

## INTISARI

### PEMODELAN PREMI MURNI ASURANSI KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA *GRADIENT BOOSTING MACHINE*, *EXTREME GRADIENT BOOSTING*, DAN *DUAL-PARAMETER EXTREME* *GRADIENT BOOSTING*

Oleh

Ukhti Erisawati

20/455437/PA/19652

Perkembangan data telematika, khususnya jarak tempuh, membuka peluang baru dalam asuransi kendaraan bermotor berbasis penggunaan, seperti *Pay-As-You-Drive Insurance* yang memberikan pemahaman lebih mendalam terkait perilaku pengendara dan hubungannya dengan faktor risiko. Saat ini, sistem perhitungan premi masih dinilai kurang adil dan relevan karena didasarkan pada faktor-faktor yang memiliki korelasi tidak langsung dengan risiko. Penelitian ini berfokus pada pemodelan premi murni berbasis jarak tempuh tahunan dan faktor-faktor tradisional lainnya dengan memanfaatkan algoritma pembelajaran mesin dan distribusi aktuarial klasik. Algoritma *Gradient Boosting Machine* (GBM) dan *Extreme Gradient Boosting* (XGBoost) diterapkan untuk membangun model prediksi yang secara umum menunjukkan performa lebih baik dibandingkan *Generalized Linear Models* (GLM). Selain itu, model *dual-parameter* XGBoost yang mengakomodasi distribusi *Zero-Inflated Poisson* (ZIP) dibangun untuk menangani distribusi klaim dengan nilai nol berlebih yang kerap menjadi tantangan di industri asuransi.

Hasil penelitian menunjukkan bahwa *machine learning* dapat mengungguli GLM. Selain itu, diperoleh bahwa model GLM ZIP dan XGBoost ZIP menunjukkan tantangan dalam mengatasi distribusi klaim dengan nilai nol berlebih. Hal ini tercermin dari nilai *deviance* yang kecil dan RMSE yang relatif tinggi. Meskipun demikian, pendekatan ini tetap menawarkan potensi untuk diterapkan pada berbagai skenario dengan penyesuaian lebih lanjut. Pemodelan premi didasarkan pada model XGBoost Poisson untuk frekuensi dan XGBoost Gamma untuk *severity* yang memberikan performa terbaik berdasarkan nilai RMSE. Berdasarkan hasil analisis, diperoleh bahwa jarak tempuh tahunan merupakan variabel yang paling signifikan baik dalam frekuensi maupun *severity*.

## **ABSTRACT**

### **MODELING AUTO INSURANCE PURE PREMIUM USING GRADIENT BOOSTING MACHINE, EXTREME GRADIENT BOOSTING, AND DUAL-PARAMETER EXTREME GRADIENT BOOSTING ALGORITHM**

By

Ukhti Erisawati

20/455437/PA/19652

The development of telematics data, particularly mileage, has opened new opportunities in usage-based motor vehicle insurance, such as Pay-As-You-Drive Insurance, which provides better insights into driver behavior and its relationship to risk factors. Currently, premium calculation systems are considered less fair and relevant as they rely on factors that have an indirect correlation with risk. This study focuses on modeling pure premiums based on annual mileage and other traditional factors by utilizing machine learning algorithms and classical actuarial distributions. The Gradient Boosting Machine (GBM) and Extreme Gradient Boosting (XGBoost) algorithms were applied to build predictive models, which generally demonstrated better performance compared to Generalized Linear Models (GLM). Additionally, a dual-parameter XGBoost model accommodating the Zero-Inflated Poisson (ZIP) distribution was developed to address the challenge of excess zero claims which often encountered in the insurance industry.

The results indicate that machine learning can outperform GLM. Furthermore, both the GLM ZIP and XGBoost ZIP models encountered challenges in managing the excess zero claim distribution, as reflected by low deviance values and relatively high RMSE. Nevertheless, these approaches still offer potential for application in various scenarios with further adjustments. Premium modeling was based on the XGBoost Poisson for frequency and the XGBoost Gamma for severity, both of which provided the best performance based on RMSE. It was found that annual mileage is the most significant variable in both frequency and severity models.