

# SmartBridge - Sentiment Analysis of Commodity News (Gold)

## 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 self-employed customer wants to know the price sentiment of commodity news like gold from the newspaper headlines.

**Problem Statement Report:** [Click Here](#)

### Activity 2: Project Proposal (Proposed Solution)

The proposed project, "Sentiment Analysis of Commodity News (Gold)" aims to leverage machine learning for more accurate price sentiment predictions. Using a comprehensive dataset including news headlines, price and conditions the project seeks to develop a predictive model optimizing Sentiment Analysis of Commodity News (Gold). This initiative aligns with our objective to enhance decision-making, reduce risks, and streamline lending operations, ultimately improving customer satisfaction and operational efficiency.

**Project Proposal Report:** [Click Here](#)

### Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a estimating and predicting sentiment analysis of gold. 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.

**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 "Sentiment Analysis of Commodity News (Gold)" 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.

**Data Collection Report:** [Click Here](#)

### **Activity 2: Data Quality Report**

The dataset for "Sentiment Analysis of Commodity News (Gold)" is sourced from Kaggle. It includes applicant details and prediction metrics. Data quality is ensured thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

**Data Quality Report:** [Click Here](#)

### **Activity 3: Data Exploration and Preprocessing**

Data Exploration involves analyzing the prediction dataset to understand patterns, distributions. 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.

**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., news headline, status etc.) for the price prediction model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern credible price prediction.

**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 price 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 sentiment price prediction 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 logistic regression 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 logistic regression 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 logistic regression as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

**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.