

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pembahasan yang telah dilakukan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan data asuransi kendaraan bermotor di Perancis, variabel yang paling berpengaruh secara signifikan terhadap prediksi frekuensi klaim di antaranya adalah *VehPower*, *DrivAge*, *VehGas*, dan *Region*.
2. Variabel terpenting (*Variable Importance*) bagi data asuransi kendaraan bermotor di Perancis dengan menggunakan model *XGBoost* adalah variabel *VehAge*, dengan nilai indikator yang lebih tinggi dibandingkan dengan variabel-variabel lainnya.
3. Model *XGBoost* merupakan model yang lebih baik dalam memprediksi frekuensi klaim pada data asuransi kendaraan bermotor di Perancis dibandingkan dengan model GLM Poisson, dengan menggunakan ukuran perbandingan deviansi Poisson dan *root mean squared error* (RMSE).

5.2 Saran

Berikut ini adalah beberapa saran yang dapat diajukan oleh penulis untuk pengembangan topik selanjutnya.

1. Penulis menyarankan menggunakan seluruh data yang terdapat pada *dataset* tersebut, yaitu dengan menggunakan variabel *BonusMalus* karena ada kemungkinan di masa mendatang bahwa Indonesia akan menerapkan prinsip layaknya *Bonus* dan *Malus*.
2. Prediksi frekuensi klaim dimodelkan menggunakan distribusi *EDM* lainnya, seperti contohnya adalah distribusi binomial, binomial negatif, dan *tweedie*.
3. Menggunakan teknik model *boosting* yang lain, seperti *gradient boost*, *adaboost*, dan *neural network*.
4. Penulis menyarankan untuk melakukan prediksi penentuan besar premi pada data asuransi kendaraan bermotor, dengan menambahkan dataset baru yang mendukung prediksi besar premi.
5. Penulis menyarankan agar terdapat hubungan antar variabel-variabel prediktor.
6. Menambahkan *dataset* baru yang berisi beberapa variabel lain untuk mempermudah prediksi frekuensi klaim.

DAFTAR REFERENSI

- [1] Shalev-Shwartz, S. dan Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Algorithms*, 1st edition. Cambridge University Press, New York.
- [2] Goldburd, M., Khare, A., Tevet, D., dan Guller, D. (2019) *Generalized Linear Models for Insurance Rating*, 2nd edition. Casualty actuarial society, Virginia.
- [3] Ohlsson, E. dan Johansson, B. (2010) *Non-Life Insurance Pricing with Generalized Linear Models*, 1st edition. Springer, Sweden.
- [4] de Jong, P. dan Heller, G. Z. (2008) *Generalized Linear Models for Insurance Data*, 1st edition. Cambridge University Press, Cambridge.
- [5] Wade, C. (2020) *Hands-On Gradient Boosting with XGBoost and scikit-learn*, 1st edition. Packt, Birmingham.
- [6] Ross, S. M. (2010) *A First Course In Probability*, 8th edition. Pearson Prentice Hall, California.
- [7] Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., dan Cochran, J. J. (2016) *Statistics for Business Economics*, 13th edition. Cengage Learning, Boston.
- [8] Grami, A. (2019) *Probability, Random Variables, Statistics, and Random Processes*, 1st edition. Wiley, New Jersey.
- [9] Stigler, S. (1981) Gauss and the invention of least squares. *The Annals of Statistics*, **9**, 465–474.
- [10] Samuel, A. L. (1959) Some studies in machine learning using the game of checkers. *IBM*, **36**, 210–229.
- [11] Lantz, B. (2013) *Machine Learning with R*, 1st edition. Packt Publishing, Birmingham.
- [12] Hastie, T., Tibshirani, R., dan Friedman, J. (2009) *The Elements of Statistical Learning*, 2nd edition. Springer, New York.
- [13] Strang, G. dan Herman, E. (2016) *Calculus Volume 2*, 1st edition. OpenStax, Rice University, Texas.
- [14] Chen, T. dan Guestrin, C. (2016) Xgboost: A scalable tree boosting system. *Knowledge Discovery and Data Mining, 22nd ACM SIGKDD Conference*, **1**, 785–794.
- [15] Pesantez-Narvaez, J. dan Guillen, M. (2019) Predicting motor insurance claims using telematics data - xgboost vs. logistic regression. *MDPI*, **7**, 1–16.
- [16] Dutang, C. dan Charpentier, A. (2016) Casdatasets: Insurance datasets. *CASDatasets with R*, **1**, 62–64.