Jurnal Komtika (Komputasi dan Informatika),
Vol. 9 No. 2 | Nov 2025

Optimasi Gradient Boosted Trees dalam Memprediksi Minat Nasabah untuk Berlangganan Pinjaman Berjangka

Refi Riduan Achmad[1*], Muhammad Zulfariansyah[2]
[1,2]Program Studi Teknik Informatika, Universitas Nahdlatul Ulama Kalimantan Timur
*email: refiriduanachmad@unukaltim.ac.id

DOI: https://doi.org/ 10.31603/komtika.v9i2.15259

Received: 06-11-2025, Revised: 13-11- 2025, Accepted: 16-11-2025

ABSTRACT

The increasing need for data-driven decision-making in the banking sector has encouraged the use of machine learning methods to predict customer interest in financial products. This study aims to optimize the parameters of the Gradient Boosted Trees (GBT) algorithm to improve prediction performance for term deposit subscription data. The research follows the CRISP-DM framework, including stages of data preprocessing, modeling, parameter optimization, and model evaluation using accuracy, precision, recall, and Area Under Curve (AUC) metrics. Experimental results demonstrate that the optimized GBT model achieved the best performance compared to Decision Tree and Random Forest, with an AUC of 0.976, accuracy of 97.50%, precision of 63.89%, and recall of 65.71%. Meanwhile, the GBT model with cross-validation achieved an AUC of 0.956 and accuracy of 95.89%, indicating more stable performance and better generalization to data variations This study confirms that systematic parameter optimization enhances model discrimination and predictive accuracy in imbalanced data scenarios. Practically, the optimized GBT model can be utilized by financial institutions to improve data-driven marketing strategies by accurately identifying potential customers.

Keywords: *Gradient Boosted Trees; Parameter Optimization; Cross-validation; Customer Interest Prediction; Banking Data.*

ABSTRAK

Kebutuhan akan pengambilan keputusan berbasis data pada sektor perbankan mendorong penerapan metode *machine learning* untuk memprediksi minat nasabah terhadap produk keuangan. Penelitian ini bertujuan untuk mengoptimasi parameter algoritma Gradient Boosted Trees (GBT) guna meningkatkan performa prediksi terhadap data langganan deposito berjangka. Proses penelitian mengikuti tahapan CRISP-DM, yang meliputi pra-pemrosesan data, pemodelan, optimasi parameter, serta evaluasi model menggunakan metrik akurasi, presisi, *recall*, dan *Area Under Curve* (AUC). Hasil eksperimen menunjukkan bahwa model GBT dengan parameter teroptimasi memberikan performa terbaik dibandingkan Decision Tree dan Random Forest, dengan nilai AUC sebesar 0.976, akurasi 97.50%, presisi 63.89%, dan recall 65.71%. Sementara itu, model GBT dengan *cross validation* menunjukkan nilai AUC sebesar 0.956 dan akurasi 95.89%, yang menggambarkan performa yang lebih stabil dan generalisasi yang lebih baik terhadap variasi data. Penelitian ini membuktikan bahwa optimasi parameter secara sistematis dapat meningkatkan kemampuan diskriminatif model dan efektivitas prediksi pada data yang tidak seimbang. Secara praktis, hasil ini dapat digunakan oleh lembaga keuangan untuk meningkatkan strategi pemasaran berbasis data dalam menargetkan calon nasabah potensial secara lebih akurat.

Keywords: Gradient Boosted Trees; Optimasi Parameter; Cross Validation; Prediksi Minat Nasabah, Data Perbankan.

PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong transformasi besar dalam dunia bisnis dan industri, termasuk sektor keuangan. Pemanfaatan data sebagai dasar pengambilan keputusan atau yang dikenal dengan konsep *data-driven decision making* telah menjadi kebutuhan penting dalam mendukung efisiensi dan keakuratan strategi bisnis [1]. Dalam konteks perbankan, kemampuan untuk memprediksi perilaku serta minat nasabah terhadap produk tertentu, seperti *term deposit* atau pinjaman berjangka, berperan penting dalam menentukan keberhasilan kampanye pemasaran dan peningkatan loyalitas pelanggan [2]. Model prediksi yang akurat memungkinkan lembaga keuangan menargetkan nasabah potensial secara lebih efektif, mengurangi biaya promosi, serta meningkatkan tingkat konversi penawaran produk [3].

Salah satu pendekatan yang banyak digunakan dalam pengembangan sistem prediksi di bidang keuangan adalah *machine learning*. Algoritma ini mampu mempelajari pola kompleks dari data historis untuk menghasilkan prediksi yang lebih andal dibandingkan metode statistik konvensional [4]. Di antara berbagai algoritma yang ada, Gradient Boosted Trees (GBT) menjadi salah satu metode yang paling populer karena kemampuannya menggabungkan sejumlah *weak learners* berbasis *decision tree* menjadi model kuat (*strong learner*) melalui proses *boosting* yang iteratif [5], [6]. Kinerja GBT sangat bergantung pada penentuan parameter optimal (*hyperparameter tuning*) seperti jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan laju pembelajaran (*learning_rate*) [7]. Penentuan parameter yang tidak tepat dapat menyebabkan model *underfitting* atau *overfitting*, sehingga menurunkan akurasi dan kemampuan generalisasi model [8]. Oleh karena itu, proses optimasi parameter menjadi bagian penting untuk memperoleh konfigurasi yang mampu memaksimalkan kinerja model prediktif.

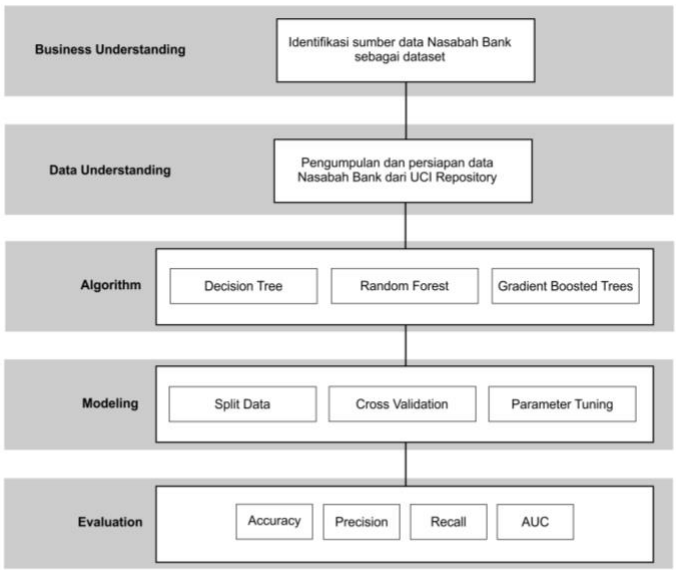
Beberapa penelitian terdahulu telah menerapkan GBT dalam konteks prediksi perbankan, namun sebagian besar belum menitikberatkan pada optimasi parameter secara sistematis. Misalnya, penelitian oleh Moro et al. [9] menggunakan *boosting* untuk prediksi keberhasilan kampanye pemasaran deposito, namun tidak melakukan eksplorasi mendalam terhadap kombinasi parameter terbaik. Demikian pula, studi oleh Vrigazova [10] menunjukkan bahwa tuning parameter dapat meningkatkan performa model hingga 5–10%, tetapi masih terbatas pada skenario data seimbang. Kondisi ini menunjukkan adanya celah penelitian terkait bagaimana optimasi parameter dapat meningkatkan kinerja prediksi pada data perbankan yang tidak seimbang (*imbalanced data*). Berdasarkan permasalahan tersebut, penelitian ini berfokus pada optimasi parameter algoritma Gradient Boosted Trees (GBT) untuk meningkatkan akurasi dan sensitivitas model dalam memprediksi minat nasabah terhadap produk pinjaman berjangka. Penelitian ini tidak hanya membangun model prediktif berbasis GBT, tetapi juga menerapkan strategi penyetelan parameter secara sistematis untuk memperoleh konfigurasi optimal. Hasil optimasi kemudian dibandingkan dengan model *Decision Tree* dan *Random Forest* untuk mengevaluasi peningkatan kinerja yang dihasilkan melalui proses tuning.

Secara teoretis, penelitian ini memperkuat bukti empiris bahwa proses optimasi parameter berperan penting dalam meningkatkan kemampuan generalisasi model *machine learning* pada data tidak seimbang [11]. Secara praktis, hasil penelitian ini diharapkan dapat

membantu lembaga keuangan dalam mengidentifikasi calon nasabah potensial serta mendukung pengambilan keputusan berbasis data yang lebih efisien dan tepat sasaran. Adapun struktur artikel ini disusun sebagai berikut: bagian kedua membahas metode penelitian yang meliputi tahapan pemrosesan data dan strategi optimasi parameter; bagian ketiga menyajikan hasil eksperimen dan analisis kinerja model; bagian keempat berisi pembahasan komparatif antar algoritma; dan bagian terakhir menyajikan kesimpulan serta arah penelitian selanjutnya.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning dengan mengikuti kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) [1]. Pendekatan ini dipilih karena menyediakan alur sistematis yang dimulai dari pemahaman bisnis hingga evaluasi model, serta telah banyak diterapkan dalam penelitian berbasis data di bidang perbankan dan pemasaran [2]. Gambar 1 memperlihatkan diagram alir tahapan penelitian yang menggambarkan alur kerja dari tahap awal hingga evaluasi hasil model.



Gambar 1. Diagram Alir Tahapan Penelitian (CRISP-DM)

Tahap *business understanding* bertujuan untuk memahami konteks dan permasalahan bisnis yang akan diselesaikan. Permasalahan utama dalam penelitian ini adalah rendahnya tingkat konversi kampanye pemasaran produk pinjaman berjangka di sektor perbankan akibat ketidaktepatan dalam mengidentifikasi calon nasabah potensial. Oleh karena itu, diperlukan model prediksi yang mampu membedakan nasabah yang berminat dan tidak berminat berdasarkan data historis. Fokus penelitian ini adalah mengoptimalkan parameter Gradient Boosted Trees (GBT) agar mampu menghasilkan kinerja prediksi yang lebih akurat dan sensitif terhadap kelas minoritas (*interested customers*).

Dataset yang digunakan bersumber dari UCI Machine Learning Repository, yaitu *Bank Marketing Data Set* [3]. Dataset ini berisi data hasil kampanye pemasaran lembaga perbankan terhadap produk deposito berjangka, terdiri dari 41.188 entri dan 20 atribut. Atribut mencakup data demografis, sosial-ekonomi, serta karakteristik kontak nasabah, seperti

age, *job*, *marital*, *education*, *balance*, dan *duration*. Distribusi kelas pada data bersifat tidak seimbang (*imbalanced*), di mana hanya sekitar 11% nasabah yang berlangganan produk. Hal ini menjadi tantangan utama karena model cenderung bias terhadap kelas mayoritas. Untuk mengatasi hal tersebut, dilakukan evaluasi menggunakan metrik tambahan seperti *recall* dan *AUC*, yang lebih sensitif terhadap kelas minoritas [4].

Tahapan data *preprocessing* bertujuan untuk menyiapkan data agar layak digunakan dalam proses pelatihan model. Langkah-langkah yang dilakukan meliputi:

1. Pembersihan Data – Menghapus atribut redundan dan menangani nilai kosong (*missing values*).
2. Transformasi Atribut Kategorikal – Menggunakan metode Label Encoding untuk mengubah variabel kategorikal menjadi numerik, sesuai kebutuhan algoritma berbasis pohon keputusan [5].
3. Normalisasi Data Numerik – Dilakukan dengan metode Min-Max Scaling untuk menyesuaikan rentang nilai antar fitur.
4. Penanganan Data Tidak Seimbang – Data dibiarkan dalam kondisi asli, tetapi evaluasi dilakukan dengan memperhatikan metrik *AUC* dan *recall* agar model tidak bias terhadap kelas mayoritas [6].

Tahapan pra-pemrosesan ini menghasilkan dataset bersih dan konsisten, siap digunakan untuk tahap pemodelan.

Tahap pemodelan dilakukan menggunakan tiga algoritma utama: Decision Tree (DT), Random Forest (RF), dan Gradient Boosted Trees (GBT). Ketiganya dipilih karena berbasis pohon keputusan, namun memiliki perbedaan dalam mekanisme pembelajaran [7].

1. Decision Tree (DT) membangun pohon tunggal berdasarkan entropi atau *Gini index*.
2. Random Forest (RF) membentuk kumpulan pohon secara acak dan menggabungkan hasil prediksi secara voting.
3. Gradient Boosted Trees (GBT) membangun model secara bertahap, di mana setiap pohon baru berfokus memperbaiki kesalahan pohon sebelumnya [8].

Proses optimasi parameter GBT dilakukan menggunakan metode grid search yang menguji kombinasi nilai terbaik dari parameter utama:

1. *n_estimators* $\in \{50, 100, 150\}$
2. *max_depth* $\in \{3, 4, 5, 6\}$
3. *learning_rate* $\in \{0.05, 0.1, 0.15\}$

Evaluasi tiap kombinasi dilakukan dengan validasi silang (*k-fold cross-validation*, k = 10) untuk mengurangi risiko *overfitting* [9].

Evaluasi model dilakukan menggunakan empat metrik utama: akurasi, precision, recall, dan Area Under the Curve (AUC) [10].

1. *Akurasi* mengukur proporsi prediksi benar terhadap total prediksi.
2. *Precision* menilai tingkat ketepatan model dalam mengidentifikasi kelas positif.
3. *Recall* mengukur kemampuan model dalam mengenali seluruh data kelas positif.
4. *AUC* menunjukkan kemampuan model dalam membedakan kelas positif dan negatif secara keseluruhan.

Seluruh proses pelatihan dan evaluasi dilakukan menggunakan perangkat lunak RapidMiner Studio, yang mendukung algoritma ensemble serta evaluasi berbasis *cross-validation* [11].

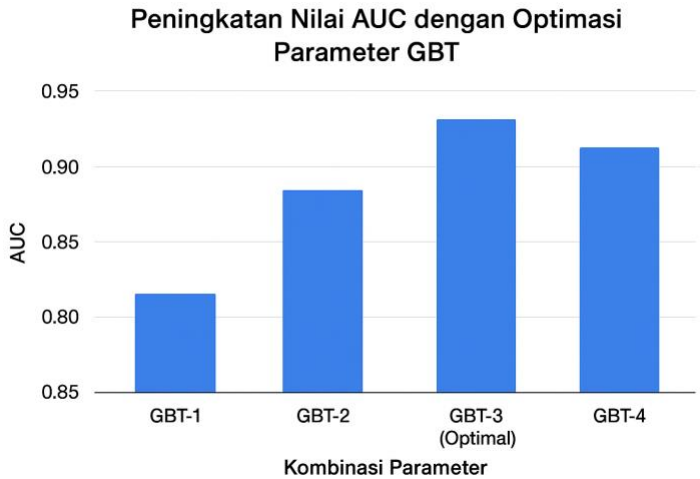
HASIL DAN PEMBAHASAN

Tahapan pemodelan dilakukan dengan menerapkan tiga algoritma pembelajaran mesin berbasis *decision tree*, yaitu Decision Tree (DT), Random Forest (RF), dan Gradient Boosted Trees (GBT). Pemilihan ketiga algoritma ini didasarkan pada kemampuannya dalam menangani data non-linear serta interpretabilitas yang tinggi dalam konteks prediksi perilaku nasabah [12]. Proses pelatihan model dilakukan menggunakan dataset hasil pra-pemrosesan, dengan teknik *10-fold cross-validation* untuk meminimalkan *overfitting* [13]. Model Decision Tree digunakan sebagai *baseline* awal, Random Forest diterapkan untuk memperbaiki stabilitas model melalui mekanisme *bagging*, sedangkan GBT digunakan untuk memperbaiki kesalahan prediksi secara iteratif melalui mekanisme *boosting* [14]. Hasil awal menunjukkan bahwa GBT tanpa optimasi sudah memberikan performa lebih baik dibandingkan DT dan RF, namun keseimbangan antara *precision* dan *recall* belum optimal. Oleh karena itu, dilakukan proses optimasi parameter seperti yang dijelaskan pada Tabel 1.

Proses optimasi dilakukan menggunakan metode grid search, yang mengevaluasi kombinasi nilai parameter utama (*n_estimators*, *max_depth*, dan *learning_rate*) untuk memperoleh konfigurasi yang menghasilkan kinerja terbaik. Hasil eksperimen menunjukkan bahwa konfigurasi terbaik diperoleh pada kombinasi *n_estimators* = 100, *max_depth* = 5, dan *learning_rate* = 0.1, dengan akurasi 97,50% dan nilai AUC sebesar 0.976. Nilai AUC tersebut menunjukkan kemampuan diskriminatif model yang sangat baik, di mana model mampu membedakan kelas positif dan negatif dengan tingkat ketepatan yang tinggi [15].

Tabel 1. Hasil Eksperimen Optimasi Parameter GBT

Konfigurasi	n_estimators	max_depth	learning_rate	Akurasi (%)	AUC
GBT-1	50	3	0.05	96.42	0.931
GBT-2	100	4	0.10	97.18	0.961
GBT-3 (Optimal)	100	5	0.10	97.50	0.976
GBT-4	150	6	0.15	97.20	0.963



Gambar 1. Nilai AUC

Visualisasi pada Gambar 2 memperlihatkan bahwa nilai AUC meningkat seiring dengan penyetelan parameter hingga konfigurasi GBT-3, kemudian sedikit menurun pada konfigurasi

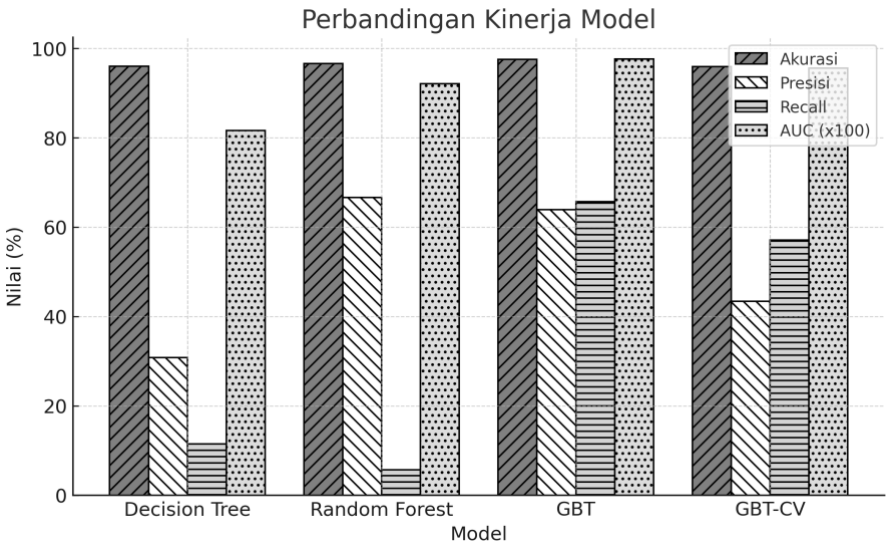
GBT-4. Fenomena ini sejalan dengan teori *bias-variance trade-off* dalam pembelajaran mesin: peningkatan kompleksitas model hanya efektif sampai titik tertentu, setelah itu model mulai *overfit* terhadap data latih [16]. Hasil optimasi ini juga mengonfirmasi bahwa pemilihan *learning_rate* 0.1 merupakan titik keseimbangan yang baik antara kecepatan pembelajaran dan kemampuan generalisasi, sebagaimana juga disarankan oleh Friedman [17] dalam teori dasar *gradient boosting machine*.

Untuk menilai peningkatan performa hasil optimasi parameter, dilakukan perbandingan antara tiga algoritma berbasis pohon keputusan, yaitu Decision Tree (DT), Random Forest (RF), dan Gradient Boosted Trees (GBT), termasuk varian GBT dengan Cross Validation (GBT-CV). Perbandingan mencakup empat metrik utama, yakni AUC, akurasi, presisi, dan recall sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Perbandingan Kinerja Model

Model	AUC	Akurasi	Presisi	Recall
Decision Tree	0.816	96.00%	30.77%	11.43%
Random Forest	0.921	96.60%	66.67%	5.71%
Gradient Boosted Trees	0.976	97.50%	63.89%	65.71%
GBT dengan Cross Validatio	0.956	95.89%	43.38%	57.14%

Perbandingan visual antar model ditunjukkan pada *Gambar 3*, yang memperlihatkan tren perbedaan nilai keempat metrik evaluasi.

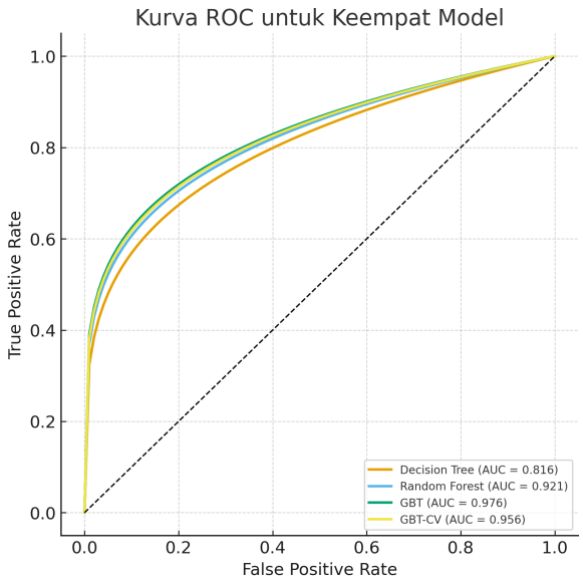


Gambar 3. Perbandingan Kinerja Model Berdasarkan Metrik Evaluasi

Berdasarkan hasil pada Tabel 2 dan Gambar 3, algoritma Gradient Boosted Trees (GBT) dengan parameter teroptimasi menghasilkan kinerja tertinggi pada hampir seluruh metrik evaluasi. Nilai AUC sebesar 0.976 dan akurasi 97.50% menunjukkan bahwa model ini mampu membedakan kelas positif dan negatif dengan sangat baik serta memberikan prediksi yang akurat. Nilai *recall* yang meningkat signifikan (65.71%) menunjukkan kemampuan model dalam mengenali lebih banyak nasabah potensial dibandingkan Decision Tree dan

Random Forest. Model GBT dengan Cross Validation (GBT-CV) menunjukkan hasil yang sedikit lebih rendah, dengan AUC 0.956 dan akurasi 95.89%, namun tetap mempertahankan nilai *recall* yang cukup tinggi (57.14%). Hal ini menunjukkan bahwa penerapan *cross-validation* menghasilkan estimasi performa yang lebih konservatif dan stabil terhadap variasi data pelatihan, sekaligus membantu mengidentifikasi potensi *overfitting* pada model GBT tanpa CV.

Sementara itu, model Random Forest memiliki nilai *presisi* tertinggi (66.67%) namun *recall* yang sangat rendah (5.71%), menandakan kecenderungan model yang lebih hati-hati dalam memprediksi kelas positif. Adapun Decision Tree sebagai model dasar memiliki kinerja terendah secara keseluruhan, yang mengonfirmasi bahwa model tunggal kurang stabil untuk data yang bersifat tidak seimbang. Untuk melengkapi analisis kinerja, *Gambar 4* memperlihatkan Kurva Receiver Operating Characteristic (ROC) dari keempat model yang diuji. ROC menggambarkan hubungan antara *True Positive Rate (TPR)* dan *False Positive Rate (FPR)* pada berbagai ambang keputusan, sehingga menunjukkan kemampuan diskriminatif setiap model.



Gambar 4. Kurva Receiver Operating Characteristic (ROC) untuk Keempat Model

Dari Gambar 4 dapat diamati bahwa kurva milik Gradient Boosted Trees (GBT) berada paling mendekati sudut kiri atas, menunjukkan performa klasifikasi terbaik dengan AUC tertinggi (0.976). Model GBT-CV memiliki kurva yang hampir serupa, dengan AUC 0.956, menandakan performa yang stabil dan generalisasi yang baik. Sementara Random Forest menempati posisi di bawah GBT, dan Decision Tree memiliki area terendah (AUC 0.816), yang mengindikasikan keterbatasan model sederhana dalam mengenali pola kompleks data perbankan. Konsistensi antara hasil Tabel 2 dan kurva ROC ini memperkuat bahwa optimasi parameter pada algoritma GBT secara efektif meningkatkan kemampuan diskriminatif model dibandingkan pendekatan *bagging* seperti Random Forest. Temuan ini juga mendukung hasil penelitian Chen dan Guestrin [18], yang menyatakan bahwa model *boosted trees* mampu memberikan hasil prediksi yang lebih unggul dan stabil pada dataset berskala besar dan tidak seimbang.

KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa optimasi parameter pada algoritma GBT secara signifikan meningkatkan kinerja model dalam memprediksi minat nasabah terhadap produk pinjaman berjangka. Berdasarkan hasil eksperimen, model GBT dengan parameter optimal mencapai AUC sebesar 0.976, akurasi 97.50%, presisi 63.89%, dan recall 65.71%, yang menunjukkan peningkatan kinerja dibandingkan Decision Tree dan Random Forest. Model GBT dengan Cross Validation (GBT-CV) menunjukkan nilai AUC sebesar 0.956 dan akurasi 95.89%, dengan nilai *recall* yang tetap tinggi (57.14%). Hasil ini menegaskan bahwa penerapan *cross-validation* menghasilkan estimasi performa yang lebih realistis dan stabil terhadap variasi data, sekaligus membantu mengurangi risiko *overfitting* yang mungkin terjadi pada model tanpa validasi silang. Untuk penelitian selanjutnya, disarankan penerapan metode optimasi adaptif seperti *Bayesian Optimization* atau *Genetic Algorithm*, serta eksplorasi integrasi model GBT dengan teknik penyeimbangan data seperti SMOTE atau ADASYN, guna meningkatkan sensitivitas terhadap kelas minoritas dan memperkuat performa prediksi di berbagai konteks data keuangan.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM) Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemdiktisaintek) atas dukungan pendanaan penelitian pada Tahun Anggaran 2024 berdasarkan kontrak nomor 113/E5/PG.02.00.PL/2024, 57/LL11/KM/2024, dan 01.07/43/KP/LPPM/UNU-KT/VI/2024. Dukungan tersebut sangat berperan dalam pelaksanaan dan penyelesaian penelitian ini. Penulis juga mengucapkan terima kasih kepada seluruh pihak yang telah memberikan kontribusi, masukan, dan dukungan selama proses penelitian hingga penulisan artikel ini.

DAFTAR PUSTAKA

[1] D. J. Hand, “Principles of data mining,” Drug Safety, vol. 30, no. 7, pp. 621–622, 2007.

[2] M. Moro, R. Laureano, and P. Cortez, “Using data mining for bank direct marketing: An application of the CRISP-DM methodology,” in Proc. European Simulation and Modelling Conf., 2011, pp. 117–121.

[3] S. Moro, P. Cortez, and P. Rita, “A data-driven approach to predict the success of bank telemarketing,” Decision Support Systems, vol. 62, pp. 22–31, 2014.

[4] H. Chen, R. H. L. Chiang, and V. C. Storey, “Business intelligence and analytics: From big data to big impact,” MIS Quarterly, vol. 36, no. 4, pp. 1165–1188, 2012.

[5] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2009.

[6] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” Frontiers in Neurorobotics, vol. 7, pp. 1–21, 2013.

[7] S. Raschka, “Model evaluation, model selection, and algorithm selection in machine learning,” arXiv preprint arXiv:1811.12808, 2018.

[8] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed. Springer, 2021.

[9] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[10] E. Vrigazova, "Gradient boosting decision trees: A powerful ensemble learning algorithm," International Journal of Data Science, vol. 5, no. 2, pp. 45–58, 2020.
[11] RapidMiner, RapidMiner Studio – Data Science Platform, version 10.1, RapidMiner GmbH, 2024. [Online]. Available: https://rapidminer.com
[12] C. Sammut and G. Webb, Encyclopedia of Machine Learning and Data Mining, Springer, 2017.
[13] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in Encyclopedia of Database Systems, Springer, 2009, pp. 532–538.
[14] L. Rokach, "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, no. 1–2, pp. 1–39, 2010.
[15] S. Zhang and Y. Ma, Ensemble Machine Learning: Methods and Applications, Springer, 2012.
[16] T. Dietterich, "Bias, variance and ensemble methods," Machine Learning, vol. 37, pp. 167–193, 1999.
[17] J. H. Friedman, "Stochastic gradient boosting," Computational Statistics & Data Analysis, vol. 38, pp. 367–378, 2002.
[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[19] Y. Zhou, "Improving ensemble learning for imbalanced data classification," Expert Systems with Applications, vol. 187, pp. 115–211, 2022.
[20] M. Moro, R. Laureano, and P. Cortez, "Predicting bank telemarketing success using data mining," Neurocomputing, vol. 84, pp. 3–15, 2012.
[21] M. Wibowo, A. P. Utomo, and D. P. Kurniadi, "Penerapan metode ensemble Gradient Boosting dalam prediksi minat nasabah terhadap deposito berjangka," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 10, no. 3, pp. 421–430, 2023.
[22] M. Wedel and W. A. Kamakura, Market Segmentation: Conceptual and Methodological Foundations, 2nd ed. Springer, 2012.
[23] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. O'Reilly Media, 2023.
[24] K. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," Advances in Neural Information Processing Systems (NeurIPS), vol. 31, 2018.
[25] L. Xu and J. Ma, "Handling imbalanced data in classification: A review," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 11, pp. 6102–6121, 2022.
[26] S. Maldonado, R. Weber, and F. Famili, "Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines," Information Sciences, vol. 286, pp. 228–246, 2014.