

Estimation of Probability of Default with Machine Learning Techniques: a comparative approach

Stefano Bonini^{*},^a, Giuliana Caivano^a

^a School of Economics, Management, and Statistics, University of Bologna, Bologna, Italy

^{*}*Corresponding author:* stefano.bonini@unibo.it

Estimation of Probability of Default with Machine Learning Techniques: a comparative approach

Default prediction through probability of default modelling has attracted lots of research interests in the past literature and recent studies have shown that Artificial Intelligence (AI) methods achieved better performance than traditional statistical methods. This paper empirically investigates the results of applying different machine learning techniques through the overall estimation process to reduce the running time, maximize – in the first stage - the predictive power and contribute of each variable to the estimation of PDs. In the second stage, we have identified the best multivariate combination of drivers by comparing the results of a set of supervised machine learning algorithm. In the last development stage, we have applied an unsupervised machine learning to calibrate parameters and rank the customers within an ordinal n-class scale. Finally, we have verified the calibration goodness through classical calibration test (e.g. binomial tests). The study has been done on big data sample with more than 800,000 Retail customers of a European Bank under ECB Supervision, with 10 years of historical information and more than 600 variables to be analysed for each customer.

Keywords: Quantitative Finance, Risk Management, Credit Risk, Machine Learning, Big Data, Data Analysis.

Subject classification codes: C38, C45, C46, C51, C52, C53, G24, G32