Lymphography Classification Using ML

Milestone 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: A researcher seeking to improve lymphography analysis and trying to classify lymph nodes accurately but the current model has low accuracy.
Because these are limited annotated datasets which makes feel motivated to develop a better classification model.

Lymphography Classification Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

The proposal report aims to revolutionize the classification of lymphography using machine learning, boosting efficiency and accuracy in diagnostic processes. It tackles system inefficiencies, promising better operations, reduced risks. Key features include a machine learning-based classification model, Real-time processing, Scalability and User-friendly interface.

Lymphography Classification Project Proposal Report: Click Here

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a lymphography image classification 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. Effective initial planning lays the foundation for a systematic and well-executed project, ensuring successful outcomes.

Lymphography Classification 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 lymphography

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.

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

The dataset for "Lymphography Classification - Medical Diagnosis" is sourced from Kaggle. It includes patient details and medical metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Lymphography Classification Data Collection Report: [Click Here](#)

Activity 2: Data Quality Report

The dataset for "Lymphography Classification" from Kaggle undergoes rigorous data quality measures. It includes a thorough verification process to ensure accuracy, addressing any missing or incomplete values. Ethical guidelines are strictly followed to protect patient information. The dataset is regularly updated to maintain its relevance and reliability. These steps ensure a solid foundation for accurate and ethical predictive modeling.

Lymphography Classification Data Quality Report: [Click Here](#)

Activity 3: Data Exploration and Preprocessing

Initial exploration of the lymphography dataset involves understanding the distribution of medical and patient details. We identify and handle missing values, ensuring the dataset's completeness. Categorical variables are encoded appropriately, and numerical features are standardized. Outliers are detected and addressed to prevent skewed model training. This preprocessing ensures the dataset is clean and ready for effective predictive modeling

Lymphography Classification Data Exploration and Preprocessing Report: [Click Here](#)

Milestone 3: Model Development Phase

In the model development phase, we experiment with various machine learning algorithms such as logistic regression, decision trees, and support vector machines. Hyperparameter tuning is performed to optimize model performance. The dataset is split into training and testing sets to evaluate accuracy and generalization. Cross-validation techniques are employed to ensure robust performance. The best-performing model is selected based on metrics like precision, recall, and F1-score, ready for deployment.

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., Age, Node Count, Tumor Size) for the lymphography classification model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern various lymphographic conditions. By carefully selecting these features, we enhance the model's performance and reliability in medical diagnosis.

Lymphography Classification 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 lymphography classification. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

Lymphography Classification 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 lymphography classification dataset, setting the foundation for predictive modeling. The subsequent Model Validation and EvaluationReport rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness

Lymphography Classification 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. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Activity 1: Hyperparameter Tuning Documentation

The Gradient Boosting 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.

Activity 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 Gradient Boosting model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing Gradient Boosting as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives.

Lymphography Classification 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.