# TITLE: EcoForecast: AI-Powered Prediction of Carbon Monoxide Levels

## **Milestone 1: Project Initialization and Planning Phase**

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

## Activity 1: Define Problem Statement

EcoForecast addresses the pressing issue of unpredictable carbon monoxide (CO) levels in the environment, which pose significant challenges for public health and environmental sustainability. Traditional methods of monitoring and forecasting CO levels often fall short due to their reliance on static models and historical data, hindering timely interventions to mitigate air pollution's adverse effects. By leveraging artificial intelligence and machine learning, EcoForecast aims to provide accurate and real-time forecasts of CO concentrations, empowering decision-makers with actionable insights to implement proactive measures and protect public health.

## Activity 2: Project Proposal (Proposed Solution)

EcoForecast proposes the development of an AI-powered prediction system that utilizes historical data, meteorological factors, and emission patterns to forecast CO concentrations in the environment. By employing sophisticated machine learning algorithms, such as neural networks and ensemble methods, the system will learn complex patterns and relationships in the data to generate accurate predictions.

## Activity 3: Initial Project Planning

The EcoForecast project utilizes AI and machine learning to predict carbon monoxide levels in the environment, aiming to mitigate air pollution and safeguard public health. It involves activities such as data collection and preparation, exploratory data analysis, model building, performance testing, and model deployment. These tasks are organized into sprints with specific functional requirements, user stories, story points, priorities, and team members assigned, ensuring efficient progress towards project goals.

### Milestone 2: Data Collection and Preprocessing Phase

There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc. In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset.

Link: <https://www.kaggle.com/datasets/anubhav3242/carbon-monoxide-ppm-data-for-regression>

As the dataset is downloaded. Let us read and understand the data properly with the help of some visualization techniques and some analyzing techniques.

Note: There are a number of techniques for understanding the data. But here we have used some of it. In an additional way, you can use multiple techniques.

The Machine Learning model cannot be trained on the imported data directly. The dataset might have randomness, we might have to clean the dataset and bring it in the right form. This activity involves the following steps:

* Handling Missing Values
* Handling Categorical Data
* Handling Imbalance Data

Data collection and preprocessing are crucial phases in AI-Powered Prediction for CO levels, ensuring the quality and relevance of input data for subsequent analysis. This phase involves gathering diverse datasets including Temperature, Time Stamp, Humidity and ppm. Data preprocessing techniques such as normalization, noise reduction, and visualization are applied to ensure consistency and accuracy. By meticulously curating and preparing the data, AI-Powered Prediction for CO levels can generate reliable insights into public health.

## Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

For a AI-Powered Prediction for CO levels project, the data collection plan outlines the methodology for gathering information related to public health, likely including parameters like Temperature, Time Stamp, Humidity and ppm.

|  |
| --- |
| The data quality report would assess the accuracy, completeness, and reliability of the collected data, ensuring it meets the standards required for AI analysis and decision-making in public health monitoring. This includes checking for any biases, errors, or missing information that could impact the effectiveness of the AI system. |
|  |

## Activity 2: Data Quality Report

The Data Quality Report for a AI-Powered Prediction for CO levels project assesses the reliability and accuracy of the collected data, crucial for ensuring the effectiveness of AI algorithms in monitoring public health. It examines factors like data completeness, consistency, accuracy, and potential biases. By identifying and addressing any issues, the report aims to enhance the trustworthiness and efficacy of the AI system.

|  |
| --- |
|  |

## Activity 3: Data Exploration and Preprocessing

Data exploration and preprocessing for a AI-Powered Prediction for CO levels project involve initial analysis and preparation of the collected data before feeding it into machine learning models. This includes tasks such as examining data distributions, identifying outliers, handling missing values, and potentially normalizing or scaling features. The goal is to ensure that the data is clean, relevant, and structured optimally for subsequent analysis, ultimately improving the performance and interpretability of the AI system in monitoring public health.

### Milestone 3: Model Development Phase

In the model development phase of a AI-Powered Prediction for CO levels project, machine learning algorithms are applied to the preprocessed data to build predictive models for monitoring public health. This phase involves selecting appropriate algorithms, training them on the prepared data, and Decision Tree Regression, Random forest regression, XGBoost, K-Nearest Neighbour, Linear Regression model parameters for optimal performance. Additionally, techniques such as cross-validation and hyperparameter tuning are employed to ensure the robustness and generalizability of the models. The ultimate goal is to develop accurate and reliable AI models capable of detecting CO levels and providing valuable insights into public well-being.

## Activity 1: Feature Selection Report

The Feature Selection Report in a AI-Powered Prediction for CO levels project evaluates the relevance and importance of different variables in predicting public health outcomes. It identifies the most informative features while potentially reducing dimensionality and computational complexity. Techniques such as statistical tests, correlation analysis, and machine learning algorithms are utilized to rank and select the most influential features. The report aims to optimize model performance, interpretability, and computational efficiency by focusing on the most relevant factors for public health monitoring.

## Activity 2: Model Selection Report

The Model Selection Report in a AI-Powered Prediction for CO levels project compares the performance of various machine learning algorithms to determine the most suitable model for predicting public health outcomes. It assesses metrics such as accuracy, precision, recall, and F1 score across different models like Decision Tree Regression, Random forest regression, XGBoost, K-Nearest Neighbour, Linear Regression model, considering factors like computational complexity and interpretability. Techniques such as cross validation and hyperparameter tuning are employed to ensure robust evaluation. The report aims to identify CO levels.

## Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code person selected algorithms on the arduino\_data dataset, setting the foundation for predictive modelling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, person metrics like accuracy and precision to ensure reliability and effectiveness in predicting outcomes.

The initial model for AI-Powered Prediction for CO levels project, a Random Forest Classifier, was trained on features extracted from public health data. Evaluation on a test set yielded an accuracy of 0.92, precision of 0.93, and F1score of 0.92. Further validation on diverse datasets is recommended for broader applicability.

### Milestone 4: Model Optimization and Tuning Phase

In the Model Optimization and Tuning Phase of the AI-Powered Prediction for CO levels project, techniques such as hyperparameter tuning and feature engineering are employed to enhance the performance of the predictive models. This involves fine-tuning model parameters, optimizing algorithms, and selecting the most relevant features to improve accuracy, precision, recall, and overall model performance. The goal is to refine the AI system for more accurate and reliable prediction of public health outcomes, thereby facilitating better decision-making.

## Activity 1: Hyperparameter Tuning Documentation

The Hyperparameter Tuning Documentation in the AI-Powered Prediction for CO levels project outlines the process of optimizing model performance by adjusting hyperparameters. Techniques such as grid search or random search are typically used to explore different combinations of hyperparameters and identify the optimal settings. The documentation includes details on the hyperparameters being tuned, the search strategy employed, and the evaluation metrics used to assess performance. The aim is to fine-tune the model for improved accuracy, precision, recall, and overall effectiveness in predicting public health outcomes.

## Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report in the AI-Powered Prediction for CO levels project compares various evaluation metrics across different models or approaches. Metrics such as accuracy, precision, recall, and F1score are typically assessed to gauge the performance of each model in predicting public health outcomes. The report summarizes the findings, highlighting the strengths and weaknesses of each approach and providing insights into the most effective methods for decision-making.

## Activity 3: Final Model Selection Justification

The Final Model Selection Justification in the AI-Powered Prediction for CO levels project is based on a comprehensive evaluation of model performance metrics such as accuracy, precision, recall, and F1-score. The selected model demonstrates the highest predictive capability and generalizability across diverse datasets. Its effectiveness in accurately predicting public health outcomes makes it the most suitable choice for deployment in clinical settings, offering valuable support for prenatal care decision making.

## Milestone 5: Project Files Submission and Documentation

For project file submission in Github, Kindly click the link and refer to the flow.

<https://github.com/srivani47/EcoForecast-AI-Powered-Prediction-of-Carbon-Monoxide-Levels.git>

For the documentation, Kindly refer to the link.

<https://github.com/srivani47/EcoForecast-AI-Powered-Prediction-of-Carbon-Monoxide-Levels.git>

## Milestone 6: Project Demonstration

In the upcoming module called Project Demonstration, individuals will be explaining about the project by recording the problem statement and running the application providing with few examples .