A Review Paper on Feature Selection in Credit Card Fraud Detection

Surbhi Bansal [1,a)] and **Reena Hooda**[2,b)],**Ajay Kumar**[2, b)]

Author Affiliations
[1]*Research Scholar CSE, Indira Gandhi University, Meerpur, Rewari, Haryana, India*
[2]*Assistant Professor CSE, Indira Gandhi University, Meerpur, Rewari, Haryana,India*

Author Emails
[a)]surbhibansal2011@gmail.com
[b)]reenah2013@gmail.com
[c)]yajay2@gmail.com

Abstract. With the increasing popularity of e-commerce and e-payment systems, credit card fraud has become a significant concern for financial institutions and their customers. To combat this problem, machine learning techniques are often used to detect fraudulent transactions. The choice of features used for credit card fraud detection is critical to the effectiveness of the machine learning model. Some of the features that are commonly used are, the circumstances of the transaction, including its place and timing of the location and amount of the transaction, and the merchant category code. Other features that may be considered include the cardholder's purchase history, the frequency of transactions, and the average transaction amount. It is also important to consider the balance between false positives and false negatives when choosing features for credit card fraud detection. False positives occur when a legitimate transaction is flagged as fraudulent, while false negatives occur when a fraudulent transaction is not detected. The balance between these two types of errors can be optimized by adjusting the threshold for classifying a transaction as fraudulent. Overall, the selection of features for credit card fraud detection requires careful consideration and analysis of the data to ensure that the machine learning model can effectively identify fraudulent transactions while minimizing false positives and false negatives. In the paper, we discuss the impact of feature selection in the credit card fraud detection model.

INTRODUCTION

A credit card is a payment card that allows its holder to make purchases on credit, meaning that they can buy goods or services without having to pay for them immediately. Instead, the cardholder is billed for the purchases they make, typically monthly, and is required to pay at least a minimum amount due by the due date to avoid late fees and interest charges. Credit cards are widely used around the world and are often issued by banks, financial institutions, or other credit card companies. They typically come with a credit limit, which is the maximum amount of credit the cardholder is allowed to use at any given time. Using a credit card is generally a straightforward process and can be very easy to use. Once you have been approved for a credit card and have received it, you can start using it immediately. You simply need to present the card at the point of sale, either by swiping it or inserting the chip, entering your PIN, or signing for the purchase. Many credit cards also offer the ability to make purchases online or over the phone, which can be convenient for those who prefer to shop from home. You typically need to submit your credit card information when making an online transaction, including the card number, expiration date, and security code. Using a credit card responsibly involves paying your bill on time, staying within your credit limit, and avoiding high-interest debt. Overall, using a credit card can be a convenient and easy way to make purchases and build credit, but it's important to use it responsibly to avoid financial problems like credit card fraud.

When someone else's credit card or credit card information is used without their permission to make purchases or get cash advances, this is referred to as credit card fraud. This can occur in a variety of ways, such as through the use

of stolen or counterfeit credit cards, Credit card information is captured via skimming devices or phishing scams that mislead consumers into giving over their credit card information. Credit card fraud can have serious consequences for both the cardholder and the card issuer. Cardholders may be liable for unauthorized charges made on their account, and may also suffer damage to their credit score if fraudulent charges are not promptly reported and resolved. Card issuers may also incur significant financial losses as a result of credit card fraud, and may be required to reimburse cardholders for fraudulent charges. To combat credit card fraud, many card issuers use advanced fraud detection systems that leverage machine learning algorithms to identify and flag suspicious transactions. These systems analyze a range of data points, such as the location of the transaction, the type of merchant, the time of day, and the purchase amount, to identify patterns and anomalies that may indicate fraud. In addition to these automated systems, card issuers also rely on manual reviews and investigations to identify and prevent fraud. Cardholders can also play a role in preventing fraud by regularly monitoring their accounts for unauthorized transactions and promptly reporting any suspicious activity to their card issuer. Overall, Credit card fraud is a severe problem that can have negative effects on people's finances and personal lives. Cardholders and card issuers need to take proactive steps to prevent fraud and protect their accounts.

Credit card fraud is a significant issue in India, and has been on the rise in recent years due to the increasing use of credit cards and e-commerce transactions. According to a report by the Reserve Bank of India (RBI), credit card frauds reported by banks increased by 46% in the fiscal year 2020-21. There are various types of credit card fraud that occur in India, including:

- Skimming: This involves stealing credit card information by using a device to read the magnetic strip on the card.
- Phishing: This involves using fake emails or websites to trick users into divulging their credit card information.
- Card cloning: This involves creating a duplicate credit card using stolen information.
- Identity theft: This includes setting up a credit card account or conducting fraudulent transactions using another individual's personal information.

To combat credit card fraud in India, banks and credit card issuers have implemented various measures, such as two-factor authentication for online transactions, SMS alerts for transactions, and fraud detection systems that use advanced analytics and machine learning algorithms. The RBI has also issued guidelines for banks and credit card companies to ensure the safety and security of credit card transactions. However, despite these measures, credit card fraud remains a significant challenge, and consumers need to take steps to protect their credit card information and monitor their accounts for any suspicious activity.

Feature selection is an important step in credit card fraud detection, as it involves identifying the most relevant and informative features that can help distinguish between fraudulent and legitimate transactions. The effectiveness and efficiency of machine learning algorithms used for fraud detection can be increased with the appropriate feature selection.

Various features can be used to detect credit card fraud, including transaction amount, transaction type, location of the transaction, time of day, and cardholder behavior. However, not all of these features are equally important or relevant for detecting fraud, and some may even introduce noise or irrelevant information into the model. Feature selection techniques can be used to identify the most important and informative features for credit card fraud detection, and remove any redundant or irrelevant features that may negatively impact the model's performance. Some common feature selection techniques used in credit card fraud detection include:

- Correlation-based feature selection: This involves selecting features that are highly correlated with the target variable (fraudulent or legitimate transaction) while removing any features that are highly correlated with each other.
- Recursive feature elimination: To get the appropriate amount of features, this entails repeatedly deleting the model's least significant features and re-fitting the model.
- Principal component analysis: This involves minimizing the number of primary components that represent the original features and capturing the most significant variance in the data.
- By using feature selection techniques, credit card fraud detection models can be optimized to achieve higher accuracy, faster processing times, and improved overall performance. Like swarm intelligence.

Feature Selection

Avariable subset selection approach is commonly referred to as feature selection (FS). One of the most prominent and important techniques for minimizing feature dimensionality without reducing the efficiency of a method is feature selection.

Some attributes, generally often referred to as irrelevant features and redundant features, fail to give any useful information. The feature selection method (FSM) may decrease the complexity of the data and the computation.

To select the most appropriate features from the given bucket of features list, it is important to establish the correlation between the data elements. To select the features from a given set of feature vectors Swarm Intelligence has been broadly utilized. SI(Swarm Intelligence) is based on the food collection behavior of different species in the world. SI was first proposed in 1991, and to date, a lot of modifications and alterations have been presented [1]. The following figure illustrated the evaluation of swarm intelligence to date[2].

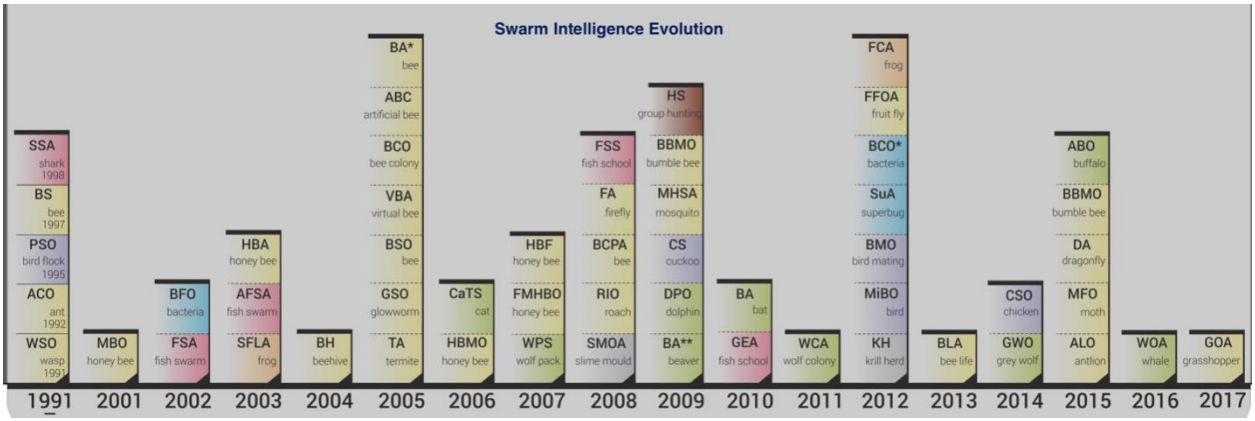

FIGURE 1.SI algorithm list[2]

LITERATURE REVIEW

It is common that fraud detection and its prevention become an unavoidable trend in different sectors such as banking institutions, insurance sector, retail industry, and many more. Therefore, fraud detector advocates and practitioners had been used numerous techniques to determine the causes of fraud. Researchers used different techniques such as ML, SVM, optimization techniques based on swarm, and metaheuristic techniques to prevent credit card fraud. The fraud detection models have been trained and numerous crucial elements are to consider such as feature pre-processing, extraction of features, and many more.

Bahnsen et al. 2016 constructed and evaluate the transactions having a frequency of fraud using the distribution technique in this research. Then, the authors used an actual fraud dataset of a credit card using an established European company. Further, a comparison has been made in which existing fraud detection methods had been compared and it is revealed that attributes affect the findings. The results revealed a 13 percent increase in savings when the recommended periodic characteristics had been included in the techniques [3].

Pushpalatha and Willson Joseph 2017 used to predict and detect fraud on credit cards. In this paper, they compare many of the algorithms like GA(Genetic Algorithm), NN(Neural Network), and other optimization models to determine the severity of fraud. These algorithms are compared based on performance. The hidden Markov Model and Ontology had better performance to detect fraud [4].

Rtayli et al. deployed the Recursive feature reduction strategy, which frequently picks the features and assesses the accuracy of the selected feature, to emphasize the importance of feature selection. In addition, the author applied SVM for hyper-plane-oriented feature optimization. As there are more than two classes in the proposed algorithm architecture, SVM is not appropriate for multi-class classification[5].

Geetha et al. in early 2019, illustrated the significance of machine learning and further in 2022 improved the selection architecture using SI by applying Artificial Bee Colony (ABC). The authors improved the group behavior and classified the optimized set using Neural Networks[6].

Meraihi et al .the authors discuss and critique GOA variations, hybrids, and applications in a variety of fields, including feature selection, scheduling, distributed generation, economic dispatch, and flood routing, which are the major topics of the work. Despite GOA's success, the authors argue that there is still much to investigate in terms of adapting the standard algorithm, combining it with other meta-heuristics, and using it to address other real-world optimization issues. They suggest more investigation into the utilization of alignment, separation, and cohesion as swarming processes in GOA in addition to developing new operators and strategies[7].

Singh et al. This study used the German dataset to test machine learning approaches and feature selection methods for the detection of application-level credit card fraud. Based on their prediction accuracy, MCC, sensitivity, specificity, precision, and recall, five machine learning methods were evaluated. The accuracy and precision of classifiers were increased by the application of filter and wrapper feature selection techniques for selecting strongly correlated features. The J48 and PART classifiers improved their prediction accuracy the most, while AdaBoost and random forest classifiers improved their precision the most[8].

Darwish proposes a model, to overcome the shortcomings in credit card fraud detection (CCFD) systems, the current research suggests an intelligent system based on the Artificial Bee Colony (ABC) optimization method. To extract pertinent characteristics from a big dataset and produce a more accurate model, the system combines an ABC optimizer and a semantically fused K-means classifier. The time needed for model construction and computation is also reduced by the suggested methodology. The goal function to be optimized is defined to have a critical influence on the quality of the resulting model. A rule engine is utilized to correlate variables received by clustering levels with their meaning for a person[9].

CONCLUSION

In the study, we find out that Credit card fraud is a big challenge for both financial institutions and clients. The detection of fraud has made extensive use of machine learning techniques, and feature selection approaches are essential for increasing the precision and effectiveness of fraud detection systems.

REFERENCES

1. G. Beni, “Swarm intelligence. Complex Social and Behavioral Systems.” *Game Theory and Agent-Based Models*,(2020), pp.791-818.
2. SI algorithms, Available online at https://www.google.com/search?q=swarm+intelligence+algorithms+list&tbm=isch&ved=2ahUKEwjYp8zKmo_4AhXi_TgGHThWBpgQ2-cCegQIABAA&oq=swarm+intelligence+algorithms+list&gs_lcp=CgNpbWcQAzIGCAAQHhAIOgQIIxAnOgYIABAeEAc6BAgAEBg6BAgAEB46BggAEB4QBVDQCFiTDWDLDmgAcAB4AIABgwGIAdoFkgEDMC42mAEAoAEBqgELZ3dzLXdpei1pbWfAAQE&sclient=img&ei=9OeYYthd4vvj4Q-4rJnACQ&bih=808&biw=1512&rlz=1C5CHFA_enIN987IN988#imgrc=2SNmVj6ChsAGNM&lns=W251bGwsbnVsbCxbMCwwLDk5LDk5XSxudWxsLG51bGwsbnVsbCxudWxsLCJFa2NLSkRaaVkyTmhZbVZtTFdWak9HRXRORE13WWkwNVl6VTBMVGt6TVRReE9Ea3lNMlkwWVJJZlFUaHhOamRVTlU5YU5UUlhXVXh1WlRaNlNUTlhTRkJzYWt0YVZVVm9adz09IixudWxsLG51bGwsbnVsbCwxXQ.
3. A. C.Bahnsen, D. Aouada, A.Stojanovic, and B.Ottersten, “Feature engineering strategies for credit card fraud detection”, *Expert Sys.with App., vol. 51, pp. 134-142,* 2016.
4. B. Pushpalatha and J.C.Willson, “Credit Card Fraud Detection Based on the Transaction by Using Data Mining Techniques”,*Inter. J. of Inn. Res. in Comp. and Comm.Engg*, vol. 5, no. 2, February 2017.
5. N. Rtayli, & N. Enneya, “Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization”, *Journal of Information Security and Applications*, vol.55, p. 102596, 2020.
6. N.Geetha, and G.Dheepa, “Transaction Fraud Detection Using Artificial Bee Colony (Abc) Based Feature Selection And Enhanced Neural Network (ENN) Classifier”, *International Journal of Mechanical Engineering*, vol. 7, no. 3, 2022.

7. Y. Meraihi, A. B. Gabis, S. Mirjalili, and A. Ramdane-Cherif, "Grasshopper Optimization Algorithm: Theory, Variants, and Applications," *IEEE Access*, vol. 9, pp. 50001–50024, Mar. 2021, doi: 10.1109/access.2021.3067597.
8. A. B. Singh and A. Jain, "Adaptive Credit Card Fraud Detection Techniques Based on Feature Selection Method," in *Advances in intelligent systems and computing*, Springer Nature, 2019, pp. 167–178. doi: 10.1007/978-981-13-6861-5_15.
9. S.M. Darwish, "A bio-inspired credit card fraud detection model based on user behavior analysis suitable for business management in electronic banking", *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp.4873-4887, 2020.