

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

Manikandan Rajagopal¹

¹ Christ University

Potential competing interests: No potential competing interests to declare.

Novelty:

1. The paper introduces a novel approach by combining data mining technology with the K-means clustering algorithm to enhance financial risk management in enterprises.
2. The integration of Random Forest Light Gradient Boosting Machine (RF-LightGBM) with the K-means algorithm for risk audit in e-commerce enterprises is a unique and innovative contribution.
3. The utilization of big data analysis technology to study the payment flow in e-commerce transactions adds a novel dimension to the research.
4. The application of the RF algorithm in constructing decision trees within the context of risk audit models is a fresh approach that enhances the accuracy of classification.
5. The consideration of the paperless status of electronic payment in e-commerce transactions sets this research apart and highlights its relevance to current technological trends.

Significance:

1. The research addresses a critical need for improved financial risk management mechanisms in enterprises, particularly in the context of e-commerce transactions.
2. The high classification accuracy achieved by the proposed model (95.46%) indicates its significant potential for enhancing risk control and safety in e-commerce audit processes.
3. The lower packet loss rate, data message leakage rate, and average delay compared to other models underscore the practical significance and effectiveness of the proposed approach.
4. The use of real-world data from Jingdong Mall to establish the risk audit model demonstrates the applicability and relevance of the research to e-commerce enterprises.
5. The paper's findings provide a strong foundation for future applications of data mining in risk audit processes across different industry sectors, indicating broader implications for risk management practices.

Literature Survey:

1. The paper provides a comprehensive overview of existing literature related to risk audit models, data mining technologies, and machine learning algorithms, establishing a solid theoretical foundation.
2. The integration of K-means clustering with RF-LightGBM for risk audit in e-commerce enterprises is well-supported by

relevant prior research in machine learning and financial risk management.

3. The discussion on the limitations of traditional risk audit methods in the context of e-commerce transactions demonstrates a clear understanding of the existing research landscape.
4. The incorporation of the ID3 algorithm within the RF algorithm framework is justified through references to its effectiveness in decision tree construction, adding credibility to the methodology.
5. The comparison of the proposed model's performance metrics with those of other models in related fields highlights the research's awareness of current advancements and benchmarks in the domain.

Methodology:

1. The detailed explanation of the RF algorithm's construction process, including training set generation and decision tree construction, enhances the paper's methodological rigor.
2. The use of bag random sampling with replacement to prevent local optima in decision trees demonstrates a thoughtful approach to mitigating algorithmic biases.
3. The description of the ID3 algorithm and its role in calculating information gain for attribute selection within RF adds clarity to the model construction process.
4. The integration of K-means clustering with RF-LightGBM is described systematically, emphasizing the rationale behind combining these techniques for risk audit in e-commerce.
5. The simulation-based evaluation of the model's performance and comparison with other models ensures robustness and reliability in assessing the proposed methodology's effectiveness.

Language:

1. The paper generally maintains a clear and formal language suitable for academic discourse, enhancing readability and comprehension.
2. However, there are instances where the description of technical concepts could be further clarified for readers less familiar with data mining and machine learning terminology.
3. The paper would benefit from additional transitions between sections to improve the flow of ideas and enhance the overall coherence of the narrative.
4. Some sentences are overly complex and could be rephrased for greater clarity and ease of understanding.
5. While the paper effectively presents the research findings, there is room for improvement in organizing the discussion section to more effectively highlight the implications and limitations of the study.

Final Comment (Accepted with Revisions): Overall, the paper makes a valuable contribution to the field by proposing an innovative approach to risk audit in e-commerce enterprises through the integration of data mining technologies and machine learning algorithms. The research demonstrates thoroughness in methodology and a strong understanding of relevant literature. The suggested revisions primarily focus on enhancing clarity in language, improving transitions between sections, and refining the discussion to better emphasize the research's significance, limitations, and future directions. With these revisions, the paper will significantly strengthen its impact and appeal to a wider audience of

researchers and practitioners in the field of financial risk management and e-commerce audit processes.