

Chapter 74

Machine Learning Based Program to Prevent Hospitalizations and Reduce Costs in the Colombian Statutory Health Care System

Alvaro J Riascos

University of los Andes and Quantil, Bogotá, Colombia

Natalia Serna

University of Wisconsin-Madison, Madison, USA

ABSTRACT

Health-care systems that rely on hospitalization for early patient treatment pose a financial concern for governments. In this article, the author suggests a hospitalization prevention program in which the decision of whether to intervene on a patient depends on a simple decision model and the prediction of the patient risk of an annual length-of-stay using machine learning techniques. These results show that the prevention program achieves significant cost savings relative to several base scenarios for program efficacies greater than or equal to 40% and intervention costs per patient of 100,000 to 700,000 Colombian pesos (i.e., approximately 14% to 100% of the average cost per patient in Colombia statutory health care system). This article also shows how tree-based methods outperform linear regressions when predicting an annual length-of-stay and the final model achieves a lower out-of-sample error compared to those of the Heritage Health Prize.

DOI: 10.4018/978-1-6684-6291-1.ch074

INTRODUCTION

Avoidable hospitalizations are a source of increased health expenditures in many health systems. Prolonged length-of-stay is costly for providers, insurers, and patients because it is associated to greater health service consumption and to the development of endangering states during the hospital stay. In the Colombian public health care system, the increase in health costs due to avoidable hospitalizations has raised many questions on whether insurers are implementing prevention programs and on whether such programs are effective. In this context, prediction of patient annual length-of-stay (LOS) is an important tool for resource allocation and improving patient health outcomes. Accordingly, the objectives of this paper are: predicting the annual length-of-stay of users in the public health care system in Colombia and estimating the potential cost savings of a preventive program whose main input is the annual LOS prediction.

Most of the literature on prediction of annual LOS has been developed from the providers' perspective rather than from the insurers' perspective. Many authors predict LOS using a sample of patients with specific acute conditions or physiological traits that are often unobserved by the insurer. For example, Chang et. al (2002) study individuals with cerebrovascular accident, Tu & Guerriere (1993) study patients that are admitted to the intensive care unit after having a cardiac surgery, Chertow, Burdick, Honour, Bonventre, & Bates, (2005) focus on patients with renal failure, and Clague, Craddock, Andrew, Horan, & Pendleton, (2002) analyze patients with hip fracture. Our study differs from the previous ones in the sense that we predict annual LOS using information that is symmetrical between insurers, providers, and the government. We do not focus on users with particular health conditions but analyze a representative sample of individuals in the public health care system with heterogenous demographic and morbidity characteristics. We also lack data regarding specific patient physiological traits and we extend our analysis to measuring the potential cost savings of a prevention program where the intervention is decided upon patient LOS prediction. With regard to the empirical techniques for predicting annual LOS, we use machine learning approaches similar to the ones used by Rezaei, Ahmadi, Alizadeh, & Sadoughi (2013) and Walsh et al. (2004), which include boosted trees, random forests, and artificial neural networks.

The remainder of this paper is structured as follows: after this introduction, section II describes the Colombian public health care system, section III provides the empirical framework, section IV describes our database and the data preprocessing, section V presents the results of machine learning techniques, section VI presents the impact of LOS on health costs, and section VII concludes.

The Colombian Public Health Care System

The Colombian public health care system consists of two regimes: contributory and subsidized. The first covers 44 percent of the population and the second the remaining 56 percent. Each regime has its own network of health insurers and health service providers, which are responsible for providing a predetermined benefits package to all enrollees, known as the "Plan Obligatorio de Salud" (POS). In the contributory regime, enrollees (formal employees and individual contractors) pay for health care services a compulsory monthly tariff proportional to their income, while the subsidized regime is fully funded by the government.

Contributions of enrollees in the contributory regime are collected by a government agency called FOSYGA. This agency redistributes contributions to insurers at the beginning of the year using a risk-adjusted premium per enrollee known as the "Unidad de Pago por Capitación" (UPC). The capitation

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/machine-learning-based-program-to-prevent-hospitalizations-and-reduce-costs-in-the-colombian-statutory-health-care-system/307518

Related Content

Challenges and Issues in Plant Disease Detection Using Deep Learning

Priyanka Sahu, Anuradha Chug, Amit Prakash Singh, Dinesh Singhand Ravinder Pal Singh (2021). *Handbook of Research on Machine Learning Techniques for Pattern Recognition and Information Security* (pp. 56-74).

www.irma-international.org/chapter/challenges-and-issues-in-plant-disease-detection-using-deep-learning/279904

Integrated Regression Approach for Prediction of Solar Irradiance Based on Multiple Weather Factors

Megha Kambleand Sudeshna Ghosh (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-12).

www.irma-international.org/article/integrated-regression-approach-for-prediction-of-solar-irradiance-based-on-multiple-weather-factors/294105

Transforming Retail Markets Through IoT and AI-Enabled Architecture to Identify COVID-19 Shoppers

Mohan N.and Sanjeev Kumar Thalari (2023). *Advanced Machine Learning Algorithms for Complex Financial Applications* (pp. 166-182).

www.irma-international.org/chapter/transforming-retail-markets-through-iot-and-ai-enabled-architecture-to-identify-covid-19-shoppers/317023

Efficient Closure Operators for FCA-Based Classification

Nida Meddouriand Mondher Maddouri (2020). *International Journal of Artificial Intelligence and Machine Learning* (pp. 79-98).

www.irma-international.org/article/efficient-closure-operators-for-fca-based-classification/257273

Automatic Multiface Expression Recognition Using Convolutional Neural Network

Padmapriya K.C., Leelavathy V.and Angelin Gladston (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-13).

www.irma-international.org/article/automatic-multiface-expression-recognition-using-convolutional-neural-network/279275