

Analisis akurasi model random forest untuk big data-studi kasus prediksi klaim severity pada asuransi mobil = Analyzing accuracy of random forest model for big data-A case study of claim severity prediction in car insurance

Karina Chandra Dewi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20493045&lokasi=lokal>

Abstrak

Klaim asuransi merupakan salah satu elemen penting dalam bidang jasa asuransi. Klaim severity mengacu pada besarnya dana yang harus dikeluarkan untuk memperbaiki kerusakan yang terjadi. Besarnya klaim asuransi dipengaruhi oleh banyak faktor. Hal ini menyebabkan volume data menjadi sangat besar. Sehingga diperlukan suatu metode yang tepat dalam memprediksi besarnya klaim severity untuk data besar. Salah satu metode yang dapat digunakan untuk menyelesaikan permasalahan tersebut yaitu Random Forest yang merupakan salah satu metode machine learning. Tesis ini mengaplikasikan model Random Forest untuk menyelesaikan masalah prediksi besarnya klaim severity pada asuransi mobil serta menganalisis pengaruh jumlah fitur yang digunakan pada model Random Forest terhadap akurasi model sebagai alternatif solusi terkait Big Data. Hasil simulasi menunjukkan bahwa model Random Forest dapat diterapkan pada kasus prediksi klaim severity yang merupakan kasus regresi dalam konteks machine learning. Dengan menggunakan 1/3 dari keseluruhan fitur yang ada, model Random Forest dapat menghasilkan akurasi yang setara dengan akurasi yang diperoleh ketika menggunakan seluruh fitur dalam membangun model, yaitu sekitar 99%. Hasil ini menunjukkan skalabilitas yang baik dari Random Forest terutama ditinjau dari jumlah fitur. Sehingga, model Random Forest dapat digunakan sebagai solusi untuk masalah Big Data terkait volume data.

<hr>

The insurance claim is one of the important elements in the field of insurance services. Claim severity refers to the amount of fund that must be spent to repair the damage. The amount of insurance claim is influenced by many factors. This causes the volume of data to be very large. Therefore, a suitable method is required. Random Forest, one of the machine learning methods can be implemented to handle this problem. This thesis applies the Random Forest model to predict the amount of this claim severity on car insurance. Furthermore, analysis on the effect of the number of features used on model accuracy is conducted. The simulation result show that the Random Forest model can be applied in cases of prediction of claim severity which is a case of regression in the context of machine learning. Only by using 1/3 of the overall features, the accuracy of Random Forest model can produce accuracy that is comparable to that obtained when using all features, which is around 99%. This result confirms the scalability of Random Forest, especially in terms of the number of features. Hence, Random Forest model can be used as a solution to Big Data problems related to data volume.