project: Electricity Price Prediction

Introduction:

- Electricity price prediction is a fascinating area of study that focuses on forecasting the future prices of electricity.
- It involves using historical data, along with various statistical and machine learning techniques, to analyze and predict the fluctuations in electricity prices.
- This prediction can be valuable for both consumers and energy providers, as it helps them make informed decisions regarding energy usage, pricing strategies, and resource allocation.
- The goal is to develop accurate models that consider factors such as market conditions, weather patterns, supply and demand dynamics, and regulatory policies to anticipate electricity price movements.
- By leveraging advanced analytics and data-driven insights, electricity price prediction can contribute to optimizing energy consumption, reducing costs, and promoting sustainability

Content for Project Phase 3 :

This phase 3 involves simply loading and preprocessing the dataset.

Loading and pre-processing a dataset involves reading the data, handling missing values, and transforming the data for analysis or model training.

Data Source:

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

Day	Month	Year	PeriodOfDay	ForecastWindProduction	SystemLoadEA	SMPEA	ORKTemperature	ORKWindspeed	CO2Intensity	ActualWindProduction	SystemLoadEP2	SMPEP2
1	11	2011	0	315.31	3388.77	49.26	6	9.3	600.71	356	3159.6	54.32
1	11	2011	1	321.8	3196.66	49.26	6	11.1	605.42	317	2973.01	54.23
1	11	2011	2	328.57	3060.71	49.1	5	11.1	589.97	311	2834	54.23
1	11	2011	3	335.6	2945.56	48.04	6	9.3	585.94	313	2725.99	53.47
1	11	2011	4	342.9	2849.34	33.75	6	11.1	571.52	346	2655.64	39.87
1	11	2011	5	342.97	2810.01	33.75	5	11.1	562.61	342	2585.99	39.87
1	11	2011	6	343.18	2780.52	33.75	5	7.4	545.81	336	2561.7	39.87
1	11	2011	7	343.46	2762.67	33.75	5	9.3	539.38	338	2544.33	39.87
1	11	2011	8	343.88	2766.63	33.75	4	11.1	538.7	347	2549.02	39.87
1	11	2011	9	344.39	2786.8	33.75	4	7.4	540.39	338	2547.15	39.87
1	11	2011	10	345.02	2817.59	33.75	4	7.4	532.3	372	2584.58	39.87
1	11	2011	11	342.23	2895.62	47.42	5	5.6	547.57	361	2641.37	39.87
1	11	2011	12	339.22	3039.67	44.31	5	3.7	556.14	383	2842.19	51.45
1	11	2011	13	335.39	3325.1	45.14	5	3.7	590.34	358	3082.97	51.45
1	11	2011	14	330.95	3661.02	46.25	4	9.3	596.22	402	3372.55	52.82
1	11	2011	15	325.93	4030	52.84	5	3.7	581.52	368	3572.64	53.65
1	11	2011	16	320.91	4306.54	59.44	5	5.6	577.27	361	3852.42	54.21
1	11	2011	17	365.15	4438.05	62.15	6	5.6	568.76	340	4116.03	58.33
1	11	2011	18	410.55	4585.84	61.81	8	7.4	560.79	358	4345.42	58.33
1	11	2011	19	458.56	4723.93	61.88	9	7.4	542.8	339	4427.29	58.33
1	11	2011	20	513.17	4793.6	61.46	?	?	535.37	324	4460.41	58.33
1	11	2011	21	573.36	4829.44	61.28	11	13	532.52	335	4493.22	58.27
1	11	2011	22	636.75	4888.29	61.63	11	22.2	534.34	372	4513.02	58.26
1	11	2011	23	683.59	4936.25	62.12	11	18.5	530.08	415	4490.71	58.26
1	11	2011	24	731.07	4995.51	62.83	11	22.2	517.55	513	4493.73	58.26
1	11	2011	25	780.23	5044.68	60.2	11	20.4	506.83	623	4481.31	58.15
1	11	2011	26	828.09	5018.8	56.25	12	20.4	513.98	683	4408.46	54.74
1	11	2011	27	873.81	4916.93	56.25	11	24.1	518.96	711	4341.14	54.74
1	11	2011	28	920.69	4933.87	56.25	12	22.2	525.69	761	4338.35	54.14
1	11	2011	29	985.09	4978.87	56.25	11	25.9	528.47	750	4294.17	53.63
1	11	2011	30	1044.37	5013.1	56.25	11	22.2	528.17	758	4318.87	53.63
1	11	2011	31	1098.97	5061.1	56.25	11	24.1	513.22	805	4375.62	53.63

Data Collection and Preprocessing:

✓ Importing the dataset: Obtain a comprehensive dataset containing relevant features such as ForeCastWind production, SystemLoadEA , SMPEA , ORKTemprature , ORKWindspeed 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 (electricity price prediction).

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

Use a data pre-processing library:

- There are a number of libraries available that can help with data preprocessing tasks, such as handling missing values, encoding categorical variables, and scaling the features.

Carefully consider the specific needs of your model:

- The best way to preprocess the data will depend on the specific machine learning algorithm that you are using.

- It is important to carefully consider the requirements of the algorithm and to preprocessthe data in a way that is compatible with the algorithm.

Validate the pre-processed data:

- It is important to validate the preprocessed data to ensure that it isin a format that can be used by the machine learning algorithm and that it is of high quality.

- This can be done by inspecting the data visually or by using statistical methods.

Validate the pre-processed data:

- It is important to validate the preprocessed data to ensure that it isin a format that can be used by the machine learning algorithm and that it is of high quality. This can be done by inspecting the data visually or by using statistical methods.

1.Loading the dataset:

- ✓ Loading the dataset using machine learning is the process of bringingthe data into the machine learning environment so that it can be used to train and evaluate a model.

- ✓ The specific steps involved in loading the dataset will vary dependingon the machine learning library or framework that is being used. However, there are some general steps that are common to most machine learning frameworks:

Identify the dataset:

- The first step is to identify the dataset that you want to load. Thisdataset may be stored in a local file, in a database, or in a cloud storageservice.

Load the dataset:

- Once you have identified the dataset, you need to load it into the machine learning environment.

- This may involve using a built-in function in the machine learning library, or it may involve writing yourown code.

Pre-process the dataset:

- Once the dataset is loaded into the machine learning environment, you may need to preprocess it before you can start training and evaluating your model. This may involve cleaning the data, transforming the data into a suitable format, and splitting the data into training and testing.

PROGRAM:

ELECTRICTY PRICE PREDICTION

IMPORTING REQUIRED PACKAGES

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import os

df = pd.read_csv("C:/Users/Lenovo/Desktop/Electricity updated.csv", low memory = False)

df
```

	HolidayFlag	DayOfWeek	WeekOfYear	Day	Month	Year	PeriodOfDay	ForecastWindProduction	SystemLoadEA	SMPEA	ORKTemperature	ORKWindspe
0	0	1	44	1	11	2011	0	315.31	3388.77	49.26	6	
1	0	1	44	1	11	2011	1	321.8	3196.66	49.26	6	1
2	0	1	44	1	11	2011	2	328.57	3060.71	49.1	5	1
3	0	1	44	1	11	2011	3	335.6	2945.56	48.04	6	
4	0	1	44	1	11	2011	4	342.9	2849.34	33.75	6	1
...	...	...	...	...	...	...	...	...	...	...	...	
38009	1	1	1	31	12	2013	43	1179.14	3932.22	34.51	6	2
38010	1	1	1	31	12	2013	44	1152.01	3821.44	33.83	5	2
38011	1	1	1	31	12	2013	45	1123.67	3724.21	31.75	4	2
38012	1	1	1	31	12	2013	46	1094.24	3638.16	33.83	5	1
38013	1	1	1	31	12	2013	47	1064	3624.25	33.83	5	1

CHECKING FOR NULL VALUES

```
df1 = df.isnull()

df1
```

	HolidayFlag	DayOfWeek	WeekOfYear	Day	Month	Year	PeriodOfDay	ForecastWindProduction	SystemLoadEA	SMPEA	ORKTemperature	ORKWinds
0	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
38009	False	False	False	False	False	False	False	False	False	False	False	
38010	False	False	False	False	False	False	False	False	False	False	False	
38011	False	False	False	False	False	False	False	False	False	False	False	
38012	False	False	False	False	False	False	False	False	False	False	False	
38013	False	False	False	False	False	False	False	False	False	False	False	

ADDING NULL VALUES

df1 = df.isnull().sum()

df1

```
Out[5]: HolidayFlag              0
        DayOfWeek                0
        WeekOfYear               0
        Day                      0
        Month                    0
        Year                     0
        PeriodOfDay              0
        ForecastWindProduction   0
        SystemLoadEA             0
        SMPEA                    0
        ORKTemperature           0
        ORKWindspeed             0
        CO2Intensity             0
        ActualWindProduction     0
        SystemLoadEP2            0
        SMPEP2                   0
        SystemLoadEP2.1          0
        SMPEP2.1                 0
        dtype: int64
```

CHECKING THE DATA TYPES

df2 = df.dtypes

df2

```
Out[6]: HolidayFlag               int64
        DayOfWeek                 int64
        WeekOfYear                int64
        Day                       int64
        Month                     int64
        Year                      int64
        PeriodOfDay               int64
        ForecastWindProduction   object
        SystemLoadEA             object
        SMPEA                    object
        ORKTemperature           object
        ORKWindspeed             object
        CO2Intensity             object
        ActualWindProduction     object
        SystemLoadEP2            object
        SMPEP2                   object
        SystemLoadEP2.1          object
        SMPEP2.1                 object
        dtype: object
```

 # REPLACING SPECIAL CHARACTER WITH NULL VALUE

```
df.replace(to_replace='?',value='0',inplace=True)
df
```

 # CHECKING FOR ANY SPECIAL CHARACTERS IN FEATURES

```
df4 = df[df['ORKWindspeed']=='?']
df4
```

 # COVERTING THE FILE AS CSV

```
df.to_csv(r'E:\file3.csv')
df
```

 # IMPORTING THE FILE

```
dtf1=pd.read_csv("E:/file3.csv")
dtf1
```

 # CONVERTING THE DATATYPES OF FEATURES

```
dtf1['ForecastWindProduction'] = dtf1['ForecastWindProduction'].astype(float)
dtf1['SystemLoadEA'] = dtf1['SystemLoadEA'].astype(float)
dtf1['SMPEA'] = dtf1['SMPEA'].astype(float)
dtf1
```

 # CHECKING THE DATA TYPES

```
dtf1.dtypes
```

```
Out[17]: Unnamed: 0                int64
         HolidayFlag               int64
         DayOfWeek                 int64
         WeekOfYear                int64
         Day                       int64
         Month                     int64
         Year                      int64
         PeriodOfDay               int64
         ForecastWindProduction  float64
         SystemLoadEA            float64
         SMPEA                   float64
         ORKTemperature            int64
         ORKWindspeed            float64
         CO2Intensity            float64
         ActualWindProduction      int64
         SystemLoadEP2           float64
         SMPEP2                  float64
         SystemLoadEP2.1         float64
         SMPEP2.1                float64
         dtype: object
```