Journal of Computers for Society

Journal homepage: https://ejournal.upi.edu/index.php/JCS

A Support Vector Machine Credit Card Fraud Detection Model based on High Imbalance Dataset

Odeyale K. Musiliudeen[1,], Oyelakin A. Moruff[2], Salau-Ibrahim T. Taofeekat[3], Saka M. Kayode[4]*

[1] Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria
[2] Department of Computer Science, College of Information and Communication Technology, Crescent University, Abeokuta, Nigeria
[3] Federal University of Lafia. Lafia, Nasarawa State, Nigeria
[4] Department of Information and Communication Technology, University of Ilorin Teaching Hospital, Kwara State, Nigeria.
*Correspondence: E-mail: kmodeyale@alhikmah.edu.ng

ABSTRACT

Credit card transactions are exposed to fraudulent activities owing to their sensitive nature. The illegal activities of the fraudsters have been reported to cost financial institutions a lot of money globally as reported in many notable research works. In the past, several machine learning-based approaches have been proposed for the detection of credit card fraud. However, little attention has been given to classification of fraud in high imbalance dataset. Generally, if a dataset is imbalanced, a learning algorithm will give a bias result in terms of the accuracy resulting in poor performance of the model. This study focuses on using Synthetic Minority Oversampling Technique (SMOTE) to address the class imbalance in the selected credit card dataset. Then, ANOVA-F statistic was applied for the selection of promising features. Both the class imbalance and attribute selection techniques were targeted at improving the SVM-based credit card fraud classification. With the balanced dataset, the study achieved an accuracy of 93.9%, recall of 97.3%, precision of 90.3%, and f1 score of 93.5% respectively. It was observed that the result of the Support Vector VM based credit card fraud detection model with class imbalance is better than that of the standard SVM. The study concluded that the class imbalance addressing and attribute selection techniques used were very promising for the credit card fraud detection tasks.

ARTICLE INFO

Article History:
Submitted/Received 3 Mar 2024
First Revised 27 Apr 2024
Accepted 6 Jun 2024
First Available Online 7 Jun 2024
Publication Date 15 Sep 2024

Keyword:
ANOVA F-Test,
Fraud detection,
Imbalanced Data
Machine learning,
SMOTE,
Support Vector Machine.

1. INTRODUCTION

A credit card is a card issued to consumers (cardholders), who can use it to make purchases up to a certain limit or withdraw cash from any location (Lee and Kwon, 2002). Using credit cards, banks provide their consumers with a variety of services. For instance, by moving it to the succeeding next bill, it permits customers to pay at a later date and time. Fraud is illegal or criminal behavior that aims to gain a financial or personal advantage (Moradi and Mokhatab, 2019). Credit cards are used for various transactions globally. Online purchasing of goods and services have become more common in everyday life. Internet payments are increasingly the most common type of online transaction. The banking system offers e-cash, e-commerce, and e-services through internet transactions. In line with a Nielsen survey carried out between 2007 and 2008, 28% of the global population uses the World Wide Web, 85% of people worldwide use the internet in order to carry out online transactions, and the frequency of running business via the internet has climbed by 40% between the years 2005 and 2008.

As credit card usage expands globally, there is also an increase in the likelihood that an attacker may steal credit card information and use it to commit fraud (Ali *et al*., 2019). Fraud is defined as any action taken to deceive to collect money without the cardholder's or issuer bank's knowledge. Many methods can be used to commit credit card fraud. By misplacing or stealing cards, manufacturing counterfeit or phony cards, copying phishing, skimming, or stealing data from a merchant's side, deleting or replacing the magnetic stripe on the card that stores the user's information (Budhram, 2012). One of the most widely used techniques is the theft of credit cards. Skimming can be carried out offline or online. In the physical technique, the credit card is swiped through the skimmer, which records the card's number, expiration date, and complete name of the user (Sivakumar and Balasubramanian, 2015). Via cyber-attacks like phishing, SQL injection, or keylogging, the credit card information is collected from the user or an e-commerce system's servers using the online method.

Thus, proposing a machine learning credit card fraud detection model based on some innovative approaches is a step in the right direction. This work focuses on using the SVM Algorithm with SMOTE in highly imbalanced dataset to detect credit card fraud.

2. RELATED WORKS

Trivedi *et al*, (2020) proposed detection of credit card fraud using SVM, With such a feedback system, the entire study piece established an efficient technique of detecting fraud based on machine learning methodologies. Using Machine Learning techniques, a model was put forward that detects fraudulent credit card transactions. The suggested model approaches detection of fraud as a binary problem of classification. The research project explored the efficiency of the analysis of the Support vector machine's Kernel; the techniques were trained using transactional data, and their performances were evaluated and compared using accuracy, specificity, and sensitivity performance measures. The model is evaluated in comparison to existing classifiers such as Naive Bayes, Decision Tree, K-Nearest Neighbor (KNN), and Logistic Regression (LR). When compared to other methodologies, the Radial Basis Function (RBF) kernel function outperforms and provides 96% accuracy and 96% sensitivity.

Kadam *et al* (2021) carried out research on Credit Card Fraud Detection with ML techniques, three classification approaches were employed which are Support Vector Machine, Logistic Regression and Random Forest, PCA was used as dimensionality reduction for the dataset. 0.172% of fraudulent transactions were able to detect due to the highly imbalanced dataset.

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

Different techniques like SVM ,RF, ANN, Tree Classifiers, NB and LR were employed, SVM outperformed other techniques with an accuracy percentage of 91.988%. The authors concede that more research should be conducted to implement a feature selection method that could improve on the accuracy of other ML methods.

Sharma *et al*, (2021) worked on A Comparison of Machine Learning Models for Detecting Credit Card Fraud. A comparison analysis was performed using several techniques such as RF, LR, SVM and AAN. PCA was used for data preprocessing, the result showed that with an F1-score of 0.91, Artificial Neural Networks performed best.

Stochastic Gradient Descent (SGD), Decision Tree (DT), Random Forest (RF), J48 and Instance Based-k (IBk) machine learning classifiers were applied. The binary issue of classification is examined in this research, whereby a transaction can be classed as either fraudulent or lawful. The idea is to use five different methods of machine learning to classify the transactions. Result showed that RF outperformed the other classifier algorithms in term of evaluation metrics.

Ahirwar *et al*, (2020) conducted a study on Credit Card Fraud Detection Using Enhanced SMOTE and Fast Random Forest Techniques. The suggested method presents a smart card fraud identification model that is capable of identifying fraud in highly skewed or nonspecific credit card transaction data. The suggested fraud detection approach contains three stages, incorporates preprocessing in which duplicate attributes are removed, then ranks qualities by their significance utilizing the rapid Random Forest algorithm. The results are compared with the UCSD FICO data mining competitions 2009 dataset, that is a generic dataset for real-world transactions made with credit cards. The suggested method can handle extremely unbalanced data. The study investigates the performance of several techniques on four datasets used for training. DF1, DF2, DF3, and DF4 have the fraud rates of 20%, 15%, 10%, and 3% respectively. The rapid Random Forests algorithm's results demonstrated that a similar strategy would prove successful in real-time.

Ata and Hazim (2020) performed an evaluation of several distributions' datasets employing data mining approaches on fraud with credit cards detection. On actual credit card transactions from European cardholders, four data mining algorithms were investigated in this research: RF, SVM, KNN, NB. All of these algorithms were employed on an under-sampled class to categorize transactions as fraud or legitimate, and then their performance metrics were evaluated and compared using a confusion matrix. The best accuracy for the NB, SVM, KNN, and RF classifiers is 97.80%, 97.46%, 98.16%, and 98.23%, respectively. The limitation of this study was that a sampling approach was used, and the technique was not applied to a huge dataset.

3. METHODS

3.1. Proposed Credit Card Fraud Detection Model

The approach and procedures required to achieve the specified objectives are described in **Figure 1** below.

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

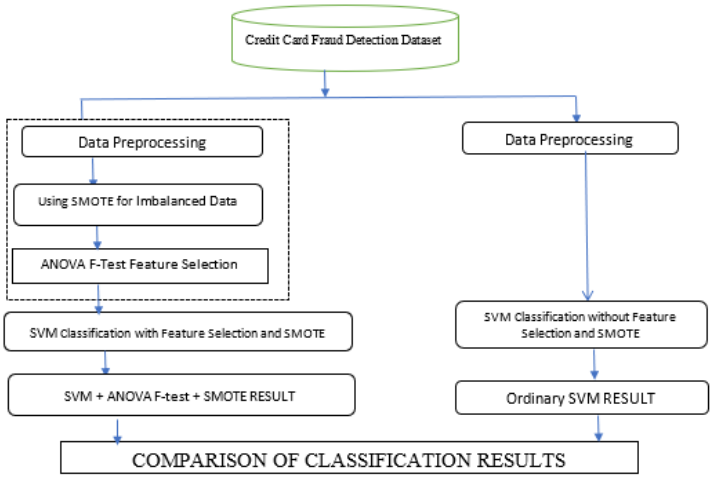


Figure 1. Proposed approach for the credit card fraud detection.

3.2. Dataset Description

The dataset used in this study contains 31 features/attributes and 284,807 samples. The target class in the dataset that can be predicted as '0' or '1', i.e., 'genuine' or 'fraudulent' transactions. In order to protect user privacy, Principal Component Analysis (PCA) has previously been utilized to transformed the dataset. The purpose of the transformation is to preserve as much variation and correlation between the attributes in the dataset as is practicable while reducing their dimensionality. The properties 'V1' through 'V28' were concealed using PCA. Thus, the feature analysis and selection of these 28 qualities were limited. The last two features are "Amount" and "Time." After feature selection, 30 attributes, including the class column, were still present.

3.3. Dataset Pre-processing

Pre-processing is the modification of the dataset prior it is fed to the algorithm. Pre-processing is required to prepare the data for modeling, including removing any errors and outliers that may be present in the data. This section concentrated on filtering away every inconsistent set from the dataset in order to improve the dataset's smooth operation for better result optimization.

3.4. Handling the Imbalance Data using SMOTE Technique

In 2002, SMOTE was introduced by chawla and some researchers. The goal was to help the classifier improve its generalization on the testing data by overcoming the overfitting caused by simple oversampling via replication. The basis of this innovative data preparation technique was to produce new minority instances rather than "weighting" data points. This technique was known as the Synthetic Minority Oversampling Technique (SMOTE) (Chawla *et al*., 2002). The SMOTE technique was based on interpolation among surrounding minority class instances. As a result, it can increase the number of minority class instances in the neighborhood by providing fresh minority class examples, supporting the classifiers in improving their generalization capacity.

SMOTE is an oversampling technique in which false samples are manufactured for the minority class. This strategy helps to overcome the overfitting problem caused by random oversampling. Several studies have indicated that if a dataset for a classification problem is imbalanced, the model will be skewed and the results would be deceptive. This strategy will

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

use the class imbalance fixing technique, which increases the likelihood of obtaining a balanced dataset (Oyelakin and Jimoh, 2020).

3.5. Feature Selection

In a Machine Learning-based predictive analysis, the feature selection algorithm focuses on choosing promising variables. We have the whole feature set in Feature Selection and then try to develop an identified feature set for the problem in the domain that we are studying (Eseye et al., 2018).

The feature selection method used in this study is filter-based selection technique called ANOVA F-Test. The procedure is chosen based on its applicability in light of the availability of numerical input variables and a categorization of target variables (Sarker, 2021).

3.6. SVM-based Credit Card Fraud Classification

For classification and pattern analysis, SVM was employed. This classifier divides patterns into two categories: fraud and valid patterns. For binary classifications, this method is employed. This is employed in pattern identification and classification tasks like text categorization, facial recognition, and bioinformatics. The experimental approaches involve combining SVM with SMOTE and without SMOTE.

3.7. Support Vector Machine

An SVM is a well-known supervised probabilistic technique that may partition data both sequentially and non-sequentially. The linear SVM is a binary classifier that classifies multi-dimensional data by creating hyper-planes using some nearest training data points of each class and maximizing the margin between them (Dutta *et al.*, 2015). SVM is a supervised method that is used to separate behavioral features that belong to different classes by converting feature vectors into high-dimensional space and locating hyperplanes (lines separating data points) to split the space. SVM algorithm can be seen in **Table 1**.

Table 1. SVM algorthm

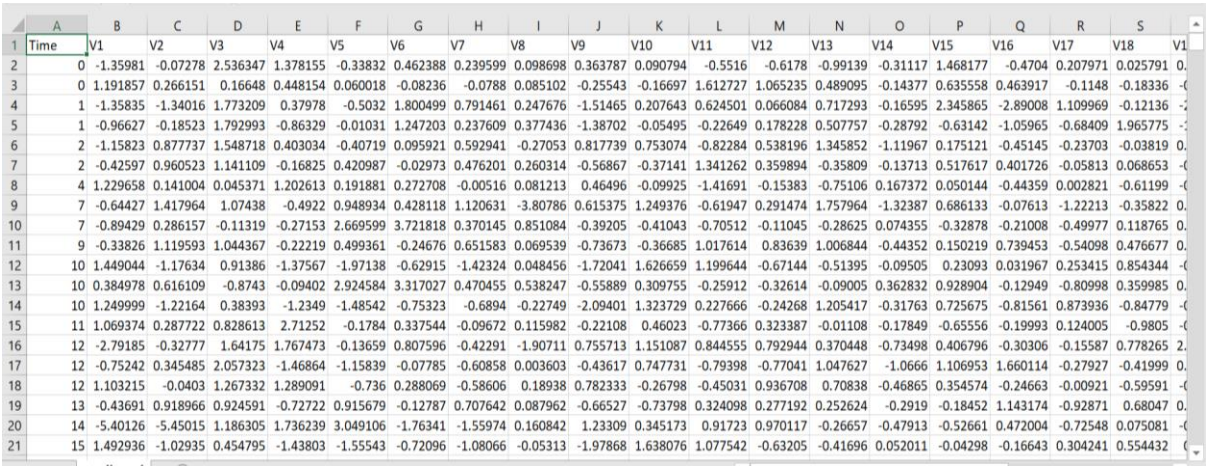
Algorithm 1 : Fraund Detection based on SVM and ANOVA F-Test	
First step	Employ ANOVA F-Test to select important features Compute importance score for feature $F = \frac{\sigma_1^2}{\sigma_2^2}$ select the threshold that maximizes the model's performance Reduce the variables
Then	Make use of the SVM linear-kernel function$\emptyset(Xi)$ Establish the separation hyperplane $W^T \emptyset(Xi) + b = 0$ Sort the data into non-fraud and fraudulent class.

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

4. RESULTS AND DISCUSSION

4.1. Screenshot of the Dataset Temperature

To detect credit card fraud as effectively as possible in highly imbalanced datasets, oversampling method was applied to the imbalanced dataset using SMOTE techniques, followed by classification to improve the performance of Support Vector Machine model. The credit card dataset can be seen in **Figure 2**.

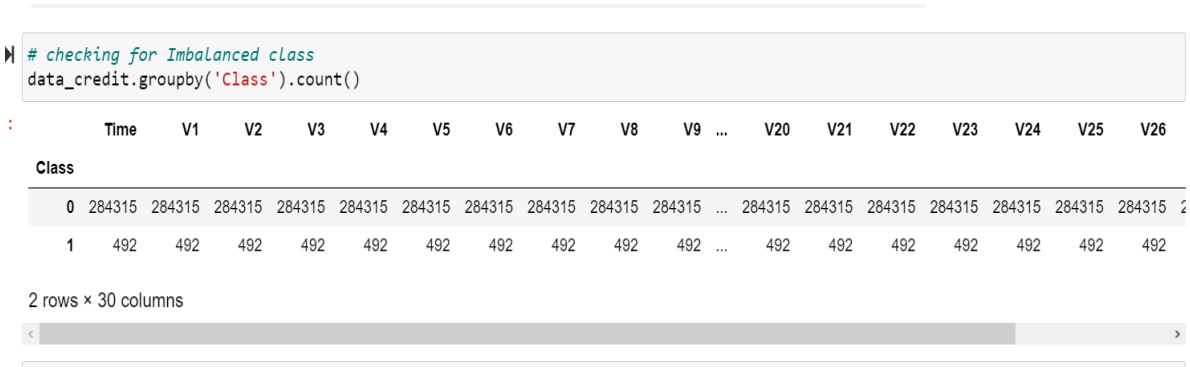


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V1
2	0	-1.35981	-0.07278	2.536347	1.378155	-0.33832	0.462388	0.239599	0.098698	0.363787	0.090794	-0.5516	-0.6178	-0.99139	-0.31117	1.468177	-0.4704	0.207971	0.025791	0.
3	0	1.191857	0.266151	0.16648	0.448154	0.060018	-0.08236	-0.0788	0.085102	-0.25543	-0.16697	1.612727	1.065235	0.489095	-0.14377	0.635558	0.463917	-0.1148	-0.18336	-(
4	1	-1.35835	-1.34016	1.773209	0.37978	-0.5032	1.800499	0.791461	0.247676	-1.51465	0.207643	0.624501	0.066084	0.717293	-0.16595	2.345865	-2.89008	1.109969	-0.12136	-2
5	1	-0.96627	-0.18523	1.792993	-0.86329	-0.01031	1.247203	0.237609	0.377436	-1.38702	-0.05495	-0.22649	0.178228	0.507757	-0.28792	-0.63142	-1.05965	-0.68409	1.965775	-1
6	2	-1.15823	0.877737	1.548718	0.403034	-0.40719	0.095921	0.592941	-0.27053	0.817739	0.753074	-0.82284	0.538196	1.345852	-1.11967	0.175121	-0.45145	-0.23703	-0.03819	0.
7	2	-0.42597	0.960523	1.141109	-0.16825	0.420987	-0.02973	0.476201	0.260314	-0.56867	-0.37141	1.341262	0.359894	-0.35809	-0.13713	0.517617	0.401726	-0.05813	0.068653	-(
8	4	1.229658	0.141004	0.045371	1.202613	0.191881	0.272708	-0.00516	0.081213	0.46496	-0.09925	-1.41691	-0.15383	-0.75106	0.167372	0.050144	-0.44359	0.002821	-0.61199	-(
9	7	-0.64427	1.417964	1.07438	-0.4922	0.948934	0.428118	1.120631	-3.80786	0.615375	1.249376	-0.61947	0.291474	1.757964	-1.32387	0.686133	-0.07613	-1.22213	-0.35822	0.
10	7	-0.89429	0.286157	-0.11319	-0.27153	2.669599	3.721818	0.370145	0.851084	-0.39205	-0.41043	-0.70512	-0.11045	-0.28625	0.074355	-0.32878	-0.21008	-0.49977	0.118765	0.
11	9	-0.33826	1.119593	1.044367	-0.22219	0.499361	-0.24676	0.651583	0.069539	-0.73673	-0.36685	1.017614	0.83639	1.006844	-0.44352	0.150219	0.739453	-0.54098	0.476677	0.
12	10	1.449044	-1.17634	0.91386	-1.37567	-1.97138	-0.62915	-1.42324	0.048456	-1.72041	1.626659	1.199644	-0.67144	-0.51395	-0.09505	0.23093	0.031967	0.253415	0.854344	-(
13	10	0.384978	0.616109	-0.8743	-0.09402	2.924584	3.317027	0.470455	0.538247	-0.55889	0.309755	-0.25912	-0.32614	-0.09005	0.362832	0.928904	-0.12949	-0.80998	0.359985	0.
14	10	1.249999	-1.22164	0.38393	-1.2349	-1.48542	-0.75323	-0.6894	-0.22749	-2.09401	1.323729	0.227666	-0.24268	1.205417	-0.31763	0.725675	-0.81561	0.873936	-0.84779	-(
15	11	1.069374	0.287722	0.828613	2.71252	-0.1784	0.337544	-0.09672	0.115982	-0.22108	0.46023	-0.77366	0.323387	-0.01108	-0.17849	-0.65556	-0.19993	0.124005	-0.9805	-(
16	12	-2.79185	-0.32777	1.64175	1.767473	-0.13659	0.807596	-0.42291	-1.90711	0.755713	1.151087	0.844555	0.792944	0.370448	-0.73498	0.406796	-0.30306	-0.15587	0.778265	2.
17	12	-0.75242	0.345485	2.057323	-1.46864	-1.15839	-0.07785	-0.60858	0.003603	-0.43617	0.747731	-0.79398	-0.77041	1.047627	-1.0666	1.106953	1.660114	-0.27927	-0.41999	0.
18	12	1.103215	-0.0403	1.267332	1.289091	-0.736	0.288069	-0.58606	0.18938	0.782333	-0.26798	-0.45031	0.936708	0.70838	-0.46865	0.354574	-0.24663	-0.00921	-0.59591	-(
19	13	-0.43691	0.918966	0.924591	-0.72722	0.915679	-0.12787	0.707642	0.087962	-0.66527	-0.73798	0.324098	0.277192	0.252624	-0.2919	-0.18452	1.143174	-0.92871	0.68047	0.
20	14	-5.40126	-5.45015	1.186305	1.736239	3.049106	-1.76341	-1.55974	0.160842	1.23309	0.345173	0.91723	0.970117	-0.26657	-0.47913	-0.52661	0.472004	-0.72548	0.075081	-(
21	15	1.492936	-1.02935	0.454795	-1.43803	-1.55543	-0.72096	-1.08066	-0.05313	-1.97868	1.638076	1.077542	-0.63205	-0.41696	0.052011	-0.04298	-0.16643	0.304241	0.554432	(

Figure 2. Credit card dataset.

4.2. Screenshot of the Dataset Temperature

The target value was divided into two categories: non-fraud(0) and fraud(1). The dataset is severely unbalanced, as shown in **Figure 3** and **Figure 4** below. The total amount of non-fraud transactions is 284325, which is much greater than the total number of fraud transactions (492).



```
# checking for Imbalanced class
data_credit.groupby('Class').count()
```

Class	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V20	V21	V22	V23	V24	V25	V26
0	284315	284315	284315	284315	284315	284315	284315	284315	284315	284315	...	284315	284315	284315	284315	284315	284315	284315
1	492	492	492	492	492	492	492	492	492	492	...	492	492	492	492	492	492	492

2 rows × 30 columns

Figure 3. Number of non-fraud and fraud transaction.

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

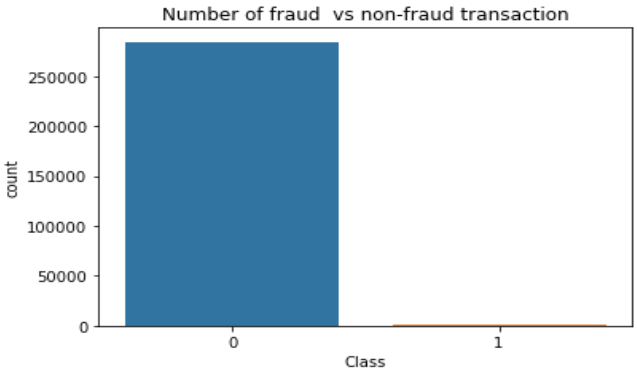


Figure 4. Graphical representation of the transactions in the Dataset.

4.3. Distribution of Classes with Feature Time

The class distribution with respect to time is shown in **Figure 6** below; the column Time needs to be removed from our modeling process because it is clear from the distribution plot below that there is no distinct pattern distinguishing between fraudulent and non-fraudulent transactions with respect to time. This demonstrates that the time feature does not aid in the identification of fraudulent transactions.

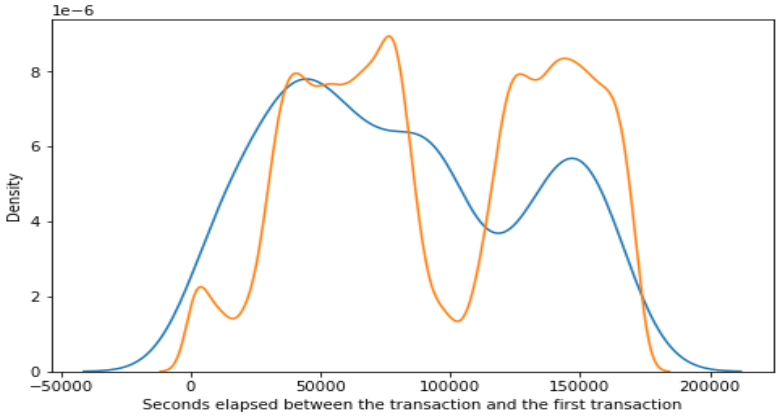


Figure 5. Graphical representation of the transactions in the Dataset.

4.4. Data Scaling with Standardscaler Method

Table 2. First four rows before the standardization of Amount feature.

Time	V1	V2	V3	Amount	Class
0	-1.35981	-0.07278	2.536347	**149.62**	0
0	1.191857	0.266151	0.16648	**2.69**	0
1	-1.35835	-0.18523	1.773209	**378.66**	0
1	-1.35835	0.877737	1.792993	**123.5**	0

Table 3. First four rows after the standardization of Amount feature.

Time	V1	V2	V3	Amount	Class
0	-1.35981	-0.07278	2.536347	**0.244964**	0
0	1.191857	0.266151	0.16648	**-0.342475**	0
1	-1.35835	-0.18523	1.773209	**1.160686**	0
1	-1.35835	0.877737	1.792993	**0.140534**	0

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

The **Table 2** above shows unscaled values of amount feature before applying data scaling method while the **Table 3** shows the scaled values of amount features after the standardization method.

4.5. Data Oversampling using SMOTE

This effort oversampled the data using the Synthetic Minority Oversampling Technique (SMOTE), ensuring that the data was balanced and that the results did not lead to false conclusions. The accuracy of the SVM model still given over 99% despite 492 of the 284,807 transactions in the dataset are fraud, this is unacceptable. SMOTE was used to enhance the SVM model's performance. Show classes before and after oversampling can be seen in the **Table 4**.

Table 4. Feature classes before and after oversampling.

Classes	Before Oversampling	After Oversampling
NON-FRAUD (0)	284315	284315
FRAUD (1)	492	284315

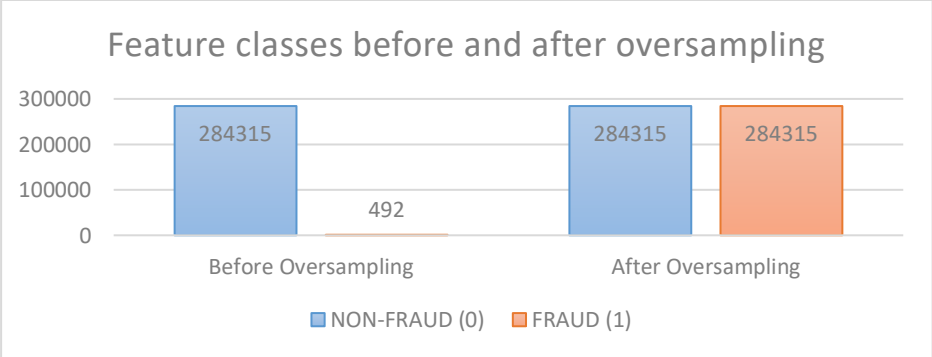


Figure 6. Feature classes before and after oversampling.

The **figure 6** shows how Synthetic Minority Oversampling Techniques (SMOTE) was used to balance the class distribution by oversampling the minority class.

Table 5. Feature Selection with ANOVA F-test.

Features	Before Feature Selection	After Feature Selection
Rows	284315	284315
Columns	29	20

The **Table 5** shows the total number of rows and columns in the dataset before the feature selection and the total number of rows and columns after some important features were selected.

4.6. Performance Evaluation of SVM Model and SVM with SMOTE

Performance evaluation of SVM Model and SVM with SMOTE can be seen in the **Table 6** and **Figure 7**.

Table 6. Performance evaluation of SVM Model and SVM with SMOTE.

Models	Accuracy	Precision score	Recall score	F1-score

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

SVM	99.872	82.051	32.653	46.715
SVM and SMOTE	93.891	97.362	90.289	93.50

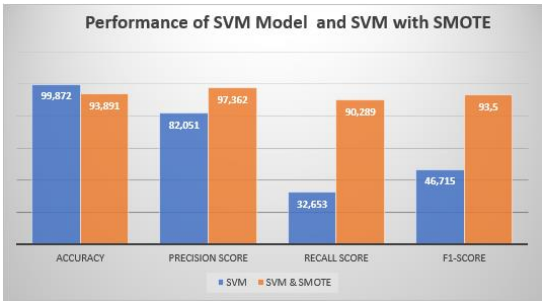


Figure 7. Performance of SVM Model and SVM with SMOTE.

4.7. Discussion of Results

The output of the Classifier (SVM) with a severely unbalanced dataset which is different from the balanced dataset produced by the SMOTE sampling method, as shown in Figure 8. The SVM model's accuracy with an unbalanced dataset is 99.9%, while its precision score is 82% and recall is 32%, indicating that it is not consistent and that it is not very good at spotting suspicious transactions. This is as a result of the model identifying every transaction as legal due to the unbalanced dataset. If the data are imbalanced, accuracy can convey incorrect information. The accuracy, precision, recall, and F1-Score of the SVM Classifier with a balanced dataset are 94.7%, 97.3%, 91.8%, and 94.5%, respectively. The SVM model with SMOTE outperformed the SVM model with imbalanced dataset based on the values of the performance metrics obtained.

5. CONCLUSION

SVM classifier incorporated with SMOTE techniques was implemented in the first stage and compared with ordinary SVM, result obtained from the model indicated that the proposed model outperformed the ordinary SVM in term of accuracy, precision score, Recall and F1 score.The output of the Classifier (SVM) with a severely unbalanced dataset is different from the balanced dataset produced by the SMOTE sampling method, as shown in Figure 8 above. The SVM model's accuracy with an unbalanced dataset is 99.9%, while its precision score is 82% and recall is 32%, indicating that it is not consistent and also it is not very good at spotting suspicious transactions. This is as a result of the model identifying every transaction as legal due to the unbalanced dataset. Along with accuracy, this study also takes into account other parameters like precision, recall, and F1-score. If the data are imbalanced, accuracy can convey incorrect information. The additional measures were also utilized to justify reliability in order to bury that element. The resulting classification model is said to be reliable if the Precision and Recall metrics are close to 1, or close to 100%. In this case, the accompanying F1-score will also be high. The SVM Classifier's accuracy, precision, recall, and F1-Score with a balanced dataset are 93.9%, 97.3%, 90.3%, and 93.5% respectively. The results of the SVM algorithm with SMOTE demonstrated the viability of such an approach in real time, and our methodology aims to provide some insight into the detection of fraud.

6. AUTHORS' NOTE

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

7. REFERENCES

Ahirwar, A., Sharma, N., and Bano, A. (2020). Enhanced SMOTE & fast random forest techniques for credit card fraud detection. *Solid State Technology, 63*(6), 4721-4733.

Ali, M. A., Azad, M. A., Centeno, M. P., Hao, F., and van Moorsel, A. (2019). Consumer-facing technology fraud: Economics, attack methods and potential solutions. *Future Generation Computer Systems, 100*, 408-427.

Ata, O., and Hazim, L. (2020). Comparative analysis of different distributions dataset by using data mining techniques on credit card fraud detection. *Tehnički vjesnik, 27*(2), 618-626.

Budhram, T. (2012). Lost, stolen or skimmed: Overcoming credit card fraud in South Africa. *South African Crime Quarterly*, *40*, 31-37.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Syntheticminority over– sampling technique.*Journal of Artificial Intelligent Research,16*, 321–357.

Dutta, R., Smith, D., Rawnsley, R., Bishop-Hurley, G., Hills, J., Timms, G., & Henry, D. (2015). Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and electronics in agriculture, 111*, 18-28.

Eseye, A. T., Lehtonen, M., Tukia, T., Uimonen, S., & Millar, R. J. (2019). Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems. *IEEE Access*, *7*, 91463-91475.

Kadam, S., Kumari, S., Trivedi, S., & Shah, V. (2021) Credit Card Fraud Detection Using Machine Learning Algorithms. *9*(6), 7496-7499.

Lee, J., and Kwon, K. N. (2002). Consumers' use of credit cards: store credit card usage as an alternative payment and financing medium. *Journal of Consumer Affairs, 36*(2), 239-262.

Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation, 5*(1), 1-27.

Oyelakin, A. M., & Jimoh, R. G. (2020). Towards building an improved botnet detection model in highly imbalance botnet dataset-a methodological framework. *Volume, 3*(3), 2020.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 160.

Sharma, P., Banerjee, S., Tiwari, D., & Patni, J. C. (2021). Machine learning model for credit card fraud detection-a comparative analysis. *Int. Arab J. Inf. Technol., 18*(6), 789-796.

Sivakumar, N., & Balasubramanian, R. (2015). Fraud detection in credit card transactions: classification, risks and prevention techniques. *International Journal of Computer Science and Information Technologies*, *6*(2), 1379-1386.

Trivedi, N. K., Simaiya, S., Lilhore, U. K., & Sharma, S. K. (2020). An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology*, *29*(5), 3414-3424.

DOI: https://doi.org/10.17509/jcs.v5i2.70802
e- ISSN 2723-4088