

# PREDICTIVE MODELING FOR CLAIMS IN AUTOMOBILE INSURANCE

Monica Andreea MARCIUC<sup>a\*</sup>

<sup>a)</sup> Babeş-Bolyai University, Faculty of Economics and Business Administration,  
Cluj-Napoca, Romania

Please cite this article as:

Article History:

Marciuc, M.A., 2024. Predictive modeling for claims in automobile insurance, *Review of Economic Studies and Research Virgil Madgearu*, 17(2), pp.79-99. doi: 10.24193/RVM.2024.17.118.

Received: 25 October 2024  
Accepted: 15 November 2024

---

**Abstract:** *The rise of advanced machine learning methods has revolutionized the landscape of predictive modeling in the automobile insurance sector. This paper presents the relevant literature review on the use of machine learning methods, including gradient boosting, random forests, and decision trees, to model claims in automobile insurance. By synthesizing findings from key studies, we conclude on the predictive performance of these methods compared to traditional actuarial models and identify emerging trends and challenges in this domain. Our analysis highlights how data-driven approaches enhance pricing accuracy, optimize risk assessment, and improve operational efficiency. Furthermore, the paper addresses critical issues such as model interpretability, fairness, and ethical considerations in adopting machine learning technologies. This literature review contributes to the ongoing discourse on improving automobile insurance practices through predictive analytics and provides a foundation for future research.*

**Key words:** machine learning; GLM model; claim; automobile insurance

**JEL Classification:** C87; C88

---

---

\* Corresponding author. E-mail address: monica.marciuc@econ.ubbcluj.ro.

## **References:**

1. Blier-Wong, C., Cossette, H., Lamontagne, L., Marceau, É., 2021. Machine Learning in P&C Insurance: A Review for Pricing and Reserving. *Risks*, 9(1), p. 4. <https://doi.org/10.3390/risks9010004>.
2. Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5-32.
3. Clemente, C., Guerreiro, G.R., Bravo, J.M., 2023. Modelling Motor Insurance Claim Frequency and Severity Using Gradient Boosting. *Risks*, 11(9), p. 163. <https://doi.org/10.3390/risks11090163>.
4. Denuit, M., Guillen, M., Trufin, J., 2018. Multivariate Credibility Modeling for Usage-Based Motor Insurance Pricing with Behavioral Data. *DetraNote*, 2018-2.
5. Diana, A., Griffin, J.E., Oberoi, J., Yao, J., 2019. Machine-Learning Methods for Insurance Applications: A Survey. Society of Actuaries Research Report.
6. Dugas, C., Bengio, Y., Chapados, N., Vincent, P., Denoncourt, G., Fournier, C., 2003. Statistical Learning Algorithms Applied to Automobile Insurance Ratemaking. *Casualty Actuarial Society Forum*, Winter 2003, pp. 179-200.
7. Goodfellow, I., Bengio, Y., Courville, A., 2014. *Deep Learning*, MIT Press.
8. Grize, Y.-L., Fischer, W., Lützelshwab, C., 2020. Machine learning applications in nonlife insurance. *Applied Stochastic Models in Business and Industry*, 36(4), pp. 523-537. <https://doi.org/10.1002/asmb.2543>.
9. Guelman, L., 2012. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, 39(3), pp. 3659-3667. <https://doi.org/10.1016/j.eswa.2011.09.058>.
10. Guillen, M., Nielsen, J.P., Pérez-Marín, A.M., Elpidorou, V., 2019. Can automobile insurance telematics predict the risk of near-miss events? *North American Actuarial Journal*, 23(2), pp. 220-235. <https://doi.org/10.1080/10920277.2019.1627221>.
11. Hanafy, M. and Ming, R., 2021. Machine Learning Approaches for Auto Insurance Big Data. *Risks*, 9(2), p. 42. <https://doi.org/10.3390/risks9020042>.
12. Henckaerts, R., Côté, M.-P., Antonio, K., Verbelen, R., 2021. Boosting insights in insurance tariff plans with tree-based machine learning

- methods. *North American Actuarial Journal*, 25(2), pp. 255-285. <https://doi.org/10.1080/10920277.2020.1745656>.
13. Kaščelan, V., Kaščelan, Lj., Novović Burić, M., 2016. A nonparametric data mining approach for risk prediction in car insurance: a case study from the Montenegrin market. *Economic Research-Ekonomska Istraživanja*, 29(1), pp. 545-558. <https://doi.org/10.1080/1331677X.2016.1175729>.
  14. Liu, Y., Wang, B.J., Lv, S.G., 2014. Using Multi-class AdaBoost Tree for Prediction Frequency of Auto Insurance. *Journal of Applied Finance & Banking*, 4(5), pp. 45-53.
  15. McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*, 2nd ed. Chapman & Hall/CRC.
  16. Noll, A., Salzmann, R. and Wüthrich, M.V., 2020. Case Study: French Motor Third-Party Liability Claims. SSRN Electronic Journal.
  17. Pesantez-Narvaez, J., Guillén, M., Alcañiz, M., 2019. Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks*, 7(2), p. 70. <https://doi.org/10.3390/risks7020070>.
  18. Poufnas, T., Gogas, P., Papadimitriou, T., Zaganidis, E., 2023. Machine Learning in Forecasting Motor Insurance Claims. *Risks*, 11(9), p. 164. <https://doi.org/10.3390/risks11090164>.
  19. Quan, Z. and Valdez, E.A., 2018. Predictive Analytics of Insurance Claims Using Multivariate Decision Trees. SSRN Electronic Journal.
  20. Staudt, Y. and Wagner, J., 2021. Assessing the performance of random forests for modeling claim severity in collision car insurance. *Risks*, 9(3), p. 53. <https://doi.org/10.3390/risks9030053>.
  21. Su, X. and Bai, M., 2020. Stochastic gradient boosting frequency-severity model of insurance claims. *PLOS ONE*, 15(8), p. e0238000. <https://doi.org/10.1371/journal.pone.0238000>
  22. Yang, Y., Qian, W., Zou, H., 2016. Insurance Premium Prediction via Gradient Tree-Boosted Tweedie Compound Poisson Models. arXiv preprint arXiv:1508.06378.