**ADS\_Electricity Prices Prediction\_phase5**

**Phase\_5-Project Documentation**

**Problem Statement:**

Electricity prices are notoriously volatile, and can fluctuate significantly from hour to hour, day to day, and week to week. This volatility can make it difficult for businesses and consumers to plan their energy budgets and make informed decisions about energy consumption. Accurate electricity price forecasting can help to mitigate this uncertainty and provide a more stable and predictable pricing environment.

**Problem:** The aim of this project is to predict future electricity prices to empower consumers with information for making informed decisions regarding their energy consumption.

**Significance:** Accurate electricity price predictions can help consumers optimize their usage, reduce costs, and make greener choices.

**Objective:** Develop a robust forecasting model for electricity prices based on historical data, considering factors such as seasonality, demand, and external variables.

**Data Gathering:** Collect historical electricity price data along with relevant information.

**Data Cleaning:** Ensure data reliability by correcting errors and handling missing values.

**Feature Creation:** Generate new features for better predictions (e.g., day of the week, past price trends).

**Choosing a Prediction Method:** Decide on a suitable prediction method (ARIMA, LSTM).

**Training the Model:** Teach the model with cleaned and enhanced data to learn from past patterns.

**Checking Accuracy:** Assess model performance using metrics like Mean Absolute Error and Root Mean Squared Error.

**Design thinking process:**

1**. Empathize:** Understand the challenges consumers face with electricity prices. Listen to their concerns, such as cost control and the impact of price fluctuations.

2. **Define:** Clearly state the problem. We aim to predict electricity prices accurately to help consumers make informed decisions about their energy usage.

3**. Ideate:** Brainstorm ideas for solving the problem. Consider data sources, modeling techniques, and user-friendly features like price alerts or budget recommendations.

4**. Prototype:** Develop and test different electricity price prediction models. Experiment with various algorithms to find the most effective one.

5. **Test:** Evaluate the models' performance. Ensure they are accurate and practical for consumers. Make improvements as needed to enhance usability and accuracy.

**DATA SET :**

The dataset used in this project used from Kaggle: https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction.

**The dataset contains :**

Date and time records.

Whether a day is a holiday or not.

Electricity-related data like usage and prices.

Weather-related data such as temperature and wind speed.

It's a dataset that can help analyze how electricity usage, pricing, and environmental factors change over time, especially on holidays and with varying weather condition

**Dataset contain:**

The following columns will be available in the Electricity prices prediction data set:

**DateTime:** Timestamp indicating the date and time of the data record.

**Holiday:** Indicates whether the day is a holiday.

**HolidayFlag:** A flag or code related to holiday information.

**DayOfWeek:** The day of the week (e.g., Monday, Tuesday).

**WeekOfYear**: Week number within the year.

**Day:** Day of the month.

**Month:** Month of the year.

**Year:** Year of the data record.

**PeriodOfDay:** Time period within a day (e.g., morning, afternoon).

**ForecastWindProduction:** Forecasted wind energy production.

**SystemLoadEA:** Electricity system load for a specific area (EA).

**SMPEA:** Spot Market Price for Electricity (SMP) for area EA.

**ORKTemperature:** Temperature measurement at a location referred to as ORK.

**ORKWindspeed:** Wind speed measurement at a location referred to as ORK.

**CO2Intensity:** Carbon dioxide (CO2) intensity or emissions related to electricity generation.

**ActualWindProduction:** Actual wind energy production.

**SystemLoadEP2:** Electricity system load for another area (EP2).

**SMPEP2:** Spot Market Price for Electricity (SMP) for area EP2.

**Data preprocessing steps:**

**1. Data set Loading:**

Load the CSV dataset into a data structure for further processing.

**2. Data Inspection:**

Examine the dataset to understand its structure, columns, and potential issues.

**3. Handling Missing Values:**

Identify and address missing values, which may involve imputation or removal.

**4. Data Format Conversion:**

Ensure that date and time columns are in the correct format for time series analysis.

**5. Data Sorting:**

Sort the data based on the time or date column to maintain the chronological order.

**6. Data Scaling/Normalization:**

Apply scaling or normalization techniques to ensure consistent ranges for different features.

**7. Feature Engineering:**

Create additional features or variables, such as lag values or moving averages, to capture seasonality and trends.

**8. Outlier Handling:**

Detect and handle outliers in the data that may negatively impact model training.

**9. Data Splitting:**

Split the dataset into training and testing sets, preserving the temporal order of data.

**10. Data Transformation:**

If necessary, apply mathematical or statistical transformations to the data, such as differencing for stationarity.

**11. Data Visualization:**

Visualize the dataset to gain insights into patterns, trends, and potential factors affecting electricity prices.

These preprocessing steps are essential to prepare the dataset for modeling and analysis in an electricity price prediction project.

**Model Training**

**Algorithm Selection:**

In the context of electricity prices prediction, the choice of a suitable time series forecasting algorithm is a pivotal decision. After careful consideration, we have selected the ARIMA (AutoRegressive Integrated Moving Average) algorithm for several reasons:

**Seasonality and Trends:** ARIMA is well-suited for modeling time series data exhibiting both seasonality and trends, which are often observed in electricity price data. It can effectively capture short-term fluctuations and long-term trends in prices.

**Statistical Foundation:** ARIMA is rooted in statistical principles, making it a robust choice for understanding and modeling the underlying data patterns. It can accommodate various patterns such as autocorrelation and seasonality.

**Flexibility:** ARIMA models can be customized with parameters like order (p, d, q) to adapt to specific time series data characteristics. This flexibility enables us to fine-tune the model for improved forecasting accuracy.

**Widely Adopted**: ARIMA is a well-established and widely adopted forecasting method in the time series analysis domain. It has a proven track record of success in various industries, including energy and finance.

**Evaluation Metrics**

**Selection of Evaluation Metrics:**

To assess the performance of our electricity price prediction model, we have chosen the following evaluation metrics:

**Mean Absolute Error (MAE):**

MAE is selected as it measures the average magnitude of errors between predicted and actual electricity prices. It provides insight into the model's accuracy in terms of price prediction.

**Mean Squared Error (MSE):**

MSE is useful in quantifying the model's predictive accuracy by penalizing larger errors more heavily. It measures the average squared differences between predicted and actual prices.

Root Mean Squared Error (RMSE):

RMSE, derived from MSE, provides an interpretable measure of the model's predictive error in the same unit as the target variable. It is particularly helpful for understanding the magnitude of forecasting errors.

Visual Inspection:

Alongside quantitative metrics, visual inspection of the model's predictions compared to actual electricity prices is essential. It provides a holistic view of how well the model captures price fluctuations, trends, and seasonality.

**Structured Code directory :**

**# Dependencies Installation:**

pip install pandas

pip install numpy

import pandas as pd

import numpy as np

**# Data Visualization**

pip install matplotlib

pip install seaborn

import matplotlib.pyplot as plt

import seaborn as sns

**# Feature Engineering**

pip install scikit-learn

from sklearn.preprocessing import StandardScaler

from sklearn.feature\_selection import SelectKBest

**# Machine Learning Models**

from sklearn.linear\_model import LinearRegression

from sklearn.ensemble import RandomForestRegressor

import xgboost as xgb

import lightgbm as lgb

from fbprophet import Prophet

**# Deep Learning (for Advanced Tasks and Neural Networks)**

pip install tensorflow

pip install torch

pip install keras

import tensorflow as tf

import torch

import keras

**# Handling Missing Data (Imputation)**

from sklearn.impute import SimpleImputer

imputer = SimpleImputer(strategy='mean')

data['column\_with\_missing\_values'] = imputer.fit\_transform(data[['column\_with\_missing\_values']])

**# Encoding Categorical Data (One-Hot Encoding)**

from sklearn.preprocessing import OneHotEncoder

encoder = OneHotEncoder(sparse=False)

data\_encoded = pd.DataFrame(encoder.fit\_transform(data[['categorical\_column']))

data = pd.concat([data, data\_encoded], axis=1)

**# Splitting the Data Set into Training and Test Sets**

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

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size=0.2, random\_state=42)

**# Feature Scaling (Standardization)**

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

**CONCLUSION**

electricity price prediction project has equipped consumers with a valuable tool for navigating the dynamic world of energy costs. Through a meticulous design thinking approach and well-structured development phases, we've crafted a robust model that accounts for seasonality, trends, and external factors.

Our choice of the ARIMA algorithm, complemented by reliable evaluation metrics like MAE, MSE, and RMSE, reflects our dedication to providing accurate forecasts. This ensures that consumers can make informed decisions about their energy consumption with confidence.

Looking ahead, our project contributes to a future where consumers have the knowledge and tools to manage their energy costs efficiently and sustainably. It's a significant step towards a brighter and more cost-effective energy landscape.