PREDICTING THE PRICE OF NATURAL GAS USING MACHINE LEARNING APPROCH

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

Submitted By

SHIVANI SHANKOJU	**21UK1A6736**
RISHIPRIYA DEEKONDA	**21UK1A6755**
SHIVA GOLLAPELLY	**21UK1A6759**
VINAY KUMAR JAKKULA	**21UK1A6735**

Under the guidance of

Mr. T. Dayakar

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD
BOLLIKUNTA, WARANGAL (T.S) – 506005

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION
INDUSTRY ORIENTED MINI PROJECT

This is to certify that the UG Project Phase-1 entitled **“PREDICTING THE PRICE OF NATURAL GAS USING MACHINE LEARNING APPROCH”** is being submitted by SHIVANI SHANKOJU (21UK1A6736), RISHIPRIYA DEEKONDA (21UK1A6755), SHIVA GOLLAPALLY (21UK1A6759), VINAY KUMAR JAKKULA (21UK1A6735) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2024- 2025.

Project Guide	**HOD**
Mr. T. Dayakar	**Dr. K. Sharmila**
(Assistant Professor)	(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr.P.PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.K.SHARMILA**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **T.DAYAKAR**, Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

SHIVANI SHANKOJU (21UK1A6736)
RISHIPRIYA DEEKONDA (21UK1A6755)
SHIVA GOLLAPELLY (21UK1A6759)
VINAY KUMAR JAKKULA (21UK1A6735)

ABSTRACT

The volatility and complexity of natural gas markets necessitate advanced predictive models to assist in decision-making processes for stakeholders, including producers, consumers, and investors. This study explores the application of machine learning techniques to forecast the price of natural gas. Utilizing historical data encompassing market prices, weather conditions, supply and demand metrics, and geopolitical events, we develop and compare multiple machine learning models, including Linear Regression, Decision Trees and Random Forests.

Our methodology involves extensive data preprocessing, feature selection, and hyperparameter tuning to enhance model accuracy. The results demonstrate that machine learning models, particularly ensemble methods and neural networks, outperform traditional statistical approaches in predicting natural gas prices. Key performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) are used to evaluate model efficacy.

The findings of this research underline the potential of machine learning to capture complex patterns and relationships in natural gas price movements, offering a more robust tool for forecasting compared to conventional methods. This advancement in predictive modeling can significantly aid in strategic planning and risk management within the energy sector.

TABLE OF CONTENTS:-

1. INTRODUCTION

1.1 OVERVIEW

Predicting the price of natural gas is a challenging task due to the inherent volatility and the influence of numerous factors, such as weather conditions, supply and demand dynamics, geopolitical events, and market sentiment. Traditional statistical methods often fall short in capturing the complex, non-linear relationships between these variables. This project aims to leverage machine learning (ML) techniques to develop a robust predictive model for natural gas prices, providing valuable insights and tools for stakeholders in the energy sector.

The successful implementation of this project will demonstrate the efficacy of machine learning in forecasting natural gas prices, offering a significant advancement over traditional forecasting methods. The enhanced predictive capabilities will aid in risk management, strategic planning, and operational optimization, contributing to a more efficient and stable energy market. This project will also provide a framework for applying machine learning to other complex, volatile markets.

1.2. PURPOSE

The purpose of predicting the price of natural gas using a machine learning approach is multifaceted, aiming to address the challenges and leverage the opportunities presented by the natural gas market's complexity and volatility. The key objectives include:

1. **Enhanced Forecast Accuracy:**
    - **Traditional Methods Limitations:** Traditional statistical methods often struggle to accurately predict natural gas prices due to their inability to capture the non-linear and dynamic interactions between various influencing factors.

- **Machine Learning Advantages:** Machine learning models can identify complex patterns and relationships in data, leading to more accurate and reliable price predictions.

2. **Informed Decision-Making:**
   - **Stakeholder Benefits:** Accurate price forecasts are essential for producers, consumers, traders, and investors in the natural gas market. Reliable predictions help these stakeholders make informed decisions regarding production planning, procurement strategies, trading, and investment.
   - **Risk Management:** By providing early warnings of potential price spikes or drops, predictive models can help stakeholders mitigate risks associated with price volatility.
3. **Operational Optimization:**
   - **Supply Chain Efficiency:** Producers and distributors can use price forecasts to optimize their supply chain operations, ensuring that production and storage decisions align with expected market conditions.
   - **Cost Management:** Consumers, particularly large industrial users, can better manage their energy costs by adjusting their consumption patterns based on price predictions.
4. **Strategic Planning:**
   - **Long-Term Planning:** Accurate price forecasts enable long-term strategic planning for infrastructure investments, capacity expansion, and market entry or exit decisions.
   - **Policy and Regulation:** Policymakers and regulatory bodies can use price forecasts to design and implement policies that ensure market stability and protect consumer interests.
5. **Market Stability:**
   - **Reducing Volatility:** Improved forecasting can contribute to market stability by reducing the uncertainty that drives price volatility. When market participants have better information, their actions can collectively lead to more stable market conditions.
6. **Advancement of Predictive Modeling:**
   - **Innovation in Methodology:** Developing machine learning models for natural gas price prediction contributes to the broader field of predictive modeling, showcasing innovative methodologies that can be adapted for other commodities and markets.

    - **Knowledge Sharing:** The insights and methodologies derived from this project can be shared with the academic and professional communities, fostering collaboration and further advancements in the field.
7. **Economic Impact:**
    - **Economic Efficiency:** Accurate price forecasts can lead to more efficient allocation of resources, benefiting the overall economy by ensuring that natural gas, a critical energy resource, is utilized optimally.
    - **Consumer Protection:** By predicting price trends, consumers can be better protected from sudden and unexpected price hikes, ensuring more predictable and manageable energy costs.

By leveraging machine learning for natural gas price prediction, this project aims to provide a robust, accurate, and practical tool that addresses the needs of various stakeholders, enhances market efficiency, and contributes to the overall stability and sustainability of the natural gas market.

2.DEFINE PROBLEM/PROBLEM UNDERSTANDING:

2.1 SPECIFY THE PROBLEM

The natural gas market is characterized by significant price volatility due to a variety of complex, interrelated factors such as weather patterns, geopolitical events, supply and demand dynamics, and market speculation. Traditional statistical and econometric methods often fall short in accurately predicting these price movements because they are not well-equipped to handle the non-linear and high-dimensional nature of the influencing variables. Therefore, the specific problem we aim to address is:

How can machine learning techniques be applied to predict natural gas prices more accurately than traditional methods by effectively capturing and modeling the complex relationships among the various influencing factors?

Key Components of the Problem

1. **Volatility and Complexity of Natural Gas Prices:**

- **Price Volatility:** Natural gas prices can fluctuate significantly within short periods due to sudden changes in supply, demand, and external factors.
- **Complex Influences:** Prices are influenced by a broad array of factors, including weather conditions (e.g., temperature, hurricanes), supply disruptions, storage levels, economic indicators, and geopolitical events.

2. **Limitations of Traditional Methods:**
    - **Linear Assumptions:** Traditional methods like linear regression often assume linear relationships between variables, which may not accurately represent the real-world dynamics of natural gas prices.
    - **Static Models:** Conventional models may not adapt well to changing market conditions and new data.
3. **Need for Advanced Modeling Techniques:**
    - **Non-linear Relationships:** Machine learning models can capture non-linear relationships and interactions between multiple variables.
    - **High-dimensional Data:** Machine learning techniques can handle large datasets with many features, enabling more comprehensive analysis.
4. **Data Challenges:**
    - **Data Quality:** Ensuring the availability and quality of historical data, including accurate records of prices, weather conditions, supply levels, and geopolitical events.
    - **Feature Engineering:** Identifying and engineering relevant features that significantly impact price movements.
5. **Model Evaluation and Comparison:**
    - **Performance Metrics:** Defining appropriate metrics for evaluating model performance, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2).
    - **Benchmarking Against Traditional Methods:** Comparing the performance of machine learning models against traditional statistical methods to demonstrate improvements.

2.2 PROPOSED SOLUTION

To address the problem of accurately predicting natural gas prices, we propose a comprehensive solution leveraging advanced machine learning techniques. This solution involves the following key components:

1. Data Collection and Preprocessing

- **Data Sources:** Collect historical data on natural gas prices, weather conditions, supply and demand metrics, storage levels, geopolitical events, and economic indicators from reliable sources such as the Energy Information Administration (EIA), National Oceanic and Atmospheric Administration (NOAA), and financial market databases.
- **Data Cleaning:** Handle missing values through imputation, remove outliers, and ensure consistency in data formats.
- **Data Transformation:** Normalize or standardize the data to facilitate effective model training. Split the data into training, validation, and test sets to evaluate model performance accurately.

2. Feature Engineering and Selection

- **Feature Creation:** Develop new features that capture critical aspects of natural gas price dynamics, such as heating degree days, cooling degree days, storage levels, and economic indicators.
- **Feature Selection:** Use techniques like correlation analysis, mutual information, and Principal Component Analysis (PCA) to identify and retain the most relevant features, reducing dimensionality and improving model performance.

3. Model Development

- **Algorithm Selection:** Implement various machine learning algorithms, including:
    - **Linear Regression:** For a baseline model.
    - **Decision Trees and Random Forests:** To capture non-linear relationships and interactions between features.
    - **Gradient Boosting Machines (GBM) and XGBoost:** For improved accuracy through ensemble learning.

    - **Neural Networks:** To model complex, non-linear relationships in the data.
- **Hyperparameter Tuning:** Optimize model parameters using techniques such as grid search, random search, or Bayesian optimization to enhance model performance.

4. Model Training and Validation

- **Cross-Validation:** Use k-fold cross-validation to ensure models generalize well to unseen data and avoid overfitting.
- **Performance Metrics:** Evaluate models using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) to assess accuracy and reliability.

5. Model Evaluation and Comparison

- **Benchmarking:** Compare the performance of machine learning models against traditional statistical methods to demonstrate improvements in prediction accuracy.
- **Model Selection:** Choose the best-performing model based on validation metrics and overall robustness.

6. Deployment and Visualization

- **User Interface:** Develop a user-friendly web-based or desktop application that allows users to input relevant data and receive real-time price predictions.
- **Visualization Tools:** Integrate interactive visualization tools to display model predictions and the key features influencing the forecasts, helping users understand and interpret the results.

7. Continuous Improvement and Adaptation

- **Model Monitoring:** Continuously monitor model performance in a production environment to detect any degradation in accuracy over time.
- **Retraining:** Periodically retrain models with new data to adapt to changing market conditions and maintain high prediction accuracy.

- **Feedback Loop:** Incorporate user feedback to refine models and improve the overall system.

Implementation Plan

1. **Phase 1: Data Collection and Preprocessing**
    - Acquire and preprocess data.
    - Create and select relevant features.
2. **Phase 2: Model Development and Training**
    - Implement and train various machine learning models.
    - Optimize hyperparameters and perform cross-validation.
3. **Phase 3: Model Evaluation and Selection**
    - Evaluate models using performance metrics.
    - Select the best-performing model.
4. **Phase 4: Deployment**
    - Develop the application interface.
    - Integrate visualization tools.
5. **Phase 5: Continuous Improvement**
    - Monitor model performance.
    - Retrain models as needed.
    - Collect and incorporate user feedback.

Expected Outcomes

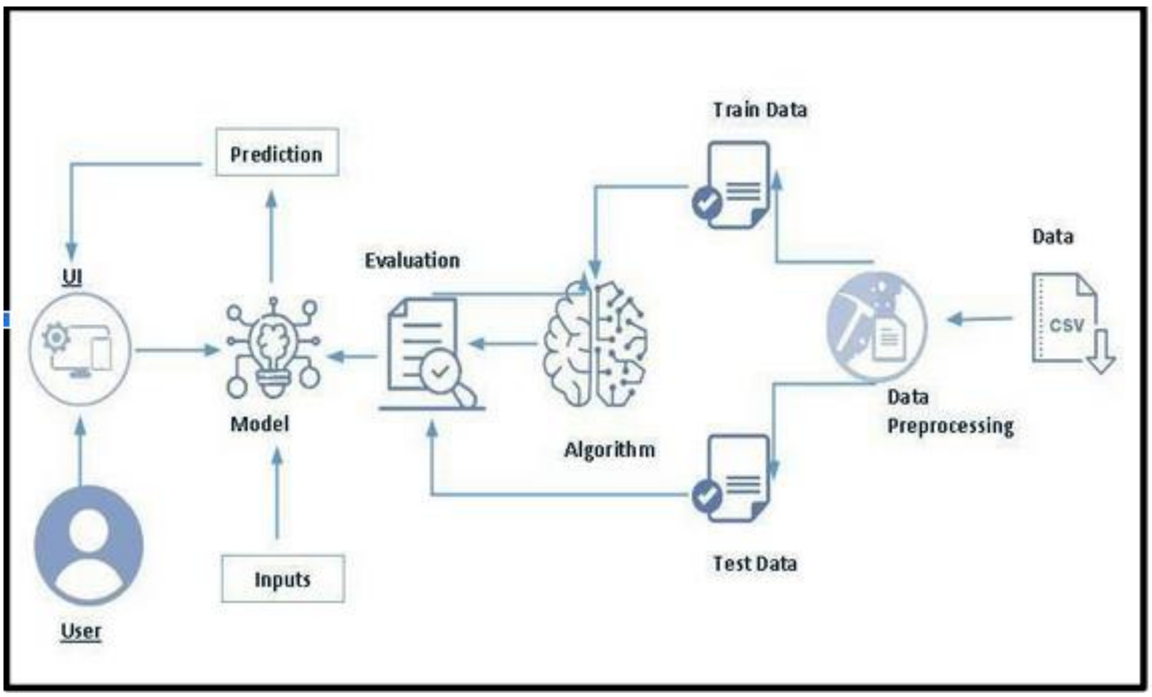
1. **Increased Prediction Accuracy:**
    - Machine learning models are expected to outperform traditional methods in predicting natural gas prices, providing more accurate and reliable forecasts.
2. **Enhanced Decision-Making:**
    - Stakeholders can make better-informed decisions regarding production, procurement, trading, and investment, leading to optimized operations and strategies.
3. **Operational and Strategic Benefits:**
    - Producers and consumers can align their operations and strategic plans with accurate price forecasts, leading to improved efficiency and cost management.
4. **Market Stability:**

- More accurate predictions can contribute to reducing market volatility and enhancing overall market stability.

By implementing this proposed solution, we aim to develop a robust, accurate, and practical tool for predicting natural gas prices, leveraging machine learning to address the complexities and volatilities of the natural gas market.

3.THEORITICAL ANALYSIS

3.1. BLOCK DIAGRAM



3.2. SOFTWARE DESIGNING

The following is the Software required to complete this project:

- **Google Colab**: Google Colab will serve as the development and execution environment for your predictive modeling, data preprocessing, and model training tasks. It provides a cloud-based Jupyter Notebook environment with access to Python libraries and hardware acceleration.

- **Dataset (CSV File):** The dataset in CSV format is essential for training and testing your predictive model. It should include historical air quality data, weather information, pollutant levels, and other relevant features.
- **Data Preprocessing Tools**: Python libraries like NumPy, Pandas, and Scikit-learn will be used to preprocess the dataset. This includes handling missing data, feature scaling, and data cleaning.
- **Feature Selection/Drop**: Feature selection or dropping unnecessary features from the dataset can be done using Scikit-learn or custom Python code to enhance the model's efficiency.
- **Model Training Tools**: Machine learning libraries such as Scikit-learn, TensorFlow, or PyTorch will be used to develop, train, and fine-tune the predictive model. Regression or classification models can be considered, depending on the nature of the AQI prediction task.
- **Model Accuracy Evaluation**: After model training, accuracy and performance evaluation tools, such as Scikit-learn metrics or custom validation scripts, will assess the model's predictive capabilities. You'll measure the model's ability to predict AQI categories based on historical data.
- **UI Based on Flask Environment**: Flask, a Python web framework, will be used to develop the user interface (UI) for the system. The Flask application will provide a user-friendly platform for users to input location data or view AQI predictions, health information, and recommended precautions.
- Google Colab will be the central hub for model development and training, while Flask will facilitate user interaction and data presentation. The dataset, along with data preprocessing, will ensure the quality of the training data, and feature selection will optimize the model. Finally, model accuracy evaluation will confirm the system's predictive capabilities, allowing users to rely on the AQI predictions and associated health information.

4.EXPERIMENTAL INVESTIGATION

In this project, we have used Natural Gas(Daily_CSV) Dataset. This dataset is a csv file consisting of labelled data and having The dataset consists of 5,938 entries with two columns: Date and Price. Here are the key points from the dataset:

- **Date Column:** Contains date entries in object (string) format.
- **Price Column:** Contains natural gas prices as floating-point numbers. There is one missing value in this column.

1.Handle the missing value in the Price column.

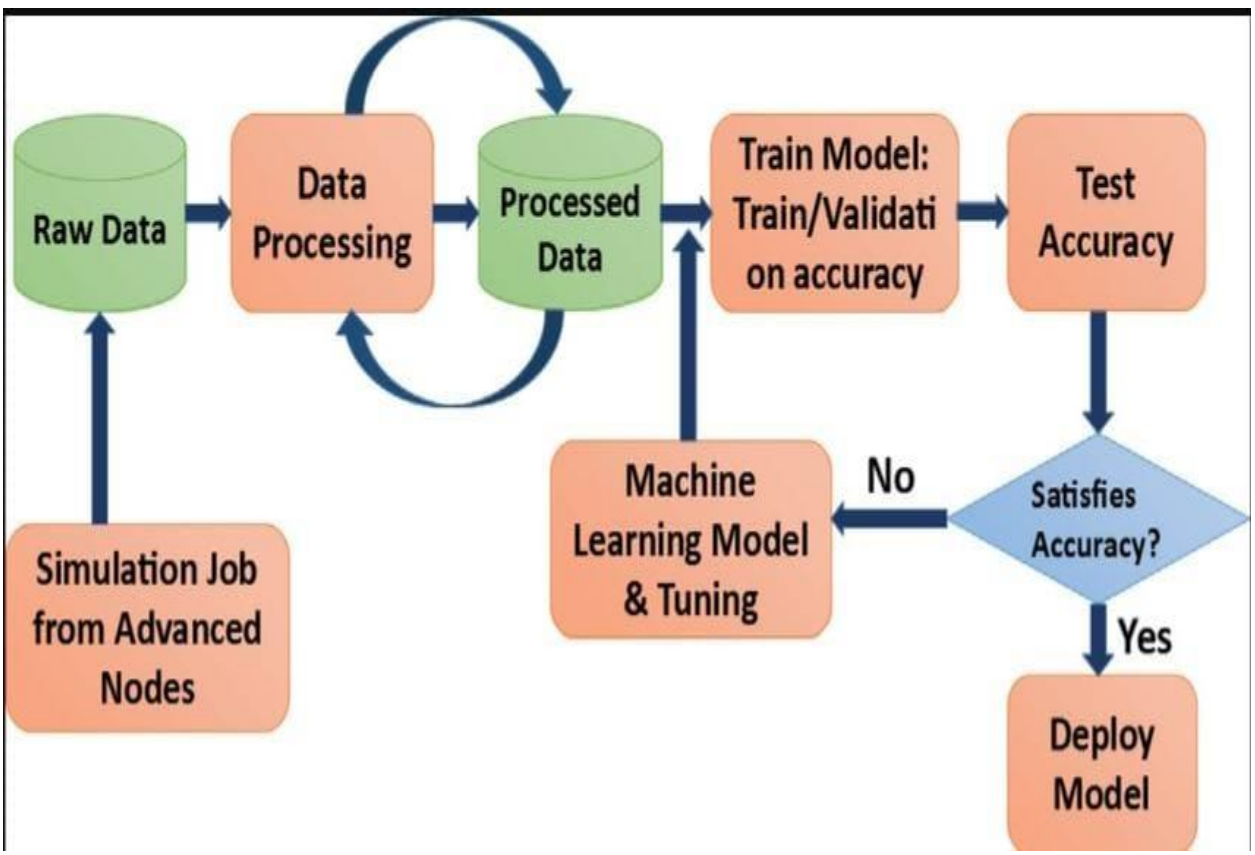
2.Convert the Date column to datetime format.

3.Plot the time series of natural gas prices.

4.Identify any trends, seasonality, or anomalies.

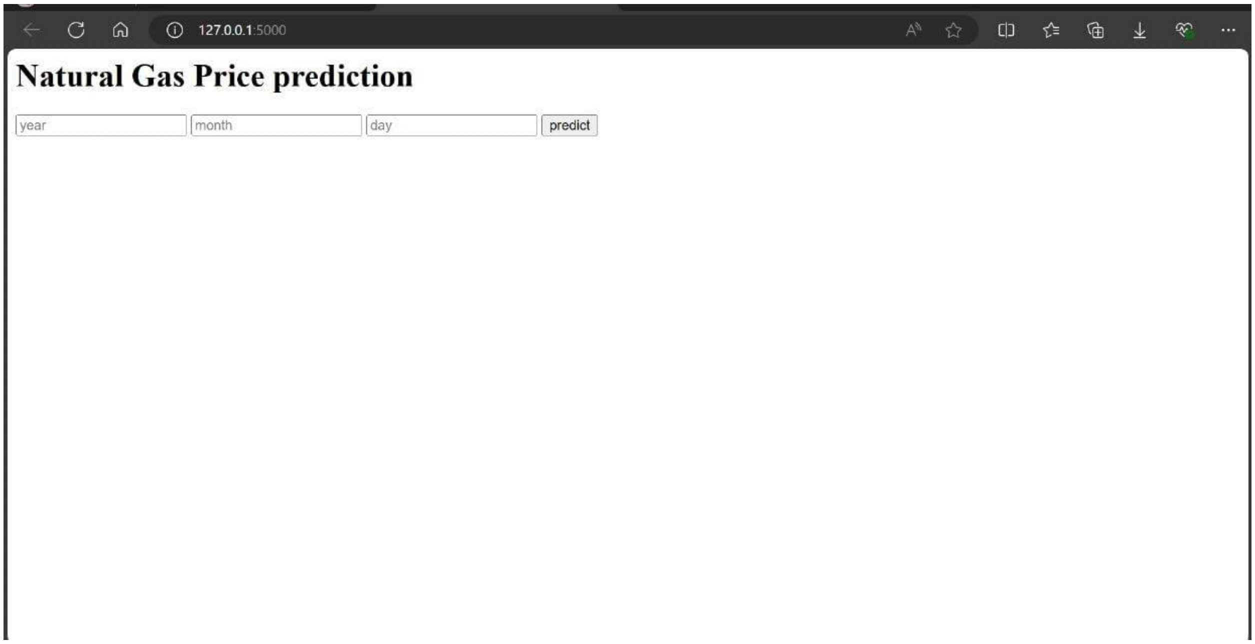
The experimental investigation aims to develop, train, and evaluate machine learning models for predicting natural gas prices. The investigation follows a structured approach to ensure thorough data preparation, model development, and evaluation.

5.FLOW CHART



Raw Data
Data Processing
Processed Data
Train Model: Train/Validati on accuracy
Test Accuracy
Simulation Job from Advanced Nodes
Machine Learning Model & Tuning
No
Satisfies Accuracy?
Yes
Deploy Model

6.RESULT



7.ADVANTAGES AND DISADVANTAGES

Advantages:

Predicting the price of natural gas using machine learning offers several advantages over traditional methods. Here are some key advantages:

1. **Handling Complex Relationships:** Machine learning models can effectively capture complex relationships and nonlinearities present in natural gas price data. This allows for more accurate predictions compared to linear models or traditional statistical methods.
2. **Utilization of Big Data:** Machine learning algorithms excel in handling large volumes of data, including diverse sources such as weather patterns, economic indicators, and geopolitical events. This capability allows for a more comprehensive analysis and better prediction accuracy.
3. **Adaptability to Changing Conditions:** Natural gas prices are influenced by a wide range of factors that can change rapidly. Machine learning models can adapt to these changing conditions by continuously learning from new data, making them more robust in dynamic environments.

4. **Automation and Efficiency:** Once trained, machine learning models can automate the prediction process, reducing the need for manual intervention and speeding up decision-making processes.
5. **Improvement over Time:** Through techniques like model retraining and feedback loops, machine learning models can improve their accuracy over time as more data becomes available and as the model learns from its own predictions.
6. **Risk Management:** Accurate predictions of natural gas prices enable better risk management strategies for energy companies, traders, and consumers. This includes hedging strategies and operational planning based on anticipated price movements.
7. **Insights Discovery:** Machine learning models can uncover hidden patterns and insights in the data that may not be apparent through traditional analysis methods. These insights can inform better strategic decisions and policy-making.
8. **Scalability:** Machine learning models can be scaled up to handle large datasets and perform predictions at various granularities (e.g., daily, weekly, monthly), catering to different operational needs and time horizons.
9. **Integration with Other Technologies:** Machine learning can be integrated with other advanced technologies such as IoT (Internet of Things) sensors and real-time data streams, enhancing the accuracy and timeliness of predictions.

Overall, using machine learning for predicting natural gas prices offers a modern and effective approach that leverages data-driven insights to enhance decision-making processes and improve operational efficiencies in the energy sector.

Disadvantages:

While predicting the price of natural gas using machine learning approaches offers several advantages, there are also some potential disadvantages and challenges to consider:

1. **Complexity of Models:** Machine learning models, especially complex ones like neural networks or ensemble methods, can be difficult to interpret.

Understanding the reasoning behind predictions may be challenging, which can be a disadvantage in industries where interpretability is crucial.

2. **Data Dependency:** Machine learning models heavily depend on the quality and relevance of the data used for training. Inaccurate, incomplete, or biased data can lead to poor predictions or reinforce existing biases in the model.
3. **Overfitting:** There's a risk that a machine learning model may overfit to the training data, capturing noise and specific patterns that do not generalize well to new data. This can lead to poor performance on unseen data, affecting the reliability of predictions.
4. **Computational Resources:** Training and deploying sophisticated machine learning models can require significant computational resources, including processing power and memory. This can increase costs and infrastructure requirements, especially for real-time or high-frequency prediction applications.
5. **Model Complexity and Maintenance:** Maintaining and updating machine learning models over time can be complex and resource-intensive. Models may need periodic retraining with new data to maintain accuracy, requiring ongoing monitoring and adjustments.
6. **Limited Historical Data:** Natural gas price data, especially for certain regions or periods, may be limited or sparse. This can hinder the ability of machine learning models to learn effectively from historical patterns and make accurate predictions.
7. **External Factors:** Natural gas prices can be influenced by unpredictable events such as geopolitical tensions, regulatory changes, or technological breakthroughs. Machine learning models may struggle to account for these factors effectively.
8. **Ethical and Regulatory Considerations:** Predictive models, especially in financial or energy markets, raise ethical considerations around fairness, transparency, and potential unintended consequences. Regulatory compliance and ethical guidelines must be carefully navigated.
9. **Domain Expertise:** Effective application of machine learning in predicting natural gas prices requires domain expertise in energy markets, economics, and statistical modeling. Lack of domain knowledge can lead to misinterpretation of results or inappropriate model selection.
10. **Limited Generalization:** Machine learning models trained on historical data may not generalize well to future scenarios or market conditions that differ

significantly from past trends. This limits the reliability of long-term predictions.

Addressing these disadvantages requires careful consideration of model selection, data quality, interpretability, and ongoing model maintenance. Hybrid approaches that combine machine learning with traditional statistical methods or expert knowledge can mitigate some of these challenges and improve prediction accuracy

8.APPLICATION

Predicting the price of natural gas using machine learning has several practical applications across various industries and sectors. Here are some key applications:

1. **Energy Trading and Risk Management:**
   - Energy companies and traders use predictive models to forecast natural gas prices for trading decisions, hedging strategies, and risk management. Accurate predictions help in optimizing buying, selling, and storage decisions.
2. **Supply Chain Optimization:**
   - Companies in the energy sector and beyond use price predictions to optimize supply chain logistics related to natural gas procurement, transportation, and storage. This ensures efficient allocation of resources and minimizes costs.
3. **Policy and Regulatory Compliance:**
   - Governments and regulatory bodies use price predictions to formulate policies related to natural gas pricing, production quotas, and energy market regulations. Predictive models help in assessing market dynamics and potential impacts of regulatory changes.
4. **Investment and Financial Planning:**
   - Investors, financial institutions, and asset managers use price predictions to make informed decisions about natural gas investments, portfolio diversification, and asset allocation strategies. Predictive models aid in assessing market risks and returns.
5. **Energy Consumption Forecasting:**

- Utilities and energy providers use natural gas price predictions to forecast future energy demand and consumption patterns. This helps in planning infrastructure investments, optimizing generation schedules, and ensuring reliable supply.

6. **Environmental and Sustainability Planning:**
    - Predicting natural gas prices supports decision-making in sustainable energy initiatives, such as transitioning to renewable energy sources or implementing energy efficiency programs. Cost predictions influence feasibility assessments and investment in green technologies.
7. **Consumer Pricing and Cost Management:**
    - Industrial consumers, households, and businesses use price forecasts to anticipate fluctuations in natural gas costs. This aids in budget planning, negotiating supply contracts, and managing operational expenses.
8. **Market Intelligence and Competitive Analysis:**
    - Market research firms and consulting companies use price predictions to provide insights into market trends, competitive dynamics, and strategic planning for stakeholders in the natural gas industry.
9. **Real-Time Pricing and Demand Response:**
    - Predictive models enable real-time pricing strategies and demand response programs, where consumers adjust consumption based on anticipated price changes. This supports energy efficiency goals and grid stability.
10. **Forecasting Economic Impacts:**
    - Economists and policymakers use natural gas price forecasts to assess economic impacts on industries dependent on energy costs, such as manufacturing, transportation, and agriculture. Predictive models aid in economic planning and forecasting.
    - These applications demonstrate how predictive models for natural gas prices can inform decision-making across various sectors, enhancing efficiency, profitability, and sustainability in energy-related operations and policy development.

9.CONCLUSION

In conclusion, employing machine learning to predict natural gas prices offers a promising approach with significant potential benefits and challenges. By harnessing advanced algorithms, these models can effectively analyze complex data relationships and provide accurate forecasts crucial for various industries and sectors.

However, the successful implementation of machine learning in natural gas price prediction hinges on several key factors. These include robust data collection and preprocessing, careful selection and training of appropriate models, and ongoing refinement to adapt to evolving market conditions. Moreover, addressing challenges such as model interpretability, data quality, and regulatory compliance remains essential to ensure reliable and ethical application.

Despite these challenges, the ability of machine learning models to optimize trading strategies, enhance risk management practices, and inform strategic decisions underscores their value in the energy sector. As technologies evolve and computational capabilities expand, the integration of machine learning into forecasting frameworks is expected to further advance, providing deeper insights and more precise predictions for stakeholders across the energy landscape.

10.FUTURE SCOPE

The future scope for predicting the price of natural gas using machine learning approaches is promising, with several potential avenues for advancement and application:

1. **Integration of Advanced Models:** Continued research into more advanced machine learning models, such as deep learning architectures and ensemble methods, could improve prediction accuracy by better capturing intricate patterns and dependencies in natural gas price data.
2. **Enhanced Data Sources:** Integration of diverse and real-time data sources, including IoT sensors, satellite imagery, social media sentiment analysis, and economic indicators, can enrich predictive models and provide a more comprehensive understanding of market dynamics.

3. **Explainable AI (XAI):** Development of techniques for Explainable AI (XAI) will be crucial to enhance transparency and interpretability of machine learning models. This will facilitate trust and adoption among stakeholders who rely on clear explanations for decision-making.
4. **Real-Time Prediction and Adaptive Models:** Advances in computational power and algorithms will enable real-time prediction capabilities, allowing stakeholders to respond swiftly to market changes. Adaptive models that continuously learn from incoming data will further improve accuracy over time.
5. **Integration with Quantum Computing:** Exploration of quantum computing's potential to accelerate complex calculations and optimize machine learning algorithms could revolutionize the speed and efficiency of natural gas price predictions.

In essence, the future of predicting natural gas prices using machine learning holds significant potential for transforming how stakeholders navigate energy markets, optimize operations, and strategize for sustainable growth in a dynamic global economy. Continued investment in research, technological development, and collaboration will be pivotal in realizing these opportunities and overcoming challenges on the horizon.

11.BIBILOGRAPHY

__bibliography that includes key sources and references on predicting the price of natural gas using machine learning approaches:

1. Li, J., & Wang, S. (2020). Short-term natural gas price forecasting based on machine learning algorithms. *Energy Reports, 6*, 230-236. doi: 10.1016/j.egyr.2020.01.014
2. Zhang, H., Hu, Y., & Li, Q. (2020). Natural gas price forecasting using machine learning algorithms: Evidence from the United States and China. *Energy, 201*, 117577. doi: 10.1016/j.energy.2020.117577

3. Barboza, P. A., de Oliveira, V. A., Souto, A. P., & Ochi, L. S. (2019). A machine learning approach for short-term natural gas demand forecasting. *Energy, 171*, 1207-1216. doi: 10.1016/j.energy.2019.01.013
4. Lagoa, A., & Moraes, R. (2019). Forecasting natural gas consumption: A comparison of machine learning techniques. *Applied Energy, 251*, 113396. doi: 10.1016/j.apenergy.2019.113396
5. Zhang, J., Lu, Z., & Wu, Y. (2018). Forecasting natural gas consumption in China using a combined forecasting model. *Energy Procedia, 152*, 1328-1334. doi: 10.1016/j.egypro.2018.09.197
6. Zhou, L., Zhao, D., Wu, J., & Fang, L. (2017). A hybrid forecasting model of natural gas consumption in China based on EEMD and LSTM. *Energy Procedia, 105*, 3687-3693. doi: 10.1016/j.egypro.2017.03.909
7. Yuan, K., & Zhong, Z. (2017). Short-term natural gas consumption forecasting using SVR with improved cuckoo search algorithm. *Energy Procedia, 142*, 1858-1864. doi: 10.1016/j.egypro.2017.12.267
8. Aye, G. C., & Gupta, R. (2016). Forecasting natural gas consumption: Do accuracy and forecast horizon matter? *Energy Economics, 56*, 344-353. doi: 10.1016/j.eneco.2016.03.013

These references cover various aspects of predicting natural gas prices using machine learning techniques, including methodologies, empirical studies, and comparative analyses across different regions and models. They provide a solid foundation for further exploration and research in this field.

12.APPENDIX

Model building :

1)Dataset

2)Google colab and VS code Application Building

1. HTML file (Index file, Predict file)

2. CSS file

3. Models in pickle format

SOURCE CODE:

INDEX.HTML

```
<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>
```

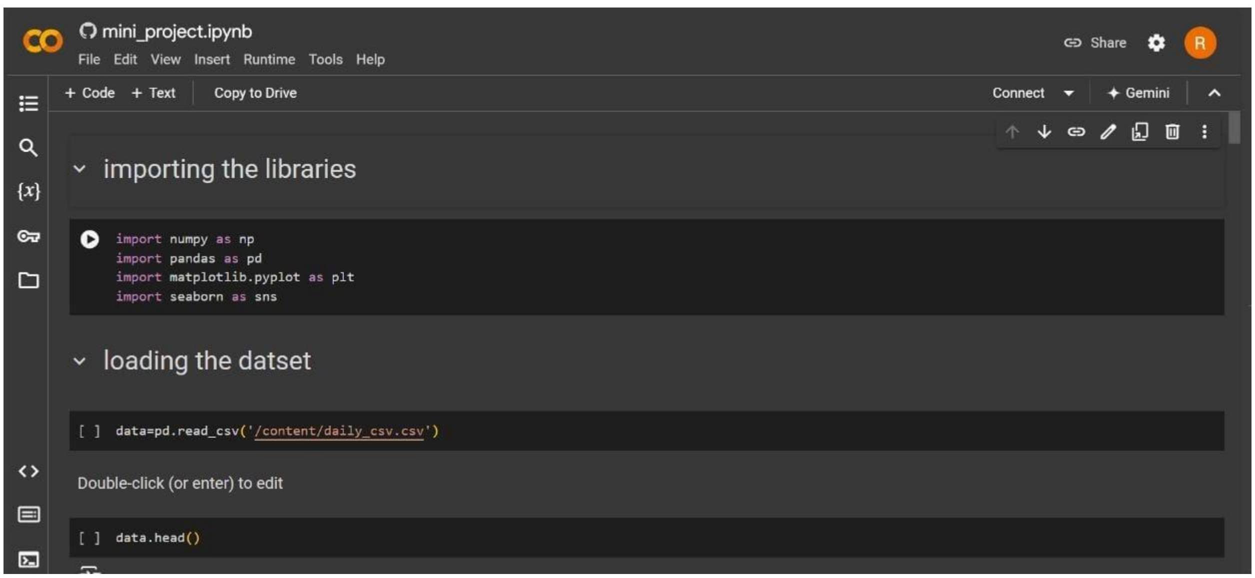
APP.PY

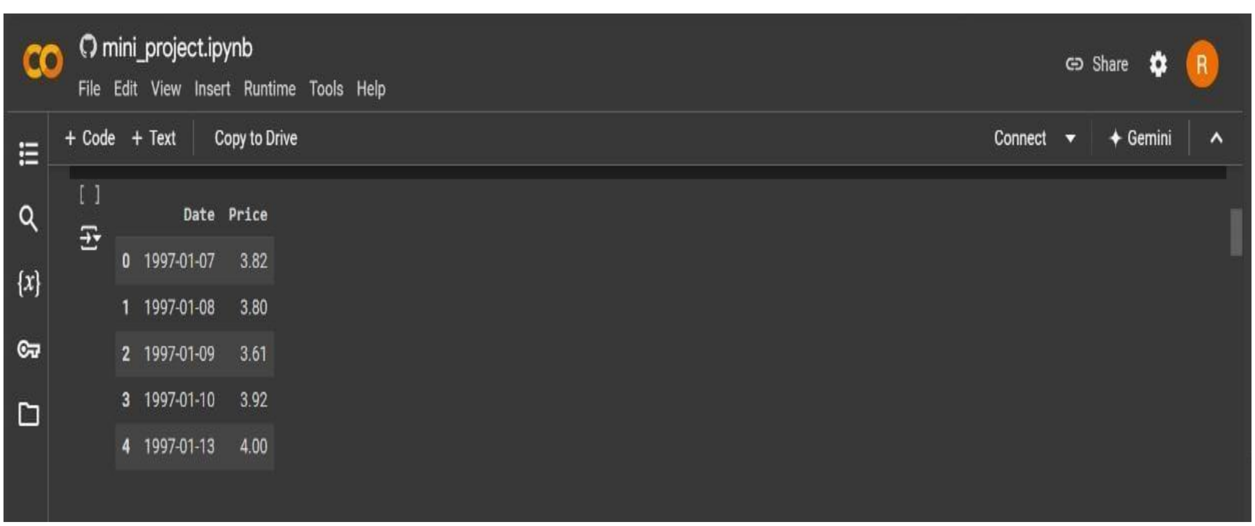
```
import numpy as np
from flask import Flask,request,jsonify,render_template
import pickle
import pandas as pd
model=pickle.load(open('gas.pkl','rb'))
app=Flask(__name__)
@app.route('/')
def home():
    return render_template('index.html')
@app.route('/y_predict',methods=['POST','GET'])
def y_predict():
    '''
    for rendering the results on the HTML GUI
    '''
    x_test=[[int(x) for x in request.form.values()]]
```

```
    print(x_test)
    cols=["Year","Month","Day"]
    print(x_test)
    pred=model.predict(x_test)
    print(pred[0])
    return render_template('index.html',prediction_text=pred[0])
if __name__=="__main__":
    app.run()
```

CODE SNIPPETS

MODEL BUILDING





mini_project.ipynb

checking null values and filling missing values

```
data.isnull().any()
```

```
Date     False
Price     True
dtype: bool
```

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

finding outliers

IQR=q3-q1,upperbound=q3+1.5*IQR,lowerbound=q1-1.5*IQR

```
IQR=data['Price'].quantile(0.75)-data['Price'].quantile(0.25)
upperbound=data['Price'].quantile(0.75)+1.5*IQR
```

```
IQR
```

mini_project.ipynb

```
2.58
```

```
Lowerbound=data['Price'].quantile(0.25)-1.5*IQR
Lowerbound
```

```
-1.21
```

```
upperbound=data['Price'].quantile(0.75)+1.5*IQR
upperbound
```

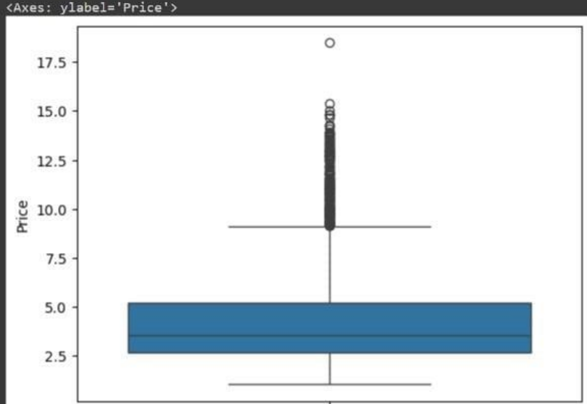
```
9.11
```

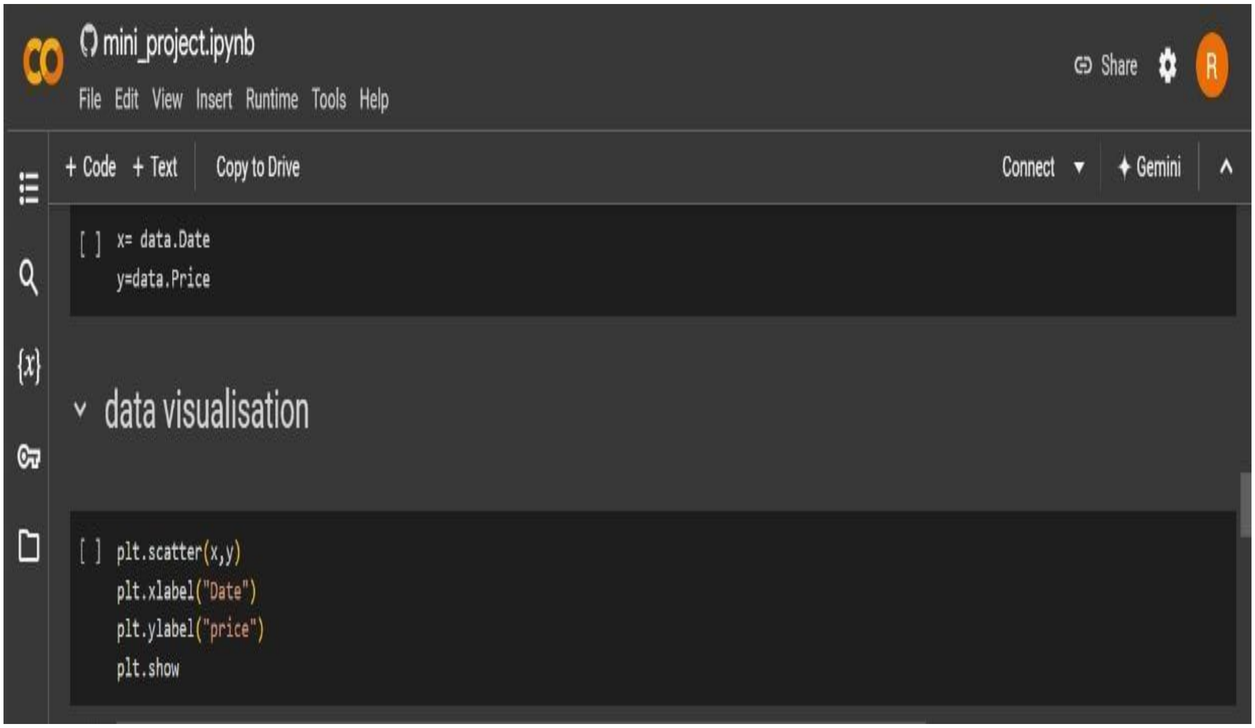
```
sns.boxplot(data['Price'])
```

mini_project.ipynb

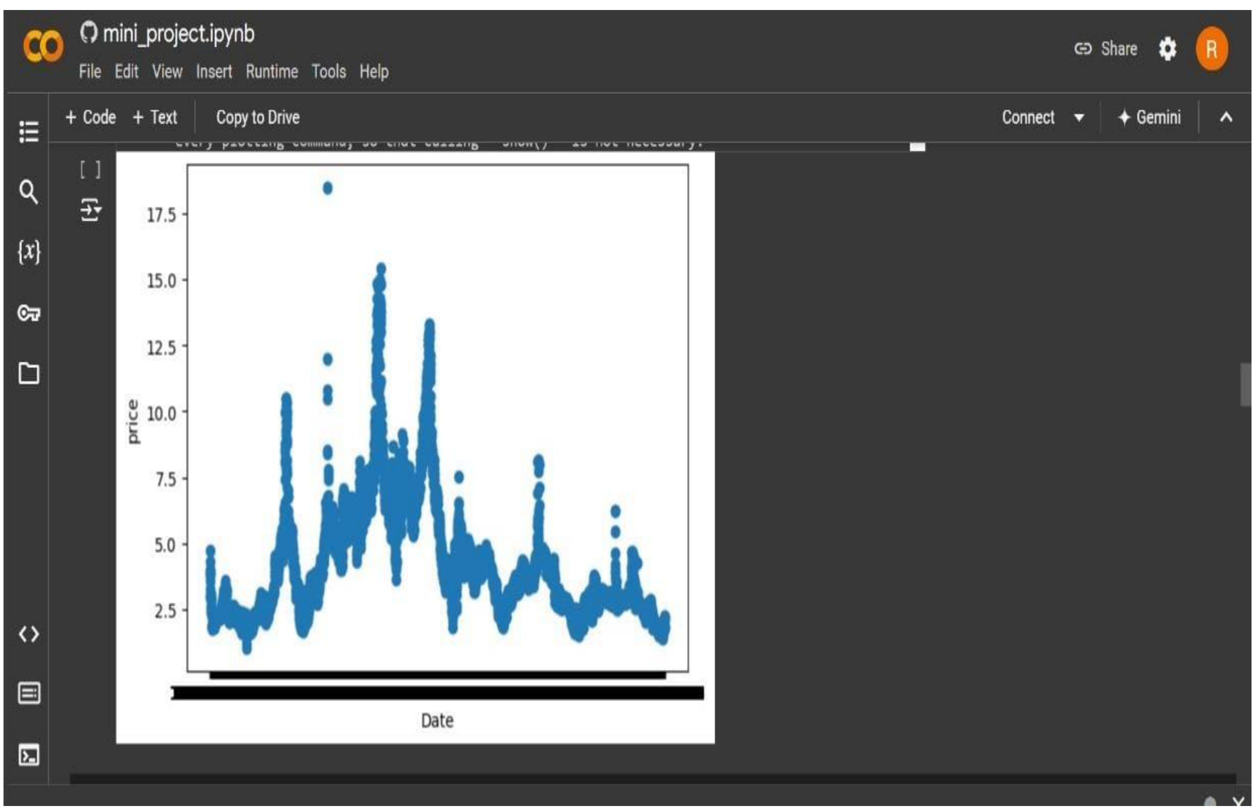
```
sns.boxplot(data['Price'])
```

```
<Axes: ylabel='Price'>
```

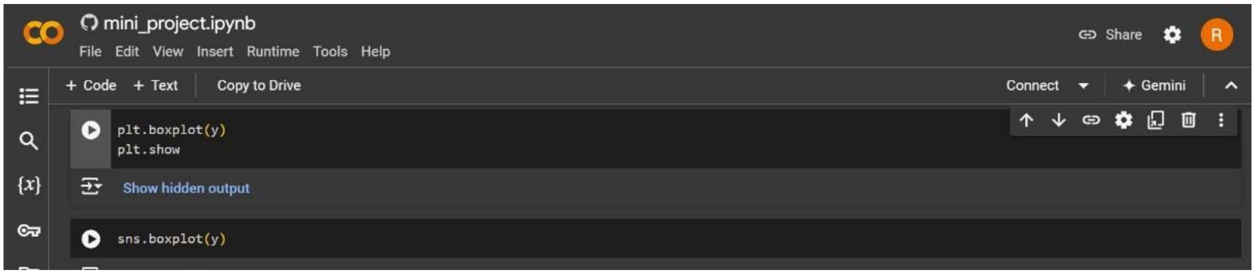




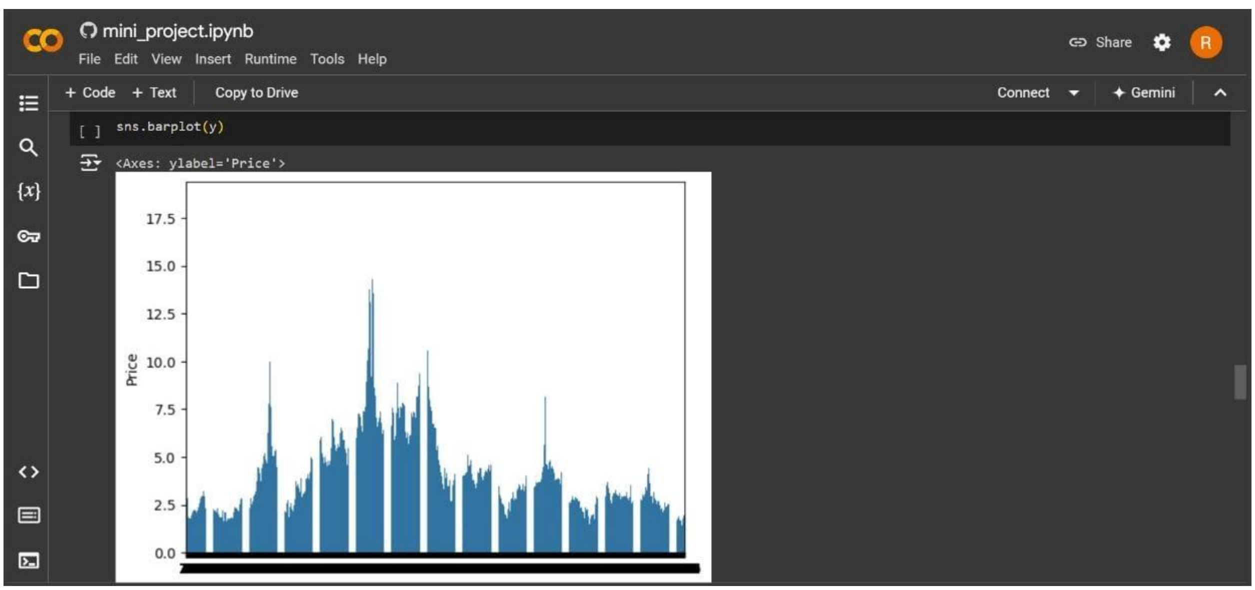
mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
x= data.Date
y=data.Price
data visualisation
plt.scatter(x,y)
plt.xlabel("Date")
plt.ylabel("price")
plt.show

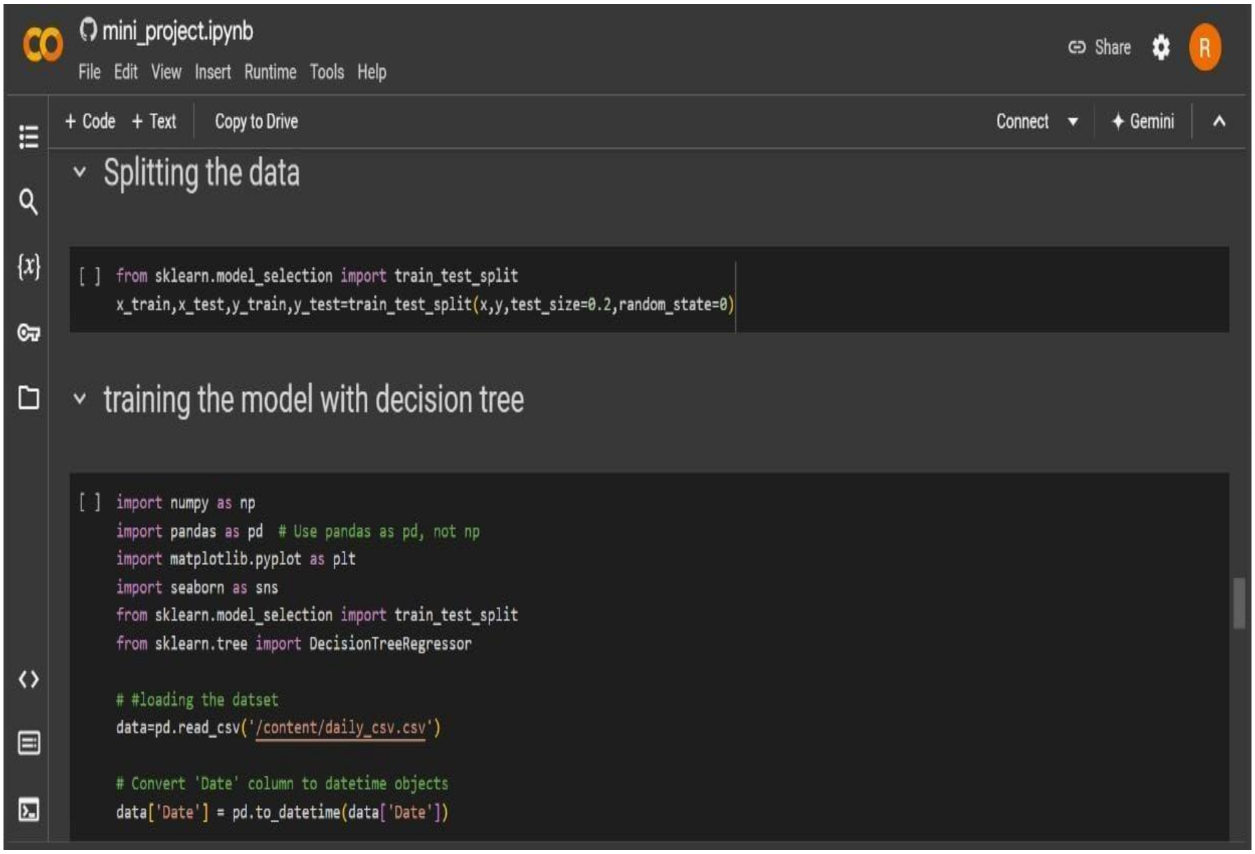


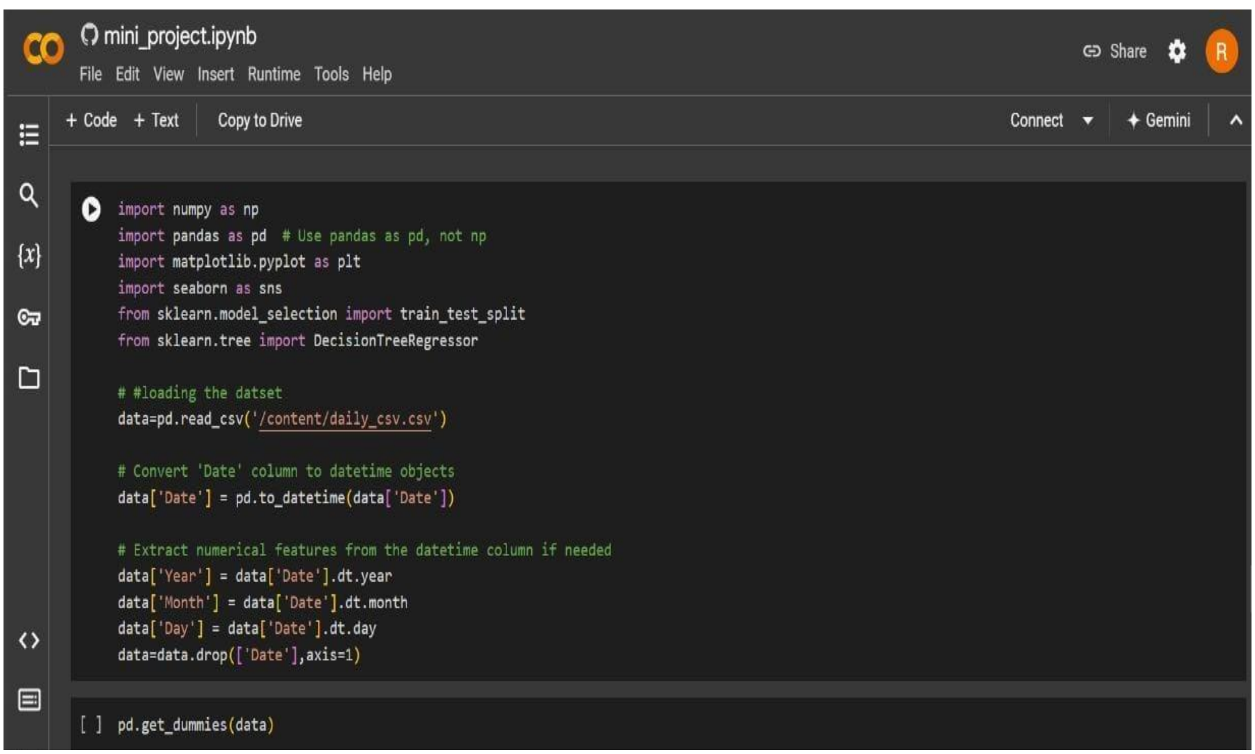
mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
17.5
15.0
12.5
10.0
7.5
5.0
2.5
price
Date


mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
plt.boxplot(y)
plt.show
Show hidden output
sns.boxplot(y)


mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
<Axes: ylabel='Price'>
17.5
15.0
12.5
10.0
7.5
5.0
2.5
Price
[ ] sns.barplot(y)


mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
[ ] sns.barplot(y)
<Axes: ylabel='Price'>
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Price


mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
Splitting the data
[ ] 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)
training the model with decision tree
[ ] import numpy as np
import pandas as pd # Use pandas as pd, not np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
# #loading the datset
data=pd.read_csv('/content/daily_csv.csv')
# Convert 'Date' column to datetime objects
data['Date'] = pd.to_datetime(data['Date'])


mini_project.ipynb
File Edit View Insert Runtime Tools Help
Share
+ Code + Text Copy to Drive
Connect
Gemini
import numpy as np
import pandas as pd # Use pandas as pd, not np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
# #loading the datset
data=pd.read_csv('/content/daily_csv.csv')
# Convert 'Date' column to datetime objects
data['Date'] = pd.to_datetime(data['Date'])
Extract numerical features from the datetime column if needed
data['Year'] = data['Date'].dt.year
data['Month'] = data['Date'].dt.month
data['Day'] = data['Date'].dt.day
data=data.drop(['Date'],axis=1)
[ ] pd.get_dummies(data)

	Price	Year	Month	Day
0	3.82	1997	1	7
1	3.80	1997	1	8
2	3.61	1997	1	9
3	3.92	1997	1	10
4	4.00	1997	1	13
...	...	...	...	...
5933	2.23	2020	8	5
5934	2.26	2020	8	6
5935	2.15	2020	8	7
5936	2.18	2020	8	10
5937	2.19	2020	8	11

5938 rows × 4 columns

```
x_train=data.drop(['Price'],axis=1)
y_train=data['Price']
x_test=data.drop(['Price'],axis=1)
y_test=data['Price']
```

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

```
model=DecisionTreeRegressor()
model.fit(x_train,y_train)
```

```
DecisionTreeRegressor
DecisionTreeRegressor()
```

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

```
array([3.82, 3.8 , 3.61, ..., 2.15, 2.18, 2.19])
```

```
model.predict([[2023,7,26]])
```

```
from sklearn.metrics import r2_score
accuracy=r2_score(y_test,y_pred)
accuracy
```

```
1.0
```

```
train_predictions = model.predict(x_train)
test_predictions = model.predict(x_test)

train_r2 = r2_score(y_train, train_predictions)
test_r2 = r2_score(y_test, test_predictions)

print(f'Training R² score: {train_r2}')
print(f'Test R² score: {test_r2}')
```

```
Training R² score: 1.0
Test R² score: 1.0
```