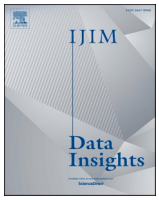
International Journal of Information Management Data Insights 5 (2025) 100328

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei



Customization of health insurance premiums using machine learning and explainable AI

Manohar Kapse [a,*], Vinod Sharma [b], Rutuj Vidhale [b], Varun Vellanki [b]

[a] Jaipuria Institute of Management Indore, India
[b] Symbiosis Centre for Management and Human Resource Development, Symbiosis International (Deemed University), Pune, India

ARTICLE INFO

Keywords:
Premium prediction
Predictive modeling
Machine learning
Health insurance
Explainable AI
XG boost
Gradient boosting
Random forest

ABSTRACT

This study presents an analysis of health insurance premiums across various customer segments. Specifically, it aims to identify the factors influencing the pricing of health insurance premiums, vis a vis their impact on different customer segments. Using a dataset from consumer surveys, coupled with multiple Machine Learning models, the study analyzed and predicted features of importance for premiums paid across various age groups, gender, health conditions, policy duration, and the number of members included in the policy. Finally, the explainable AI was used to predict the weightage of each variable in determining the price of the insurance policy for the individuals. The findings provide crucial insights into the factors such as demographic factors and lifestyle that effectively influence the pricing of health insurance premiums vis a vis their impact on various customer segments. The results of this study will assist prospective buyers and decision-makers in choosing the best health insurance plans.

Introduction

Health insurance is necessary for everyone to cover the sudden and expensive medical costs. With rising medical costs throughout the world, including India, it acts as a safety net for many low- and middle-income families. When comparing the first half of 2023 to the same time in the previous fiscal year, the standalone health insurance premiums increased by 25.24 % (GVR, 2023). This rising healthcare cost has long been a source of anxiety for people, reflected through the costs of health insurance, owing to continual innovation in the medical field to improve patient care and treatment outcomes (Bhardwaj & Anand, 2020). Several studies in the past, have tried to specifically identify the factors that influence health insurance and their claims e.g. (Joji Rao & Pandey, 2013; Quan & Valdez, 2018; Rawat et al., 2021; Ringshausen et al., 2021; Orji & Ukwandu, 2024).

It is, however, quite difficult to create an appropriate forecast model for medical insurance costs because of the multiplicity of elements and their inherent complexity. Several factors could significantly affect the anticipated costs of health insurance, including service characteristics, lifestyle choices, health conditions, regional accessibility, and demographics (Orji & Ukwandu, 2024). The extent of coverage, plan type, deductible, and the customer's age upon enrollment are all important variables that influence the possible cost of health insurance. This study delves into the complex dynamics of health insurance rates while exploring various factors that go on to affect the premium of the policy.

It is impossible to overstate the significance of having an efficient system that can help in predicting individuals' healthcare premiums (Adeniran et al., 2024; Danda et al., 2024). Insurance firms are increasingly using machine learning (ML) techniques to enhance their policies and premium settings since ML algorithms have demonstrated the ability to accurately forecast high-cost, high-need patient expenditures (AbdElminaam et al., 2024; Kolambe & Kaur, 2024; Orji & Ukwandu, 2024; Yang et al., 2018a). ML has the potential to improve the effectiveness of policy wording in the insurance industry. ML algorithms in the healthcare industry are very effective in forecasting high-demand, high-cost patient costs (Bhadouri &, Singh2024; ul Hassan et al., 2021; Yang et al., 2018b). According to Orji et al. (2022a), the most used ML feature for industrial applications is now predictive analytics.

Panay et al. (2019) have observed that the black-box character of ML algorithms somewhat counteracts their excellent performance in healthcare. Predictive analytics may be biassed if the clinical and personal data about patients is not clearly understood or communicated

* Corresponding author.
E-mail addresses: m10oct@gmail.com (M. Kapse), sharmavins@gmail.com (V. Sharma), rutuzvidhale14@gmail.com (R. Vidhale), varunvellanki38@gmail.com (V. Vellanki).

https://doi.org/10.1016/j.jjimei.2025.100328

Available online 7 February 2025