PROJECT OVERVIEW STATEMENT	Project Name: "DeepGuard": High-Accuracy Credit Card Fraud Detection	Student Name: Piyush Gupta

Problem/Opportunity:

The increasing rate of credit card fraud poses significant financial and security risks to consumers and financial institutions. A robust and automated system is urgently required to detect and flag fraudulent transactions in real time.

Goal:

The goal of "DeepGuard" is to achieve a minimum detection accuracy of 98% for identifying fraudulent credit card transactions. DeepGuard will employ a spectrum of machine learning algorithms to reach this ambitious target, such as Logistic Regression, Naive Bayes, Random Forest, K-Neighbor, and Support Vector Machines. In addition, we will push the boundaries of traditional machine learning by integrating a Convolutional Neural Network (CNN) developed in TensorFlow. This multifaceted approach aims to create a system that is not only highly accurate but also robust and adaptable to emerging fraud patterns. We will also test the efficacy of under-sampling and over-sampling methods to address data imbalance issues. To bring the project to its culmination, the DeepGuard model will deployed through GitHub Actions and AWS EC2 instances with enhancements to optimize speed, performance, and reliability. The project is targeted for completion by the first week of December.

Objectives:

- Acquire and clean a dataset containing a minimum of 200,000 credit card transactions through data collection and preprocessing activities, targeting a fully processed dataset as the measure, with an expected completion date of September 30, 2023.
- Assess the efficacy of under-sampling and over-sampling techniques by implementing, testing, and analyzing both methods, targeting comparison metrics to determine which method yields better results, with an expected completion date of October 10, 2023.
- Implement and evaluate machine learning algorithms like logistic regression, Naive Bayes, random forest, K-neighbor, and support vector machines through coding, algorithmic development, and evaluation. Performance metrics like accuracy, AUC score, precision, recall, F1-score, and confusion matrix will be used as a measure, and the project is expected to be done by October 20, 2023.
- Design, implement, and initially test a Convolutional Neural Network (CNN) model using TensorFlow through the actions of developing and evaluating the deep learning model, targeting successful development, implementation, and initial testing as the measure, with an expected completion date of October 30, 2023.
- Deploy the finalized model using GitHub actions and AWS EC2 instances by handling model deployment and operational testing, targeting successful model deployment and real-time fraud detection capabilities as the measure, with an expected completion date of November 10, 2023.

Success Criteria:

- A fraud detection system with an accuracy rate of at least 98%
- Effective flagging of fraudulent transactions with minimal false positives
- Successful application of multiple machine learning algorithms and a CNN model
- Successful deployment using GitHub actions and AWS EC2 instances

Assumptions, Risks, Obstacles:

Assumption: The dataset is assumed to be representative of real-world transactions. Additionally, some features are hidden through PCA for confidentiality reasons and are named generically, such as V1, V2, and many more.

Risk: Aiming for a high-accuracy target may introduce the risk of overfitting, which could negatively affect the model's applicability to new data.

Obstacle: The class imbalance in the dataset may skew the model's effectiveness in identifying fraud accurately, which could reduce the model's utility in real-world applications.

Prepared By	Date	Approved By	Date
Piyush Gupta	09/25/2023		