Indonesian Journal of Computer Science
ISSN 2302-4364 (*print*) dan 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam, No. 1, Padang, Indonesia, Telp. (0751) 7056199, 7058325
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Dampak Pengambilan Sampel Data untuk Optimalisasi Data Tidak Seimbang pada Klasifikasi Penipuan Transaksi E-Commerce

Wowon Priatna
wowon.priatna@dsn.ubharajaya.ac.id
Universitas Bhayangkara Jakarta Raya

Informasi Artikel

Diterima : 31 Jan 2024
Direview : 23 Feb 2024
Disetujui : 1 Apr 2024

Kata Kunci

E-commerce, Imbalance Class, SMOTE, SVM, KNN

Abstrak

Tujuan dari penelitian ini adalah untuk mengatasi masalah pengklasifikasian dan prediksi data yang tidak seimbang terkait dengan kondisi transaksi E-Commerce. Menjamurnya transaksi e-commerce menimbulkan potensi permasalahan: penipuan dalam pembelian e-commerce. Kasus penipuan e-niaga terus meningkat setiap tahun sejak tahun 1993. Menurut survei tahun 2013, untuk setiap $100 transaksi e-niaga, terdapat kerugian sebesar 5,65 sen akibat penipuan. Mendeteksi penipuan merupakan pendekatan yang efektif untuk meminimalkan terjadinya aktivitas penipuan dalam transaksi e-commerce. Pembelajaran menjadi metode yang semakin dapat diandalkan untuk memprediksi keadaan. Tidak adanya keseimbangan antara data yang curang dan tidak curang mengakibatkan klasifikasi menjadi bias. Algoritma SMOTE diperlukan untuk mencapai keseimbangan data. Selanjutnya peristiwa transaksi akan diklasifikasikan menggunakan algoritma Support Vector Machine, K-Nearest Neighbor, Naive Bayes, dan C45, dengan mempertimbangkan hasil penyeimbangan data. Di antara algoritma SVM, KNN, dan C45, metode Naive Bayes menunjukkan nilai akurasi tertinggi. Oleh karena itu, disarankan untuk menggunakan teknik ini untuk tujuan mengidentifikasi kondisi e-commerce..

Keywords

E-commerce, Imbalance Class, SMOTE, SVM, KNN

Abstract

This research aims to solve the problem of classification and prediction of unbalanced data related to E-Commerce transaction conditions. The large number of e-commerce transactions raises the possibility of a new problem—fraud in e-commerce transactions. E-commerce fraud has increased every year since 1993. A 2013 report stated that every $100 in e-commerce turnover in fraud causes a loss of 5.65 cents. One way to reduce the amount of fraud in e-commerce transactions is to detect fraud. Increasingly, learning is one way to predict conditions. Fraudulent data is not balanced with non-fraudulent data so that the classification will be biased. So, the SMOTE algorithm should be used to balance the data. Then, based on the results of data balancing, the Support vector machine, K-Nearst Neighbor, Naive Bayes, and C45 algorithms will be used to classify transaction events. The Naive Bayes algorithm has the highest accuracy value compared to the SVM, KNN, and C45 algorithms. Therefore, this algorithm is recommended for use to detect the state of e-commerce.

A. Pendahuluan

Peran pasar sangat penting dalam masyarakat di mana sebagian besar komunikasi terjadi secara online, dan lingkungan virtual kita dipenuhi dengan iklan yang menarik untuk berbagai produk dan layanan. Namun, banyak penjahat yang mencoba mengeksploitasi situasi ini dengan menggunakan penipuan dan perangkat lunak berbahaya untuk membahayakan data pelanggan. Statistik menunjukkan prevalensi penipuan e-commerce yang signifikan. Perkiraan kerugian akibat aktivitas penipuan diperkirakan mencapai $16 miliar, sedangkan perkiraan penjualan e-commerce kemungkinan akan melampaui $630 miliar pada tahun 2020. Amazon mendominasi pasar e-commerce di Amerika Serikat, mencakup hampir seluruh transaksi. Tingkat pertumbuhan tahunannya berkisar antara lima belas persen hingga dua puluh persen. Pada tahun 2019, terjadi peningkatan signifikan sebesar 57% pada belanja konsumen di toko ritel fisik dibandingkan dengan pembelian online di Amerika Serikat. Menurut pengguna, perusahaan e-commerce memiliki potensi aktivitas penipuan dan cara untuk mengatasinya. Hasilnya, pengguna merasa lebih nyaman dan percaya diri saat melakukan pembayaran online.

Semakin banyak orang di Indonesia yang menggunakan Internet mendorong para pelaku pasar untuk mencoba mengembangkan bisnis mereka melalui media Internet. Ini digunakan untuk membahas bisnis e-dagang. Menurut data statistik yang dihimpun oleh Statista.com, penjualan ritel e-commerce di Indonesia diperkirakan meningkat 133,5% dari posisi tahun 2017 menjadi US$16,5 miliar, atau sekitar Rp 219 triliun, pada tahun 2022. Untuk mendorong pertumbuhan ini, pelanggan semakin mudah berbelanja berkat kemajuan teknologi. Penipuan e-commerce adalah masalah baru yang muncul sebagai akibat dari banyaknya transaksi e-commerce. Penipuan e-commerce telah meningkat setiap tahun sejak tahun 1993. Menurut laporan tahun 2013, penipuan menyebabkan kerugian sebesar 5,65 sen per $100 omset perdagangan e-niaga. Pada tahun 2019, penipuan telah mencapai lebih dari 70 triliun dolar. Mendeteksi keadaan adalah salah satu cara untuk mengurangi jumlah kejadian dalam transaksi e-commerce. Deteksi kondisi adalah salah satu cara untuk memprediksi kondisi pembelajaran mesin untuk kondisi transaksi e-commerce[1][2]dan penipuan kartu kredit[3]. Klasifikasi adalah salah satu jenis algoritma pembelajaran mesin[4][5].

Salah satu kelemahan algoritma klasifikasi adalah jika data atau kelas yang ditargetkan tidak seimbang maka hasilnya akan bias[6][7], dan sifat karakteristik data[8]. Penelitian [9]yang menggunakan algoritma SMOTE untuk mengurangi dampak gangguan data pada dataset kondisi kartu kredit menangani ketidakseimbangan data pada lima dataset dari berbagai aplikasi[10][11][12].

Beberapa penelitian telah menggunakan algoritma SMOTE sebelum proses klasifikasi untuk mengatasi ketidakseimbangan data. Misalnya pada penelitian ini [13] menggunakan Support Vector Machine (SVM) untuk menangani ketidakseimbangan data sebelum klasifikasi, K-Nearst Neighbor (KNN) dalam pemilihan mahasiswa menggunakan data sampling sebelum klasifikasi[14], optimasi akurasi C.45 untuk klasifikasi penyakit jantung [15][16], serta SVM untuk menangani ketidakseimbangan data sebelum analisis sentimen tentang pemilu 2024[17].

Tujuan dari penelitian ini adalah untuk melakukan optimasi akurasi algoritma klasifikasi dari model klasifikasi fraud dalam transaksi e-commerce. Data e-commerce mempunyai data yang tidak seimbang maka akan di gunakan algoritma sampling untuk menyeimbangkan data class, sehingga algoritma KNN, SVM, Niave Bayes dan C.45 diharapkan mendapatkan akurasi yang lebih baik.

B. Metode Penelitian

Salah satu bidang ilmu data yang menggunakan metode pembelajaran mesin adalah penambangan data. Data mining digunakan sebagai sumber penelitian di banyak domain penelitian[18]. Menurut penelitian, prediksi kinerja karyawan ditingkatkan dengan menggunakan data mining, memungkinkan mereka membuat keputusan yang lebih tepat[19]. Kondisi di sektor keuangan [20] dan investasi di pasar transaksi [21] dapat ditentukan dengan menggunakan temuan klasifikasi. Kinerja klasifikasi dipengaruhi oleh penyeimbangan data kelas yang lebih baik[20][22]. Nilai akurasi sebesar 14% diperoleh dengan melakukan oversampling data sehingga mengurangi kesalahan dalam kategorisasi data. Nilai yang diperoleh C45 lebih unggul dibandingkan Bayesian, jaringan syaraf tiruan, dan pohon keputusan.

1. Dataset

Penelitian ini menggunakan dataset penipuan e-commerce yang bersumber dari Kaggle. Dataset terdiri dari 1700 record, dataset yang tergolong fraud sebanyak 624 record, dan rasio data fraud sebesar 0,36. SMOTE (Synthetic Minority Oversampling Technique)[23] meminimalkan ketidakseimbangan kelas dalam dataset transaksi fraud dengan menghasilkan data sintesis, sehingga total data terdiri dari 1378 record, dataset yang tergolong fraud adalah 689 record, rasio data fraud adalah 0,5. Hasil dari sebelum dan smote ditunjukan pada gambar I.

2. Pre-Prosesing Data

Pra-pemrosesan adalah fase berikutnya. Data akan melalui dua tahap pengolahan yaitu metode min max scaler digunakan untuk menskalakan data pada tahap pertama [24]. Pelatihan algoritma Naïve Bayes, KNN, C45, dan Support Vector Machine akan menjadi tantangan karena periode yang panjang ini.

3. Data Sampling

Synthetic Minority Oversampling (SMOTE) adalah metode pengambilan sampel data yang digunakan. SMOTE adalah pendekatan oversampling yang populer untuk pemecahan masalah dan relaksasi. Akibatnya, poin data buatan dari kelas minoritas dihasilkan [25]. Mereka menganggap bahwa dataset pelatihan berisi n sampel dan ruang fitur memiliki m fitur. Pertama, sampel acak x dipilih dari kelas minoritas. Ini juga mencakup k tetangga terdekat (yaitu tetangga yang memiliki jarak Euclidean terkecil) dalam ruang fitur. Selanjutnya, tetangga acak terdekat Anda dipilih dan ditunjuk sebagai KNN. Persamaan (1) digunakan untuk menghitung titik data sintetik baru.

$$X^{SMOTE} = x + u * (X^{NN} - x) \quad (1)$$

Misalkan u adalah bilangan acak yang diambil dari Distribusi seragam, yang berkisar antara 0 hingga 1. Prosedur berlanjut hingga mencapai jumlah sampel buatan yang ditentukan.

4. Klasifikasi

Pada tahapan ini adalah klasifikasi akan dibuat model klasifikasi dari data transaksi e-commerce yang telah dilakukan penyamaan class oleh algoritma SMOTE*. Model klasifikasi yang akan dibangun untuk deteksi fraud dalam e-commerce adalah Naieve bayes, SVM, C.45 dan KNN.*

5. Confusion Matrix

Confusion Matrix mengevaluasi kinerja klasifikasi[6]. False Positive (FP) dan False Negative (FN) mengacu pada contoh di mana kelas positif dan negatif salah diklasifikasikan. Sebaliknya, True Negative (TN) mengacu pada contoh di mana kelas positif dan negatif diklasifikasikan dengan benar. Matriks kebingungan dapat digunakan untuk menghitung ukuran kinerja seperti akurasi, presisi, dan perolehan. Akurasi adalah kriteria yang sering digunakan untuk mengevaluasi kinerja klasifikasi. Namun demikian, jika digunakan dalam kelompok kelas yang berbeda, kriteria ini akan kurang relevan karena kelompok minoritas tidak akan memberikan kontribusi besar terhadap kebenaran kriteria tersebut. Perolehan kembali dan presisi direkomendasikan sebagai kriteria evaluasi. Persamaan (3), (2), dan (5) menyajikan rumus perhitungan recall, akurasi, dan skor F1. Metode-metode ini digunakan untuk meningkatkan akurasi kategorisasi penelitian ini.

$$Akurasi = \frac{True\ Positif + True\ Negatif}{True\ Positif + True\ Negatif + False\ Negatif + False\ Positif} \quad (2)$$

$$Recall = \frac{True\ Positif}{True\ Negatif + False\ Positif} \quad (3)$$

$$Precision = \frac{True\ Positif}{True\ Positif + False\ Positif} \quad (4)$$

$$F1 - Score = \frac{2\ x\ Precicion\ x\ Recall}{Precision + Recall} \quad (5)$$

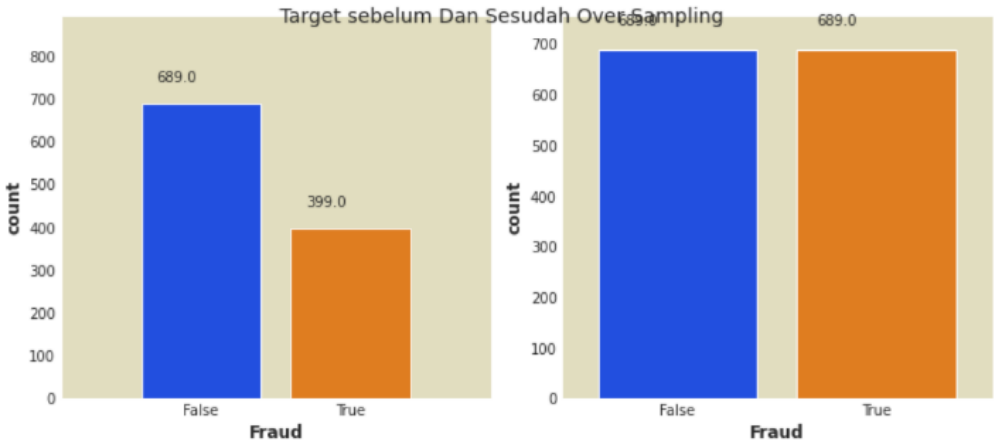
C. Hasil dan Pembahasan

Dataset yang digunakan untuk penipuan transaksi E-Commerce terdiri dari 1700 catatan. Sebelum melakukan proses klasifikasi menggunakan beberapa skenario yang telah ditentukan, kumpulan data harus sah dan bebas dari gangguan apa pun. Kumpulan data harus mematuhi spesifikasi dan persyaratan algoritma Naïve Bayes, C45, KNN, dan SVM dan harus bebas dari masalah kumpulan data apa pun seperti data interval[26].

1. Support Vector Machine

Setelah tahap pra-pemrosesan data selesai, percobaan awal melibatkan penerapan algoritma klasifikasi Support Vector Machine (SVM). Percobaan dilakukan sebelum dan sesudah data diseimbangkan menggunakan Support Vector

Machines (SVM), seperti digambarkan pada Gambar 1. Tabel 1 tersebut menampilkan kategorisasi yang dihasilkan oleh Support Vector Machine (SVM).



Gambar 1. Visualisasi Sebelum dan Sesudah Oversampling

Setelah seluruh proses selesai, kelas/target mencapai keseimbangan, seperti yang digambarkan pada Gambar 1. Sebelum oversampling, data target terdiri dari 624 instance berlabel True dan 1076 instance berlabel False. Setelah proses oversampling, SMOTE menghasilkan 689 instance berlabel True dan 689 instance berlabel False. Tabel I menampilkan hasil kategorisasi Support Vector Machine (SVM) sebelum dan sesudah pemerataan data menggunakan SMOTE. Seperti yang ditunjukkan oleh matriks konfusi, hasil pengujian menunjukkan dampak penggunaan SMOTE pada klasifikasi SVM. Teknik augmentasi ini meningkatkan perolehan, presisi, akurasi, dan nilai F1.

Tabel 1. Hasil Klasifikasi SVM

Algorithm	Recall	Presisi	Accurracy	F1 Score
SVM	0.74	0.85	0.86	0.79
SVM + SMOTE	0.95	0.83	0.91	0.88

2. Naïve Bayes

Tabel 2 menampilkan hasil klasifikasi Naive Bayes sebelum dan sesudah penyeimbangan data menggunakan SMOTE. Seperti yang ditunjukkan oleh matriks chaos, hasil pengujian menunjukkan bahwa akurasi klasifikasi algoritma Naive Bayes relatif lebih buruk ketika SMOTE diterapkan dibandingkan ketika SMOTE tidak digunakan.

Tabel 2. Hasil Klasifikasi Naïve Bayes

Algorithm	Recall	Presisi	Accuracy	F1 Score
Naïve Bayes	0.38	0.86	0.75	0.53
Naïve Bayes + SMOTE	0.39	0.70	0.66	0.50

3. K-Nearst Neighbor

Tabel 3 menampilkan hasil klasifikasi Naive Bayes sebelum dan sesudah penyeimbangan data menggunakan SMOTE. Temuan pengujian, seperti yang ditunjukkan oleh matriks chaos, menunjukkan bahwa pemanfaatan SMOTE dalam proses klasifikasi KNN menghasilkan peningkatan dalam recall, presisi, akurasi, dan skor F1.

Tabel 3. Hasil Klasifikasi KNN

Algorithm	Recall	Presisi	Accuracy	F1 Score
KNN	0.86	0.98	0.94	0.92
KNN + SMOTE	0.98	0.99	0.99	0.99

4. Decision Tree C45

Table 4 displays the outcomes of the C45 Decision Tree categorization both before and following data balancing by SMOTE. According to the data in table 4, the utilization of the confusion matrix reveals that the implementation of SMOTE improves the recall, precision, accuracy, and F1 scores of the C45 classification results..

Tabel 4. Hasil Klasifikasi

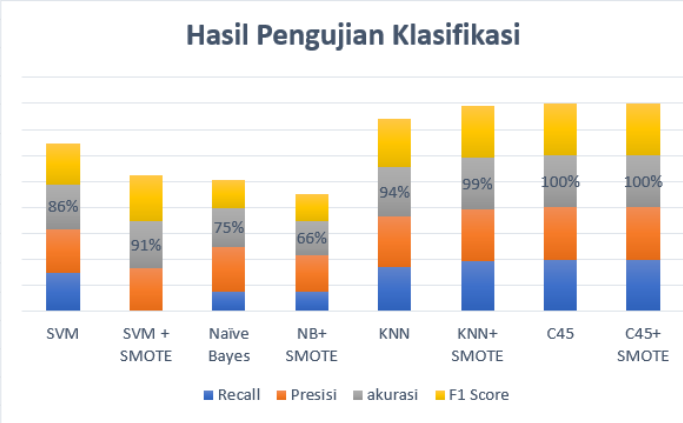
Algorithm	Recall	Presisi	Accuracy	F1 Score
C45	1.00	1.00	1.00	1.00
C45 + SMOTE	1.00	1.00	1.00	1.00

5. Hasil Evaluasi Klasifikasi

Hasil klasifikasi yang diperoleh setelah oversampling data menggunakan SMOTE selanjutnya dievaluasi melalui uji matriks konfusi untuk menilai kinerja pengklasifikasi SVM, Naïve Bayes, KNN, dan C45 berdasarkan Recall Value, Precision, Accuracy, dan F1 Score. Nilai terkait disajikan pada Tabel 5.

Tabel 5. Hasil Klasifikasi

	SVM	SVM + SMOTE	Naïve Bayes	NB+ SMOTE	KNN	KNN+ SMOTE	C45	C45+ SMOTE
Recall	74%	74%	38%	39%	86%	98%	100%	100%
Presisi	85%	83%	86%	70%	98%	99%	100%	100%
akurasi	86%	91%	75%	66%	94%	99%	100%	100%
F1 Score	79%	88%	53%	50%	92%	99%	100%	100%



Gambar 2. Hasil Visualisasi Klasifikasi

Hasil yang dicapai dengan metode klasifikasi sebelum dan sesudah penggunaan SMOTE dapat dilihat pada Gambar 2.

- Hasil klasifikasi yang diperoleh sebelum implementasi algoritma SMOTE menunjukkan bahwa SVM mencapai tingkat akurasi sebesar 86%, Naïve Bayes mencapai 75%, KNN mencapai 94%, dan C45 mencapai 100%. Hasil klasifikasi setelah menggunakan algoritma SMOTE adalah SVM mendapatkan akurasi=91%, Naïve Bayes=66%, KNN=99%, C45=100%.
- Hasil klasifikasi yang diperoleh setelah penerapan algoritma SMOTE adalah sebagai berikut: SVM mencapai akurasi 91%, Naïve Bayes mencapai 66%, KNN mencapai 99%, dan C45 mencapai 100%.
- Hasil klasifikasi yang diperoleh setelah penerapan algoritma SMOTE adalah sebagai berikut: SVM mencapai recall rate sebesar 74%, Naïve Bayes mencapai 39%, KNN mencapai 98%, dan C45 mencapai 100%.
- Presisi klasifikasi SVM setelah penerapan algoritma SMOTE adalah 83%, sedangkan Naïve Bayes mencapai presisi 70%, KNN mencapai 94%, dan C45 mencapai 100%.
- Hasil klasifikasi yang diperoleh sebelum penerapan algoritma SMOTE menunjukkan bahwa SVM memperoleh F1 Score sebesar 79%, Naïve Bayes memperoleh 53%, KNN memperoleh 92%, dan C45 memperoleh 100%.
- Hasil klasifikasi yang diperoleh setelah penerapan algoritma SMOTE adalah sebagai berikut: SVM memperoleh F1 Score sebesar 88%, Naïve Bayes memperoleh 50%, KNN memperoleh 99%, dan C45 memperoleh 100%.
- Pendekatan Decision Tree menunjukkan nilai akurasi paling fantastis dalam mengklasifikasikan penipuan E-Commerce setelah dilakukan oversampling menggunakan SMOTE, melampaui algoritma SVM, KNN, dan Naïve Bayes.

D. Simpulan

Kumpulan data tersebut merupakan skenario e-niaga yang ditandai dengan kelas/target yang tidak seimbang dan frekuensi yang bervariasi antara kelas penipuan dan non-penipuan. Algoritma Synthetic Minority Over Sampling Technique (SMOTE) mencapai keseimbangan data. Empat metode klasifikasi, yaitu SVM, Naïve Bayes, KNN, dan C45 digunakan untuk mengkategorikan dataset kondisi E-Commerce. Temuan yang diperoleh dari penelitian ini adalah sebagai berikut: Dataset yang digunakan adalah skenario e-commerce dengan kelas/target yang

tidak seimbang antara kategori penipu dan non-penipuan. Algoritma Synthetic Minority Over Sampling Technique (SMOTE) mencapai keseimbangan data. Dataset kondisi E-Commerce diklasifikasikan menggunakan empat algoritma klasifikasi: SVM, Naïve Bayes, KNN, dan C45. Temuan yang diperoleh dari penelitian ini adalah sebagai berikut:Sebelum digunakan SMOTE class Fraud=624 dan non fraud=1076, setelah digunakan algoritma SMOTE class menjadi seimbang dengan fraud menjadi 689 dan non fraud menjadi 689.

Metode C45 menunjukkan akurasi yang unggul dibandingkan dengan algoritma SVM, KNN, dan Naïve Bayes ketika mengkategorikan situasi E-Commerce setelah menerapkan oversampling SMOTE.

E. Referensi

[1] R. Jhangiani, D. Bein, and A. Verma, "Machine Learning Pipeline for Fraud Detection and Prevention in E-Commerce Transactions," *2019 IEEE 10th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf. UEMCON 2019*, pp. 0135–0140, 2019, doi: 10.1109/UEMCON47517.2019.8992993.

[2] J. Liu, X. Gu, and C. Shang, "Quantitative Detection of Financial Fraud Based on Deep Learning with Combination of E-Commerce Big Data," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/6685888.

[3] A. M. Babu and A. Pratap, "Credit Card Fraud Detection Using Deep Learning," *2020 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2020*, pp. 32–36, 2020, doi: 10.1109/RAICS51191.2020.9332497.

[4] J. A. Choi and K. Lim, "Identifying machine learning techniques for classification of target advertising," *ICT Express*, vol. 6, no. 3, pp. 175–180, 2020, doi: 10.1016/j.icte.2020.04.012.

[5] R. A. L. Torres and M. Ladeira, "A proposal for online analysis and identification of fraudulent financial transactions," *Proc. - 19th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2020*, pp. 240–245, 2020, doi: 10.1109/ICMLA51294.2020.00047.

[6] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00684-w.

[7] P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib, and S. Alam, "Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques," *Procedia Comput. Sci.*, vol. 218, pp. 2575–2584, 2023, doi: 10.1016/j.procs.2023.01.231.

[8] V. I. Tomescu, G. Czibula, and stefan Niticâ, "A study on using deep autoencoders for imbalanced binary classification," *Procedia Comput. Sci.*, vol. 192, pp. 119–128, 2021, doi: 10.1016/j.procs.2021.08.013.

[9] Y. E. Kurniawati and Y. D. Prabowo, "Model optimisation of class imbalanced learning using ensemble classifier on over-sampling data," *IAES Int. J. Artif. Intell.*, vol. 11, no. 1, pp. 276–283, 2022, doi: 10.11591/ijai.v11.i1.pp276-283.

[10] I. D. Mienye and Y. Sun, "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection," *IEEE Access*, vol. 11, no. February, pp. 30628–30638, 2023, doi: 10.1109/ACCESS.2023.3262020.

[11] T. H. Lin and J. R. Jiang, "Credit card fraud detection with autoencoder and probabilistic random forest," *Mathematics*, vol. 9, no. 21, pp. 4–15, 2021, doi:

10.3390/math9212683.

[12] P. Mrozek, J. Panneerselvam, and O. Bagdasar, "Efficient resampling for fraud detection during anonymised credit card transactions with unbalanced datasets," *Proc. - 2020 IEEE/ACM 13th Int. Conf. Util. Cloud Comput. UCC 2020*, pp. 426–433, 2020, doi: 10.1109/UCC48980.2020.00067.

[13] N. S. Ramadhanti, W. A. Kusuma, and A. Annisa, "Optimasi Data Tidak Seimbang pada Interaksi Drug Target dengan Sampling dan Ensemble Support Vector Machine," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1221, 2020, doi: 10.25126/jtiik.2020762857.

[14] S. Mutrofin, A. Mu'alif, R. V. H. Ginardi, and C. Fatichah, "Solution of class imbalance of k-nearest neighbor for data of new student admission selection," *Int. J. Artif. Intell. Res.*, vol. 3, no. 2, 2019, doi: 10.29099/ijair.v3i2.92.

[15] E. Prasetyo and B. Prasetiyo, "Peningkatan Akurasi Klasifikasi Algoritma C 4.5 Menggunakan Teknik Bagging pada Diagnosis Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 1035–1040, 2020, doi: 10.25126/jtiik.2020752379.

[16] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM," *Techno.Com*, vol. 20, no. 3, pp. 352–361, 2021, doi: 10.33633/tc.v20i3.4762.

[17] W. Silalahi and A. Hartanto, "Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 7, no. 2, p. 245, 2023, doi: 10.30595/jrst.v7i2.18133.

[18] C. Liu, E. Fakharizadi, T. Xu, and P. S. Yu, "Recent advances in domain-driven data mining," *Int. J. Data Sci. Anal.*, vol. 15, no. 1, pp. 1–7, 2023, doi: 10.1007/s41060-022-00378-1.

[19] S. S. Pangastuti, K. Fithriasari, N. Iriawan, and W. Suryaningtyas, "Data Mining Approach for Educational Decision Support," *EKSAKTA J. Sci. Data Anal.*, vol. 2, no. 1, pp. 33–44, 2021, doi: 10.20885/eksakta.vol2.iss1.art5.

[20] Y. Chen, "Financial Statement Fraud Detection based on Integrated Feature Selection and Imbalance Learning," vol. 8, no. 3, pp. 1–3, 2023.

[21] P. Craja, A. Kim, and S. Lessmann, "Deep learning for detecting financial statement fraud," *Decis. Support Syst.*, vol. 139, p. 113421, 2020, doi: 10.1016/j.dss.2020.113421.

[22] H. Peng and J. Wang, "Unbalanced Data Processing and Machine Learning in Credit Card Fraud Detection," 2022.

[23] Ri. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor," *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018, [Online]. Available: https://ejournal-medan.uph.edu/index.php/isd/article/view/177/63.

[24] K. Bin Saboor, Q. Ul, A. Saboor, L. Han, and A. S. Zahid, "Predicting the Stock Market using Machine Learning: Long short-term Memory," *Electron. Res. J. Eng. Comput. Appl. Sci. www.erjsciences.info*, vol. 2, no. January 2021, p. 202, 2020, [Online]. Available: https://ssrn.com/abstract=3810128.

[25] I. A. Mondal, M. E. Haque, A. M. Hassan, and S. Shatabda, "Handling Imbalanced Data for Credit Card Fraud Detection," *24th Int. Conf. Comput. Inf. Technol. ICCIT 2021*, no. February 2022, 2021, doi:

10.1109/ICCIT54785.2021.9689866.

[26] M. R. Givari, M. R. Sulaeman, and Y. Umaidah, "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," *Nuansa Inform.*, vol. 16, no. 1, pp. 141–149, 2022, doi: 10.25134/nuansa.v16i1.5406.