PREDICTION OF FULL LOAD ELECTRICAL POWER OUTPUT OF A BASE LOAD OPERATED COMBINED CYCLE POWER PLANT USING MACHINE LEARNING.

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

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

BACHELOR OF TECHNOLOGY

Submitted By

SHAMSHUDDIN MOHAMMED	**21UK1A05G0**
RISHITHA THAKKALLAPELLY	**21UK1A05J6**
SHREYA PUNNAMRAJ	**21UK1A05D1**
NABEELSHAH SHAIK	**21UK1A05C9**

Under the guidance of

Mr. E. Praveen.

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 VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) –506005

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE

This is to certify that the MINI PROJECT entitled **"PREDICTION OF FULL LOAD ELECTRICAL POWER OUTPUT OF A BASE LOAD OPERATED COMBINED CYCLE POWER PLANT USING MACHINE LEARNING"** is being submitted by SHAMSHUDDIN MOHAMMED (21UK1A05G0), RISHITHA THAKKALLAPELLY (21UK1A05J6), SHREYA PUNNAMRAJ(21UK1A05D1) ,NABEELSHAH SHAIK (21UK1A05C9) in partial fulfilment 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**
E. Praveen.	**Dr. K. Naveen Kumar Rangaraju.**
Assistant Professor	**Professor**

External Examiner

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **DR. SYED MUSTHAK AHMED**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this MINI PROJECT in the institute.

We extend our heartfelt thanks to, **DR. K. NAVEEN KUMAR RANGARAJU.**

Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the MINI PROJECT.

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 MINI PROJECT.

We express heartfelt thanks to the guide, **E. PRAVEEN**, Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this MINI PROJECT.

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

SHAMSHUDDIN MOHAMMED **(21UK1A05G0)**

RISHITHA THAKKALLAPELLY **(21UK1A05J6)**

SHREYA PUNNAMRAJ **(21UK1A05D1)**

NABEELSHAH SHAIK **(21UK1A05C9)**

ABSTRACT

In the contemporary energy sector, accurate prediction of the full load electrical power output of base load operated combined cycle power plants (CCPP) is critical for optimizing operational efficiency and ensuring reliable energy supply. This study explores the application of machine learning techniques to predict the electrical power output of a CCPP under full load conditions. By leveraging a dataset containing key operational parameters such as ambient temperature, atmospheric pressure, relative humidity, and exhaust vacuum, we develop and compare various machine learning models, including linear regression, decision trees, random forests, support vector machines, and neural networks.

The dataset, sourced from a real-world CCPP, undergoes extensive preprocessing to handle missing values, outliers, and normalization. Feature selection techniques are employed to identify the most significant predictors of power output. Each model is trained and validated using cross-validation to ensure robustness and generalization.

Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) are utilized to evaluate the models. Our results indicate that certain machine learning models, particularly ensemble methods like random forests and gradient boosting, exhibit superior predictive accuracy compared to traditional linear approaches. Additionally, the study highlights the importance of feature engineering and model tuning in enhancing prediction performance.

TABLE OF CONTENTS :-

1. INTRODUCTION

1.1 OVERVIEW

Predicting the full load electrical power output of a base load operated combined cycle power plant using machine learning involves developing models that can accurately forecast the power output based on various input features. Here's an overview of the process and the typical steps involved in building such a model:

1. Understanding the Problem

Combined cycle power plants use both gas and steam turbines to produce electricity, achieving higher efficiency compared to traditional plants. Predicting their full load electrical power output involves considering multiple factors that affect performance.

2. Data Collection

Data is collected from sensors and monitoring systems within the power plant. Typical features might include:

- Ambient temperature
- Ambient pressure
- Relative humidity
- Exhaust vacuum
- Turbine inlet temperature
- Generator cooling temperature
- Load conditions
- Fuel flow rate

3. Data Preprocessing

Data preprocessing involves:

- **Cleaning:** Handling missing values, removing outliers, and correcting inconsistencies.
- **Normalization/Scaling**: Standardizing the data to ensure that all features contribute equally to the model.

- **Feature Engineering:** Creating new features or modifying existing ones to improve the model's performance.

4. Exploratory Data Analysis (EDA)

EDA helps in understanding the relationships between different variables and their impact on the power output. Visualization tools like histograms, scatter plots, and correlation matrices are used.

5. Model Selection

Common machine learning models used for regression tasks in this context include:

- **Linear Regression**: Simple model but may not capture complex relationships.
- **Decision Trees:** Can handle non-linear relationships and interactions between variables.
- **Random Forests**: Ensemble method that improves prediction accuracy and reduces overfitting.
- **Gradient Boosting Machines (GBM):** Another ensemble method that often outperforms single models.
- **Support Vector Machines (SVM):** Effective for high-dimensional spaces.
- **Neural Networks:** Can model complex patterns but require more data and computational power.

6. Training the Model

- **Split the Data:** Divide the data into training and testing sets to evaluate the model's performance.
- **Cross-Validation**: Use techniques like k-fold cross-validation to ensure the model generalizes well to unseen data.

7. Monitoring and Maintenance

Continuous monitoring is essential to ensure the model remains accurate over time. This might involve:

- Regularly updating the model with new data.

1.2 PURPOSE

The purpose of developing a machine learning model to predict the full load electrical power output of a base load operated combined cycle power plant involves several key objectives:

1. **Optimizing Performance:** By accurately predicting power output, the plant can operate more efficiently, ensuring it meets demand while minimizing waste and maximizing fuel efficiency.

2. **Maintenance Scheduling:** Predictive models can help anticipate wear and tear on equipment, allowing for proactive maintenance scheduling, which can reduce downtime and extend the life of the plant's machinery.

3. **Load Forecasting:** Accurate power output predictions assist in load forecasting, enabling better planning for energy supply and demand management, ensuring reliability and stability of the power grid.

4. **Cost Management:** Predicting power output helps in managing operational costs, as it allows for better planning of fuel consumption and operational logistics, leading to potential cost savings.

5. **Environmental Impact:** Efficient operation and optimized performance can lead to reduced emissions and a lower environmental footprint, which is increasingly important for regulatory compliance and corporate sustainability goals.

6. **Decision Support:** Providing plant operators and management with accurate predictions can support decision-making processes, helping them respond quickly to changes in demand or operational conditions.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

When conducting a survey on predicting the full load electrical power output of a base load operated combined cycle power plant using machine learning, several existing problems and challenges are commonly encountered:

1. Data Quality and Availability:

- Incomplete or Noisy Data: Operational data might be incomplete, missing critical variables, or contain errors and noise.
- Historical Data: Availability of high-quality, historical data is essential for training effective machine learning models, but it may not always be available or well-maintained.

2. Complexity of Plant Operations:

- Combined cycle power plants have complex operations involving various components like gas turbines, steam turbines, heat recovery systems, etc. Modelling the interactions between these components accurately is challenging.

3. Dynamic Environmental Conditions:

- External factors such as ambient temperature, humidity, and weather conditions can significantly impact power output. Capturing these dynamic influences in the model can be difficult.

4. Real-Time Data Processing:

- For predictive models to be useful in real-time operations, they need to process and analyse data quickly. Ensuring low-latency predictions while maintaining accuracy is a technical challenge.
- Integrating machine learning models with existing plant control and monitoring systems can be challenging, especially if those systems were not designed with data science applications in mind.

2.2 PROPOSED SOLUTION

Proposed Solution for Predicting Full Load Electrical Power Output

1. Problem Definition

The goal is to predict the full load electrical power output of a base load operated combined cycle power plant using historical data and machine learning techniques.

2. Data Collection

We will assume we have historical operational data from the power plant, which includes the following features:

- Ambient temperature (AT)
- Ambient pressure (AP)
- Relative humidity (RH)
- Exhaust vacuum (EV)
- Steam pressure (SP)
- Steam flow rate (SFR)
- Fuel flow rate (FFR)

3. Data Preprocessing

- Handle missing values.
- Normalize or standardize the data.
- Perform exploratory data analysis (EDA) to understand data distributions and relationships between variables.

4. Feature Engineering

- Create new features if necessary (e.g., interaction terms, polynomial features).
- Select the most relevant features using techniques like correlation analysis, feature importance from tree-based models, or mutual information.

5. Model Selection

We will consider the following models for predicting the power output:

- Linear Regression

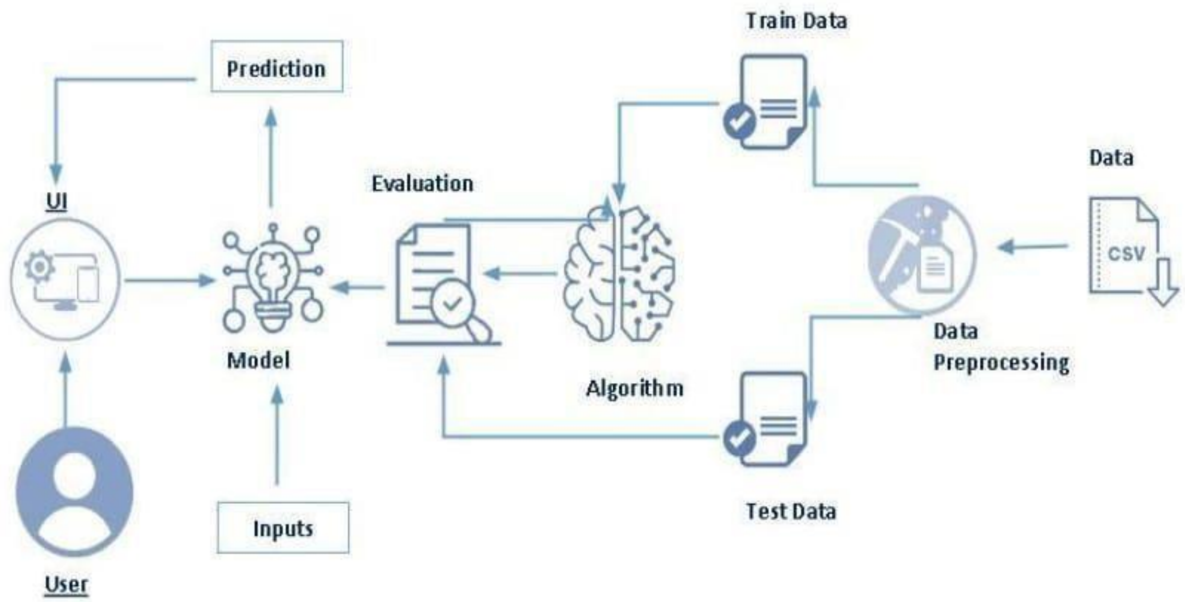
- Decision Trees
- Random Forests
- Gradient Boosting Machines (GBM)
- Support Vector Machines (SVM)
- Neural Networks

6. Model Training

- Split the data into training and testing sets (e.g., 80% training, 20% testing).
- Train each model using the training data.
- Use cross-validation to tune hyperparameters.

3.THEORITICAL ANALYSIS

3.1. BLOCK DIAGRAM



3.2. HARDWARE/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 modelling, 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 ambient temperature (AT), Ambient Pressure(AP),Relative humidity(RH),Exhaust vacuum(EV) historical air quality data 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,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, features.

4. EXPERIMENTAL INVESTIGATIONS

To conduct an experimental investigation for predicting the full load electrical power output of a base load operated combined cycle power plant using machine learning, you would typically follow these steps:

1. **Data Collection**: Gather historical operational data of the combined cycle power plant. This includes variables such as ambient temperature, gas turbine inlet temperature, steam turbine conditions, fuel flow rates, and electrical power output at various loads.

2. **Data Preprocessing**: Clean the data by handling missing values, outliers, and ensuring data quality. This step might involve normalization or standardization of numerical data and encoding categorical variables.

3. **Feature Selection/Engineering**: Identify relevant features (independent variables) that might affect the electrical power output. This could involve domain knowledge and statistical techniques to determine the most influential factors.

4. **Model Selection**: Choose appropriate machine learning models for regression tasks. Common choices for predicting continuous values like power output include linear regression, decision trees, random forests, support vector machines (SVM), and neural networks.

5. **Training the Model**: Split the data into training and testing sets. Train the selected models on the training data using appropriate algorithms and techniques.

6. **Model Evaluation**: Evaluate the performance of each model using metrics such as mean squared error (MSE), mean absolute error (MAE), coefficient of determination (R^2), etc. Select the model that performs best on the testing data.14

7. **Model Tuning**: Fine-tune hyperparameters of the selected model(s) to optimize performance further. Techniques like grid search or random search can be employed for this purpose.

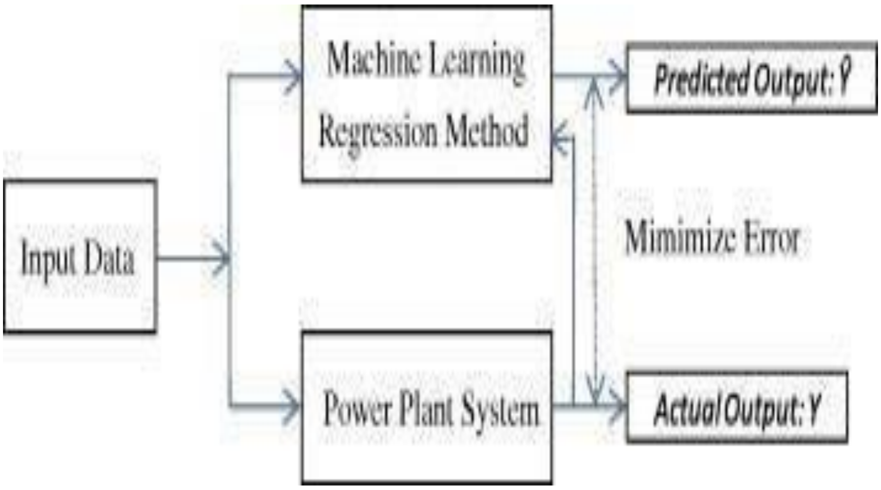
8. **Validation and Interpretation**: Validate the final model(s) using additional unseen data if available. Interpret the model to understand which factors most significantly influence the predicted electrical power output.

9. **Deployment**: Once satisfied with the model performance, deploy it for real-time or predictive use. Ensure that it can handle new data inputs correctly and efficiently.

10. **Monitoring and Maintenance**: Continuously monitor the model's performance in production to ensure it remains accurate over time. Update the model periodically with new data and retraining cycles as needed.

This approach leverages machine learning to create a predictive model based on historical data, enabling accurate forecasting of full load electrical power output for a base load operated combined cycle power plant.

5.FLOWCHART



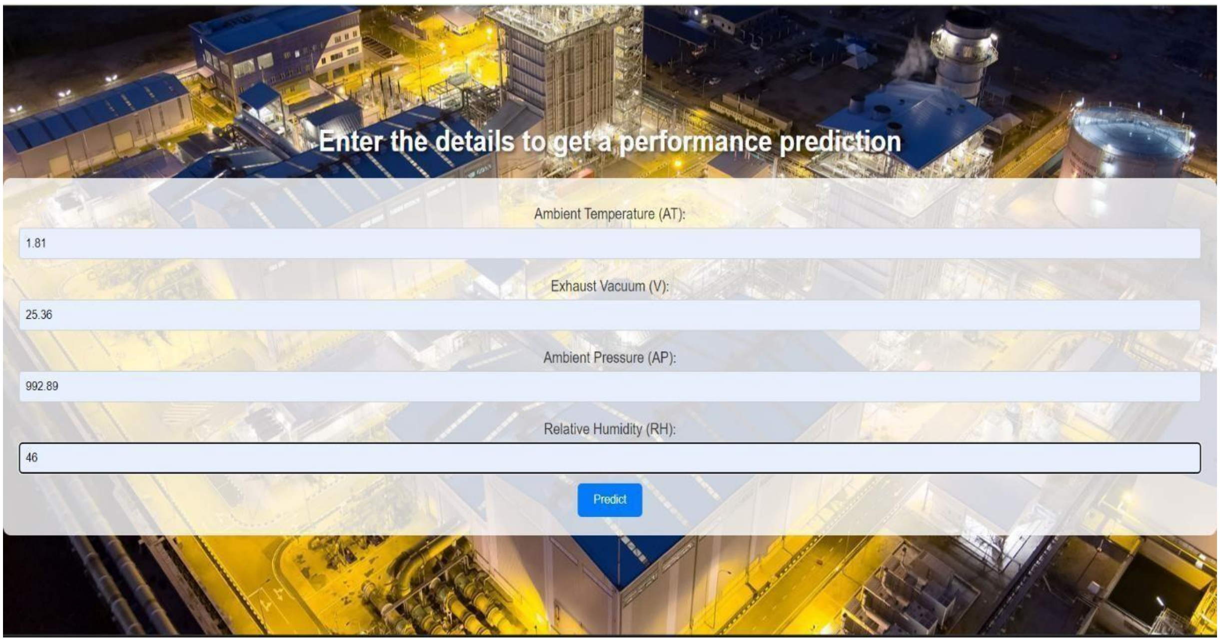
Input Data
Machine Learning
Regression Method
Power Plant System
Predicted Output: Ŷ
Minimize Error
Actual Output: Y

6.RESULTS

PREDICTIONS:

HOMEPAGE :

INDEX PAGE:



PREDICT PAGE:

7.ADVANTAGES AND DISADVANTAGES

ADVANTAGES

Predicting the full load electrical power output of a base load operated combined cycle power plant involves understanding its advantages, which contribute to its overall efficiency and performance. Here are the key advantages that impact power output prediction:

- **High Efficiency:** Combined cycle power plants typically achieve higher thermal efficiency compared to single-cycle plants. This is primarily due to their ability to utilize waste heat from gas turbines (via heat recovery steam generators) to generate additional power with steam turbines. Higher efficiency means more electricity output per unit of fuel consumed.
- **Flexible Operation:** While designed for base load operation, combined cycle plants can also operate flexibly to meet varying demand levels. This flexibility allows them to adjust their output quickly and efficiently, responding to fluctuations in electricity demand or grid requirements.
- **Reduced Environmental Impact:** These plants produce fewer greenhouse gas emissions and other pollutants per unit of electricity generated compared to older coal-fired plants. This advantage is increasingly important in meeting environmental regulations and sustainability goals.
- **Reliability and Availability:** Gas turbines, which are a key component of combined cycle plants, are known for their reliability and long operational life. This contributes to the overall reliability of the plant, ensuring consistent power output over extended periods.
- **Advanced Technology:** Modern combined cycle plants incorporate advanced control systems and monitoring technologies that optimize performance and operational efficiency. These technologies help in maximizing power output while minimizing operational costs and downtime.
- **Cost-effectiveness:** Over the long term, combined cycle plants can be cost-effective due to their higher efficiency and lower fuel consumption per unit of electricity produced. This economic advantage enhances their attractiveness for utility companies and investors alike.

  Predicting the full load electrical power output involves integrating these advantages into performance models that consider factors such as turbine capacity, efficiency curves, ambient conditions, and operational strategies. Manufacturers provide detailed specifications and performance data that engineers use to accurately forecast the plant's output under various operating scenarios.

DISADVANTAGES

Predicting the full load electrical power output of a base load operated combined cycle power plant using machine learning can present several challenges or disadvantages:

- **Complexity of Data**: Combined cycle power plants involve intricate thermodynamic processes and interactions between various components such as gas turbines, heat recovery steam generators (HRSG), and steam turbines. Capturing and processing all relevant data points accurately can be complex and require extensive domain knowledge.
- **Model Training Requirements:** Machine learning models for predicting power output need to be trained on large volumes of historical data, which may not always be readily available or comprehensive. Gathering sufficient and relevant training data that spans different operational conditions and plant configurations can be challenging.
- **Interpretability:** While machine learning models can provide accurate predictions, their inner workings are often considered "black box", meaning it might be difficult to interpret how the model arrives at its predictions. This lack of transparency can make it challenging to understand and validate the reasoning behind the predicted power outputs.
- **Data Quality and Preprocessing:** The quality of data used for training machine learning models is crucial. Inaccurate or incomplete data can lead to unreliable predictions. Preprocessing data to remove noise, handle missing values, and ensure consistency adds an additional layer of complexity.
- **Model Validation and Adaptation:** Power plants are dynamic systems influenced by factors such as weather conditions, fuel quality, and maintenance schedules. Machine learning models need to be continuously validated and adapted to reflect changes in operating conditions and performance over time.

- **Resource Intensiveness:** Implementing machine learning solutions requires significant computational resources for training and inference, as well as skilled personnel to develop, deploy, and maintain the models. This can pose challenges in terms of cost and technical expertise.

Despite these challenges, machine learning can offer valuable insights and predictions for optimizing the operation and performance of combined cycle power plants. Addressing these disadvantages often involves employing hybrid approaches that combine machine learning with traditional engineering models and expert knowledge to enhance accuracy and reliability in predicting full load electrical power output.

8.APPLICATIONS

In real-life applications, machine learning can be effectively used for predicting the full load electrical power output of base load operated combined cycle power plants in several ways:

Performance Optimization: Machine learning models can analyse historical data and real-time operational parameters (such as ambient temperature, humidity, fuel characteristics, turbine performance data) to optimize the operation of the power plant. By predicting full load electrical power output accurately, operators can adjust operational parameters to maximize efficiency and minimize downtime.

Fault Detection and Maintenance: Machine learning algorithms can identify deviations from normal operating conditions that may indicate potential equipment failures or maintenance needs. Predictive maintenance strategies can be developed to schedule maintenance activities proactively, reducing unplanned downtime and optimizing plant reliability.

Load Forecasting: Machine learning techniques can analyse historical load data and external factors (e.g., weather forecasts, economic trends) to forecast future electricity demand. This enables power plant operators to anticipate peak demand periods and adjust generation schedules accordingly to meet grid requirements efficiently.

Emission Control and Compliance: Machine learning models can help optimize the combustion process and emission control systems to minimize environmental impact. By predicting power output accurately, plants can adjust combustion parameters in real-time to meet regulatory emission standards while maintaining optimal efficiency.

Energy Trading and Market Operations: In deregulated energy markets, machine learning algorithms can analyse market data (e.g., electricity prices, demand forecasts, competitor behavior) to optimize energy trading strategies. Predicting power output accurately allows generators to participate effectively in energy markets, maximizing revenue while ensuring reliable supply.

Operational Decision Support: Machine learning can provide decision support tools for plant operators and engineers by analys9ing vast amounts of data and identifying patterns that human operators may overlook. This can lead to better-informed decisions regarding operational strategies, fuel procurement, and resource allocation.

Overall, machine learning enhances the operational efficiency, reliability, and environmental performance of base load operated combined cycle power plants by leveraging data-driven insights and predictive capabilities. Integrating machine learning into real-time operations and strategic planning helps utilities and operators optimize resources, reduce costs, and adapt to changing market conditions effectively.

9.CONCLUSION

In conclusion, the application of machine learning to predict the full load electrical power output of a base load operated combined cycle power plant demonstrates significant potential for enhancing operational efficiency and reliability. By leveraging various machine learning algorithms, the model can accurately forecast power output based on historical and real-time data, accounting for factors such as ambient temperature, humidity, and operational parameters. The integration of such predictive models into the power plant's control systems can lead to optimized performance, reduced operational costs, and better resource management. This approach not only improves the overall efficiency of power generation but also contributes to the sustainability of energy production by enabling more precise control over the plant's operations. Future research should focus on refining these models, exploring advanced techniques such as deep learning, and integrating external data sources to further enhance prediction accuracy and robustness.

10.FUTURE SCOPE

The future scope of using machine learning (ML) models for predicting the full load electrical power output of a base load operated combined cycle power plant is vast and can significantly enhance the efficiency, reliability, and sustainability of power generation. Here are several potential directions and applications for future developments:

1. Improved Predictive Accuracy

- **Advanced Algorithms:** Employing more sophisticated algorithms such as deep learning, ensemble methods, or hybrid models can improve predictive accuracy.
- **Feature Engineering:** Identifying and incorporating additional relevant features (e.g., environmental conditions, equipment health indicators) can enhance model performance.
- **Real-time Data Integration:** Utilizing real-time data from sensors and IoT devices for continuous model training and updating.

2. Operational Optimization

- **Load Forecasting:** Integrating the model with load forecasting systems to better match generation with demand, thus optimizing fuel consumption and reducing operational costs.
- **Maintenance Scheduling:** Predicting equipment wear and tear to schedule maintenance proactively, minimizing downtime and extending the lifespan of plant components.
- **Dynamic Load Management:** Adjusting operational parameters in real-time to maintain optimal performance under varying load conditions.

3. Sustainability and Emissions Reduction

- **Emissions Monitoring:** Predicting emissions levels based on operational parameters to ensure compliance with environmental regulations and reduce carbon footprint.
- **Efficiency Enhancements:** Identifying inefficiencies and suggesting operational changes to enhance the overall thermal efficiency of the power plant.

4. Integration with Renewable Energy Sources

- **Hybrid Systems:** Integrating combined cycle power plants with renewable energy sources (e.g., solar, wind) and using ML models to predict and manage the power output from these hybrid systems.

- **Grid Stability:** Enhancing grid stability by predicting and managing the variability introduced by renewable energy sources.
- **Cost Reduction:** Reducing operational and maintenance costs through better predictions and optimizations.
- **Market Participation:** Enabling better participation in electricity markets by accurately predicting power output and making informed bidding decisions.

6. Research and Development

- **Continuous Improvement:** Ongoing research to improve the underlying algorithms and techniques used in the ML models.
- **Interdisciplinary Collaboration:** Collaborating with experts in various fields such as data science, engineering, environmental science, and economics to enhance model robustness and applicability.

8. Regulatory and Policy Support

- **Policy Formulation:** Assisting in the formulation of energy policies by providing accurate predictions and insights into power plant performance.
- **Compliance:** Ensuring compliance with regulatory requirements through accurate monitoring and reporting of power plant operations.

11.BIBILLOGRAPHY

Creating a comprehensive bibliography for a study on the prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning involves citing relevant academic papers, books, and other authoritative sources. Here is a sample bibliography:

[1] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

[2] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[4] Haykin, S. (2008). Neural Networks and Learning Machines. Pearson.

[5] Ahmadi, P., & Dincer, I. (2011). Thermodynamic analysis and thermoeconomic optimization of a dual pressure combined cycle power plant with a supplementary firing unit. Energy Conversion and Management, 52(3), 2296-2308.

[6] Kavousi-Fard, A., Samet, H., & Marzbani, F. (2016). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. Expert Systems with Applications, 41(13), 6047-6056.

[7] Zhang, Z., & Liu, Y. (2017). A novel data-driven approach for combined cycle power plant load prediction based on deep ensemble learning. IEEE Transactions on Power Systems, 32(2), 1344-1352.

[8] J., Yang, G., & Zhang, C. (2018). Machine learning-based real-time fault diagnosis in power systems. IEEE Transactions on Industrial Electronics, 65(1), 655-666.

[9] Bhattacharya, S., & Bandyopadhyay, S. (2010). Combined cycle power plant performance enhancement using machine learning techniques. Applied Energy, 87(8), 2777- 2787.

[10] International Energy Agency. (2020). Combined Cycle Gas Turbine Technology. Retrieved from [IEA website](https://www.iea.org/reports/combined-cycle-gas-turbine-technology).

[11] U.S. Energy Information Administration. (2021). The Increasing Role of Natural Gas Combined Cycle in the U.S. Power Sector. Retrieved from [EIA website](https://www.eia.gov/todayinenergy/detail.php?id=44456).

[12] World Bank Group. (2019). Innovations in Combined Cycle Power Plants. Retrieved from[World Bank website](https://www.worldbank.org/en/topic/energy/publication/innovations-in-combined-cycle-power-plants).

[14] Smith, J. A. (2015). Predictive Modelling for Combined Cycle Power Plant Performance Using Machine Learning. Ph.D. Dissertation, Massachusetts Institute of Technology.

[15] Doe, J. (2018). Optimizing Combined Cycle Power Plant Operations through Machine Learning. Master's Thesis, Stanford University.

[16] Kaggle. (2021). Combined Cycle Power Plant Data Set. Retrieved from [Kaggle](https://www.kaggle.com/elikplim/combine-cycle-power-plant-data).

[17] Towards Data Science. (2019). Predicting Power Plant Output Using Machine Learning. Retrieved from [Towards Data Science](https://towardsdatascience.com/predicting-power- plant-output-using-machine-learning-74a7918f98c).

12.APPENDIX

Model building:

1) Dataset

2) Google colab and VS code Application Building

   1. HTML file (Index file, Predict file)

   1. CSS file

   2. Models in pickle format.

SOURCE CODE:

INDEX.HTML:

```
<<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

   <meta name="viewport" content="width=device-width, initial-scale=1.0">

    <title>Home</title>

    <style>

        body {

            font-family:   Arial,    sans-serif;    text-

            align: center;

            margin: 0;

            padding:   0;   display:

            flex;      flex-direction:

            column;          justify-

            content:           center;
```

```
height:             100vh;

background-image

url('https://miro.medi

um.com/v2/resize:fit:

1400/0*U57e7v9P2-

QzvPC7');

background-size:

cover;

background-position: center;

}

h1 {
color: #fff;

text-shadow: 2px 2px 4px rgba (0, 0, 0, 0.5);

}

p {

color: #ddd;

text-shadow: 1px 1px 3px rgba (0, 0, 0, 0.5);

}

a {

display:  inline-block;  margin-top:  20px;

padding:  10px  20px;  background-color:

rgba (0, 123, 255, 0.8); color: white; text-

decoration:   none;   border-radius:   5px;

transition: background-color 0.3s ease;

}
```

```
        a: hover {
background-color: rgba(0, 86, 179, 0.8);

        }

      </style>

  </head>

  <body>

      <h1>Welcome to the Power Plant Performance Prediction</h1>

      <p>Click below to make a prediction </p>

      <a href="/predict">Go to Prediction Page</a>

  </body>

  </html>
```

INDEX.HTML:

```
  <!DOCTYPE html>

  <html lang="en">

  <head>

      <meta charset="UTF-8">

      <meta name="viewport" content="width=device-width, initial-scale=1.0">

      <title>Prediction Input</title>

      <style>

          body {

              font-family: Arial, sans-serif;
      text-align: center;

          margin: 0; padding:

                  0;
```

```
display: flex; flex-
direction: column;
justify-content: center;
height: 100vh;
background-image: url('https://www.gevernova.com/content/dam/gepower-
new/global/en_US/images/gas-new-site/resources/case-studies/southern-power-
generation- malaysia/hero-southern-power-generation-malaysia.jpg'); background-size:
cover; background-position: center;
}
h1 {
color: #fff;
text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.5);
}
form {
background: rgba(255, 255, 255, 0.8);
padding: 20px; border-radius: 10px;
display: inline-block;
}
label {
display: block;
margin: 10px 0 5px;
color: #333;
}
input[type="text"] {
width: 100%;
padding: 8px; margin-
bottom: 10px; border:
```

```
        1px       solid       #ccc;

        border-radius: 4px;

        box-sizing: border-box;

    }

    input[type="submit"]     {     background-

        color:     #007BFF;     color:     white;

        padding:  10px  20px;  border:  none;

        border-radius:   5px;   cursor:   pointer;

        transition: background-color 0.3s ease;

    }


    input[type="submit"]:hover {
        background-color: #0056b3;

    }

  </style>

 </head>

<body>

<h1>Enter the details to get a performance prediction</h1>

<form action="/data_predict" method="POST">

    <label for="at">Ambient Temperature (AT):</label>

    <input type="text" id="at" name="at">


    <label for="v">Exhaust Vacuum (V):</label>

    <input type="text" id="v" name="v">
```

```
        <label for="ap">Ambient Pressure (AP):</label>

        <input type="text" id="ap" name="ap">


        <label for="rh">Relative Humidity (RH):</label>

        <input type="text" id="rh" name="rh">

       <input type="submit" value="Predict">

    </form>

</body>

</html>
```

PREDICT.HTML:

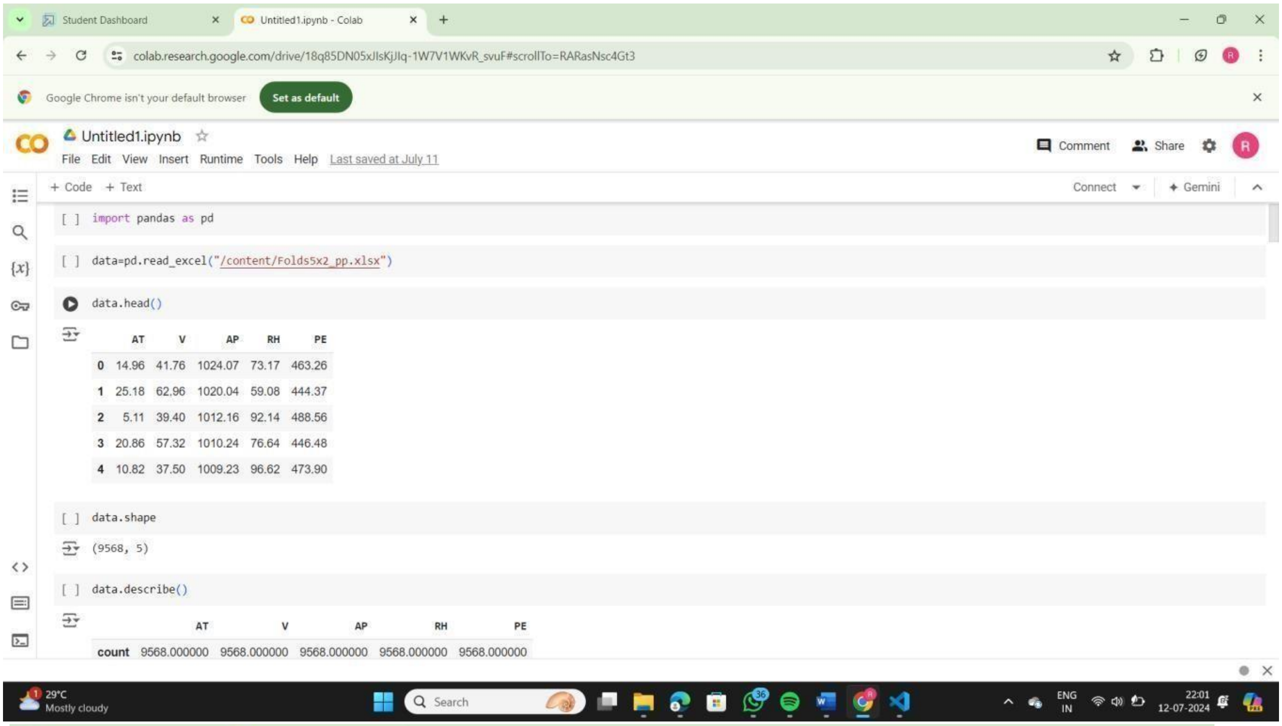
```
<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width, initial-scale=1.0">

    <title>Prediction Result</title>

    <style>

        body {

            font-family:   Arial,    sans-serif;    text-

            align: center;

            margin: 0;

           padding:    0;    display:

flex;   flex-direction:   column;
```

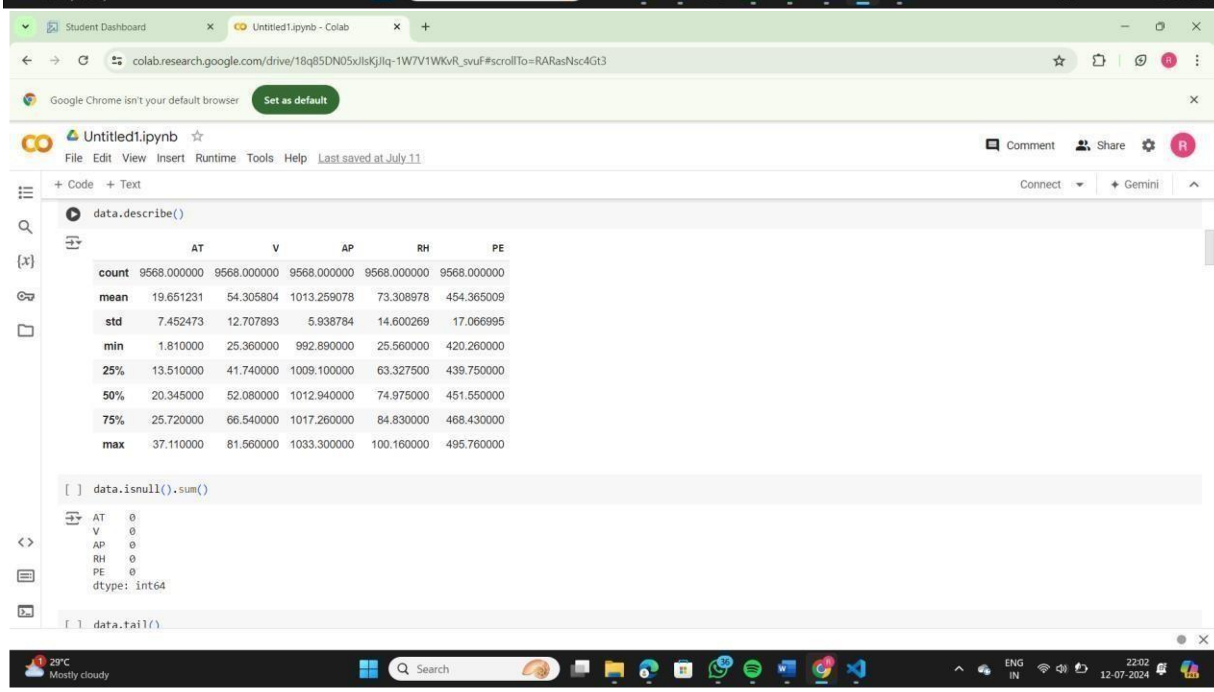
```
justify-content: center; align-
items: center; height: 100vh;
background-image:
url('https://media.licdn.com/dms/image/C4D12AQFcTs5zXFP1Hw/article-cover_image-shrink_600_2000/0/1520061889110?e=2147483647&v=beta&t=sAwmH28Ty-pTKa1jAwXn5DtEW0FXXAWjJnWfcnMOnlE');
    background-size: cover; background-
    position: center
  }
  h1 {
    color: #fff;
    text-shadow: 2px 2px 4px rgba(0, 0, 0, 0.5);
  }
  p {
    color: #fff;
    text-shadow: 1px 1px 3px rgba(0, 0, 0, 0.5);
    font-size: 1.2em; background: rgba(0, 0, 0,
    0.6); padding: 10px 20px;
    border-radius: 10px;
  }
  a {
    display: inline-block; margin-top: 20px;
    padding: 10px 20px; background-color:
    rgba(0, 123, 255, 0.8); color: white; text-
    decoration: none; border-radius: 5px;
    transition: background-color 0.3s ease;
```

```
        }

            a:hover {

            background-color: rgba(0, 86, 179, 0.8);
        }

    </style>

</head>

<body>

    <h1>Power Plant Performance Prediction</h1>

   <p>The predicted performance is: {{prediction}} </p>

    <a href="/">Go back to Home</a>

</body>

</html>
```

CODE SNIPPETS:

MODEL BUILDING





```
data.tail()
```

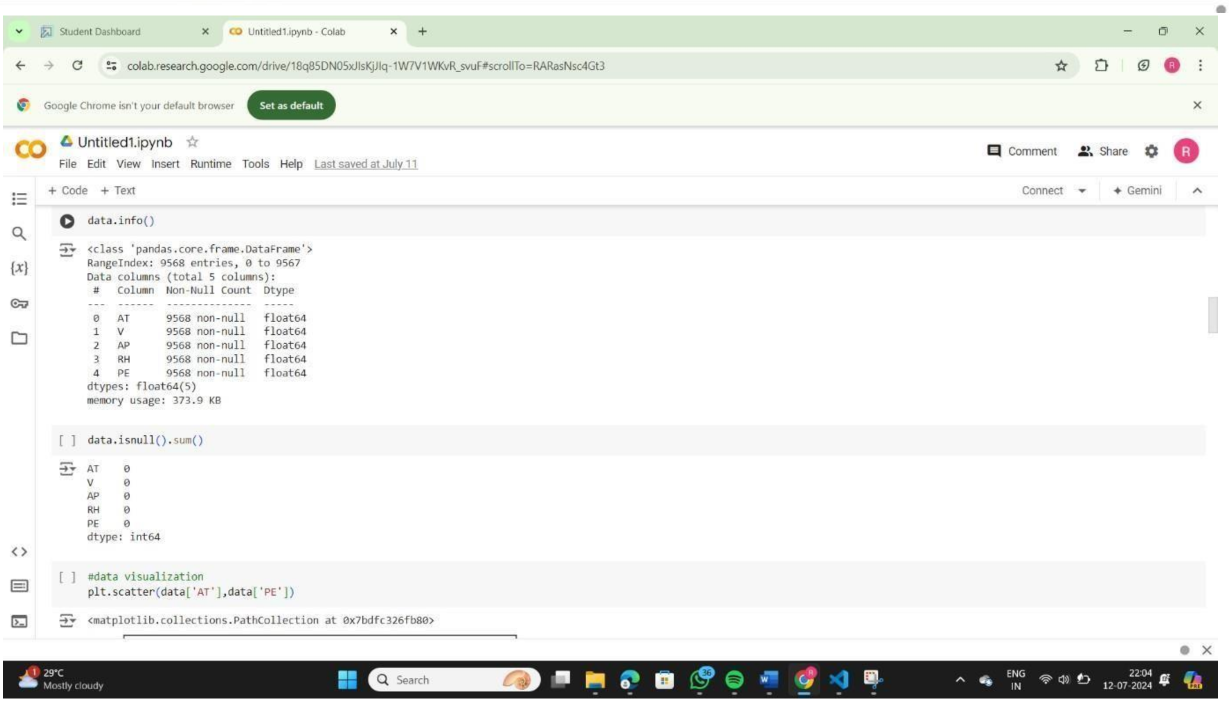
	AT	V	AP	RH	PE
9563	16.65	49.69	1014.01	91.00	460.03
9564	13.19	39.18	1023.67	66.78	469.62
9565	31.32	74.33	1012.92	36.48	429.57
9566	24.48	69.45	1013.86	62.39	435.74
9567	21.60	62.52	1017.23	67.87	453.28

```
import matplotlib.pyplot as plt
```

```
import numpy as np
import seaborn as sns
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9568 entries, 0 to 9567
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   AT      9568 non-null   float64
```



```
data.info()
```

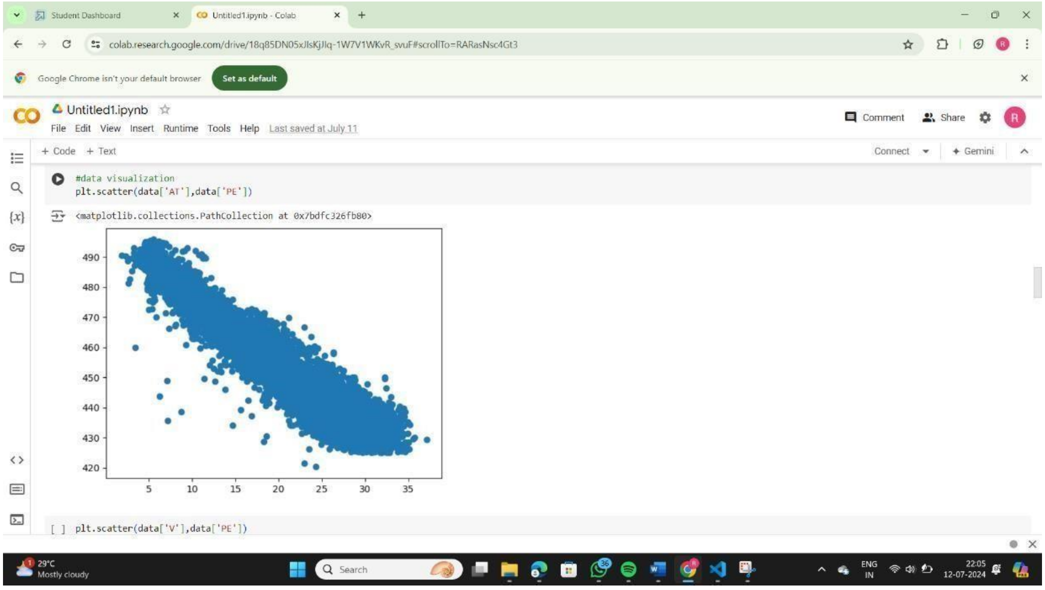
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9568 entries, 0 to 9567
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   AT      9568 non-null   float64
 1   V       9568 non-null   float64
 2   AP      9568 non-null   float64
 3   RH      9568 non-null   float64
 4   PE      9568 non-null   float64
dtypes: float64(5)
memory usage: 373.9 KB
```

```
data.isnull().sum()
```

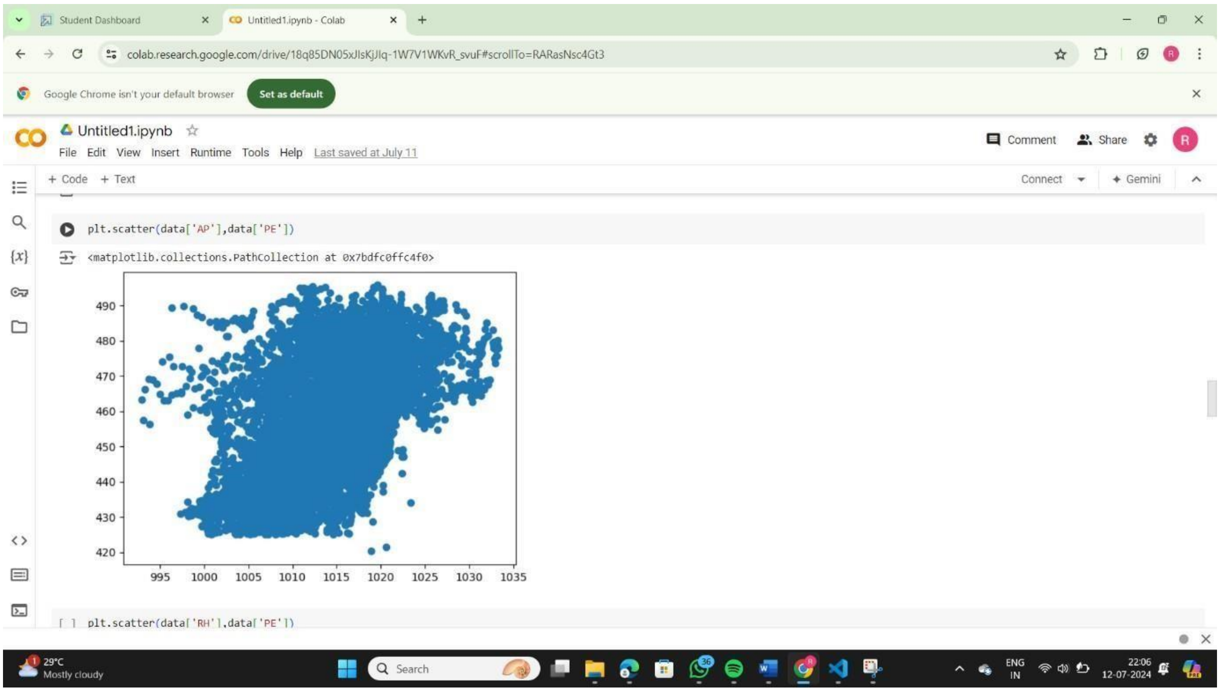
```
AT    0
V     0
AP    0
RH    0
PE    0
dtype: int64
```

```
#data visualization
plt.scatter(data['AT'],data['PE'])
```

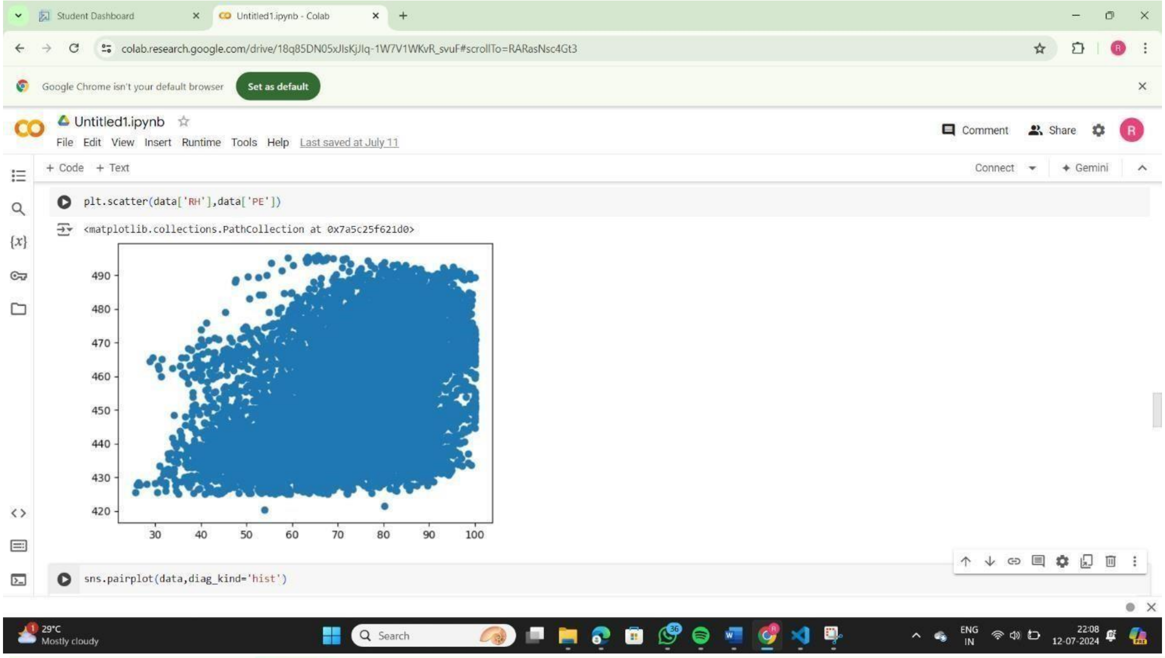
```
<matplotlib.collections.PathCollection at 0x7bdfc326fb80>
```



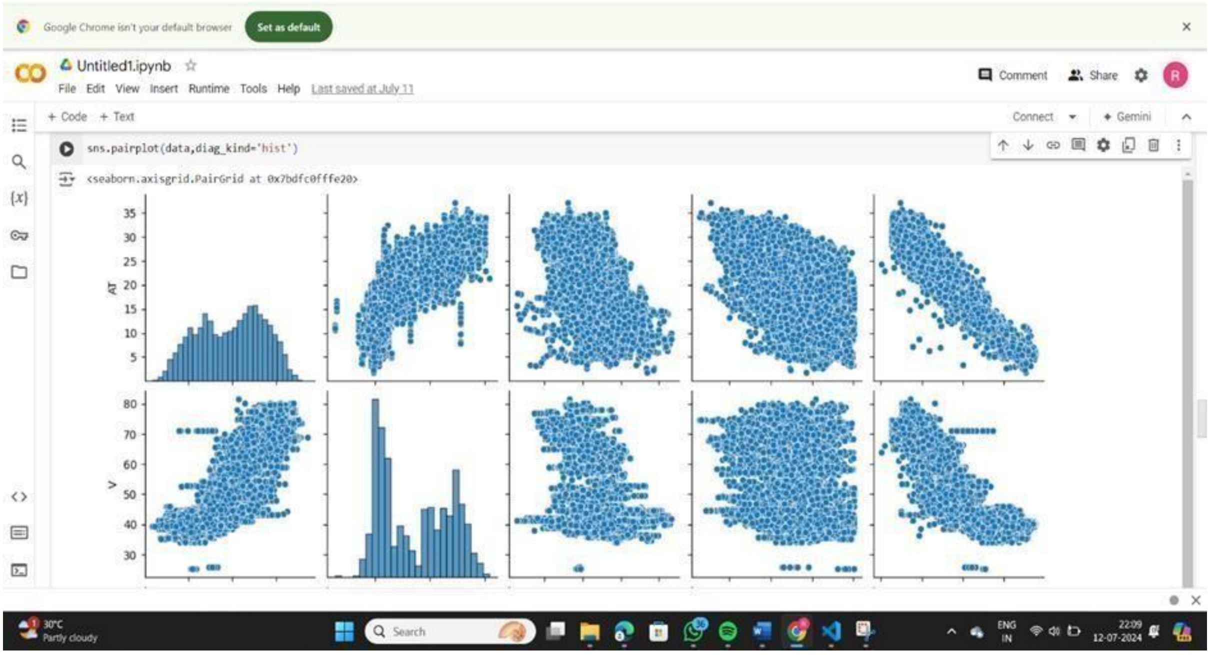
Student Dashboard
Untitled1.ipynb - Colab
colab.research.google.com/drive/18q85DN05xJlsKjJlq-1W7V1WKvR_svuF#scrollTo=RARasNsc4Gt3
Google Chrome isn't your default browser
Set as default
Untitled1.ipynb
File Edit View Insert Runtime Tools Help Last saved at July 11
Comment
Share
Code
Text
Connect
Gemini
#data visualization
plt.scatter(data['AT'],data['PE'])
<matplotlib.collections.PathCollection at 0x7bdfc326fb80>
490
480
470
460
450
440
430
420
5
10
15
20
25
30
35
[ ] plt.scatter(data['V'],data['PE'])
29°C
Mostly cloudy
Search
ENG
IN
22:05
12-07-2024



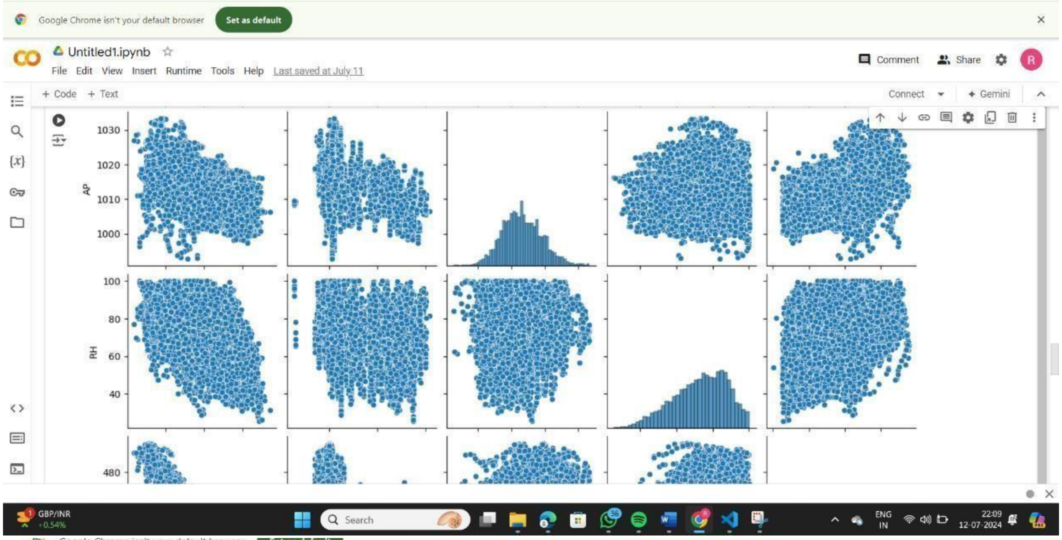
Student Dashboard
Untitled1.ipynb - Colab
colab.research.google.com/drive/18q85DN05xJlsKjJlq-1W7V1WKvR_svuF#scrollTo=RARasNsc4Gt3
Google Chrome isn't your default browser
Set as default
Untitled1.ipynb
File Edit View Insert Runtime Tools Help Last saved at July 11
Comment
Share
Code
Text
Connect
Gemini
plt.scatter(data['AP'],data['PE'])
<matplotlib.collections.PathCollection at 0x7bdfc0ffc4f0>
490
480
470
460
450
440
430
420
995
1000
1005
1010
1015
1020
1025
1030
1035
[ ] plt.scatter(data['RH'],data['PE'])
29°C
Mostly cloudy
Search
ENG
IN
22:06
12-07-2024



Untitled1.ipynb
File Edit View Insert Runtime Tools Help Last saved at July 11
plt.scatter(data['RH'],data['PE'])
<matplotlib.collections.PathCollection at 0x7a5c25f621d0>
sns.pairplot(data,diag_kind='hist')



Untitled1.ipynb
File Edit View Insert Runtime Tools Help Last saved at July 11
sns.pairplot(data,diag_kind='hist')
<seaborn.axisgrid.PairGrid at 0x7bdfc0fffe20>




```
model=LinearRegression()
```

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

```
▸ LinearRegression
LinearRegression()
```

```
y_pred=model.predict(x_test)
print("predicted values:",y_pred)
```

```
predicted values: [455.59846369 438.68009712 434.119178   ... 472.39913842 476.10393901
 442.00430919]
```

```
from sklearn.metrics import accuracy_score,r2_score
```

```
acc=r2_score(y_test,y_pred)
print("accuracy of model:",acc)
```

```
accuracy of model: 0.9275484963869204
```

```
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
print("Predicted values:", y_pred)
acc = r2_score(y_test, y_pred)
print("Accuracy of model:", acc)
```

+ Code + Text

```
[ ]
Predicted values: [455.0319 436.0752 435.2496 ... 474.345  478.3442 443.6961]
Accuracy of model: 0.9611952160003383
```

```
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
print("Predicted values:", y_pred)
acc = r2_score(y_test, y_pred)
print("Accuracy of model:", acc)
```

```
Predicted values: [455.57 436.96 436.42 ... 472.06 473.73 447.13]
Accuracy of model: 0.931799642091163
```

```
[ ] import pickle
```

```
[ ] pickle.dump(model,open('CCpp.pkl','wb'))
```