MSc Project Report

PREMIUM PREDICTION USING RISK ASSESSMENT TO GENERATE SMART CONTRACTS FOR THE HEALTH INSURANCE SECTOR

Kanishka Lahiru Wijendra

2021

A report submitted as part of the requirements for the degree of MSc Big Data Analytics at Robert Gordon University, Aberdeen, Scotland

Abstract

Insurance is a mode of transferring the risk and it comprises of several products. Health insurance sector being the foundation of this study, has various activities that are integrated in its process. Many of those being manual activities with lengthy processes, often leads to inefficiencies. Through this study, a basic solution will be developed to overcome the inefficiencies in the health insurance process. An automation of this process will be developed as a solution. Smart contract technology of the blockchain paradigm will be used in this automation due to its characteristic of immutability of transactions. The decentralized framework of blockchain networks will allow the solution to be secured. The smart contracts will be generated to replace the traditional health insurance policies with this application. Replacement of traditional health insurance policies with an automated approach will help in enabling faster insurance claim processes and avoid inconsistencies. Process automation will require the insurance premium to be predicted prior to generating a smart contract. Several predictive algorithms will be evaluated to arrive at best fit model. Regression approaches such as multiple linear regression, lasso regression, ridge regression, regression tree method and gradient boosting will be evaluated through this study. Fitted models will be evaluated using evaluation criteria such as AIC, RMSE, and adjusted R squared. Further the best fit model will be validated using K-fold cross validation approach. This study considers gradient boosting as the best fit model for predicting health insurance premium with an accuracy of 83.7% obtained through K-fold cross validation. Being a study that is focused in helping to overcome existing issues in the domain of health insurance, this will be a value addition for the insurance industry.

Keywords: blockchain, smart contract, gradient boosting, multiple linear regression, lasso and ridge regression, regression tree method, k-fold cross validation