**Electricity Prices Prediction - Innovation**

**1.Electricity prices prediction**

Predicting electricity prices refers to forecasting the future costs of electrical energy. It's a crucial task for energy companies, traders, and consumers to make informed decisions. This prediction involves analysing various factors, such as supply and demand dynamics, weather conditions, fuel prices, and infrastructure.To predict electricity prices, data-driven models, like machine learning and statistical techniques, are commonly used. These models analyze historical data to identify patterns and trends, allowing for the estimation of future prices. The goal is to provide accurate price forecasts, helping stakeholders manage costs and plan their energy usage efficiently.

**2. Dataset and its detail:**

The dataset used in this project is available on 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.
* Carbon dioxide emissions related to electricity generation.

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

**3. Details about columns :**

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

1. **DateTime**: Timestamp indicating the date and time of the data record.
2. **Holiday**: Indicates whether the day is a holiday.
3. **HolidayFlag**: A flag or code related to holiday information.
4. **DayOfWeek**: The day of the week (e.g., Monday, Tuesday).
5. **WeekOfYear**: Week number within the year.
6. **Day**: Day of the month.
7. **Month**: Month of the year.
8. **Year**: Year of the data record.
9. **PeriodOfDay**: Time period within a day (e.g., morning, afternoon).
10. **ForecastWindProduction**: Forecasted wind energy production.
11. **SystemLoadEA**: Electricity system load for a specific area (EA).
12. **SMPEA**: Spot Market Price for Electricity (SMP) for area EA.
13. **ORKTemperature**: Temperature measurement at a location referred to as ORK.
14. **ORKWindspeed**: Wind speed measurement at a location referred to as ORK.
15. **CO2Intensity**: Carbon dioxide (CO2) intensity or emissions related to electricity generation.
16. **ActualWindProduction**: Actual wind energy production.
17. **SystemLoadEP2**: Electricity system load for another area (EP2).
18. **SMPEP2**: Spot Market Price for Electricity (SMP) for area EP2.

This diverse dataset contains information related to electricity generation, consumption, market prices, weather conditions, and time-related details. It can be used for various analyses, including predicting electricity prices, forecasting energy demand, and assessing the environmental impact of electricity generation.

**4.Libraries used in the electricity prices prediction :**

**Data Manipulation and Analysis**:

* **Pandas**: For data loading, cleaning, and manipulation. It's essential for handling the CSV dataset.
* **NumPy**: For numerical operations and efficient array handling.

**To download and use pandas and numpy :**

Pip install pandas

Pip install numpy

import pandas as pd

import numpy as np

**Data Visualization** (optional, for data exploration):

* **Matplotlib** and **Seaborn**: For creating visualizations and exploring data distributions.

**To download and use matplotlib and seaborn:**

Pip install matplotlib.pyplot

Pip install seaborn

import matplotlib.pyplot as plt

import seaborn as sns

**Feature Engineering**:

* + **Scikit-Learn (sklearn)**: Provides tools for feature selection, scaling, and transformation.

**Use command to download and use:**

pip install scikit-learn

from sklearn.preprocessing import StandardScaler

from sklearn.feature\_selection import SelectKBest

**Machine Learning Models**:

* **Scikit-Learn (sklearn)**: Offers various regression models like Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor for price prediction.
* **XGBoost** or **LightGBM**: Popular gradient boosting libraries for regression tasks.
* **Prophet**: A time series forecasting library developed by Facebook.

**Use command** :

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):

* **TensorFlow** or **PyTorch**: Deep learning libraries for building neural networks.
* **Keras**: High-level API that works with both TensorFlow and PyTorch.

**Use commands to download and use**:

Pip install tensorflow

Pip install torch

Pip install keras

import tensorflow as tf

import torch

import keras

**5. To train and test the data set for electricity price prediction:**

1. Split the dataset into training and testing sets: The dataset will be split into two sets: a training set and a testing set. The training set will be used to train the models, and the testing set will be used to evaluate the performance of the models on unseen data.
2. Preprocess the data: The data will be pre-processed by scaling it and removing any outliers.
3. Train the forecasting models: The forecasting models will be trained on the training set.
4. Evaluate the forecasting models: The forecasting models will be evaluated on the testing set.

**6. Rest of explanation:**

The following advanced forecasting techniques will be used in this project:

* Prophet: Prophet will be used to forecast electricity prices using a statistical model that accounts for seasonality and other trends in the data.
* Deep learning models: A variety of deep learning models, such as RNNs and LSTMs, will be used to forecast electricity prices. Deep learning models can learn complex patterns in the data and make predictions based on those patterns.

**7. Metrics used for the accuracy check:**

The following metrics will be used to evaluate the accuracy of the forecasting models:

* Mean squared error (MSE): MSE is a measure of the average squared difference between the predicted and actual values.
* Mean absolute error (MAE): MAE is a measure of the average absolute difference between the predicted and actual values.
* Median absolute error (MedAE): MedAE is a measure of the median absolute difference between the predicted and actual values.
* R-squared: R-squared is a measure of the goodness of fit of the model. It ranges from 0 to 1, with higher values indicating a better fit.

**Conclusion:**

This document has outlined a plan for using advanced forecasting techniques, such as Prophet and deep learning models, to predict future electricity prices. The proposed approach will be evaluated using a variety of accuracy metrics, including MSE, MAE, MedAE, and R-squared.

* In addition to the features listed above, other features, such as weather data and demand data, may also be used to improve the accuracy of the forecasting models.
* The performance of the forecasting models may be affected by the length of the training set. It is important to use a training set that is long enough to capture the seasonal patterns and other trends in the data.
* The forecasting models can be used to predict electricity prices for different time horizons, such as hourly, daily, weekly, and monthly. The appropriate time horizon will depend on the specific needs of the user.