**Project Initialization and Planning Phase**

|  |  |
| --- | --- |
| Date | 07 JULY 2024 |
| Team ID | 739850 |
| Project Title | Air Quality Index analyzer using ML |
| Maximum Marks | 3 Marks |

**Project Proposal (Proposed Solution) template**

To explain an Air Quality Index (AQI) analyzer using machine learning (ML), you can structure it similarly to the project proposal template shown in the image. Here's an outline:

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Overview** |  | | |
| Objective | Develop a machine learning system to analyze and predict air quality index levels in real-time. | | |
| Scope | Implement a scalable fraud detection system that can analyze large volumes of transaction data to identify and flag potentially fraudulent activities. The project will cover data collection, model training, and integration with existing transaction systems. | | |
| **Problem Statement** |  | | |
| Description | Air pollution is a significant issue affecting public health and the environment. Accurate and timely predictions of AQI can help in mitigating health risks and informing the public and authorities about air quality. | | |
| Impact | Solving this problem will enhance public health safety, inform policy decisions, and improve the overall quality of life. It will also help in identifying pollution sources and trends, facilitating better environmental management. | | |
| **Proposed Solution** |  | | |
| Approach | Utilize supervised machine learning techniques, such as regression models, decision trees, and neural networks, to analyze historical and real-time environmental data to predict AQI levels. The solution will involve data preprocessing, feature engineering, model training, and evaluation. | | |
| **Resource Type** | | **Description** | **Specification/Allocation** |
| **Hardware** | |  | |
| Computing Resources | | High-performance CPUs/GPUs | e.g., 2 x NVIDIA V100 GPUs |
| Memory | | Sufficient RAM for large datasets | e.g., 32 GB |
| Storage | | Large storage for data, models, and logs. | e.g., 1 TB SSD |
| **Software** | |  | |
| Frameworks | | Python frameworks | e.g., Tenser flow, sklearn, keras. |
| Libraries | | Pandas, NumPy, Matplotlib for data manipulation and  visualization | e.g., numpy, pandas. |
| Development Environment | | Jupyter Notebooks, IDEs | e.g., Pycharm |
| **Data** | |  | |
| Data | | Source: Government and private environmental monitoring agencies, open data | e.g., Kaggle |

|  |  |
| --- | --- |
| Data collection: | Sources: Meteorological data, pollution sensors, satellite data  Types: Temperature, humidity, wind speed, particulate matter (PM2.5, PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO) |
| Data preprocessing: | Cleaning: Handle missing values, remove outliers  Transformation: Normalize/standardize data  Feature Engineering: Create new features from raw data |
| Model Training: | Algorithms: Linear regression, random forest, gradient boosting, deep learning models  Evaluation: Cross-validation, performance metrics (RMSE, MAE, R2 score)  Integration: Real-time data ingestion and prediction  Visualization: Dashboards and alerts for AQI levels |

**Resource Requirements**

|  |  |  |
| --- | --- | --- |
|  | platforms (e.g., Kaggle) Size: Varies depending on the region and time span Format: CSV, JSON, realtime API feeds |  |