PROJECT OVERVIEW STATEMENT	Project Name: 'InvisiFraud': Fraud Intelligence System	Student Name: Sai Bhere

Problem/Opportunity:

With the rise of online transactions and the widespread usage of credit cards, fraud has become a serious and growing issue for both financial institutions and consumers. Traditional rule-based fraud detection systems are frequently ineffective, resulting in a high number of false positives, missed fraud instances, and significant financial losses. The opportunity is to use modern machine learning and data analytics techniques to create a more efficient, accurate, and scalable fraud detection system that can identify fraudulent transactions in real time, thereby minimizing false alarms and the danger of undetected fraud.

Goal:

InvisiFraud attempts to detect fraudulent credit card transactions with a detection accuracy of at least 90% using multiple machine learning methods. The research also uses a Convolutional Neural Network (CNN) constructed in TensorFlow to create a system that is both extremely accurate and flexible to new fraud trends. To solve the dataset's class imbalance, the system will experiment with under-sampling and over-sampling strategies.

Objectives:

- **Data collection and preprocessing**.
  Outcome: Gather and clean a dataset of at least 200,000 credit card transactions to assure data accuracy and integrity.
  Timeline: Complete by October 01, 2024.
  Measure: Success will be defined as having a fully preprocessed dataset available for analysis.
  Action: Carry out data collecting, cleaning, and preprocessing tasks such as managing missing values, normalizing data, and ensuring a balanced representation of transaction kinds.

- **Handle class imbalance.**
  Outcome: Assess the efficacy of under- and over-sampling approaches in addressing class imbalance.
  Timeline: Complete on October 08, 2024.
  Measure: Compare the performance of models trained with both strategies using metrics like accuracy, precision, recall, and AUC score.
  Action: Apply, test, and analyze under- and over-sampling algorithms to the dataset to see which produces superior fraud detection findings.

- **Development of a Machine Learning Model**
  Outcome: Apply and assess machine learning methods such as Logistic Regression, Naive Bayes, Random Forest, K-Nearest Neighbors, and Support Vector Machines.
  Timeline: Complete on October 15, 2024.
  Measure: Performance will be evaluated using metrics such as accuracy, AUC score, precision, recall, F1-score, and a confusion matrix.
  Action: Create and test each algorithm on the preprocessed data to find which model performs best.

- **Testing and Designing the CNN Model**
  Outcome: Create and test a Convolutional Neural Network (CNN) using TensorFlow to discover fraud patterns.
  Timeline: Complete by October 29, 2024.
  Measure: Measure the CNN's accuracy and performance parameters to ensure it achieves or exceeds 90% accuracy in fraud detection.
  Action: Create and test the CNN model with deep learning approaches, fine-tuning the network for the best detection performance.

- **Model Deployment**.
  Outcome: Use the finished model for real-time fraud detection.
  Timeline: Complete on November 12, 2024.
  Measure: Success will be determined by the model's operational performance, which includes its ability to

handle transactions in real time with high accuracy and low latency.
Action: Manage model deployment, integration, and operational testing to assure dependability and scalability in a real-world setting.

Success Criteria:

- The project's success is determined by achieving a minimum detection accuracy of 90% in fraudulent transactions using machine learning algorithms and the Convolutional Neural Network (CNN), which is consistently met or exceeded during testing and real-world deployment.
- The model's success is determined by its ability to handle class imbalance (fraudulent vs. legitimate transactions) using under-sampling or over-sampling techniques without compromising performance metrics. High precision, recall, F1-score, and AUC score indicate balanced detection of both fraud and legitimate transactions.
- The project's success hinges on the successful deployment of a fraud detection system that demonstrates reliable real-time detection capabilities, ensuring minimal latency and speed in transactions, as confirmed by real-world testing.
- The project must be completed within specified deadlines, including data acquisition, model implementation, and deployment. This includes completing preprocessing, model development, CNN testing, and deployment by the first week of December, indicating the project's completion by the specified dates.
- The project's success is determined by meeting or exceeding the expectations of key stakeholders, including professors, collaborators, and potential clients, in terms of performance, innovation, and applicability to real-world fraud detection, as confirmed by stakeholder reviews and evaluations.

Assumptions, Risks, Obstacles:

- **Assumption**: The dataset is expected to accurately represent real-world credit card transactions. Furthermore, several attributes have been encrypted using PCA for anonymity, with generic designations such as V1, V2, and so on.
- **Risk**: Pursuing a high accuracy aim raises the possibility of overfitting, which may impair the model's capacity to generalize to new, unknown data.
- **Obstacle**: The dataset's significant class imbalance could affect the model's capacity to detect fraud accurately, potentially limiting its practical application in real-world circumstances.

Prepared By	Date	Approved By	Date
Sai Bhere	09/24/2024		