ENHANCING PREDICTIVE MODELING IN CUSTOMER RETENTION RATE ANALYSIS: A STACKED APPROACH INTEGRATING C4.5 DECISION TREES AND FEED-FORWARD NEURAL NETWORKS

Rationale

Accurately predicting customer retention rates remains a significant challenge in customer relationship management (CRM) across diverse industries. Lee et al. (2022) argue that by employing decision trees, businesses gain the ability to analyze data effectively and forecast outcomes. This facilitates the formulation of well-informed decisions for strategic planning and operational management. By employing advanced computational techniques, the research seeks to unpack the complexities of consumer decision-making processes, ultimately contributing to a more comprehensive understanding of market behavior (Greene et al. 2017). Traditional methods often struggle to capture the nuances of customer behavior, hindering the development of optimal retention strategies and hindering efforts to foster longterm customer relationships. Existing predictive modeling techniques may lack the necessary complexity to effectively analyze large and intricate datasets, leading to diminished predictive accuracy and limited actionable insights. This gap in research highlights the need for a more robust and sophisticated approach to customer retention rate analysis. However, current approaches also face limitations, including scalability challenges with expanding datasets, interpretability issues in complex models, and difficulties in handling noisy or incomplete data.

To bridge the identified research gap, this study proposes a novel stacked ensemble approach that integrates C4.5 decision trees and feed-forward neural networks. This integration aims to leverage the complementary strengths of these distinct modeling techniques, ultimately enhancing the predictive accuracy and comprehensiveness of customer retention rate analysis. C4.5 decision trees offer the advantage of interpretable rule generation and efficient handling of categorical variables. Feed-forward neural networks, on the other hand, demonstrate proficiency in capturing non-linear relationships and intricate patterns within the data. (Džeroski, S., & Ženko, B. (2004). Stacking classifiers presents a promising strategy for achieving superior predictive accuracy compared to single-model selection. This ensemble approach leverages the complementary strengths and perspectives of multiple classifiers, harnessing their collective intelligence. Stacking mitigates biases inherent to individual models and improves overall prediction performance. By combining these methods within a stacked framework, we exploit their respective strengths to generate more reliable and actionable predictions of customer retention rates.

The adoption of the proposed stacked ensemble approach has the potential to generate substantial benefits for businesses by improving customer retention strategies and enhancing overall profitability. This approach would enable organizations to accurately predict customer churn, identify key factors influencing retention, and proactively implement targeted interventions to mitigate customer attrition and cultivate long-term customer loyalty.

Furthermore, the enhanced predictive modeling capabilities could facilitate the development of personalized marketing campaigns, tailored product recommendations, and optimized resource allocation, ultimately leading to increased customer satisfaction and sustainable revenue growth. However, it is important to acknowledge that while this approach addresses many limitations inherent to traditional methods, it may still encounter challenges when handling exceptionally large datasets or datasets exhibiting significant class imbalance.

This study presents a novel stacked ensemble approach that integrates C4.5 decision trees and feed-forward neural networks as a promising solution for the challenges in customer retention rate analysis. By capitalizing on the strengths of these advanced modeling techniques, the proposed approach aims to equip businesses with actionable insights that can inform the development of more effective retention strategies and foster stronger customer relationships. This research initiative holds significant value for the field of computer science by advancing the frontiers of predictive analytics and empowering organizations to leverage the full potential of their customer data. While acknowledging potential limitations, the authors express confidence that the proposed approach will contribute meaningfully to the ongoing endeavors to improve customer retention practices and drive business success within the digital landscape.

Significance of the Study

The significance of conducting this study lies in its potential to revolutionize business approaches to customer retention strategies and elevate customer relationship management (CRM) practices. By addressing shortcomings in existing customer retention rate prediction models, this research offers a unique opportunity to unlock valuable insights into customer behavior and preferences. These insights can empower businesses to develop more targeted and effective retention strategies, leading to increased customer satisfaction, loyalty, and ultimately, sustainable business growth. Furthermore, by employing advanced techniques like stacked ensemble modeling, the study contributes to the advancement of predictive analytics in CRM, paving the way for more accurate and actionable predictions in the future. The impact of this research transcends individual businesses. Improved customer retention practices have

the potential to stimulate economic growth, foster innovation, and enhance overall market competitiveness. Therefore, investing in this study is not only beneficial for businesses seeking to optimize customer relationships but also holds broader significance for the economy and society as a whole.

The proposed hybrid model transcends its application in consumer behavior analysis, exhibiting potential for broader applicability. Its adaptability and effectiveness suggest the possibility of tailoring the model for developing domain-specific predictive technologies across various fields. This innovation holds the potential for far-reaching impacts, fostering advancements in healthcare, finance, environmental monitoring, and beyond. By establishing a robust framework for predictive modeling, this study paves the way for the development of customized technologies that can address the complexities of data analysis in various industries.

The enhanced predictive accuracy and efficiency offered by the stacked approach can contribute to achieving several SDGs. The broader applicability of the proposed model across various sectors holds promise for achieving Goal 8 (Decent Work and Economic Growth – Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all) (UN, 2023) through innovation and job creation (UN, 2023). The improved decision-making facilitated by the model aligns with Goal 9 (Industry, Innovation, and Infrastructure – Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation) by promoting technological advancements (UN, 2023). Furthermore, by fostering informed practices that minimize waste and optimize resource utilization, the approach contributes to Goal 12 (Responsible Consumption and Production – Ensure sustainable consumption and production patterns) (UN, 2023). By empowering businesses with more precise consumer behavior insights, this study contributes to the overarching agenda of sustainable development by promoting economic growth, fostering efficiency, and encouraging responsible consumption practices.

Scope and Limitations

This study focuses on improving predictive modeling for customer retention rate analysis by integrating C4.5 decision trees and feed-forward neural networks within a stacked ensemble framework. The research aims to develop a more accurate and interpretable model capable of identifying key factors influencing customer retention. The study encompasses data preprocessing, model implementation, optimization, and evaluation using real-world customer

behavior datasets. The scope includes analyzing the impact of various model parameters and features on predictive performance while ensuring robustness and scalability.

However, the study has certain limitations. While the stacked approach enhances predictive accuracy, it may face challenges in handling exceptionally large datasets, especially those with significant class imbalances. The interpretability of the feed-forward neural network component may also pose difficulties in extracting explicit decision rules. Additionally, the study is constrained by the availability and quality of datasets, as incomplete or noisy data could affect model performance. The research is also limited to the specific machine learning libraries used, such as scikit-learn and TensorFlow, and may not explore other advanced deep learning architectures. Despite these constraints, the study aims to contribute to the advancement of predictive analytics in customer retention strategies.

Objectives of the Study

This research aims to investigate the efficacy of a stacked approach integrating C4.5 decision trees and feed-forward neural networks in enhancing predictive modeling for customer retention rate analysis. Specifically, the objectives of the study are as follows:

Specifically, the study seeks to:

1. To develop a comprehensive understanding of customer retention rate analysis methodologies, including the factors influencing customer churn and the importance of accurate predictive modeling in CRM.
2. To implement and optimize a stacked approach that integrates C4.5 decision trees and feed-forward neural networks for customer retention rate prediction, ensuring robustness and scalability.
3. To evaluate the performance of the proposed stacked model in terms of predictive accuracy, sensitivity to various features, and interpretability, using real-world customer retention datasets.
4. To provide insights into the underlying factors driving customer churn and identify actionable strategies for improving customer retention based on the findings of the enhanced predictive modeling approach.

Expected Outputs

The expected outputs of this research encompass both theoretical and practical advancements in predictive modeling for customer retention rate analysis:

- A refined and optimized stacked model integrating C4.5 decision trees and feed-forward neural networks, demonstrated to enhance predictive accuracy and efficiency in customer retention rate analysis.
- Empirical results showcasing the performance improvement achieved by the proposed model compared to traditional predictive modeling techniques, along with insights into the nuanced factors influencing customer churn and retention.

Material Used:

- This research will leverage existing machine learning libraries such as scikit-learn and TensorFlow for implementing the C4.5 decision trees and feed-forward neural networks.
- Real-world consumer behavior datasets from various industries will be utilized for training and testing the predictive models.
- Software tools for data preprocessing, model evaluation, and visualization for model implementation and evaluation will be employed to ensure robust analysis and interpretation of results.

References

Predictive Analytics in Business Analytics: Decision tree. (2022). *Advances in Decision Sciences*, *26*(1), 1–30. https://doi.org/10.47654/v26y2022i1p1-30

Greene, M. N., Morgan, P. L., & Foxall, G. R. (2017). NEURAL networks and Consumer behavior: NEURAL models, logistic Regression, and the Behavioral Perspective Model. *Behavior Analyst*, *40*(2), 393–418. https://doi.org/10.1007/s40614-017-0105-x

Džeroski, S., & Ženko, B. (2004). Is Combining Classifiers with Stacking Better than Selecting the Best One? *Machine Learning*, *54*(3), 255–273. https://doi.org/10.1023/b:mach.0000015881.36452.6e