

# Optimasi Algoritma Random Forest Dengan Fitur Seleksi Backward Elimination Untuk Penilaian Kelayakan Kredit

Amrin<sup>1\*</sup>, Omar Pahlevi<sup>2</sup>, Sismadi<sup>3</sup>

<sup>1,2,3</sup>Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Jl. Kramat Raya No. 98 Senen, Jakarta Pusat-10420, Indonesia

E-mail: <sup>1</sup>amrin.ain@bsi.ac.id, <sup>2</sup>omar.opi@bsi.ac.id, <sup>3</sup>sismadi.ssm@bsi.ac.id

(\*: corresponding author)

**Abstrak**— Kredit sekarang menjadi tren di masyarakat. Problem kredit adalah sejarah penggunaan kartu kredit yang salah. Dampak yang ditimbulkan dapat menyebabkan kredit macet. Jika pelanggan tidak membayar utang yang telah disepakati dengan bank, mereka dapat meningkatkan risiko kredit mereka. Dalam penelitian ini, peneliti menerapkan algoritma *Random Forest* tanpa optimasi dan Algoritma *Random Forest* dengan Optimasi Fitur Seleksi *Backward Elimination* untuk mengklasifikasikan status kelayakan kredit. Peneliti menggunakan 481 catatan kredit kendaraan dengan ulasan "bad" dan "good". Variabel independen digunakan dalam penelitian adalah status tanggungan, usia, pendidikan terakhir, status pernikahan, pekerjaan, status perusahaan, pendapatan, status pekerjaan, kondisi rumah, lama tinggal dan uang muka. Dari hasil penelitian dan pengujian, performa model random forest tanpa backward elimination untuk penilaian kelayakan kredit memberikan tingkat akurasi kebenaran sebesar 78,60% dengan nilai area under the curve (AUC) sebesar 0,907. Sedangkan Performa model random forest dengan backward elimination memberikan tingkat akurasi kebenaran sebesar 89,81% dengan nilai area under the curve (AUC) sebesar 0,922. Hal ini membuktikan bahwa optimasi dengan backward elimination dapat meningkatkan kinerja metode klasifikasi yang digunakan.

**Kata Kunci**— *random forest, backward elimination, confusion matrix, ROC curva.*

**Abstract**— *Credit is now a trend in society. Credit problems are the history of incorrect use of credit cards. The impact can cause bad credit. If customers do not pay the debt that has been agreed with the bank, they can increase their credit risk. In this study, researchers applied the Random Forest algorithm without optimization and the Random Forest Algorithm with Backward Elimination Feature Selection Optimization to classify creditworthiness status. Researchers used 481 vehicle credit records with "bad" and "good" reviews. The independent variables used in the study were dependent status, age, last education, marital status, occupation, company status, income, employment status, house condition, length of stay and down payment. From the results of the study and testing, the performance of the random forest model without backward elimination for creditworthiness assessment provided a truth accuracy level of 78.60% with an area under the curve (AUC) value of 0.907. While the performance of the random forest model with backward elimination provided a truth accuracy level of 89.81% with an area under the curve (AUC) value of 0.922. This*

*proves that optimization with backward elimination can improve the performance of the classification method used.*

**Keyword**— *random forest, backward elimination, confusion matrix, ROC curva.*

## I. PENDAHULUAN

Kredit memiliki sejarah yang panjang. Dokumen pertama tentang pemberian kredit berasal dari Mesir kuno, Babilonia, dan Asyur sekitar 3.000 tahun yang lalu. Kredit menjadi lebih populer di Eropa selama Abad Pertengahan karena pedagang dari seluruh benua memperdagangkan barang dari berbagai negara [1]. Data keuangan termasuk laporan arus kas, laporan laba rugi, neraca, dan rasio keuangan antara lain: kredit, modal kerja, pertumbuhan, dan profitabilitas [2]. Risiko kredit adalah salah satu masalah terbesar yang dihadapi organisasi keuangan di seluruh dunia.

Tahapan analisa kredit mulai dijalankan pada saat proses pemberian pinjaman atau kredit. Karena banyaknya dokumen yang masuk, diperlukan waktu yang cukup lama untuk menganalisa data nasabah. Prinsip *Character, Capital, Capacity, Collateral, dan Condition* (5C) memerlukan verifikasi lapangan yang profesional [3].

Tidak semua konsumen atau pelanggan mampu membayar tagihan kredit mereka tepat waktu setiap bulannya selama kegiatan perkreditan berlangsung, oleh karena itu masalah-masalah tagihan yang berkembang dapat mempengaruhi kemampuan perusahaan untuk bertahan [4]. Selain itu, data penilaian kredit memiliki dimensi yang tinggi. Selain itu, fitur yang berlebihan atau tidak relevan sering kali menyebabkan *overfitting* pada model, sehingga mengganggu kinerja [5]. Semakin tidak mungkin untuk mengabaikan pentingnya peringkat kredit bisnis di sektor keuangan. Hal ini penting karena melindungi uang investor dan mengurangi ketidakseimbangan pengetahuan yang ada di antara bisnis, lembaga keuangan, dan masyarakat [6].

Saat ini, banyak industri, seperti bisnis, perbankan, dan serikat kredit, menggunakan *data mining* [7]. Model klasifikasi pada *data mining* yang berfungsi sebagai model keputusan, biasanya didukung oleh metode diskritisasi fitur, resampling data, dan seleksi fitur. Teknik-teknik di atas dapat digunakan dalam berbagai publikasi. Ketika subset fitur yang relevan dipilih, beban komputasi dapat dikurangi dan efisiensi

dan pemahaman model dapat ditingkatkan [8]. Tahapan mengidentifikasi dan memprediksi nasabah secara baik dan benar dapat dilakukan pada saat sebelum proses pemberian pinjaman dengan cara memeriksa data historis pinjaman nasabah tersebut. Kegiatan ini merupakan upaya yang dilakukan oleh industri perbankan pada masa ini dalam menghadapi masalah resiko kredit [9].

Beberapa penelitian terdahulu yang terkait dengan tema metode-metode yang pernah digunakan untuk menyelesaikan prediksi kelayakan kredit, diantaranya penelitian yang dilakukan oleh [10], penelitian ini menganalisa Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. Hasil pengujian menunjukkan bahwa dengan penggunaan optimasi GA ataupun Bagging belum mampu meningkatkan akurasi dari algoritma RF untuk klasifikasi set data Bank Marketing, dimana dengan menggunakan optimasi ataupun tidak akurasi yang diperoleh adalah sebesar 88,30%. Optimasi yang dilakukan belum berhasil disebabkan data yang digunakan terlalu imbalance, sehingga ketika dilakukan proses bagging ataupun GA sebaran data yang dihasilkan masih cukup banyak. Selanjutnya penelitian yang dilakukan oleh [11], penelitian ini tentang Optimasi Decision Tree untuk Risiko Kredit KMG Bank DKI. Hasil dari penelitian ini adalah optimalisasi Particle Swarm Optimization (PSO) pada analisis risiko data kredit multiguna di Bank DKI menghasilkan nilai akurasi sebesar 99,13 %. Sementara itu, penggunaan Decision Tree tanpa optimalisasi PSO hanya mencapai akurasi sebesar 97,83 %. Hal ini menunjukkan bahwa terdapat peningkatan akurasi yang signifikan dengan penggunaan PSO.

Berikutnya penelitian oleh [12], penelitian ini membahas tentang Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit. Kesimpulan dari penelitian ini adalah bahwa Algoritma C4.5 terbukti efektif dalam memprediksi kelayakan kredit dengan tingkat akurasi sebesar 79,11%. Metode Forward Selection dan Stratified Sampling terbukti berhasil meningkatkan akurasi algoritma C4.5 sebesar 9,2% dalam memprediksi kelayakan kredit. Kemudian penelitian oleh [13] yang membahas tentang Peningkatan Akurasi Kelayakan Kredit Menggunakan Particle Swarm Optimization. Hasil dari penelitian ini adalah bahwa akurasi dari masing-masing teknik klasifikasi meningkat ketika digabungkan dengan PSO pada pembobotan atribut. Model SVM-PSO memiliki akurasi dan presisi paling baik diantara teknik lain.

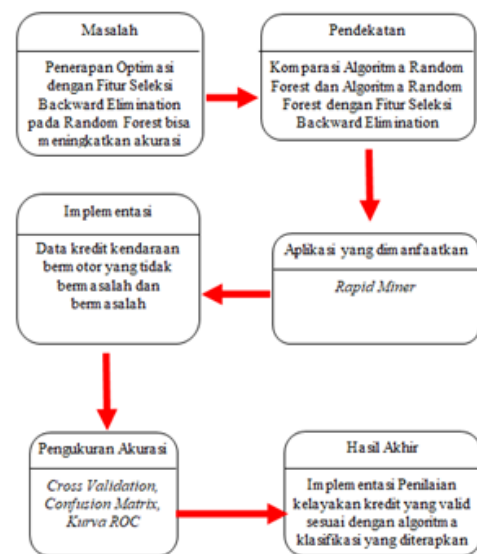
Penelitian terdahulu selanjutnya yang dilakukan oleh [14], penelitian ini membahas tentang Data Mining Optimization Using Sample Bootstrapping and Particle Swarm Optimization in the Credit Approval Classification. Dari hasil penelitian dan pengujian, model Decision Tree yang menggunakan algoritma C4.5 dengan optimasi PSO dan Sample Bootstrapping mendapatkan nilai 94.44%. Sementara hasil pengujian model Naive Bayes menggunakan optimasi PSO dan Sample Bootstrapping mendapatkan nilai akurasi 90,10%. Penelitian berikutnya oleh [15] tentang

Optimasi K-Nearst-Neighbor (K-NN) Dengan Particle Swarm Optimization Pada Klasifikasi Nasabah Kredit.

Berdasarkan hasil eksperimen dan evaluasi penelitian yang dilakukan dengan dataset yang berbeda pada penelitian sebelumnya, maka dapat disimpulkan bahwa eksperimen dengan tingkat akurasi tertinggi pada metode k-Nearest Neighbor (K-NN) pada particle swarm optimization sebesar 90.73% dan nilai Area Under Curve (AUC) sebesar 0.464 berhasil memperbaiki akurasi penelitian sebelumnya yakni 86.09%. Dalam penelitian ini, peneliti akan menerapkan algoritma *Random Forest* tanpa optimasi dan Algoritma *Random Forest* dengan Optimasi Fitur Seleksi *Backward Elimination* untuk mengklasifikasikan status kelayakan kredit.

## II. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahap seperti terlihat pada kerangka pemikiran Gambar 1. Permasalahan pada penelitian ini adalah Apakah Penerapan Optimasi dengan Fitur Seleksi Backward Elimination pada Random Forest bisa meningkatkan akurasi? Untuk itu dibuat approach (model) yaitu algoritma Random Forest dan Algoritma Random Forest dengan Fitur Seleksi Backward Elimination untuk memecahkan permasalahan kemudian dilakukan pengujian terhadap kinerja dari metode tersebut. Pengujian menggunakan metode *Cross Validation*, *Confusion Matrix* dan kurva ROC. Untuk mengembangkan aplikasi (development) berdasarkan model yang dibuat, digunakan Rapid Miner.



Sumber: Hasil Penelitian (2025)

Gambar 1. Kerangka Pemikiran Penelitian

## III. HASIL DAN PEMBAHASAN

### A. Analisa Data

Dataset yang digunakan sebanyak 481 data yang *balance* kredit kendaraan bermotor baik yang bermasalah maupun yang tidak bermasalah. Variabel input pada penelitian ini terdiri dari tiga belas variabel, yaitu: 1) Status Perkawinan, 2) Jumlah Tanggungan, 3) Umur 4) Status Tempat Tinggal, 5) Kepemilikan Rumah, 6) Pekerjaan, 7) Status Pekerjaan, 8)

Status Perusahaan, 9) Penghasilan, 10) Uang Muka, 11) Pendidikan, 12) Lama Tinggal, dan 13) Kondisi Rumah. Gambar 1 menunjukkan sampel dataset yang digunakan untuk menguji algoritma yang akan diuji.

status perkawinan	jumlah tanggungan	pendidikan terakhir	usia	kepemilikan rumah	lama tinggal	kondisi rumah
menikah	2-3	SLTP	21-55	ortu	>5	permanen
belum menikah	tdk ada	SLTA	21-55	ortu	1-3	permanen
menikah	2-3	SLTA	21-55	KPR	3-5	permanen
belum menikah	tdk ada	SLTA	<21>60	ortu	1-3	permanen
menikah	>3	SD	21-55	atas nama	>5	non permanen
menikah	2-3	SLTP	21-55	milik sendiri	>5	permanen
menikah	2-3	SLTA	21-55	milik sendiri	>5	permanen

Gambar 2. Sampel Dataset

### B. Pengujian Model

Penelitian ini dilakukan dengan eksperimen pengujian pada model yang diusulkan. Kemudian dilakukan evaluasi dan validasi model untuk menghasilkan nilai accuracy dan AUC. Pengujian menggunakan Rapidminer dengan operator 10-fold cross-validation untuk mendapatkan hasil accuracy dan AUC pada setiap algoritma yang diuji. Evaluasi yang dilakukan adalah dengan Confusion Matrix dan ROC Curve atau Area Under Curve (AUC).

#### 1. Confusion Matrix

##### a. Algoritma Random Forest

Tabel 1 adalah *confusion matrix* untuk algoritma Random Forest. Diketahui 92 data diklasifikasi “bad” diprediksi sesuai dengan data sebenarnya, lalu 87 data diprediksi “good” tetapi ternyata “bad”. Kemudian 286 data diklasifikasi “good” diprediksi sesuai, dan 16 data diprediksi “bad” ternyata “good”.

Tabel 1. Model *Confusion Matrix* untuk Random Forest

accuracy: 78.60% +/- 8.46% (mikro: 78.59%)			
	true bad	true good	class precision
pred. bad	92	16	85.19%
pred. good	87	286	76.68%
class recall	51.40%	94.70%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

##### b. Algoritma Random Forest dengan Backward Elimination

Tabel 2 adalah *confusion matrix* untuk algoritma *Random Forest* dengan Backward Elimination. Diketahui 278 data diklasifikasi “good” diprediksi sesuai dengan data sebenarnya, lalu 24 data diprediksi “bad” tetapi ternyata “good”. Kemudian 154 data diklasifikasi “bad” diprediksi sesuai, dan 25 data diprediksi “good” ternyata “bad”.

Tabel 2. Model *Confusion Matrix* untuk Metode *Random Forest* dengan Backward Elimination.

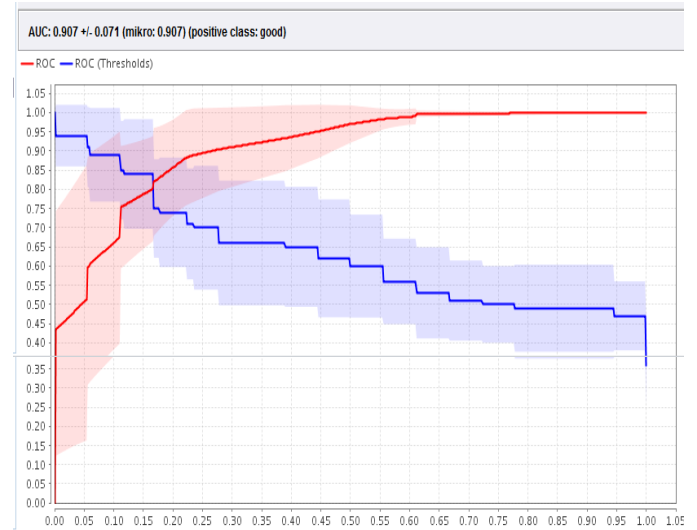
accuracy: 89.81%			
	true bad	true good	class precision
pred. bad	154	24	86.52%
pred. good	25	278	91.75%
class recall	86.03%	92.05%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

## 2. Kurva ROC

### a. Algoritma Random Forest

kurva ROC untuk algoritma random forest seperti ditunjukkan oleh Gambar 3.

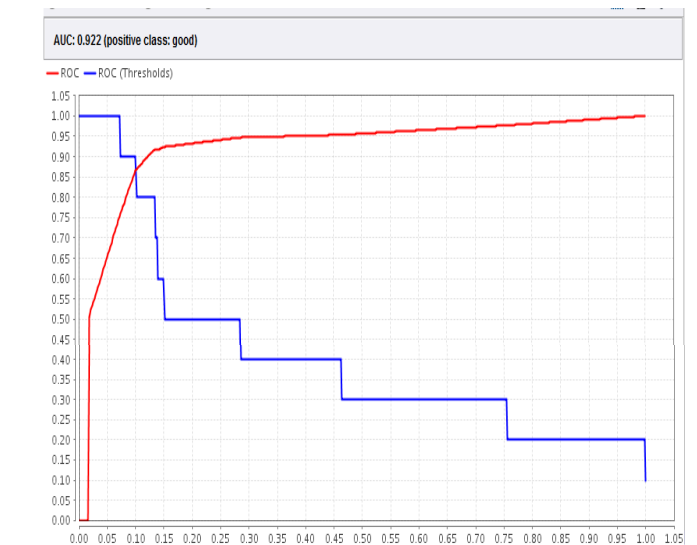


Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)  
 Gambar 3. Kurva ROC Algoritma Random Forest

Kurva ROC pada gambar 2 mengekspresikan *confusion matrix*. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.

### b. Algoritma Random Forest dengan Backward Elimination

Kurva ROC untuk algoritma *Random Forest* dengan Backward Elimination seperti ditunjukkan oleh Gambar 4 berikut:



Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)  
 Gambar 4. Kurva ROC Metode *Random Forest* dengan Backward Elimination.

### 3. Analisis Hasil

Perbandingan nilai *accuracy*, *precision*, *recall* dan *AUC* untuk algoritma *Random Forest* dan *Random Forest* dengan *Backward Elimination* ditunjukkan oleh tabel 3.

Tabel 3. Evaluasi dan Validasi Model

Algoritma	Accuracy	Precision	Recall	AUC
<i>Random Forest</i>	78.60%	78.20%	94.72%	0.907
<i>Random Forest dengan Backward Elimination</i>	89.81%	91.75%	92.05%	0.922

Sumber: Hasil Pengolahan Menggunakan RapidMiner (2025)

Tabel 3 membandingkan *accuracy*, *precision*, *recall* dan *AUC* dari algoritma *random forest* tanpa *backward elimination* dan *random forest* dengan *backward elimination*. Terlihat bahwa nilai *accuracy algoritma random forest* dengan *backward elimination* memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma *random forest* tanpa *backward elimination*. Begitupun dengan nilai *AUC*, jelast bahwa nilai *AUC algoritma random forest* dengan *backward elimination* memiliki nilai yang lebih tinggi dibandingkan dengan algoritma *random forest* tanpa *backward elimination* Untuk klasifikasi *data mining*, nilai *AUC* dapat dibagi menjadi beberapa kelompok [16].

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan di atas maka dapat disimpulkan bahwa model algoritma *random forest* dan *random forest* dengan *backward elimination* termasuk katagori klasifikasi sangat baik.

#### IV. KESIMPULAN

Dari hasil penelitian dan pengujian, performa model *random forest* tanpa *backward elimination* untuk penilaian kelayakan kredit memberikan tingkat akurasi kebenaran sebesar 78,60% dengan nilai *area under the curva (AUC)* sebesar 0,907. Sedangkan Performa model *random forest* dengan *backward elimination* memberikan tingkat akurasi kebenaran sebesar 89,81% dengan nilai *area under the curve (AUC)* sebesar 0,922. Hal ini membuktikan bahwa optimasi dengan *backward elimination* dapat meningkatkan kinerja metode klasifikasi yang digunakan dengan peningkatan akurasi sebesar 11,21%.

#### REFERENSI

[1] T. Himberg, "Loan Default Prediction with Machine Learning," 2020. Doi: 10.1007/978-3-030-37720-5\_4.  
[2] D. A. A. Al-Sayed, W. A. Q. Awad, and M. T. M. Salem, "a Comparative Study of Forecasting Corporate Credit Ratings Using Artificial Neural Networks, Support Vector Machine, Random Forest, the Naive Bayes, Decision Tree and -Nearest Neighbor," *Adv. Appl.*

*Stat.*, vol. 91, no. 2, pp. 125–139, 2023. Doi: 10.17654/0972361724010.  
[3] M. Lasena and S. R. Ahmad, "Sistem Pendukung Keputusan Kelayakan Pemberian Kredit Nasabah Dengan Metode Electre," *Bull. Inf. Technol.*, vol. 4, no. 2, pp. 232–238, 2023. Doi: 10.47065/bit.v4i2.690.  
[4] A. B. Alfian and A. H. D. Nugroho, "Analisis Sistem Pengendalian Internal Terhadap Efektivitas Pemberian Kredit Kendaraan Bermotor di PT. Indomobil Finance Indonesia Cabang Semarang," *J. Econ. Bussines Account.*, vol. 7, no. 4, pp. 9071–9084, 2024. Doi: 10.31539/costing.v7i4.8902.  
[5] Y. Jin, W. Zhang, X. Wu, Y. Liu, and Z. Hu, "A Novel Multi-Stage Ensemble Model with a Hybrid Genetic Algorithm for Credit Scoring on Imbalanced Data," *IEEE Access*, vol. 9, pp. 143593–143607, 2021. Doi: 10.1109/ACCESS.2021.3120086.  
[6] Y. Song, "Enterprise Credit Rating Prediction Model Based on Data Mining Algorithm," in *Innovative Computing Vol 1 - Emerging Topics in Artificial Intelligence*, J. C. Hung, J.-W. Chang, and Y. Pei, Eds., Singapore: Springer Nature Singapore, 2023, pp. 745–751.  
[7] Y. Oktafriani, G. Firmansyah, B. Tjahjono, and A. M. Widodo, "Analysis of Data Mining Applications for Determining Credit Eligibility Using Classification Algorithms C4.5, Naïve Bayes, K-NN, and Random Forest," *Asian J. Soc. Humanit.*, vol. 1, no. 12, pp. 1139–1158, 2023. Doi: 10.59888/ajosh.v1i12.119.  
[8] P. Ziemba, J. Becker, A. Becker, A. Radomska-Zalas, M. Pawluk, and D. Wierzba, "Credit decision support based on real set of cash loans using integrated machine learning algorithms," *Electron.*, vol. 10, no. 17, pp. 1–22, 2021. Doi: 10.3390/electronics10172099.  
[9] A.- Amrin and O.- Pahlevi, "Implementation of Logistic Regression Classification Algorithm and Support Vector Machine for Credit Eligibility Prediction," *J. Informatics Telecommun. Eng.*, vol. 5, no. 2, pp. 433–441, 2022. Doi: 10.31289/jite.v5i2.6220.  
[10] Y. Religia, A. Nugroho, and W. Hadikristanto, "Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, 2021. Doi: 10.29207/resti.v5i1.2813.  
[11] J. B. E. Putry, A. T. Sasongko, and W. Hadikristanto, "Optimasi Decision Tree Menggunakan Particle Swarm Optimization (PSO) pada Risiko Kredit KMG Bank DKI," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1403–1410, 2024. Doi: 10.57152/malcom.v4i4.1521.  
[12] I. Ubaedi and Y. M. Djaksana, "Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit," *JSil (Jurnal Sist. Informasi)*, vol. 9, no. 1, pp. 17–26, 2022. Doi: 10.30656/jsii.v9i1.3505.  
[13] S. I. Novichasari, "Peningkatan Akurasi Kelayakan Kredit Menggunakan Particle Swarm Optimization," *J. Multimatrix*, vol. 3, no. 1, pp. 86–90, 2021, [Online]. Available: <https://jurnal.unw.ac.id/index.php/mm/article/view/1533%0Ahttps://jurnal.unw.ac.id/index.php/mm/article/view/1533/992>  
[14] A. A. Agustian and A. Bisri, "Data Mining Optimization Using Sample Bootstrapping and Particle Swarm Optimization in the Credit Approval Classification," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 1, pp. 18–27, 2019. Doi: 10.24014/ijaidm.v2i1.6299.  
[15] R. Racmatl and R. D. Likliwati, "Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Pada Klasifikasi Nasabah Kredit Kendaraan," *J. JTRISTE*, vol. 6, no. 1, pp. 9–16, 2019, [Online]. Available: <https://jurnal.kharisma.ac.id/jtriste/article/view/89>  
[16] F. Gorunescu, *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer, 2011.