DETECTION OF PHISHING WEBSITES FROM URLS

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: Detecting phishing websites from URLs is a crucial task in cybersecurity. To tackle this challenge, several problem statements can be defined. Firstly, a classification problem can be formulated, where a system is designed to categorize URLs as phishing or legitimate based on features extracted from the URL itself. Another approach is anomaly detection, which involves identifying URLs that significantly deviate from legitimate URL patterns, indicating potential phishing attempts.

DETECTION OF PHISHING WEBSITE FROM URLS
Problem Statement Report: **Click Here**

Activity 2: Project Proposal (Proposed Solution)
The proposed project, A project proposal report for detecting phishing websites from URLs outlines a comprehensive plan to develop a system that identifies and flags fraudulent websites. The report begins by highlighting the significance of phishing detection, citing the growing number of cyber attacks and financial losses. The problem statement defines the scope, emphasizing the need for accurate URL analysis to distinguish legitimate sites from phishing ones.
DETECTION OF PHISHING WEBSITE FROM URLS
Project Proposal Report: **Click Here**

Activity 3: Initial Project Planning
Initial Project Planning involves outlining key objectives, defining scope, and identifying and predicting phishing website from urls. 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.

DETECTION OF PHISHING WEBSITE FROM URLS
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 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.

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

The dataset for "Detection of phishing website from urls" is sourced from kaggle. It includes applicant details and financial metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

DETECTION OF PHISHING WEBSITE FROM URLS Data Collection Report: ClickHere

Activity 2: Data Quality Report

The dataset for " Detection of phishing website from urls " is sourced from Kaggle. It includes applicant details and financial metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

DETECTION OF PHISHING WEBSITE FROM URLS Data Quality Report: Click here

Activity 3: Data Exploration and Preprocessing

Data Exploration involves predicting the phishing website from urls. 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 loan approval project.

DETECTION OF PHISHING WEBSITE FROM URLS
Data Exploration and Preprocessing Report: Click Here

Milestone 3: Model Development Phase

The Model Development Phase entails crafting a predictive model for loan approval. 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.

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., Gender, Married, Credit History) for the loan approval model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern credible loan applicants.

DETECTION OF PHISHING WEBSITE FROM URLS

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 loan approval prediction. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

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 loan approval 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.

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, predicting the phishing website from urls

Detection of phishing websites using urls 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.

Link : **Click here**