

INTISARI

ANALISIS KLASIFIKASI DENGAN METODE *RANDOM FOREST*, LOGITBOOST, DAN XGBOOST UNTUK MEMPREDIKSI STATUS KLAIM ASURANSI

Oleh

Ester Nadia Emor
19/440785/PA/19105

Perusahaan asuransi di Indonesia terus mengalami pertumbuhan yang ditandai dengan jumlah pemegang polis mengalami kenaikan setiap tahunnya. Hal tersebut menggambarkan data dalam perusahaan asuransi yang semakin besar. Data yang sangat besar pada perusahaan asuransi akan membutuhkan waktu yang panjang apabila diproses secara manual oleh tenaga manusia, sehingga dibutuhkan bantuan teknologi komputasi. Seiring dengan perkembangan teknologi, pendekatan *machine learning* sering digunakan untuk menganalisis data dalam jumlah yang besar dan membantu memecahkan permasalahan dalam perusahaan asuransi. Salah satu permasalahan yang dapat dibantu oleh teknologi *machine learning* adalah penilaian klaim asuransi. Dengan bantuan *machine learning*, waktu proses penilaian klaim dapat menjadi lebih singkat.

Pada penelitian ini, analisis klasifikasi yang merupakan salah satu metode *machine learning* digunakan untuk membuat prediksi status klaim pada perusahaan asuransi kesehatan. Beberapa metode yang digunakan adalah metode klasik seperti *Random Forest* serta metode *boosting* seperti LogitBoost dan XGBoost. Berdasarkan hasil analisis, diperoleh bahwa metode *boosting*, secara khusus model XGBoost memiliki performa yang paling baik dalam melakukan klasifikasi untuk memprediksi status klaim dibandingkan dengan model lainnya, ditandai dengan metrik *accuracy*, *precision*, *F1-score*, dan *F2-score* yang paling tinggi. Model XGBoost memperhatikan kelas positif dan kelas negatif dengan sama baiknya walaupun data bersifat tidak seimbang. Dalam kasus ini, kelas positif merupakan klaim ditolak dan kelas negatif merupakan klaim diterima. Model yang diperoleh kemudian digunakan untuk memberikan rekomendasi status klaim.

Kata kunci: Pertumbuhan perusahaan asuransi, klaim asuransi, penilaian klaim, status klaim, *machine learning*, analisis klasifikasi, *Random Forest*, LogitBoost, XGBoost.

ABSTRACT

CLASSIFICATION ANALYSIS WITH RANDOM FOREST, XGBOOST, AND LOGITBOOST TO PREDICT INSURANCE CLAIM STATUS

By

Ester Nadia Emor
19/440785/PA/19105

Insurance companies in Indonesia continue to experience growth as the number of policyholders increases every year. This reflects the growing size of data in insurance companies. The huge data in insurance companies will consume a long time if processed manually by human work, thus requiring the assistance of computational technology. As the technology continues to evolve, machine learning approaches are often used to analyze large amounts of data and help solve problems in insurance companies. One of the problems that can be assisted by machine learning technology is insurance claim adjudication. With machine learning, the claim adjudication processing time can be reduced.

In this research, classification analysis which is one of the machine learning methods is used to make predictions of claim status in a health insurance company. Some of the methods used are classic methods such as Random Forest and boosting methods such as LogitBoost and XGBoost. Based on the analysis results, it is found that boosting methods, specifically the XGBoost model, has the best performance in performing classification to predict claim status compared to other models, indicated by having the highest accuracy, precision, F1-score, and F2-score metrics. The XGBoost model considers both positive and negative classes equally well despite the imbalanced data. In this case, positive classes are denied claims and negative classes are approved claims. The model obtained is then used to provide recommendations for claim status.

Keyword: Insurance growth, insurance claim, claim adjudication, claim status, machine learning, classification analysis, Random Forest, LogitBoost, XGBoost.