

Review of: "Application of Data Mining Combined with K-means Clustering Algorithm in Enterprises' Risk Audit"

Mohammad Mahdi Ershadi¹

¹ Amirkabir University of Technology

Potential competing interests: No potential competing interests to declare.

This study focused on enhancing enterprise financial risk management through the utilization of data mining technology, specifically the application of the K-means clustering algorithm in risk audit processes. The paper investigates the integration of the K-means clustering algorithm in assessing the paperless status of electronic payment within e-commerce enterprises. Moreover, it presents a risk audit model for e-commerce entities employing a combination of the K-means algorithm and the Random Forest Light Gradient Boosting Machine (RF-LightGBM). The model entails stages such as data preparation, preprocessing, construction, application, and evaluation to scrutinize payment flows using big data analysis technology. Through simulation-based evaluation, the model exhibits a classification accuracy of 95.46%, with a stable data message delivery rate of approximately 81.54%. Furthermore, the model demonstrates lower data message leakage, packet loss, and average delay rates compared to alternative approaches. Thus, the audit model ensures high data transmission security alongside prediction accuracy, offering a foundation for enhancing safety and risk control in e-commerce audit processes. Although it is an interesting study, there are some points as follows:

- * Separate the introduction, background, and literature review sections. Add graphical abstract, research flow, future works, and limitation sections.
 - * Redesign figures in this paper with better resolutions.
 - * This paper does not emphasize the unique contributions and novelties compared to the existing literature review. It is better to compare this study with other similar studies in this field.
 - * Represent more details about the dataset in this study. Utilize statistical tests to represent the validity of the sample size and the duration of analysis. Besides, list the advantages of the utilized methodologies in this paper in comparison with other similar methods for this purpose.
 - * The literature review needs to be improved with more recent similar studies to be more informative.
 - * It is better to update references using up-to-date papers. In this paper, 7 of 31 references are published after 2020, and the rest of them are before 2020. A table to summarize the literature review is suggested, as well.
 - * The discussion, limitations, and future works sections are missed in this study.
 - * In addition, the text of this paper needs to be revised in a way that is more interesting to readers.
- I hope these comments improve the quality of this paper in the next revisions.

