SMS SPAM DETECTION

Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the beginning of the SMS Spam Detection project, defining its goals, scope, and identifying stakeholders. This critical phase establishes the project's foundation, sets realistic timelines, allocates resources, and includes a risk assessment and mitigation strategy. Proper initiation ensures that the project remains organized, efficient, and aligned with its objectives, preparing for any challenges in the development phase.

Activity 1: Define Problem Statement

Problem Statement: With the rise of unsolicited and malicious SMS messages, users face a constant influx of spam. This project aims to develop an SMS spam detection system that accurately classifies messages as either spam or not spam, providing a solution to this growing issue by leveraging machine learning for text classification.

Ref. template: Click Here

SMS Spam Detection Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

The proposed project, "SMS Spam Detection" aims to leverage machine learning models like Naive Bayes to accurately classify SMS messages. Using a labeled dataset containing spam and non-spam messages, the project seeks to develop a web application that provides real-time spam detection. The initiative aligns with the objective of improving user experience and enhancing message security.

Ref. template: Click Here

SMS Spam Detection Project Proposal Report: Click Here

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives and defining the scope for the spam detection system. It includes setting timelines, allocating resources,

and determining the overall project strategy. The team establishes a clear understanding of the dataset, formulates goals for analysis, and plans the workflow for text preprocessing and model development. Effective initial planning ensures a smooth and systematic project execution.

Ref. template: Click Here

SMS Spam Detection Project Planning Report: Click Here

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase includes gathering SMS spam data from public sources like Kaggle. Ensuring data quality through proper verification and addressing missing values is critical. Preprocessing involves cleaning 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 "SMS Spam Detection" is sourced from Kaggle. It includes labeled SMS messages with spam and non-spam classifications. Data quality is ensured by addressing missing values and ensuring proper cleaning, forming a strong foundation for building the predictive model.

Ref. template: Click Here

SMS Spam Detection Data Collection Report: Click Here

Activity 2: Data Quality Report

The dataset for "SMS Spam Detection" is verified for quality. Data cleaning includes removing irrelevant characters, handling duplicate messages, and addressing text noise to ensure the dataset is of high quality for accurate modelling.
Ref. template: Click Here
SMS Spam Detection Data Quality Report: Click Here

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the SMS message dataset to understand
distributions, patterns, and trends. Preprocessing includes text cleaning, tokenization, and vectorization using TF-IDF to convert text into numerical format. These steps are essential for enhancing data quality and improving the model's performance in spam classification.

Ref. template: Click Here

SMS Spam Detection Data Exploration and Preprocessing Report: Click Here

Milestone 3: Model Development Phase

The Model Development Phase involves developing a machine learning model for SMS spam detection. It encompasses feature extraction using TF-IDF, model selection (Naive Bayes, Decision Tree, Logistic Regression), and training the models. The phase includes validating and evaluating the model performance using accuracy, precision, and F1-score to ensure effective classification.

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features like word frequency and message length for the SMS spam detection model. It evaluates the relevance and impact of these features on predictive accuracy, ensuring the model captures key patterns to classify spam messages effectively.

Ref. template: Click Here

SMS Spam Detection Feature Selection Report: Click Here

Activity 2: Model Selection Report

The Model Selection Report details the models evaluated for SMS spam classification, including Naive Bayes, Decision Tree, and Logistic Regression. The report explains the strengths and weaknesses of each model in handling text classification, with Naive Bayes being the most effective in this context.

Ref. template: Click Here

SMS Spam Detection Model Selection Report: Click Here

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

Report

The Initial Model Training Code implements the selected algorithm (Naive Bayes) on the SMS dataset, setting the foundation for prediction. The Model Validation and Evaluation Report rigorously assesses performance using metrics like accuracy, precision, recall, and F1-score to ensure the model effectively classifies spam messages.

Ref. template: Click Here

SMS Spam Detection Model Development Phase Template: Click Here

Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining the Naive Bayes model for peak performance. It includes tuning hyperparameters, comparing performance metrics, and justifying the final model selection to ensure optimal predictive accuracy and efficiency.

Activity 1: Hyperparameter Tuning Documentation

The Naive Bayes model was selected for its simplicity and strong performance. Tuning focused on adjusting the alpha parameter to optimize smoothing, improving the model's ability to handle spam classification. Hyperparameter tuning increased accuracy while maintaining low computational cost.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts baseline and tuned metrics for the Naive Bayes model, highlighting improvements in precision and recall. The report provides a comprehensive understanding of the model's refined predictive capabilities through hyperparameter tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing Naive Bayes as the final model due to its exceptional performance in text classification. Its simplicity, speed, and accuracy make it the most suitable choice for real-time

spam detection in this project.

Ref. template: Click Here
SMS Spam Detection Model Optimization and Tuning Phase Report: Click Here

Milestone 5: Project Files Submission and Documentation

For project file submission, refer to the GitHub link below to upload all related project files. Additionally, the documentation should follow the prescribed format for clarity and completeness.

Ref. template: Click Here

Milestone 6: Project Demonstration

In the final Project Demonstration, the team will record a video demonstration of the working project. During the presentation, the team will explain the project goals, workflow, and demonstrate the application of the machine learning model to classify SMS messages.