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

Alex Junior da Cunha Coelho[1]

1 Federal University of São João del-Rei

The manuscript presents the application of Data Mining by combining the Random Forest Algorithm with K-means in enterprise Risk-Audit. The author analyzes the application of the proposed methodology in a real dataset provided by Jingdong Mall. The use of Machine Learning techniques in this sort of problem is very interesting, and I congratulate the author for all the hard work in preparing this article.

Aiming to improve the quality of the paper and clarify some aspects, I suggest the author revise the points detailed below:

- The Abstract section needs to emphasize the motivation of the research;

- In the literature review, it's interesting to point out the advantages and disadvantages of other related work. Also, it would be great to insert a table
summarizing the literature review;

- The author states a data division into a training set and test set with a ratio of 8:2. Why is the training set too big compared with the test set? Usually, when applying Machine Learning techniques, the training set is a small sample of the input data that characterizes information well; thus, after the training process, the methodology can classify a large set of samples into K-clusters very quickly;

- The analysis of the results should better compare why the proposed methodology outperforms the other investigated methodologies;

- It's interesting to elaborate on the potential areas and improvements for future work;

Finally, here are some minor comments:

- There are some language issues, such as: Stimulation instead of Simulation, anther instead of another, RD instead of RF. Please review typos along the text;

- Following the pattern of other Figures, the caption of Figure 6 should be below the figure and not above it.