Final Project Report - AnemiaSense

1. Introduction

a. Project Overview

AnemiaSense is a machine learning-based web application developed to detect anemia using essential clinical parameters. It enables real-time prediction through a simple,user-friendly Flask web interface, making it a useful tool for early detection and management support.

b. Objectives

i. Build a predictive model for anemiadetection using machinelearning.
ii. Develop a web-based interface to collectpatient data and display results.
iii. Ensure the system is efficient, accurate, and easy to use for both medical and non-medical users.

2. Project Initialization and Planning Phase

a. Define Problem Statement

Anemia, a condition marked by low hemoglobin levels, often goes undetected until symptoms worsen. Manualdetection can be slow and inconsistent. This project aims to automateand improve the accuracy of anemia detection using clinical data and machine learning.

b. Project Proposal (ProposedSolution)

Develop an ML model that predicts whether a patient is anemic basedon health metrics. This model will be deployed through a Flask web application to allow instant predictions.

c. Initial Project Planning

i. Select relevant clinicaldataset
ii. Preprocess and clean data
iii. Train and evaluatemultiple models
iv. Deploy the best-performing model
v. Build a responsive web UI

3. Data Collection and Preprocessing Phase

a. Data Collection Plan and Raw Data SourcesIdentified

Dataset sourcedfrom a publicly available medicaldatabase containing featuressuch as age, hemoglobin, hematocrit, MCV, MCH, MCHC, RBC count, and serum ferritin.

b. Data Quality Report

i. Checked for missingvalues
ii. Handled outliers usingstatistical methods
iii. Normalized continuous valueswhere needed
iv. Encoded categorical labels

c. Data Exploration and Preprocessing

i. Used Pandas and Seaborn for EDA
ii. Found strong correlation between anemia and hemoglobin/RBC count
iii. Applied standard scalingand label encoding
iv. Split data into training and test sets (80:20)

4. Model Development Phase

a. Feature Selection Report

Selected features based on correlation and medical relevance:
-Age, Hemoglobin, Hematocrit, MCV,MCH, MCHC, RBC Count, Serum Ferritin

b. Model Selection Report

Trained and evaluated the following models:

i. Decision Tree Classifier
ii. Random Forest
iii. Logistic Regression

Decision Tree was selecteddue to its simplicity, speed, and strong accuracy.

c. Initial Model Training Code, Model Validation and Evaluation Report

i. Trained models using scikit-learn
ii. Evaluated using accuracy,precision, recall, and F1 score
iii. Decision Tree accuracy: -92%

5. Model Optimization and Tuning Phase

a. Hyperparameter Tuning Documentation

Performed GridSearchCV on:

i. Max depth
ii. Min samples split
iii. Criterion (gini vs entropy)

b. Performance MetricsComparison Report

Model	Accuracy	F1 Score	Precision	Recall
Decision Tree	92%	0.91	0.90	0.93
Random Forest	94%	0.93	0.92	0.94
Logistic Regression	88%	0.86	0.84	0.88

c. Final Model Selection Justification

Decision Tree chosen for its balance of performance, interpretability,and low latency, ideal for real-time predictions in the web app.

6. Results

a. Output Screenshots

(Screenshots of the runningweb app displaying prediction results can be inserted here)

7. Advantages & Disadvantages

Advantages

1. Fast, accurate prediction
2. Simple UI for non-technical users
3. Deployable on any systemwith Python

Disadvantages

4. Model accuracy depends ondata quality

5. Doesn't yet supportmulti-class classification or severity grading

8. Conclusion

AnemiaSense successfully demonstrates how machine learningcan aid in the early diagnosis of anemia. It bridges the gap between medical data and actionable insights through an intuitive interface.

9. Future Scope

1. Expand model to classify anemiaseverity
2. Integrate more clinicalfeatures (e.g., iron levels, vitamin B12)
3. Add multilingual support for broader accessibility

4. Deploy to cloud platforms (Render,Heroku, etc.)