

Unsupervised Clustering of Debt Portfolios Followed by Expert Strategy Allocation: An Exploratory Data Analysis*

Beata BASIURA, Łukasz JANKOWSKI and Rafał JANKOWSKI

AGH University of Krakow, Krakow, Poland

Correspondence should be addressed to: Beata BASIURA, bbasiura@agh.edu.pl

* Presented at the 45th IBIMA International Conference, 25-26 June 2025, Cordoba, Spain

Abstract

This study addresses the challenge of managing mass debt portfolios by integrating unsupervised machine learning techniques with expert knowledge to enhance debt recovery processes. Existing literature highlights the need for more tailored strategies in debt collection, especially given the heterogeneity of debtors and receivables. To fill this gap, we apply the K-Modes clustering algorithm to a large dataset of over 870,000 receivables, characterized by categorical features such as debt size, legal form, geographic region, and communication history. The objective was to identify homogeneous groups within the portfolio to enable differentiated collection strategies. After validating various clustering approaches, K-Modes with three clusters was selected as the most representative model. Each cluster revealed distinct debtor and debt characteristics, which were then evaluated by industry experts to assign optimized collection strategies. The findings demonstrate that aligning targeted collection methods—ranging from automated, low-cost campaigns to more intensive, personalized negotiation—can significantly increase recovery rates while reducing operational costs. Furthermore, the study illustrates the practical value of this approach by showing how new debt portfolios can be quickly classified into established segments, enabling immediate application of the most effective strategies. The integration of data mining techniques with expert assessment not only improves efficiency but also creates a foundation for future automation and personalization in mass debt collection. This research offers a scalable framework that can adapt to dynamic market conditions and supports responsible debt management practices.

Keywords: mass debt collection, K-Modes clustering, portfolio segmentation, machine learning, expert strategy allocation