SmartBridge - Ecommerce Shipping Prediction Using Machine Learning

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

An eCommerce company seeks to predict shipping times for orders based on various factors such as order details, customer location, and historical shipping data. The goal is to enhance customer satisfaction by providing accurate delivery estimates and optimizing logistics operations

link:https://drive.google.com/file/d/1KMKgS4RZDUN62fTU-93AmeFYvUhJGuqV/view?usp=sharing

Activity 2: Project Proposal (Proposed Solution)

Develop a machine learning model to predict eCommerce shipping times using historical shipping data, customer demographics, and order specifics. The model will be trained and validated using various algorithms, such as Random Forest and Gradient Boosting, to ensure accuracy. By implementing this predictive system, the company aims to provide more reliable delivery estimates, streamline logistics, and improve overall customer satisfaction. This will involve data preprocessing, feature engineering, and rigorous model evaluation to ensure robustness.

link:https://drive.google.com/file/d/1UBeyjiTNlY83GYt1zExQKWKOCLvFnCWX/view?usp=sharing

Activity 3: Initial Project Planning

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

link: https://drive.google.com/file/d/19pX3g3O9nHeDgSCYjjE8d690uofuOZ_-/view?usp=sharing

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 data collection plan involves using the `train.csv` dataset from Kaggle, which includes comprehensive order details, customer information, and shipping data. Key features such as order ID, product type, order date, customer location, shipping method, and delivery dates will be utilized. A data quality report will be generated to assess completeness, handle missing values, and identify outliers. Data preprocessing will involve cleaning, feature engineering, normalization, and splitting into training and validation sets. Validation will include cross-referencing with historical shipping data and performing exploratory data analysis.
Link: https://drive.google.com/file/d/1i-etrKnYzG6AVNGFI0-dbszLy_TZQP32/view?usp=sharing

Activity 2: Data Quality Report

The dataset for the eCommerce shipping prediction model is sourced from Kaggle and includes order details, customer information, and shipping data. To ensure data quality, the dataset undergoes thorough verification, including handling missing values, removing duplicates, and identifying outliers. Adherence to data integrity and privacy guidelines is maintained throughout the process. This meticulous approach establishes a reliable foundation for developing an accurate and effective shipping time prediction model.
Link:
https://drive.google.com/file/d/1dCh2HXnIsKq661dBEchPjArGE2tSnPK9/view?usp=sharing

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the eCommerce shipping dataset to understand patterns, distributions, and outliers. Preprocessing includes handling missing values, scaling numerical features, and encoding categorical variables. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in the shipping time prediction project.
Link: https://drive.google.com/file/d/1znVtnFPaJR66iQN2FffCQl19yc3qbERW/view?usp=sharing

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., order date, shipping method, customer location) for the shipping time prediction model. It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to accurately predict shipping times.
Link: https://drive.google.com/file/d/1lstdUxpILMSFaveiPkOcY44ylLsUoOpk/view?usp=sharing

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing the Random Forest model for the shipping time prediction project. It considers Random Forest's strengths in handling complex relationships, its robustness against overfitting, and its overall predictive performance. This model is selected for its ability to capture intricate patterns in the data and provide reliable shipping time predictions, aligning with the project's objectives.

Link: https://drive.google.com/file/d/1lstdUxpILMSFaveiPkOcY44ylLsUoOpk/view?usp=sharing

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

the Initial Model Training Code applies the Random Forest algorithm to the eCommerce shipping dataset, establishing the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, using metrics such as Mean Absolute Error (MAE) and R-squared to ensure accuracy and reliability in predicting shipping times.

Link:https://drive.google.com/file/d/1crB3Hv0GSODRMsXuYipNBDmlSboer8Wt/view?usp=sharing

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

Link:https://drive.google.com/file/d/1oEb6YN6xehK1yp53bYuiUoLlEuDH468G/view?usp=sharing

Milestone 5: Project Files Submission and Documentation

For project file submission in Github, Kindly click the link and refer to the flow. **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.https://drive.google.com/file/d/1EXBwPDzQM8xwWPpNoQDIJgMFiaCeihdt/view?usp=sharing