## 1. Introduction

In today's digital economy, credit card transactions have become an essential aspect of global financial activities. The convenience of digital payments, however, has also introduced a significant rise in fraudulent activities. As consumer reliance on online shopping and card-based transactions increases, so does the opportunity for cybercriminals to exploit these platforms. According to the Nilson Report, global card fraud losses exceeded $28 billion in 2020 alone. This staggering figure highlights the scale of the problem and underscores the urgent need for efficient methods to detect fraudulent transactions. Fraudulent credit card transactions not only result in monetary losses but also damage customer trust and brand reputation, further emphasizing the importance of timely fraud detection.

This project leverages the power of data mining and machine learning techniques to detect fraudulent credit card transactions. The dataset used in this study is highly imbalanced, with fraudulent transactions comprising a very small proportion of the total transactions. The core aim of the project is to apply a variety of data preprocessing and machine learning techniques to identify and predict fraudulent activities, thereby providing a comprehensive approach to detecting anomalies in the dataset. By exploring this problem domain, the project aims to provide hands-on experience with crucial techniques like data preprocessing, model evaluation, and pattern recognition, as well as practical exposure to the challenges posed by imbalanced datasets.

**Problem Statement:**

Fraudulent transactions represent a significant challenge for the financial industry, with credit card companies, banks, and e-commerce platforms facing substantial financial losses due to fraudulent activities. Detecting these fraudulent transactions in real-time is crucial for preventing financial damage, but the problem is complicated by the fact that fraudulent transactions are rare compared to legitimate ones, leading to class imbalance. This project focuses on developing a predictive model that can effectively identify fraudulent credit card transactions from a highly imbalanced dataset, utilizing a combination of data mining techniques such as classification, anomaly detection, and resampling methods. Can machine learning and data mining techniques effectively identify fraudulent credit card transactions from a highly imbalanced dataset?

**Research Question:**

How can machine learning models be effectively applied to predict fraudulent credit card transactions in a highly imbalanced dataset, and what are the most effective techniques for handling class imbalance while maximizing model performance?

## 2. Dataset Description

The dataset used for this project is sourced from Kaggle, specifically the "Credit Card Fraud Detection" dataset (available at: https://www.kaggle.com/mlg-ulb/creditcardfraud). It contains a total of 284,807 credit card transactions, including both legitimate and fraudulent entries. The dataset is highly imbalanced, with only 492 fraudulent transactions, representing just 0.172% of the total dataset, which presents a significant challenge for traditional machine learning models that tend to favor the majority class.

The dataset contains several key features:

- **28 anonymized principal components (V1 to V28):** These features represent transformations of the original data using Principal Component Analysis (PCA), ensuring that the data remains confidential. The exact nature of these components is not disclosed due to privacy concerns, which is a standard practice in credit card fraud detection to protect customer data.
- **Time:** This feature captures the number of seconds elapsed between the current transaction and the first transaction in the dataset. This time variable helps to establish patterns that could reveal suspicious behaviors over time.
- **Amount:** This feature represents the transaction amount, which can be crucial in detecting anomalies. Fraudulent transactions often involve higher or unusually low amounts compared to legitimate transactions.
- **Class:** This is the target variable, where '0' indicates a legitimate transaction and '1' indicates a fraudulent transaction.

The dataset was collected in September 2013 and remains one of the most referenced datasets in the field of fraud detection. Its popularity stems from its real-world relevance and the challenges it poses, particularly the class imbalance and the need for effective anomaly detection.

## 3. Methodology

The project follows a structured workflow, encompassing data preprocessing, exploratory data analysis (EDA), handling class imbalance, model training, evaluation, and interpretability using SHAP (SHapley Additive exPlanations). Each step is critical to ensuring that the machine learning pipeline is effective and reliable.

### a. Data Preprocessing

Data preprocessing is a crucial step to prepare the dataset for modeling. The first step in preprocessing was the removal of duplicate records, with 108 duplicate entries identified

and eliminated to avoid skewing the model's results. Next, **feature scaling** was applied to the 'Amount' and 'Time' variables, which are on different scales. Standardization was chosen to normalize these variables, ensuring that all features contribute equally to the model.

After standardization, the original 'Amount' and 'Time' columns were dropped from the dataset, as the newly scaled versions were retained. The final step involved removing duplicate entries from the dataset using the drop_duplicates() function, which ensured the integrity of the data.

```
scaler = StandardScaler()

df['norm_amount'] = scaler.fit_transform(df['Amount'].values.reshape(-1,1))

df['norm_time'] = scaler.fit_transform(df['Time'].values.reshape(-1,1))

df.drop(['Amount', 'Time'], axis=1, inplace=True)

df.drop_duplicates(inplace=True)
```

**b. Exploratory Data Analysis (EDA)**

EDA was performed to better understand the distribution of the dataset and identify potential patterns. A significant finding from this analysis was the severe **class imbalance**, with only 0.172% of the transactions being fraudulent. A **correlation heatmap** was used to visualize relationships among the principal component features (V1 to V28), revealing that features such as V14, V10, and V17 were strongly correlated with fraudulent transactions, which could guide feature selection and model training.

**c. Handling Class Imbalance**

The class imbalance in the dataset is a major challenge. To address this, the **SMOTE (Synthetic Minority Over-sampling Technique)** algorithm was used to oversample the minority class (fraudulent transactions). SMOTE generates synthetic examples by interpolating between existing instances of the minority class, helping balance the class distribution and improving the model’s ability to detect fraud.

```
python

CopyEdit

sm = SMOTE()

X_resampled, y_resampled = sm.fit_resample(X, y)
```

**d. Predictive Modeling**

Several machine learning models were tested for their effectiveness in detecting fraudulent transactions. Among them, the **Random Forest Classifier** was selected as the final model due to its:

- High **precision** and **recall**, ensuring that the model is both accurate and able to detect a high proportion of fraudulent transactions.
- **Robustness to overfitting**, which is particularly important when working with imbalanced datasets.
- **Interpretability** through feature importance, allowing stakeholders to understand which features are most influential in detecting fraud.

**e. Clustering (KMeans)**

While not directly used for prediction, **KMeans clustering** was employed to detect potential outliers and explore inherent clusters within the transaction data. Although clustering did not directly enhance fraud detection, it provided valuable insights into the data structure and helped identify anomalies that could potentially signal fraudulent activity.

**4. Evaluation and Results**

The performance of the Random Forest model was evaluated using several key metrics, including **accuracy**, **precision**, **recall**, **F1-score**, and **ROC-AUC**. The model achieved the following results:

- **Accuracy:** 0.999 – This shows that the model correctly identified legitimate transactions with high accuracy.
- **Precision:** 0.94 – This indicates that 94% of the transactions predicted as fraudulent were indeed fraudulent.
- **Recall:** 0.91 – This means that the model detected 91% of all fraudulent transactions, which is critical in fraud detection, where missing a fraudulent transaction can result in significant financial loss.
- **F1-Score:** 0.925 – This metric balances precision and recall, providing a good overall measure of model performance.
- **ROC AUC:** 0.98 – This indicates that the model has excellent ability to distinguish between fraudulent and legitimate transactions.

The **confusion matrix** provided further insights into the model's performance:

| | Predicted 0 | Predicted 1 |
|---|---|---|
| Actual 0 | 85,293 | 3 |
| Actual 1 | 62 | 120 |

The high recall rate (91%) is particularly important in fraud detection, where minimizing false negatives (i.e., missed fraudulent transactions) is more critical than minimizing false positives.

**5. Visualizations**

To enhance the understanding of the model's performance and the dataset structure, several visualizations were created:

- **Class Imbalance Bar Chart**: Highlighted the severe imbalance between legitimate and fraudulent transactions.
- **Correlation Heatmap**: Showed the relationships among the PCA features.
- **Feature Importance Bar Plot**: Illustrated which features were most influential in predicting fraud.
- **SHAP Summary Plot**: Provided an explanation of the model's decision-making process.
- **ROC Curve**: Displayed the trade-off between true positive rate and false positive rate.

**6. Explainability Using SHAP**

To interpret the Random Forest model's predictions, SHAP was used. SHAP values allowed the project to break down the contribution of each feature to the model's decision-making process. The top contributing features to fraud predictions were V14, V10, and V17. SHAP summary plots and force plots were generated for both global and individual predictions, helping stakeholders understand how the model arrived at its conclusions and ensuring transparency in the decision-making process.

**7. Challenges Faced**

Several challenges were encountered during the course of the project:

- **Severe Class Imbalance**: SMOTE was used to address this issue, but managing the imbalance required careful tuning and model adjustments.
- **PCA-Anonymized Features**: The transformation of features using PCA made it difficult to interpret the real-world meaning of the features, complicating the interpretability of the model.
- **Overfitting Risk**: The risk of overfitting was mitigated by employing hyperparameter tuning and cross-validation to ensure the model generalized well to unseen data.
- **Computational Load**: Training the model and evaluating it with SHAP plots were resource-intensive, requiring substantial computational power.

**8. Discussion**

The project successfully demonstrated how data mining and machine learning techniques could be applied to credit card fraud detection. The combination of preprocessing, oversampling, and robust classifiers yielded a model with impressive recall and precision. While PCA-transformed features posed some interpretability challenges, the model's application remains highly valuable for fraud detection in other financial datasets. KMeans clustering, although not directly useful for fraud prediction, provided useful insights into the structure of the data and suggested potential future areas for improvement, such as using unsupervised anomaly detection techniques.

**9. Conclusion**

Credit card fraud is an ever-growing issue in the digital age, and data mining techniques offer an effective approach to combating it. This project successfully implemented a full pipeline from preprocessing to model training, evaluation, and interpretation. The results show that machine learning, particularly Random Forest, can provide highly accurate predictions for detecting fraudulent credit card transactions. The project met all requirements for the CS 619 course and highlighted the importance of handling class imbalance and ensuring model interpretability.

**Key Takeaways:**

- Handling class imbalance is essential for realistic fraud detection.
- Model interpretability is crucial for stakeholder adoption.
- SHAP provides transparency to otherwise opaque machine learning models.

**Future enhancements may include:**

- Developing **real-time fraud detection systems** to detect fraudulent transactions as they occur.
- Exploring **ensemble learning** techniques to combine the strengths of multiple models.
- Implementing **neural networks** for modeling complex patterns and behaviors associated with fraud.
-

**References**


1. Kaggle Dataset: https://www.kaggle.com/mlg-ulb/creditcardfraud
2. Chawla et al. (2002), "SMOTE: Synthetic Minority Over-sampling Technique"
3. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. NIPS.