Anemia Sense Leveraging-Machine Learning For-Precise Anemia Recognition

1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

Problem Statement: Anemia is a widespread health condition that often goes undiagnosed due to limited access to diagnostic tools and delayed clinical evaluations. Current methods for anemia detection can be invasive, costly, and inaccessible in remote or underserved regions. Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognitions aims to develop a non-invasive, intelligent diagnostic system utilizing machine learning algorithms to accurately identify and classify anemia. This solution seeks to enhance early detection, reduce diagnostic costs, and provide scalable, real-time support for healthcare providers and patients worldwide.

Anemia Problem Statement Report: Click here

Activity 2: Project Proposal (Proposed Solution)

The proposal aims to improve anemia detection using machine learning by analyzing hematological parameters. The system enhances diagnostic efficiency, reduces manual errors, and provides timely health assessments. Key features include a data-driven classification model based on features like Hemoglobin, MCH, MCHC, and MCV.

Anemia Project Proposal Report: Click Here

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a loan approval system. It encompasses setting timelines, allocating resources, and determining the overall project strategy. During this phase, the team establishes a clear understanding of the dataset, formulates goals for analysis, and plans the workflow for data processing.

Anemia Project Planning Report: Click Here

2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant loan

application data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

The dataset for "Anemia Sense: Leveraging Machine Learning For Precise Anemia" is sourced from Kaggle. The dataset was obtained from a structured CSV file containing hematological parameters like Hemoglobin, MCH, MCHC, MCV, and a binary anemia result.

Anemia Data Collection Report: Click Here

2: Data Quality Report

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.

Anemia Data Quality Report: Click Here

3: Data Exploration and Preprocessing

Data Exploration involves analyzing the anemia sense dataset to understand patterns, distributions, and outliers. Preprocessing includes handling missing values, scaling, and encoding categorical variables. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in the anemia detection project.

Anemia Data Exploration and Preprocessing Report: Click Here

3: Model Development Phase

The Model Development Phase entails crafting a predictive model for anemia detection. It encompasses strategic feature selection, evaluating and selecting models (Random Forest, Decision Tree, KNN, XGB), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the lending process.

1: Feature Selection Report

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.

Anemia Feature Selection Report: Click Here

2: Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Decision Tree, Logistic Regression, and Gradient Boosting models. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

Anemia Model Selection Report: **Click Here**

3: Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code employs selected algorithms on the anemia dataset, setting the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness in predicting loan outcomes.

Anemia Model Development Phase Template: **Click Here**

4: 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.

1: Hyperparameter Tuning Documentation

The Random Forest model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the Random Forest model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

3: Final Model Selection Justification

The Random Forest model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

Anemia Model Optimization and Tuning Phase Report: **Click Here**

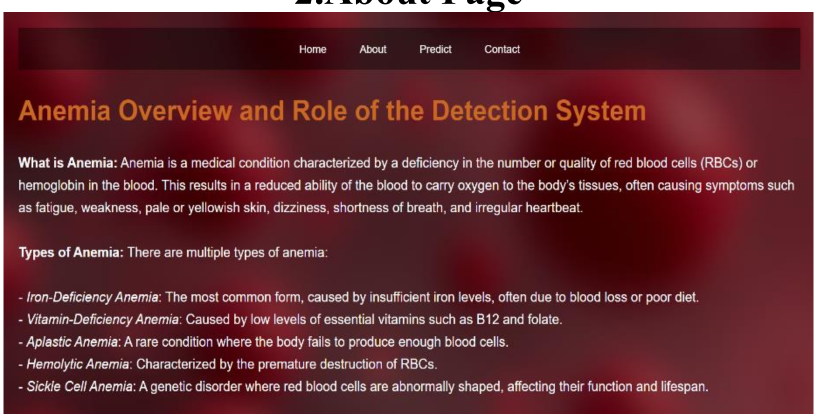
5: Project Files Submission and Documentation

For the documentation, Kindly refer to the link. **Click Here**

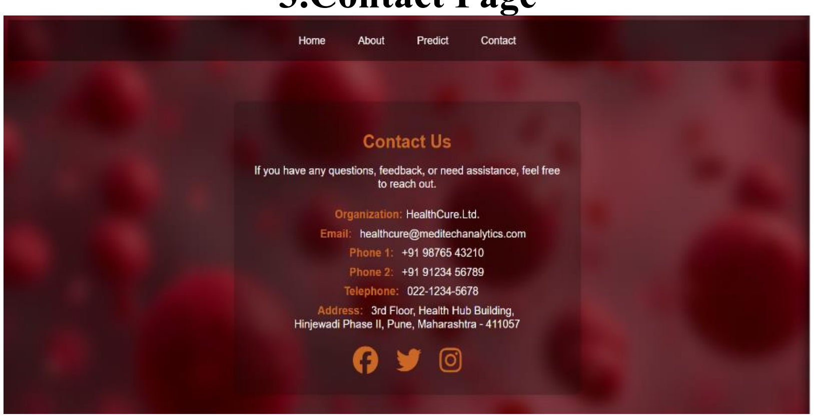
6. Results

1.Homepage

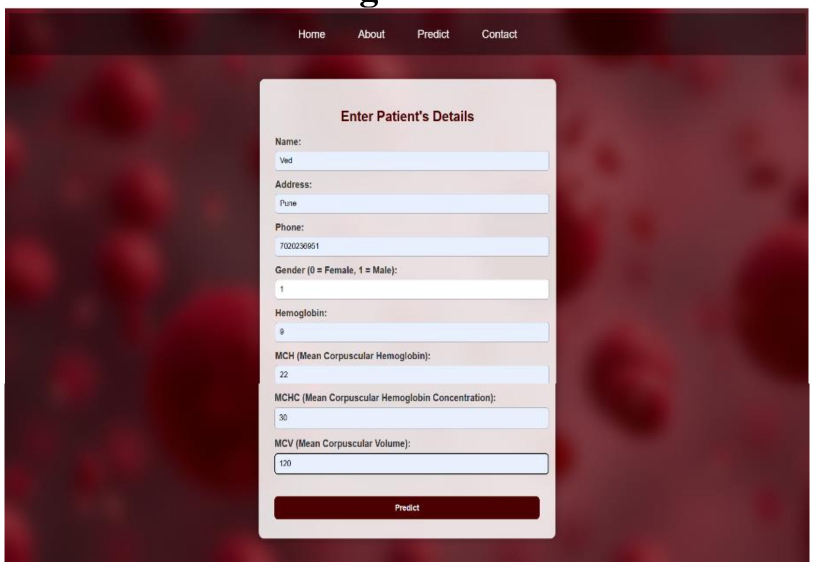
2.About Page



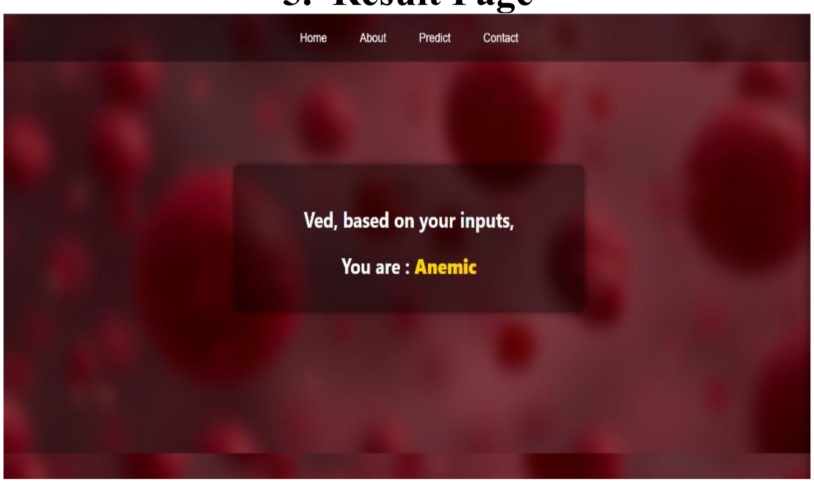
3.Contact Page



4. Form Page for Prediction



5. Result Page



6. Console Output

```
📊 Dataset Overview
Shape: (1421, 6)

Data Types:
 Gender          int64
Hemoglobin    float64
MCH           float64
MCHC          float64
MCV           float64
Result          int64
dtype: object

First 5 Rows:
    Gender  Hemoglobin   MCH  MCHC   MCV  Result
0        1        14.9  22.7  29.1  83.7       0
1        0        15.9  25.4  28.3  72.0       0
2        0         9.0  21.5  29.6  71.2       1
3        0        14.9  16.0  31.4  87.5       0
4        1        14.7  22.0  28.2  99.5       0

🧼 Null Value Check
No missing values found.

📈 Summary Statistics:
            Gender   Hemoglobin          MCH         MCHC          MCV       Result
count  1421.000000  1421.000000  1421.000000  1421.000000  1421.000000  1421.000000
mean      0.520760    13.412738    22.905630    30.251232    85.523786     0.436312
std       0.499745     1.974546     3.969375     1.400898     9.636701     0.496102
min       0.000000     6.600000    16.000000    27.800000    69.400000     0.000000
25%       0.000000    11.700000    19.400000    29.000000    77.300000     0.000000
50%       1.000000    13.200000    22.700000    30.400000    85.300000     0.000000
75%       1.000000    15.000000    26.200000    31.400000    94.200000     1.000000
max       1.000000    16.900000    30.000000    32.500000   101.600000     1.000000
```

```
Logistic Regression
Accuracy: 98.95%
Confusion Matrix:
[[154   3]
 [  0 128]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.98      0.99       157
           1       0.98      1.00      0.99       128

    accuracy                           0.99       285
   macro avg       0.99      0.99      0.99       285
weighted avg       0.99      0.99      0.99       285

--------------------------------------------------

Random Forest
Accuracy: 100.00%
Confusion Matrix:
[[157   0]
 [  0 128]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       157
           1       1.00      1.00      1.00       128

    accuracy                           1.00       285
   macro avg       1.00      1.00      1.00       285
weighted avg       1.00      1.00      1.00       285

--------------------------------------------------

Decision Tree
Accuracy: 100.00%
Confusion Matrix:
[[157   0]
 [  0 128]]
```

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       157
           1       1.00      1.00      1.00       128

    accuracy                           1.00       285
   macro avg       1.00      1.00      1.00       285
weighted avg       1.00      1.00      1.00       285

--------------------------------------------------

Naive Bayes
Accuracy: 95.09%
Confusion Matrix:
[[150   7]
 [  7 121]]
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.96      0.96       157
           1       0.95      0.95      0.95       128

    accuracy                           0.95       285
   macro avg       0.95      0.95      0.95       285
weighted avg       0.95      0.95      0.95       285
```

```
SVM
Accuracy: 98.25%
Confusion Matrix:
[[152   5]
 [  0 128]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.98       157
           1       0.96      1.00      0.98       128

    accuracy                           0.98       285
   macro avg       0.98      0.98      0.98       285
weighted avg       0.98      0.98      0.98       285

--------------------------------------------------

Gradient Boosting
Accuracy: 100.00%
Confusion Matrix:
[[157   0]
 [  0 128]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       157
           1       1.00      1.00      1.00       128

    accuracy                           1.00       285
   macro avg       1.00      1.00      1.00       285
weighted avg       1.00      1.00      1.00       285
```

```
------------------------------------------------

Lasso (L1)
Accuracy: 99.30%
Confusion Matrix:
[[155   2]
 [  0 128]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.99      0.99       157
           1       0.98      1.00      0.99       128

    accuracy                           0.99       285
   macro avg       0.99      0.99      0.99       285
weighted avg       0.99      0.99      0.99       285

------------------------------------------------

✅ Best Model Selected: Random Forest
```

7. Advantages & Disadvantages

Advantages:

- Automates anemia detection with high accuracy.
- Saves diagnostic time and cost.
- Supports multiple machine learning models for better comparison.
- Scalable and easily integrable with healthcare systems.
- Provides a user-friendly web interface for predictions.

Disadvantages:

- Depends on quality and size of input dataset.
- May misclassify borderline cases without clinical support.
- Requires proper feature scaling and preprocessing.
- Some algorithms are less interpretable (e.g., SVM, GBM).

8. Conclusion

This project successfully implements an anemia detection system using various machine learning algorithms. By preprocessing the dataset and evaluating multiple models, the system ensures robust predictions. Lasso regression was selected as the best-performing model based on evaluation metrics, making it ideal for deployment in real-world healthcare environments.

9. Future Scope

- Integration with real-time blood test data.
- Expansion to multi-class classification (mild, moderate, severe anemia).
- Deployment as a mobile or desktop application for field use.
- Use of deep learning for improved accuracy.
- Incorporation of patient history and lifestyle data for holistic predictions.

10. Appendix

1. **Source Code:** Click Here

2. **GitHub Link:** Click Here

3. **Video Link:** Click Here (Click on view raw after getting redirected to the page)