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

Venugopal Jothi Prakash

Potential competing interests: No potential competing interests to declare.

This research presents a comprehensive study aimed at enhancing the prediction of customer churn within the banking sector. It utilizes the Synthetic Minority Over-sampling Technique (SMOTE) to address the imbalance in datasets and compares the performance of several classification algorithms, including Random Forest (RF), K-Nearest Neighbors (KNN), AdaBoost, and Artificial Neural Networks (ANN). Here's a detailed review, highlighting the strengths and areas for improvement:

Strengths:

1. **Relevance and Timeliness:** The study addresses a critical issue in the banking sector, where understanding and predicting customer churn is vital for maintaining competitiveness and profitability. Given the dynamic nature of customer preferences and the competitive landscape, this research is both timely and relevant.
2. **Comprehensive Methodology:** The use of SMOTE to balance the dataset, followed by the application of genetic algorithms for feature selection, provides a robust methodology for enhancing model performance. This approach is well-justified and systematically implemented, offering a thorough investigation into churn prediction.
3. **Detailed Comparative Analysis:** By comparing multiple classification algorithms, the paper provides a comprehensive overview of how each model performs in the context of churn prediction. This comparative analysis is invaluable for readers seeking to understand the strengths and weaknesses of various machine learning techniques in predicting customer churn.
4. **Extensive Evaluation Metrics:** The use of multiple metrics (accuracy, precision, recall, and F1-score) for model evaluation ensures a well-rounded assessment of the models' performance. This multi-metric evaluation provides a deeper insight into the models' ability to predict churn accurately.

Areas for Improvement:

1. **Model Interpretation and Insights:** While the paper excellently details the performance of each model, it could further benefit from a more detailed interpretation of the results. Insights into why certain models performed better than others, considering the characteristics of the banking dataset, would provide added value.
2. **Handling of Imbalanced Data:** The application of SMOTE is well-justified; however, the paper could explore and discuss alternative methods for handling imbalanced data. A comparison of SMOTE with other techniques, such as ADASYN or Tomek Links, could offer a broader perspective on managing class imbalance.

3. **Impact of Feature Selection:** The use of genetic algorithms for feature selection is a highlight of the study. However, the paper could delve deeper into the impact of this selection on model performance. A discussion on which features were deemed most informative and why could enrich the reader's understanding of churn prediction factors.
4. **Practical Implications:** While the study is robust in its technical execution, expanding on the practical implications of the findings would enhance its relevance. Suggestions on how banks can implement these models in real-world scenarios, including potential challenges and strategies for deployment, would be beneficial.
5. **Future Research Directions:** The conclusion mentions future research will focus on examining customer churn characteristics. Elaborating on other potential avenues for future research, such as exploring unsupervised learning methods for identifying unknown patterns in churn or integrating customer sentiment analysis, could inspire further studies.
6. **Literature Review:** Elaborate the related works to include DL and Transformer based approaches such as the ones listed below:

[1]V. Jothi Prakash and N. K. Karthikeyan, "Enhanced Evolutionary Feature Selection and Ensemble Method for Cardiovascular Disease Prediction," Interdisciplinary Sciences: Computational Life Sciences, vol. 13, no. 3, pp. 389–412, Sep. 2021, doi: 10.1007/s12539-021-00430-x.

[2]V. J. Prakash and N. K. Karthikeyan, "Dual-Layer Deep Ensemble Techniques for Classifying Heart Disease," Information Technology and Control, vol. 51, no. 1, pp. 158–179, Mar. 2022, doi: 10.5755/j01.itc.51.1.30083.

[3]A. A. V. S and J. P. V, "A Modified Firefly Deep Ensemble for Microarray Data Classification," The Computer Journal, vol. 65, no. 12, pp. 3265–3274, Dec. 2022, doi: 10.1093/comjnl/bxac143.

[4]J. P. Venugopal, A. A. V. Subramanian, and J. Peatchimuthu, "The realm of metaverse: A survey," Computer Animation and Virtual Worlds, vol. 34, no. 5, Sep. 2023, doi: 10.1002/cav.2150.

[5]A. A. V. Subramanian and J. P. Venugopal, "A deep ensemble network model for classifying and predicting breast cancer," Computational Intelligence, vol. 39, no. 2, pp. 258–282, Apr. 2023, doi: 10.1111/coin.12563.

[6]J. P. V and A. A. V. S, "Cross-lingual Sentiment Analysis of Tamil Language Using a Multi-stage Deep Learning Architecture," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 12, pp. 1–28, Dec. 2023, doi: 10.1145/3631391.

[7]G. K. P, A. A. V. S, J. P. V, A. Paul, and A. Nayyar, "A context-sensitive multi-tier deep learning framework for multimodal sentiment analysis," Multimedia Tools and Applications, Dec. 2023, doi: 10.1007/s11042-023-17601-1.

[8]J. P. V. and A. A. V. S., "A multi-aspect framework for explainable sentiment analysis," Pattern Recognition Letters, vol. 178, pp. 122–129, Feb. 2024, doi: 10.1016/j.patrec.2024.01.001.

[9]J. P. V., A. A. V. S., G. K. P., and K. N.K., "A novel attention-based cross-modal transfer learning framework for predicting cardiovascular disease," Computers in Biology and Medicine, vol. 170, p. 107977, Mar. 2024, doi: 10.1016/j.compbiomed.2024.107977.

Overall, this paper makes a significant contribution to the field of churn prediction within the banking sector. Its methodological rigor, extensive comparative analysis, and the application of SMOTE for data balancing are commendable aspects. Enhancing the document with more detailed interpretations, broader discussions on data imbalance handling, insights into feature selection impact, practical implementation guidance, and expanded future research directions could further strengthen its contribution. This study provides a valuable resource for academics and practitioners alike, offering insights and methodologies applicable beyond the banking industry to any sector facing customer retention challenges.