

Chapter 11

Prediction of the Disappearance of Companies From the Market in Bogotá, Colombia Using Machine Learning

William Stive Fajardo-Moreno

EAN University, Colombia

Rubén Dario Acosta Velásquez

EAN University, Colombia

Ivan Dario Castaño Pérez

Ruta N, Colombia

Leonardo Espinosa-Leal

 <https://orcid.org/0000-0001-6861-8024>

Arcada University of Applied Sciences, Finland

ABSTRACT

In this chapter, the results concerning the modeling of companies' disappearance from Bogotá's market using machine learning methods are presented. The authors use the available information from Bogotá's Chamber of Commerce, where the companies are registered yearly. The dataset comprises the years 2017 to 2020 with almost 3 million registries. In this work, a deep analysis of the different features of the data is presented and explained. Next, four state-of-the-art machine learning models are trained for comparison: logistic regression (LR), extreme learning machine (ELM), random forest (RF), and extreme gradient boosting (XGBoost), all with five-fold cross-validation and 50 steps in the randomized grid search. All methods showed excellent performance, with an average of 0.895 in the area under the curve (AUC), being the latter algorithm the best overall (0.97). These results are in agreement with the state-of-the-art values in the field and will be of paramount importance to assess companies' stability for Bogotá's local economy.

DOI: 10.4018/978-1-7998-8185-8.ch011

INTRODUCTION

The recent advances in machine learning and data availability have increased the capacity to create computational models for decision-making in the corporate realm. These models will have the ability to reduce uncertainty and make informed decisions for the organizations' benefits and the people in general. Economic stability is of paramount importance for society, and the use of these methods can be critical from both private and public perspectives.

The previous global financial crisis has shown how the lack of control from governmental institutions can hardly impact all societal network levels, especially the lower layers. Many regulatory frameworks arise afterward, targeting mainly financial institutions and insurance operators (Acemoglu, 2015). Thoughtful modeling of companies' future capabilities to remain competitive in the market has become a necessary element, not only from the management of the individual companies but also from governmental and financial institutions. Companies can use these models to establish future business strategies or strengthen the current ones. Banks, public institutions, and private investors can predict future capabilities or assess future business risks, capital investing, loans, subsidies, among others. The goal is to keep the market's stability, generate profit, and in general, produce a steady and continuous wealth for the whole society.

The rapid development of centralized information databases and the processing power and the accessible specialized software have shifted the classical analysis of companies from statistical modeling to more data-centric approaches since Altman's seminal work (Altman, 1968). A decade later, Ohlson (1980) pioneered machine learning strategies for corporate bankruptcy modeling. Nowadays, most predictions are made by adopting optimal classification algorithms, ranging from logistic regression, support vector machines, and decision trees to more complex deep neural network modeling (Qu, 2019). Moreover, other methods such as genetic algorithms (Jiang, 2009), particle swarm optimization (Chen, 2013), or even reinforcement learning (Espinosa-Leal, 2020) algorithms have been proposed in combination with machine learning to improve the models' prediction capacity. Furthermore, an additional challenge is finding the right models where explainability and interpretability balance with the prediction accuracy and the possibility of the ethical assessment to avoid biases in the algorithms' outcomes.

In this work, we study and present the results concerning the prediction of the disappearance of companies registered in the commerce of Bogota, Colombia, during the period 2016 until 2019. This data is updated every year. Therefore, if a company fails or does not add its name in the chambers' registry, it is automatically declared closed or out of business starting that year. This fact can be considered an arbitrary generalization. The original information contains other valuable information such as size, legal type, economic sector, number of employees, as well as other features. This data can be used for analytics purposes using cutting-edge statistical and artificial intelligence methods. Hence, we model companies' disappearance following the registry, as mentioned earlier, if they have or have not been renewed in a specific year. Upon a full statistical description of the dataset, where each variable is presented and described, a feature selection process of the data is performed. In the final stage, we split the dataset into two subsets, one for training and the other for testing, then fit four different machine learning methods: Logistic Regression (LR), Extreme Learning Machine (ELM), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) with a complete parameter's optimization to ensure the final model is optimal. The obtained scores are presented with the testing subset.

This chapter is organized as follows: In the background section, a compilation of the most recent and relevant works in the topic of machine learning modeling for predicting bankruptcy or insolvency in companies, as well as their predictive capabilities is presented and discussed, then in the research

18 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/prediction-of-the-disappearance-of-companies-from-the-market-in-bogot-colombia-using-machine-learning/285878

Related Content

Financial Risk Management: An Introduction

Mirela-Madalina Stoian and Rares-Gabriel Stoian (2016). *Managing Project Risks for Competitive Advantage in Changing Business Environments* (pp. 301-311).

www.irma-international.org/chapter/financial-risk-management/154333

Integrating the Information Systems Success Model With Project Success Management Process: Position Paper

Ali Varshosaz, João Varajão and Nilton Takagi (2021). *International Journal of Applied Management Theory and Research* (pp. 1-13).

www.irma-international.org/article/integrating-the-information-systems-success-model-with-project-success-management-process/279651

Wasta, the Impact of Preferment on Organizational Culture and Human Resource Management

Elizabeth Kassab Sfeir (2019). *International Journal of Applied Management Theory and Research* (pp. 16-31).

www.irma-international.org/article/wasta-the-impact-of-preferment-on-organizational-culture-and-human-resource-management/227054

Strategic Leadership Competency Development

Paul G. Putman (2017). *Encyclopedia of Strategic Leadership and Management* (pp. 1495-1520).

www.irma-international.org/chapter/strategic-leadership-competency-development/173613

Measuring Digital Marketing Performance: A Balanced Scorecard Approach

Tasnia Fatin and Nayem Rahman (2020). *International Journal of Applied Management Theory and Research* (pp. 1-15).

www.irma-international.org/article/measuring-digital-marketing-performance/244216