**MEASUREENERGYCONSUMPTION**

**NAME: SANJAY. S REGNO:110521106020**

**PROJECTTITLE:MEASUREENEGRYCONSUMPTION**

**PHASE04:DevelopmentPart2**

**TOPIC: To measure energy consumption by performing differentactivitieslikefeatureengineering,modeltraining,evaluationetc.**



**MACHINELEARNINGOPERATIONS**

Modelling&

Evaluation

DataEngineering

|  |  |  |  |
| --- | --- | --- | --- |
| DistributionTransformations | FeatureImportance | SyntheticData | AdvancedTechniques |

# INTRODUCTION:

Feature

engineering

Measureenergyconsumptionisusefulforperformingandprocessing of machine learning operations like data engineering,feature engineering, modelling & evaluation is also essential inthe field of data science. In the above flow diagram as shownseveraloperations.

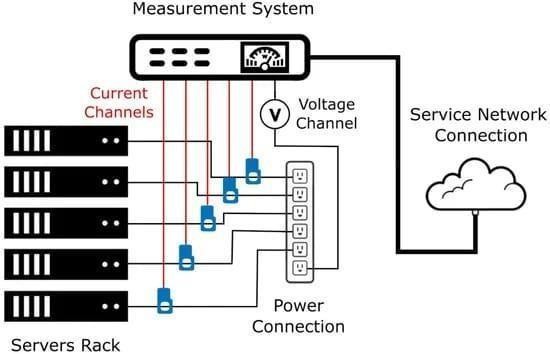
Itcontainsacrucialconsumptionmodelsandgainingvaluesinsightsfrom data.

# Datacollection:

The first step is to gather relevant data from various sources,ensuringit'sclean andwell-structured.

Data collection is used to gathering and storing large datasets canconsume energy, especially when dealing with data centers andcloud services. Using efficient data collection techniques can helpreduceenergy usage.

It also produces a digital transition that drives the new industrialrevolution is largely driven by the application of intelligence anddata.Thisboostleadstoanincreaseinenergy centers.



Thefigureshowsthedataconnectionarchitecturediagram.Themeasurementsystemisconnectedtotheserver’spowerinputto measure their energy consumption and send the informationovera computernetworks.

# Necessarysteptofollow:

1.Importlibraries:startbyimportingthenecessarylibraries.

## Program:

importpandasaspdimportnumpyasnp

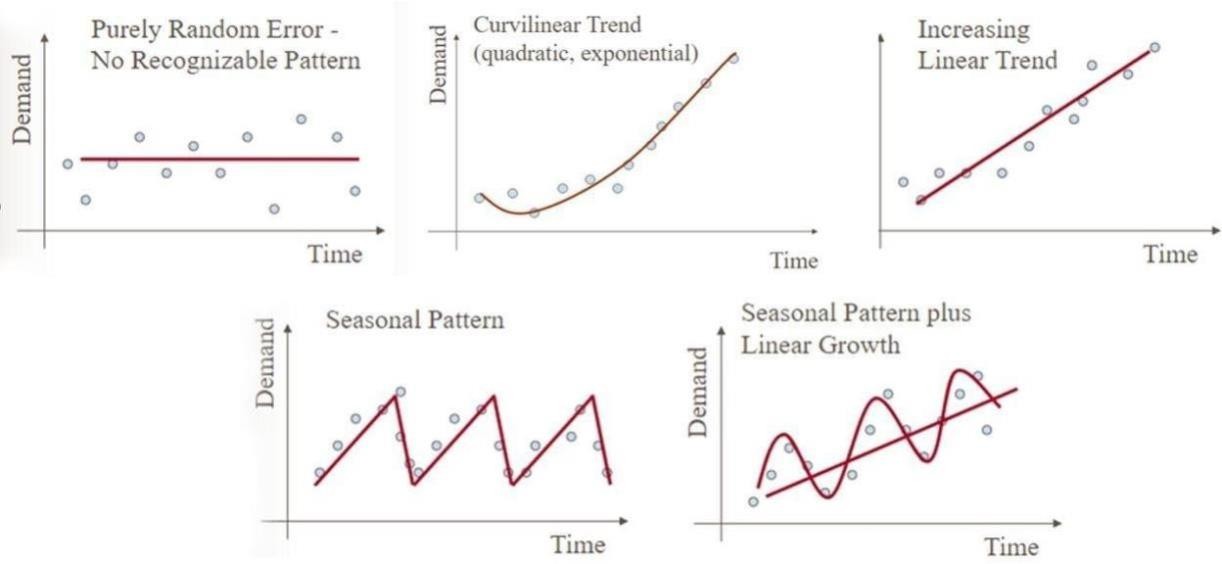
importmatplotlib.pyplotaspltimportseabornassns

mport pandas as pdimportxgboostasxgb

Timeseriesdatabyusingmean\_square\_red\_error:

importxgboostasxgb

fromsklearn.metricsimportmean\_squared\_errorcolor\_pal=sns.color\_palette()plt.style.use('fivethirtyeight')



The figure showsthatthe different types of time series dataforecastingwithmachinelearning.

**DataProcessing:**

Dataprocessinginvolvestaskslikedatacleaning,handlingmissingvalues,andencodingcategoricalvariablestopreparethedataforanalysis.

df = pd.read\_csv('../input/hourly-energy-consumption/PJME\_hourly.csv')

df = df.set\_index('Datetime')df.index=pd.to\_datetime(df.index)df.plot(style='.',

figsize=(15, 5),color=color\_pal[0],

title='PJMEEnergyUseinMW')plt.show()

**DatasetforDateTimePJMWMW:**

|  |  |  |
| --- | --- | --- |
| 0 | 2002-12-31  01:00:00 | 5077.0 |
| 1 | 2002-12-31  02:00:00 | 4939.0 |
| 2 | 2002-12-31  03:00:00 | 4885.0 |
| 3 | 2002-12-31  04:00:00 | 4857.0 |
| 4 | 2002-12-31  05:00:00 | 4930.0 |

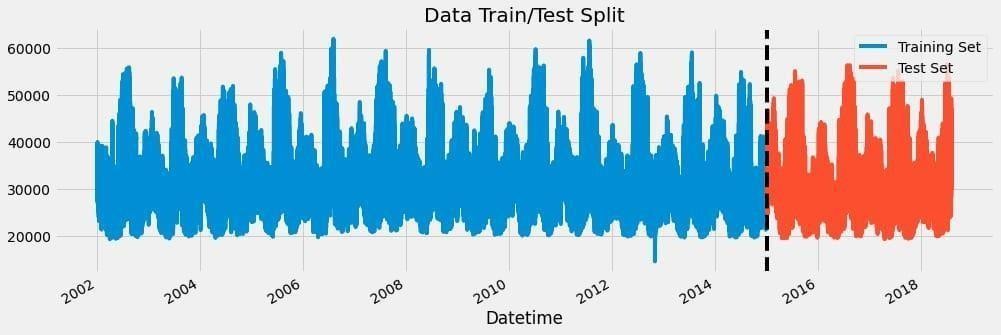
# Trainingandtestingofdatasplit:

Splittingdataintotrainingandtestingsetsiscrucialformeasuringenergyconsumptionusingtimeseriesinmachinelearning.Ithelpsevaluatethemodel'sperformanceonunseendata.

fig,ax=plt.subplots(figsize=(15,5))

train.plot(ax=ax,label='TrainingSet',title='DataTrain/TestSplit')

test.plot(ax=ax, label='Test Set')ax.axvline('01-01-2015',color='black',ls='--')ax.legend(['Training Set', 'Test Set'])plt.show()



# Featurecreation:

It involves creating new features from the existing data toimprovetheaccuracy oftheenergy consumptionprediction.

Inordertomeasureenergyconsumptionandforecasttimeseriesdata,playsa vital roleoffeatureengineering.

defcreate\_features(df):

df=df.copy()

df['hour'] = df.index.hourdf['dayofweek']=df.index.dayofweekdf['quarter'] = df.index.quarterdf['month'] = df.index.monthdf['year'] = df.index.yeardf['dayofyear'] = df.index.dayofyeardf['dayofmonth'] =df.index.day

df['weekofyear'] = df.index.isocalendar().weekreturndf

df=create\_features(df)

//creationoftimeseriesfeaturesbasedontimeseriesindex.

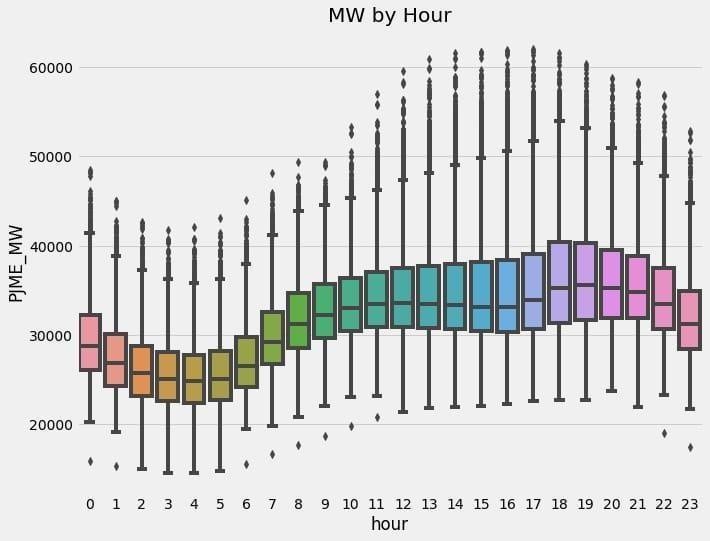
# Visualizationoffeatureandtargetrelationship:

Ithelpstounderstandpatternandcorrelations.

It also allows to identify the features of impact in energyconsumption.

fig, ax = plt.subplots(figsize=(10, 8))sns.boxplot(data=df, x='hour', y='PJME\_MW')ax.set\_title('MWby Hour')

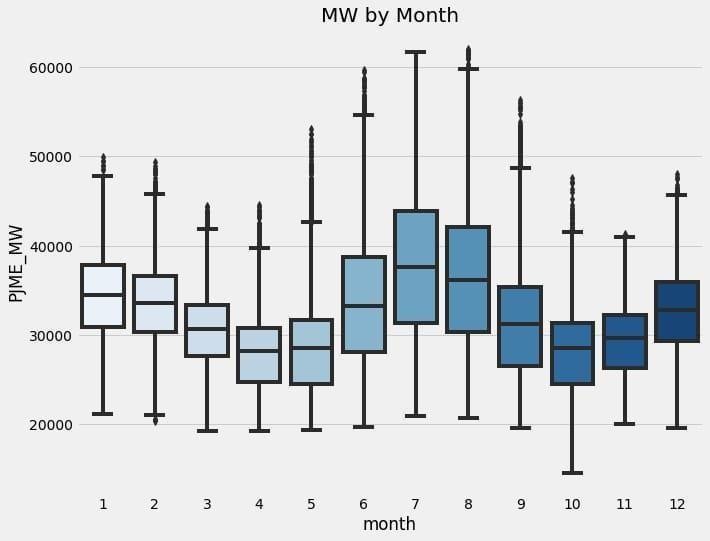
plt.show()



fig,ax=plt.subplots(figsize=(10,8))

sns.boxplot(data=df, x='month', y='PJME\_MW',palette='Blues')

ax.set\_title('MWbyMonth')plt.show()



# Creationofmodelsandtraining:

Creation of models is choosing an appropriate machine learningalgorithmthatmatchestheproblemathand,andfine-tuninghyperparametersforoptimalperformance.

Training the selected model on the preprocessed data to learnpatternsand relationshipswithinthe data.

train=create\_features(train)test=create\_features(test)

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

TARGET='PJME\_MW'

X\_train=train[FEATURES]y\_train=train[TARGET]

## Programbyusingregressionfunction:

X\_test=test[FEATURES]y\_test=test[TARGET]

reg=xgb.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)

## Output:

XGBRegressor(base\_score=0.5, booster='gbtree',callbacks=None,

colsample\_bylevel=1, colsample\_bynode=1,colsample\_bytree=1,

early\_stopping\_rounds=50,enable\_categorical=False,

eval\_metric=None,gamma=0,gpu\_id=-1,grow\_policy='depthwise',

importance\_type=None,interaction\_constraints='',

learning\_rate=0.01, max\_bin=256,max\_cat\_to\_onehot=4,

max\_delta\_step=0,max\_depth=3,max\_leaves=0,min\_child\_weight=1,

missing=nan, monotone\_constraints='()',n\_estimators=1000,

n\_jobs=0,num\_parallel\_tree=1,objective='reg:linear',predictor='auto',random\_state=0,reg\_alpha=0,...)

# Featureimportanceofmodel:

Featureimportancereferstodeterminingthecontributionofeachfeatureinpredictingenergyconsumption.

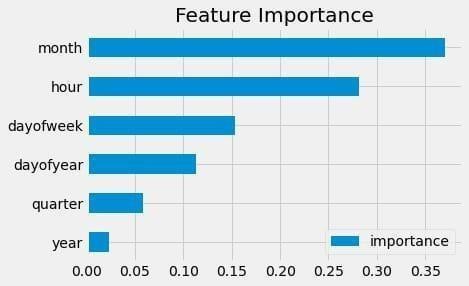
It helps identify which factors have the most influence on energyconsumptionpatterns.Thisinformationcanguidedecision-makingand helpprioritizeactionstooptimize energy usage.

Various techniques, such as permutation importance or featureimportancefromtree-based models, canbeusedtocalculatefeatureimportance.

fi = pd.DataFrame(data=reg.feature\_importances\_,index=reg.feature\_names\_in\_,columns=['importance'])

fi.sort\_values('importance').plot(kind='barh', title='FeatureImportance')

plt.show()



# Forecastingondataandtestprediction:

Thisallowsyoutoevaluatethemodel'sperformanceandaccuracyinpredictingenergy consumption.

Forecasting a data is also used for model on historical data andthentestitspredictingenergy consumption.

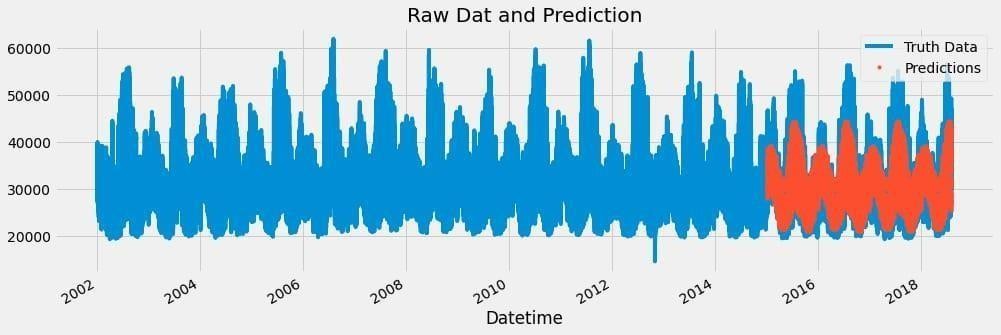
Bycomparingthemodel'spredictionswiththeactualenergyconsumption values, you can assess how well the model is able toforecastfuture energy consumptionpatterns.

It'saneffectivewaytovalidateandfine-tunethemachinelearningmodelforaccurateenergyconsumptionpredictions.

test['prediction']=reg.predict(X\_test)

df=df.merge(test[['prediction']],how='left',left\_index=True,right\_index=True)

ax = df[['PJME\_MW']].plot(figsize=(15, 5))df['prediction'].plot(ax=ax, style='.')plt.legend(['Truth Data', 'Predictions'])ax.set\_title('Raw Dat and Prediction')plt.show()



ax=df.loc[(df.index>'04-01-2018')&(df.index<'04-08-2018')]['PJME\_MW'] \

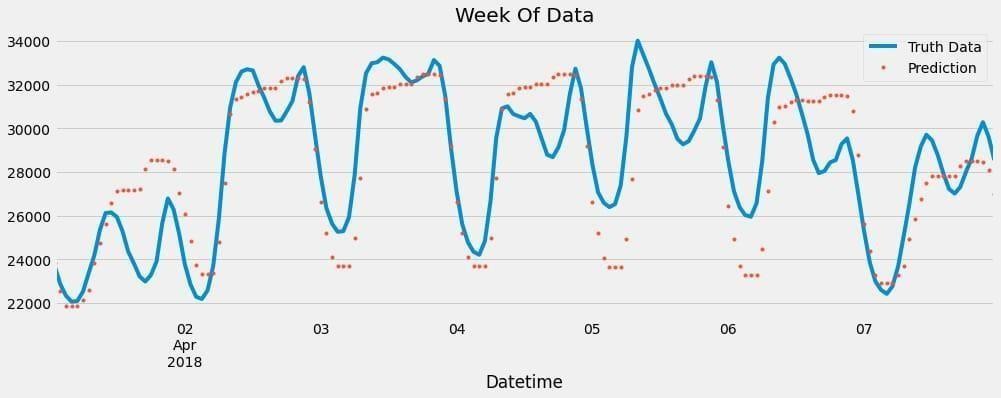
.plot(figsize=(15,5),title='WeekOfData')

df.loc[(df.index>'04-01-2018')&(df.index<'04-08-2018')]['prediction']\

.plot(style='.')

plt.legend(['TruthData','Prediction'])

plt.show()



# EvaluateandCalculatingErrors:

To evaluate and calculate the error in energy consumptionprediction,

Itperformsavariousmetricslikeabsoluteerror(MAE),rootmeansquarederror(RMSE),ormeanabsolutepercentageerror(MAPE).

These metrics help quantify the differencebetween thepredictedenergyconsumptionvaluesandtheactualvalues.

By calculating the error, it also assess the accuracy andperformance ofthepredictionmodel.

test['error']=np.abs(test[TARGET]-test['prediction'])

test['date']=test.index.date

test.groupby(['date'])['error'].mean().sort\_values(ascending=False).head(10)

|  |  |
| --- | --- |
| **Output:** |  |
| date |  |
| 2016-08-13 | 12839.595459 |
| 2016-08-14 | 12780.209554 |
| 2016-09-10 | 11356.302002 |
| 2015-02-20 | 10965.976237 |
| 2016-09-09 | 10864.953451 |
| 2018-01-06 | 10506.844889 |
| 2016-08-12 | 10124.050618 |
| 2015-02-21 | 9881.798503 |
| 2015-02-16 | 9781.549805 |
| 2018-01-07 | 9739.143555 |

Name:error,dtype:float64

**Conclusion:**

In conclusion, performing different activities like featureengineering, model training, and evaluation is crucial foraccurate energyconsumption prediction.

Featureengineeringhelps inselectingand transformingrelevant variables, such as time-based features, weatherdata,andlaggedvariables.

Modeltraininginvolvestrainingthemachinelearningmodelonhistoricaldatatomakeaccuratepredictions.

Evaluation is done by comparing the model's predictionswith the actual energy consumption values using metricslike MAEorRMSE.

Byfollowingthisprocess,wecanbuildareliablemodelthatcaneffectivelypredictenergyconsumptionpatterns.