Predicting the Health Insurance Premium by Analyzing the Customer Feature Importance with Random Forest

Hendra Achmadi, Kim Sun Suk, Isana Meranga, Golrida Karyawati, Sylvia Samuel

Fakultas Ekonomi dan Bisnis Universitas Pelita Harapan, Indonesia

ABSTRACT

This research aims to predict the Heath care Insurance Premi by analyzing the customer characteristics Random Forest Algorithm. Secondly, this research will determine feature importance from customer characteristics that drive the closing decision. This research methodology is quantitative, and data mining methodology and data will be derived from primary data, and 202 data will be processed and cleaned into 148 data. This research also uses supervisory learning and using the random forest algorithm. The novelty of this research, Firstly, the health care insurance premium can be predicted by the machine learning random forest algorithm. Secondly, the sequence of importance from the customer characteristics can be determined by the feature of importance function at the random forest. A random forest algorithm can predict the prospect data test with an accurate value of 63,33 %. First, the crucial question that has to be made is how much the income is; second is age, is between 45-55 years old, and the third is how many dependent. The decision tree graph helps construct the questioner's level, leading to a monthly target premium of 3 million to 5 million.

Keywords: Health care Insurance Premi Prediction, Random forest Algorithm, The feature Importance.