Credit Card Fraud Analysis and Detection

Final Report
1924 - Fall 2024 - Foundations of Data Science
9th December 2024
Group 6

Group Members: Romaisa Nadeem, Megan Ciglich, Natalia Lebedeva, Markus Jeganathan, Onesime Dianganzi

Introduction

In today's digital age, fraudulent credit card transactions can pose a significant challenge to financial security. This project aims to analyze key factors of fraudulent transactions and build a reliable classification model to differentiate between fraudulent and authentic transactions.

The primary objective of our project is to explore and analyze the features in our dataset that indicate fraudulent transactions. Based on the insights gained from our data analysis, we will implement and evaluate a classification model to validate the predictive power of these features in distinguishing between authentic and fraudulent transactions.

Dataset Overview

The dataset was freely available on Kaggle and was downloaded without additional permissions after agreeing to the terms of use. The link to the dataset used in this project can be found here.

This dataset contains 284,807 transactions from European cardholders over two days in September 2013. It consists of 30 features.

- **Principal Component Features (V1, V2, ..., V28): T**hese are anonymized, PCA-transformed variables. Due to confidentiality reasons, the exact nature of these features was not disclosed.
- **Time:** Seconds elapsed since the first transaction, providing a temporal perspective
- **Amount:** The transaction amount, crucial for cost-sensitive fraud detection
- **Class:** This is the target variable, indicating whether a transaction is fraudulent (1) or authentic (0).

Hypothesis

We hypothesized that certain features, likely a subset of the PCA-transformed components (V1-V28), along with 'Amount', and 'Time' are more predictive of fraudulent behaviour. For example, transactions with anomalous amounts or occurring during specific time periods may have a higher likelihood of being fraudulent. By analyzing these features and uncovering patterns, we expect to identify trends that distinguish fraud from legitimate transactions.

Approach

Using our dataset, we analyzed which features most effectively indicate fraudulent behavior. We conducted data analysis and explored patterns in the PCA-transformed components (V1-V28) and the features "Amount" and "Time" to determine if any values or combinations of values are consistently associated with fraud. After a thorough data analysis, we built, evaluated and compared fraud detection across two classification models.

Our approach included three individual parts as follows:

- Part 1: Exploring and Refining the Dataset
- Part 2: Data Analysis and Visualization
- Part 3: Classification Model Training and Evaluation

In Part 1: Exploring and Refining the Dataset, we essentially loaded the dataset, examined its size and structure and checked for missing and duplicate values. We also computed statistical summaries for the features to understand their distributions and prepared the data for subsequent analysis.

For Part 2: Data Analysis and Visualization, we prepared a series of questions to better understand the features of our data with our key variable, "class". We employed both statistical and visualization tools to answer these questions and investigate patterns within our dataset.

In Part 3: Classification Model Training and Evaluation, we built and evaluated classification models using logistic regression and KNN to detect fraud. The two models were compared for overall performance.

Data Analysis and Findings

Our dataset is fairly large, with 284, 807 rows and 31 columns. It had a consistent format and did not require cleaning. There were no missing values, and duplicate values were dropped. Outliers were present in both authentic and fraudulent transactions, which were included to maintain data integrity and to preserve its real-life applicability.

We asked the following questions for our data analysis, findings and observations for which are detailed below.

How many authentic and fraudulent transactions are in the dataset?

As a first step, we wanted to determine the distribution of authentic and fraudulent transactions in our dataset. A pie-chart, shown below, was used to visualize the distribution between the two classes.

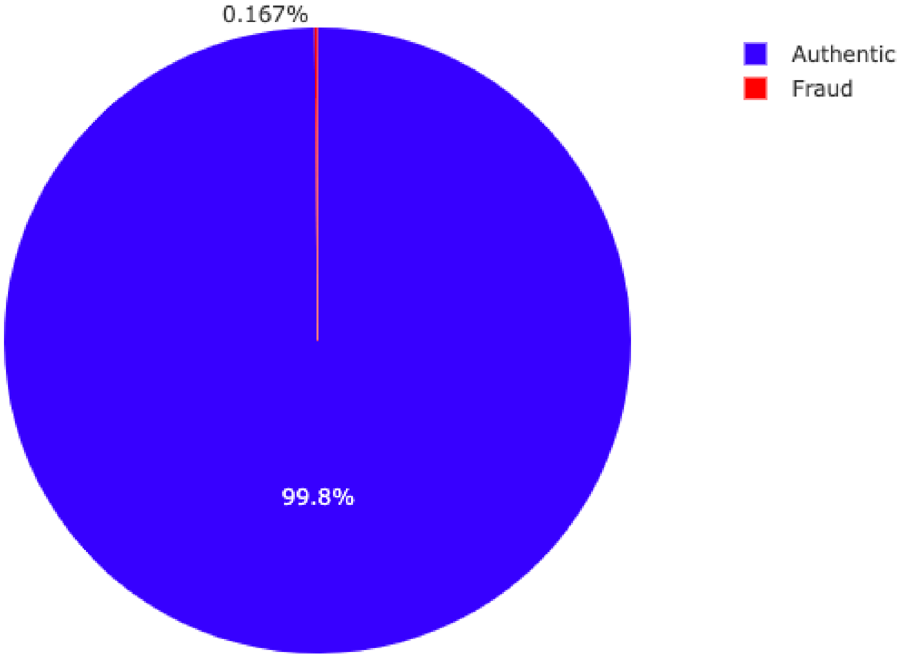


Figure 1: *Class Distribution of Fraud Detection Dataset: Authentic (99.8%) vs Fraud (0.167%)*

The pie chart reinforces the imbalance within our dataset, with only 0.167% of transactions being fraudulent and 99.8% being authentic. Our next step was feature analysis.

Are there any strong correlations within the features in our dataset that could indicate multicollinearity?

To investigate inter-features correlations within our dataset, we created heatmaps, as shown in figure 2, to visualize the correlation matrix of the dataset.

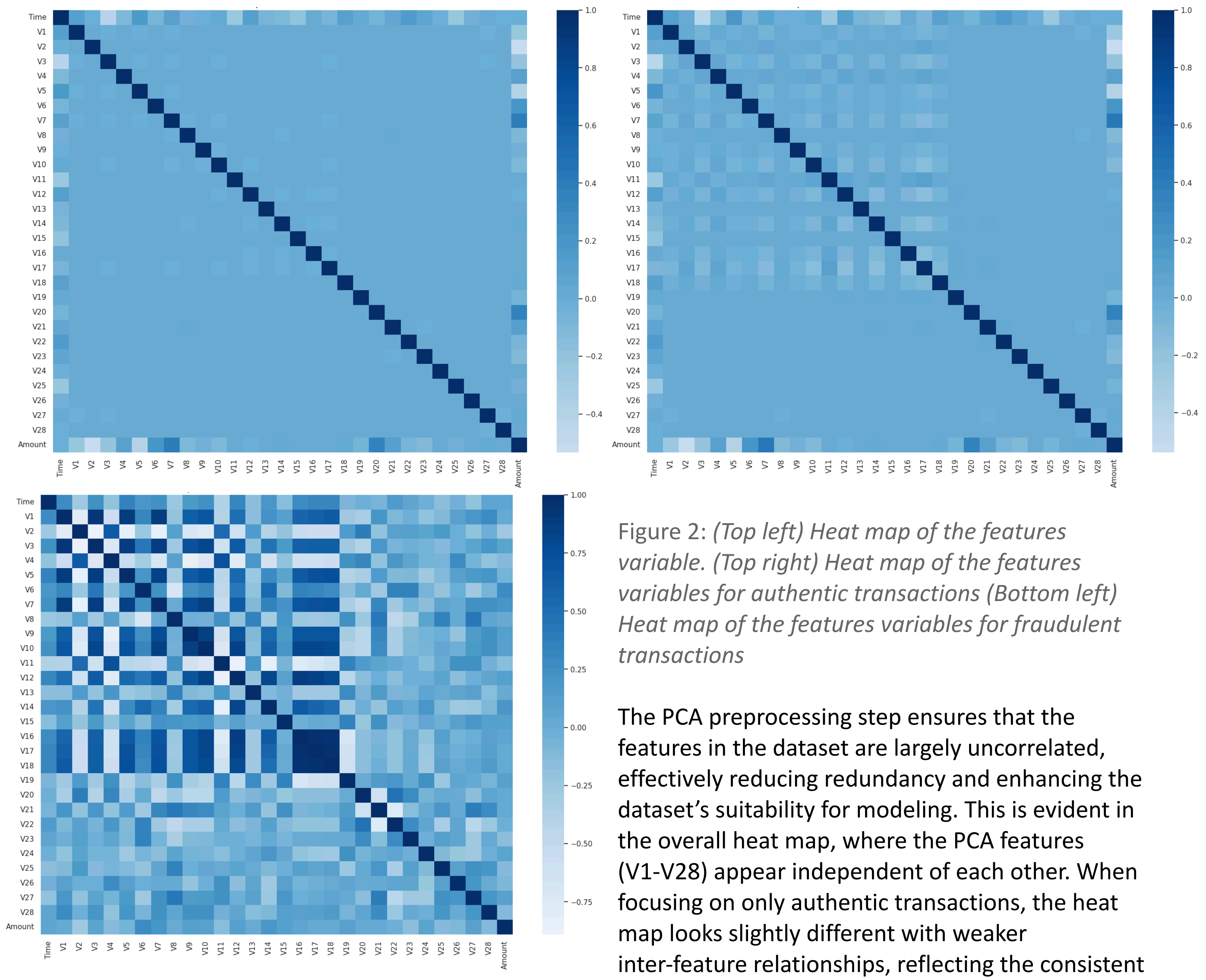


Figure 2: *(Top left) Heat map of the features variable. (Top right) Heat map of the features variables for authentic transactions (Bottom left) Heat map of the features variables for fraudulent transactions*

The PCA preprocessing step ensures that the features in the dataset are largely uncorrelated, effectively reducing redundancy and enhancing the dataset's suitability for modeling. This is evident in the overall heat map, where the PCA features (V1-V28) appear independent of each other. When focusing on only authentic transactions, the heat map looks slightly different with weaker inter-feature relationships, reflecting the consistent patterns of authentic transactions. In comparison, the heat map for only fraudulent transactions looks very different, displaying distinct and stronger correlations inter-features, suggesting unique behavioral patterns in fraudulent activity.

What are the correlation coefficients between each feature in the dataset and the 'Class' variable (authentic vs. fraudulent transactions)?

Continuing with our feature analysis, we calculated correlation coefficients and visualized the relationship of the "Class" variable with each feature of our dataset. A bar plot, visualizing the most prominent positive and negative correlations can be seen below.

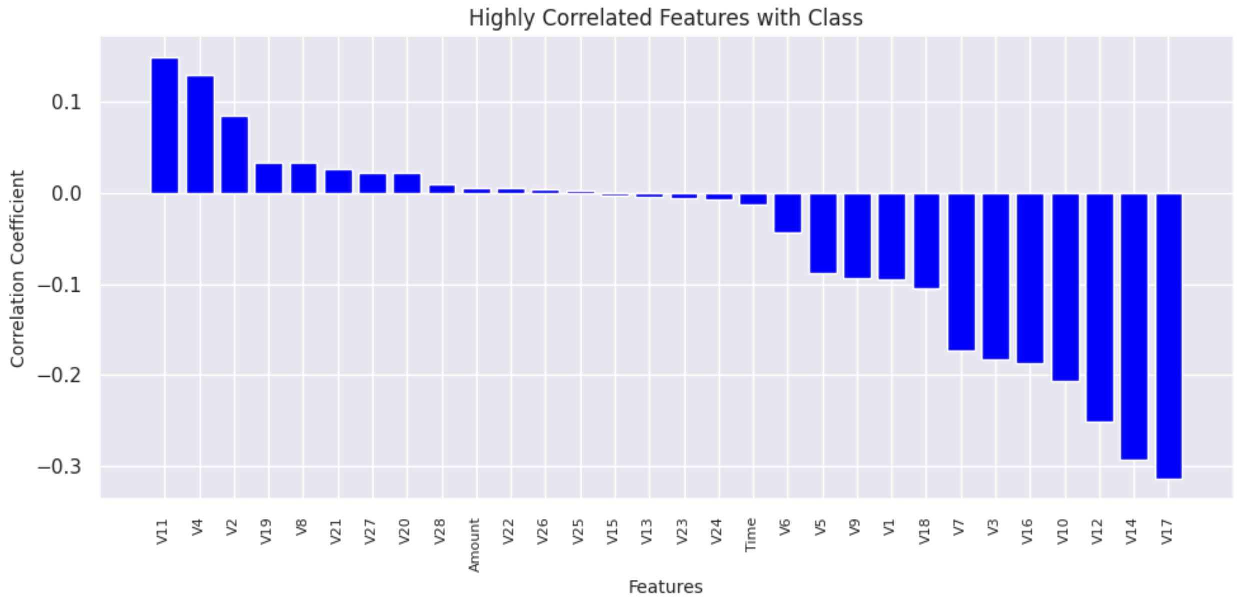


Figure 3: Correlation of features with class: key indicators of fraudulent and authentic transactions.

There is no 'strong' linear correlation - PCA transformation of the features generates new components that are optimized to catch variance but not necessarily to preserve the relationships with the target variable. Despite the PCA transformation, there are some apparent patterns.

Features V11, V4, and V2 have positive correlations with the class variable suggesting that they are likely associated with fraudulent transactions .

Features V17, V14, and V12 show negative correlation with class, suggesting their relevance to authentic transactions.

It is also notable that 'Amount' and 'Time' are not PCA transformed features, but both do not show any linear correlation with either authentic or fraudulent transactions. This suggests that while there is no linear correlation, their relationship with our target variable (Class) may be complex and/or influenced by non-linear patterns.

Which PCA components (features) are significantly different between authentic and fraudulent transactions?

To determine if the PCA components differ significantly between the two transaction types, we conducted a z-score analysis with a threshold of 3.37 (p-value ≈ 0.001, 99.95% confidence level). As identified by z-score analysis, significant PCA features identified include **V1–V12, V14, V16–V20, and V23–V28**. This indicates these features are particularly different between fraudulent and authentic transactions. In comparison, features **V13, V15, V22, V25, and V26** were determined to be insignificant, suggesting little difference between the two classes.

What are the temporal patterns in transaction activity for authentic versus fraudulent transactions?

Our dataset had a column for “time”, which were seconds from which the very first data observation took place. We converted this to hours of the day to investigate temporal patterns in transaction activity for authentic and fraudulent transactions. Figure 4, shows the hourly distribution of fraudulent and authentic transactions using line and bar plots for visualization, highlighting distinct temporal behaviors between the two classes.

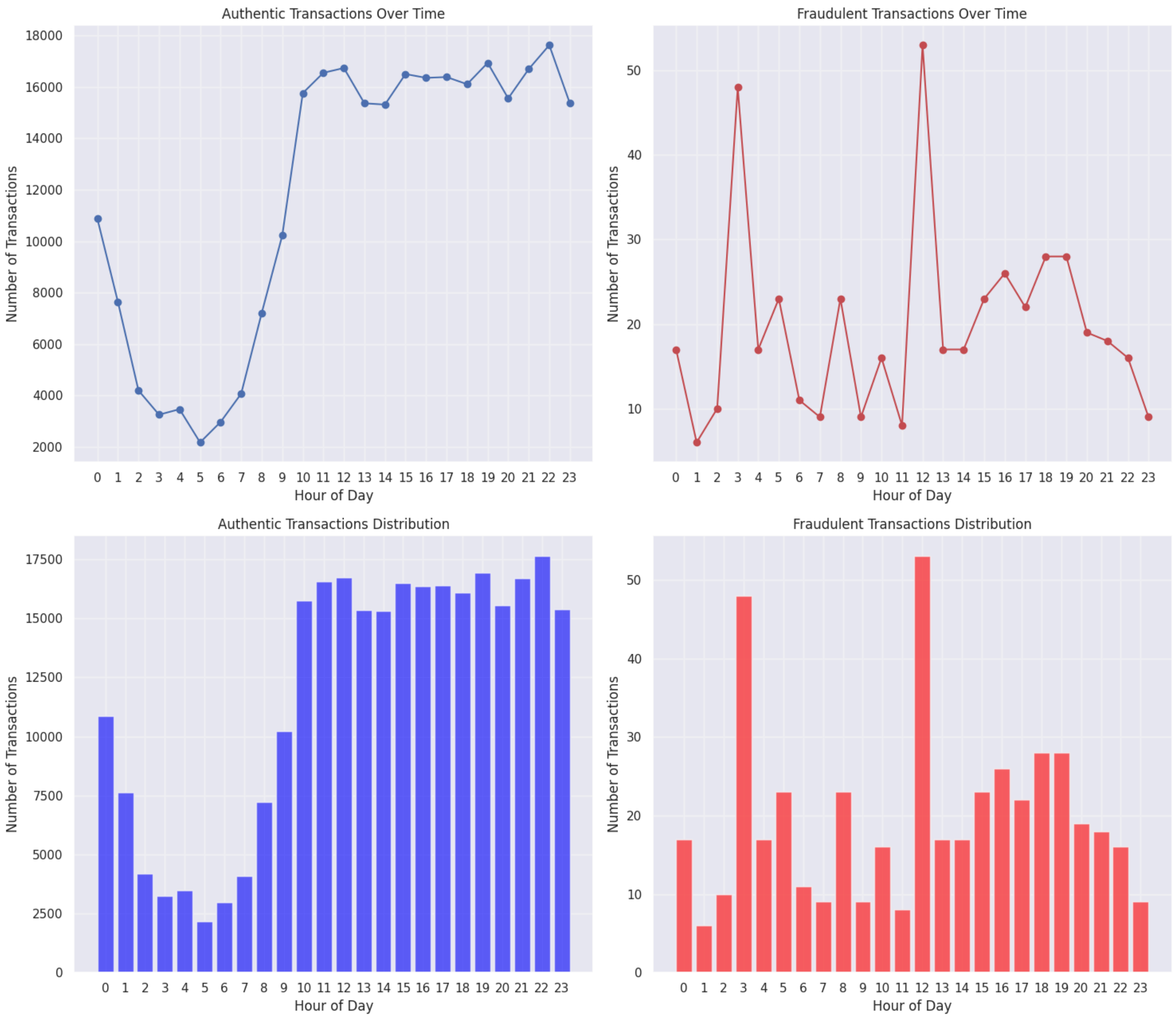


Figure 4: Hourly Distribution of Transactions: Authentic transactions show consistent patterns, while fraudulent transactions show irregular patterns.

As seen in figure 4, our analysis of transaction activity by hour revealed very distinct patterns. Authentic transactions show consistent patterns with lower activity in early morning hours. Authentic transactions were overall high from 9 AM to 11 PM, peaking at 10 PM, and showed a sharp decline in activity during early morning hours (12 AM–6 AM), with the lowest activity being at 5 AM.

However, fraudulent transactions showed an inconsistent pattern with two prominent spikes at 3:00 AM and 12:00 PM. Our findings are very relevant to real-world scenarios as well - with fraudulent transactions generally remaining sparse - fewer than 30 transactions per hour for most hours of the day.

Do fraudulent transactions occur during high or low transaction periods?

To further our analysis and to build on relevant temporal patterns, we examined whether fraudulent transactions occur during periods of high or low transaction activity within our dataset. The total of authentic and fraudulent transactions were analyzed across different hours of the day to detect any patterns in fraudulent behavior. We specifically wanted to investigate if fraudulent activity is more likely to align with high transaction periods, such as business hours, or with low transaction periods, like early morning hours.

To address the imbalance in our data (where only 0.167% of transactions are fraudulent), we utilized Min-Max Scaling which normalizes the number of overall transactions and fraudulent transactions. This normalization was then visualized (see Figure 7). Figure 7 highlights whether fraudulent transactions align with high or low transaction hours.

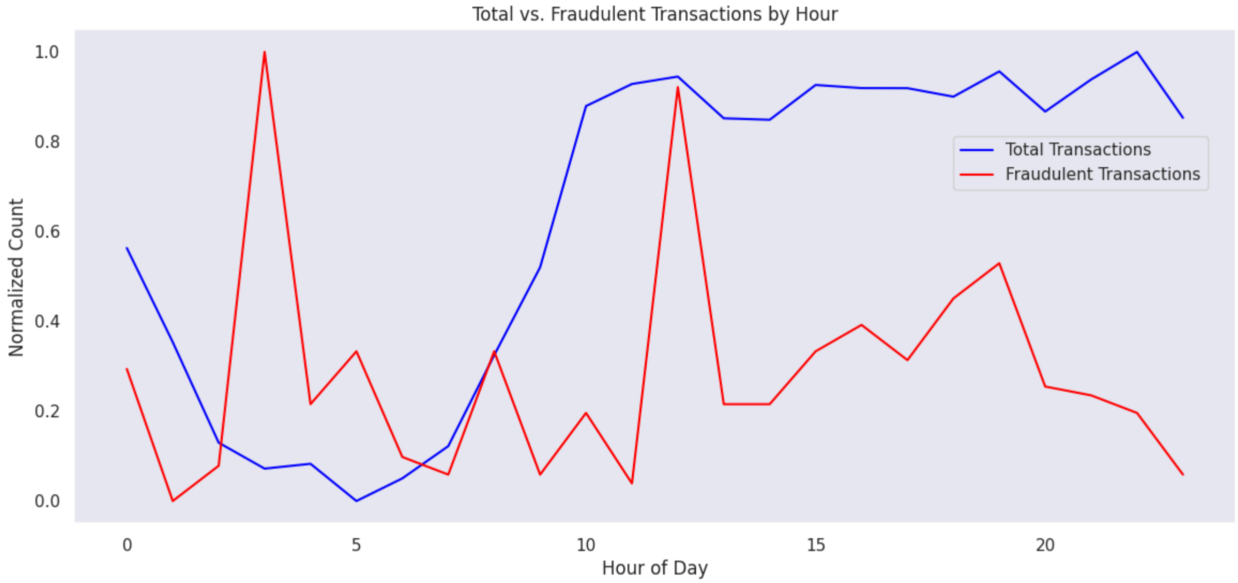


Figure 5: Normalized hourly distribution: Fraudulent transactions exhibit irregular spikes during both, high and low transaction periods, where as total transactions follow a consistent pattern

As visualized in figure 5, total transactions follow a predictable daily routine. This aligns with our finding of authentic transactions following consistent patterns, given that 99.8% of our data is authentic transactions.

We already know that fraudulent transactions have a sporadic behaviour and inconsistent patterns, but this can be especially seen in Figure 6. Irregular spikes can be seen in both, high transaction periods and well as low transaction periods.

Spikes in fraudulent activity in **low** transaction periods happen during 3 AM, 4 AM, 5 AM.

Spikes in fraudulent activity during **high** transaction periods happen at 12 PM.

Interestingly, during early morning hours, when total transactions seem to significantly drop, fraudulent transactions seem to rise, suggesting a potential relationship between overall low activity and fraudulent behaviour. They also seem to coincide with a higher activity period at 12 PM.

These patterns suggest that while fraudulent transactions are less frequent and more sporadic compared to the consistent pattern of authentic transactions, they likely exploit low activity periods while also aligning with high-activity periods to blend in.

How does the "fraud rate" vary across hours of the day?

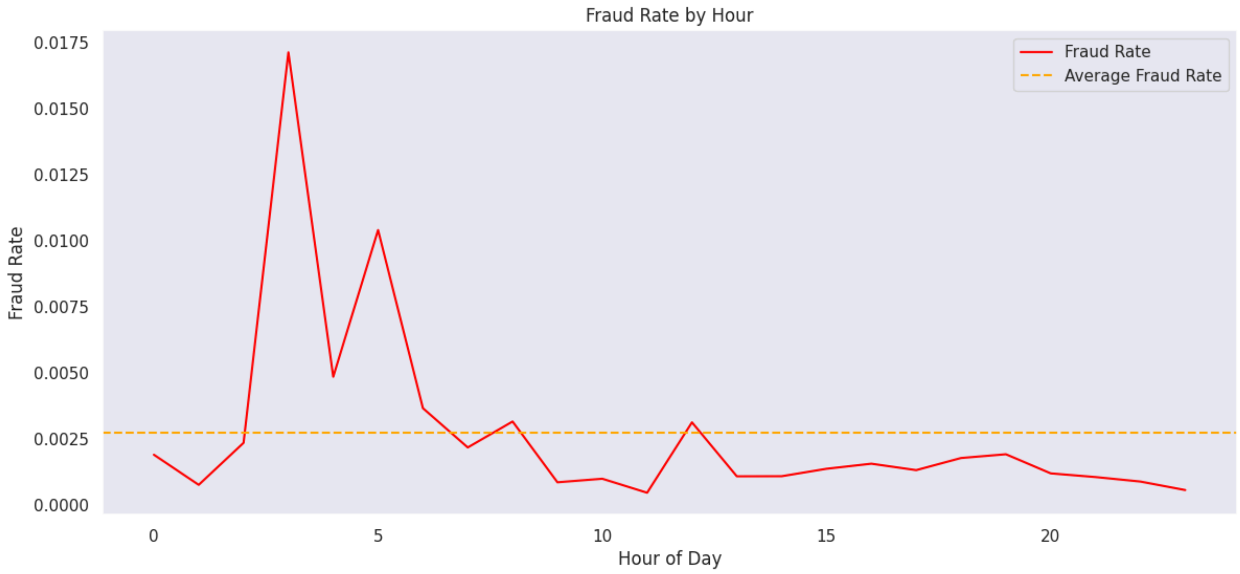


Figure 6: Fraud rate by hour: fraud rates peak during early morning hours, exceeding the average rate, and remain low during the rest of the day.

The fraud rate for each hour of the day was calculated by dividing the counts of fraudulent transactions by the counts of total transactions. Fraud by hour was then plotted for visualization.

As evident in figure 6, fraud rate peaks at 3 AM, being significantly higher than average, with higher rates continuing from 3 AM to 5 AM, indicating overall high fraudulent activity during early morning hours.

A smaller spike can be seen at 12 PM, which also coincides with higher transaction volumes during business hours.

Except for the small spike at 12 PM, after 6 AM fraud rates consistently remain below average, with particularly minimal activity in the late afternoon and late night hours. Given that, fraud activity seems to be focused during very specific low and high activity transaction periods.

How do transaction amounts differ between authentic and fraudulent transactions across different hours of the day?

We decided to further investigate how transaction amounts may be different between authentic and fraudulent transactions during different times of the day. Transaction amounts were standardized and a violin plot was used to visualize the distribution by class by hour of the day.

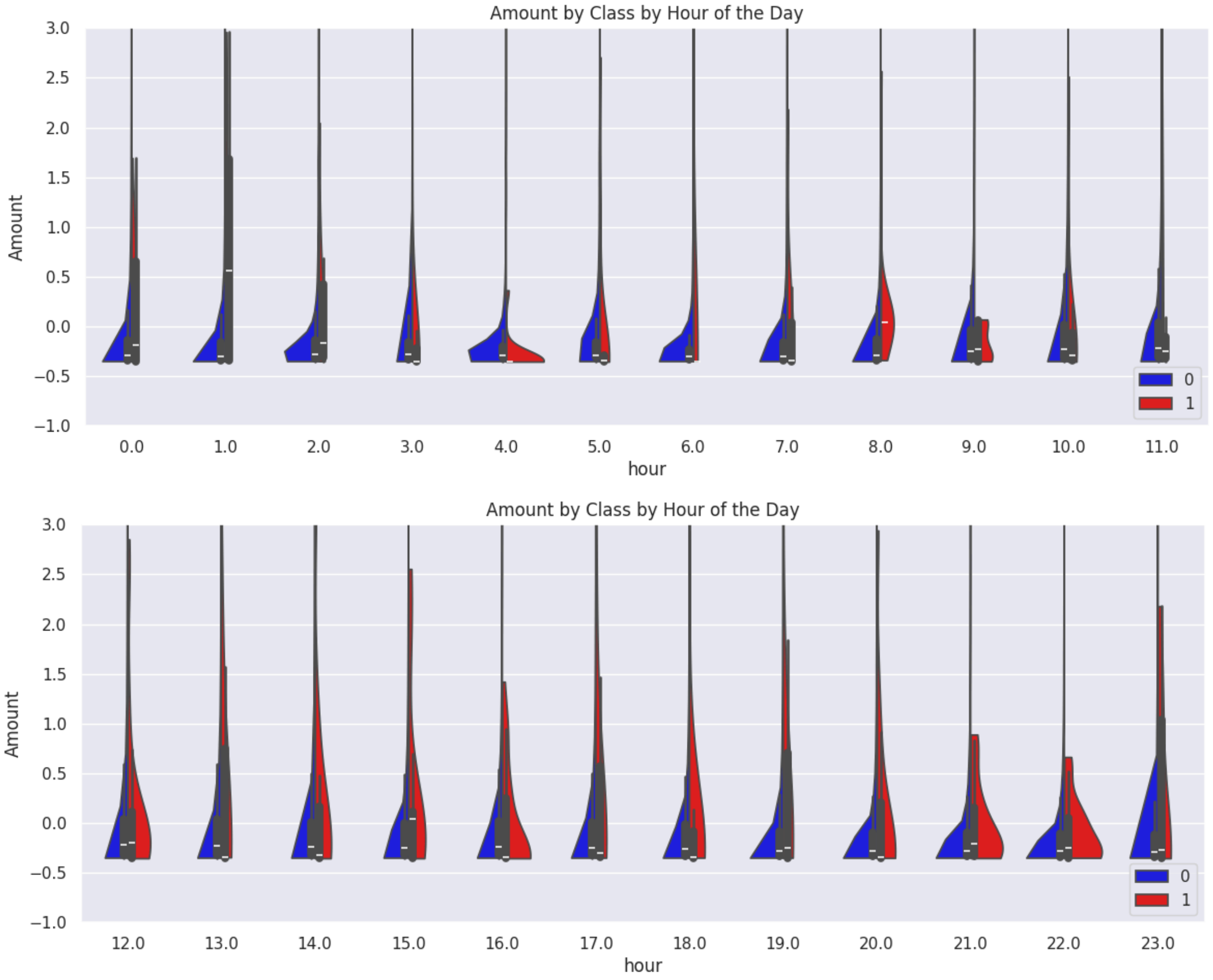


Figure 7: Hourly Distribution of Transactions Amounts by Class

As seen above in Figure 7, the density patterns for fraudulent transactions (red) seem to vary significantly, while authentic transactions (blue) show more consistent patterns. This was also observed earlier, in Figure 5, where the number of transactions for both classes followed specific time patterns.

For many hours of the day, the amount distribution for the type transaction classes overlaps, suggesting that the amounts are similar.

From midnight to early morning, during 12 AM to 5 AM, both transaction classes show a smaller distribution of high amount values, with most of the transactions being low-value amounts.

Between 12 PM and 11 PM, fraudulent transactions appear to have more variability in the transaction amounts, potentially suggesting high amount values during these times. It is very interesting to note that while fraudulent transactions generally peak during early morning hours, the amount remains relatively low and is instead more varied in the afternoon and late night hours.

What are the differences in the distribution of transaction amounts between authentic and fraudulent transactions? Particularly, do fraudulent transactions have higher amounts?

We further performed descriptive statistics, which revealed notable differences between authentic and fraudulent transactions. A snapshot of our descriptive statistics can be seen in appendix A, figure 11. Fraudulent transactions had a higher mean transaction amount ($123.87 vs. $88.41), but their median value ($9.82) was much lower than that of authentic transactions ($22.00). This suggests that while fraudulent transactions sometimes involve larger sums, most of them are for smaller amounts. Most fraudulent transactions ranged from $1 to $105.89, as observed by our 25th and 75th quartiles. Although infrequent, there are larger sums of up to $2125.87.

Both classes show variability, with very similar standard deviations (fraudulent: $260.21, authentic: $250.37), though fraudulent transactions show a slightly larger range. The maximum value for authentic transactions ($25,691.16) was much higher than that of fraudulent transactions ($2125.87), and outliers were more prevalent among authentic transactions.

Patterns noticed with fraudulent transaction amounts could definitely help flag fraud activity. For instance, more common and smaller transactions of $1 to $105 could be closely monitored during early mornings, identified to be high-risk periods. Larger transactions that are less frequent could be monitored during high activity timing

A boxplot below visualizing the distribution of transaction amounts in authentic and fraudulent transactions.

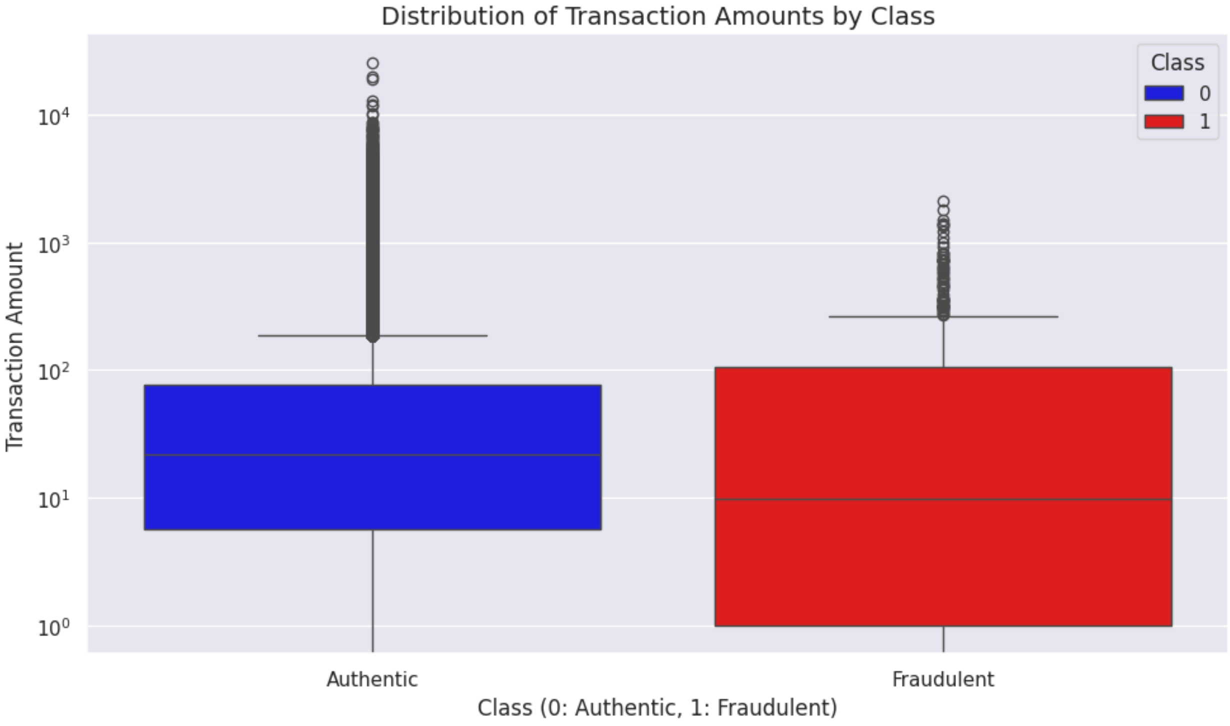


Figure 8: Boxplot comparing transaction amounts: fraudulent transactions show lower medians but higher variability than authentic transactions

Classification Model Training and Evaluation

Part 3 of our approach involved building a classification model and evaluating its performance.

Pre-processing our data

Our data was split into an 80:20 split for training and testing, with stratification applied to ensure that the class distribution in both sets also reflects the original dataset. The dataset's high imbalance, with 99.8% of transactions being authentic and only 0.167% classified as fraudulent poses a challenge for machine learning models. These models often prioritize optimizing overall accuracy, which can lead to poor detection of the minority class. For instance, a model that predicts all transactions as authentic would achieve an impressive accuracy of 99.8%, but it would fail to identify fraudulent transactions, making it ineffective for its intended purpose.

Handling class imbalance with SMOTE

To address the imbalance, we implemented the Synthetic Minority Oversampling Technique (SMOTE), which synthesizes new examples of the underrepresented fraudulent class. Unlike simple oversampling, which duplicates existing examples, SMOTE generates synthetic examples by interpolating between actual fraudulent transactions. This approach introduces diversity into the minority class, enabling the model to learn generalizable patterns of fraud without overfitting.

SMOTE was applied only to the training set to ensure that the test set remained unaltered, preserving real-world class imbalance for evaluation purposes. After applying SMOTE, the training set was balanced, with 226,652 fraudulent transactions synthesized to match the number of legitimate transactions. This balanced dataset allowed the model to learn effectively from both classes, improving its ability to identify fraudulent transactions while minimizing bias toward the majority class.

Feature Scaling

In addition to addressing class imbalance, the features within our dataset were scaled so that all variables contribute equally to the model. Our dataset included features whose range varied widely, for example, PCA-transformed components V1–V28, transaction amount, and time. Without scaling, these features with large ranges, such as transaction amount, might dominate the learning process of the model, thus making biased predictions. We used the StandardScaler for normalization of feature values, scaling them within one range and improving the performance and interpretability of the model.

Final pre-processing steps

The combination of SMOTE with feature scaling ensured that our model was trained on a balanced and normalized dataset. In this way, the model could learn effectively from both authentic and fraudulent transactions from the training data, while a stratified test set allows for a realistic evaluation of its performance under imbalanced conditions.

Logistic regression: model development and evaluation metrics

Overall, our logistic regression model performed well, achieving a 85.26% recall and 99.12% accuracy on the testing set.

The classification report, created to evaluate the model, highlighted the strengths and weaknesses of our model. For authentic transactions, (Class 0), the model performed with near-perfect precision and recall at 1.00 each; this shows how good the model is at correctly classifying legitimate transactions reliably. On the other hand, for fraudulent transactions, (Class 1), the model showed a low score for precision of 0.14, indicating a high false-positive rate, but yet still showed strong recall of 0.85, catching most of the fraudulent cases.

The confusion matrix, as shown below in figure 8 provided further insights.

- True Positives (TP): 81 fraudulent transactions correctly identified.
- False Negatives (FN): 14 fraudulent transactions missed.
- False Positives (FP): 482 authentic transactions incorrectly flagged as fraudulent.
- True Negatives (TN): 56,169 authentic transactions correctly identified.

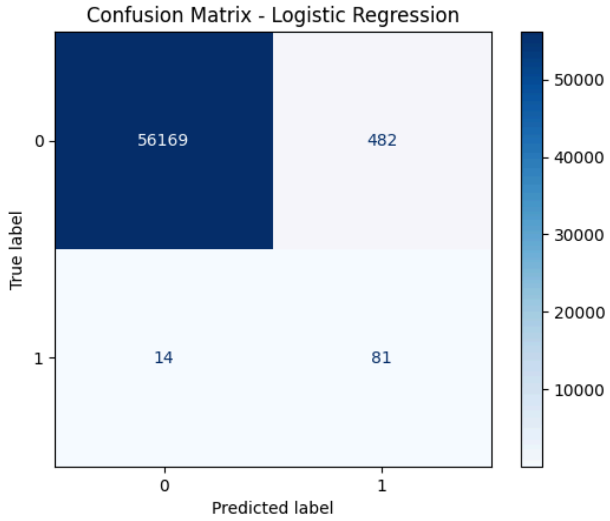


Figure 9: Confusion matrix for logistic regression model

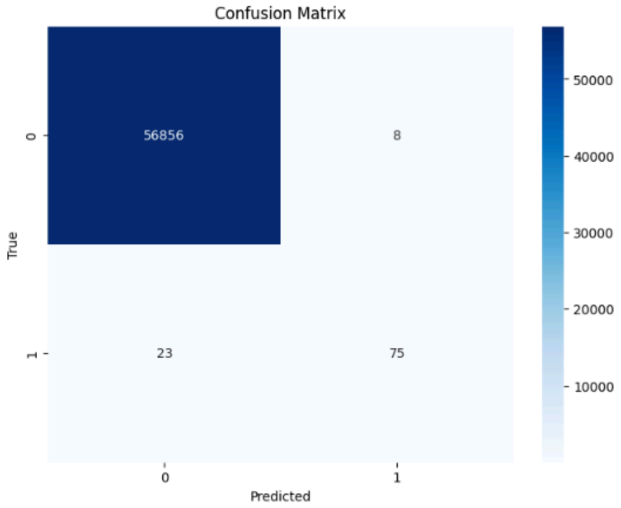
KNN regression: model development and evaluation metrics

A KNN model was developed using the same training and testing splits. Our KNN model had very good performance on authentic transactions and robust results for fraudulent transactions.

A classification report and a confusion matrix provided insights into the model's performance. For authentic transactions (Class 0), the KNN model had a precision, recall and F-1 score of 1.00 on 56,864 transactions. This shows that the model is very reliable in classifying authentic transactions. As for the fraudulent transactions (Class 1), the model had a precision of 0.90, thereby reducing false positives, and a recall of 0.77, identifying most of the fraudulent cases. The F-1 score was 0.83.

A confusion matrix, as seen in figure 8, provides the following insights:

- **True Positives (TP):** 75 fraudulent transactions correctly identified.
- **False Negatives (FN):** 23 fraudulent transactions missed.
- **False Positives (FP):** 8 authentic transactions incorrectly flagged as fraudulent.
- **True Negatives (TN):** 56,856 authentic transactions correctly identified.



Real world applicability: Logistic regression vs. K-Nearest Neighbors (KNN)

Both the logistic regression and KNN models demonstrated strong performance, each with their own strengths and weaknesses, making them suitable for different real-world applications. Logistic regression had better recall on fraudulent transactions - it performed well at identifying most cases of fraud, which is particularly important in a real-world scenario. On the other hand, it resulted in many false positives, which may lead to unwanted alerts in practical applications.

The KNN model in comparison shows a big decrease in false positives and still maintains a good recall. It did miss more fraudulent transactions relative to logistic regression, which may be a limitation in practical applications, since missing fraudulent transactions carries a high cost.

In real-world fraud detection systems, the choice between these models depends on the priorities of its application. If identifying fraudulent cases is a priority, then logistic regression may be favoured due to its recall. If reducing false positives is favoured due to unnecessary alerts, KNN offers an efficient solution.

Challenges

Throughout our project, working with a highly imbalanced dataset presented several challenges. The proper visualization of our data was tricky because of the very high prevalence of the majority class. Traditional visualizations techniques lead misleading interpretations where the minority class was not fairly represented. To overcome this, we applied logarithmic scales to balance the scale of our visual representations. This can be seen in Figure 5. Transaction amounts were standardized for better visualization in Figure 6. Counts were normalized to better illustrate Figure 7. We also employed sophisticated visualization tools such as heat maps for correlation analysis, box plots and violin plots to overcome the visualization challenges that arise with a highly skewed dataset and discern patterns representative of fraudulent transactions.

Another significant challenge that we encountered was the interpretability of the PCA transformed features: by design, they are anonymized and are uncorrelated. Given the anonymization and the nature of PCA itself - it was challenging to understand what each component represents. The clear lack of definitions hindered our ability to directly link specific features with fraudulent transactions. To address this challenge, we utilized correlation analysis as well as z-score analysis. Furthermore, due to such an imbalance in our dataset, it was important that we implemented resampling techniques such as SMOTE, particularly to improve the representation of the minority class and to reduce the model being biased towards the majority class.

Conclusion

In this modernized and tech driven world, we are able to identify features and create classification models to validate the predictive power for identifying fraudulent activities in credit card transactions. Throughout this analysis we were able to gain insight about the relationships between PCA-transformed components (V1-V28), Time, Amount and the Class of each transaction. In addition, we compared classification models using logistic regression and KNN to validate the predictive powers of these features. Although there were many findings drawn from our various comparisons and analysis, we did observe very particular trends that align with identification of fraudulent credit card transactions.

We noticed that there was a larger inter-correlation between the individual PCA transformed components (V1-V28) when comparing the fraudulent data with the non fraudulent data, which was an early indicator of an apparent relationship. When calculating the correlation coefficients of the PCA-transformed components (V1-V28) features and the transaction class, there wasn't a strong linear correlation between the two, which is expected since they are PCA transformed. However, there was an indicator that components V11, V4 and V2 have a positive correlation with class, which indicated that they are strongly related with detecting fraudulent transactions. While components V17, V14 and V12 have a negative correlation with class, indicating that they are relevant to detecting authentic transactions. Although the nature of the features were not disclosed, we are able to identify that various PCA transformed components (V1-V28) were more correlated than others in the identification of fraud transactions.

We found that time did not have a linear correlation with class but instead has very specific-temporal patterns that correlate with transaction type. Specifically, we found that most authentic transactions happened between 9AM to 11PM, with the highest activity at 10PM. In contrast, the least transactions happened between 12AM to 6AM, with the lowest point at 5AM. On the other hand, we found that the highest fraudulent transactions happened at 3AM and 12PM. For the remaining hours of the day, fraudulent activities were fewer than 30 transactions per hour. During early morning hours specifically 3AM to 5AM, when total transactions seem to significantly drop, fraudulent transactions seem to rise. However, after 6AM the fraud rates consistently remain below average except for a spike at 12PM. Therefore, after examining the temporal pattern, we were able to conclude that most fraud transactions from the data set happened during 3AM to 5AM and 12PM.

There was an apparent difference between the transaction amounts of fraudulent and authentic transactions. To begin with, the average transaction amount for fraudulent transactions was $123.87, while the average authentic transaction amount was $88.41. The respective standard deviation is $260.211 (fraud) and $250.379 (non fraud). This shows that fraudulent transactions average a higher amount with a slightly larger variability when compared with the authentic transactions. Additionally, the median for fraud transaction amounts was $9.82 and $22.00 for the authentic transactions. This shows that while fraudulent transactions sometimes involve larger sums, most of them are for smaller amounts. Finally, we are able to see that authentic transaction amounts range from [$0, $25691.16]. Fraudulent transaction amounts range from [$0, $2125.87], with most of them falling between $1 to $105. To sum up, we can see that fraud transactions have less extreme transaction amounts, with occasional, less frequent amounts as high as $2125.87.

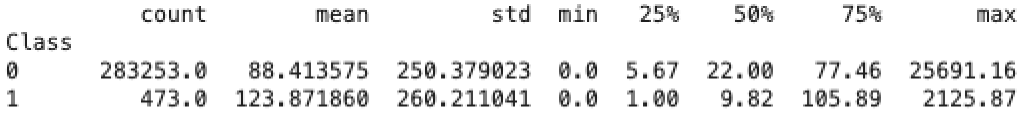
Furthermore, we found that there was an apparent relationship between the transaction amount, fraudulent activity and time of day. Firstly, there was no straightforward linear relationship between the transaction amount and class. This suggests that their relationship is more complex. Secondly, fraud transaction amounts vary significantly throughout the day while authentic transaction amounts show a more consistent pattern. Thirdly, from 12AM to 5AM, both types of transactions are low-value amounts. Fourthly, from 12PM to 11PM, fraudulent transactions have a higher variability in amount showing large amounts are being transacted. To summarize, we noticed that there were small amount-values of fraud activity during the early morning hours, and larger fraud transaction amount-values during busy hours.

After our analysis, we created two classification models to evaluate the performance of the features ability to detect fraudulent transactions. We used SMOTE to create synthetic new examples of the underrepresented fraudulent class and used standard scalar normalize feature values to improve interpretability. Both of our classification models had their own unique strengths and weaknesses. Our logistic regression model identified more fraudulent transactions, but resulted in false positives which can cause customers to receive a false alert. On the other hand, our KNN model showed a big decrease in false positives but identified less fraudulent transactions in comparison, which can be costly in real-world applications. The choice between the two models depends on their priorities for its applications.

For our next steps, testing with other models, such as random forest or neural networks may improve fraud detection. Fine-tuning the classification threshold as well as addition of other features like precise demographic details and consumer behaviour patterns would also significantly improve detection of fraudulent activity. Furthermore, testing the model in real-world scenarios could refine its performance and improve its applicability.

To conclude, we found the temporal patterns that align with both, the number of fraudulent transactions as well as their amount value, we would recommend financial institutions to implement time-specific monitoring systems for fraud detection. Lower fraudulent transaction amounts in the early morning hours could be focused for closer monitoring. Similarly, fraud activity with higher fraudulent transaction amount variability could be closely monitored during peak hours. Additionally, if the nature of the PCA components was disclosed, we would recommend to use PCA transformed components (V11, V4 and V2) as it has a positive correlation with class (transaction type), which indicates that they are relevant to detecting fraudulent transactions. Finally, after comparing both classification models, in a real-world application, logistic regression may be a good choice if identifying fraudulent transactions is a top priority while KNN may be more suitable if reducing false and unnecessary alerts is a priority.

Appendix A



```
          count        mean         std  min   25%    50%     75%       max
Class
0      283253.0   88.413575  250.379023  0.0  5.67  22.00   77.46  25691.16
1         473.0  123.871860  260.211041  0.0  1.00   9.82  105.89   2125.87
```

Figure 11: Descriptive statistics for transaction amounts for Class 0 (authentic) and Class 1 (fraudulent)

References

Kaggle. (2018). *Credit Card Fraud Detection*. Www.kaggle.com.

https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud