BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Algoritma prediksi berbasis *machine learning* seperti *Random Forest*, KNN, *Naïve Bayes*, dan SVM telah berhasil diimplementasikan untuk memprediksi kemungkinan *dropout* mahasiswa. Dari hasil penelitian yang dilakukan, *Random Forest* terbukti paling efektif dalam memprediksi *dropout*, menunjukkan tingkat akurasi dan kemampuan diskriminatif yang tinggi terhadap kelas target. Algoritma ini efisien dalam menangani data yang tidak seimbang dan memiliki *robustness* terhadap variasi data, terutama ketika dilakukan *hyperparameter tuning* dan ketika menghadapi *noise* tambahan dalam dataset.

Tingkat akurasi algoritma prediksi dalam penerapan prediksi dropout mahasiswa sangat bervariasi tergantung pada jenis algoritma dan kondisi data. *Random Forest* menunjukkan akurasi sangat tinggi yaitu 0.99725 tanpa modifikasi, 0.99718 dengan *hyperparameter tuning* dan 0.99127 dengan penambahan *noise*. KNN dan *Naïve Bayes* juga menunjukkan peningkatan performa dengan *hyperparameter tuning*. Sementara itu, SVM, meskipun menunjukkan peningkatan dengan tuning, tetap memiliki performa yang lebih rendah dibandingkan algoritma lainnya. Ini menunjukkan bahwa beberapa algoritma lebih sensitif terhadap perubahan dalam data dan membutuhkan penyesuaian parameter yang lebih cermat untuk mencapai optimalisasi prediksi.

Analisis *feature importance* yang dilakukan menunjukkan bahwa variabel seperti keaktifan mahasiswa di perpustakaan, penggunaan LMS, poin mahasiswa, dan jumlah SKS semester 2 adalah faktor kunci dalam prediksi dropout. Penemuan ini memberikan landasan untuk intervensi yang lebih ditargetkan dan perencanaan akademik yang lebih efisien, sehingga potensi *dropout* dapat diminimalisir.

6.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran untuk penelitian selanjutnya adalah melakukan penelitian tentang:

1. Pengaruh variabel lain yang mungkin mempengaruhi *dropout*, seperti dukungan sosial, kesehatan mental, dan faktor ekonomi, untuk memperkaya dalam prediksi mahasiswa *dropout*.
2. Eksplorasi dan penerapan bentuk prediksi lain *seperti Deep Learning* atau *ensemble methods* untuk melihat apakah ada peningkatan dalam akurasi dan efektivitas.
3. Kemungkinan integrasi algoritma prediksi ke dalam sistem informasi mahasiswa yang ada untuk fasilitasi peringatan dini dan intervensi yang lebih efektif.

DAFTAR PUSTAKA

[1] Bashir Barthos, *Perguruan Tinggi Swasta di Indonesia: Proses Pendirian Penyelenggaraan & Dunia*.

[2] "Pangkalan Data Pendidikan Tinggi." Accessed: May 06, 2024. [Online]. Available: https://pddikti.kemdikbud.go.id/data_pt/OEZCMjVDN0MtQUZBQS00MDM4LTg5NDktODQ1RDU3NEZEQUVF

[3] T. Dharmawan, H. Ginardi, and A. Munif, "Dropout Detection Using Non-Academic Data," 2018.

[4] B. Pérez, C. Castellanos, and D. Correal, "Predicting student drop-out rates using data mining techniques: A case study," in *Communications in Computer and Information Science*, Springer Verlag, 2018, pp. 111–125. doi: 10.1007/978-3-030-03023-0_10.

[5] M. Vaarma and H. Li, "Predicting student dropouts with machine learning: An empirical study in Finnish higher education," *Technol Soc*, vol. 76, Mar. 2024, doi: 10.1016/j.techsoc.2024.102474.

[6] S. R. Sihare, "Student Dropout Analysis in Higher Education and Retention by Artificial Intelligence and Machine Learning," *SN Comput Sci*, vol. 5, no. 2, Feb. 2024, doi: 10.1007/s42979-023-02458-w.

[7] M. M. Tamada, R. Giusti, and J. F. de M. Netto, "Predicting Students at Risk of Dropout in Technical Course Using LMS Logs," *Electronics (Switzerland)*, vol. 11, no. 3, Feb. 2022, doi: 10.3390/electronics11030468.

[8] M. M. Rofi, F. A. Setiawan, and F. Riana, "Perbandingan Metode K-Nn Dan Random Forest Pada Klasifikasi Mahasiswa Berpotensi Dropout," *INFOTECH journal*, vol. 10, no. 1, pp. 84–89, Mar. 2024, doi: 10.31949/infotech.v10i1.8856.

[9] D. Andrade-Girón *et al.*, "Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 5, pp. 1–11, 2023, doi: 10.4108/eetsis.3586.

[10] M. Labib Mu'tashim *et al.*, "Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)," 2023.

[11] C. H. Cho, Y. W. Yu, and H. G. Kim, “A Study on Dropout Prediction for University Students Using Machine Learning,” *Applied Sciences*, vol. 13, no. 21, p. 12004, Nov. 2023, doi: 10.3390/app132112004.

[12] A. B. Urbina-Nájera and L. A. Méndez-Ortega, “Predictive Model for Taking Decision to Prevent University Dropout,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 4, pp. 205–213, 2022, doi: 10.9781/ijimai.2022.01.006.

[13] J. Kabathova and M. Drlik, “Towards predicting student’s dropout in university courses using different machine learning techniques,” *Applied Sciences (Switzerland)*, vol. 11, no. 7, Apr. 2021, doi: 10.3390/app11073130.

[14] R. Lottering, R. Hans, and M. Lall, “A Machine Learning Approach to Identifying Students at Risk of Dropout: A Case Study,” 2020. [Online]. Available: www.ijacsa.thesai.org

[15] A. A. Biswas, A. Majumder, M. J. Mia, I. Nowrin, and N. A. Ritu, “Predicting the enrollment and dropout of students in the post-graduation degree using machine learning classifier,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 3083–3088, Sep. 2019, doi: 10.35940/ijitee.K2435.0981119.

[16] F. Agrusti, G. Bonavolontà, and M. Mezzini, “University dropout prediction through educational data mining techniques: A systematic review,” *Journal of E-Learning and Knowledge Society*, vol. 15, no. 3, pp. 161–182, Oct. 2019, doi: 10.20368/1971-8829/1135017.

[17] “Berita Negara Republik Indonesia,” 2024. [Online]. Available: www.peraturan.go.id

[18] D. E. , & H. J. H. Goldberg, “Genetic algorithms and Machine Learning”, Accessed: May 05, 2024. [Online]. Available: file:///C:/Users/asus/Downloads/7951-22464-1-PB.pdf

[19] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.

[20] J. L. Berral, R. Gavaldà, and J. Torres, “Power-aware Multi-DataCenter Management using Machine Learning,” 2013.

[21] T. Hastie, R. Tibshirani, and J. Friedman, “Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction.”

[22] F. Rohmawati, G. Rohman, and S. Mujilahwati, “Sistem Prediksi Jumlah Pengunjung Wisata Wego Kec.Sugio Kab.Lamongan Menggunakan Metode Fuzzy Time Series,” 2017.

[23] A. Azlina Putri, “Resolusi : Rekayasa Teknik Informatika dan Informasi Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT. Central Brastagi Utama),” *Media Online)*, vol. 1, no. 6, pp. 354–361, 2021, [Online]. Available: https://djournals.com/resolusi

[24] D. Y. , A. R. , & H. M. A. Putri, “Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study : Information Systems Students of Telkom University,” 2018.

[25] T. K. Ho, “Random Decision Forest. Proceedings of the 3rd International Conference on Document Analysis and Recognition,” 1995.

[26] L. Breiman, “Random Forests,” 2001. doi: 10.1023/A:1010933404324.

[27] M. van Wezel and R. Potharst, “Improved customer choice predictions using ensemble methods,” *Eur J Oper Res*, vol. 181, no. 1, pp. 436–452, Aug. 2007, doi: 10.1016/j.ejor.2006.05.029.

[28] Bustami, “Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi,” 2014.

[29] T. M. Cover and P. E. Hart, “Approximate formulas for the information transmitted bv a discrete communication channel,” 1967.

[30] R. O. , H. P. E. , & S. D. G. Duda, “Pattern Classification,” 2001. [Online]. Available: https://www.researchgate.net/publication/228058014

[31] · N S Altman and N. S. Altman, “BU-1065MA An Introduction to Kernel and Nearest Neighbor Nonparametric Regression An Introduction to Kernel and Nearest Neighbor Nonparametric Regression,” 1991.

[32] D. W. Aha, D. Kibler, M. K. Albert, and J. R. Quinian, “Instance-Based Learning Algorithms,” 1991.

[33] L. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009, doi: 10.4249/scholarpedia.1883.

[34] C. Cortes, V. Vapnik, and L. Saitta, “Support-Vector Networks Editor,” Kluwer Academic Publishers, 1995.

[35] V. N. Vapnik, “An Overview of Statistical Learning Theory,” 1999.

[36] B. Schölkopf, A. S. Rsise, A. J. Smola, K. Algorithms, B. Schölkopf, and A. J. Smola, “Support Vector Machines and Kernel Algorithms,” 2002.

[37] J. Bergstra, J. B. Ca, and Y. B. Ca, “Random Search for Hyper-Parameter Optimization Yoshua Bengio,” 2012. [Online]. Available: http://scikit-learn.sourceforge.net.

[38] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” 1995. [Online]. Available: http//roboticsStanfordedu/"ronnyk

[39] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions,” 1974.

[40] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit Lett*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[41] D. M. W. Powers and Ailab, “Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness & Correlation,” 2011.

[42] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf Process Manag*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[43] A. P. B, “The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms Draft Only 2,” 1997.

[44] T. Fawcett, “ROC Graphs: Notes and Practical Considerations for Researchers,” 2004. [Online]. Available: http://www.purl.

[45] S. García, J. Luengo, and F. Herrera, “Intelligent Systems Reference Library 72 Data Preprocessing in Data Mining,” 2015. [Online]. Available: http://www.springer.com/series/8578

[46] D. Pyle, D. D. Cerra, and M. Kaufmann, “Data Preparation for Data Mining,” 1999.

[47] E. Rahm and H. H. Do, “Data Cleaning: Problems and Current Approaches,” 2000. [Online]. Available: http://dbs.uni-leipzig.de

[48] M. Kuhn and K. Johnson, “Applied Predictive Modeling,” 2013.

[49] M. Kuhn and K. Johnson, “Feature Engineering and Selection; A Practical Approach for Predictive Models; Edition 1,” 2013. doi: 10.4324/9781315108230.

[50] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” 2002.

[51] C. Bunkhumpornpat, E. Boonchieng, V. Chouvatut, and D. Lipsky, “FLEX-SMOTE: Synthetic over-sampling technique that flexibly adjusts to different minority class distributions,” *Patterns*, Nov. 2019, doi: 10.1016/j.patter.2024.101073.

[52] H. Han, W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning,” 2005.

[53] F. Grina, Z. Elouedi, and E. Lefevre, “Learning from Imbalanced Data Using an Evidential Undersampling-Based Ensemble,” pp. 235–248, 2022, doi: 10.1007/978-3-031-18843-5_16ï.

[54] K. R. Harker, J. Rowe, and C. Hargis, “Predicting Student Success with and without Library Instruction,” 2024.