Review of: "Bank Customer Churn Prediction Using SMOTE: A Comparative Analysis"

Elton Alves[1]

1 Federal University of Southern and Southeastern Pará

Potential competing interests: No potential competing interests to declare.

I consider that the work needs some adjustments, as there are some points that need to be better clarified. My considerations are:

1. It is not clear what the work contributes.
2. In related works, the authors present some works, but are unable to show the difference in the proposal presented.
3. Remove the information: For the evaluation of bank customer churn, it is essential to devise a methodology, as all work needs a methodology.
4. Regarding AG, which the authors mention the use of, there is a lot of obscure information:

- What type of selection and crossover rate is used?
- Authors must show the evolution of fitness to prove their results.
- How to ensure that the information in Table 1 was the best for the AG?
- The authors mention a mutation rate of 3.3%, the literature recommends up to 1%, how can they prove that there was no loss of diversity?
- Where is the proof of the use of AG in the proposal?

5. What justifies the choice of four algorithms: Random Forest (RF), K-Nearest Neighbor (KNN), AdaBoost, and Artificial Neural Network (ANN)?
6. What percentage of training and testing data? What type of validation was used?
7. There are two sections for results with SMOTE.
8. What are the SMOTE parameters used?
9. I don't think it's prudent to put a code snippet like figure 6.
10. The authors do not comment on the results of the confusion matrix.
11. Place only one ROC graph for all results.
12. Some results were close, such as RF and KNN, so I don't think it's prudent to point to KNN as being the best; it requires a deeper statistical analysis to make this statement.