NAAN MUTHALVAN

IBM COLLABARATE

ARTIFICIAL INTELLIGENCE

PROJECT TITLE

MEASURE ENERGY CONSUMPTION

NAME: LOGESH E

DEPT & YEAR : CSE & III yr REG.NO : 712221104004

COLLEGE: PARK COLLEGE OF ENGINEERING AND

TECHNOLOGY

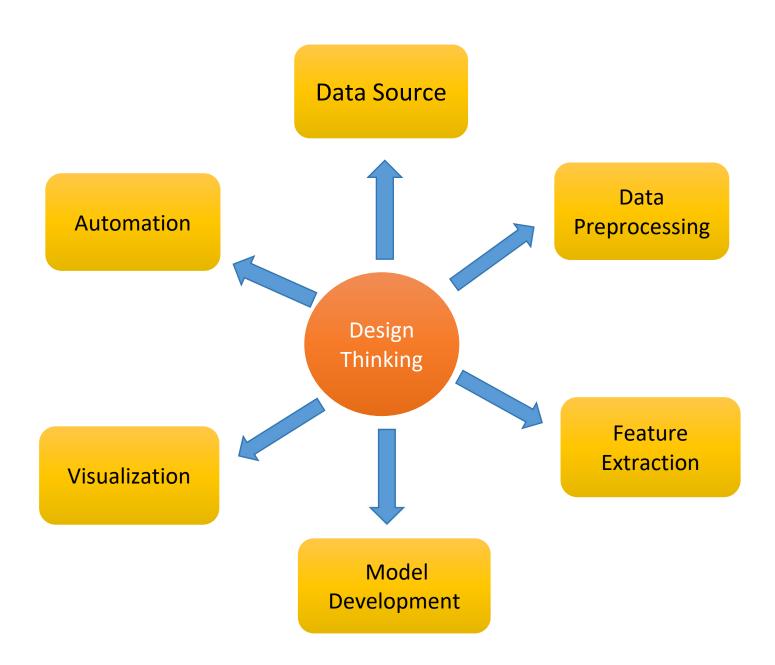
PHASE 1

PROBLEM DEFINITION AND DESIGN THINKING

PROBLEM DEFINITION

The problem at hand is to create an automated system that measures energy consumption, analysis the data, and provides visualizations for informed decision making. This solution aims to enhance efficiency, accuracy, and ease of understanding in managing energy consumption across various sectors.

DESIGN THINKING



PHASE 2

PROJECT INNOVATION IDEA

Steps:

- 1. Data Collection
- 2. Data Preprocessing
- 3. Build Machine Learning Model
- 4. Create Real Time Monitoring System
- 5. Build User Engagement
- 6. Data Analytics And Data Visualization
- 7. Deployment the Project

PHASE 3

DATA ANALYSIS AND PREPROCESSING

STEP 1:

Import library:

The first step is import the library files.

The library files are numpy for array calculation, pandas for data visualization, matplotlib for plotting, seaborn for statics graphics and also load the dataset using pandas.

STEP 2:

Reformat the Date Time Columns

Hence ,the large amout of dataset are reformated for required analysis and also include date time colums for time series process.

STEP 3:

Cleaning the dataset

The data cleaning process is must for analysis because the dataset have unordered ,anamoly and other soure data . Hence the data cleaning process is used to reduce this problem.

STEP 4:

Transforming the data

The data are reformatted for best analysis .this reduce the analysis process in less time.

STEP 5:

Show the Energy Consumption Each Year

The main process of analysis are done this section. The time series analysis is used to analyse energy consumption in each year.

Hence the analysis are shown by statistical plotting perspective.

STEP 6:

Data validation

The final process of data preprocessing is validate the data and show the data using plotting.

PHASE 4

BUILDING MODEL FOR ENERGY PREDICTION

- Hence, we are going to build our prediction model algorithm for measure energy consumption.
- It consist the process of loading the dataset,data transformation,feature importance,train/test split,Visualize feature to target relationship,modeling,Forecast on Test,Outlier Analysis,Reviewing: Train/Test Split,Feature Horizon,Lag Features,Train Using Cross Validation,Fold Analysis,Retraining on all Data and Predicting Future.

1.IMPORTING LIBRARIES AND DATA SET LOADING

CODE:

```
import numpy as np # linear algebra import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv) import seaborn as sns import matplotlib.pyplot as plt import xgboost as xg
```

from sklearn.metrics import mean_squared_error

```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
   for filename in filenames:
      print(os.path.join(dirname, filename))
```

```
df =
pd.read_csv("C:\\Users\\CYPHER\\Desktop\\archive\\PJME_hourly.csv")
df = df.set_index("Datetime")

df.index = pd.to_datetime(df.index)
df.head()
```

O/P:

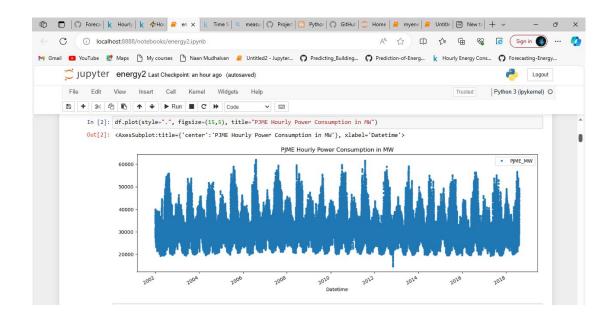
```
| File | Edit View | Insert | Cell | Kernel | Widgets | Help | Trusted | Python 3 (jpykernel) | Out[1]:

| PyME_MWW | Datetime | Datetime | 2002-12-31 03:00:000 | 245940 | 2002-12-31 03:00:000 | 245940 | 2002-12-31 03:00:000 | 245940 | 2002-12-31 05:00:000 | 245940 | 2002-12-31 05:00:000 | 24596.0 | 2002-12-31 05:00:000 | 24560.0 | 24560.0 | 24560.0 | 2002-12-31 05:00:000 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 24560.0 | 245600.0 | 24560.0 | 245600.0 | 245600.0 | 245600.0 | 245600.0 | 245600.0 | 245600.0 | 245600.0 | 2
```

2.TRAIN TEST SPLIT

CODE:

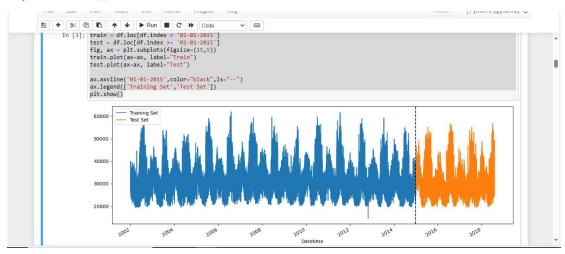
df.plot(style=".", figsize=(15,5), title="PJME Hourly Power Consumption in MW")



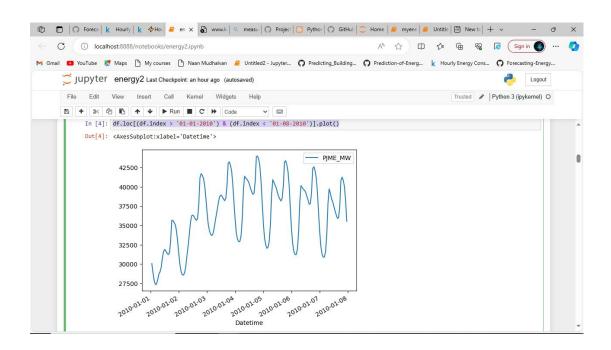
train = df.loc[df.index < '01-01-2015']
test = df.loc[df.index >= '01-01-2015']
fig, ax = plt.subplots(figsize=(15,5))
train.plot(ax=ax, label="Train")
test.plot(ax=ax, label="Test")

ax.axvline('01-01-2015',color="black",ls="--")

ax.legend(['Training Set','Test Set'])
plt.show()



df.loc[(df.index > '01-01-2010') & (df.index < '01-08
-2010')].plot()</pre>



3. FEATURE CREATION

```
def create_time_series_features(dataframe):
    df = dataframe.copy()
    df['hour'] = df.index.hour
    df['dayofweek'] = df.index.dayofweek
    df['quarter'] = df.index.quarter
    df['month'] = df.index.month
    df['year'] = df.index.year
    df['dayofyear'] = df.index.dayofyear

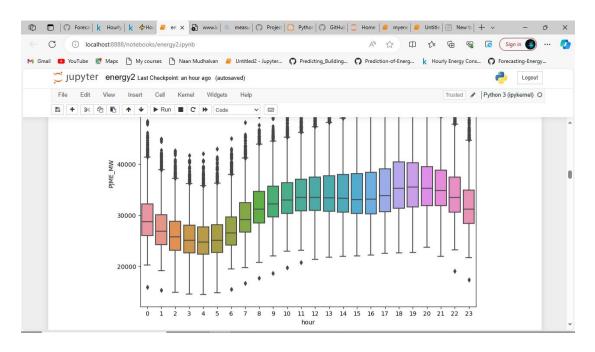
return df
```

df = create_time_series_features(df)

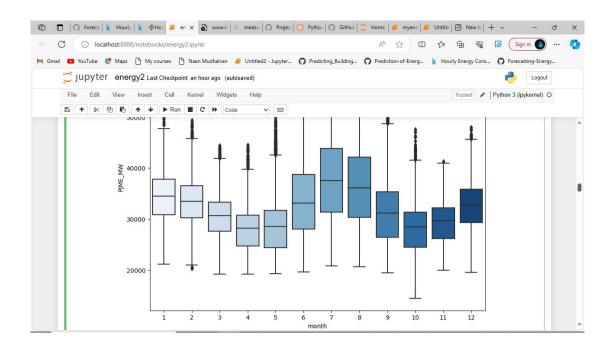
4. VISUALIZE FEATURE TO TARGET RELATIONSHIP

```
fig, ax = plt.subplots(figsize=(10,8))
sns.boxplot(data=df, x="hour",y="PJME_MW")
```

ax.set_title("MW By Hour") plt.show()

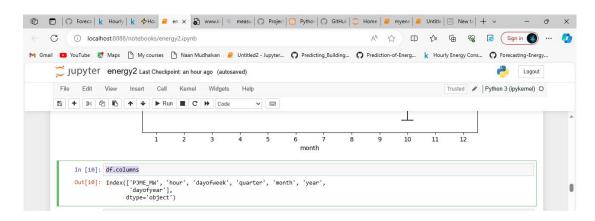


fig, ax = plt.subplots(figsize=(10,8))
sns.boxplot(data=df, x="month",y="PJME_MW",palette="Blues")
ax.set_title("MW By Month")
plt.show()



5. MODELING

df.columns



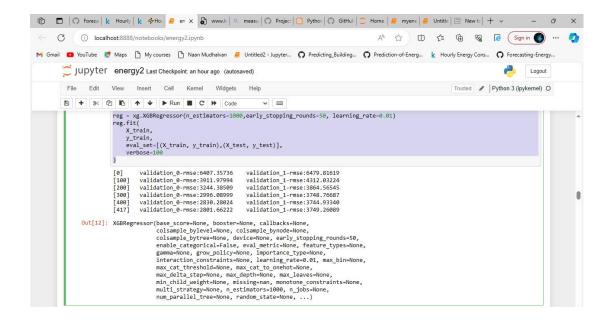
```
FEATURES = ['hour', 'dayofweek', 'quarter', 'month', 'year', 'dayofyear']
```

```
OUTPUT = ['PJME_MW']

train = create_time_series_features(train)
test = create_time_series_features(test)
X_train = train[FEATURES]
y_train = train[OUTPUT]

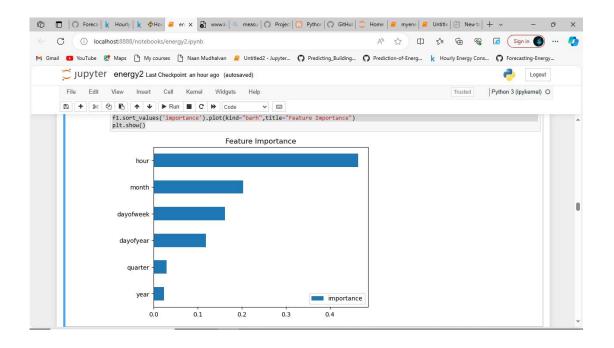
X_test = test[FEATURES]
y_test = test[OUTPUT]
```

```
reg =
xg.XGBRegressor(n_estimators=1000,early_stopping_rounds=5
0, learning_rate=0.01)
reg.fit(
    X_train,
    y_train,
    eval_set=[(X_train, y_train),(X_test, y_test)],
    verbose=100
)
```



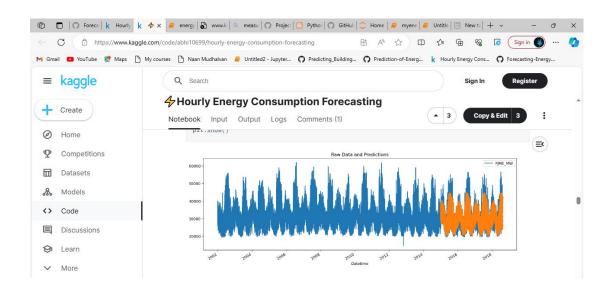
6.FEATURE IMPORTANCE

f1 = pd.DataFrame(data=reg.feature_importances_,
index=reg.feature_names_in_, columns=['importance'])
f1.sort_values('importance').plot(kind="barh",title="Feature
Importance")
plt.show()

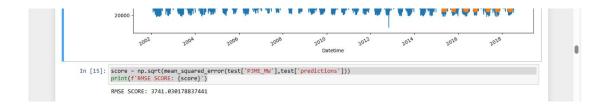


7.FEATURE FORECAST ON TEST

```
test['predictions'] = reg.predict(X_test)
df = df.merge(test[['predictions']], how='left',left_index=True,
right_index=True)
ax = df[['PJME_MW']].plot(figsize=(15,5))
df['predictions'].plot(ax=ax, style=".")
ax.set_title("Raw Data and Predictions")
plt.show()
```



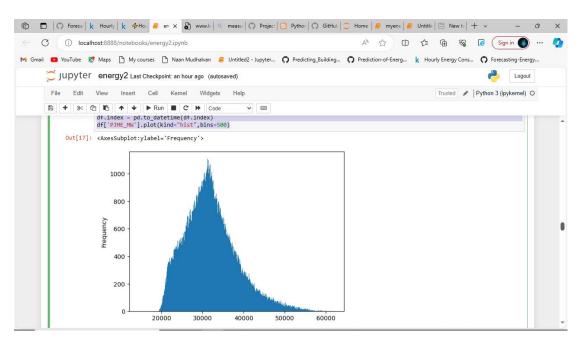
score =
np.sqrt(mean_squared_error(test['PJME_MW'],test['prediction
s']))
print(f'RMSE SCORE: {score}')



8. OUTLIER ANALYSIS

```
df =
pd.read_csv("C:\\Users\\CYPHER\\Desktop\\archive\\PJME_ho
urly.csv")
df = df.set_index("Datetime")
```

df.index = pd.to_datetime(df.index)
df['PJME_MW'].plot(kind="hist",bins=500)



 $df = df.query('PJME_MW > 19_000').copy()$

9. REVIEWING TRAIN AND TEST SPLIT

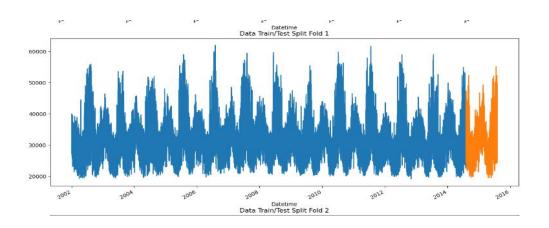
from sklearn.model_selection import TimeSeriesSplit tss = TimeSeriesSplit(n_splits=5, test_size=24*365*1,gap=24) df = df.sort_index() fig, axs = plt.subplots(5,1,figsize=(15,35))

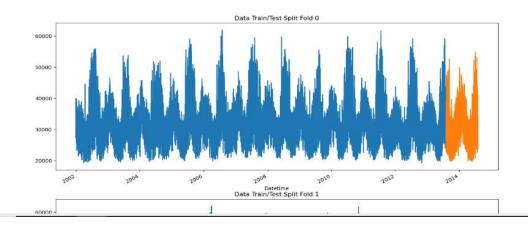
fold = 0
for train_idx, val_idx in tss.split(df):

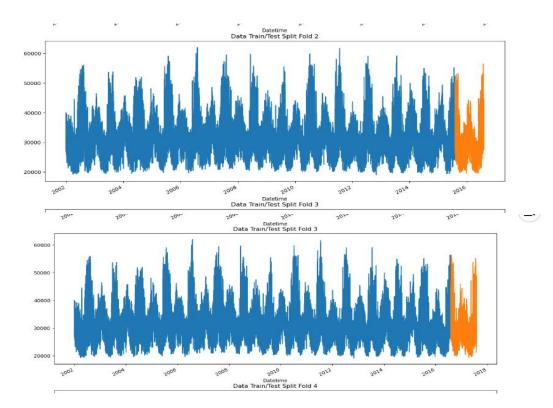
```
train = df.iloc[train_idx]
test = df.iloc[val_idx]

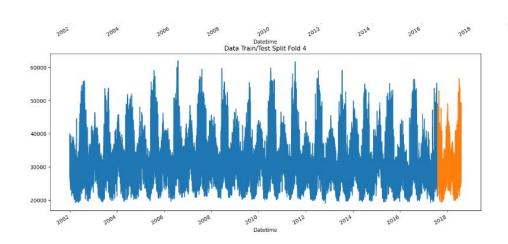
train['PJME_MW'].plot(
    ax=axs[fold],
    label="Training Set",
    title=f"Data Train/Test Split Fold {fold}"
)
test['PJME_MW'].plot(
    ax=axs[fold],
    label="Test Set",
)
```

fold += 1









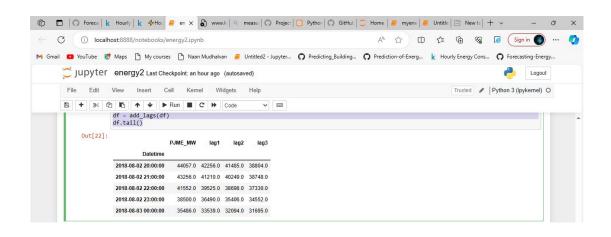
10. FEATURE HORIZON AND LAG FEATURES

```
df = create_time_series_features(df)

target_map = df['PJME_MW'].to_dict()
def add_lags(dframe):
    df = dframe.copy()
    df['lag1'] = (df.index - pd.Timedelta('364
days')).map(target_map)
    df['lag2'] = (df.index - pd.Timedelta('728
days')).map(target_map)
    df['lag3'] = (df.index - pd.Timedelta('1092
days')).map(target_map)

    return df

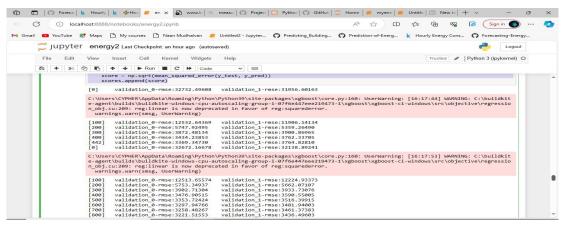
df = add_lags(df)
df.tail()
```

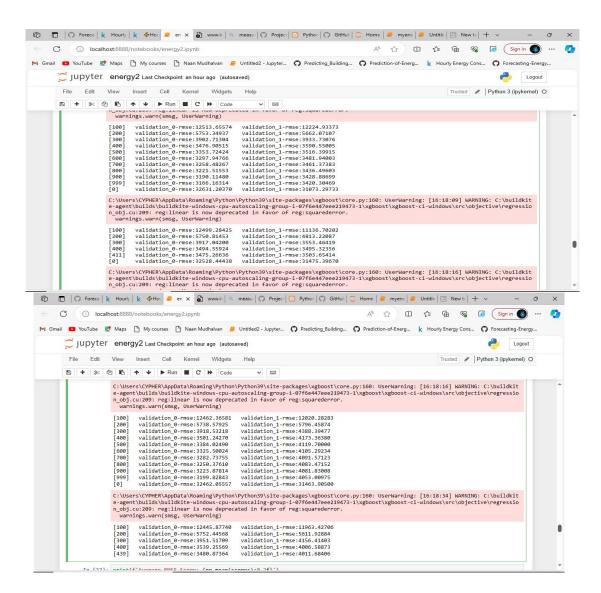


11. TRAIN USING CROSS VALIDATION

```
tss = TimeSeriesSplit(n_splits=5, test_size=24*365*1,gap=24)
df = df.sort_index()
df.columns
```

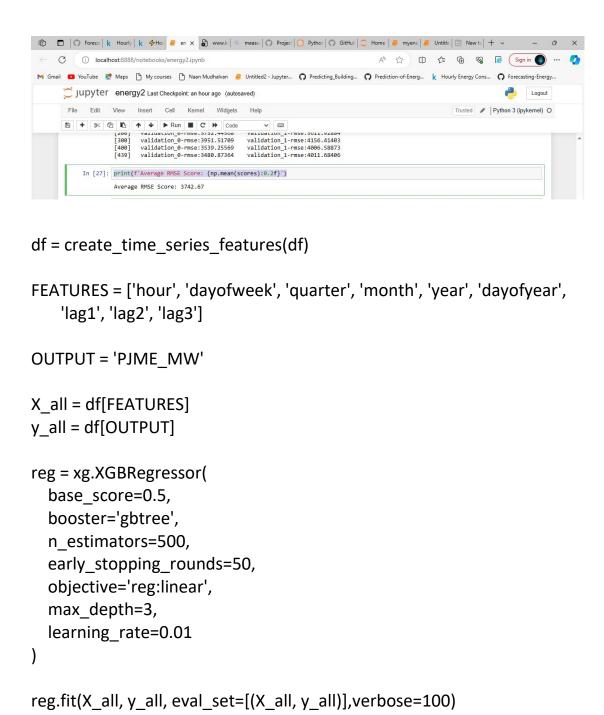
```
y train = train[OUTPUT]
X test = test[FEATURES]
y test = test[OUTPUT]
reg = xg.XGBRegressor(
  base score=0.5,
  booster='gbtree',
  n estimators=1000,
  early_stopping_rounds=50,
  objective='reg:linear',
  max depth=3,
  learning rate=0.01
reg.fit(
  X_train,
  y_train,
  eval_set=[(X_train, y_train),(X_test, y_test)],
  verbose=100
y_pred = reg.predict(X_test)
preds.append(y_pred)
score = np.sqrt(mean squared error(y test, y pred))
scores.append(score)
```

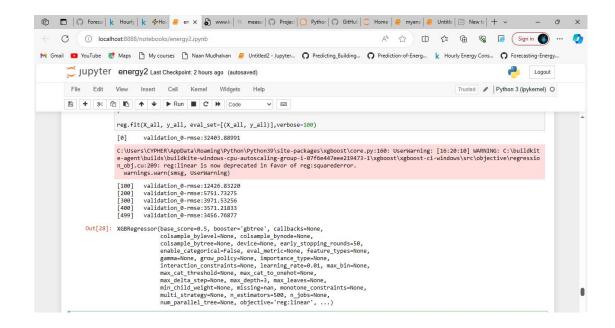




12. FOLD ANALYSIS & RETRAINING ON ALL DATA

print(f'Average RMSE Score: {np.mean(scores):0.2f}')





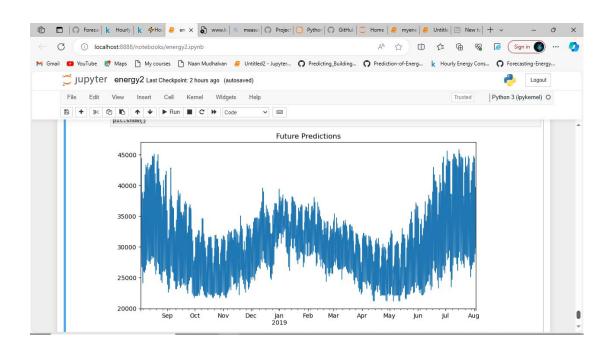
13.PREDICTING FUTURE

df.index.max()

```
File Edit View Insert Cell Kemel Widgets Help

| Company | Company
```

```
future = pd.date range('2018-08-03','2019-08-03',freq='1h')
  future_df = pd.DataFrame(index=future)
  future_df['isFuture'] = True
  df['isFuture'] = False
  df_and_future = pd.concat([df, future_df])
  df and future = create time series features(df and future)
  df and future = add lags(df and future)
  future_w_features = df_and_future.query('isFuture').copy()
  future_w_features['pred'] =
reg.predict(future w features[FEATURES])
  future_w_features['pred'].plot(
    figsize=(10,5),
    ms=1,
    lw=1,
    title="Future Predictions"
  )
  plt.show()
```



DEPLOYMENT THE PROJECT

- This is the final process of my project MEASURE ENERGY CONSUMPTION.
- Hence ,my project is running in real world environment using watson cloud or python flask.
- This shows my energy consumption prediction approximately related to original result in graphical manner.
- Then it is finally used for future purposed.