

**University of Hertfordshire**  
**School of Physics, Engineering and Computer Science**

**Research Methods (7COM1085-0105-2022)**

**Using logistic regression, decision tree and SVMs to identifying credit card fraud**

**Group: a\_group 117**

Z Pavan Sunder

Husnain raza

Z Mohammed Rehaan

Muhammad Usman Shahzad

12 May 2023

## Table of Contents

Introduction .....	3
Background Research .....	4
Research Question and Methods .....	5
Results .....	6
Synthesis.....	7
Conclusion .....	7
Bibliography .....	8

## Introduction

Credit card fraud is a growing concern for financial institutions and customers alike (Chen & Li, 2019). Fraudsters are constantly developing new tactics to steal credit card information and use it for illegal purposes, making it difficult to detect and prevent fraudulent transactions. To address this issue, machine learning algorithms have been developed to analyse large amounts of data and detect patterns that may indicate fraudulent activity (Bolton & Hand, 2002).

There are several machine learning algorithms that can be utilized to detect credit card fraud. Decision trees can create a tree model to predict outcomes based on a data set (Liu, Ma, & Wang, 2019). Logistic regression is another commonly used algorithm that can predict the likelihood of an event (Bolton & Hand, 2002). Support vector machines (SVMs) can be used to decompose data into different classes based on their characteristics (Bolton & Hand, 2002).

One of the main advantages of using machine learning algorithms to detect credit card fraud is their ability to analyse large amounts of data in a short time (Chen & Li, 2019). This can lead to more accurate and efficient fraud detection, ultimately saving financial institutions and their customers money in the long run. Additionally, these algorithms can be tuned to detect new types of fraud as they emerge, providing a scalable and flexible solution to fraud detection (Bolton & Hand, 2002).

Machine learning algorithms have the potential to revolutionize the way credit card fraud is detected and prevented. By utilizing these algorithms, financial institutions can stay ahead of fraudsters and protect their customers from financial loss.

## Background Research

Machine learning algorithms such as SVMs, decision trees, and logistic regression have been shown to be effective in detecting credit card fraud (Kumar et al., 2021; Neha et al., 2020; Zied, 2018). Decision trees are useful because they can handle combinations of categorical and continuous variables and can identify the most important variables in a data set (Neha et al., 2020). Logistic regression is a probabilistic classification model that can be used to identify fraudulent behaviour by modelling the likelihood of fraudulent behaviour based on input variables (Kumar et al., 2021).

SVMs are efficient at handling multidimensional data and can separate data into classes, making them useful in fraud detection (Zied, 2018). SVMs have been shown to be effective in detecting credit card fraud, with one study achieving an accuracy of 99.6% (Zied, 2018). Another study found that logistic regression achieved an accuracy of 96.7% in detecting credit card fraud (Kumar et al., 2021). SVMs have also been shown to outperform other machine learning algorithms in detecting credit card fraud, including decision trees and neural networks (Zied, 2018).

One of the main advantages of using machine learning algorithms to detect credit card fraud is their ability to analyse large amounts of data in a short time (Chen & Li, 2019). This can lead to more accurate and efficient fraud detection, ultimately saving financial institutions and their customers money in the long run. Additionally, these algorithms can be tuned to detect new types of fraud as they emerge, providing a scalable and flexible solution to fraud detection (Bolton & Hand, 2002).

## Research Question and Methods

### Research Question

In what ways can decision trees, logistic regression, and SVMs be utilized to enhance the performance of identifying credit card fraud?

### PICO Elements

**Population:** identifying credit card fraud, SVMs, decision trees and logistic regression.

**Intervention:** SVMs, Decision trees and logistic regression.

**Comparison:** Current techniques for identifying credit card fraud.

**Outcome:** Improved performance in identifying credit card fraud.

**Context:** Fraud in Credit card is a serious problem for economic institutions and individuals. Traditional rule-based systems are often insufficient to detect increasingly sophisticated fraud techniques. There are some algorithms in Machine learning, such as SVMs (support vector machine, decision trees and logistic regression, this have been applied to improve credit card fraud recognition performance. Decision trees are useful for detecting complex relationships between features, logistic regression is efficient for dealing with large data sets, and SVMs can accurately classify data into distinct categories. By using these machine learning algorithms, the accuracy of credit card fraud identification can be improved, enabling early detection, and preventing financial loss.

The database used for Data collection is IEEE Xplore

**Search String:** ("credit card fraud" OR "fraud detection") AND ("decision trees" OR "logistic regression" OR "support vector machines" OR "SVM")

### Number of papers found after applying Inclusion and Exclusion criteria

**Number of papers:** 92

#### Inclusion criteria:

1. adding decision trees
- 2.adding logistic regression
- 3.adding credit card fraud

#### Exclusion criteria:

1. 2022 to 2023
- 2.Exclude paper not written in English
- 3.Exclude duplicate papers
- 4.Early access articles

## Results

The Answer to the Research question is

Decision trees can be used to identify the most important features that contribute to credit card fraud detection. This can be done by constructing a decision tree model and analyzing the feature importance scores. By doing this, the model can be optimized to focus on the most important features, which can improve its accuracy in detecting fraud (Li et al., 2018).

Logistic regression can be used to estimate the probability of credit card fraud. By analyzing the probability estimates generated by the model, the risk of fraud can be evaluated, and appropriate actions can be taken. Additionally, logistic regression can be used in combination with other models, such as decision trees, to further improve the accuracy of fraud detection (Chen et al., 2017)

SVMs can be used to classify transactions as fraudulent or non-fraudulent by constructing a hyperplane that separates the two classes. By optimizing the hyperplane, the SVM model can accurately classify transactions and improve the performance of credit card fraud detection (Zhang et al., 2017).

**Total number of papers accepted after Include title is: 49.**

**Total number of papers accepted after Include abstract is: 49.**

**Total number of papers excluded is: 42.**

**Total number of papers accepted after accept content is: 49.**

## Synthesis

The studies have shown that support vector machines (SVMs), logistic regression, and decision trees are all efficient machine learning algorithms for detecting credit card fraud. These techniques, in particular, have been demonstrated to increase the accuracy of fraud detection models and decrease the losses brought on by fraudulent actions. The study also discovered that decision trees can spot complex fraud patterns that other techniques might overlook. It has been demonstrated that SVMs are useful for detecting credit card fraud, particularly when the data cannot be separated linearly. Additionally, some studies have looked into the application of ensemble methods, which mix various machine learning approaches to increase the precision of fraud detection models. The accuracy of this strategy is found to be higher than using any one method alone, for instance, when decision trees, logistic regression, and SVMs were integrated in a combined model.

Overall, the results indicate that SVMs, decision trees, and logistic regression can all be useful for detecting credit card fraud. Financial organizations can create effective fraud detection systems that are able to react to new and developing fraud patterns by combining these methods and using ensemble methodologies.

## Conclusion

Decision trees, logistic regression, and SVMs are viable machine learning procedures that can be utilized to enhance the execution of recognizing credit card false transaction. These models can recognize the foremost vital features that contribute to fraud detection, assess the likelihood of fraud, and classify transactions as false or non-fraudulent. By optimizing these models and analyzing their yields, the exactness of extortion discovery can be progressed, and the hazard of false exchanges can be minimized. in this it illustrate the adequacy of these models in identifying credit card fraud, and give an establishment for further investigate in this area.

## Bibliography

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235-255.

Chen, X., & Li, Y. (2019). Credit card fraud detection model based on machine learning algorithm. *Journal of Physics: Conference Series*, 1204(1), 012029.

Liu, J., Ma, C., & Wang, Q. (2019). Credit card fraud detection based on decision tree algorithm. *Journal of Physics: Conference Series*, 1180(1), 012075

Kumar, A., Singh, S., & Yadav, A. (2021). A comparative study of machine learning algorithms for credit card fraud detection. *Journal of Big Data*, 8(1), 1-22.

Neha, S., Chakraborty, S., & Das, S. (2020). Credit card fraud detection using machine learning algorithms. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

Zied, H. (2018). Credit card fraud detection using support vector machines. *Journal of Intelligent & Fuzzy Systems*, 35(4), 4487-4495.

Chen, X., Xu, L., Liu, G., & Liu, X. (2017). A hybrid credit card fraud detection model based on decision tree and logistic regression. *Journal of Computational Science*, 22, 37-43.

Li, Y., Zhang, X., & Zhang, Y. (2018). A credit card fraud detection model based on decision tree algorithm. In *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)* (pp. 58-62). IEEE.

Zhang, J., Zhang, F., & Chen, X. (2017). Credit card fraud detection based on SVM optimized by improved genetic algorithm. In *2017 36th Chinese Control Conference (CCC)* (pp. 10673-10677). IEEE.