**Module 7: Portfolio Project- Paper**

Alexander Verbrugge

Colorado State University- Global

24WA-MIS581-01: Capstone- Business Intelligence and Data Analytics

Dr. Jamia Mills

Due Date: 5 January, 2025

**Abstract**

E-commerce fraud is an epidemic in today’s society and has become one of the toughest challenges in the digital economy with the increasing growth of online transactions and the complexity of fraudulent schemes. This study uses data analysis to help detect and mitigate e-commerce fraud. With the use of various data analysis techniques and modeling, it aims to uncover patterns that will lead to detecting fraudulent behavior. Statistical hypothesis testing was conducted to investigate the relationship between transactional features like payment methods and frequency of payments with the likelihood of fraud occurring. Using publicly available datasets, exploratory data analysis was done to uncover trends and anomalies. This was followed up with predictive analysis to find the trends. This research attempted to find actionable ideas for e-commerce organizations to reduce financial and customer losses and maintain customer satisfaction.

**Introduction**

E-commerce has grown exponentially over the last few years with companies like Etsy and Temu alongside quicker shipping that allows consumers to purchase goods from all over the world. The world is literally at almost everyone’s fingertips and *their* world is at the fingertips of fraudsters as the internet and e-commerce opens up everyone to the risk of potential fraudulent activities without their knowledge if they are not careful and or protect one’s self. The consumers often do not know their information is being used by someone else and the company does not know it’s not the correct person on the other side. There is an annual loss of an organization’s Gross Domestic Product (GDP) of 5% due to fraudulent behavior (Wang et al., 2024). One may not say that is not a lot, but when one is talking worldwide e-commerce, it can be in the millions, billions, or even trillions of dollars a year in fraudulent activities. E-commerce companies are at high risk of taking significant losses and equally important, loss of reputation and business due to these fraudulent transaction attempts. Detecting these fraudulent transactions in an efficient and accurate manner is crucial in order to minimize these losses and provide the customer with an amazing shopping experience to ensure they return and shop again.

**Objectives**

The objective of this study is to identify some key features and patterns in behavior that may indicate fraudulent transactions in e-commerce. With the use of statistical analysis and machine learning, the study looks to provide insight in which organizations can use to take action to help minimize the fraud risk at the same time lower the issues and inconvenience of the real, non-fraudulent customers.

**Overview**

The project will use two datasets called c*ustomer\_DF* and *cust\_transaction\_details* to analyze the past fraudulent and non-fraudulent transactions in order to find patterns in the dataset as it has a lot of data on transactions and the customers making the purchases. The amount of data provides a strong foundation for investigating meaningful insights within different kinds of information regarding transactions, method of payments, among other information in regards to the transactions that might indicate patterns. The research paper will outline two research questions and hypotheses. It will talk about the background and supporting details from other articles and other studies done around the subject of fraud and more specifically, e-commerce fraud. It will also talk about the methodologies and methods that will be used in the various test and the results of the various analyses to draw conclusions that hopefully will answer the research question and help solve the overall issue of detecting e-commerce fraud. When collecting the data and analyzing it, it is critical to keep in mind its limitations and ethical considerations. This paper to the limitations of the dataset and ethical implications through out the whole process and at times dropped unnecessary data like IP Address, physical addresses, and phone numbers in order to clean the data before analysis as it was not going to be looked at in this case. This was to ensure a responsible and transparent approach to the research.

**Research Questions and Hypotheses**

**Number of Payments vs. Fraud Research Questions and Hypotheses**

In order to figure out if there are any triggers that might signify a fraudulent transaction, or even a potential fraudulent transaction that would require more research, the following question was formed. *Is there a significant relationship between the number of payments and if it can predict fraud?* The element of time has always been a factor in the prediction of fraudulent behavior. Studies have shown that a good, non-fraudulent, user spends more time on reviewing products online before purchasing, while fraudulent would try to mirror the behavior, but will actually spend way less time reviewing the products before making a purchase (Yin et al., 2022). To see if this is true in regard to the amount of payments indication fraud occurring the following hypothesis was developed:  
**H0**: There is not a significant relationship between the number of payments and fraud being true.

**H1**:There is a significant relationship between the number of payments and fraud being true.

Finding the answer to this research question can help the organization determine further steps or strategies to mitigate fraud. For example, if there was to be a significant relationship, plans like keeping the number of payments to a minimum or requiring a certain number of payments for purchases over a given amount or implementing extra verification steps to ensure the legitimacy of the identity to the consumer to help reduce risk.

**Bitcoin/Apple Pay vs. Transaction Failing Questions and Hypotheses**

Having transactions failures could also be a sign of fraudulent activity. With the new coming of age, there have been concerns with having virtual wallets and/or cryptocurrency being used as a form of payment. Cryptocurrencies are supposed to use secure networks, electronic signatures, and encryption to make transactions or “cash out” without a third party such as banks (Hajr et al., 2023). People often wonder if it is safe to use, or in the way of cryptocurrency, is it stable enough where the organization would not lose money. It has become convenient to use, but it also comes with unique risks. From the end of 2020 to March 2021, over 7000 people became victims of cryptocurrency fraud worth over $80 billion (Kutera, 2022). The following research question that can be asked addresses these concerns. *Is there a significant relationship between the use of bitcoin or apple pay and the transaction failing?* If there is a strong or significant relationship between the payment method and transaction failing, then the organization can perform further risk assessments concerning the payment types and if they deem the payment methods too risky the organization can opt not to accept them. The following hypothesis can help test this.  
**H0**: There is not a significant relationship between the use of bitcoin or apple pay and the transaction failing.

**H1:**There is a significant relationship between the use of bitcoin or apple pay and the transaction failing.

A study has found that there was more than a 30% increase in online shoppers using Bitcoin and other cryptocurrencies for their purchases (The Intersection of Cryptocurrencies, 2023). Understanding the relationship can help the organization make informed decisions about balancing the convenience for customers and their forms of payment with the security and the prevention of fraud

**Literature Review**

**E-commerce and Fraud**

E-commerce has integrated itself in a large manner in the financials of all organizations and businesses. Whether they know it or not. If customers cannot find you on the internet, there is a good chance your company will not last long. Global retail stores made 4.28 trillion dollars in sales in 2020 while it is projected to do over 6.39 trillion in 2024 (Wu et al., 2024). In China, their Gross Merchandise Volume (GMV) change was an increase of one trillion dollars alone of transactions from April 2019 to March of 2020 and they tried to use detection techniques to prevent fraud on many fronts by identifying fraudster seeds (Zhang et al., 2022). Many different sources dispute exactly how much in dollars customers and organizations are losing per year exactly, but they all seem to agree that it is in the billions if not trillions of dollars a year. There are a lot of avenues that fraud can happen plays in to the statistics of e-commerce fraud and they are all not the same.

**Types of E-Commerce Fraud**

Three of the most common types of e-commerce fraud are refund fraud, account takeover fraud, and payment fraud. Within each fraud type are subcategories. Some of the more popular refund fraud schemes that get heard about the most includes a process called wardrobing and receipt fraud. Wardrobing is where someone buys a piece of clothing and wears it with the tag on and then tries to return it as if it was never worn to get the full refund back. Receipt fraud is where stolen or fakes receipts are used to request refunds for items that were never purchased at the store. This area has grown fast as artificial intelligence has gotten better. Fraudsters would use deepfakes to help create receipts and/or get around verification procedures in order to get a refund (*Artificial Intelligence to push e-commerce fraud to $107b.* 2024). AI can play a part in account takeover fraud as well. Account takeover is essentially someone taking over an existing account by purchasing stolen credentials on the dark web. This is being done to people’s wallets by being hacked into or the holder is being phished into releasing login details. The fraudster is then able to use those funds in a malicious manner. Fake or compromised accounts can be created with stole identities as well for a combo identity fraud and a version of account takeover. However, the most popular type of e-commerce fraud is payment fraud, or fraud where someone else uses credit card information, or other forms of payment, for personal gain. The most common way people use to make purchases online is by credit/debit card. With the increase in digital payment methods there, has without a doubt an increase in digital payment fraud (Chang et al., 2024). There are 2 main types of payment fraud, card present and card-not-present. Card-not-present makes up almost 75% of the total payment fraud in the United States with an increase from 57% in 2019 (McDonald, 2023). This is what most people think of when all they have to do is click the card they want to use for their purchases or for their subscriptions that charge them every month without having to click pay each time.

**Research Design**

**Methodology**

A mixed method approach was used in the research to investigate the impact of the various variables on the prediction of fraudulent transactions and the vulnerability of certain payment methods, specifically Apple Pay and Bitcoin. Doing this allows for a better and stronger approach by analyzing both quantitative and qualitative data sources for a more complete understanding of the data. The datasets were found through Kaggle and were previously collected. However, they were broken down to make for easier review and analysis. For an example, some aspects, like addresses, we split up to provide more information in case they were needed for additional analysis like identifying patterns or correlations that were not apparent at first. The qualitative information was taken from the customer filling out billing addresses, credit card/ payment information, and the device they used. The quantitative information was probably taken from the individual purchase themselves which includes the number of orders, the number of payments, and the number of transactions. These records could help identify the behavior and bring pattern to the light that could indicate fraud or identify the riskier payment methods. The analysis will include both sets of analysis to help strengthen the case by having one of the types of analysis (quantitative or qualitative) supported and validated by the other in its findings. If they do not agree, then further analysis could be done to determine the cause of the discrepancy. This can lead to more informed decisions and recommendations for improved fraud detection and lowering the risks for the organization and their customers.

**Methods**

In the *Number of Payments vs. Fraud* analysis, one of the important decisions is how to treat the Number of Payments variable as there is a lot of different numbers within the variables. Since there is flexibility between having individual counts of payments, they could also be set up as categorical (low, medium, high) just to cluster the number into a narrow few options to simplify the analysis and help increase the count per category which can improve the results of the test. Categorizing the data could make the results stronger and more robust. It may even make it easier to find patterns and fraudulent behaviors. For the latter, a Chi-Square could be run to see if there is a relationship between the fraud or non-fraud status and the newly created category of number of payments. This would provide an idea if there were a significant relationship between the frequency and range of the payments as a predictor of the likelihood of fraud occurring. However, on the other hand, if the number of payments was to stay numerical, then a logistical regression analysis could be performed as it would allow the continuous variable to be assessed and how it relates to the likelihood of a transaction being fraudulent. It can be assumed an organization will use machine learning models for fraud prediction R studio will be used in this paper for analysis. R Studio supports advanced analysis and machine learning algorithms, an ideal system to perform analysis for the data sets presented here and for Chi-square and logistical regression testing. The analysis will be using Synthetic Minority Oversampling Technique (SMOTE) to help the imbalance of non-fraud vs fraud counts, where fraudulent transactions will most likely will get greatly outnumbered by the valid transactions. SMOTE creates examples for the lesser class to make the sample larger a k-nearest values are calculated (Zou et al., 2018). This technique will ensure that the model is not biased toward predicting non-fraudulent transactions, allowing for more accurate fraud detection.

The *Bitcoin and Apple Pay vs. Transaction Failing* case will use R Studio for Chi-square and logistic regression as well to look into the relationship between all the payment methods (Bitcoin, Apple Pay, card, Paypal) and their transaction outcomes whether it be failed, pending, or fulfilled. A failed transaction can be a sign of a potential fraud attempt and therefore needs to be treated as so. A pending status may imply further information, may be necessary to ask for additional security questions or verification steps in order to complete the transaction. Once all the analysis has been run, then the Bitcoin and Apple Pay results could be compared to the more traditional payments, like credit/debit cards, to see if there is a significant difference. A decision tree could also be an option here as it can find the relationships between the various transaction characteristics and the failure results. The visualizations like histograms, box plots, and exploratory data analysis tools will mostly be done in SAS Studio as it allows for a better, cleaner understanding of the data represented and the data’s distribution visually. Doing this will make it easier to identify trends, anomalies, and/or patterns that might depict fraudulent behaviors. The hope is that using multiple statistical tests, machine learning techniques, along with data visualization will provide a in-depth analysis of the different factors that comes part of fraud detection and transaction failures that will lead to more of an effective fraud prevention strategies.

**Limitations**

There are some limitations of the data and the research process that should be acknowledged to help create a firm understanding of the results and what they may lead to in creating fraud prevention strategies and the fraud detection process an organization currently has. First, the description of the datasets on Kaggle were not entirely clear on the context and what each variable represented. This means that some logical assumptions were taken and used as interpretations during the analysis. Evry effort was made to ensure the assumptions were grounded with supporting facts, there is a chance that certain misinterpretations may tend to affect the analysis and therefore the conclusions and interpretations of the data. There may be some information in the data set that might cloud out the analysis, also known as noise, and be deemed unnecessary but may be the primary key that makes the dataset unique so it would be unwise to delete even if it does not play a direct role in the analysis. It may show any relationships that were not previously seen between other variables. One thing one should be aware of is that leaving noisy data in could also bring in unintended biases. Another limitation is the fact that the *customer\_DF* dataset is a smaller dataset size, especially when looking at fraud related transactions. The imbalance of fraud/non-fraud might pose an issue for an accurate model training if not treated and adjusted properly. It could lead to results in models that would be extremely biased towards the majority class, in this case non-fraud. Even though a technique like SMOTE has been used to decrease the issue in the past, it remains an issue when trying to generalize the results. Another sign of potential fraud is the frequency of purchases in a short period of time as if one wasn’t caring what they were ordering. However, the dataset does not have a measure of time like timestamp, dates, or seasons in which a calculation can be done to determine if a customer is purchasing “too fast”. This lack of data can introduce a sampling bias because the dataset may not fully represent all of the customer’s demographics. The bias may also prevent the analysis of the data set from being generalized into the real-world context.

**Ethical Considerations**

Since these datasets have a lot of personal data like physical addresses including military bases, phone numbers, IP Address, and emails to name a few, they need to be protected. This starts at the beginning with the data collection. There was no reason to collect the Social Security Numbers or indicative data on the consumers that has no bearing on the analysis. It is also important that the data that is collected does not get leaked or used for anything outside of the organization to protect the consumer and their privacy. It is important to look at and/or use the data carefully to eliminate any bias like regions or ethnic groups that could cause incorrect predictions for another group, especially when dealing with fraud analysis. Fairness constraints can be applied to ensure that the outcomes for the demographics are equal. It should be noted that people often view that there is a trade-off for accuracy with fairness constraints, but that is untrue (Sharma & Deshpande). The analysis and dataset must also meet compliance and legal regulations. The data being released must have the consent of the individual and they must also have a way to access, change, and delete the data if they so choose. Transparency is everything and should be included as well as well documented. It would be a good idea for the organization to set up an accountability system or framework like audits and regulatory checks to protect themselves and ensure fair results. A separate team who works with the analysis team and the legal team would be an option here if the budget allows it. They would not be involved in the day-to-day functions but would rather add a third-party view to catch anything that may not have been caught. To help further the protection of the customer, they need to have the option to withdrawal their data at any time.

The next section with be the actual analysis testing on the datasets and will illustrate the detailed findings of the analysis. The section will show the relationships, if any, the variables have with one another. It will show if the null hypotheses would be rejected or fail to reject to answer the research questions as well as solve the problem of e-commerce payment fraud. This will help determine further action the organization can take to improve experience as well as mitigate future fraudulent attacks.

**Findings**

Looking at the question regarding if the number of payments was an indicator of fraud, the first thing that was looked at in the dataset was to see if it was balanced and if there were any missing values. As seen in Figure 1, There were no missing values, however the amount of non-fraud drastically outnumbered the amount of fraud. Figure 2 and Figure 3 supported this and to confirm that a fix was needed. Since there is such an imbalance, SMOTE was used to try and help balance it out.   
**Figure 1** *Results of missing values and the imbalance of fraud* **A white background with blue text

Description automatically generated**

*Note:* Created in R Studio

**Figure 2** *Histogram showing before SMOTE*

A graph of a bar

Description automatically generated

*Note:* Created in R Studio

**Figure 3** *Histogram showing after SMOTE*A graph of a bar

Description automatically generated

*Note:* Created in R Studio

The SMOTE ended up creating additional “True” samples under the Fraud category for a total of 122 samples and achieving a better balance although not perfect. The purpose was to balance the samples and address any misbalancing issues. After investigating the data a little bit more under the Number of Payments variable, it was soon realized that they could be set up in categories using low, medium, and high. Doing this allowed for a clearer understanding of the relationship between Fraud and the Number of Payments. Figure 4 shows the disbursement of the different ranges. It is obvious that there are a lot lower number of payments than the other two.

**Figure 4** *Disbursement of number of payments*  
*A white background with black text

Description automatically generated*

*Note:* Created in R Studio

After this a correlation analysis as performed between the Number of Payments and Fraud to see the relationship as seen in Figure 5. The correlation value is .361595 or just .36. This means that Fraud and Number of Payments have a weak positive correlation. To further test the relationship, a Chi-square test was run as well and had the results of X-square 17.134 with the degrees of freedom of 2. This shows that there was some deviation of payments. Since the maximum number of payments within the dataset was 15 payments. The X-square value was more than the max which can show there being a noticeable discrepancy among the data. The deviation is very large in comparison and could be caused by the large range within the Number of Payments variable.

**Figure 5** *Correlation analysis*  
*A close up of numbers

Description automatically generated*

*Note:* Created in R Studio

The p-value equaled .0001903, which is significantly low. Since the value is so low, it would strongly suggest the null hypothesis will most likely be rejected as usually a low p-value indicates a strong relationship. However, based on the logistic regression between Number of Payments and Fraud variables, as seen in Figure 6, showed a weaker relationship aligning with the Chi-square coeffiecent and it would not be wise to reject the null hypothesis as the results remains inconclusive and do not support the idea of rejecting the null hypothesis. The result does not show enough evidence to support the rejection of the null hypothesis as Figure 6 depicts a weaker relationship that aligns more with the Chi-square coefficient. The difference between the two scores shows the significance of having multiple testing methods done before making conclusions.

**Figure 6** *Logistic regression  
A graph with a red line

Description automatically generated  
Note:* Created in R Studio

Next the relationships between Bitcoin and Apple Pay within the Payment Method Type variable with the transaction failure within the Order State variable were evaluated. A correlation matrix was drawn up to see if there was enough correlation between the payment method as a whole to try to see if it was worth moving on with the analysis. Figure 7 shows the results of the matrix, but it produced a value of -0.00458. This means there is very little, if any correlation between the variables and the analysis could probably stop here.   
**Figure 7** *Correlation matrix of payment method type and transaction fail.   
A number on a white background

Description automatically generated*

*Note:* Created in R Studio

To double check this thought process, a Chi-square test was conducted in SAS Studio. The test broke down the scores for each type of payment for easy comparison. The results seen in Figure 8 shows that each payment type has a mid to high p-value. Since the p-value is way above the significance level, it is safe to say that the null hypothesis should not be rejected as the evidence from multiple tests show there is not sufficient evidence to support the idea of a meaningful relationship between the Payment Method type and the Order State variables.

**Figure 8** *Chi-square test on payment method type*A table with numbers and text

Description automatically generated

*Note:* Created in SAS Studio

**Conclusion**

It was learned pretty quickly that the null hypothesis for both research questions was not going to be rejected at this time. For the question about if the Number of Payments having significant influence on the determination of fraud, even with SMOTE, showed that there was significant evidence to support rejecting the null. But, while doing the test, an error was given about the Chi-square maybe incorrect and other factors including a Chi-square value of 17.134 and a weak correction of .36 shows that the null hypothesis should fail to be rejected at this time. The tests answered the question of whether or not the Payment Method Type was a good predictor of a transaction failing. proved to show that it was not. The correlation coefficient also had a lower correlation score of -.0045. This shows that there is very little to no correlation between the variables. Chi-square test further supported this conclusion with the p-values are in the middle to high range so it would be inconclusive and complicated if one were to reject the null hypothesis. These results point back to the lack of significant relationship between the variables and ensuring the conclusion of not rejecting the null hypothesis. In both cases further tests will need to be done in order to draw more conclusive results to further test the hypothesis. While some patterns and potential relationships were seen, based on the current results remain inconclusive and insufficient for rejecting the null hypothesis.

**Recommendations**

There are plenty of recommendations that could be made. The overall results show that the data is inconclusive, and no decision should be made to reject the null hypothesis for either question proposed in the paper. One recommendation would be to find a larger data set that has more records. A larger data set would help with the imbalance of the data. If a larger dataset remains unbalanced, then the larger analysis could be split into several smaller ones. The card transactions could be removed, and a smaller analysis of the alternative payments could be evaluated to identify trends without being overshadowed by the sheer numbers of the card records. The ones used in this study had very few “True” under the Fraud variable compared to the “False” numbers. Even with Synthetic Minority Oversampling Technique (SMOTE), there was still very little data to work with at times. Other techniques could be used as well like Adaptive Synthetic Smoothing or Sampling (ADASYN). Machine learning algorithms do not perform well when on set in a classification. ADASYN is an improvement on SMOTE as “it is run for two sets of data sets. For Minority and Majority classes, the data in the Minority class number approximates the data in the majority class” (Alhudhaif, 2021). ADASYN focuses on difficult to learn samples which can improve the generalization of the data and has a lower risk of overfitting than SMOTE. Overfitting may have occurred in this analysis. The number of payments in the low category could be investigated and adjusted as well. Adjusting the ranges and/or exploring alternative categorization can help uncover potential stronger patterns that may help detect and predict fraudulent behavior. The customer data transaction data set had more records, but again the spread was just too wide between card usage and the rest. There were about 400 more cards used than Bitcoin, Apple Pay, and PayPal combined. For the purpose of this analysis, only failed transactions under Order State were looked at, so a recommendation would be to look into is how would the results change if pending status was added in with the failed transactions. If one was to remove all the card records, the data set would comprise of maybe 120 records. The data set could be used for credit card fraud. It could be used to look at the regional use and how many payments it would take based on the number of orders or transactions. One can also look at the time of day or the type of device triggers the most fraud or gets the most use in general. Once a better idea of how the data works with more of it then it could be tested and validated with a Receiver Operating Curve (ROC) that tests the false positive rate versus the true positive rate. The ROC method would help judge the model’s performance by looking at the false positive rate to the true positive rate to show the test is reliable. Data collection strategies could be better to collect more diverse to help represent each facet of the data more completely. There was a lot of data that could have been useful to help understand the vast demographics the datasets alluded to.

**References**

Artificial intelligence to push e-commerce fraud to $107b. (2024, October 14). *Nation*. Retrieved

December 15, 2024, from <https://eds-p-ebscohost-com.csuglobal.idm.oclc.org/eds/detail/detail?vid=29&sid=cbdb4b54-daa0-4290-a685-2ca1c954a888%40redis&bdata=JkF1dGhUeXBlPWlwLHVpZCZzaXRlPWVkcy1saXZl#AN=GXLFNIG20241014.00027.2&db=n5h>.

Alhudhaif, A. (2021). A novel multi-class imbalanced EEG signals classification based on the

adaptive synthetic sampling (ADASYN) approach. *PeerJ Computer Science*, *7*.

https://doi.org/10.7717/peerj-cs.523

Chang, V., Ali, B., Golightly, L., Ganatra, M. A., & Mohamed, M. (2024). Investigating credit

card payment fraud with detection methods using advanced machine learning. *Information*, *15*(8), 478–498. https://doi.org/10.3390/info15080478

Hajr, L., Katamoura, S., & Mirza, A. (2023). Bitcoin cryptocurrency and electronic commerce in

Saudi Arabia. *Sage Open*, *13*(4). https://doi.org/10.1177/21582440231218513

Intersection of cryptocurrencies and E-commerce, The. (2023, December 12). The Baltic

Times. https://link-gale-com.csuglobal.idm.oclc.org/apps/doc/A776174037/AONE?

Kutera, M. (2022). Cryptocurrencies as a subject of financial fraud. *Journal of Entrepreneurship,*

*Management and Innovation*, *18*(4), 45–77. <https://doi.org/10.7341/20221842>

McDonald, J. (2023, June 5). CNP fraud losses represent nearly three-fourths of total card

payment fraud losses in the us. *The Fraud Practice*.

Sharma, M., & Deshpande, A. (2023). *How Far Can Fairness Constraints Help Recover From*

*Biased Data?* <https://doi.org/10.48550/arXiv.2312.10396>

Wang, B., Bian, J., Xu, Y., & Liu, D. (2024). The role of internal control on financial fraud:

Evidence of pre-fraud and post-punishment from Chinese firms. *Emerging Markets*

*Finance and Trade*, 1–14. https://doi.org/10.1080/1540496x.2024.2421913

Wu, T., Peng, C.-H., Sia, C., & Lu, Y. (2024). Website localization strategies to promote global

e-commerce: The moderating role of individualism and collectivism. *MIS Quarterly*, *48*(1), 31–66. https://doi.org/10.25300/misq/2022/15542

Yin, H., Zhang, Z., Wang, Z., Ozyurt, Y., Liang, W., Dong, W., Zhao, Y., & Shan, Y. (2022a).

Behavioral graph fraud detection in e-commerce. *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1–8. https://doi.org/10.1109/icdmw58026.2022.00105

Zhang, G., Li, Z., Huang, J., Wu, J., Zhou, C., Yang, J., & Gao, J. (2022). EFRAUDCOM: An

e-commerce fraud detection system via competitive graph neural networks. *ACM Transactions on Information Systems*, *40*(3), 1–29. https://doi.org/10.1145/3474379

Zou, X., Feng, Y., Li, H., & Jiang, S. (2018). Improved over-sampling techniques based on

sparse representation for imbalance problem. *Intelligent Data Analysis*, *22*(5), 939–958. https://doi.org/10.3233/ida-173534