An

Internship Report on

Natural Gas Price Prediction System Using IBM Watson Machine Learning Service

Submitted
in partial fulfillment of the requirements for
the award of the degree of

BACHELOR OF TECHNOLOGY

in

Electronics and Communication Engineering

By
Gottimukkula Harini
19R11A0462

Department of Electronics and Communication Engineering
GEETHANJALI COLLEGE OF ENGINEERING AND TECHNOLOGY
(UGC AUTONOMOUS)
Cheeryal (V), Keesara (M), Medchal Dist, Hyderabad– 501 301
(Affiliated to Jawaharlal Nehru Technological University, Hyderabad, accredited by NAAC and NBA,New Delhi)

2019-2023

GEETHANJALI COLLEGE OF ENGINEERING & TECHNOLOGY

Department of Electronics and Communication Engineering

CERTIFICATE

This is to certify that the internship report titled **Natural Gas Price Prediction System Using IBM Watson Machine Learning Service** being submitted by **Gottimukkula Harini** bearing roll number **19R11A0466** in partial fulfillment of the requirements for the award of the **Degree of Bachelor of Technology** in ***Electronics and Communication Engineering*** is a record of bonafide work carried out through online mechanism.

Prof. K. Somasekhara Rao
Professor &Dean-Student Affairs

Dr. P. Srihari
Professor

Dr. S. Suryanarayana
HoD-ECE

Prof. B. Hari Kumar
Dean, SE&CE

This is to certify that

G. Harini

has completed all courses in the learning path

Natural gas price prediction system using IBM Watson machine learning service

a learning path on smartbridge.skillsnetwork.site

Powered by **IBM Developer Skills Network**

Issued by

The SmartBridge

Natural gas price prediction system using IBM Watson machine learning service consists of the following courses:

- Design Thinking
- DevOps Fundamentals
- Agile Training Module

Issued on:

June 16, 2021

Rav Ahuja
Program Director
Skills Network, IBM

Authenticity of this certificate can be validated by going to:
https://smartbridge.skillsnetwork.site/certificates/067619d1-4c12-4669-a338-8cde288b8c86

Amarender Katkam
Founder & CEO
The SmartBridge

Smart Internz

N·E·A·T
National Educational Alliance for Technology

Externship Certificate

This to certify that

Gottimukkula Harini

has successfully completed the externship program on **Artificial Intelligence powered by IBM** from 23 April 2021 to 31 May 2021 and fulfilled the project work requirements.

Certificate ID: Ext-AI-2021-4830

August 23, 2021

Issued Date

Jayaprakash. Ch

Program Manager

AKNOWLEDGEMENT

I, the student of Geethanjali College of Engineering of technology in the stream Electronics and Communication Engineering would like to express my sincere gratitude to the Principal **Dr. S. UdayaKumar**, and Chairman **Dr. G. R. Ravinder Reddy** for their encouragement to take up this internship.

I also would like to thank **Dr. S. Suryanarayana**, Head of the Department of Electronics and Communication Engineering; **Dr. P. Srihari** (Professor); **Dr. Shoukath Ali** (Associate Professor), whoconstantly provided us guidance to complete the internship successfully. I feel delighted to be under theroof of the most intellectual faculty members. A sincere gratitude towards the college management for providing the best education which ever field we are interested in.

Regards,

Gottimukkula Harini(19R11A0466)

CONTENTS

TOPIC	PAGE NUMBER
Abstract	i
List of figures	ii
Symbols or Abbreviations	ii
CHAPTER 1. INTRODUCTION	
1.1 Organization Background	1
1.2 Training Objective	2
1.3 Domain of Internship	2
CHAPTER 2. TECHNICAL DETAILS	
2.1 Block diagram	3
2.2 Flow chart	4
2.3 Methodology	5
2.4 Skills acquired during Internship	13
CHAPTER 3. RESULTS	14
CHAPTER 4. CONCLUSION	15
CHAPTER 5. FUTURE SCOPE	16
REFERENCES	

ABSTRACT

Artificial Intelligence and Machine Learning is a burning topic in the tech industry. Artificial Intelligence (AI) is impacting our daily lives in many ways. AI is everywhere, from gaming stations to maintaining complex information at work. Computer Engineers and Scientists are working hard to impart intelligent behavior in the machines making them think and respond to real-time situations. Tech giants like Google and Facebook have placed huge bets on Artificial Intelligence and Machine Learning and are already using it in their products. Machine Learning (ML) is a subset of Artificial Intelligence. ML is a science of designing and applying algorithms that are able to learn things from past cases.

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.

Accurate natural gas price forecasting not only provides an important guide for effective implementation of energy policy and planning, but also is extremely significant in economic planning, energy investment, and environmental conservation.

The aim of this project is to build data-driven machine learning models for natural gas price forecasting.

LIST OF FIGURES

Figno:	Figure Name	Page Number
1.	Evolution of Artificial Intelligence	2
2.	Block Diagram of the project	3
3.	Flow chart for the project	4
4.	Project folder with all the necessary files	10
5.	Output/Result	14

SYMBOLS/ ABBREVIATIONS

AI-Artificial Intelligence

ML-Machine Learning

OpenCV- Open Source Computer Vision Library

HTML-Hyper Text Markup Language

CSS-Cascading Style sheets

CHAPTER 1

INTRODUCTION

1.1 Organization Background

Geethanjali College of Engineering and Technology has taken the initiative to train their students in their respective fields of engineering in the best possible way. The college has geared itself to impart and equipped to take on the responsibility of providing the student community with the technical education on par with the latest advances in the respective field with a great ambience. It moulds the students to face any sort of competencies.

In this journey to accomplish and integrate their vision and mission towards perfection, the college has collaborated and established the excellent partnership with the esteemed institution**"Smart Bridge"**.

Smart Bridge is an educational and technological institution whose vision is to bridge the gap between academics and industry. They provide the learning programs on the emerging technologies like Artificial Intelligence, Internet of things, Robotic Process Automation, Cloud Computing and many more. The other wing of the smart bridge **" Smart Internz",** has provided the opportunity for the students to work on the real time projects.

They have always motivated the students to develop the quality of teamwork, leadership, ethics and team responsibility. They have excellent mentors for respective technology. They imparted innovation and gave a great exposure on how the industry level working seems like.

Their **methodology** is as follows:

1. **LEARN**: Choosing the emerging technology and imparting the best possible knowledgeabout the tools in regard to the field.
2. **PRACTICE**: "Practice makes man perfect"; the popular saying goes. While going through the internship; they also provide assignments to assess ourselves how far we areable to cope up with it. They encourage the students to express their doubts.

3. **INTERN**: A team based real time projects are provided to the students after the training

1.2 Training Objective

I have chosen the Artificial Intelligence and Machine Learning. I belong to the Electronics and Communication Engineering, but I also wanted to concentrate towards IT and software field other than my core area. As Artificial Intelligence and Machine Learning was most emerging andtrending technology, I was attracted towards this. The main objective of this internship is to understand the the basics of ML, Deep Learning and AI and Neural Networks understanding and programming .

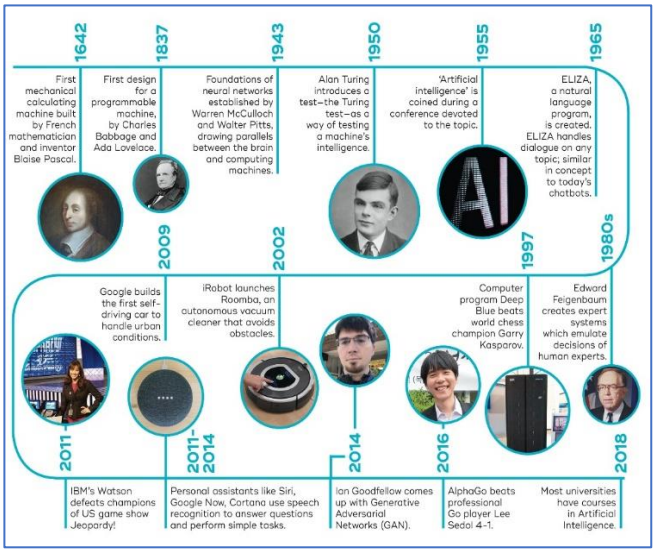
It is definitely a wonder how the world is revolving around AI.
These things have increased the fascination to learn this technology and no wonder I got to know many things from this internship program.

To make place for myself in this constantly progressing world ,it would be beneficial for me. Taking up an internship would be the best thing where I can witness the real- world scenarios.

1.3 Domain of Internship

AI is an introductory course in Artificial Intelligence. The goal is to acquire knowledge on intelligent systems and agents, formalization of knowledge, reasoning with and without uncertainty, machine learning and applications at a basic level. The main research topics in AI include: problem solving, reasoning, planning, natural language understanding, computer vision,automatic programming, machine learning, and so on. Of course, these topics are closely relatedwith each other.

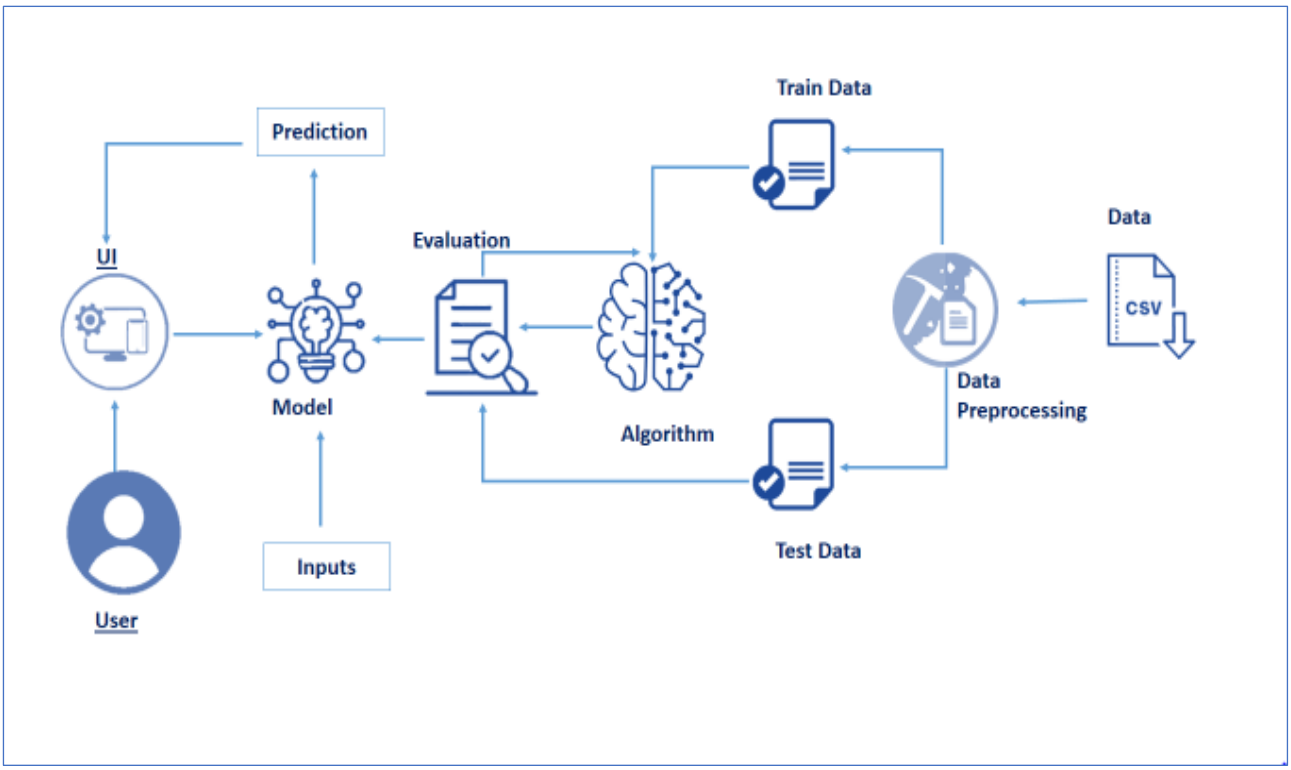
(fig 1)



CHAPTER 2
TECHNICAL DETAILS OF INTERNSHIP DONE

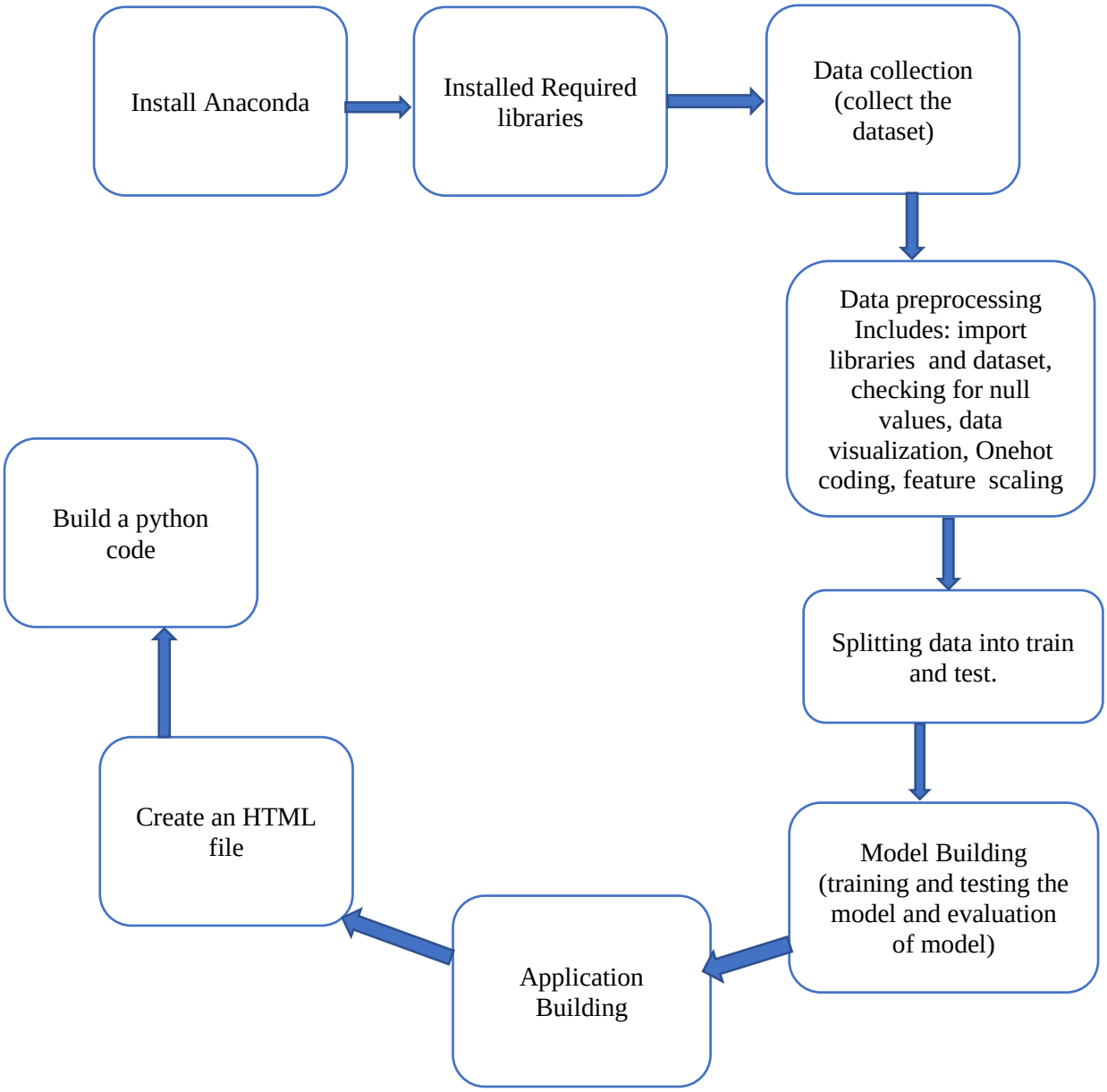
2.1 Block Diagram

(fig 2)



2.2 Flow Chart

(fig 3)



2.3 Methodology followed

To complete this project you should need the following Packages and libraries

- Anaconda
- Jupyter Notebook
- All necessary Python packages

❖ Data Collection

ML depends heavily on data, without data, it is impossible for an "AI" to learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions.

❖ Download Dataset

There exist a great number of factors that affect the natural gas price, e.g., crude oil, heating oil, drilling activity, temperature, natural gas supply, demand, storage, imports, etc. To better forecast the gas price, we should consider as many relevant factors as possible. According to the project, we are going to predict the price of gas based on day, month, and year of historical data.

You can collect datasets from different open sources like kaggle.com, data.gov, UCI machine learning repository, etc. We use a dataset that is created on our own.

❖ Data Preprocessing

Now that we have downloaded the data let's start training the machine with collected , but before training we must ensure the data is in proper structure. so let us preprocess the data.
Data pre-processing is a process of cleaning the raw data i.e. the data is collected in the real world and is converted to a clean data set. In other words, whenever the data is gathered from different sources it is collected in a raw format and this data isn't feasible for analysis.
Therefore, certain steps are executed to convert the data into a small clean data set, this part of the process is called as data pre-processing Follow the following steps to process your Data

- Import the Libraries
- Importing the dataset
- Taking care of Missing Data
- Label encoding
- One Hot Encoding
- Feature Scaling
- Splitting Data into Train and Test

❖ Importing The Libraries

The first step is usually importing the libraries that will be needed in the program.
Import numpy and pandas to your Jupyter Notebook.

Note: It's conventional to refer to alias. When you add the alias name at the end of your import statement, your Jupyter Notebook understands that from this point on every time you type alias name, you are actually referring to the particular library.

```
import numpy as np
import pandas as pd
```

❖ Reading The Dataset

You might have your data in .csv files, .excel files or .tsv files or something else. But the goal is the same in all cases. If you want to analyze that data using pandas, the first step will be to read it into a data structure that's compatible with pandas.

Let's load a .csv data file into pandas. There is a function for it, called read_csv(). We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program).

Pathnames on Windows tend to have backslashes in them. But we want them to mean actual backslashes, not special characters.

```
dataset=pd.read_csv(r'D:\daily_csv.csv')
```

Note: r stands for "raw" and will cause backslashes in the string to be interpreted as actual backslashes rather than special characters.
If the dataset in the same directory as your program, you can directly read it, without giving raw as r.
To check the first five rows of dataset, we have a function call head( ).

```
dataset.head()
```

	Date	Price
0	1997-01-07	3.82
1	1997-01-08	3.80
2	1997-01-09	3.61
3	1997-01-10	3.92
4	1997-01-13	4.00

```
dataset['year'] = pd.DatetimeIndex(dataset['Date']).year
dataset['month'] = pd.DatetimeIndex(dataset['Date']).month
dataset['day'] = pd.DatetimeIndex(dataset['Date']).day
```

Drop Date Column

Often, a DataFrame will contain columns that are not useful to your analysis. Such columns should be dropped from the DataFrame to make it easier for you to focus on the remaining columns. The columns can be removed by specifying label names and corresponding axis, or by specifying index or column names directly.
As we are already extracted the day, month, and year from the Date column, this column can be dropped using the .drop() method.

```
dataset.drop('Date', axis=1, inplace=True)
```

Here axis refers to the dimension of the array, in this case, axis=0 is the dimension that points to the row-wise operation and axis=1 is the dimension that points to the column side.
When inplace = True, the data is modified in place, which means it will return nothing and the dataframe is now updated.
When inplace = False, which is the default, then the operation is performed and it returns a copy of the object.

❖ Handling Missing Values

Sometimes you may find some data are missing in the dataset. We need to be equipped to handle the problem when we come across them.
One of the most common ideas to handle the problem is to take a mean of all the values of the same column and have it to replace the missing data.
We will be using isnull().any() method to see which column has missing values.

```
dataset.isnull().any()
```

```
Price      True
year      False
month     False
day       False
dtype: bool
```

```
dataset['Price'].fillna(dataset['Price'].mean(),inplace=True)
```

❖ Data Visualization

Data visualization is where a given dataset is presented in a graphical format. It helps the detection of patterns, trends, and correlations that might go undetected in text-based data. Understanding your data and the relationship presents within it is just as important as any algorithm used to train your machine learning model. Machine learning models will perform poorly on data that wasn't visualized and understood properly.

To visualize the dataset we need libraries called Matplotlib and Seaborn. The Matplotlib library is a Python 2D plotting library that allows you to generate plots, scatter plots, histograms, bar charts, etc.

❖ Splitting The Dataset Into Dependent And Independent Variable.

we need to split our dataset into the matrix of independent variables and the vector or dependent variable. Mathematically, Vector is defined as a matrix that has just one column.

To read the columns, we will use iloc of pandas (used to fix the indexes for selection) which takes two parameters — [row selection, column selection].
Let's split our dataset into independent and dependent variables.

```
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)
```

❖ Split The Dataset Into Train Set And Test Set

When you are working on a model and you want to train it, you obviously have a dataset. But after training, we have to test the model on some test dataset. For this, you will need a dataset that is different from the training set you used earlier. But it might not always be possible to have so much data during the development phase. In such cases, the solution is to split the dataset into two sets, one for training and the other for testing.

To help us with this task, the Scikit library provides a tool, called the Model Selection library. There is a class in the library which is, 'train_test_split.' Using this we can easily split the dataset into the training and the testing datasets in various proportions.
The train-test split is a technique for evaluating the performance of a machine learning algorithm.

- Train Dataset: Used to fit the machine learning model.
- Test Dataset: Used to evaluate the fit machine learning model.

```
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)
```

❖ Model Building

There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms that you can choose according to the objective that you might have it may be Classification algorithms or Regression algorithms. As the dataset which we are using is a Regression dataset so you can use the following algorithms

1. Linear Regression.
2. Logistic Regression.
3. Random Forest Regression / Classification.
4. Decision Tree Regression / Classification.

You will need to train the datasets to run smoothly and see an incremental improvement in the prediction rate.

❖ Train The Model With Descision Tree Algorithm

Now that we have our data preprocessed, lets train the model. WE make use of decision tree for training
The general motive of using Decision Tree is to create a training model which can use to predict the class or value of
To know more about the Decision Tree algorithm, please refer to the link provided

```
#import decision tree regressor
from sklearn.tree import DecisionTreeRegressor
dtr=DecisionTreeRegressor()
#fitting the model or training the model
dtr.fit(x_train,y_train)
```

❖ Test The Model

Once the model is trained, it's ready to make predictions. We can use the predict method on the model and pass x_test as a parameter to get the output as y_pred.
Notice that the prediction output is an array of real numbers corresponding to the input array.

```
y_pred=dtr.predict(x_test)
y_pred
```

❖ Model evaluation

- Finally, we need to check to see how well our model is performing on the test data. We evaluate our model by finding the R2 Score and produced by the model.
- We can use the r2_score method on the model and pass y_test and y_pred as a parameter to get the accuracy.

- In regression models, R2 corresponds to the squared correlation between the observed outcome values and the predicted values by the model. The higher the R-squared, the better the model.

```
#import r2_score evaluation metric from sklearn
from sklearn.metrics import r2_score
dtr_accuracy=r2_score(y_test,y_pred)
dtr_accuracy
```

```
import pickle
pickle.dump(dtr,open('gas.pkl','wb'))
```

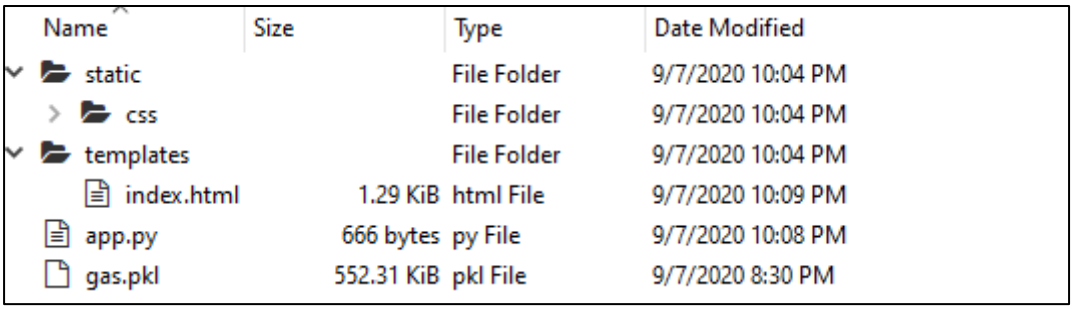
❖ Application Building

we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

To build this you should know the basics of "HTML, CSS, Bootstrap, flask framework and python" Create a project folder which should contain

- A python file called app.py.
- Model file (gas.pkl).
- Templates folder which contains the index.HTML file.
- Static folder which contains CSS folder which contains styles.css

(fig 4)



❖ Build HTML Page

This is the basic HTML page for our Project. H1 tag is used to give heading to the project. As I have mentioned there are 3 input or 3 independent variables we have created 3 text input fields in the html page. A button is used to send these values to the model files this functionality will be written in the python file app.py. The model predicts the value and is displayed on the {{prediction_text}} filed.

```
<body>
 <div class="login">
    <h1>Natural Gas Price Prediction</h1>
     <!-- Main Input For Receiving Query to our ML -->
    <form action="{{ url_for('y_predict')}}"method="post">
        <input type="text" name="year" placeholder="year" required="required" />
        <input type="text" name="month" placeholder="month" required="required" />
        <input type="text" name="day" placeholder="day" required="required" />

        <button type="submit" class="btn btn-primary btn-block btn-large">Predict</button>
    </form>
   <br>
   <br>
   {{ prediction_text }}
 </div>
</body>
</html>
```

❖ Build Python Code

We will be using python for server-side scripting. Let's see step by step process for writing backend code.
Create app.py python code file in the project folder using spyder

```
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle

app = Flask(__name__)
model = pickle.load(open('gas.pkl', 'rb'))
```

Routing to HTML Page:

```
@app.route('/')
def home():
    return render_template('index.html')
```

Showcasing prediction on UI:

```
@app.route('/y_predict',methods=['POST'])
def y_predict():
    '''
    For rendering results on HTML GUI
    '''
    x_test = [[int(x) for x in request.form.values()]]

    prediction = model.predict(x_test)
    print(prediction)
    pred=prediction[[0]]
    #output = np.round(pred[0],3)
    #output = str(output)+'$ Dollors'

    return render_template('index.html', prediction_text='Gas Price is {} Dollors'.format(pred))
```

Main Function
This is used to run the application on the localhost.

```
if __name__ == "__main__":
    app.run(debug=True)
```

❖ Run The App

- Open the anaconda prompt from the start menu.
- Navigate to the folder where your app.py resides.
- Now type “python app.py” command.
- It will show the local host where your app is running on http://127.0.0.1.5000/
- Copy that localhost URL and open that URL in the browser. It does navigate to your web page.
- Enter the values, click on the predict button, and see the result/prediction on the web page.

```
(base) C:\Users\DELL>E:

(base) E:\>cd "Guided Projects"

(base) E:\Guided Projects>cd "Natural Gas Price Prediction Using Machine Learning"

(base) E:\Guided Projects\Natural Gas Price Prediction Using Machine Learning>cd flask-1

(base) E:\Guided Projects\Natural Gas Price Prediction Using Machine Learning\flask-1>python app.py
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 833-051-520
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

❖ Train The Model On IBM

In this milestone, you will learn how to build a Machine Learning Model and deploy it on the IBM Cloud.

2.4 Skills acquired during Internship

By the end of this project:

- You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.
- You will be able to know how to pre-process/clean the data using different data preprocessing techniques.
- You will able to analyze or get insights into data through visualization.
- Applying different algorithms according to the dataset and based on visualization.
- You will able to know how to find the accuracy of the model.
- You will be able to know how to build a web application using the Flask framework.

CHAPTER 3
RESULTS

(fig 5)

If we give the year, month and day then it will predict the gas price at that particular instant.

CHAPTER 4
CONCLUSION

It has always been a difficult task to predict the exact daily price of natural gas price. Many factors such as political events, general economic conditions, and traders' expectations may have an influence on it. But here, based on the past and present traits, we were able to achieve up to 97% accuracy in predicting the price of any given date. Albeit, its impossible to predict unexpected scenarios such as acts of warfare or terrorism. But, the benefits of having reliable information of what the price of natural gas could be at any given time is paramount, it could make or break economies. And in this case, as this project points out data-driven machine learning models deserve all the attention it could ever garner and even more.

Applications:

Natural gas accounts for 1/4 of the global demand and roughly 1/3 of the US energy demand. After oil, Natural gas is the most dominant sort of energy. So, being about to improve natural gas demand prediction is extremely valuable. The accurate prediction of energy price is critical to the energy market orientation, and it can provide a reference for policymakers and market participants. In practice, energy prices are affected by external factors, and their accurate prediction is challenging.Being able to forecast natural gas price benefits various stakeholders and has become a very valuable tool for all market participants in competitive natural gas markets. Machine learning algorithms have gradually become popular tools for natural gas price forecasting.

CHAPTER 5
FUTURE SCOPE

The project has been built using 2 models of prediction namely the Decision Tree method and Random Forest method with the accuracy score of over 97% on both the models (97.4% on Decision Tree and 97.74% on Random Forest Method). By doing some further research and learning the accuracy can be uplifted upto 100% which would be an ideal prediction real- time application which would be much more helpful in the trading sector.

REFERENCES

- https://smartinternz.com/Student/guided_project_info/4885#
- https://www.osti.gov/servlets/purl/908487
- https://www.mdpi.com/1996-1073/12/9/1680/pdf