INLIGHN TECH

Website: https://www.inlighntech.com/
LinkedIn: https://www.linkedin.com/company/inlighn-tech/
Contact: support@inlighntech.com

Project Title:

Air Quality Prediction Using Machine Learning

Technologies:

Python, Google Colab

Difficulty Level:

Medium

Project Description:

This project focuses on predicting air quality using machine learning techniques. You will build a model to analyze and forecast air pollution levels based on historical data. This hands-on experience will help you understand how to preprocess time-series data, build predictive models, and evaluate their performance.

Project Requirements:

Tools and Environment

- **Platform:** Use Google Colab to run your code. It is pre-installed with necessary libraries and simplifies collaboration and sharing.
- **Dataset:** Download the **Air Quality Dataset** from the UCI Machine Learning Repository.

Libraries Required:

1. **Data Manipulation and Analysis:**
    - pandas
    - numpy
2. **Data Visualization:**
    - matplotlib
    - seaborn

3. **Machine Learning:**
    - scikit-learn
4. **Forecasting:**
    - fbprophet

Step-by-Step Guide:

1. Set Up Your Workspace

- Open Google Colab.
- Create a new notebook and name it **Air Quality Prediction**.
- Upload the dataset to Colab.

2. Data Understanding

- Load the dataset into a pandas DataFrame.
- Explore the data:
    - View the first few rows using .head().
    - Check for missing values using .isnull().sum().
    - Display data types and statistics using .info() and .describe().

3. Data Preprocessing

- Handle missing values:
    - Replace invalid or missing values (-200) with NaN.
    - Use the column mean to fill missing values.
- Convert Date and Time columns into a single datetime column. Ensure it follows the YYYY-MM-DD HH:MM:SS format.

4. Feature Engineering

- Retain important columns, such as pollutant levels and weather conditions.
- Scale the data using StandardScaler or MinMaxScaler.

5. Model Development

- Split the data into training and testing sets using train_test_split.
- Build a time-series forecasting model using fbprophet:
    - Rename the datetime column to ds and the target column to y.
    - Train the model on ds and y.
    - Predict future air quality levels.

6. Model Evaluation

- Use evaluation metrics such as:
    - **Mean Absolute Error (MAE)**
    - **Root Mean Square Error (RMSE)**
    - **R-squared (R^2)**
- Visualize actual vs. predicted pollutant levels using matplotlib.

Extra Challenges (Optional):

1. Predict air quality for a specific future period (e.g., one week ahead).
2. Create a dashboard for interactive visualization.
3. Use additional models like LSTM for comparison.

Submission Guidelines:

1. **Repository Structure:**
    - Create a folder or GitHub repository containing the following:
        - .ipynb file of your Colab notebook
        - The dataset used
2. **Documentation:**
    - Add markdown cells in your notebook to explain each step.
    - Ensure your code is well-commented.
3. **Testing:**
    - Ensure your model runs end-to-end without errors.
    - Validate that your model generates accurate predictions.

Assessment Criteria:

1. **Code Quality:** Clear, modular, and well-commented code.
2. **Data Preprocessing:** Appropriate handling of missing values and feature transformations.
3. **Model Accuracy:** Use metrics to evaluate performance.
4. **Visualization:** Effective plots to explain data and results.
5. **Submission Quality:** Structured and error-free notebook submission.

Deliverables:

1. .ipynb notebook with all required code and outputs.
2. Visualizations of predicted vs. actual air quality levels.
3. Explanation of results.

This project will prepare you for real-world applications of machine learning in environmental analysis. Let's create impactful predictions! Good luck!