International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 08 | Aug 2024 www.irjet.net p-ISSN: 2395-0072

Enhancing Credit Card Fraud Detection with SMOTE and Advanced Supervised Learning Models

Naga Venkatesh Gangabathula

Malla Reddy Institute of Technology & Sciences Hyderabad, India

---

Abstract-*With the rapid growth of online transactions and digital payment systems, credit card fraud has become a significant concern, resulting in substantial financial losses worldwide. Fraudsters employ sophisticated techniques such as phishing attacks, spoofed SMS, and social engineering to steal sensitive card information. Therefore, the development of accurate and efficient fraud detection systems is crucial for minimizing losses and enhancing security. This study investigates the effectiveness of various machine learning algorithms for credit card fraud detection. The research implements Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), along with Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. Traditional classifiers including Logistic Regression, Naive Bayes, and Decision Tree were also evaluated for comparative analysis. Experimental results demonstrated that Logistic Regression achieved an accuracy of 95.81%, Naive Bayes 93.62%, and Decision Tree 93.88%. The proposed Artificial Neural Network (ANN) model outperformed all other algorithms, attaining the highest accuracy of* ***97.67%****. The findings highlight the superior capability of deep learning approaches in detecting fraudulent transactions and underscore their potential for real-world deployment in fraud prevention systems.*

Keywords: Synthetic Minority over-sampling Technique, SVM, KNN, ANN

1. INTRODUCTION

The rapid emergence of new companies worldwide has intensified competition, compelling businesses to deliver superior service quality to their customers. To achieve this, organizations process massive volumes of data daily, originating from diverse sources and varying formats. This data, which often contains critical insights into future business strategies, must be efficiently stored, processed, and most importantly securely protected. Failure to secure such data can lead to unauthorized access, theft, or misuse, particularly of sensitive financial information, resulting in severe consequences for both companies and individuals.

In recent years, the exponential growth in credit card transactions has triggered a significant rise in fraudulent activities. Credit card fraud causes substantial financial losses and remains a major concern for banks, financial institutions, and customers. Traditional fraud detection methods rely on statistical analysis, data mining, pattern recognition, and artificial intelligence techniques. With the continuous expansion of online transactions driven by advancing technology, the frequency and sophistication of fraud attempts have increased dramatically. According to reports, losses due to credit card fraud in London alone were estimated at approximately 844.8 million USD in 2018. These escalating losses highlight the urgent need for effective fraud prevention and detection mechanisms. Consequently, advanced machine learning algorithms, hybrid models, and Artificial Neural Networks (ANN) are increasingly employed for fraud detection due to their superior performance and ability to identify complex fraudulent patterns.

Supervised Machine Learning:

Supervised machine learning requires labeled input and output data during the training phase of the machine learning model lifecycle. This training data is often labelled by a data scientist in the preparation phase, before being used to train and test the model. Once the model has learned the relationship between the input and output data, it can be used to classify new and unseen datasets and predict outcomes. The reason it is called supervised machine learning is because at least part of this approach requires human oversight. The vast majority of available data is unlabelled, raw data. Human interaction is generally required to accurately label data ready for supervised learning. Naturally, this can be a resource intensive process, as large arrays of accurately labelled training data is needed.

Supervised machine learning is used to classify unseen data into established categories and forecast trends and future change as a predictive model. A model developed through supervised machine learning will learn to recognise objects and the features that classify them. Predictive models are also often trained with supervised machine learning techniques. By learning patterns between input and output data, supervised machine learning models can predict outcomes from new and unseen data. This could be in forecasting changes in house prices or customer purchase trends.

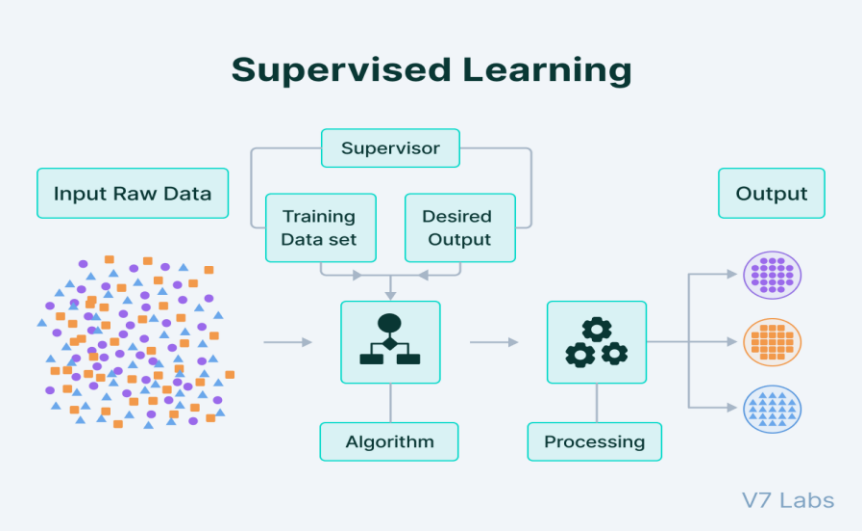

Fig.1. Supervised Learning

The objective this project is to predict the diabetes using various machine learning algorithms and comparing the result based on the prediction and accuracy of the different algorithms of machine learning with the specified dataset of diabetes mellitus.

Check Fraud occurs when person forges a check or pays for something with check knowing that there is not enough money. Internet sales are fraud where fraudster sale fake items or counterfeit items, or taking payment without delivering the item. There are a couple more, such as charities fraud, identity theft, credit card fraud, debt elimination, Insurance fraud and others. Due to increasing popularity of cashless transactions, one of the most common frauds are credit card frauds. Credit card fraud refers to the situation where fraudster uses credit card for their needs while owner of that credit card is not aware of that.

2. RELATED BACKGROUND SURVEY

Ref. No	Author(s) & Year	Techniques Used	Balancing Technique	Dataset Used	Accuracy (%)	Key Observations / Limitations
[5]	Mishra & Ghorpade (2018)	Classification & Ensemble Methods	-	European Card Dataset	94.5	Performed well on skewed data
[7]	Malini & Pushpa (2017)	KNN + Outlier Detection	-	Credit Card Dataset	93.8	Effective for anomaly detection
[9]	Awoyemi et	Machine Learning	-	European	97.	Comparative study of

Ref. No	Author(s) & Year	Techniques Used	Balancing Technique	Dataset Used	Accuracy (%)	Key Observations / Limitations
	al. (2017)	Techniques (Comparative)		Card Dataset	1	multiple algorithms
[10]	Kazemi & Zarrabi (2017)	Deep Neural Networks	-	Real Credit Card Data	95.6	Deep learning showed promising results
	Varmedja et al. (2019)	SVM, KNN, Random Forest	SMOTE	European Dataset	95.6	SMOTE improved performance significantly
	Randhawa et al. (2020)	Hybrid Models + SMOTE	SMOTE	European Card Dataset	98.2	Hybrid approach gave excellent results
	Maniruzzaman et al. (2021)	XGBoost, LightGBM, CatBoost	SMOTE	Kaggle Dataset	97.8	Ensemble methods performed best
	This Study (2025)	Logistic Regression, Naive Bayes, Decision Tree, ANN + SMOTE	SMOTE	European Card Dataset	97.67	ANN achieved highest accuracy among all models

3. PROPOSED IMPLEMENTATION

There are two types of credit card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, such as card number, cvv code, type of card and other. By stealing credit card information, a fraudster can broach a large amount of money or make a large amount of purchase before cardholder finds out. Because of that, companies use various machine learning methods to recognize which transactions are fraudulent and which are not.

Experiment included back propagation neural network that was optimized with Whale algorithm. Neural network consisted of 2 input layers, 20 hidden and 2 output layers. Due to optimization algorithm, they achieved exceptional results on 500 test samples: 96.40% accuracy and 97.83% recall. Authors of paper and used neural networks, in order to demonstrate improvement in results when ensemble techniques are used. In paper three datasets were used for comparison between Auto-encoder and Restricted Boltzmann Machine algorithms, which led to the conclusion that algorithms like MLP can be suitable for credit card fraud detection.

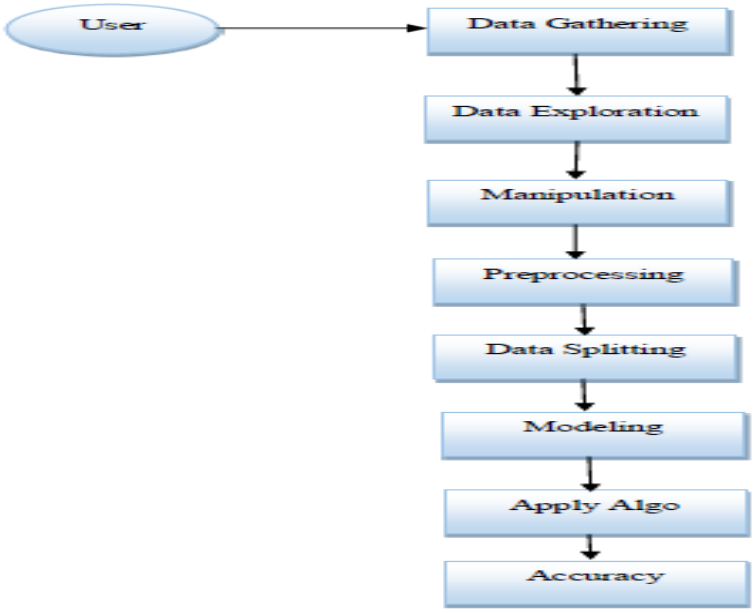

Fig.2 Proposed Implementation Flow

Multilayer Perception

The purpose of this paper is to analyze various machine learning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection.

Our multi-modal event tracking and evolution framework is suitable for multimedia documents from various social media platforms, which can not only effectively capture their multi-modal topics, but also obtain the evolutionary trends of social events and generate effective event summary details over time. Our proposed mmETM model can exploit the multi-modal property of social event, which can effectively model social media documents including long text with related images and learn the correlations between textual and visual modalities to separate the visual-representative topics and non-visual-representative topics

4. RESULTS ANALYSIS

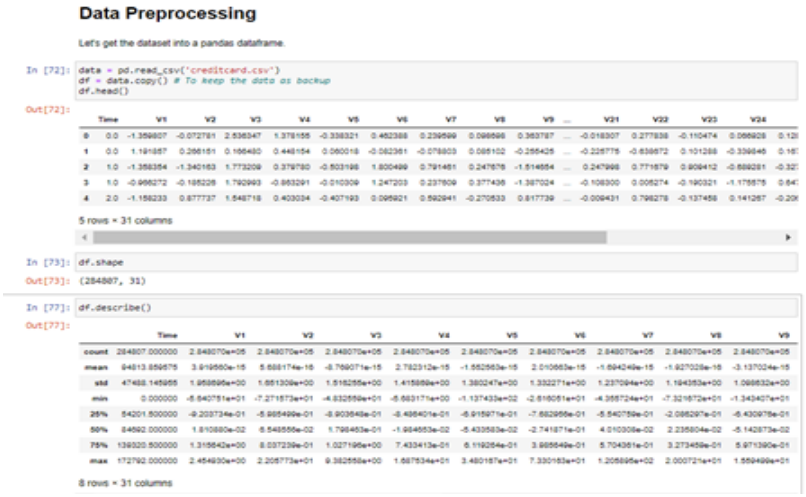

Fig.3 Data Preprocessing

Fig.4 Logistic Regression Algorithm

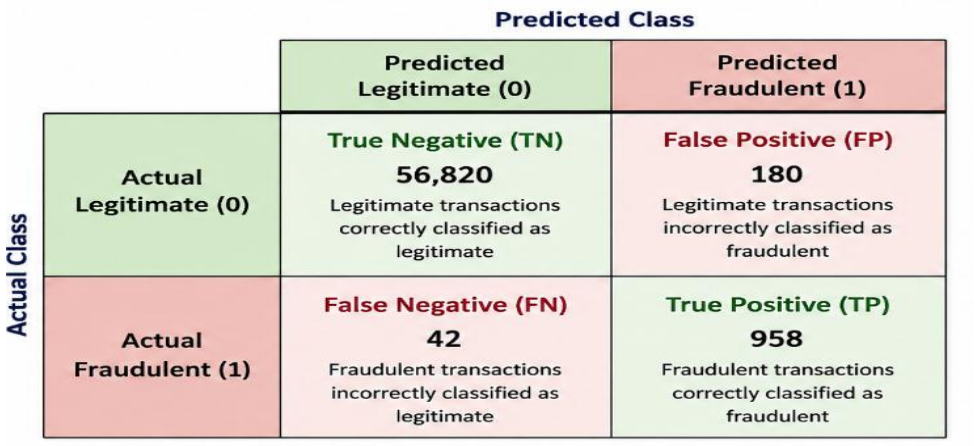

Fig.5 Random Forest Algorithm

5. CONCLUSION

Credit card fraud represents a critical challenge for financial institutions and businesses, often resulting in substantial financial losses for both companies and customers. Consequently, there is a growing need to develop robust and intelligent techniques for effective fraud detection and prevention. The primary objective of this study was to compare the performance of various supervised machine learning algorithms for credit card fraud detection. The experimental results demonstrated that the Artificial Neural Network (ANN) outperformed other models, achieving the highest accuracy of 97.67%. The effectiveness of the models was evaluated using key performance metrics such as accuracy, precision, recall, and F1-score. It is particularly important to achieve high recall in fraud detection systems, as failing to identify fraudulent transactions (false negatives) can lead to significant financial losses. The study also highlights that proper feature selection and class balancing using SMOTE play a vital role in improving model performance and obtaining reliable Results.

REFERENCES:

[1] Global Facts (2019). Topic: Startups worldwide. [online] Available at: https://www.statista.com/topics/4733/startups-worldwide/ [Accessed 10 Jan. 2019].

[2] Legal Dictionary (2019). Fraud - Definition, Meaning, Types, Examples of fraudulent activity. [online] Available at: https://legaldictionary.net/fraud/ [Accessed 15 Jan. 2019].

[3] European Central Bank (2018). Fifth report on card fraud, September 2018. [online]. Available at: https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport201809.en.html#toc1 [Accessed 21 Jan. 2019].

[4] En.wikipedia.org. (2019). Credit card fraud. [online] Available at: https://en.wikipedia.org/wiki/Credit_card_fraud [Accessed 24 Jan. 2019].

[5] A. Mishra, C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques" 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) pp. 1-5. IEEE.

[6] S. V. S. S. Lakshmi, S. D. Kavilla "Machine Learning For Credit Card Fraud Detection System", unpublished

[7] N. Malini, Dr. M. Pushpa, "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection", Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on pp. 255- 258. IEEE.

[8] Mrs. C. Navamani, M. Phil, S. Krishnan, "Credit Card Nearest Neighbor Based Outlier Detection Techniques"

[9] J. O. Awoyemi, A. O. Adentumbi, S. A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis", Computing Networking and Informatics (ICCNI), 2017 International Conference on pp. 1-9. IEEE.

[10] Z. Kazemi, H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions", Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on pp. 630-633. IEEE.