

INTISARI

APLIKASI *RANDOM FOREST* DAN ADABOOST PADA PREDIKSI BESAR KLAIM ASURANSI KESEHATAN

Oleh

Salsabila Ari Tridianti

20/459255/PA/19916

Perusahaan asuransi kesehatan memiliki kewajiban untuk membayar klaim apabila pemegang polis mengalami kerugian keuangan setelah melakukan perawatan kesehatan. Besar klaim asuransi kesehatan dapat dipengaruhi oleh beberapa faktor seperti jenis kelamin, usia peserta, jenis fasilitas kesehatan rujukan, tingkat pelayanan, segmentasi peserta, kelas rawat, kode ICD, tarif regional dan lama kunjungan. Oleh sebab itu, diperlukan metode yang tepat dalam memprediksi besar klaim asuransi kesehatan. Metode MLT (Model Linear Tergeneralisasi) merupakan metode yang paling sering digunakan. Namun, metode MLT memiliki kelemahan yaitu tidak dapat menangani data yang memiliki sifat non-linear, sehingga metode *machine learning* digunakan sebagai alternatif untuk menangani masalah tersebut. Metode *machine learning* yang digunakan untuk memprediksi besar klaim adalah metode *Random Forest* dan AdaBoost karena metode tersebut memiliki kemampuan menangani data non-linear serta memiliki kemampuan untuk menangani dataset yang besar dan kompleks. Hasil simulasi dengan menggunakan metode *machine learning* ini kemudian dibandingkan dengan metode MLT. Perbandingan dilakukan dengan cara mengevaluasi kinerja model dengan menggunakan RMSE, MAE dan MAPE. Pada penelitian ini dilakukan prediksi besar biaya klaim asuransi kesehatan untuk penderita Tuberkulosis (TB) dengan data yang digunakan adalah data sampel asuransi kesehatan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan tahun 2021-2022. Dengan menggunakan metode MLT, *Random Forest* dan AdaBoost, diperoleh bahwa metode *Random Forest* memiliki nilai RMSE, MAE dan MAPE yang lebih rendah dari model yang menggunakan metode lain. Dengan kata lain, metode *Random Forest* menghasilkan prediksi besar klaim yang lebih baik dibandingkan dengan metode MLT dan AdaBoost.

ABSTRACT

APPLICATION OF RANDOM FOREST AND ADABOOST IN PREDICTING THE COST OF HEALTH INSURANCE CLAIMS

By

Salsabila Ari Tridianti

20/459255/PA/19916

Health insurance companies must pay claims if the policyholder experiences financial losses after undergoing health care. The amount of health insurance claims can be influenced by factors such as gender, age of participants, type of referral health facility, level of service, participant segmentation, treatment class, ICD code, regional rates, and length of visit. Therefore, an appropriate method is needed to predict the amount of health insurance claims. The GLM (Generalized Linear Model) method is the most commonly used method. However, the GLM method has a weakness, namely that it cannot handle data that has non-linear properties, so the machine learning method is used as an alternative to deal with this problem. The machine learning methods used to predict the cost of claims are the Random Forest and AdaBoost methods because these methods can handle non-linear data and large and complex datasets. The simulation results using this machine learning method are then compared with the GLM method. The comparison is made by evaluating the performance of the model using RMSE, MAE, and MAPE. In this study, a prediction of the cost of health insurance claims for Tuberculosis (TB) sufferers was carried out with the data used being sample data from the Asuransi Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan in 2021-2022. By using GLM, Random Forest, and AdaBoost methods, it is obtained that the Random Forest method has lower RMSE, MAE, and MAPE values than models using other methods. In other words, the Random Forest method produces better large claim predictions compared to the GLM and AdaBoost methods.