# By:

Akkaldevi Rakshith Aytha Bhoomika Kanumuri Ajith Varma Varad Wankhade

# Final Project Report Template

## Introduction :

* 1. Project Overview:

Anemiasense leverages machine learning algorithms to provide precise recognition and management of anemia, a condition characterized by a deficiency of red blood cells or hemoglobin. Here are three general scenarios illustrating its use case:

Scenario 1:Early Detection and Diagnosis:

Anemiasense utilizes machine learning models trained on vast datasets of blood parameters and patient profiles to detect early signs of anemia. By analyzing key indicators such as hemoglobin levels, red blood cell counts, and other relevant biomarkers, the system can flag potential cases for further investigation by healthcare professionals. Early detection enables timely interventions and treatment plans, improving patient outcomes.

Scenario 2: Personalized Treatment Plans

Machine learning algorithms in Anemiasense can analyze diverse patient data, including genetic factors, lifestyle habits, and medical history, to generate personalized treatment plans. By considering individual variations and responses to different treatments, the system helps healthcare providers tailor interventions for optimal results. This personalized approach enhances the effectiveness of anemia management and reduces the risk of complications.

Scenario 3:Remote Monitoring and Follow-Up

Anemiasense supports remote monitoring of patients with anemia through wearable devices or digital health platforms. Machine learning algorithms continuously analyze real-time data such as hemoglobin levels, activity levels, and medication adherence to provide insights to both patients and healthcare providers. This remote monitoring capability facilitates proactive management, enables timely adjustments to treatment regimens, and reduces the need for frequent in-person visits, particularly beneficial for patients in rural or underserved areas.

* 1. **Objectives:**

**Analyse the data and Find the best suitable model.Build an website using flask for the customer to upload their test results to recognise anemia.**

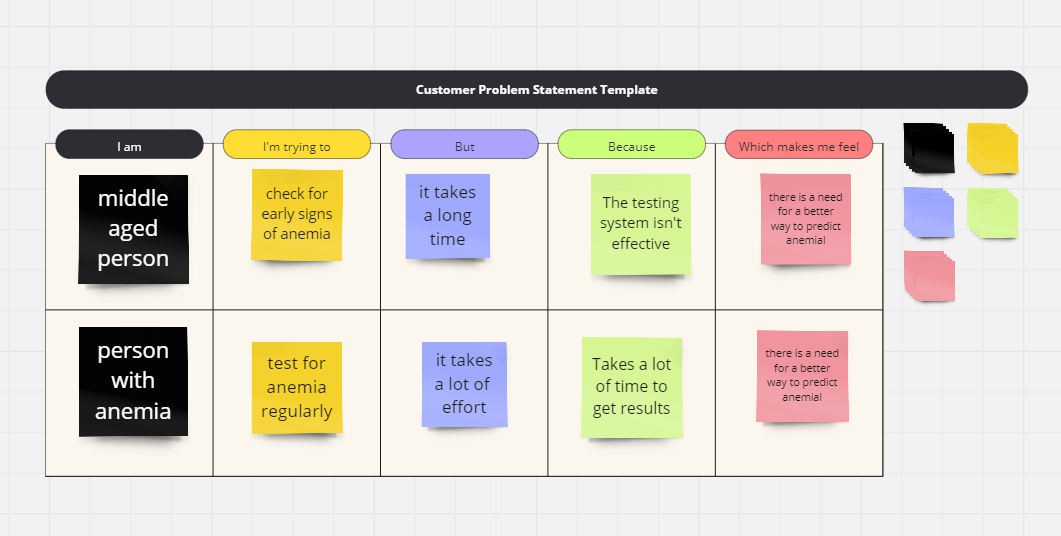
* 1. **Project Initialization and Planning Phase**

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Name | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 3 Marks |

**Define Problem Statements (Customer Problem Statement Template):**

**Anemiasense leverages machine learning algorithms to provide precise recognition and management of anemia, a condition characterized by a deficiency of red blood cells or hemoglobin. Here are three general scenarios illustrating its use case:**

**Scenario 1: Early Detection and Diagnosis Scenario 2: Personalized Treatment Plans Scenario 3:Remote Monitoring and Follow-Up**



## Project Initialization and Planning Phase

|  |  |
| --- | --- |
| Date | 10 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 3 Marks |

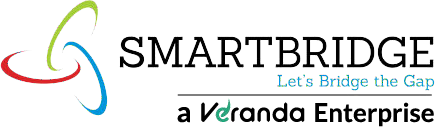
**Project Proposal (Proposed Solution) template**

This project proposal outlines a solution to address a specific problem. 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 | Leveraging Machine Learning for Precise Anemia Recognitions |
| Scope | * Data Analysis * Pattern Recognition * Personalized Medicine |
| **Problem Statement** | |
| Description | **Scenario 1:** Early Detection and Diagnosis **Scenario 2:** Personalized Treatment Plans **Scenario 3:** Remote Monitoring and Follow-Up |
| Impact | Helps in precise recognition of Anemia |
| **Proposed Solution** | |
| Approach | Train and test the data using different types of models and obtain the most efficient model and use it for Anemia Recognition. |
| Key Features | Use of Random forest model, Logistic Regression Model, Decision Tree Model,etc. |

**Resource Requirements**

|  |  |  |
| --- | --- | --- |
| **Resource Type** | **Description** | **Specification/Allocation** |
| **Hardware** | | |
| Computing Resources | Intel i3, 2 Cores | Minimum Intel Iris Xe |
| Memory | RAM specifications | minimum 8 GB |
| Storage | 40 mb | Minimum 256 GB SSD |
| **Software** | | |
| Frameworks | Python frameworks | e.g., Flask |
| Libraries | Additional libraries | e.g., scikit-learn, pandas, NumPy, matplotlib, seaborn |
| Development Environment | IDE, version control | e.g., Jupyter Notebook, Git |
| **Data** | | |
| Data | Source, size, format | e.g., Kaggle dataset, 10,000 images |

## Initial Project Planning Template

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Name | Anemia Sense: Leveraging Machine Learning  For Precise Anemia Recognitions |
| Maximum Marks | 4 Marks |

**Product Backlog, Sprint Schedule, and Estimation (4 Marks)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Functional**  **Requirement (Epic)** | **User**  **Story Number** | **User Story / Task** | **Story Points** | **Priority** | **Team Members** | **Sprint Start Date** | **Sprint End**  **Date (Planned)** |
| Sprint-1 | Data  Collection & Preparation | USN-1 | Collected Data and check for errors. | 2 | Medium | Aytha Bhoomika | 10th July | 10th July |
| Sprint-1 | Exploratory Data Analysis | USN-2 | Did Univariate,Bivariate and  Multivariate Analysis. | 1 | Medium | Akkaldevi Rakshith | 10th July | 10th July |
| Sprint-2 | Model Building | USN-3 | Tested the data using many models. | 3 | Medium | Akkaldevi Rakshith, Ajith Varma,  Varad Wankadhe, Aytha Bhoomika | 10th July | 10th July |
| Sprint-2 | Performance Testing & Hyperparameter  Tuning | USN-4 | Checking the performance and choosing the best model. | 2 | Medium | Kanumuri Ajith Varma | 11th July | 11th July |
| Sprint-3 | Model Deployment | USN-5 | Build the Web-Pages and  deploy. | 2 | Medium | Varad Wankadhe | 11th July | 11th July |

## Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | Anemiasense leverages machine learning algorithms to provide precise recognition and management of anemia, a condition characterized by a deficiency of red blood cells or hemoglobin. |
| Data Collection Plan | The dataset was taken from the SmartInternz platfrom. |
| Raw Data Sources Identified | The data set comprised values of  **Gender** **Hemoglobin** **MCH** **MCHC** **MCV** **Result**  Of each patient. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
|  | Contains all the |  |  |  |  |
| Dataset 1:  Anemia.csv | primary readings  required for | Link of Dataset 1 | CSV | 33.8kb | Public |
|  | detecting anemia. |  |  |  |  |

## Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 2 Marks |

**Data Quality Report Template**

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Dataset | Mention the issues faced in the selected dataset. | Low/ Moderate  / High | Give the solution for that issue technically. |
| Anemia.csv | No issue | Nil | Nill |

## Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | Shape: (1421, 6) |
| Univariate Analysis |  |

|  |  |
| --- | --- |
| Bivariate Analysis |  |
| Multivariate Analysis |  |
| Outliers and Anomalies | Identification and treatment of outliers. |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |

|  |  |
| --- | --- |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering |  |
| Save Processed Data | Code to save the cleaned and processed data for future use. |

## Model Development Phase Template

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 5 Marks |

**Feature Selection Report Template**

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Description** | **Selected (Yes/No)** | **Reasoning** |
|  | The average |  |  |
|  | volume of a red |  |  |
| Mean Corpuscular Volume  (MCV) | blood cell, measured in femtoliters (fL).  This helps classify | Yes | MCV measures the average volume of red blood cells and is a critical parameter for classifying anemia types. |
|  | the type of |  |  |
|  | anemia |  |  |
|  | The average |  |  |
| Mean | amount of |  | MCH measures the average amount of |
| Corpuscular  Hemoglobin | hemoglobin per  red blood cell, | Yes | hemoglobin per red blood cell and helps  determine the hemoglobin content within |
| (MCH) | measured in |  | each cell. |
|  | picograms (pg). It |  |  |
|  | helps in |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | understanding the hemoglobin content per cell. |  |  |
|  | The average |  |  |
|  | concentration of |  |  |
|  | hemoglobin in a |  |  |
| Mean Corpuscular Hemoglobin Concentrati on | given volume of packed red blood cells, measured in grams per deciliter (g/dL). It indicates  the hemoglobin | **Yes** | MCHC measures the average concentration of hemoglobin in a given volume of packed red blood cells, providing insight into the hemoglobin saturation within cells. |
|  | concentration |  |  |
|  | within red blood |  |  |
|  | cells. |  |  |

## Model Development Phase Template

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | xxxxxx |
| Project Title | xxxxxx |
| Maximum Marks | 6 Marks |

**Model Selection Report**

In the forthcoming Model Selection Report, various models will be outlined, detailing their descriptions, hyperparameters, and performance metrics, including Accuracy or F1 Score. This comprehensive report will provide insights into the chosen models and their effectiveness.

**Model Selection Report:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Description** | **Hyperparameters** | **Performance Metric (e.g., Accuracy, F1 Score)** |
| Model 1: Random Forest Classifier | Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees | * n\_estimators * max\_depth | * **Accuracy**: Overall accuracy of predictions. * **Precision**: Measures the proportion of true positive predictions out of all positive predictions. * **F1** **Score**: Harmonic mean of precision and recall, balancing both metrics. |

|  |  |  |  |
| --- | --- | --- | --- |
|  | A Decision Tree is |  | * **Accuracy**: Overall accuracy of predictions. * **Precision**: Measures the proportion of true positive predictions out of all positive predictions. * **F1** **Score**: Harmonic mean of precision and recall, balancing both metrics. |
|  | a supervised |  |
|  | learning algorithm |  |
|  | used for both |  |
| Model 2: | classification and |  |
| Decision Tree  Model | regression tasks.  It splits the data into subsets | * Criterion * splitter |
|  | based on the |  |
|  | most significant |  |
|  | attribute, creating |  |
|  | a tree structure |  |
|  | Gaussian Naive |  | * **Accuracy**: The proportion of correctly predicted instances out of the total instances. * **Precision**: The proportion of true positive predictions out of all positive predictions. * **F1** **Score**: The harmonic mean of precision and recall, providing a balance between the two |
| **Model 3:** | Bayes is a variant  of the Naive |  |
| Guassian Naive | Bayes algorithm that assumes the | * var\_smoothing |
| Bayes | features follow a  normal (Gaussian) |  |
|  | distribution. |  |

## Model Development Phase Template

|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognitions |
| Maximum Marks | 4 Marks |

**Initial Model Training Code, Model Validation and Evaluation Report**

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

**Initial Model Training Code:**

Paste the screenshot of the model training code

**Model Validation and Evaluation Report:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Classification Report** | **Accuracy** | **Confusion Matrix** |
| Model 1 |  | Accuracy Value : 1.0 | Screenshot of the confusion matrix: |
| Model 2 |  | Accuracy Value : 1.0 | Screenshot of the confusion matrix : |

|  |  |  |  |
| --- | --- | --- | --- |
| Model 3 |  | Accuracy Value: 0.97  1774 | Screenshot od confusion matrix: |

## Model Optimization and Tuning Phase Template

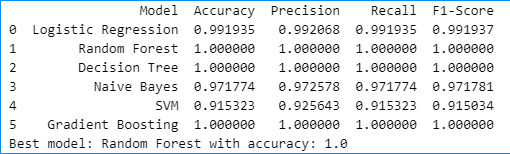
|  |  |
| --- | --- |
| Date | 11 July 2024 |
| Team ID | SWTID1720099206 |
| Project Title | Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions |
| Maximum Marks | 10 Marks |

* 1. **Model Optimization and Tuning Phase**

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

**Hyperparameter Tuning Documentation (6 Marks):**

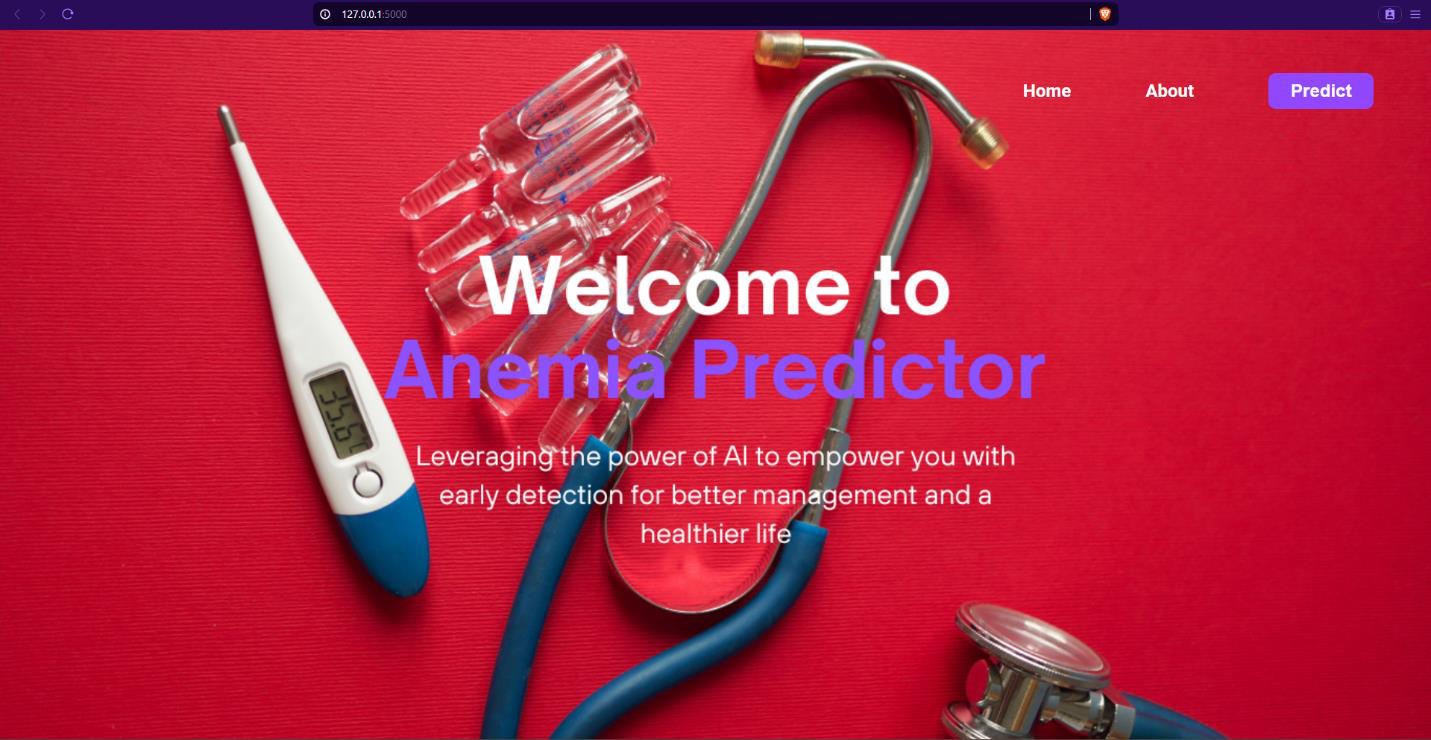
* + - **Hyperparameter Tuning was not performed as 100% accuracy was achieved.**
  1. **Performance Metrics Comparison Report (2 Marks):**

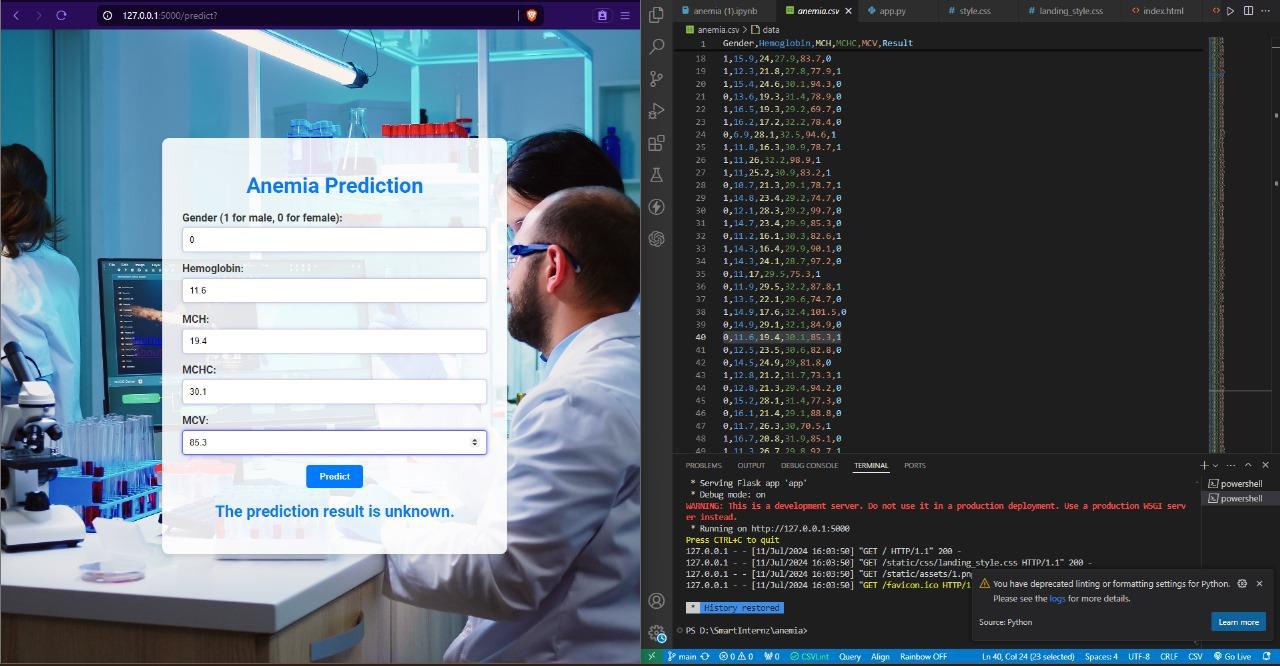


* 1. **Final Model Selection Justification (2 Marks):**

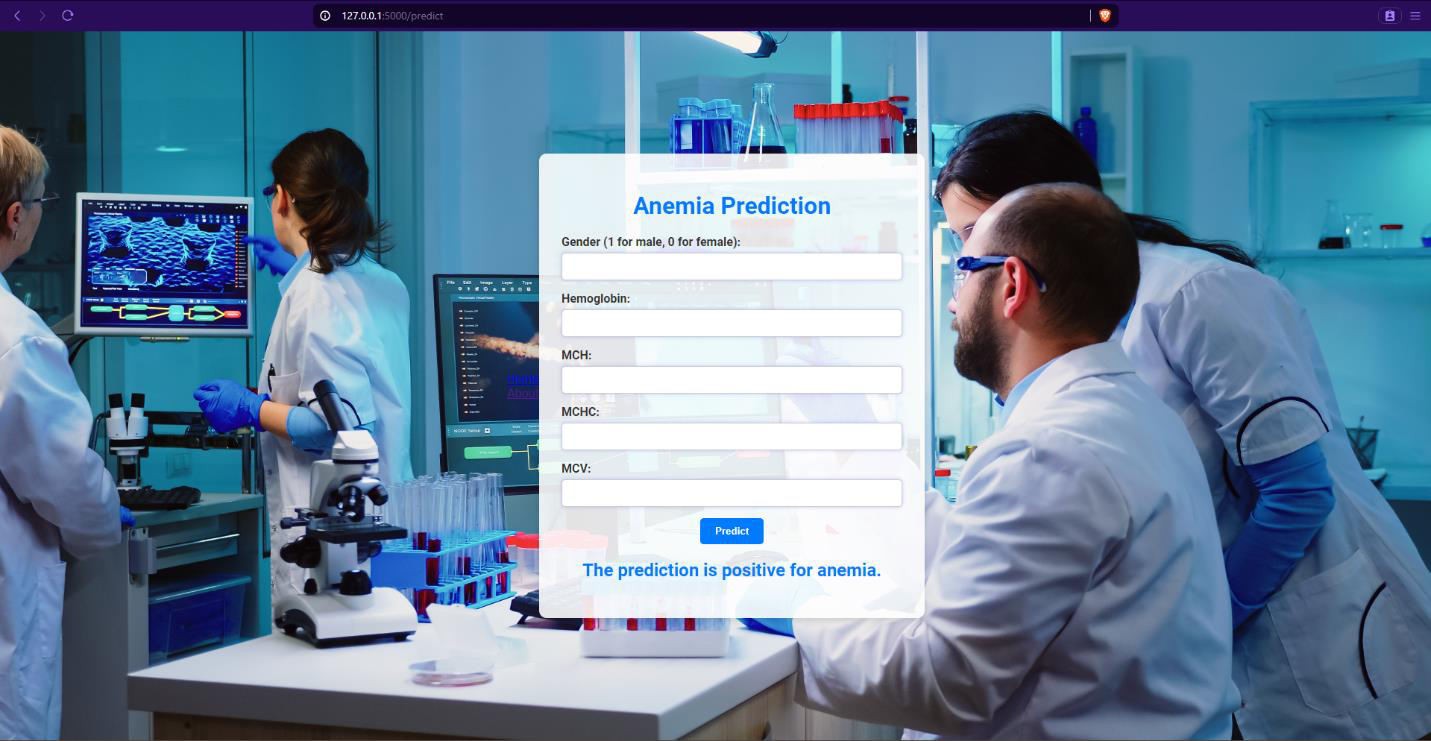
|  |  |
| --- | --- |
| **Final Model** | **Reasoning** |
|  | * 100% accuracy |
|  | * 100% Precision |
| Random Forest | * 100% Recall |
| Model | * F1 -score : 1 |

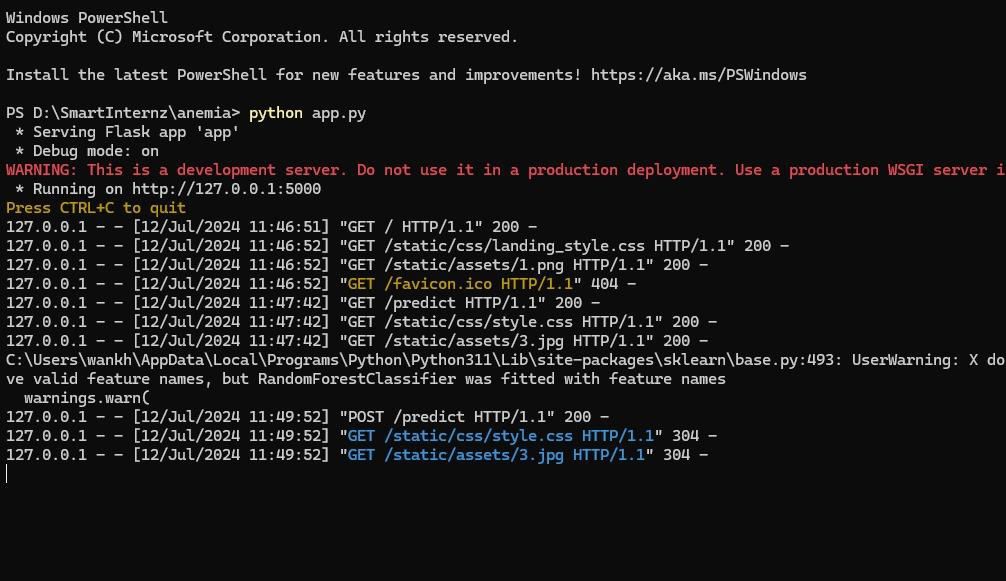
1. **Output Screenshots:**





**Result:**



**7: Advantages & Disadvantages:**

# Advantages

* 1. **Improved Accuracy**:
     + **Pattern** **Recognition**: Machine learning (ML) algorithms can identify subtle patterns in data that may not be visible to human experts, leading to more accurate anemia diagnoses.
     + **Consistency**: ML models provide consistent results, reducing human error and variability in diagnosis.
  2. **Early Detection**:
     + **Proactive** **Interventions**: ML models can detect anemia at an earlier stage, allowing for timely interventions and improved patient outcomes.
  3. **Efficiency**:
     + **Automation**: Automating the analysis of blood tests and other diagnostic procedures can save time for healthcare professionals, allowing them to focus on more complex cases.
     + **High** **Throughput**: ML can analyze large volumes of data quickly, which is beneficial in high-demand healthcare settings.
  4. **Personalization**:
     + **Tailored** **Treatments**: ML can help in developing personalized treatment plans by analyzing individual patient data, leading to more effective management of anemia.
  5. **Cost-Effective**:
     + **Reduced** **Costs**: By automating routine diagnostic tasks, ML can help reduce the costs associated with manual labor and potentially lower healthcare expenses.
  6. **Continuous Improvement**:
     + **Learning** **and** **Adaptation**: ML models can continuously learn from new data, improving their accuracy and reliability over time.

# Disadvantages

1. **Data Quality and Availability**:
   * **Data** **Dependency**: The accuracy of ML models depends heavily on the quality and quantity of data available. Poor data quality can lead to inaccurate predictions.
   * **Bias**: If the training data is biased, the model may produce biased results, leading to disparities in healthcare.
2. **Complexity**:
   * **Interpretability**: ML models, especially deep learning models, can be complex and difficult to interpret, making it hard for clinicians to understand the basis of a diagnosis.
   * **Integration**: Integrating ML solutions into existing healthcare systems can be challenging and may require significant changes to workflows.
3. **Regulation and Compliance**:
   * **Regulatory** **Hurdles**: The use of ML in healthcare is subject to strict regulations, which can slow down the development and deployment of new models.
   * **Privacy** **Concerns**: Handling sensitive patient data requires strict adherence to privacy laws and regulations, which can be complex and costly.
4. **Resource Intensive**:
   * **Computational** **Resources**: Training and deploying ML models can require significant computational resources, which may be a barrier for some healthcare providers.
   * **Expertise**: Developing and maintaining ML models requires specialized expertise that may not be readily available in all healthcare settings.
5. **Ethical and Legal Issues**:
   * **Accountability**: Determining accountability for errors made by ML systems can be challenging, raising ethical and legal concerns.
   * **Transparency**: Ensuring transparency in how ML models make decisions is crucial for maintaining trust among patients and healthcare providers.

# Conclusion

While the application of machine learning for anemia recognition has the potential to revolutionize diagnostics with enhanced accuracy, efficiency, and personalization, it also presents challenges related to data quality, model interpretability, regulatory compliance, and resource requirements. Careful consideration and mitigation of these disadvantages are essential to fully leverage the benefits of ML in healthcare.

## Future Scope:

### Advanced Diagnostic Tools

* **Integration** **with** **Wearables**: Development of non-invasive, wearable devices that continuously monitor vital signs and blood parameters, providing real-time anemia detection and monitoring.
* **Multimodal** **Data** **Integration**: Combining data from various sources, such as electronic health records (EHRs), genetic information, and imaging, to create comprehensive diagnostic models.

### Personalized Medicine

* **Tailored** **Treatment** **Plans**: Using ML to create personalized treatment plans based on a patient’s unique genetic makeup, lifestyle, and response to previous treatments.
* **Predictive** **Analytics**: Implementing models that predict the risk of anemia based on patient history and lifestyle, enabling proactive preventive measures.

### Global Health Applications

* **Low-Cost** **Solutions**: Developing cost-effective, portable diagnostic tools for use in resource- limited settings, improving access to accurate anemia diagnosis worldwide.
* **Telemedicine**: Integrating ML-based diagnostic tools with telemedicine platforms to provide remote consultation and management of anemia.

### Research and Development

* **Drug** **Discovery**: Utilizing ML to accelerate the discovery and development of new treatments for anemia by identifying potential therapeutic targets and predicting drug efficacy.
* **Clinical** **Trials**: Enhancing the design and analysis of clinical trials through better patient stratification and real-time monitoring of patient responses.

### Enhanced Algorithms and Techniques

* **Explainable** **AI**: Developing models that provide clear, interpretable insights into how they make decisions, increasing trust among healthcare providers and patients.
* **Continuous** **Learning** **Systems**: Implementing systems that continuously learn from new data, improving accuracy and adaptability over time.

## Appendix:

**Github Link:** [**https://github.com/rakshith-rk20/anemia/tree/main**](https://github.com/rakshith-rk20/anemia/tree/main)

**Demo Link:** [**https://drive.google.com/file/d/1EWVi5l5XcX2AxRH46k97GdQzaL64nfwK/view?usp=drive\_link**](https://drive.google.com/file/d/1EWVi5l5XcX2AxRH46k97GdQzaL64nfwK/view?usp=drive_link)