METODE *HYBRID DECISION TREE – ADAPTIVE BOOSTING*

(Studi Kasus: Klasifikasi *Credit Scoring*)

SKRIPSI

Diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar
Sarjana Matematika

Oleh:
Nurul Aini
NIM 2003693

PROGRAM STUDI MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA

2024

LEMBAR HAK CIPTA

METODE *HYBRID DECISION TREE – ADAPTIVE BOOSTING*
(Studi Kasus: Klasifikasi *Credit Scoring*)

Oleh:
Nurul Aini
2003693

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh
gelar Sarjana Matematika
pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia
Agustus 2024

LEMBAR PENGESAHAN

NURUL AINI

METODE *HYBRID DECISION TREE – ADAPTIVE BOOSTING*
(Studi Kasus: Klasifikasi *Credit Scoring*)

Disetujui dan disahkan oleh:
Pembimbing I,

Dr. Hj. Entit Puspita, S.Pd., M.Si.
NIP. 196704081994032002

Pembimbing II,

Hj. Dewi Rachmatin, S.Si., M.Si.
NIP. 196909291994122001

Mengetahui,
Ketua Prodi Matematika

Dr. Kartika Yulianti, S.Pd., M.Si.
NIP. 198207282005012001

ABSTRAK

Bank Indonesia mengungkapkan bahwa adanya indikasi peningkatan dalam penyaluran kredit baru oleh lembaga perbankan pada bulan Agustus 2023. Semakin tinggi jumlah transaksi kredit, maka akan semakin tinggi juga risiko kredit bermasalah. Oleh karena itu, perusahaan pemberi kredit harus lebih cermat dalam memilih calon peminjam yang berkualitas agar dapat mengurangi risiko kredit. Salah satu cara dalam mengurangi risiko kredit yaitu dengan melakukan *credit scoring*. *Credit scoring* merupakan suatu sistem penilaian risiko kredit yang banyak digunakan untuk membantu lembaga keuangan atau perusahaan pemberi kredit dalam mengevaluasi calon peminjam, baik individu ataupun perusahaan yang kemungkinan gagal melakukan pembayaran. Penelitian ini bertujuan untuk menentukan model terbaik dengan menghitung tingkat akurasi dari model *Decision Tree – AdaBoost* dan model *Logistic Regression – AdaBoost* dalam menentukan klasifikasi *credit scoring* pada perusahaan *Home Credit*. Berdasarkan hasil dari evaluasi model, model *Decision Tree – AdaBoost* menunjukkan performa terbaik dengan keseimbangan yang baik antara akurasi, *precision*, *recall*, *F1-Score*, dan *ROC-AUC*. Model ini berhasil mengungguli model *Logistic Regression – AdaBoost*. Tingkat akurasi model terbaik dari *Decision Tree – AdaBoost* dalam menentukan klasifikasi *credit scoring* pada perusahaan *Home Credit* yaitu sebesar 70% yang menunjukkan bahwa model *Decision Tree – AdaBoost* sudah cukup baik dalam menentukan klasifikasi *credit scoring*.

Kata Kunci: Klasifikasi, *Credit Scoring*, *Decision Tree*, *Logistic Regression*, *AdaBoost*

ABSTRACT

Bank Indonesia revealed indications of an increase in new credit disbursements by banking institutions in August 2023. As the number of credit transactions rises, the risk of problematic loans also increases. Therefore, credit providers must be more careful in selecting high-quality borrowers to reduce credit risk. One way to reduce credit risk is through credit scoring. Credit scoring is a widely used risk assessment system that helps financial institutions or credit providers evaluate potential borrowers, individuals, and companies, who may fail to repay their loans. This study aims to identify the best model by calculating the accuracy levels of the Decision Tree – AdaBoost model and the Logistic Regression – AdaBoost model in classifying credit scoring for Home Credit. Based on the model evaluation results, the Decision Tree – AdaBoost model demonstrated the best performance with a good balance between accuracy, precision, recall, F1-Score, and ROC-AUC. This model outperformed the Logistic Regression - AdaBoost model. The accuracy level of the best Decision Tree – AdaBoost model in classifying credit scoring for Home Credit is 70%, indicating that the Decision Tree – AdaBoost model is quite effective in determining credit scoring classifications.

Keywords: *Clasification, Credit Scoring, Decision Tree, Logistic Regression, AdaBoost*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

DAFTAR PUSTAKA

Abdullah, T., & Wahjusaputri, S. (2018). *Bank dan Lembaga Keuangan.* Jakarta: Mitra Wacana Media.

Aggarwal, C. C. (2015). *Data Classification Algorithm and Applications.* New York: CRC Press.

Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud Detection in Credit Cards using Logistic Regression. *International Journal of Advanced Computer Science and Applications (IJACSA), 11*(12). doi:https://doi.org/10.14569/ijacsa.2020.0111265

Ambarsita, L. (2013). Analisis Penanganan Kredit Macet. *Manajemen Bisnis, 3*(1).

Amin, R. K., Indwiarti, I., & Sibaroni, Y. (2015). Implementasi Klasifikasi Decision Tree Dengan Algoritma C4. 5 Dalam Pengambilan Keputusan Permohonan Kredit Oleh Debitur (Studi Kasus: Bank Pasar Daerah Istimewa Yogyakarta). *eProceeding of Engineering, 2*(1).

Barros, R. C., De Carvalho, A. C., & Freitas, A. A. (2015). *Automatic design of decision-tree induction algorithms.* Cham, Switzerland: Springer.

Bastos, J. A. (2022). Predicting Credit Scores with Boosted Decision Trees. *Forecasting, 4*, 925-935. doi:https://doi.org/10.3390/forecast4040050

Brzozowska, J., Pizon, J., Baytikenova, G., Gola, A., Zakimova, A., & Piotrowska, K. (2023). Data Engineering In Crisp-Dm Process Production Data – Case Study. *Applied Computer Science, 19*(3), 83-95. doi: 10.35784/acs-2023-26

Budianto, E. W. (2023). Pemetaan Penelitian Seputar Risiko Kredit pada Perbankan Syariah dan Konvensional: Studi Bibliometrik VOSviewer dan Literature Review. *BANCO: Jurnal Manajemen dan Perbankan Syariah, 5*(1), 20-34. doi:https://doi.org/10.35905/banco.v5i1.4987

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16*,321-357. doi:https://doi.org/10.1613/jair.953

Chopra, A., & Bhilare, P. (2018). Application of Ensemble Models in Credit Scoring Models. *Business Perspectives and Research, 6*(2), 129-141. doi:DOI: 10.1177/2278533718765531

Daniel, T. L., & Chantal, D. L. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining.* United Stases of America: John Wiley & Sons, Inc.

Djuarni, W., & Ratnasari, R. (2022). Implementasi Prinsip 5C Dalam Menentukan Kelayakan Pemberian Kredit Pada Nasabah. *Ar-Rihlah: Jurnal Keuangan dan Perbankan Syariah, 2*(2), 99-113.

Eprianti, N. (2019). Penerapan Prinsip 5C Terhadap Tingkat Non Performing Finance (NPF). *Amwaluna: Jurnal Ekonomi dan Keuangan Syariah, 3*(2), 252-266. doi:https://doi.org/10.29313/amwaluna.v3i2.4645

Firmanto, F. (2019). Penyelesaian Kredit Macet di Indonesia. *Jurnal Pahlawan, 2*(2). doi:https://doi.org/10.31004/jp.v2i2.577

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences, 55*(1), 119-139. doi:https://doi.org/10.1006/jcss.1997.1504

Garson, G. D. (2014). *Logistic Regression: Binary & Multinomial.* Asheboro: Statistical Associates Publishing.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics.* New York: McGraw-Hill/Irwin.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques Third Edition.* Waltham, MA: Morgan Kaufmann.

Herrhyanto, N., & Gantini, T. (2009). *Pengantar Statistika Matematis.* Bandung, Indonesia: Yrama Widya.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2019). *Introduction to Mathematical Statistics.* Boston MA: Pearson.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression.* Hoboken, New Jersey: John Wiley & Sons.

Jadhav, S. D., & Chane, H. P. (2016). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. *International Journal of Science and Research (IJSR), 5*(1). doi:https://doi.org/10.21275/v5i1.nov153131

Naufal, M. F., Subrata, Susanto, A. F., Kansil, C. N., & Huda, S. (2023). Analisis Perbandingan Algoritma Machine Learning untuk. *Techno.COM, 22*(1), 1-11. doi:http://publikasi.dinus.ac.id/index.php/technoc/article/view/7302

Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. *Augmented Human Research, 5*(12). doi:https://doi.org/10.1007/s41133-020-00032-0

Silva, E. C., Lopes, I. C., Correia, A., & Faria, S. (2020). A Logistic Regression Model for Consumer Default Risk. *Journal of Applied Statistics*. doi:https://doi.org/10.1080/02664763.2020.1759030

Sriwati, N. K. (2017). Analisis Tingkat Risiko Kredit Pada Bank Rakyat Indonesia Cabang Poso. *Jurnal EKOMEN, 17*(1).

Susilawati, & Putri, D. (2019). Analisis Pengaruh Transaksi Non Tunai Dan Suku Bunga Bi Terhadap Pertumbuhan Ekonomi Di Indonesia. *Jurnal Kajian Ekonomi dan Pembangunan, 1*(2), 667-678. doi:http://dx.doi.org/10.24036/jkep.v1i2.6294

Tian, Z., Xiao, J., Feng, H., & Wei, Y. (2020). Credit Risk Assessment based on Gradient Boosting Decision Tree. *Procdia Computer Science, 174*, 150-160. doi:https://doi.org/10.1016/j.procs.2020.06.070

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems, 26*, 61-68. doi:https://doi.org/10.1016/j.knosys.2011.06.020

Xiao, J., Wang, Y., Chen, J., Xie, L., & Huang, J. (2021). Impact of Resampling Methods and Classification Models on the Imbalanced Credit Scoring Problems. *Information Sciences, 569*, 508-526. doi:https://doi.org/10.1016/j.ins.2021.05.029

Zhang, X., & Chen, X. (2021). Research on breach prediction for big data through hybrid ensemble learning and logistic regression. *Journal of Physics: Conference Series, 1982*(2021). doi: https://doi.org/10.1088/1742-6596/1982/1/012049