Review of: "Supply Chain Fraud Prediction with Machine Learning and Artificial intelligence"

Suryoday Basak

Potential competing interests: No potential competing interests to declare.

This paper presents an application of machine learning to supply chain fraud. The authors use a large dataset and try three different types of models on the data: logistic regression, random forests and neural networks. The authors explain how they prune their dataset to eliminate non-essential features, and how they perform a feature-importance analysis. The application of the methods is interesting but I believe that the scope and the breadth of the paper can be improved. Here is a list of improvements that I would recommend to the authors.

1. In general, the language used in the paper is generic. While this makes the paper easy to read, which is a good thing, precision is lacking in some places. For example, the authors liberally use the term ‘AI’. At times, it is unclear what they are implying – what they are implying, however, is ‘deep neural network’ or ‘artificial neural network, whereas AI in itself is a subfield of computer science that is much broader than neural networks.
2. With regard to the results, I would recommend increasing the breadth of the analysis presented in the paper. The dataset that the authors have used aggregates ‘around 180000 observations taken from supply chain transactions that have been mined over three years’.
    1. The description of the dataset provided should be improved to include the sources of the data and which companies are included in the dataset.
    2. Then, a company-specific analysis for multiple companies should be presented.
3. The explanations on statistical indicators such as true positive, false positive, etc. need not be as extensive as presented in the paper. These concepts are canonical across all domains of data analysis. Similarly, the diagram for the confusion matrix is not necessary. The authors should, instead, focus their efforts on performing more experiments and presenting more results.

I would encourage the authors to attempt improving their article based on the above suggestions.