Prosperity Prognosticator :Machine Learning for Startup Success Prediction

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: Startups are companies started by an individual or a small group of people based on a unique business idea. Most of the startups face high failure rates whereas few startups are very successful such as Facebook, Uber, and Snapchat etc. This work proposes a predictive method to determine the outcome of start-ups based on several key factors that contribute to the success such as Acquisitions, Investments, Total rounds of funding etc.

2: Project Proposal (Proposed Solution)

Startups play a major role in economic growth. They bring new ideas, spur innovation, create employment thereby moving the economy. There has been an exponential growth in startups over the past few years. Predicting the success of a startup allows investors to find companies that have the potential for rapid growth, thereby allowing them to be one step ahead of competition. venture capitalists and angel investors can use predictive models to identify promising startups. Governments and NGOs can design better support programs for startups based on predictive insights.

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a Startup Success. 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.

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 "Machine Learning for Startup Success Prediction" is sourced from Kaggle. It includes funding year and rounds, relationships. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Activity 2: Data Quality Report

The dataset for " Machine Learning for Startup Success Prediction " is sourced from Kaggle. It includes Funding details and milestones metrics. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the Startup Companies dataset to understand patterns, distributions, and outliers. 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 Startup Success project.

Milestone 3: Model Development Phase

The Model Development Phase entails crafting a predictive model for Startup Success. It encompasses strategic feature selection, evaluating and selecting models (Random Forest, Gradient Boosting, XGB), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the Success prediction process.

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features (e.g., Funding rounds, Relationships, Rupees) for the Startup Success model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern Startup Success.

Activity 2: Model Selection Report
The Model Selection Report details the rationale behind choosing Random Forest, Gradient Boosting, AdaBoost, and XGB models for Startup Success 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.

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report
The Initial Model Training Code employs selected algorithms on the Startup companies 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 Startup Success.

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 Random Forest 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 Random Forest 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 Random Forest as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

Milestone 5: Project Files Submission and Documentation
For project file submission in Github, Kindly click the link and refer to the flow.

For the documentation, Kindly refer