Health Insurance Premium Prediction

By: Thandava Krishna M

Date: 27/5/22

Abstract :

Health insurance is insurance that covers the whole or a part of the risk of a person incurring medical expenses, spreading the risk over a large number of persons. By estimating the overall risk of health care and health system expenses over the risk pool, an insurer can develop a routine finance structure, such as a monthly premium or payroll tax, to provide the money to pay for the health care benefits specified in the insurance agreement. The benefit is administered by a central organization such as a government agency, private business, or not-for-profit entity. According to the Health Insurance Association of America, health insurance is defined as "coverage that provides for the payments of benefits as a result of sickness or injury. It includes insurance for losses from accident, medical expense, disability, or accidental death and dismemberment".The objective of this report is to predict the premium amount that has to be paid by the buyer given some of his personal information. Decison Tree, Random forest regressor, and multiple linear regression are the models used, and found that the Random Forest regressor model has given the best R2 score among the models used.

1.0 Introduction

Health Insurance is a type of insurance that covers medical expenses. A person who has taken a health insurance policy gets health insurance coverage by paying a particular premium amount. There is growing awareness of the fact that ill-health perpetuates poverty. In order to prevent the negative downward spiral of poverty and illness, more people in recent years are increasingly choosing various models of health insurance to increase access to health care for their households. To help them choose the best health care premium according to their current situation this model is trained to predict the premium price of the Health care plan.

2.0 Business Needs Assessment

This model can be used by any Health Insurance company to help the customer better understand the amount they have to pay for the premium to avail of the benefits in case of a health emergency.

3.0 Target Specification and Characterisation

People who have severe illness due to some disease and people who don't want to run out of savings in case of emergency as it did to a lot of people during COVID 19 period.

Factors considered in this model are :

1. Age - Age of the person
2. Gender - Gender of the person
3. BMI - Body mass index of that person
4. Children _ How many children does he/she have?
5. Smoker - Whether the person is a smoker or not?
6. Region - to which Region the person belongs to
7. Charges - Amount has to be paid by the person for Health Insurance premium

4.0 External Search

The Dataset for this model is downloaded from kaggle.com.Which consists of 1338 rows with 7 variables. Where the Premium Price is our Target Variable.Analysis was done using three machine learning techniques.

5.0 Bench Marking Alternate Products

Star health insurance, Reliance health insurance, Apollo Munich health insurance, and the latest startup ditto insurance all of these companies can use and a few already use this model to help the customers and make more sales by offering additional benefits.

6.0 Applicable Patents

US4491725A - Medical insurance verification and processing system
This patent reads medical information of the patient's background medical and insurance information. This is required to get the information from different patients to train our model

7.0 Applicable Constraints

Data collection of previous patients and their information related to medical insurance can be tough to get from the hospitals.

Marketing has to be done aggressively to make people use this model to buy or not to buy health insurance.

8.0 Business opportunity

We can use this Product to increase the sales of the health insurance companies, Health insurance econtribites to 6% of India's GDP, much higher than most other countries with the same level of economic development. Of that, 4.7% is private and the rest is public. In private insurance, buyers are willing to pay a premium to an insurance company that pools people with similar risks and insures them for health expenses.

9.0 Concept Generation

This Product can be generated by using Machine Learning Algorithms which can be used for predicting a continuous variable. This is a regression problem. In this Model, the Random forest regressor turned out to have a high r2 score compared to other regression algorithms.

10.0 Concept Development

Given Dataset was explored initially to make sure there are no null or missing values or outliers and a few were made to better understand the relationship between the variables. The model was scaled and split into train and test data to use for predicting and finding the r2 score of the model.

11.0 Code Implementation

The link to the jupyter notebook is given below :
https://drive.google.com/file/d/1U_sPvXmL2_GQO7DwUJTG1ix7eq9G5_ve/view?usp=sharing

12. Conclusion

By using this model we can help the customer know how much premium amount he/she has to pay to avail of healthcare benefits from the Insurance company and the Insurance company also can increase their sales by using this model.