PREDICTING ELECTRICITY PRICE PREDICTION USING DEEP LEARNING

TEAM LEADER

723721243038: PRAJEESH P

Phase 4 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 4:

- This phase 4 involves feature engineering, model training, and evaluation.

Feature engineering, model training and evaluation of a dataset involves identifying target variables ,explore the date , remove redundant and irrelevant features .

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

Overview of the process:

The following is an overview of the process of building a electricity price prediction model by feature selection, model training, and evaluation:

1. Prepare the data: This includes cleaning the data, removing outliers, and handling missing values.

2. Perform feature selection: This can be done using a variety of methods, such as correlation analysis, information gain, and recursive feature elimination.

3. Train the model: There are many different machine learning algorithms that can be used for electricity price prediction. Some popular choices include linear regression, random forests, and gradient boosting machines.

4. Evaluate the model: This can be done by calculating the mean squared error (MSE) or the root mean squared error (RMSE) of the model's predictions on the held-out test set.

5. Deploy the model: Once the model has been evaluated and found to be performing well, it can be deployed to production so that it can be used to predict the electricity prices.

PROCEDURE:

Feature selection:

1. Identify the target variable. This is the variable that you want to predict, such as electricity price.
2. Explore the data. This will help you to understand the relationships between the different features and the target variable. You can use data visualization and correlation analysis to identify features that are highly correlated with the target variable.
3. Remove redundant features. If two features are highly correlated with each other, then you can remove one of the features, as they are likely to contain redundant information.
4. Remove irrelevant features. If a feature is not correlated with the target variable, then you can remove it, as it is unlikely to be useful for prediction.

Feature Selection:

We are selecting numerical features which have more than 0.50 or less than -0.50 correlation rate based on Pearson Correlation Method—which is the default value of parameter "method" in corr() function. As for selecting categori

PROGRAM:

```
import pandas as pd
import numpy as np
# importing dataset
dft1 = pd.read_csv("C:/Users/Lenovo/Downloads/Electricity.csv" ,
low_memory = False)
dft1
correlation_matrix = dtf1.corr()

# Get the absolute correlation values for the target variable
correlation_with_target = abs(correlation_matrix['SystemLoadEP2.1'])

# Sort the features based on their correlation with the target variable
sorted_features = correlation_with_target.sort_values(ascending=False)

# Select the top k features
k = 5  # Select top 5 features
selected_features = sorted_features[1:k+1]
# Exclude the target variable

# Print the selected feature names and their correlation values
print("Selected features and their correlation with the target variable:")
for feature, correlation in selected_features.items():
    print(f"{feature}: {correlation}")
```

```
Selected features and their correlation with the target variable:
SystemLoadEP2: 1.0
SystemLoadEA: 0.9725930860315541
PeriodOfDay: 0.5939813356813026
SMPEA: 0.5354531998087732
SMPEP2: 0.516934044199139
```

Model training:

1. Choose a machine learning algorithm. There are a number of different machine learning algorithms that can be used for electricity price prediction, such as linear regression, ridge regression.

PROGRAM:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Separate the features (X) and the target variable (y)
X = dataset.drop('SMPEP2', axis=1)
y = dataset['SMPEP2.1']

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a linear regression model
model = LinearRegression()

# Train the model on the training data
model.fit(X_train, y_train)

# Make predictions on the testing data
y_pred = model.predict(X_test)
```

 # Calculate the mean squared error

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)

```
Mean Squared Error: 1.4856130887245058e-26
```

RIDGE REGRESSION :

PROGRAM:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

 # Create a ridge regression model

model = Ridge(alpha=0.5) # You can adjust the alpha value

 # Train the model on the training data

model.fit(X_train, y_train)

 # Make predictions on the testing data

y_pred = model.predict(X_test)

 # Calculate the mean squared error

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)

```
Mean Squared Error: 3.802783296886913e-13
```

LASSO REGRESSION:

model = Lasso(alpha=0.5) # You can adjust the alpha value

 # Train the model on the training data

model.fit(X_train, y_train)

 # Make predictions on the testing data

y_pred = model.predict(X_test)

 # Calculate the mean squared error

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)

```
Mean Squared Error: 0.00024767657356100355
```

SUPPORT VECTOR MACHINE:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

 # Create a Support Vector Machine (SVM) model

model = SVR(kernel='linear') # You can choose different kernels like 'linear', 'rbf', 'poly', etc.

 # Train the model on the training data

model.fit(X_train, y_train)

 # Make predictions on the testing data

y_pred = model.predict(X_test)

 # Calculate the mean squared error

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)

```
Mean Squared Error: 0.006036091532677277
```

RANDOM FOREST REGRESSION:

 # Split the dataset into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

 # Create a Random Forest Regression model

model = RandomForestRegressor(n_estimators=100, random_state=42) # You can adjust the number of estimators

 # Train the model on the training data

model.fit(X_train, y_train)

 # Make predictions on the testing data

y_pred = model.predict(X_test)

 # Calculate the mean squared error

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)

Model evaluation:

ϖ Model evaluation is the process of assessing the performance of a machine learning model on unseen data. This is important to ensure that the model will generalize well to new data.

ϖ There are a number of different metrics that can be used to evaluate the performance of a electricity price prediction model.

Some of the most common metrics include:

• **Mean squared error (MSE):** This metric measures the average squared difference between the predicted and actual electricity prices.

- **Root mean squared error (RMSE):** This metric is the square root of the MSE.
- **Mean absolute error (MAE):** This metric measures the average absolute difference between the predicted and actual electricity prices.
- **R-squared:** This metric measures how well the model explains the variation in the actual electricity prices. In addition to these metrics, it is also important to consider the following factors when evaluating a electricity price prediction model:
- **Bias:** Bias is the tendency of a model to consistently over- or underestimate electricity prices.
- **Variance:** Variance is the measure of how much the predictions of a model vary around the true electricity prices.
- **Interpretability:** Interpretability is the ability to understand how the model makes its predictions. This is important for electricity price prediction models, as it allows users to understand the factors that influence the predicted