AY2021 CBA Homework Assignment

ML Generated Insurance Premium

Introduction

An insurance company wants to use Machine Learning to generate insurance premium payable and propose the insurance offer to the client. This will shorten the sales process and eliminate effort and time to process the request. Currently, the sales cycle relies on insurance agent, actuaries, admin staff, and take too much time and effort.

To evaluate the feasibility, a sample of hospitalization insurance premium and client profile data (premium2.csv) was handled to you. A data dictionary is provided in Appendix A.

1. Create the BMI variable based on CDC definition[1]. Show your code.

2. There are many categorical variables with integer coded values (e.g. Diabetes, HighBloodPressure, Transplant…etc.) Is it necessary to convert them to factor datatype in R?

3. Explore the data and report on your key findings.

4. Using 1 SE optimal CART and one other technique learnt in this course:

   a. What is the 10-fold cross validation RMSE and number of splits in the 1SE Optimal CART?

   b. Identify the key predictors of premium.

   c. Is BMI or Gender important in determining premium?

   d. Evaluate and compare the predictive accuracy of the two techniques on a 70-30 train-test split. Present testset RMSE results in a table.

5. Explain the limitations of your analysis. [Max 1 page.]

6. Is CART successful in this application? Explain. [Max 1 page.]

---

[1] https://www.cdc.gov/healthyweight/assessing/index.html

Age: Age of the client (years).

Diabetes: Presence (1) or Absence (0) of the disease.

HighBloodPressure: Presence (1) or Absence (0) of the condition.

Transplant: Organ Transplant Surgery Ever Done (1: Yes, 0: No).

ChronicDisease: Other Chronic Disease besides Diabetes or High Blood
                            Pressure (1: Yes, 0: No).

Height: Height (cm).

Weight: Weight (kg).

Allergy: Known Allergy (1: Yes, 0: No).

CancerInFamily: Does any family member had/have Cancer (1: Yes, 0: No).

NumMajorSurgeries: Number of Major Surgeries Done, excluding Organ
                            Transplant.

Gender: (1: Male, 0: Female).

Premium: Annual Premium Payable by client (Singapore Dollars)