Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions

Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the formal launch of the project, setting a clear vision and roadmap for its execution. During this stage, the team defines the project's objectives, scope, and success criteria, while identifying key stakeholders and assigning responsibilities. Resources are allocated efficiently, timelines are established, and potential risks are assessed with mitigation strategies in place. This phase ensures that all participants are aligned, expectations are clear, and the groundwork is laid for a smooth, structured, and goal-driven machine learning project.

Activity 1: Define Problem Statement

Problem Statement: A healthcare provider seeks to detect anemia in patients using basic blood test parameters such as Hemoglobin, MCH, MCHC, and MCV. The challenge lies in accurately identifying anemia cases from non-anemia cases, considering variations across gender and other influencing factors, to support early diagnosis and treatment.

AnemiaSense Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

The proposed project, **"AnemiaSense"**, aims to apply machine learning techniques to classify patients as anemic or non-anemic based on clinical blood parameters. By analyzing a dataset containing gender and various hematological indices, the model will provide an efficient, data-driven decision-support tool for healthcare practitioners. This initiative seeks to improve diagnostic accuracy, enable early intervention, and contribute to better patient outcomes while streamlining medical workflows.

AnemiaSense Project Proposal Report: Click Here

Activity 3: Initial Project Planning

Initial Project Planning focuses on setting key objectives, defining the scope, and identifying stakeholders for the anemia detection system. This includes establishing timelines, assigning resources, and outlining the data processing and modeling workflow. The team will thoroughly understand the dataset, set performance goals for the classification model, and develop a strategy for data preprocessing, analysis, and evaluation. Effective initial planning will ensure that the project is methodically executed and achieves reliable, actionable results in detecting anemia.

Anemia Detector Project Planning Report: Click Here

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant anemia-related health data from Kaggle, ensuring data quality through verification and addressing any inconsistencies. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

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

The dataset for *AnemiaSense – Predicting Anemia Status from Health Parameters* is sourced from Kaggle. It includes demographic, lifestyle, and health-related metrics. Data quality is ensured through thorough verification, confirming the absence of missing values, and adhering to ethical guidelines, establishing a reliable foundation for predictive modeling.

AnemiaSense Data Collection Report: Click Here

Activity 2: Data Quality Report

The dataset for *AnemiaSense – Predicting Anemia Status from Health Parameters* is sourced from Kaggle. It includes demographic, lifestyle, and health-related metrics. Data quality is ensured through verification, confirming no missing values, and documenting data imbalance between anemic and non-anemic records, with imbalance addressed through random undersampling.

AnemiaSense Data Quality Report: Click Here

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the anemia dataset to understand patterns, distributions, and potential outliers. Preprocessing includes handling class imbalance, encoding categorical variables, and scaling numerical features. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in predicting anemia status.

AnemiaSense Data Exploration and Preprocessing Report: Click Here

Milestone 3: Model DevelopMoment 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 to ensure accurate anemia status predictions.

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., hemoglobin level, gender, age, dietary habits) for the anemia detection model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to accurately classify anemia status.

AnemiaSense Feature Selection Report: Click Here

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Decision Tree, KNN, and XGB models for anemia prediction. It considers each model's strengths in handling imbalanced data, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

AnemiaSense Model Selection Report: Click Here

Activity 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 using metrics such as accuracy, F1 score, and confusion matrix to ensure reliability and effectiveness in predicting anemia status.

AnemiaSense Model Development Phase Template: Click Here

Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. Typically, this includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency. However, in this project, hyperparameter tuning was not performed because certain models, such as Decision Tree and Gradient Boosting, achieved 100% accuracy during initial training. This decision was made to avoid overfitting and ensure that the reported performance is realistic and generalizable.

Activity 1: Hyperparameter Tuning Documentation

Due to the exceptional performance (100% accuracy) achieved by multiple models during initial training, hyperparameter tuning was skipped. This approach mitigates the risk of overfitting and ensures the model remains robust when applied to unseen anemia diagnosis data.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the accuracy and F1 scores of all trained models. Both Decision Tree and Gradient Boosting achieved perfect accuracy, while other models such as Logistic Regression and Gaussian Naïve Bayes also demonstrated high performance. This comparison confirms that the chosen models already operate at peak efficiency without further tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification outlines the rationale for choosing Gradient Boosting as the ultimate model for anemia prediction. Its ability to handle complex relationships, consistently deliver high accuracy, and maintain performance without hyperparameter adjustments makes it ideal for the project's goals.

AnemiaSense Model Optimization and Tuning Phase Report: Click Here

Milestone 5: Project Files Submission and Documentation

For project file submission in Github, Kindly click the link and refer to the flow. Click Here

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

Milestone 6: Project Demonstration

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation