

# Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval

## *Implementation of Random Forest Algorithm in The Classification of Credit Approval Dataset*

Nur Widjiyati

Program Sistem Informasi, Universitas Amikom Yogyakarta  
e-mail : nurwiwid@amikom.ac.id

Received:	Revised:	Accepted:	Available online:
19.03.2021	30.03.2021	04.01.2021	07.04.2021

**Abstrak:** *Knowledge Discovery in Database (KDD)* atau yang lebih dikenal dengan data mining merupakan suatu penyelesaian masalah dengan melakukan analisis terhadap data yang disajikan dalam database. *Data mining* dapat dikolaborasikan dengan ilmu lainnya. Salah satunya pada bidang ekonomi. Pemanfaatan data mining dalam bidang ekonomi seringkali digunakan untuk proses *credit approval* atau pemberian kredit kepada nasabah. Pada penelitian yang dilakukan sebelumnya mayoritas menggunakan single classifier, hal ini akan menimbulkan sebuah permasalahan baru ketika dalam dataset *credit approval* terdapat kondisi high dimensional dan ketidakseimbangan kelas. Keberadaan kondisi tersebut dapat menyebabkan kinerja single classifier menjadi tidak maksimal. Oleh karena itu pada penelitian ini akan menggunakan metode *ensemble* yaitu algoritme Random Forest. Berdasarkan pengujian yang dilakukan menunjukkan algoritme Random Forest mampu menghasilkan akurasi sebesar 87.3 %, TPR sebesar 86 % dan TNR sebesar 89 %. Dengan demikian algoritme yang diusulkan yaitu Random Forest mampu menghasilkan nilai akurasi, TPR, dan TNR yang lebih baik dibandingkan single classifier lainnya. Dengan adanya peningkatan nilai tersebut diharapkan penelitian ini mampu menjadi referensi untuk pengembangan berbagai sistem yang mendukung dan memaksimalkan tingkat keberhasilan proses pemberian kredit kepada nasabah menggunakan *data mining*.

**Kata kunci:** Data Mining, Klasifikasi, Random Forest, Credit Approval

**Abstract:** *Knowledge Discovery in Database (KDD)* or better known as data mining is a problem solving by analyzing the data presented in the database. *Data mining* can collaborate with other sciences. One of them is in the economic sector. The utilization of data mining in the economic sector is often used for the credit approval process or providing credit to customers. In previous research, the majority used a single classifier, this would create a new problem when the credit approval dataset contained high dimensional conditions and class imbalance. The existence of this condition can cause the performance of the single classifier to be not optimal. Therefore, this study will use the ensemble method, namely the Random Forest algorithm. Based on the tests conducted, it shows that the Random Forest algorithm is able to produce an accuracy of 87.3%, a TPR of 86% and a TNR of 89%. Thus the proposed algorithm, namely Random Forest, is able to produce better accuracy, TPR, and TNR values than other single classifiers. With this increase in value, it is hoped that this research will be able to become a reference for the development of various systems that support and maximize the success rate of the process of providing credit to customers using data mining.

**Keywords:** Data Mining, Classification, Random Forest, Credit Approval

## 1. PENDAHULUAN

*Knowledge Discovery in Database (KDD)* atau yang lebih dikenal dengan data mining merupakan suatu penyelesaian masalah dengan melakukan analisis terhadap data yang disajikan dalam database. Selain itu data mining juga digunakan untuk mengetahui pola data, dimana setiap pola memiliki karakteristik masing-masing yang dapat memberikan informasi penting dari data tersebut [1]. Data mining dapat diartikan sebagai berbagai macam cabang ilmu pengetahuan yang menjadi satu, terdiri atas sistem basis data, statistika, machine learning, visualization, dan informasi pengetahuan. Data mining telah berhasil diterapkan diberbagai bidang ilmu seperti ekonomi, bioinformatika, genetika, kedokteran, pendidikan dan lain sebagainya [2]. Salah satu contoh penerapan data mining pada bidang ekonomi ialah untuk klasifikasi nasabah bank dalam menentukan keputusan pemberian kredit. Hal tersebut menunjukkan bahwa data mining dapat memberikan solusi alternatif bagi para pengambil keputusan dalam memecahkan permasalahan pada bidang ekonomi khususnya pemberian kredit pada nasabah [3]. Salah satu Teknik data mining yang umum digunakan pada bidang ekonomi ialah klasifikasi [4]. Klasifikasi merupakan proses untuk menemukan sebuah model atau pola yang dapat menggambarkan serta membedakan kelas pada suatu dataset. Tujuannya agar model tersebut dapat digunakan untuk memprediksi obyek dengan label kelas yang tidak

diketahui. Model tersebut didasarkan pada analisis data latih. Model dari hasil klasifikasi dapat dimanfaatkan untuk mengklasifikasikan serta memprediksi tren data masa depan [5].

Ada beberapa algoritme klasifikasi yang sering digunakan pada bidang pemberian kredit kepada nasabah yang telah diujikan pada penelitian sebelumnya. Algoritme tersebut antara lain Naïve Bayes [6], C4.5 [7], Neural Network (NN) [8], Support Vector Machine (SVM) [9], dan k-Nearest Neighbor (k-NN) [10]. Namun pada penerapan data mining sering kali dijumpai dataset yang memiliki dimensi yang tinggi. Ciri dataset dengan dimensi yang tinggi ialah memiliki jumlah attribute atau feature yang banyak [11]. Dataset dengan dimensi tinggi tentunya akan memiliki pengaruh terhadap proses penerapan teknik data mining itu sendiri baik klasifikasi, klasterisasi maupun prediksi. Beberapa permasalahan yang sering disebabkan oleh dataset berdimensi tinggi antara lain kinerja algoritme klasifikasi baik dari sisi waktu komputasinya maupun dari sisi akurasi [12]. Selain itu penyebabnya ialah beberapa attribute feature yang tidak memiliki relevansi dengan attribute class. Sehingga secara tidak langsung berpengaruh terhadap algoritme klasifikasi yang digunakan. Karena pada umumnya algoritme klasifikasi bekerja dengan mengenali pola pada attribute feature untuk memprediksi maupun mengklasifikasi attribute classnya [13].

Selain itu dataset credit approval sering dijumpai adanya ketidakseimbangan kelas pada dataset tersebut. Dataset dikatakan kelasnya tidak seimbang apabila jumlah kelas positif dan negatif berbeda [5]. Keberadaan distribusi kelas yang tidak seimbang dapat mempengaruhi performa dari suatu algoritme klasifikasi, karena suatu algoritme klasifikasi bekerja dengan mengasumsikan distribusi kelas pada dataset relatif seimbang dan biaya kesalahan klasifikasi yang sama [16]. Hal tersebut tentunya dapat menimbulkan resiko terjadinya kesalahan klasifikasi (missclassification) terhadap dataset, sehingga berakibat pada kinerja suatu algoritme klasifikasi menjadi tidak maksimal [17]. Oleh karena itu diperlukan suatu teknik atau metode yang dapat menyelesaikan permasalahan terkait distribusi kelas yang tidak seimbang pada dataset. Selain itu penanganan distribusi kelas yang tidak seimbang pada suatu data merupakan salah satu tahapan yang penting untuk meminimalisir resiko yang akan timbul dalam proses data mining. Berdasarkan penelitian yang telah dilakukan terkait penanganan terhadap distribusi kelas yang tidak seimbang pada dataset, terdapat dua pendekatan yang dapat diterapkan yaitu pendekatan pada level data dan pendekatan pada level algoritmik. Pendekatan level data biasanya dilakukan pada tahap pra pemrosesan data dengan mengubah atau memperbaiki kecondongan distribusi kelas yang terdapat pada dataset. Metode yang sering dipakai dalam pendekatan pada level data ialah menerapkan teknik resampling maupun sintesis data. Pada pendekatan level algoritmik cara kerjanya ialah menyesuaikan operasi algoritme yang ada untuk membuat suatu classifier lebih kondusif terhadap klasifikasi kelas minoritas atau dengan kata lain dilakukan modifikasi maupun penggabungan (ensemble) dari beberapa algoritme. Beberapa metode yang sering digunakan dalam pendekatan ini ialah cost sensitive learning, bagging, boosting dan stacking [18]. Dua masalah tersebut menyebabkan algoritme single classifier seperti Naïve Bayes, SVM, NN, C4.5 dan k-NN menjadi kurang optimal pada proses klasifikasi.

Oleh karena itu pada penelitian ini diusulkan algoritme ensemble untuk mengatasi permasalahan tersebut. Pada penelitian ini algoritme ensemble yang diusulkan ialah Random Forest. Tujuannya untuk mengatasi permasalahan high dimensional dataset dan ketidakseimbangan kelas pada dataset. Kontribusi pada penelitian ini antara lain pertama penelitian ini dapat menjadi solusi untuk mengetahui apakah algoritme Random Forest mampu menangani permasalahan high dimensional dan ketidekseimbangan kelas. Kedua, penelitian ini dapat menjadi referensi bagi penelitian selanjutnya penanganan permasalahan high dimensional dan ketidekseimbangan kelas.

## 2. TINJAUAN PUSTAKA

Berbagai penelitian telah dilakukan terkait penerapan data mining untuk klasifikasi pemberian kredit pada nasabah. Kurniawan dkk [14] melakukan penelitian terkait klasifikasi pemberian kredit. Pada penelitian tersebut algoritme yang digunakan k-Nearest Neighbor (kNN). Pada penelitian lainnya Sari Dewi [10] melakukan perbandingan antara algoritme Naïve Bayes dan kNN untuk proses klasifikasi pemberian kredit. Hasil dari kedua penelitian tersebut cukup baik untuk nilai akurasi. Pada penelitian yang dilakukan Menarianti [9] melakukan perbandingan algoritme klasifikasi untuk melakukan pemberian kredit kepada nasabah. Algoritme yang dibandingkan antara

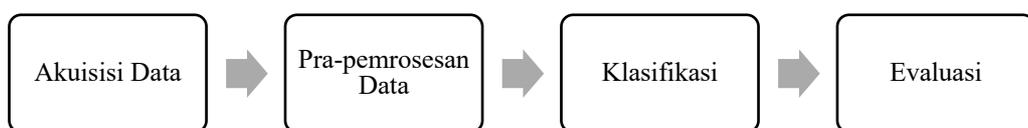
lain K-Nearest Neighbour, Naive Bayes, Decision Tree, Neural Network dan Support Vector Machine. Hasilnya algoritme Neural Network dan Decision Tree menghasilkan nilai akurasi yang paling baik. Disisi lain pada penelitian yang dilakukan oleh Nurajijah dkk [15] menyebutkan bahwa algoritme Support Vector Machine sangat baik untuk klasifikasi pemberian kredit. Pada penelitian tersebut Support Vector Machine dibandingkan dengan algoritme Naive Bayes dan Decision Tree C4.5. Namun pada penerapan data mining sering kali dijumpai dataset yang memiliki dimensi yang tinggi. Ciri dataset dengan dimensi yang tinggi ialah memiliki jumlah attribute atau feature yang banyak [11]. Dataset dengan dimensi tinggi tentunya akan memiliki pengaruh terhadap proses penerapan teknik data mining itu sendiri baik klasifikasi, klusterisasi maupun prediksi. Beberapa permasalahan yang sering disebabkan oleh dataset berdimensi tinggi antara lain kinerja algoritme klasifikasi baik dari sisi waktu komputasinya maupun dari sisi akurasinya [12]. Selain itu penyebabnya ialah beberapa attribute feature yang tidak memiliki relevansi dengan attribute class. Sehingga secara tidak langsung berpengaruh terhadap algoritme klasifikasi yang digunakan. Karena pada umumnya algoritme klasifikasi bekerja dengan mengenali pola pada attribute feature untuk memprediksi maupun mengklasifikasi attribute classnya [13].

Selain itu dataset credit approval sering dijumpai adanya ketidakseimbangan kelas pada dataset tersebut. Dataset dikatakan kelasnya tidak seimbang apabila jumlah kelas positif dan negatif berbeda [5]. Keberadaan distribusi kelas yang tidak seimbang dapat mempengaruhi performa dari suatu algoritme klasifikasi, karena suatu algoritme klasifikasi bekerja dengan mengasumsikan distribusi kelas pada dataset relatif seimbang dan biaya kesalahan klasifikasi yang sama [16]. Hal tersebut tentunya dapat menimbulkan resiko terjadinya kesalahan klasifikasi (missclassification) terhadap dataset, sehingga berakibat pada kinerja suatu algoritme klasifikasi menjadi tidak maksimal [17]. Oleh karena itu diperlukan suatu teknik atau metode yang dapat menyelesaikan permasalahan terkait distribusi kelas yang tidak seimbang pada dataset. Selain itu penanganan distribusi kelas yang tidak seimbang pada suatu data merupakan salah satu tahapan yang penting untuk meminimalisir resiko yang akan timbul dalam proses data mining. Berdasarkan penelitian yang telah dilakukan terkait penanganan terhadap distribusi kelas yang tidak seimbang pada dataset, terdapat dua pendekatan yang dapat diterapkan yaitu pendekatan pada level data dan pendekatan pada level algoritmik. Pendekatan level data biasanya dilakukan pada tahap pra pemrosesan data dengan mengubah atau memperbaiki kecondongan distribusi kelas yang terdapat pada dataset. Metode yang sering dipakai dalam pendekatan pada level data ialah menerapkan teknik resampling maupun sintesis data. Pada pendekatan level algoritmik cara kerjanya ialah menyesuaikan operasi algoritme yang ada untuk membuat suatu classifier lebih kondusif terhadap klasifikasi kelas minoritas atau dengan kata lain dilakukan modifikasi maupun penggabungan (ensemble) dari beberapa algoritme. Beberapa metode yang sering digunakan dalam pendekatan ini ialah cost sensitive learning, bagging, boosting dan stacking [18].

Dua masalah tersebut menyebabkan algoritme single classifier seperti Naive Bayes, SVM, NN, C4.5 dan k-NN menjadi kurang optimal pada proses klasifikasi. Oleh karena itu pada penelitian ini diusulkan algoritme ensemble untuk mengatasi permasalahan tersebut. Pada penelitian ini algoritme ensemble yang diusulkan ialah Random Forest. Tujuannya untuk mengatasi permasalahan high dimensional dataset dan ketidakseimbangan kelas pada dataset.

**3. METODE**

Penelitian dilakukan meliputi 4 tahapan yaitu Akuisisi Data, Pra Pemrosesan Data, Klasifikasi dan Evaluasi. Berikut Gambar 1 merupakan tahapan pada penelitian ini.



Gambar 1. Tahapan Penelitian

### 3.1. Akuisisi Data

Pada penelitian ini, dataset yang digunakan terkait Credit Card Approval yang memiliki 690 instances, 15 attribute fitur dan 1 attribute class. Selain itu terdapat 67 record missing values pada 7 attribute fitur. Sedangkan distribusi kelasnya untuk kelas positif 307 instances dan kelas negatif 383 instances, dengan kondisi demikian dataset ini dapat dikategorikan tidak seimbang kelasnya. Dataset tersebut merupakan public dataset yang diperoleh dari UCI Machine Learning Repository [19]. Untuk alasan privasi, semua nama atribut dan nilai telah ditransformasikan menjadi simbol tertentu. Tabel 1 berikut menunjukkan informasi dari dataset yang digunakan.

**Tabel 1.** Informasi Attribute Pada Dataset [19]

Attribute	Type	Missing Values
A1	Kategorik	12 Records
A2	Numerical	12 Records
A3	Numerical	-
A4	Kategorik	6 Records
A5	Kategorik	6 Records
A6	Kategorik	9 Records
A7	Kategorik	9 Records
A8	Numerical	-
A9	Kategorik	-
A10	Kategorik	-
A11	Numerical	-
A12	Kategorik	-
A13	Kategorik	-
A14	Numerical	13 Records
A15	Numerical	-
Class	Kategorik (Yes, No)	-

### 3.2. Pra Pemrosesan Data

Pada tahap preprocessing data dilakukan pembersihan data. Proses pembersihan meliputi pengisian data yang kosong, menghilangkan duplikasi data, memeriksa inkonsistensi data, dan memperbaiki kesalahan pada data. Biasanya data yang kosong disebabkan oleh adanya data baru yang belum ada informasinya [20].

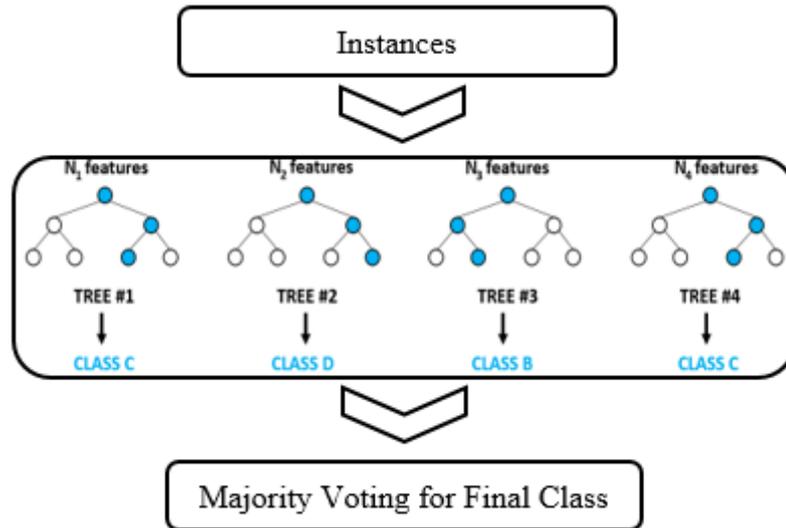
Pada tahap preprocessing kami melakukan pengisian missing values pada dataset. Keberadaan missing values pada dataset tentunya juga akan berpengaruh pada hasil klasifikasi dataset itu sendiri. Oleh karena itu pada tahap ini kami melakukan pengisian missing values menggunakan suatu nilai konstan. Pada data numerik missing values diganti dengan nilai rata-rata, sedangkan untuk data kategorik maka missing values diganti dengan nilai modus.

### 3.3. Klasifikasi

Pada tahap ini algoritme klasifikasi yang digunakan ialah Random Forest. Kemudian akan dibandingkan hasilnya dengan algoritme bertipe single classifier berikut ini Naïve Bayes, SVM, NN, C4.5 dan k-NN. Random Forest (RF) merupakan salah satu jenis metode bootstrap aggregating yang memiliki cara kerja dengan membangkitkan sejumlah tree dari data sample dimana pembuatan satu tree pada saat proses training tidak bergantung terhadap tree sebelumnya kemudian dalam pengambilan keputusannya diambil berdasarkan voting terbanyak [21].

Dua konsep yang menjadi dasar dari random forest adalah membangun ensemble dari tree via bagging dengan replacement dan penyeleksian fitur secara acak untuk tiap tree yang dibangun. Berikut merupakan penjelasan dari konsep dasar pada algoritme Random Forest [22]. Pertama, setiap sample yang diambil dari dataset untuk proses training setiap tree dapat digunakan lagi untuk proses training tree yang lain. Kedua, fitur yang digunakan saat proses training pada setiap tree merupakan subset dari fitur yang dimiliki oleh dataset. Random forest mempunyai dua parameter utama, yaitu parameter  $m$  yang merupakan presentasi dari jumlah tree yang akan dipakai dan parameter  $k$  yang merupakan representasi dari banyaknya fitur maksimal yang dipertimbangkan ketika proses percabangan pada tree. Semakin banyak nilai parameter  $m$  maka semakin bagus hasil klasifikasi,

sedangkan untuk nilai k direkomendasikan sebesar akar kuadrat atau logaritma dari jumlah total fitur pada dataset. Berikut gambar 2 merupakan ilustrasi dari algoritme Random Forest.



Gambar 2 Ilustrasi Random Forest

### 3.3. Evaluasi

Tahap terakhir pada penelitian ini ialah melakukan evaluasi kinerja dari algoritme stacking kemudian hasilnya akan dibandingkan dengan single classifier lainnya. Hasil dari sebuah proses klasifikasi data dapat dikategorikan menjadi empat jenis yaitu False positif (FP) adalah jumlah record negatif yang diklasifikasikan sebagai positif. False negatif (FN) adalah jumlah record positif yang diklasifikasikan sebagai negatif, True negatif (TN) adalah jumlah record negatif yang diklasifikasikan sebagai negatif. True Positif (TP) adalah jumlah record positif yang diklasifikasikan sebagai positif. Hasil tersebut didapatkan dari confusion matrix berikut tabel 2 merupakan tabel confusion matrix [23].

**Tabel 2. Confusion Matrix**

Aktual	Prediksi	
	True	False
True	TP	FN
False	FP	TN

Indikator evaluasi yang digunakan antara lain akurasi, true positive rate (TPR) dan true negative rate (TNR)[24]. Berikut persamaan untuk menghitung indicator tersebut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$TPR = \frac{TP}{TP+FN} \quad (2)$$

$$TNR = \frac{TN}{TN+FP} \quad (3)$$

## 4. HASIL DAN PEMBAHASAN

Pada penelitian ini masing-masing dataset dibagi ke dalam dua bagian yaitu sebagai data training dan data testing. Kami menggunakan skenario pembagian data sebagai berikut, 80% untuk melatih model dan 20% digunakan untuk validasi model atau testing model. Pada penelitian ini algoritme yang diusulkan yaitu Random Forest akan dibandingkan dengan Naïve Bayes, SVM, NN, C4.5 dan k-NN. Berikut Tabel 3 merupakan evaluasi proses klasifikasi antar algoritme klasifikasi yang digunakan.

**Tabel 3.** Hasil Evaluasi Algoritme Klasifikasi

Algoritme	Akurasi	TPR	TNR
K-NN	69.9%	55.3%	85%
C4.5	82.6%	77.8%	86.4%
SVM	84.9%	86%	83.3%
NN	85.2%	84%	86.1%
Naïve Bayes	85.9%	81.1%	88%
Random Forest	87.3%	86%	89%

Tabel 3 menunjukkan bahwa algoritme K-NN menghasilkan akurasi sebesar 69.9 %, TPR sebesar 55.3 % dan TNR sebesar 85 %. Algoritme C4.5 menghasilkan akurasi sebesar 82.6 %, TPR sebesar 77.8 % dan TNR sebesar 86.4 %. Algoritme SVM menghasilkan akurasi sebesar 84.9 %, TPR sebesar 86 % dan TNR sebesar 83.3 %. Algoritme NN menghasilkan akurasi sebesar 85.2 %, TPR sebesar 84 % dan TNR sebesar 86.1 %. Algoritme Naïve Bayes menghasilkan akurasi sebesar 85.9 %, TPR sebesar 81.1 % dan TNR sebesar 88 %. Sedangkan algoritme Random Forest mampu menghasilkan akurasi sebesar 87.3 %, TPR sebesar 86 % dan TNR sebesar 89 %. Dengan demikian algoritme yang diusulkan yaitu Random Forest mampu menghasilkan nilai akurasi, TPR, dan TNR yang lebih baik dibandingkan single classifier lainnya. Hal ini dikarenakan oleh keberadaan high dimensional dan ketidakseimbangan kelas pada dataset. Oleh karena itu metode atau algoritme Random Forest yang diusulkan dapat menjadi solusi untuk dataset yang memiliki kondisi high dimensional data dan ketiaksimbangan kelas pada dataset.

## 5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menunjukkan bahwa data mining dapat digunakan sebagai solusi untuk klasifikasi pemberian kredit kepada nasabah. Akan tetapi keberadaan high dimensional data dan ketiaksimbangan kelas pada dataset pada dataset credit approval menyebabkan kinerja single classifier menjadi tidak maksimal. Oleh karena algoritme yang diusulkan yaitu Random Forest dapat menjadi solusi untuk permasalahan high dimensional dan ketidakseimbangan kelas. Berdasarkan pengujian yang dilakukan menunjukkan bahwa algoritme Random Forest mampu menghasilkan kinerja dari sisi akurasi TPR, dan TNR yang lebih baik dibandingkan single classifier lainnya. Dengan adanya peningkatan nilai tersebut diharapkan penelitian ini mampu menjadi referensi untuk pengembangan berbagai sistem yang mendukung dan memaksimalkan tingkat keberhasilan proses pemberian kredit kepada nasabah menggunakan data mining

## DAFTAR PUSTAKA

- [1] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- [2] Fadhilah Ahmad, N. H. Ismail, and Azwa Abdul Aziz, "The prediction of students' academic performance using classification data mining techniques," *Appl. Math. Sci.*, vol. 9, no. 129, pp. 6415–6426, 2015, doi: 10.12988/ams.2015.53289.
- [3] Y. Pristyanto, S. Adi, and A. Sunyoto, "The effect of feature selection on classification algorithms in credit approval," *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 451–456, 2019, doi: 10.1109/ICOIACT46704.2019.8938523.
- [4] H. D. Honesqi, "Klasifikasi Data Mining Untuk Menentukan Tingkat Persetujuan Kartu Kredit," *J. Teknoif*, vol. 5, no. 2, pp. 57–62, 2017, doi: 10.21063/jtif.2017.v5.2.57-62.
- [5] Y. Pristyanto, "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset," *J. TEKNOINFO*, vol. 13, no. 1, pp. 11–16, 2019, doi: 10.33365/jti.
- [6] D. A. Kurniawan and D. Kriestanto, "Penerapan Naive Bayes Untuk Prediksi Kelayakan Kredit," *JIKO (Jurnal Inform. dan Komputer)*, vol. 1, no. 1, pp. 19–23, 2016, doi: 10.26798/jiko.2016.v1i1.10.
- [7] N. Iriadi and N. Nuraeni, "Kajian Penerapan Metode Klasifikasi Data Kelayakan Kredit Pada Bank," *J. Tek. Komput. AMIK BSI*, vol. II, no. 1, pp. 132–137, 2016.
- [8] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163–170, 2019,

- doi: 10.33480/pilar.v15i2.658.
- [9] I. Menarianti, "Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, pp. 1–10, 2015, [Online]. Available: <http://ejournal.upgrismg.ac.id/index.php/JITEK/article/view/836>.
- [10] S. Dewi, "Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 59–65, 2019, doi: 10.31294/jki.v7i1.5744.
- [11] K. Gao, T. Khoshgoftaar, and A. Napolitano, "Exploring software quality classification with a wrapper-based feature ranking technique," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2009, pp. 67–74, doi: 10.1109/ICTAI.2009.24.
- [12] A. Winursito, R. Hidayat, A. Bejo, and M. N. Y. Utomo, "Feature Data Reduction of MFCC Using PCA and SVD in Speech Recognition System," in *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, 2018, pp. 1–6.
- [13] W. Punlumjeak and N. Rachburee, "A Comparative Study of Feature Selection Techniques for Classify Student Performance," in *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2015, pp. 425–429, doi: 10.1109/ICMLA.2010.27.
- [14] Y. I. Kurniawan and T. I. Barokah, "Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan K-Nearest Neighbor," *J. Ilm. Matrik*, vol. 22, no. 1, pp. 73–82, 2020, doi: 10.33557/jurnalmatrik.v22i1.843.
- [15] N. Nurajijah and D. Riana, "Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah," *J. Teknol. dan Sist. Komput.*, vol. 7, no. 2, pp. 77–82, 2019, doi: 10.14710/jtsiskom.7.2.2019.77-82.
- [16] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognit.*, vol. 40, no. 12, pp. 3358–3378, 2007, doi: 10.1016/j.patcog.2007.04.009.
- [17] R. Longadge and S. Dongre, "Class Imbalance Problem in Data Mining Review," *Int. J. Comput. Sci. Netw.*, vol. 2, no. 1, pp. 83–87, 2013, doi: 10.1109/SIU.2013.6531574.
- [18] D. Zhang, W. Liu, X. Gong, and H. Jin, "A Novel Improved SMOTE Resampling Algorithm Based on Fractal," in *Advanced Data Mining and Applications - 7th International Conference*, 2011, pp. 83–96.
- [19] D. Dua and E. Karra Taniskidou, "UCI Machine Learning Repository." 2017, [Online]. Available: <http://archive.ics.uci.edu/ml>.
- [20] Y. Pristyanto and I. Pratama, "Missing values estimation on multivariate dataset: Comparison of three type methods approach," *2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019*, pp. 342–347, 2019, doi: 10.1109/ICOIACT46704.2019.8938472.
- [21] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [22] T. Lu, Y. Huang, W. Zhao, and J. Zhang, "The Metering Automation System based Intrusion Detection Using Random Forest Classifier with SMOTE+ENN," in *Proceedings of IEEE 7th International Conference on Computer Science and Network Technology, ICCSNT 2019*, 2019, pp. 370–374, doi: 10.1109/ICCSNT47585.2019.8962430.
- [23] J. Han, M. Kamber, and J. Pei, *Data Mining: Concept and Techniques*, Third Edit. Massachusetts: Morgan Kaufman, 2011.
- [24] A. Nurmasani and Y. Pristyanto, "ALGORITME STACKING UNTUK KLASIFIKASI PENYAKIT JANTUNG PADA DATASET IMBALANCED CLASS Atik," *J. Pseudocode*, vol. VIII, no. Februari, pp. 21–26, 2021.