**PROJECT INTIALIZATION DOCUMENT**

**1. DETAILS**

**Team Details:** Team consists of two people, they are Nijgururaj Ashtagi([nijgururaja12@gmail.com](mailto:nijgururaja12@gmail.com)) and Vaishnavi Patange(vrp345678@gmail.com)

**Project Name:** Life Expectancy Prediction

**2. DEFINING PROJECT AND ITS SCOPE**

**Understanding of the project:**

The Life Expectancy prediction project aims to build a machine learning model that can predict the life expectancy of a person based on various socio-economic and health-related factors. The project will involve collecting and analysing data from multiple sources, identifying relevant features, and building a predictive model that can accurately estimate life expectancy. The model will be trained on historical data and tested on new data to evaluate its performance.

**Reason for choosing this project:**

There are several reasons why the Life Expectancy prediction project is relevant and important. Firstly, it can help healthcare professionals and policymakers identify high-risk populations and develop targeted interventions to improve their life expectancy. Secondly, it can provide individuals with valuable information about their health and encourage them to adopt healthier lifestyles. Lastly, the project presents an opportunity to develop and apply advanced machine learning techniques to a real-world problem, contributing to the field of data science.

**Most challenging aspect of the project statement:**

One of the most challenging aspects of the Life Expectancy prediction project is identifying the most relevant features to include in the predictive model. There are numerous socio-economic and health-related factors that can affect life expectancy, and it can be difficult to determine which ones are most important. Additionally, there may be confounding factors or interactions between different features that need to be accounted for in the model. Overcoming these challenges will require careful data analysis and feature selection, as well as expertise in machine learning algorithms and techniques.

**3. APPROACH OF PROBLEM CHOSEN**

**Approach:**

1. Data Collection: Gather data from various sources, such as government statistics, health surveys, and medical databases. Ensure that the data is relevant, reliable, and contains information on socio-economic and health-related factors that can affect life expectancy.

2. Data Pre-processing: Clean the data and perform necessary pre-processing tasks, such as handling missing values, removing duplicates, and encoding categorical variables. Also, perform exploratory data analysis (EDA) to gain insights into the data and identify any outliers or anomalies.

3. Feature Selection: Identify the most relevant features that can predict life expectancy using techniques such as correlation analysis, feature importance, and domain knowledge. Perform feature engineering to create new features that may enhance the predictive power of the model.

4. Model Selection: Select an appropriate machine learning algorithm that can handle the data and produce accurate predictions. Compare the performance of various models using metrics such as accuracy, precision, recall, and F1 score. Consider using techniques such as cross-validation and hyperparameter tuning to optimize the model's performance.

5. Model Evaluation: Test the model's performance on a new dataset to evaluate its accuracy and generalization ability. Calculate evaluation metrics such as root mean squared error (RMSE), mean absolute error (MAE), and R-squared to measure the model's performance.

6. Model Deployment: Deploy the model into a production environment where it can be used to predict life expectancy for new data. Consider using technologies such as APIs or web applications to provide a user-friendly interface for accessing the model's predictions. Model

7. Maintenance: Continuously monitor the model's performance and update it as necessary to ensure that it remains accurate and relevant. Consider using techniques such as retraining the model with new data or using ensemble models to improve the model's performance over time.

These steps can provide a framework for your project and help ensure that you address all the necessary aspects of the Life Expectancy prediction problem.

**Diagram/Flowchart:**

**Platform/Coding Language/Frameworks (if using):** Python, Jupyter Notebook, Pandas, Scikit-learn, Matplotlib, Seaborn, Django

**4. TEAMS ABILITY TO IMPLEMENT WINNING SOLUTION**

**Background of team members/individual:** Nijgururaj Ashtagi is currently pursuing third year Computer Engineering from DYPIEMR, Akurdi. Vaishnavi Patange is currently pursuing third year EnTC Engineering from Bharati Vidyapeeth's College of Engineering for Women, Pune. We both are currently studying data science and have built basic projects in the field. We both have basic knowledge of web development.

**Major Expertise of team members/individual:** Vaishnavi has an expertise in web development and is also interested in Data science and Machine Learning. Nijgururaj on the other hand has a major expertise in the field of data science and visualization.