Project Initialization and Planning Phase

Date	8 July 2024
Team ID	SWTID1720080033
Project Title	**Anemia Sense**: Leveraging Machine Learning For Precise Anemia Recognitions
Maximum Marks	3 Marks

Project Proposal

This project proposal outlines a solution to the labor-intensive process of detecting anemia in blood. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	This project's main goal is to create a machine learning model that, given certain parameters and medical data, can correctly diagnose anaemia in patients. The model aims to assist healthcare professionals in early diagnosis and management of anemia by providing a reliable and efficient tool for analysing patients.
Scope	• Collect blood test data from reliable sources such as Kaggle. • Pre-Processing and Data cleaning • Identify and select the most relevant features that contribute to anemia detection. • Explore and select suitable machine learning algorithms for classification • Test the model with a separate validation dataset to ensure it generalizes well to unseen data. • Develop a user-friendly web application for healthcare professionals to input patient data and receive anemia diagnosis results. • Maintain comprehensive documentation throughout the project, including data sources, preprocessing steps, model development, evaluation metrics, and deployment procedures. **Boundaries:** • The project will focus solely on anemia detection and will not extend to diagnosing other blood disorders. • The model will use a predefined set of blood parameters and

	will not incorporate additional tests or patient history. • The deployment will be limited to a software tool without hardware integration.
Problem Statement	
Description	Anemia is a common blood disorder characterized by a deficiency of red blood cells or hemoglobin, leading to reduced oxygen transport in the body. It affects millions of people worldwide, causing fatigue, weakness, and other health complications. Early detection and diagnosis of anemia are crucial for effective management and treatment. Traditional diagnostic methods often require multiple laboratory tests and can be time-consuming. With machine learning, there is an opportunity to develop an automated, efficient, and accurate tool to assist healthcare professionals in diagnosing anemia using readily available blood parameters.
Impact	• A reliable and accurate machine learning model for anemia detection. • Improve patient diagnosis and management of anemia. • Streamline the process of detecting anemia by reducing dependency on traditional medical practices. • An easy-to-use interface for healthcare professionals. • Lower overall healthcare costs associated with managing anemia-related complications.
Proposed Solution	
Approach	• Obtain data from Kaggle and other reliable sources • Handle imbalanced data to create a robust model which is not biased. • Visualize the data in order to find correlation among the features. • Evaluate various algorithms suitable for the task (Linear Regression, decision trees, random forests, Naïve Bayes, support vector machines, Gradient Boosting). • Assess and choose the best model in-order to provide the most accurate results. • Create a web application using flask and integrate the model.
Key Features	• Utilizes machine learning algorithms to accurately detect anemia based on comprehensive analysis of blood parameters.

	• Identifies and leverages crucial blood parameters (e.g., hemoglobin levels) specifically relevant to anemia detection. • Provides detailed reports and visualizations to users, communicating findings, methodology, and insights effectively. • Web Application allowing users to themselves to enter their data and check if they are anemic or not.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU specifications, number of cores	Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz, 2112 Mhz, 4 Core(s), 8 Logical Processor(s)
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	500 GB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	sklearn, pandas, numpy, pickle, matplotlib, seaborn
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Kaggle, 34KB, CSV	Kaggle dataset – 1421 dataentries