**Proposal on a Comparative Analysis of Linear Regression and Random Forest Machine Learning Algorithms for Automated Car Insurance Quoting**

|  |  |
| --- | --- |
| **Oguzhan Ozbek**  **Student of Applied Science & Technology**  **Sheridan College**  **ozbeko@sheridancollege.ca** | **Raghav Dheer**  **Student of Applied Science & Technology**  **Sheridan College**  **dheerr@sheridancollege.ca** |

**I. Research Objective**

The objective of this project is to conduct a comparative analysis of two widely used machine learning algorithms, Linear Regression and Random Forest in the context of automated car insurance quoting. This application domain is particularly relevant as the insurance industry increasingly adopts machine learning techniques to streamline operations and improve pricing accuracy. The goal of this research is to evaluate the effectiveness of these algorithms in predicting car insurance premiums, focusing on their accuracy, interpretability, and ability to handle real-world data.

Through this analysis, we aim to identify which algorithm offers better performance and suitability for the task, with the intention of providing actionable insights for automating and improving the car insurance quoting process.

**II. Expected Outcomes and Impact**

The expected outcomes of this project are twofold:

1. Comparative Performance: We anticipate that Random Forest will outperform Linear Regression in terms of predictive accuracy, given its ability to capture complex, non-linear relationships in the data. However, we also expect that Linear Regression will be more transparent and easier to interpret, with clear relationships between the features and the target variable.
2. Practical Insights for Car Insurance Quoting: This research aims to provide insights into the trade-offs between model complexity and interpretability in automated car insurance quoting systems. If Random Forest shows a significant improvement in accuracy, it could be a valuable tool for insurers looking to optimize their quoting processes. On the other hand, if Linear Regression performs comparably, it may offer a simpler and more transparent alternative that is easier to deploy and explain to customers.

Ultimately, this study will contribute to a deeper understanding of how machine learning algorithms can be applied to real-world problems like car insurance pricing, offering practical recommendations for selecting the most suitable algorithm based on the specific needs of the industry.

**References**

Pamba, F. (2019). Application of Machine Learning for Estimating Motor Vehicle Insurance Premium. KCA University Institutional Repository. Retrieved April 1, 2025, from <https://repository.kcau.ac.ke/handle/123456789/484>

Guo, Shuang. (2024). Features to Involve and Collect in Future Auto Insurance Pricing. Advances in Economics, Management and Political Sciences, 113, 17-24. doi: <https://doi.org/10.54254/2754-1169/113/2024LD0103>

Reil, J.P.C. (2024). Beyond Generalized Linear Models: Advancing Insurance Pricing through Interpretable and Explainable Machine Learning. University of Twente. Retrieved April 1, 2025, from <https://purl.utwente.nl/essays/103673>

Bekkaye, C., & Zari, T. (2023). The use of Machine Learning in non-life insurance: Literature review. International Journal of Accounting, Finance, Auditing, Management and Economics, 4(2-1), 307-319. <https://doi.org/10.5281/zenodo.7827163>

Jonkheijm, T. (2023). Forecasting Insurance Claim Amounts in the Private Automobile Industry Using Machine Learning Algorithms. Tilburg University. Retrieved April 1, 2025, from <https://arno.uvt.nl/show.cgi?fid=173524>

Komperla, R.C.A. (2022). Car Insurance Risk Prediction Using Artificial Intelligence and Machine Learning. Global Journal of Advanced Engineering Technologies and Sciences, 9(4), 1-8. doi: <https://doi.org/10.29121/gjaets.2022.04.01>

Spedicato, G.A., Dutang C., & Petrini, L. Machine Learning Methods to Perform Pricing Optimization. A Comparison with Standard GLMs. Variance, 2018, 12(1), 69-89. Retrieved April 1, 2025, from <https://hal.science/hal-01942038v2>

…?

…?

…?