Project Initialization and Planning Phase

Date	14 July 2024
Team ID	739949
Project Title	Blood Donation Prediction
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) Report:

The proposal report aims to predict blood donation patterns using machine learning, enhancing efficiency and accuracy. It addresses inefficiencies in the current blood donation system, promising improved operations, reduced shortages, and increased donor satisfaction.

Project Overview	
Objective	To develop a machine learning model that can predict whether an individual is likely to donate blood based on historical donation data and demographic information.
Scope	• **Data Collection:** Gather and preprocess data from reliable sources. • **Model Development:** Train and evaluate various machine learning models. • **Feature Selection:** Identify key features that influence blood donation. • **Implementation:** Deploy the model in a user-friendly interface for predictions. • **Evaluation:** Monitor and refine the model based on performance metrics.
Problem Statement	
Description	Blood donation is a critical aspect of healthcare that ensures the availability of blood for transfusions and other medical procedures. However, predicting when individuals are likely to donate blood can be challenging due to various influencing factors. By leveraging machine learning techniques, we aim to create a predictive model that helps blood banks and healthcare organizations forecast blood donations more accurately.
Impact	• **Healthcare Efficiency:** Improve the management of blood donation drives.

	• **Availability:** Ensure a steady supply of blood, reducing shortages. • **Targeting:** More effectively target potential donors with reminders and campaigns. • **Cost-Effectiveness:** Reduce the costs associated with blood donation campaigns by focusing efforts on likely donors.
Proposed Solution	
Approach	• **Data Collection:** Collect historical blood donation data from sources like UCI Machine Learning Repository or local blood banks. • **Data Preprocessing:** Clean and preprocess the data to handle missing values, normalize features, and encode categorical variables. • **Exploratory Data Analysis (EDA):** Perform EDA to understand the data distribution and identify significant patterns. • **Model Training:** Train multiple machine learning models such as logistic regression, decision trees, random forests, and support vector machines. • **Model Evaluation:** Evaluate models using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to determine the best performing model. • **Deployment:** Develop a user-friendly interface where users can input data and get predictions about blood donation likelihood.
Key Features	• **User-Friendly Interface:** Simple and intuitive interface for entering data and viewing predictions. • **Accuracy:** High prediction accuracy using advanced machine learning techniques. • **Real-Time Predictions:** Provide instant predictions for new data inputs. • **Scalability:** Ability to handle large datasets and new incoming data for predictions.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	T4 GPU
Memory	RAM specifications	8 GB
Storage	Disk space for data, models,	1 TB

	and logs	
Software		
Frameworks	Python frameworks	e.g., Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, seaborn
Development Environment	IDE	Jupyter Notebook, pycharm
Data		
Data	Source, size, format	Kaggle dataset, 614, csv 1year dataset, 690, csv