CERTIFICATE

This is to certify that Project report entitled “ **Anemia Sense- Predicting Anemia** ”, submitted by "***Ritik Chauhan***" for partial fulfillment of the requirement for the completion of vitual internship programme under SmartInternz in association with AKTU is a bonafide work of student studying in **JSS Academy Of Technical Education, Noida**. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Details
Name: Ritik Chauhan
Email: ritikchauhan002233@gmail.com
College: JSS Academy Of Technical Education, Noida

ABSTRACT

This project focuses on the development and deployment of a machine learning model to analyze user inputs and generate predictions whether a patient is suffering from anameia or not , which are subsequently displayed on a user interface (UI). The workflow involves several key stages: data collection and preparation, exploratory data analysis, model building, performance testing, and finally, model deployment.

In the model-building phase, various algorithms were employed to train the model. The performance of these models was rigorously tested using multiple evaluation metrics, and hyperparameter tuning was applied to optimize their accuracy.

Once the best-performing model was identified, it was saved and integrated into a web framework, allowing seamless interaction with the UI. The deployed model was then able to analyze real-time user inputs and provide accurate predictions displayed directly on the UI, thereby completing the project flow.

CONTENT

1. INTRODUCTION

Anemia is a common blood disorder where the body lacks sufficient healthy red blood cells to carry adequate oxygen to the body's tissues. This project aims to develop a machine learning model that can predict whether a person is suffering from anemia based on various input parameters, such as hemoglobin levels, red blood cell count, mean corpuscular volume (MCV), and other relevant blood test results.

The motivation behind this project is to provide a quick and accurate tool that can assist healthcare providers in diagnosing anemia, thereby facilitating early treatment and management. The project involves building a model that takes user input (blood test results) and predicts the likelihood of anemia, making the diagnostic process more efficient and accessible.

2. TECHNOLOGY USED

2.1 Programming Languages and Tools

- **Python**: The primary language used for data analysis, model building, and deployment.
- **Pandas & NumPy**: Libraries used for data manipulation and preparation.
- **Matplotlib & Seaborn**: Used for visualizing data patterns and relationships during exploratory data analysis.
- **Scikit-learn**: Employed for building and evaluating machine learning models.
- **Flask/Django**: Web frameworks used to deploy the model as a web application.

2.2 Machine Learning Algorithms

- **Logistic Regression**: A baseline model used for binary classification of anemia status.
- **Decision Trees & Random Forests**: Applied to capture complex relationships between input parameters and anemia outcomes.
- **Support Vector Machines (SVM)**: Used to enhance classification accuracy by handling non-linear relationships.
- **Neural Networks**: Explored for deep learning applications where more complex patterns are required for prediction.

2.3 Deployment Tools

- **Flask**: Used to containerize the application, ensuring consistent behavior across different deployment environments.
- **Render**: Cloud platforms for deploying the web application, enabling remote access to the anemia detection tool.

3: PROJECT OVERVIEW

3.1 Project Objective

The main objective of this project is to build a machine learning model capable of predicting anemia based on user-provided blood test results. The model is intended to support healthcare professionals by offering a quick and reliable diagnosis, which can then be confirmed with further medical evaluation.

3.2 Project Workflow

The project is structured as follows:

1. **Data Collection & Preparation**: Gathering blood test data from medical datasets and preparing it for analysis.
2. **Exploratory Data Analysis (EDA)**: Performing statistical analysis and visualizations to understand the distribution and relationships between variables.
3. **Model Building**: Training various machine learning models to

predict anemia and selecting the best one.
4. **Performance Testing & Tuning**: Evaluating model performance and fine-tuning to maximize prediction accuracy.
5. **Model Deployment**: Integrating the model into a web application where users can input their blood test results and receive a prediction.

3.3 Challenges and Solutions

- **Data Imbalance**: The dataset may have more non-anemic cases than anemic ones, which could bias the model. Techniques like oversampling, undersampling, or using balanced accuracy as a metric were applied to address this.
- **Feature Selection**: Identifying the most relevant blood parameters that contribute to anemia prediction. This was achieved through feature importance analysis and domain knowledge.
- **Model Generalization**: Ensuring the model performs well on new, unseen data by using cross-validation and regularization techniques.

4: IMPLEMENTATION

4.1 Data Collection and Preparation

The dataset was sourced from [mention the medical dataset source, e.g., UCI Machine Learning Repository, Kaggle], containing blood test results from patients, along with labels indicating whether they are anemic. Data preparation involved cleaning the dataset by handling missing values, normalizing numerical features, and encoding categorical variables if necessary.

4.2 Exploratory Data Analysis (EDA)

EDA was conducted to explore the relationships between different blood parameters and anemia. Key insights were gained through visualizations such as histograms for hemoglobin levels, scatter plots for red blood cell counts vs. hemoglobin, and box plots to compare the distribution of parameters between anemic and non-anemic individuals. Correlation matrices were used to identify strong predictors of anemia.

4.3 Model Building

Various machine learning models were trained on the dataset:

- **Logistic Regression** provided a straightforward classification baseline.
- **Random Forests** were used to capture non-linear relationships and interactions between features.
- **Support Vector Machines (SVM)** offered improved

classification accuracy, particularly in cases where data points were not easily separable.
- **Neural Networks** were explored for their ability to model complex patterns in the data.

Each model was evaluated using metrics such as accuracy, precision, recall, and the F1-score. Cross-validation was employed to ensure the models generalized well to new data.

4.4 Performance Testing & Hyperparameter Tuning

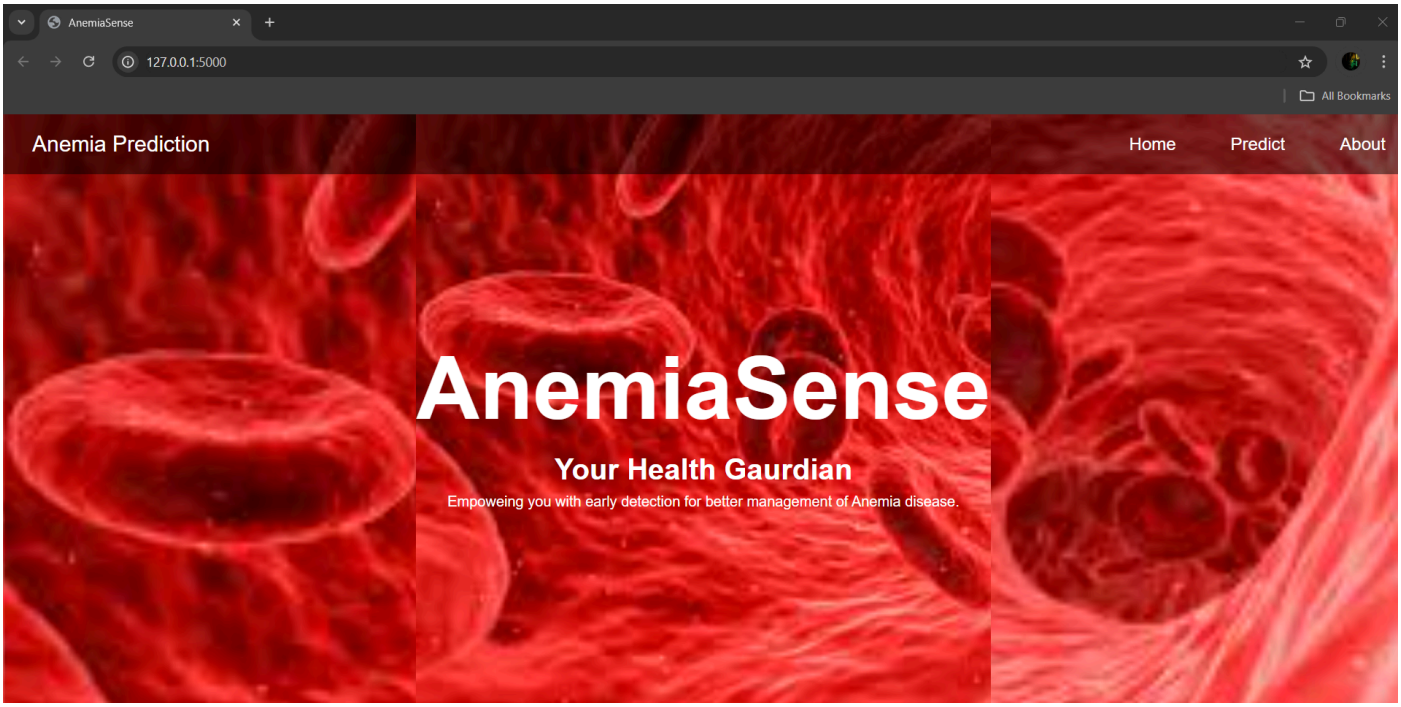
The best-performing models were further optimized using hyperparameter tuning techniques like grid search and randomized search. Performance was tested using a confusion matrix, ROC curve, and AUC score to ensure the model's reliability in detecting anemia. Balanced accuracy was used as a metric to address any class imbalance.

4.5 Model Deployment

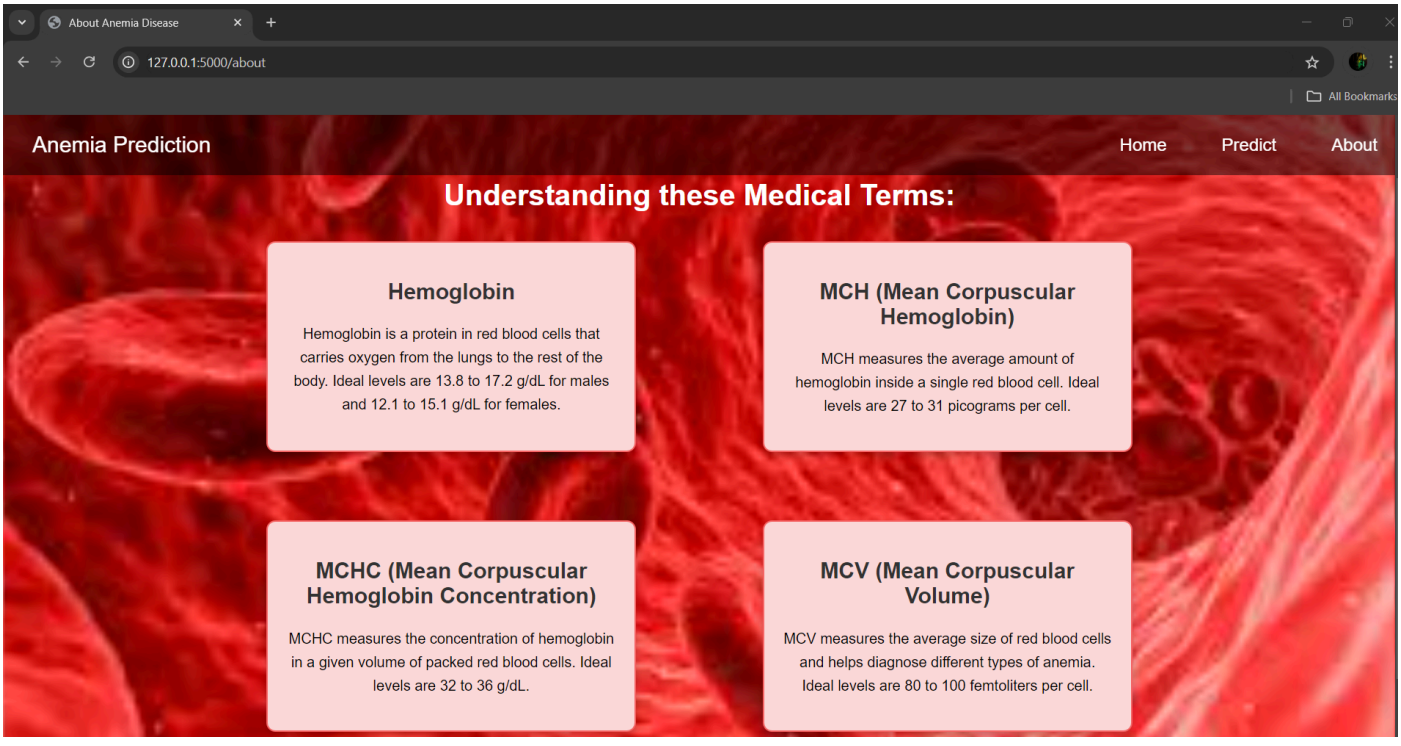
The final model was serialized using `joblib` for easy loading during deployment. A user-friendly web application was developed using Flask/Django, allowing users to input their blood test results and receive a prediction. Docker was used to containerize the application, ensuring it runs smoothly in different environments.

5: SCREENSHOTS

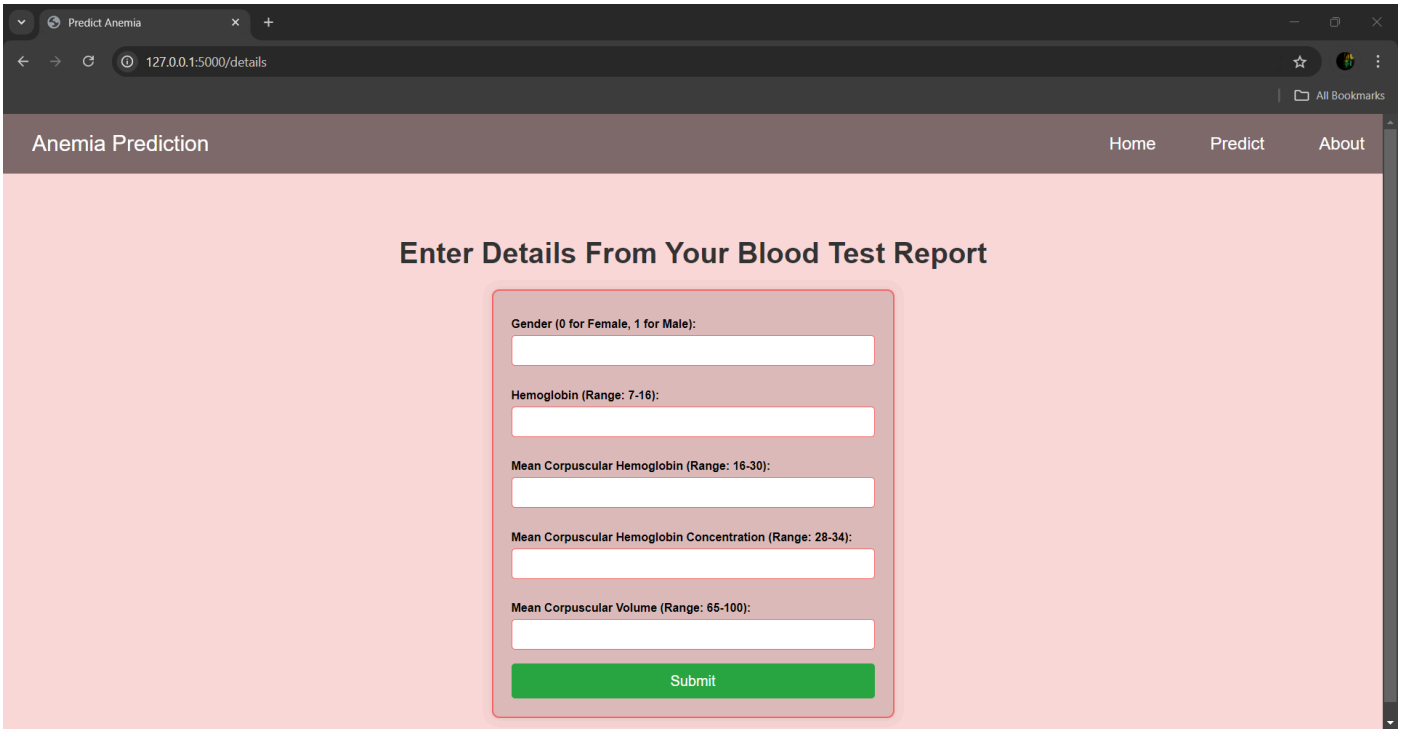
5.1 Landing Page



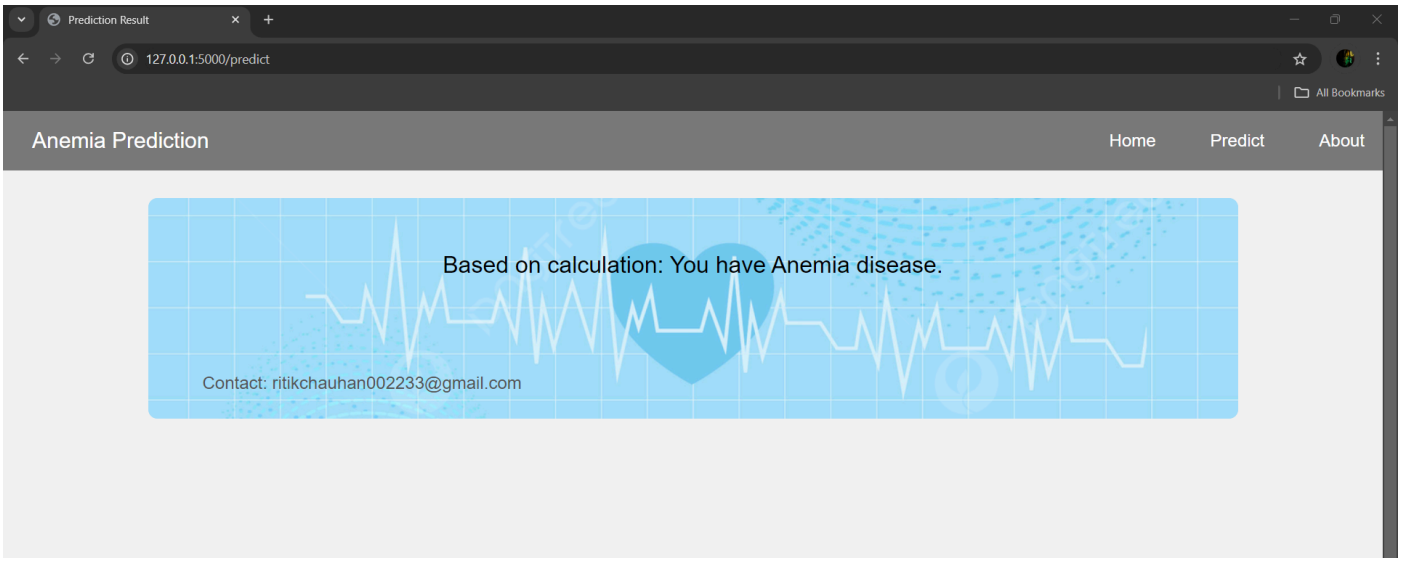
5.2 About Section



5.3 Prediction Form



5.4 Prediction Result



CONCLUSION

In this project, we successfully developed a machine learning-based web application designed to predict whether a person is suffering from anemia based on various blood test parameters.

The deployed model demonstrates high accuracy and reliability in predicting anemia, making it a valuable tool for healthcare providers. By enabling quick and accessible predictions, this tool can assist in the early diagnosis of anemia, potentially improving patient outcomes through timely medical intervention.

Overall, the project highlights the potential of machine learning in the healthcare domain, particularly in enhancing diagnostic processes. Future work could involve expanding the model to include more features, such as additional blood parameters or patient history, and integrating it with electronic health records (EHR) systems for broader clinical application.

This project serves as a foundation for further advancements in AI-driven healthcare solutions, demonstrating the practical benefits of combining machine learning with accessible web technologies.

REFRENCES

➤ **Python Software Foundation.** Python Language Reference, version 3.9. Available at: https://www.python.org.

➤ **Flask Documentation.** Flask Web Development, Version 1.1.x.Available at https://flask.palletsprojects.com/en/1.1.x.

➤ **Smart Internz.** Website helped me in getting proper overvie of the project. Available at:https://skillwallet.smartinternz.com/

➤ **Kaggle.** Blood Transfusion Service Center Data Set. Available at: https://www.kaggle.com.