

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan pada skripsi ini akan menjawab tujuan skripsi yang tertera pada Bab 1.3. Maka dari itu, kesimpulan dapat dijabarkan sebagai berikut:

- Pada skripsi ini, dijelaskan bagaimana menerapkan algoritma LightGBM untuk menyelesaikan permasalahan *fraud* klaim asuransi. Penyelesaian dilakukan dengan mencari dugaan awal, gradien, bobot *node* pohon, sampai ditemukan hasil iterasi terakhir dan mengubahnya kembali ke bentuk semula untuk menjadi penentuan *fraud* atau tidaknya sebuah klaim asuransi.
- Perbandingan performa model hasil algoritma LightGBM dengan model hasil algoritma XGBoost dilakukan dengan *dataset* klaim asuransi kendaraan Amerika Serikat dengan metode *5-fold CV*. Variabel-variabel *dataset* dan parameter kedua algoritma disamakan untuk melakukan perbandingan yang jelas.
- Secara umum, performa algoritma LightGBM ditemukan lebih baik yang terbukti dari ukuran-ukuran evaluasi (AUC, *accuracy*, *precision*, *recall*, dan *F1-score*) pada LightGBM lebih tinggi dibandingkan dengan XGBoost. Visualisasi dengan kurva AUC dan kurva *precision-recall* juga dilakukan untuk memperkuat bukti.

5.2 Saran

Ada beberapa hal-hal yang mungkin dapat dikembangkan dari skripsi ini, terutama pada penggunaan dan evaluasi data. Data yang digunakan di sini memiliki jumlah sedikit dibandingkan jumlah klaim asuransi yang dilakukan di dunia nyata. Ada beberapa variabel bebas yang harus dikaji lebih lanjut dengan analisis eksploratif yang lebih efisien untuk data dengan jumlah kolom relatif banyak untuk mengetahui hubungannya dengan terlapornya *fraud*. Maka dari itu, model yang dihasilkan dengan algoritma LightGBM belum tentu dapat mengklasifikasikan *fraud* atau tidaknya klaim asuransi pada dunia nyata secara tepat. Pengembangan dari algoritma seperti perubahan parameter harus dieksplorasi lebih lanjut, terutama dengan *dataset* yang lebih kompleks.

DAFTAR REFERENSI

- [1] Firozabadi, B. S., Tan, Y.-H., dan Lee, R. M. (1998) Formal definitions of fraud. *Norms, Logics and Information Systems-new Studies in Deontic Logic and Computer Science*, **3**, 275–288.
- [2] Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., Smith, G. M., Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., dan Smith, G. M. (2009) Limitations of linear regression applied on ecological data. *Mixed Effects Models and Extensions in Ecology with R*, **2**, 19–22.
- [3] Quinto, B. (2020) *Next-Generation Machine Learning with Spark: Covers XGBoost, LightGBM, Spark NLP, Distributed Deep Learning with Keras, and More*. Apress.
- [4] Zhang, D. dan Gong, Y. (2020) The comparison of lightgbm and xgboost coupling factor analysis and prediagnosis of acute liver failure. *IEEE Access*, **8**, 220990–221003.
- [5] Taha, A. A. dan Malebary, S. J. (2020) An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. *IEEE Access*, **8**, 25579–25587.
- [6] Asosiasi Asuransi Jiwa Indonesia (2022) Kebangkitan asuransi indonesia. <https://aaji.or.id/Articles/kebangkitan-asuransi-indonesia>.
- [7] Viaene, S. dan Dedene, G. (2004) Insurance fraud: Issues and challenges. *The Geneva Papers on Risk and Insurance-Issues and Practice*, **29**, 313–333.
- [8] Suthaharan, S. dan Suthaharan, S. (2016) *Decision Tree Learning*. Springer.
- [9] Breiman, L. (2017) *Classification and Regression Trees*. Routledge.
- [10] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An Introduction to Statistical Learning*. Springer.
- [11] Friedman, J. H. (2002) Stochastic gradient boosting. *Computational statistics & data analysis*, **38**, 367–378.
- [12] Friedman, J. H. (2001) Greedy function approximation: A gradient boosting machine. *Annals of statistics*, **29**, 1189–1232.
- [13] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., dan Liu, T.-Y. (2017) Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, **30**.
- [14] Machado, M. R., Karray, S., dan de Sousa, I. T. (2019) Lightgbm: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *2019 14th International Conference on Computer Science & Education (ICCSE)*, pp. 1111–1116. IEEE.
- [15] Lee, J.-S. (2019) Auc4. 5: Auc-based c4. 5 decision tree algorithm for imbalanced data classification. *IEEE Access*, **7**, 106034–106042.
- [16] Powers, D. M. (2020) Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness Correlation. *Journal of Machine Learning Technologies*, **2**, 37–63.

- [17] He, H. dan Ma, Y. (2013) *Imbalanced Learning: Foundations, Algorithms, and Applications*. John Wiley & Sons.
- [18] Maitra, S. (2019) Fraud-detection-insurance. https://github.com/saritmaitra/Fraud-detection--Insurance/blob/master/data/insurance_claims.csv.

