

Klasifikasi Kelayakan Pemberian Kredit Menggunakan Metode Decision Tree dengan Seleksi Fitur (Studi Kasus: PT. Adira Finance Cabang Kota Ternate)

Rosihan¹, Muhammad Fhadli², Alfanugrah A. Hi. Usman³

^{1,2,3}Program Studi Informatika, Universitas Khairun

e-mail: rosihan@unkhair.ac.id¹, mfhadli@unkhair.ac.id², nugrahalfa@gmail.com³

Abstrak

Pemberian kredit merupakan alternatif yang signifikan bagi masyarakat untuk memenuhi kebutuhan finansial mereka. Namun, evaluasi aplikasi kredit yang tepat dan pemantauan terhadap nasabah yang telah mengajukan kredit menjadi kritikal dalam menghindari risiko kredit yang sering muncul. Penelitian ini bertujuan untuk mengklasifikasikan status kredit nasabah PT. Adira Finance Cabang Kota Ternate menjadi dua kategori, yaitu "Lancar" dan "Macet," berdasarkan parameter kolektabilitas. Dalam penelitian ini, kami memanfaatkan teknik decision tree dengan melibatkan seleksi fitur sebagai pendekatan utama. Hasil eksperimen menunjukkan bahwa metode select k-best memberikan nilai F1 tertinggi sebesar 0.416, diikuti oleh L1 Based Feature Selection dengan nilai F1 sebesar 0.409, dan penghapusan fitur dengan varians rendah dengan nilai F1 sebesar 0.404. Seleksi fitur menggunakan feature important from coefficient memberikan nilai F1 terendah sebesar 0.310.

Kata kunci: *Klasifikasi, Kelayakan Pemberian Kredit, Decision Tree, Seleksi Fitur, PT Adira Finance Cabang Kota Ternate*

Abstract

Providing credit is a significant alternative for people to meet their financial needs. However, proper evaluation of credit applications and monitoring of customers who have applied for credit are critical in avoiding credit risks that often arise. This research aims to classify the credit status of PT customers. Adira Finance Ternate City Branch into two categories, namely "Current" and "Loss," based on collectability parameters. In this research, we utilize the decision tree technique involving feature selection as the main approach. Experimental results show that the k-best select method provides the highest F1 value of 0.416, followed by L1 Based Feature Selection with an F1 value of 0.409, and feature removal with low variance with an F1 value of 0.404. Feature selection using the feature importance from coefficient gives the lowest F1 value of 0.310.

Keywords : *Classification, Credit Eligibility, Decision Tree, Feature Selection, PT Adira Finance Ternate City Branch*

PENDAHULUAN

Dalam konteks dinamika ekonomi saat ini, kredit telah muncul sebagai pilihan alternatif yang signifikan bagi masyarakat untuk memenuhi kebutuhan finansial mereka. Oleh karena itu, perluasan dan perbaikan evaluasi aplikasi kredit (Hadianto et al., 2019), serta pemantauan terhadap nasabah yang telah mengajukan kredit dengan memanfaatkan data historis, telah menjadi kritikal dalam mewujudkan keputusan yang tepat (Moradi & Mokhatab Rafiei, 2019). Hal ini penting untuk menghindari risiko kredit yang sering muncul dalam situasi saat ini (Sugesti & Abdul Mukid, 2019). Dalam konteks ini, perlu juga diakui bahwa

perilaku kredit macet kadang-kadang timbul akibat perilaku nasabah yang sering kali terlambat membayar kredit (Chen et al., 2012).

Penelitian ini akan membahas pengelompokan status kredit nasabah PT. Adira Finance Kota Ternate menjadi dua kategori, yaitu "Lancar" dan "Macet," berdasarkan pada parameter kolektabilitas. Solusi dalam bentuk data mining dipertimbangkan sebagai pendekatan yang mampu mengakses informasi tersembunyi dalam himpunan data. Data Mining, sebagai proses yang mengungkap pola dan relasi dalam data dengan tujuan klasifikasi (Pritalia, 2022), memiliki peran sentral dalam analisis data. Klasifikasi, sebagai elemen dasar dalam analisis data, mengidentifikasi serangkaian data dengan struktur yang serupa, dan akan mengklasifikasikan data tersebut dalam kategori yang sesuai (Kurniawan, 2018).

Sebelumnya, penelitian serupa telah berhasil mengklasifikasikan nasabah kredit dengan memanfaatkan teknik data mining (Lubis, 2022). Terdapat beragam jenis algoritma klasifikasi data mining, seperti Random Forest, C.45, Naive Bayes, Radial Basis Function, K-nearest neighbors, dan Neural Network, yang telah digunakan dalam usaha pengklasifikasian data nasabah (Miftahusalam et al., 2023).

Dalam skenario penelitian ini, pendekatan yang akan diterapkan adalah menggunakan teknik decision tree untuk mengklasifikasikan status kredit nasabah PT Adira Finance Kota Ternate. Pendekatan ini akan dilakukan dengan melibatkan seleksi fitur. Metode yang berhasil menghasilkan tingkat akurasi tertinggi akan direkomendasikan sebagai bagian integral dalam proses pengklasifikasian nasabah kredit di PT Adira Finance Kota. Melalui perpaduan antara teknik yang efektif dan seleksi fitur yang cermat, penelitian ini diharapkan akan memberikan kontribusi berharga dalam pengembangan pemahaman tentang pengambilan keputusan terkait pemberian kredit. Selain itu, penelitian ini memiliki potensi untuk memperkaya strategi lembaga keuangan dalam mengelola risiko dan menyediakan solusi yang lebih transparan bagi para pelanggan.

Kredit dan Pengambilan Keputusan

Metode Random OverUnder Sampling Random Forest dapat diterapkan untuk meningkatkan akurasi dalam klasifikasi penilaian kredit. Hasilnya, akurasi Random Forest mencapai 0,76 atau 76%. Namun, dengan penerapan metode Random Over-under Sampling Random Forest, akurasi bisa meningkat menjadi 0,901 atau 90,1%, yang merupakan peningkatan sebesar 14,1%. Hasil penelitian menunjukkan bahwa penerapan resampling metode Random Over-Under Sampling pada algoritma Random Forest dapat meningkatkan efektivitas kinerja akurasi dalam klasifikasi penilaian kredit yang tidak seimbang (Syukron & Subekti, 2018).

Metode klasifikasi juga dapat dibandingkan, seperti Mutual K-Nearest Neighbor (MKNN) dan K-Nearest Neighbor (KNN). Kinerja klasifikasi divalidasi melalui 10-fold cross validation dan diukur menggunakan nilai Gemometric-Mean dari sensitivitas dan spesifisitas. Hasil analisis menunjukkan bahwa nilai K optimal untuk MKNN adalah 9 dengan G-Mean sebesar 0,702, dan nilai K optimal untuk KNN adalah 3 dengan G-Mean sebesar 0,718. Berdasarkan hasil ini, alternatif terbaik untuk mengklasifikasikan kelayakan kredit dalam penelitian ini adalah menggunakan algoritma K-Nearest Neighbor (KNN) dengan nilai K=3 (Sugesti & Abdul Mukid, 2019).

Decision Tree

Decision tree merupakan salah satu algoritma machine learning yang memiliki konsep pembuatan model seperti pohon. Algoritma ini membantu dalam membuat keputusan dan prediksi berdasarkan kondisi-kondisi tertentu yang didefinisikan dalam setiap node pohon. Decision tree digunakan untuk pemecahan masalah klasifikasi dan regresi (Mandala et al., 2012).

Pada setiap node dalam pohon, algoritma decision tree melakukan splittings atau pembagian berdasarkan atribut yang memiliki information gain tertinggi. Algoritma ini memilih atribut yang membantu membuat pembagian yang lebih baik dalam memprediksi target dari data. Setelah proses splitting selesai dilakukan, maka akan terbentuk sub-node dari setiap node yang berisi data yang memiliki kondisi yang sama. Proses splitting ini akan terus

dilakukan hingga tercapai kondisi tertentu, seperti kondisi dalam membatasi profunditas pohon atau jumlah data pada setiap node yang dianggap sudah cukup (S. Brid, 2018).

Implementasi decision tree banyak tersedia dalam berbagai library pemrograman, seperti scikit-learn (Pedregosa & Varoquaux, 2011) di Python. Dalam implementasi, pengguna dapat memilih beberapa parameter, seperti jenis criterion yang digunakan dalam memilih atribut splittings, profunditas pohon, jumlah minimum data pada setiap node, dan lain-lain.

Seleksi Fitur

Seleksi fitur adalah suatu proses pemilihan atribut dari data yang dipakai dalam suatu proses pembelajaran mesin. Tujuan dari proses ini adalah memilih atribut-atribut yang memiliki kontribusi yang signifikan dalam menentukan hasil akhir dan membuang atribut-atribut yang tidak bermanfaat (Piramuthu, 2004). Proses seleksi fitur dapat meningkatkan akurasi dan efisiensi model, dan juga dapat membantu untuk mengurangi overfitting (Mohammad Abualigah et al., 2020).

Ada beberapa metode yang dapat digunakan untuk melakukan seleksi fitur, seperti pengurangan dimensi melalui teknik filter, pengurangan dimensi melalui teknik wrapper, dan pengurangan dimensi melalui teknik embedded (Zebari et al., 2020). Keuntungan dan kelemahan dari setiap metode akan bervariasi tergantung pada konteks penerapannya. Oleh karena itu, pemilihan metode seleksi fitur yang tepat sangat penting untuk mencapai hasil yang optimal (Nguyen et al., 2020).

METODE

Pengumpulan Data

Pada tahap ini, data nasabah yang telah meminjam motor di Kantor Adira Finance Kota Ternate dikumpulkan. Peneliti hanya mengambil 11 atribut dan 1 label variabel, yaitu pekerjaan, DP net, OTR, tenor, gender, status pernikahan, tanggungan, pendidikan, status tempat tinggal, lamanya tinggal, pendapatan, dan status nasabah (label). Data dikumpulkan melalui tanya jawab antara peneliti dan narasumber (karyawan Adira) untuk memperoleh informasi yang diperlukan untuk penelitian.

Pembersihan Data

Pada tahap ini, peneliti akan mengolah data sehingga sesuai dengan metode analisis yang akan digunakan. Proses ini meliputi transformasi data, yaitu mengubah variabel-variabel tertentu seperti gender, status pernikahan, pendidikan, status tempat, pekerjaan, dan status nasabah menjadi nilai numerik. Kemudian, normalisasi data akan dilakukan untuk menyesuaikan nilai data pada rentang yang sama (0.0 hingga 1.0). Proses ini akan dilakukan dengan teknik min-max scaling dan label encoding menggunakan library scikit learn di Python. Tahapan min-max scaling dapat dilakukan dengan menggunakan rumus yang ditunjukkan pada persamaan 2.22, di mana x merupakan data yang akan dinormalisasi dan x' adalah hasil data yang telah dinormalisasi, sedangkan $\min x$ dan $\max x$ masing-masing adalah nilai minimum dan maksimum dari seluruh data.

Proses Training

Decision Tree adalah algoritma pembelajaran mesin yang populer digunakan untuk memecahkan masalah klasifikasi dan regresi. Proses training decision tree melibatkan penentuan pohon keputusan dari data latih yang digunakan untuk membuat prediksi pada data uji. Dalam hal ini, Python memiliki library scikit-learn yang menyediakan implementasi algoritma Decision Tree.

Pertama, data latih harus disiapkan dengan menentukan fitur-fitur yang akan digunakan sebagai input dan label yang akan diprediksi. Kemudian, data latih akan dibagi menjadi dua bagian, yaitu data latih dan data validasi untuk melakukan evaluasi model.

Berikutnya, objek Decision Tree Classifier dari library scikit-learn dapat dibuat dan di-fit pada data latih. Dalam proses ini, beberapa parameter seperti jenis pembuatan pohon keputusan (misalnya Gini atau entropy sebagai kriteria pemisahan), profunditas maksimal pohon, dan minimum jumlah sampel pada node dapat ditentukan.

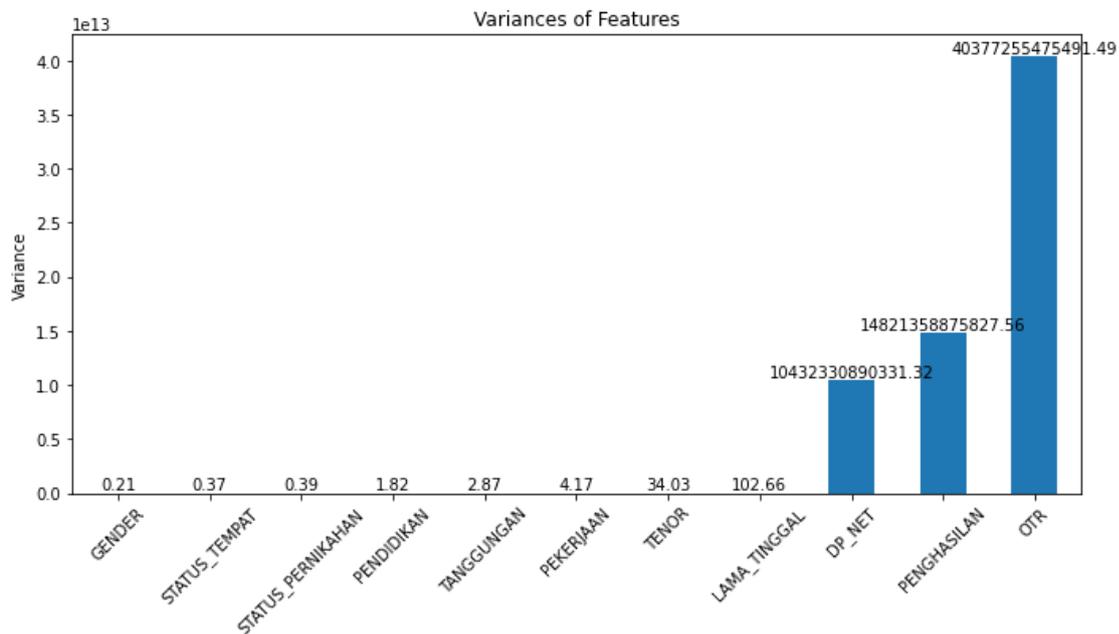
Setelah proses training selesai, model dapat digunakan untuk membuat prediksi pada data uji. Hasil prediksi dapat dicompare dengan label sebenarnya untuk menilai kinerja model dan melakukan perbaikan jika diperlukan.

Penggunaan Decision Tree Classifier dari library scikit-learn membuat proses training dan pembuatan model sangat mudah dan efisien. Dengan adanya fitur-fitur terintegrasi seperti evaluasi kinerja, visualisasi pohon keputusan, dan pemodelan hyperparameter, membuat Decision Tree menjadi algoritma yang sangat baik untuk memecahkan masalah klasifikasi dan regresi.

HASIL DAN PEMBAHASAN

Menghapus Varians Rendah

Seleksi fitur yang menjadi baseline pada penelitian ini adalah dengan cara menghapus nilai varian terendah. Varian dari setiap fitur pada data training ditunjukkan pada Gambar 1.



Gambar 1. Varians seluruh fitur

Berdasarkan Gambar 1, fitur yang memiliki nilai varians terkecil adalah gender dengan nilai varians 0.21. Tabel 1 menunjukkan nilai F1 tertinggi yang didapatkan dengan menggunakan teknik ini adalah 0.404 yang berasal dari penghapusan fitur gender. ketika ditambah dengan penghapusan fitur lain, maka nilai F1 akan semakin menurun.

Tabel 1. Performa dengan Penghapusan Varians Rendah

No	Fitur yang dihapus	Akurasi	Precision	Recall	F1
1	gender	81.1	0.385	0.439	0.404
2	gender, status_tempat	80.1	0.363	0.413	0.378
3	gender, status_tempat, status_pernikahan	79.1	0.334	0.400	0.357
4	gender, status_tempat, status_pernikahan, pendidikan	79.7	0.347	0.413	0.372
5	gender, status_tempat, status_pernikahan, pendidikan, tanggungan	76.9	0.296	0.406	0.342
6	gender, status_tempat,	77.6	0.283	0.333	0.301

7	status_pernikahan, pendidikan, tanggungan, pekerjaan gender, status_tempat, status_pernikahan, pendidikan, tanggungan, pekerjaan, tenor	77.3	0.277	0.346	0.304
8	gender, status_tempat, status_pernikahan, pendidikan, tanggungan, pekerjaan, tenor, lama_tinggal	76.4	0.274	0.353	0.307
9	gender, status_tempat, status_pernikahan, pendidikan, tanggungan, pekerjaan, tenor, lama_tinggal, dp_net	75.8	0.248	0.286	0.262
10	gender, status_tempat, status_pernikahan, pendidikan, tanggungan, pekerjaan, tenor, lama_tinggal, dp_net, penghasilan	80.4	0.239	0.146	0.172

Select K-Best

Dalam eksperimen yang menggunakan pendekatan Select K-Best, ditemukan bahwa performa maksimal yang diukur dengan nilai F1 adalah sebesar 0.416. Hasil ini dicapai dengan menggunakan semua fitur dalam proses training, kecuali fitur status pernikahan. Hasil ini menunjukkan bahwa Select K-Best menganggap bahwa 10 fitur lainnya dianggap sebagai 10 fitur terbaik yang paling efektif untuk digunakan dalam proses training. Dengan demikian, fitur-fitur tersebut memberikan kontribusi yang lebih signifikan dalam meningkatkan performa model daripada fitur yang berkaitan dengan status pernikahan.

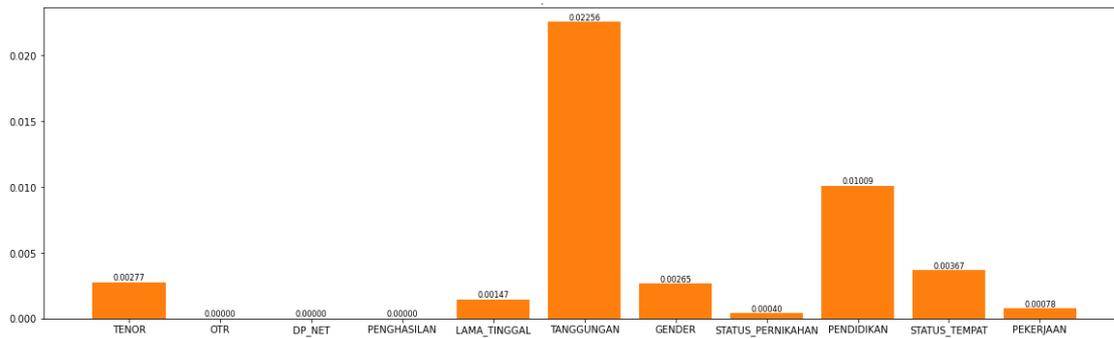
L1 Based Feature Selection

Pada metode ini tahap awalnya adalah mentransformasikan nilai masukan dari data latih menggunakan LSCV (Least Squares Cross-Validation). Setelah proses transformasi dilakukan pemilihan fitur berdasarkan metode L1 (L1-based feature Selection). Hasil penggunaan metode ini diperoleh nilai F1 sebesar 0,409. Nilai tersebut menunjukkan tingkat performa model yang diperoleh dengan menggunakan metode pemilihan fitur berbasis L1.

Namun perlu diperhatikan bahwa nilai F1 ini tidak lebih tinggi dari dua pendekatan sebelumnya yang telah diuji, yaitu pendekatan penghilangan fitur dengan varian rendah dan pendekatan Select K-Best. Hal ini menunjukkan bahwa meskipun pemilihan fitur berbasis L1 digunakan dalam eksperimen ini, metode ini gagal menghasilkan peningkatan performa model yang signifikan dibandingkan dua pendekatan sebelumnya. Hasil tersebut memberikan pemahaman bahwa dalam konteks data dan permasalahan tertentu, strategi penghapusan fitur berdasarkan varian rendah atau Select K-Best mungkin lebih efektif atau tepat dibandingkan menggunakan seleksi fitur berbasis L1.

Feature Importance from Coefficient

Tahap pertama dari pendekatan ini adalah dengan cara menentukan koefisien yang menunjukkan seberapa penting suatu fitur. Gambar 2 menunjukkan koefisien derajat kepentingan pada setiap fitur.



Gambar 2. Derajat kepentingan seluruh fitur

Pada eksperimen ini, peneliti melakukan training model terhadap setengah dari jumlah fitur dengan nilai derajat kepentingan tertinggi, yaitu: Tanggungan, Pendidikan, Status Tempat, Tenor, Gender. Hasil training menunjukkan nilai F1 tertinggi didapat dengan menggunakan seluruh fitur tersebut, dengan nilai F1 0.310

Kombinasi Seluruh Fitur

Setelah mencoba 4 teknik sebelumnya untuk melakukan seleksi fitur, peneliti juga menggunakan seluruh kombinasi dari setiap fitur untuk mendapatkan fitur dengan performa terbaik. 5 kombinasi fitur dengan performa F1 terbaik ditunjukkan pada table 2. Tabel 2 menunjukkan kombinasi fitur dengan nilai F1 yang paling tinggi dibandingkan dengan empat metode seleksi fitur sebelumnya dimana nilai F1 tertinggi adalah 0.4198. Sehingga dari seluruh metode seleksi fitur yang telah dicoba pada eksperimen ini, hasil yang mendekati performa terbaik ditunjukkan oleh metode seleksi fitur menggunakan metode select k-best dengan nilai F1 0.416.

Tabel 2. Top 5 performa kombinasi seluruh fitur

No	Fitur	Akurasi	Precision	Recall	F1
1	Tenor, OTR, Penghasilan, Lama Tinggal, Tanggungan, Gender, Status Pernikahan, Status Tempat, Pekerjaan	81.6	0.3935	0.4666	0.4198
2	OTR, Penghasilan, Tanggungan, Status Tempat, Pekerjaan	81.4	0.4010	0.4466	0.4183
3	Tenor, OTR, DP Net, Penghasilan, Lama Tinggal, Tanggungan, Gender, Status Pernikahan, Status Tempat, Pekerjaan	81.3	0.3932	0.4599	0.4182
4	Tenor, OTR, Penghasilan, Tanggungan, Gender, Pendidikan, Pekerjaan	80.4	0.3767	0.4733	0.4173
5	Tenor, OTR, Penghasilan, Lama Tinggal, Tanggungan, Gender, Status Tempat, Pekerjaan	81.1	0.3846	0.4733	0.4170

SIMPULAN

Penggunaan metode seleksi fitur dalam melakukan training membuat proses training menjadi lebih efisien. Sayangnya, metode seleksi fitur belum tentu memberikan hasil terbaik. Pada penelitian ini, penulis melakukan eksperimen seleksi fitur pada klasifikasi kelayakan pemberian kredit menggunakan 4 pendekatan seleksi fitur dengan hasil nilai F1 tertinggi 0.416 yang didapatkan menggunakan metode select k-best, diikuti oleh L1 Based Feature Selection dengan nilai F1 0.409 dan penghapusan fitur dengan varians rendah dengan nilai

F1 0.404. Seleksi fitur menggunakan feature important from coefficient berada pada posisi terakhir dengan nilai F1 0.310. Penelitian ini diharapkan dapat memberikan kontribusi berharga dalam pemahaman tentang pengambilan keputusan terkait pemberian kredit dan memperkaya strategi lembaga keuangan dalam mengelola risiko. Selain itu, hasil penelitian ini memiliki potensi untuk menyediakan solusi yang lebih transparan bagi pelanggan dalam konteks pemberian kredit.

DAFTAR PUSTAKA

- Chen, W., Xiang, G., Liu, Y., & Wang, K. (2012). Credit risk Evaluation by hybrid data mining technique. *Systems Engineering Procedia*, 3, 194–200. <https://doi.org/10.1016/j.sepro.2011.10.029>
- Hadianto, N., Novitasari, H. B., & Rahmawati, A. (2019). Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network. *Jurnal Pilar Nusa Mandiri*, 15(2), 163–170. <https://doi.org/10.33480/pilar.v15i2.658>
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455. <https://doi.org/10.25126/jtiik.201854803>
- Lubis, C. P. (2022). Perbandingan Metode Radial Basis Function dan Multilayer Perceptron Terhadap Resiko Kredit Sepeda Motor. *INFOSYS (INFORMATION SYSTEM) JOURNAL*, 7(1), 25. <https://doi.org/10.22303/infosys.7.1.2022.25-33>
- Mandala, I. G. N. N., Nawangpalupi, C. B., & Praktikto, F. R. (2012). Assessing Credit Risk: An Application of Data Mining in a Rural Bank. *Procedia Economics and Finance*, 4, 406–412. [https://doi.org/10.1016/s2212-5671\(12\)00355-3](https://doi.org/10.1016/s2212-5671(12)00355-3)
- Miftahusalam, A., Pratiwi, H., & Slamet, I. (2023). Perbandingan Metode Random Forest dan Naive Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile. *PROSIDING SEMINAR NASIONAL “PERAN TEKNOLOGI PENDIDIKAN MENUJU PEMBELAJARAN MASA DEPAN: TANTANGAN DAN PELUANG.”*
- Mohammad Abualigah, L., Al-diabat, M., Al Shinwan, M., Dhou, K., Alsalibi, B., Said Hanandeh, E., & Shehab, M. (2020). Hybrid Harmony Search Algorithm to Solve the Feature Selection for Data Mining Applications. In *Recent Advances in Hybrid Metaheuristics for Data Clustering* (pp. 19–37). Wiley. <https://doi.org/10.1002/9781119551621.ch2>
- Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation*, 5(1). <https://doi.org/10.1186/s40854-019-0121-9>
- Nguyen, B. H., Xue, B., & Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*, 54. <https://doi.org/10.1016/j.swevo.2020.100663>
- Pedregosa, F., & Varoquaux, G. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. *European Journal of Operational Research*, 156(2), 483–494. [https://doi.org/10.1016/S0377-2217\(02\)00911-6](https://doi.org/10.1016/S0377-2217(02)00911-6)
- Pritalia, G. L. (2022). Analisis Komparatif Algoritme Machine Learning pada Klasifikasi Kualitas Air Layak Minum. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 2(1).
- S. Brid, R. (2018). *Decision Trees — A simple way to visualize a decision*. Medium.Com. <https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb>
- Sugesti, A., & Abdul Mukid, M. (2019). Perbandingan Kinerja Mutual K-nearest Neighbor (Mknn) Dan K-nearest Neighbor (Knn) Dalam Analisis Klasifikasi Kelayakan Kredit. *JURNAL GAUSSIAN*, 8(3), 366–376. <http://ejournal3.undip.ac.id/index.php/gaussian>
- Syukron, A., & Subekti, A. (2018). Penerapan Metode Random Over-Under Sampling dan Random Forest untuk Klasifikasi Penilaian Kredit. *JURNAL INFORMATIKA*, 5(2).

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56–70. <https://doi.org/10.38094/jastt1224>