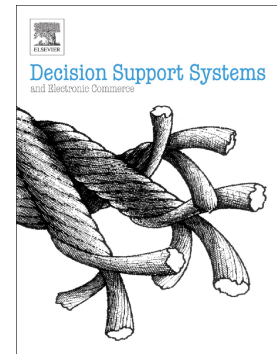


# Accepted Manuscript

An investigation of bankruptcy prediction in imbalanced datasets

David Veganzones, Eric Séverin



PII: S0167-9236(18)30108-8  
DOI: doi:[10.1016/j.dss.2018.06.011](https://doi.org/10.1016/j.dss.2018.06.011)  
Reference: DECSUP 12969  
To appear in: *Decision Support Systems*  
Received date: 19 December 2017  
Revised date: 30 May 2018  
Accepted date: 29 June 2018

Please cite this article as: David Veganzones, Eric Séverin , An investigation of bankruptcy prediction in imbalanced datasets. *Decsup* (2018), doi:[10.1016/j.dss.2018.06.011](https://doi.org/10.1016/j.dss.2018.06.011)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# An investigation of bankruptcy prediction in imbalanced datasets

David Veganzones<sup>\*a,b</sup> • Eric Séverin<sup>a</sup>

a) Université de Lille, IAE Lille, 104 Avenue de Peuple Belge, 59000, Lille, France, Laboratoire Rime Lab. EA7396

b) Université de Paris Nanterre, IUT de Ville d'Avray, 200 Avenue de la République, 92001, Nanterre, France

---

## Abstract

Previous studies of bankruptcy prediction in imbalanced datasets analyze either the loss of prediction due to data imbalance issues or treatment methods for dealing with this issue. The current article presents a combined investigation of the degree of imbalance, loss of performance, and treatment methods. It determines which imbalanced class distributions jeopardize the performance of bankruptcy prediction methods and identifies the recovery capacities of treatment methods. The results show that an imbalanced distribution, in which the minority class represents 20%, significantly disturbs prediction performance. Furthermore, the support vector machine method is less sensitive than other prediction methods to imbalanced distributions, and sampling methods can recover a satisfactory portion of performance losses. Accordingly, this study provides a better understanding of the data imbalance issue in the field of corporate failure and serves as a methodological guide for designing bankruptcy prediction methods in imbalanced datasets.

**Keywords:** bankruptcy prediction, imbalanced dataset, finance

JEL: C53, G33

---

---

\* Authors' email addresses: [david.veganzones@gmail.com](mailto:david.veganzones@gmail.com); [e.severin@wanadoo.fr](mailto:e.severin@wanadoo.fr)  
Corresponding Author: Veganzones: <sup>o</sup>+34 633 246 629; Séverin: +33 06 77 50 72 86

Download English Version:

<https://daneshyari.com/en/article/6948331>

Download Persian Version:

<https://daneshyari.com/article/6948331>

[Daneshyari.com](https://daneshyari.com)