**Project design - Phase - II**

**Stakeholder and Customer Requirements**

1. **End Users:**

Users should be able to access the web application easily and receive accurate predictions regarding cancer mortality and incidence rates based on their inputs.

1. **Healthcare Professionals:**

The system should provide reliable and interpretable predictions to assist healthcare professionals in decision-making and patient management.

1. **Data Privacy and Security:**

Ensuring the confidentiality and integrity of patient data by implementing appropriate security measures.

1. **Scalability and Performance:**

Designing the system to handle a potentially large number of users and providing efficient and responsive predictions.

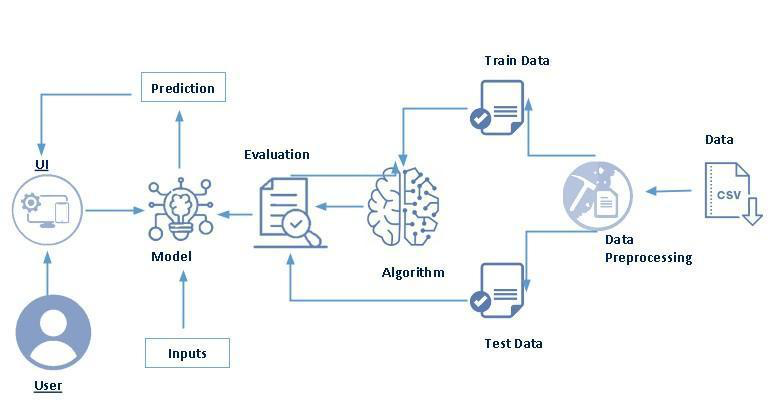
1. **User-Friendly Interface:**

Creating an intuitive and visually appealing web interface for ease of use.

**Requirement Analysis**

* **Dataset:** A dataset containing patient information and health conditions is required for training and testing the machine learning model.
* **Machine Learning Algorithms:** Selection and implementation of appropriate classification algorithms such as Logistic Regression, Decision Trees, Random Forest, or SVM.
* **Data Preprocessing:** Handling missing values, encoding categorical variables, and scaling numerical features if necessary.
* **Exploratory Data Analysis:** Generating visualisations and summary statistics to understand the dataset and identify patterns.
* **Model Evaluation:** Using suitable evaluation metrics to assess the performance of the trained model.
* **Flask Web Framework:** Developing a web application using Flask to create the user interface and handle user requests and responses.
* **Python:** Implementing the project using the Python programming language.

**Technical Architecture**

****

**Open Source Frameworks**

**Machine Learning Algorithms:** classification algorithms such as Logistic Regression, Decision Trees, Random Forest, or SVM.

**Data Preprocessing:** Scikit-learn: Scikit-learn is a widely used machine learning library in Python. It provides a range of classification algorithms, data preprocessing tools, and evaluation metrics, making it suitable for model development and evaluation.

**Exploratory Data Analysis:** NumPy is a fundamental library for scientific computing in Python. It provides efficient numerical operations and data structures, which are essential for working with arrays and manipulating data. Pandas is a data manipulation and analysis library. It offers powerful data structures and data analysis tools, making it useful for data preprocessing, exploratory data analysis, and feature engineering.

**Data Visualization :** Matplotlib and Seaborn: Matplotlib and Seaborn are libraries for creating visualizations in Python. They provide a wide range of plotting functions and customization options, enabling the generation of informative visualizations during the exploratory data analysis phase.