IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111205

E-Commerce Fraud Detection Using Support Vector Machine and Naïve Bayes Algorithm

Wowon Priatna[1], Joni Warta[2], Tyastuti Sri Lestari[3]

Department of Computer Science, Universitas Bhayangkara Jakarta Raya, Jakarta, Indonesia[1,2]

Department of Industrial, Universitas Bhayangkara Jakarta Raya, Jakarta, Indonesia[3]

Abstract: The prevalence of online fraud cases, including e-commerce fraud, is rising as a result of technological advancements and the speed with which cybercriminals can change their methods of operation. Scams are nothing new, but as the frequency of transactions without currency rises, so does the trend of online fraud. People are purchasing more goods online as a result of the COVID-19 quarantine because they want to be safe or because the items, they require are hard to get in the shuttered local stores. The best course of action in this circumstance is to implement a fraud prevention service that automatically identifies fraudulent behaviour patterns, associated with the time, place, and device name associated with the login or transaction. This will prevent fraudsters from using the data they stole. You can halt fraudsters before they start a transaction by spotting suspicious activity on an account. Through relevant historical data from databases and machine learning techniques, this study aims to identify fraud patterns in e-commerce transactions. Based on email, payment methods, payment method providers, and transaction volume, this research will train a computer or system that can predict fraud patterns. Machine learning must be used to improve fraud protection in e-commerce since it allows machines to be analysed using learning algorithms. Support vector machine and naive Bayes will be the algorithms employ.

Keywords: Machine Learning, Support Vector Machine, Naïve Bayes, Classification, Fraud E-Commerce.

I. INTRODUCTION

In a world when most communication occurs online and our virtual environment is saturated with adverts for exciting goods and services to purchase, it's difficult to undervalue the significance of the marketplace. Meanwhile, it is obvious that many criminals are attempting to profit from it by infiltrating user data with malware and phishing schemes. Statistics indicate a high rate of e-commerce fraud. By 2020, it is predicted that e-commerce sales would total $630 billion (or more), while fraud will cost the economy an estimated $16 billion. Nearly a third of all American e-commerce transactions take place on Amazon, and its sales are growing by 15% to 20% annually. The third time, ecommerce expenditure climbed by 57% between 2018 and 2019, For the third time in US history, more money was spent online than in physical stores, thanks to a surge in ecommerce of 57%. User feedback and surveys demonstrate that e-commerce companies have tools and solutions to address fraud concerns, which makes users feel much more at ease and confident when making online payments.

he stats above are impressive. In addition, it shows that we can observe a very serious lack of opportunity to solve this problem. One method to prevent financial fraud is to use machine learning techniques with the main features of this technology enabling it to prevent, detect, and combat fraud more effectively [1]. Research [2] uses machine learning for fraud detection of e-commerce transactions by classification, identification of credit card fraud using KNN and nave Bayes [3][4], mapping of forms of fraud and prevention has been carried out [5], Machine learning techniques can also be used for pattern detection tire fraud using the support vector machine algorithm [1].

Research [6] detecting fraudulent credit card transactions using nave Bayes and rewarding data using smote, using nave bayes detection for fraud on call data [7], for fraud detection by verifying transactions using the Support Vector Machine Algorithm [8], predicting financial reporting using Support vector Machine [9]

Several studies to compare the accuracy of the Vector Machine and Nave Bayes support algorithms include predicting employee recruitment[10], sentiment analysis for gadgets[11], classification of diabetics[12] and classification of electrical grid stability resulting in SVM accuracy of 98.9% while Naïve Bayes by 97.64%.

From the problems and research above that the Support vector machine and nave Bayes have never been used to predict or classify fraud in E-Commerce, this research will solve the problem in detecting E-Commerce fraud using the Support vector machine and nave Bayes as proven from several studies in solve prediction and classification problems

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111205

II. METHODOLOGY

The following are the stages in this research:

1) Future choice. The identification of process characteristics in the data set collected from Kaggle with 7 variables and 157 data records is the first step in the classification process.
2) Pre-processing is the process of extracting, modifying, normalizing, and scaling new features for use by machine learning algorithms. To turn unprocessed data into high-quality data, pre-processing is performed. PCA (Principal Component Analysis) is used in this study's pre-processing along with feature extraction, transformation, normalization, and scaling.
3) The SMOTE technique builds duplicate synthetic data as many times as the desired proportion between k randomly chosen and positive classes, then locates the k nearest neighbors for positive classes..
4) Produce the support vector machine and the naive Bayes model. The Nave Bayes algorithm and support vector are used to forecast online shopping fraud.
5) Test data and accuracy. Using a confusion matrix, it is now necessary to determine whether the model that has been created has the correct accuracy by testing the data.

III. RESULT AND DISCUSSION

E-Commerce transaction fraud dataset used is 167 records. The dataset must be free from noise and valid before the classification process is carried out with several scenarios that have been prepared.

The dataset must be in accordance with the design and requirements of the Naïve Bayes algorithm and SVM free from dataset problems such as data intervals.

A. Naïve Bayes Classification and Support Vector Machine

The results of testing the Naïve Bayes and SVM classifications use a confusion matrix consisting of precision, recal and accuracy. Then the test results are shown in Table 1.

TABLE I SVM AND NAÏVE BAYES CLASSIFICATION TEST RESULTS

Algorithm	Recall	Precision	Accuracy
SVM	100%	68,8%	71%
Naïve Bayes	95%	70%	61%

From Table 1 shows that the classification generated by SVM gets an accuracy value of 71% better than Naïve Bayes which obtains an accuracy of 61%.

B. Naïve Bayes +PCA Classification and Support Vector Machine +PCA

rom the results of data processing to determine the tendency of E-Commerce fraud using the Support Vector Machine and Naïve Bayes algorithms by using feature dimension reduction with Principal Component Analysis.

The following comparison of the results of accuracy, precision and recall is shown in Table II.

TABLE II I SVM+PCA AND NAÏVE BAYES +PCA CLASSIFICATION TEST RESULTS

Algorithm	Recall	Precision	Accuracy
SVM+PCA	100%	66%	68%
NB+PCA	95%	66%	64%

The results from table 2 can be concluded that the SVM algorithm with PCA combination still gets a better accuracy value than the Naïve Bayes + PCA algorithm with 68% accuracy SVM + PCA and 64% accuracy NB + PCA respectively.

C. Naïve Bayes Classification and Support Vector Machine with PCA+SMOTE Combination

From the results of data processing to determine the tendency of E-Commerce fraud using the Support Vector Machine and Naïve Bayes algorithms by using feature dimension reduction with Principal Component Analysis and balancing classes with SMOTE. The following comparison of the results of accuracy, precision and recall is shown in Table III.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111205

TABLE IIIII COMPARISON OF SVM AND NB CLASSIFICATION RESULTS WITH PCA+SMOTE

Algorithm	Recall	Precision	Accuracy
SVM+PCA+SMOTE	72%	40%	77%
NB+PCA+SMOTE	81%	60%	59%

The results from Table III can be concluded that the SVM algorithm with PCA combination still gets better accuracy values than the Naïve Bayes + PCA algorithm with 77% accuracy SVM + PCA and 59% accuracy NB + PCA.

D. Overall Classification Test Results

Table IV and Fig. 1 show that the Support Machine algorithm using Principal Component Analysis (PCA) has no effect on classification performance because before using PCA the accuracy value of 71% after using PCA decreased 3% to 68%. For the effect of using oversampling, balancing class/target data has an effect so that the accuracy of SVM before using SMOTE is 71% and after using SMOTE it becomes 77%, it increases by 6%.

For Naïve Bayes, the use of PCA has the effect of increasing the accuracy value by 3% from 61% before using PCA, increasing to 64%. Meanwhile, the use of SMOTE for Naïve Bayes did not decrease the accuracy value from 61% to 59%, down 4%.

From the results of Table IV, it can be used as a reference for the classification of fraud in E-Commerce transactions, it is recommended to use the Support Vector Machine Algorithm with optimization using the SMOTE Algorithm which functions as over sampling that can balance class/target.

TABLE IVV CLASSIFICATION PERFORMANCE TEST COMPARISON RESULTS

	SVM	NB	SVM+ PCA	NB+ PCA	SVM+ PCA+ SMOTE	NB+ PCA+ SMOTE
Recall	100%	95%	100%	95%	72%	81%
Precision	68.8%	70%	66%	66%	40%	60%
Accuracy	71%	61%	68%	64%	77%	59%

.

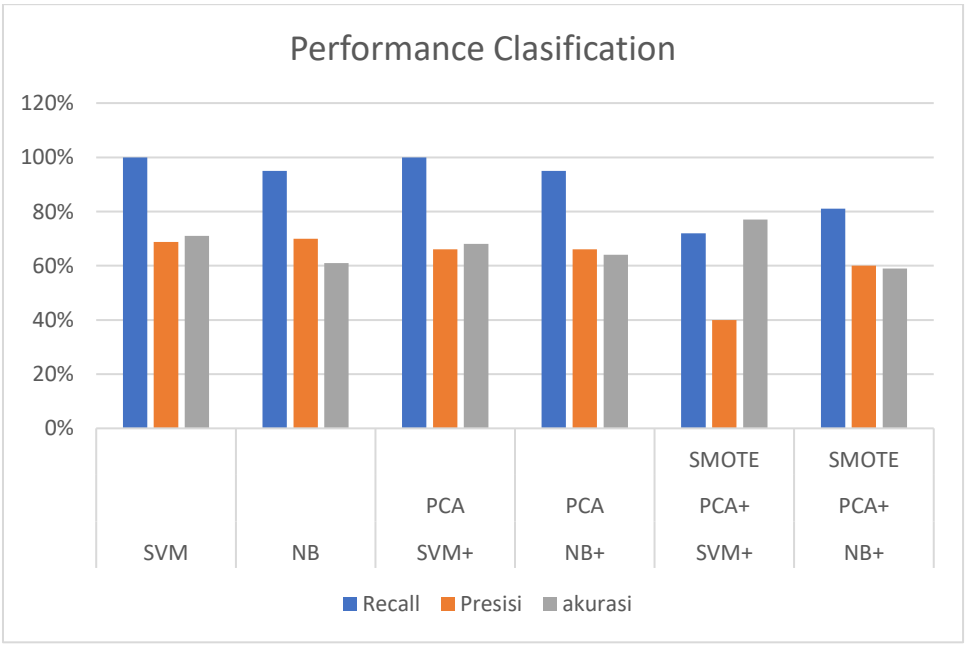

Fig. 1 Performance Classification

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111205

IV. CONCLUSION

The conclusion of this research is:

1. The application of the Support Vector Machine algorithm has a better accuracy value than nave bayase, where SVM gets 71% accuracy while Nave bayase is 61%. So to classify e-commerce fraud data, the SVM algorithm is recommended to be used.
2. Support Machine algorithm using Principal component Analysis (PCA) has no effect on classification performance because before using PCA the accuracy value is 71% after using PCA it drops 3% to 68% while for Naïve Bayes PCA use has the impact of increasing the accuracy value by 3% from before using PCA by 61% increased to 64%.
3. The application of SMOTE in Classification for SVM and Naïve Bayes algorithms is for SVM before using SMOTE by 71% and after using SMOTE it becomes 77% an increase of 6% while the use of SMOTE for Nave Bayes does not have an impact on the accuracy value from 61% to 59% down 4%.

REFERENCES

[1] N. K. Gyamfi and J. D. Abdulai, "Bank Fraud Detection Using Support Vector Machine," 2018 IEEE 9th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2018, no. November, pp. 37–41, 2019, doi: 10.1109/IEMCON.2018.8614994.
[2] H. Animesh, K. M. Subrata, G. Amit, M. Arkomita, and A. Mukherje, "Heart Disease Diagnosis and Prediction Using Machine Learning," Adv. Comput. Sci. Technol., vol. 10, no. 7, pp. 2137–2159, 2017, [Online]. Available: http://www.ripublication.com.
[3] D. Kaur, "Machine Learning Approach for Credit Card Fraud Detection (KNN & Naïve Bayes)," Mach. Learn. Approach Credit Card Fraud Detect. (KNN Naïve Bayes)(March 30, 2020), 2020.
[4] W. Priatna and R. Purnomo, "Comparation of Support Vector Machine and Artificial Neural Network Algorithm for Lecturer Performance Classification," Ijarcce, vol. 10, no. 9, pp. 7–11, 2021, doi: 10.17148/ijarcce.2021.10901.
[5] N. K. Arista Dewi and L. P. Mahyuni, "Pemetaan Bentuk Dan Pencegahan Penipuan E-Commerce," E-Jurnal Ekon. dan Bisnis Univ. Udayana, vol. 9, p. 851, 2020, doi: 10.24843/eeb.2020.v09.i09.p03.
[6] M. Y. Sahroni, N. A. Setifani, and D. N. Fitriana, "Analisis perbandingan algoritma Naïve Bayes, k-Nearest Neighbor dan Neural Network untuk permasalahan class-imbalanced data pada kasus credit card fraud dataset," Teknologi, vol. 11, no. 2, pp. 69–73, 2021, doi: 10.26594/teknologi.v11i2.2393.
[7] N. P. Ilna, A. Rachmadita, and S. Edi, "ANALISIS DAN DETEKSI FRAUD PADA DATA PANGGILAN MENGGUNAKAN ALGORITMA NAÏVE BAYES PADA PT XYZ," e-Proceeding Eng., vol. 7, no. 2, pp. 6647–6655, 2020.
[8] Y. Yazid and A. Fiananta, "Mendeteksi Kecurangan Pada Transaksi Kartu Kredit Untuk Verifikasi Transaksi Menggunakan Metode Svm," Indones. J. Appl. Informatics, vol. 1, no. 2, pp. 61–66, 2017.
[9] Puspandoyo et al., "Prediksi Kualitas Laporan Keuangan Kementerian Negara / Lembaga Menggunakan," vol. 3, pp. 15–36, 2022.
[10] N. A. Sinaga, B. H. Hayadi, and Z. Situmorang, "Perbandingan Akurasi Algoritma Naïve Bayes, K-Nn Dan Svm Dalam Memprediksi Penerimaan Pegawai," J. Tek. Inf. dan Komput., vol. 5, no. 1, p. 27, 2022, doi: 10.37600/tekinkom.v5i1.446.
[11] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
[12] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," J. Inf. Technol. Ampera, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.isssue3.year2020.page133-143.

