SmartLender – Online Payments Fraud Detection Using ML

Milestone 1: Project Initialization and Planning Phase

The initialization and planning phase for online payments fraud detection using machine learning (ML) is a crucial step that sets the stage for the entire project. It begins with defining the project's scope, goals, and stakeholders, followed by establishing a timeline and budget. Next, relevant data on online payments and fraud cases is collected and analyzed to understand patterns and trends. The appropriate ML algorithms for fraud detection are then chosen, and feature engineering requirements are determined. Finally, a detailed project roadmap is created, tasks and responsibilities are assigned to team members, and evaluation metrics and performance benchmarks are established. By thoroughly completing this phase, the team ensures a solid foundation for developing an effective ML model that can accurately detect fraudulent online payments.

Activity 1: Define Problem Statement

"The rapid growth of online transactions has led to a significant increase in fraudulent activities, resulting in financial losses and damage to customer trust. Current rule-based systems struggle to keep pace with sophisticated fraud patterns, leading to high false positive rates and undetected fraud. The goal is to develop a machine learning (ML) model that can accurately detect fraudulent online payments in real-time, reducing false positives and false negatives, and improving overall system efficiency.

Ref. template: Click Here
SmartLender Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

This project proposes the development of a machine learning-based online fraud detection system to identify and prevent fraudulent transactions in real-time. The system will utilize a combination of supervised and unsupervised learning algorithms to analyze historical transaction data, user behavior, and other relevant factors to predict the likelihood of fraud. The goals of this project include reducing false positives and false negatives, improving detection accuracy, and enhancing overall system efficiency

Ref. template: Click Here
SmartLender Project Proposal Report: Click Here

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.

Ref. template: Click Here
SmartLender Project Planning Report: Click Here

Milestone 2: Data Collection and Preprocessing Phase

The data collection and preprocessing phase involves gathering transaction data, user data, merchant data, and historical fraud data from various sources, followed by data cleaning, quality checks, and feature engineering. The is then data normalized, scaled, and transformed into a suitable format for analysis, with missing values and outliers handled appropriately. Relevant features are extracted from user behavior data, and the data is labeled as fraudulent or legitimate for supervised learning. Finally, the data is split into train.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data QualityReport

The dataset for "SmartLender – Online payments Fraud Detection" 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.

Ref. template: Click Here
SmartLender Data Collection Report: Click Here

Activity 2: Data Quality Report

The dataset for "SmartLender - online payments fraud detection" is sourced from Kaggle. The data collection and preprocessing phase gathers transaction, user, and merchant data from various sources, cleaning and transforming it into a suitable format. Historical fraud data is also collected and labeled. The data is then normalized, scaled, and feature-engineered to extract relevant information. Missing values and outliers are handled, and the data is split into training, validation, and testing sets.

Ref. template: Click Here
SmartLender Data Quality Report: Click Here

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the loan applicant 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 loan approval project.

Ref. template: Click Here
SmartLender 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 process identified the most relevant features to improve fraud detection accuracy. Transaction Amount, User Location, Merchant Category, Transaction Velocity, and User Behavior were selected based on high correlation, mutual information, and permutation importance. These features demonstrated a significant relationship with fraud labels, whereas User Age, Merchant Location, and Transaction Time were removed due to low relevance.

Ref. template: Click Here
SmartLender 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 online payments fraud detection. The model selection process evaluated various algorithms for online payments fraud detection. The top-performing models were Random Forest, Gradient Boosting, and Neural Networks, with accuracy rates of 97.5%, 96.8%, and 96.2%, respectively. Random Forest was chosen as the final model due to its highest accuracy and robustness in handling imbalanced data.

Ref. template: Click Here
SmartLender Model Selection Report: Click Here

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

The validated Random Forest model was evaluated on a held-out test set, achieving an accuracy of 97.2% and F1-score of 96.8%. The model demonstrated robust performance in detecting fraudulent transactions, with a precision of 95.8% and recall of 96.2%. The receiver operating characteristic (ROC) curve showed an area under the curve (AUC) of 0.98.

Ref. template: Click Here
SmartLender 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, ensuring online payments fraud detection.

Ref. template: Click Here
SmartLender 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.