**NATURAL GAS**

**PRICE PREDICTION USING MACHINE LEARNING**

# PROFESSIONAL TRAINING REPORT

**at**

**Sathyabama Institute of Science and Technology (Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering

By

## Avula Nanda Gopal

**REG. NO. 39110101**

****

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI – 600119, TAMILNADU**

**APRIL2021**

|  |  |  |
| --- | --- | --- |
| **Picture 12** | **SATHYABAMA**  **INSTITUTE OF SCIENCE AND TECHNOLOGY**  **(DEEMED TO BE UNIVERSITY)** **Accredited with Grade “A” by NAAC**  (Established under Section 3 of UGC Act, 1956)  JEPPIAAR NAGAR, RAJIV GANDHI SALAI  CHENNAI– 600119  [**www.sathyabama.ac.in**](http://www.sathyabama.ac.in) | Picture 11 |

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **Avula Nanda Gopal (Reg. No: 39110101)** who carried out the project entitled “**Natural Gas Price Prediction Using Machine Learning**” under my supervision from Jan 2022 to May 2022.

## Internal Guide

## Dr. D.Usha Nandini

**Head of the Department**

## Submitted for Viva voce Examination held on

**InternalExaminer ExternalExaminer**

**DECLARATION**

I, **Avula Nanda Gopal** hereby declare that the project report entitled **NATURAL GAS PRICE PREDICTION USING MACHINE LEARNING** done by me under the guidance of

**Dr.D.Usha Nandini.,** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering.

## DATE: Picture 22

**PLACE: SIGNATURE OF THE CANDIDATE**

**CERTIFICATE**

**ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D**, **Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of **Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.D.Usha Nandini.,**for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

**ABSTRACT**

In [machine learning](https://builtin.com/machine-learning), there’s something called the “[No Free Lunch](http://www.no-free-lunch.org/)” theorem. In a nutshell, it states that no one machine learning algorithm works best for every problem, and it’s especially relevant for supervised learning (i.e. predictive modeling).

As a result, you should try many different algorithms for your problem, while using a hold-out “test set” of [data](https://builtin.com/data-science) to evaluate performance and select the winner.Of course, the algorithms you try must be appropriate for your problem, which is where picking the right machine learning task comes in.

Machine learning algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y): Y = f(X)

This is a general learning task where we would like to make predictions in the future (Y) given new examples of input variables (X). We don’t know what the function (f) looks like or its form. If we did, we would use it directly and we would not need to learn it from data using machine learning algorithms.

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| FIGURE NO | FIGURE NAME | PAGE NO |
| 1.1 | Decision Tree | 02 |
| 3.1 | Tools used for programming | 5 |
| 4.1 | Initialising dataset | 10 |
| 4.2 | Printing data values | 11 |
| 4.3 | Importing packages | 12 |
| 4.4 | Analysing data and developing graphs | 13 |
| 4.5 | Pair plot | 14 |
| 4.6 | Splitting dataset | 15 |
| 4.7 | Accuracy and prediction | 15 |
| 4.8 | Result | 17 |

|  |  |  |
| --- | --- | --- |
| **CHAPTER No.** | **TITLE** | **PAGE NO** |
|  | **ABSTRACT** | i |
| **LIST OF FIGURES** | ii |
| **1.** | **INTRODUCTION** | **1** |
|  | 1.1 Machine Learning | 1 |
|  | 1.2 Decision tree regression | 2 |
| **2.** | **AIM AND SCOPE** | **3** |
|  | 2.1 Aim | 3 |
|  | 2.2 Scope | 3 |
| **3** | **METHODS AND STEPS USED** | 4 |
|  | 3.1 Ideation map | 4 |
|  | 3.2 Software Used | 5 |
|  | 3.3 Implementation | 6 |
|  | 3.4Algorithm | 6 |
|  | 3.5 Comment lines | 7 |
|  | 3.5.1 For gas price prediction | 7 |
|  | 3.5.2 Page creation | 9 |
|  | 3.6 Final outcome | 10 |
| **4** | **RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS** | 11 |
|  | 4.1 Result Analysis | 11 |
|  | 4.2 Performance Analysis | 16 |
| **5.** | **SUMMARY** | 17 |
| **6.** | **CONCLUSIONS** | 17 |
| **7** | **REFERENCES** | 19 |
| **8** | **BIBILIOGRAPHY** | 20 |

**1.INTRODUCTION**

**1.1 MACHINE LEARNING**

Machine learning is a branch of artificialintelligence(AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

Machine Learning can be a Supervised or Unsupervised. If you have lesser amount of data and clearly labelled data for training, opt for Supervised Learning. Unsupervised Learning would generally give better performance and results for large data sets. If you have a huge data set easily available, go for deep learning techniques.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

**1.2 DECISION TREE REGRESSION**

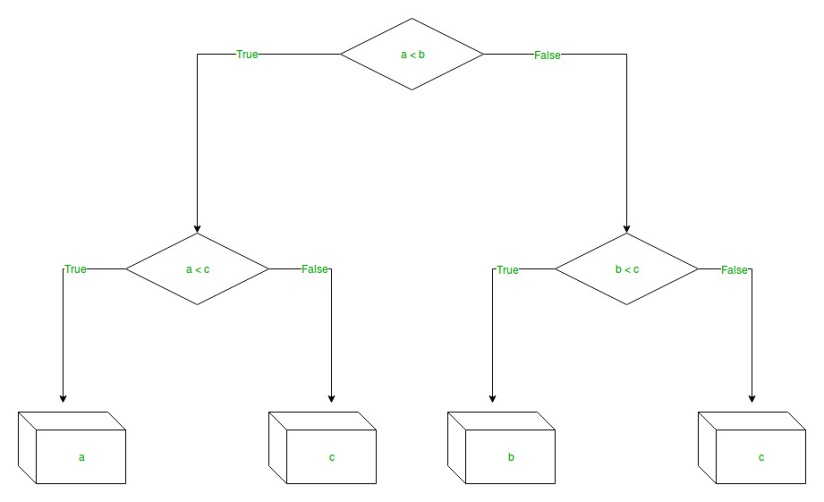
Decision Tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs, and utility.  
Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables.

Fig 1.1 Decision tree example

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

Decision Tree is one of the most commonly used, practical approaches for supervised learning. It can be used to solve both Regression and Classification tasks with the latter being put more into practical application. It is a tree-structured classifier with three types of nodes.

Decision Trees (DTs) are a non-parametric supervised learning method. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

**2 AIM AND SCOPE**

**2.1 AIM**

* + The project is about predicting the price of natural gas using machine learning.
  + We try to achieve a result above 90% score as they depict how good our machine is.
  + We try to achieve this by removing most of the null values in the data sheet.
  + To predict Natural gas price using machine learning.
  + To understand decision tree regression algorithm.
  + To execute python code to predict.
  + To use python packages to plot and predict price changes.

**2.2 SCOPE**

* + - * + Result above 90% score as they depict how good our machine is.
        + Forecasting natural gas prices is a powerful and essential tool which has become more important for different stakeholders in the natural gas market, allowing them to make better decisions for managing the potential risk, reducing the gap between the demand and supply, and optimizing the usage of resources based on accurate predictions.
        + The machine can predict from seven different cover types in four different wilderness areas. Therefore the result varies from 1 to 7.

**3 METHODS AND STEPS USED**

**3.1 IDEATION MAP**

START

INITIALIZE DATASET

IMPORT required packages

Analysing the data and removing the null values.

Training the machine using train-split algorithm.

Predicting the value using natural gas price prediction.

Accuracy is checked

STOP

**3.2 SOFTWARE USED**

Operating System - Windows

Coding Language - Python

Tool -

Jupyter Notebook

Anaconda

IDLE

Libraries -

webbrowser

pandas

numpy

sklearn

DecisionTreeClassifier

seaborn

train\_test\_split

accuracy\_score

matplotlib.pyplot,

****

Fig 3.1 tools used for programming - anaconda

**3.3 IMPLEMEMTATION**

**Steps involved in this project :**

* + - * + Collection of data and analysing the data.
        + Accuracy of the model is important so removing of the null values is important, thus analyzing is important.
        + The analysed data is then used to train our model.
        + Using the train – test split algorithm to train our algorithm.
        + Using the trained model to predict the price of the gas.
        + Using the algorithm to predict the rate of natural gas the result.
        + Calculation of the accuracy value, which if higher than 90% tells us that our model work great.

**3.4 ALGORITHM**

START

Step 1: Import the required libraries.

Step 2: Initialise and print the Dataset.

Step 3: Select all the rows and column from the dataset to “X”

Step 4: Select all of the rows and column from the dataset to “y”.

Step 5: Fit decision tree regressor to the dataset

Step 6: Predicting a new value

Step 7: Visualising the result

STOP

**3.5 COMMENT LINES**

*3.5.1 CODE FOR PRICE PREDICTION*

*import pandas as pd*

*rd=pd.read\_csv(r"/content/daily\_csv.csv")*

*rd.head()*

*rd.tail()*

*rd.info()*

*print(rd.size)*

*print(len(rd))*

*rd.dtypes*

*print(rd.values)*

*print(rd.isnull().sum())*

*#finding outliers for each attributes*

*import pandas as pd*

*import numpy as np*

*from sklearn.tree import DecisionTreeClassifier*

*import seaborn as sns*

*m=np.mean(rd['Price'])*

*print("")*

*print("mean of Price: ", m)*

*std=np.std(rd['Price'])*

*print("Satandard deviation of Price: ",std)*

*out=[]*

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

*from sklearn.metrics import accuracy\_score*

*x = rd.iloc[:, :-1].values*

*y = rd.iloc[:, -1].values*

*x\_train, x\_test, y\_train, y\_test= train\_test\_split(x,y,test\_size=0.10, random\_state=10)*

*from sklearn.tree import DecisionTreeClassifier*

*rd\_model= DecisionTreeClassifier()*

*print(x\_train)*

*print(y\_train)*

*import matplotlib.pyplot as plt*

*rd['Price'].plot(figsize=(10,6), grid=True)*

*plt.title('PRICE PREDICTION')*

*plt.show()*

*rd['year'] = pd.DatetimeIndex(rd['Date']).year*

*rd['month'] = pd.DatetimeIndex(rd['Date']).month*

*rd['day'] = pd.DatetimeIndex(rd['Date']).day*

*rd*

*rd.drop('Date',axis=1, inplace=True)*

*rd.isnull().any()*

*import matplotlib.pyplot as plt*

*fig=plt.figure(figsize=(5,5))*

*plt.scatter(rd['day'],rd['Price'],color='blue')*

*plt.xlabel('Day')*

*plt.ylabel('Price')*

*plt.title('PRICE OF NATURAL GAS ON THE BASIS OF DAYS OF A MONTH')*

*plt.bar(rd['month'],rd['Price'],color='green')*

*plt.xlabel('Month')*

*plt.ylabel('Price')*

*plt.title('PRICE OF NATURAL GAS ON THE BASIS OF MONTHS OF A YEAR')*

*import seaborn as sns*

*sns.lineplot(x='year',y='Price',data=rd,color='red')*

*sns.lineplot(x='month',y='Price',data=rd,color='red')*

*sns.lineplot(x='day',y='Price',data=rd,color='red')*

*fig=plt.figure(figsize=(8,4))*

*plt.scatter(rd['year'],rd['Price'],color='purple')*

*plt.xlabel('Month')*

*plt.ylabel('Price')*

*plt.title('PRICE OF NATURAL GAS ON THE BASIS OF MONTHS OF A YEAR')*

*sns.pairplot(rd)*

*plt.show()*

*x=rd.iloc[:,1:4].values*

*y=rd.iloc[:,0:1].values*

*x*

*y*

*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=0)*

*x\_train.shape*

*y\_train.shape*

*from sklearn.tree import DecisionTreeRegressor*

*dtr=DecisionTreeRegressor()*

*import numpy as np*

*x\_train[np.isnan(x\_train)] = np.median(x\_train[~np.isnan(x\_train)])*

*y\_train[np.isnan(y\_train)] = np.median(y\_train[~np.isnan(y\_train)])*

*#x\_pred[np.isnan(x\_pred)] = np.median(x\_pred[~np.isnan(x\_pred)])*

*dtr.fit(x\_train,y\_train)*

*y\_pred=dtr.predict(x\_test)*

*y\_pred*

*from sklearn.metrics import r2\_score*

*accuracy=r2\_score(y\_test,y\_pred)*

*print(accuracy)*

*acc\_precent= accuracy\*100*

*print(acc\_precent,"%")*

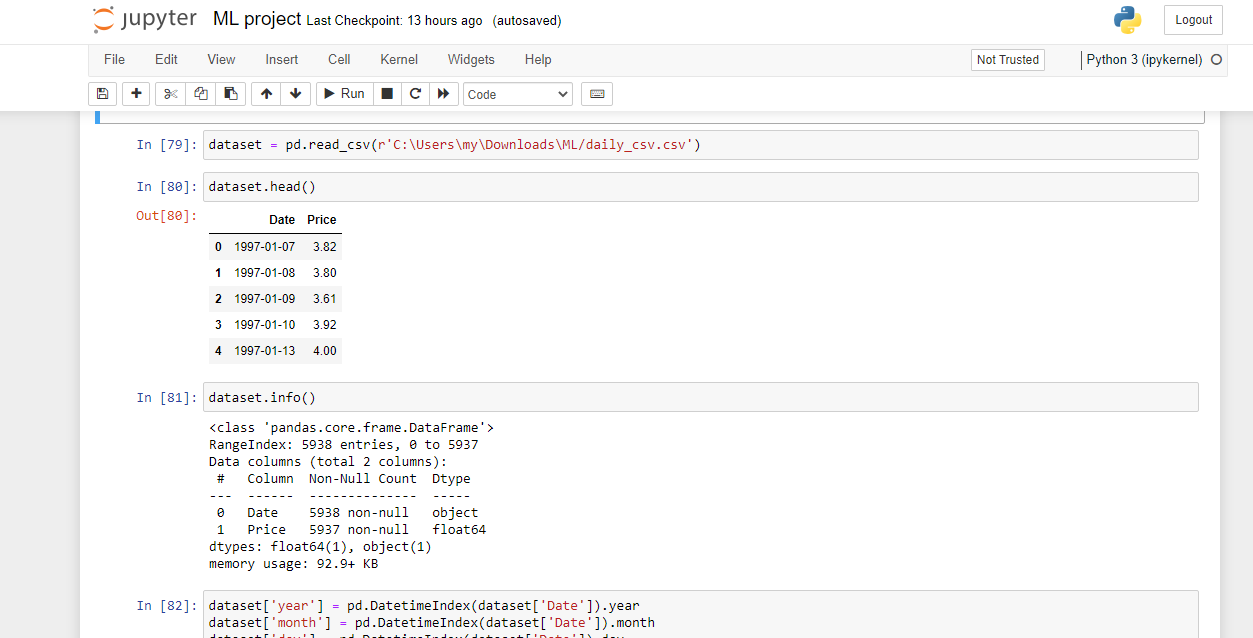
*y\_p=dtr.predict([[year,mon,day]])*

*y\_p*

**3.6 FINAL OUTCOME**

* + Our model was able to use the data properly and achieve a respectable result
  + As the accuracy score of the prediction is 97% we can say that the model works good and has achieved the goal we visualized
  + The accuracy of the model can be still improved if the data is more analysed which is out the scope as it requires much precision in collection of the data
  + The random classifier has an attribute named n-estimators which increases the accuracy when increased, but its tough to use as the model takes more time than usual

**4 RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS**

****

4.1 RESULT ANALYSIS

Fig 4.1 Initialising data set

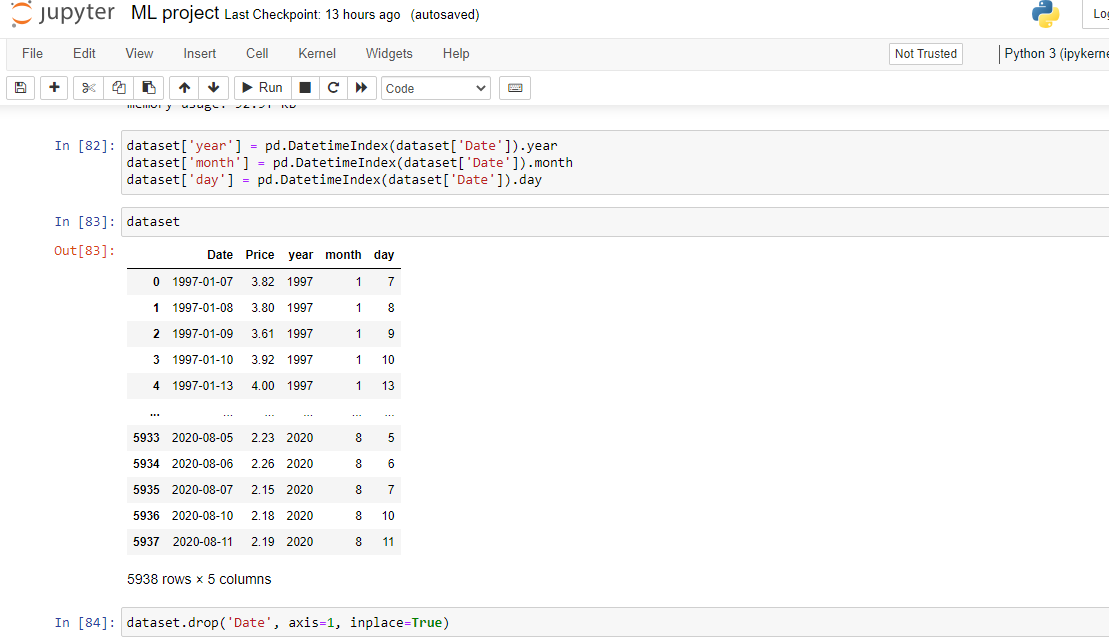


Fig 4.2 printing data values

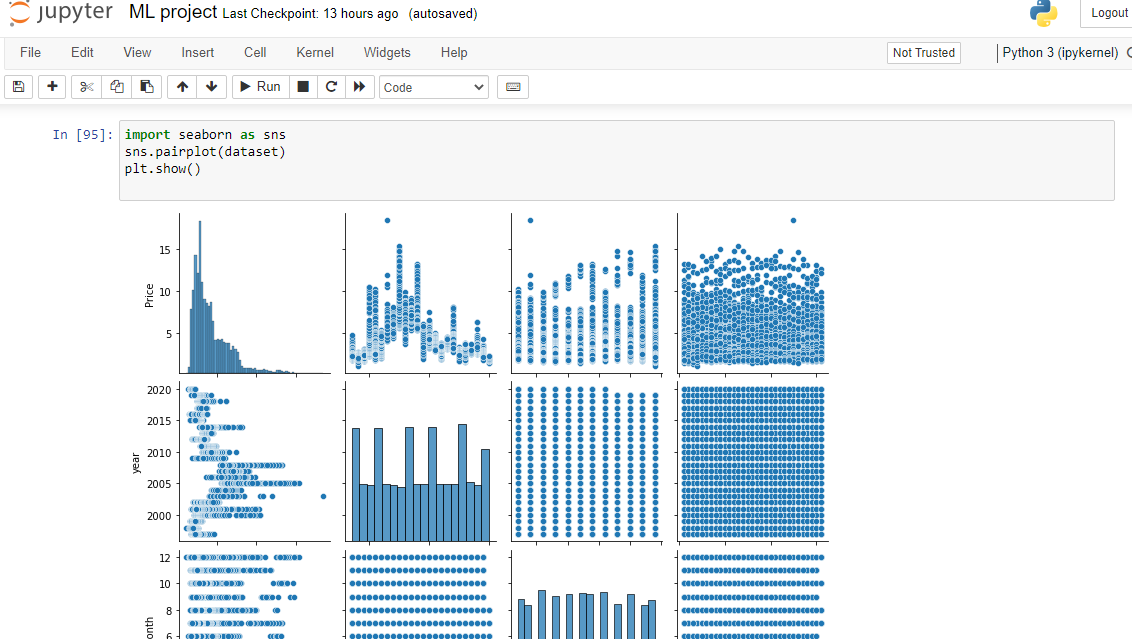
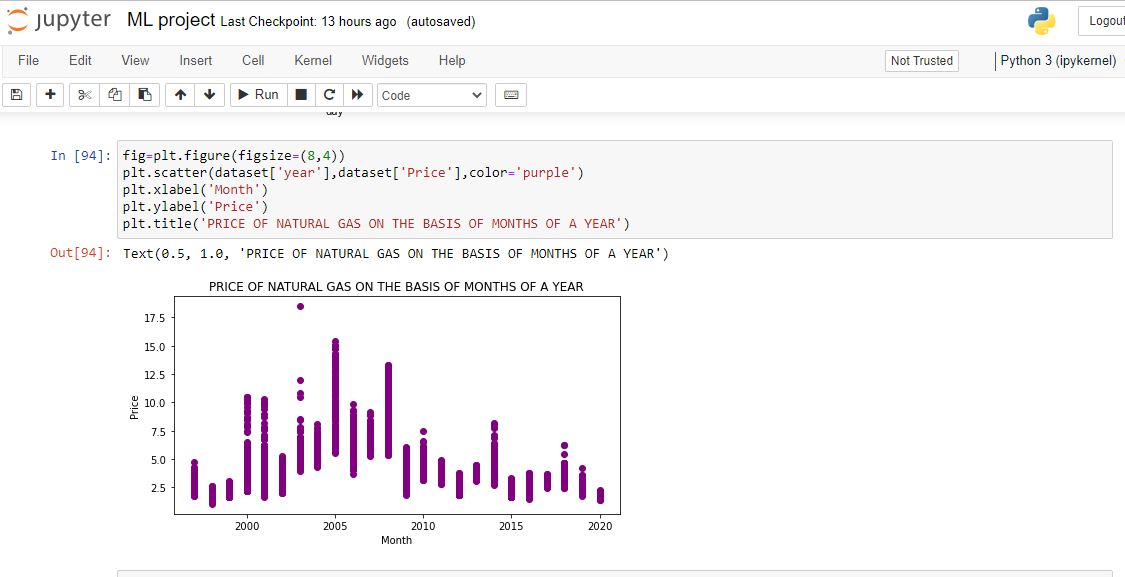
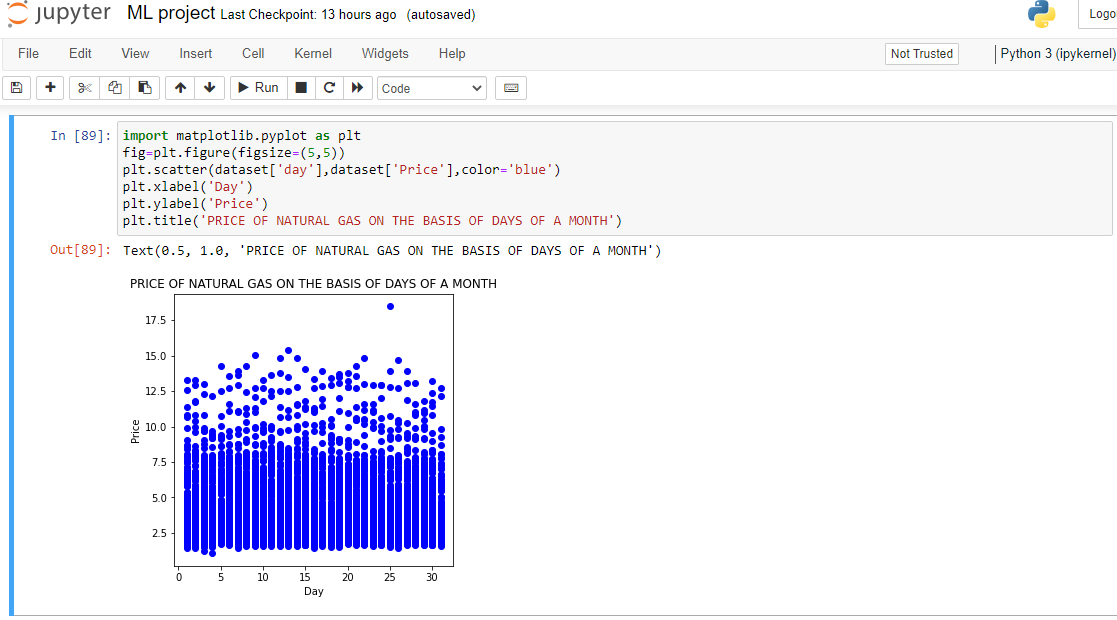
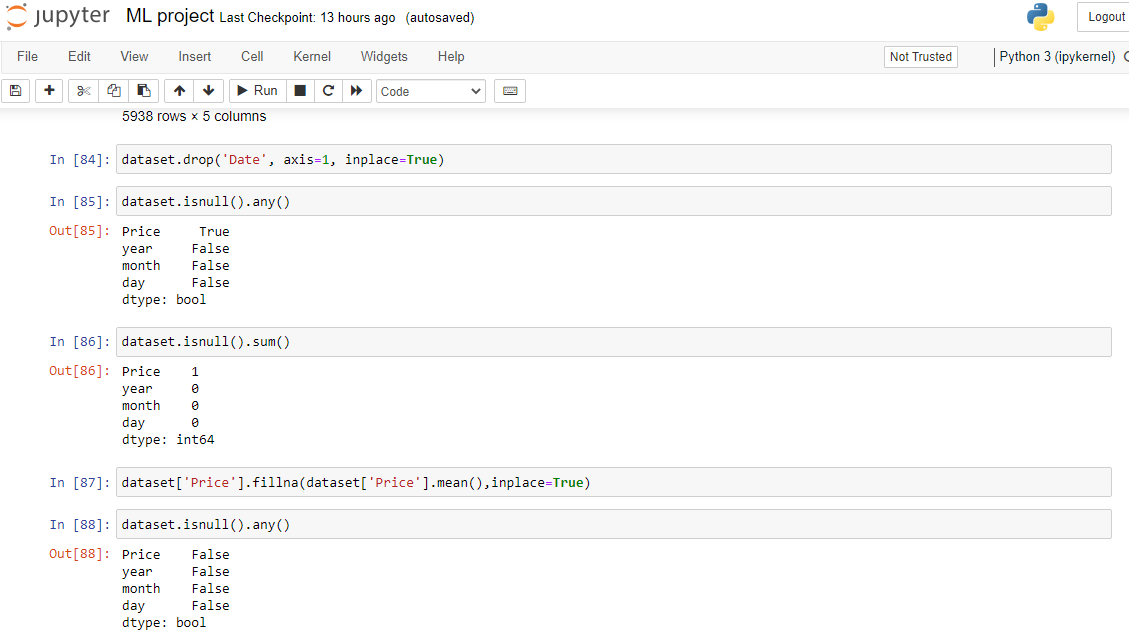


Fig 4.4 Analysing data and developing graphs

Fig 4.3 Importing packages



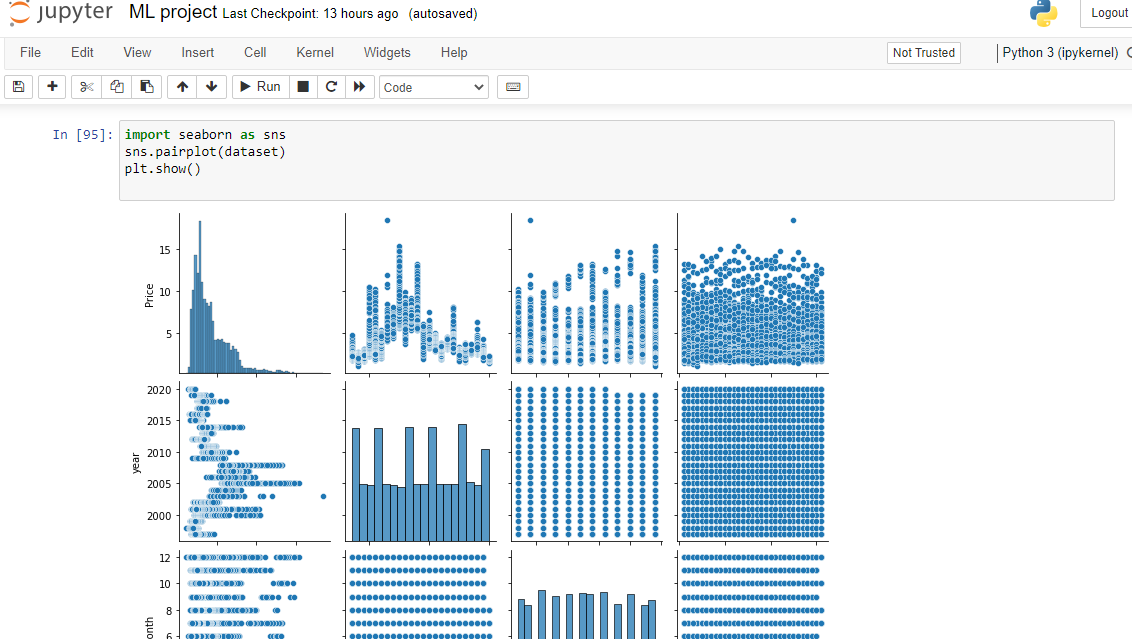


Fig 4.5 Pair plot- comparing the year month and date with price of natural gas.

**4.2 PERFORMANCE ANALYSIS :**

Fig 4.6 Splitting datasets

Fig 4.7 finding accuracy and prediction of the model



RESULT:

* + Our model was able to use the data properly and achieve a respectable result
  + As the accuracy score of the prediction is 97% we can say that the model works good and has achieved the goal we visualized



Fig 4.8 Result

**5 SUMMARY**

* In our project we used predict() function of scikit-learn to predict our model taking X\_test inside to predict() function and called it y\_predict.i.e. y\_predict=predict(X\_test)
* For our sake we checked the length of our y\_predict which gave us 1,45,253 which is equal to the length of X\_test before training and predicting our model.
* These predictions tell us how well our model performed and how well they predicted the final result.
* Finally we can see that the model has an accuracy score of more than 97% which tells us that our model performs well with a training score of 1.
* We further use precision, recall ,classification report, confusion matrix to see how good our model works after the prediction

**6 CONCLUSION**

* A price prediction model that predicts the price of natural gas on that particular day.
* Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output.
* Machine Learning is a technique of training machines to perform the activities a human brain can do, albeit bit faster and better than an average human-being.
* As the accuracy score of the prediction is 97% we can say that the model works good and has achieved the goal we visualised.
* The accuracy of the model can be still improved if the data is more analysed which is out the scope as it requires much precisionin collection of the data.

**7 REFERENCES**

* Indian GDP growth dropped to 5.7% in 2009, well below the previous years’ levels according to the IMF’s World Economic Outlook from July 2010.The IMF expects Indian GDP to grow back to 9.4% and 8.4% in 2010 and 2011 respectively.
* Working paper is based on some of the material in, Anil Jain, Natural Gas in India: Liberalisation and Policy, forthcoming, Oxford University Press, 2011.
* Blackard, Jock A., Dean, Denis J. “Comparative accuracies of artifical neural networks and discriminant analysis in predicting foorest cover types from cartogrpahic variables”. Computers and Electronics in Agriculture (1999). \
* Bache, K. Lichman, M. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science (2013).

8 BIBILIOGRAPHY

**8 BIBILIOGRAPHY**

* + [www.oxfordenergy.org](http://www.oxfordenergy.org)
  + python-course.eu/machine-learning
  + [www.tutorialspoint.com](http://www.tutorialspoint.com)
  + [colab.research.google.com](http://colab.research.google.com)
  + [office.com](http://office.com)