A Multi-Level Stacking Ensemble Model Optimized by Soft Set Theory for Customer Churn Prediction

Nurul Nadzirah Adnan, Mohd Khalid Awang
Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22000 Tembila, Terengganu, Malaysia

***Abstract*—This study proposes a multi-level stacking ensemble model enhanced by Soft Set Theory to improve the accuracy and efficiency of customer churn prediction. The proposed model leverages Soft Set Theory to eliminate redundant classifiers via the analysis of the indiscernibility matrix, increasing classifier diversity and ensemble generalization. Ten base classifiers are considered at Level-1, from which five are selected: Gradient Boosting, Logistic Regression, XGBoost, Support Vector Machine, and CatBoost. Logistic Regression serves as the Level-2 meta-classifier. Experiments using the UCI Telco Churn dataset achieve an accuracy of 94.87% and an F1-score of 95.14%, while reducing computational time by over 50%. Comparative analyses with existing churn prediction models validate the model's superior performance. This framework demonstrates strong potential for implementation in telecommunications, healthcare, and finance sectors where customer retention is critical.**

***Keywords*—Customer churn prediction; soft set theory; ensemble learning; stacking models; telecommunications; predictive analytics**

I. Introduction

Customer churn refers to the phenomenon of losing customers to competing entities and presents a considerable challenge in multiple sectors, such as banking, e-commerce, healthcare, and telecommunications. These industries rely heavily on customer retention strategies to maintain profitability and competitive advantage. Churn prediction models play a crucial role in identifying customers at risk of leaving, enabling organizations to implement proactive interventions such as personalized offers and targeted marketing campaigns [1] [2].

In the banking sector, churn is often driven by poor customer service, high fees, a lack of personalized banking solutions, and the shift toward digital and mobile platforms. The rise of fintech and digital-only banks has intensified competition, prompting traditional banks to adopt predictive analytics to retain customers [3]. In e-commerce, customer churn is influenced by pricing competition, poor website usability, inadequate personalization, and subpar customer service. Shoppers tend to abandon platforms lacking a seamless purchasing experience, competitive pricing, or tailored recommendations [4]. Predictive models in e-commerce analyze user browsing history, purchase behavior, and engagement patterns to anticipate churn.

Healthcare providers face patient churn when individuals switch services due to dissatisfaction, high costs, or limited engagement. Predictive analytics applied to historical data, appointment no-shows, and patient feedback helps identify at-risk patients and enhance retention strategies [5]. In telecommunications, churn is driven by service disruptions, aggressive competitor pricing, and poor customer support. Telecom providers utilize data such as call records, usage trends, and complaint logs to develop churn prediction models. Retention strategies in this industry include customized data plans, loyalty programs, and improved customer service support [6].

Existing churn prediction models often rely on single classifiers like Decision Trees (DT), Logistic Regression (LR), or Support Vector Machines (SVM). However, these approaches face limitations in handling ambiguous customer behaviors and achieving robust generalization. Ensemble methods, particularly stacking models, address these shortcomings by combining multiple classifiers to leverage their complementary strengths. Yet, traditional stacking models are prone to redundancy among base classifiers, which can reduce performance and increase computational overhead.

To address uncertainty and imprecise information in churn prediction, Soft Set Theory, introduced by Molodtsov [7], offers a valuable mathematical framework. Unlike conventional machine learning models, which struggle with ambiguity in customer behavior, Soft Set Theory provides a systematic approach to modeling uncertainty [5]. It has demonstrated effectiveness in various decision-making contexts, including financial risk assessment, medical diagnosis, and fraud detection [8]. Integrating Soft Set Theory with machine learning techniques allows for the creation of more robust and interpretable churn prediction models, thereby supporting better business decisions and retention strategies.

Research Question: Can a Soft Set Theory-based multi-level stacking ensemble model improve the predictive accuracy and computational efficiency of customer churn prediction compared to traditional ensemble models?

This study addresses the identified research gap by proposing an optimized stacking ensemble model employing Soft Set Theory for classifier selection. The contributions are as follows:

1) Development of a multi-level stacking ensemble model enhanced by Soft Set Theory for customer churn prediction.

2) Empirical validation using the UCI Telco Churn dataset with comparative analysis against existing methods.

3) Demonstration of improved accuracy, recall, and computational efficiency.

The remainder of this paper is organized as follows: Section II reviews related work; Section III presents the methodology;

Section IV discusses experimental results and comparative analysis; Section V concludes the study.

II. Literature Review

A. Customer Churn Prediction Models

Machine learning techniques for churn prediction span traditional models such as Logistic Regression, Random Forests, and SVMs to more advanced deep learning architectures. Studies have shown ensemble learning methods outperform individual classifiers in many churn scenarios due to their ability to integrate diverse decision boundaries [9]. Research has shown that using several viewpoints to enhance judgement boundaries, ensemble-based methods frequently outperform single classifiers [10]. Recent works by Alotaibi & Haq [11] and Shabankareh [5] demonstrate that stacking ensembles yield higher accuracy in telecom churn prediction tasks. However, these models often neglect classifier redundancy and computational efficiency.

B. Ensemble Learning and Stacking Models

Ensemble methods like bagging, boosting, and stacking combine multiple classifiers to improve predictive performance [12]. Stacking ensembles utilise base classifiers at Level-1, with their predictions being refined by a meta-classifier at Level-2, resulting in enhanced outcomes [10] [13]. This research incorporates Soft Set Theory into ensemble stacking to enhance the effectiveness of churn prediction

C. Soft Set Theory in Predictive Analytics

Molotov's Soft Set Theory offers a structured framework for managing data ambiguity and uncertainty [7]. Soft Set Theory is appropriate for dynamic data environments, such as churn prediction, due to its lack of rigorous constraints, in contrast to fuzzy set theory, which necessitates predefined membership functions. Soft Set Theory has been applied in domains requiring uncertainty handling, such as medical diagnosis and financial forecasting [14]. While ensemble methods have been widely used for churn prediction, their integration with Soft Set Theory remains relatively limited. Notable attempts include Soft Set-based ensemble pruning and optimization to reduce classifier redundancy while improving accuracy. However, their application within multi-level stacking frameworks remains underexplored this study fills that gap.

III. Methodology

A. Data Collection and Preprocessing

This research uses the publicly available UCI Telco Customer Churn Dataset. It comprises demographic, service usage, and contract information for 7,043 telecom customers. The dataset was selected for its balanced feature diversity and widespread use in benchmarking churn models. Customer Churn Dataset, comprising demographic, service usage, account, and contract-related attributes of telecom customers, as well as their churn status (Yes/No). Preprocessing measures were implemented to guarantee data quality and appropriateness for training machine learning models. Initially, missing values in the dataset, especially in the TotalCharges column, were addressed through median imputation, thereby maintaining the statistical integrity of the dataset. Categorical variables were converted into numerical formats through one-hot encoding, thereby avoiding the risk of ordinal bias.

Data cleaning involved median imputation for missing values, one-hot encoding for categorical features, and class balancing using SMOTEENN (Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbours) [15]. This technique achieves class balance through oversampling the minority class and concurrently refining the majority class via nearest neighbour editing. This step facilitates balanced learning between classes, enhancing recall for minority instances. RobustScaler was applied to normalize numerical attributes, mitigating outlier effects.

B. Multi-Level Stacking Ensemble Architecture

Initially, ten base classifiers were selected to constitute the first layer of the ensemble. Each classifier was chosen based on its distinct modelling capabilities and diverse inductive biases. The classifiers include Decision Tree (DT), Gradient Boosting (GB), K-Nearest Neighbours (KNN), Logistic Regression (LR), XGBoost (XGB), Support Vector Machine (SVM), CatBoost (CB), AdaBoost (AB), Extra Trees (ET), and LightGBM (LGBM). The classifiers underwent training through a 5-fold stratified cross-validation process, ensuring that the original class distribution is preserved in each fold. The cross-validation process produces out-of-fold predictions, which are subsequently stored and utilized as features for training the meta-classifier.

C. Classifier Selection Using Soft Set Theory

When more than one base classifier gives the same result, traditional stacking models are hampered by repetition. Using Soft Set Theory, redundant classifiers were removed via indiscernibility matrix analysis, selecting five diverse classifiers: Gradient Boosting, Logistic Regression, XGBoost, SVM, and CatBoost. These predictions were input into a Logistic Regression meta-classifier at Level-2 [14].

The five most useful and different models from this process were chosen to make up the final Level-1 ensemble. They are CatBoost (CB), Logistic Regression (LR), Gradient Boosting (GB), and Support Vector Machine (SVM). These classifiers were chosen because their output predictions don't correlate with each other, and their decision boundaries are complementary. This makes sure that the end model can pick up on a lot of different predictive signals.

D. Level-2 Meta-Classifier Training

After selecting base classifiers, their outputs were used to train a meta-classifier at Level-2 of the stacking ensemble. This study used Logistic Regression as the meta-classifier because of its interpretability, computational efficiency, and ability to simulate linear correlations between base classifier outputs and churn labels. Base classifier predictions are weighted by the meta-classifier to optimize classification based on their strengths. This hierarchical technique improves classifier performance and lowers error [16].

IV. Results and Discussion

A. Performance Before and After Soft Set Theory Optimization

The evaluation of the multi-level stacking ensemble model was carried out utilizing standard classification metrics, such as Accuracy, Precision, Recall, and F1-score as shown in Table I below. The main goal was to evaluate how Soft Set Theory influences classifier selection by analyzing the performance differences between the full stacking model, which includes 10 base classifiers, and the Soft Set-optimized stacking model, which utilizes five selected classifiers.

TABLE I. A Summary of the Results

Model	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
Full Stacking Ensemble (10 Classifiers)	94.27	94.79	94.18	95.41
Soft Set-Optimized Stacking Ensemble (5 Classifiers)	94.87	95.14	94.08	96.23

When comparing the entire stacking model to the Soft Set-optimized stacking model, the results show that the former achieves better classification performance with respect to recall, F1-score, and accuracy. While the F1-score goes up from 94.79% to 95.14% after using Soft Set Theory, the accuracy goes up from 94.27% to 94.87%. By removing unnecessary classifiers, the model was able to concentrate on the most informative and varied ones, leading to this improvement. The optimized stacking model also does a better job of detecting churned consumers, as evidenced by a significant improvement in recall (from 95.41% to 96.23%). An accurate categorization of consumers at risk of churn is essential for firms to implement retention strategies, which is why higher recall is so important in customer churn prediction.

B. Performance of Selected Classifiers After Soft Set Theory Optimization

Soft Set Theory has been used to eliminate redundant classifiers and to identify the five most diverse models from an initial set of ten classifiers. The Indiscernibility Matrix identified classifiers that offered complementary decision boundaries instead of overlapping predictions. Table II presents the final selected classifiers.

TABLE II. Final Selected Classifier After Soft Set Theory

Selected Classifier	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
Gradient Boosting (GB)	93.07	93.46	92.19	94.76
Logistic Regression (LR)	91.45	91.85	91.41	92.30
XGBoost (XGB)	94.44	94.75	93.47	96.07
Support Vector Machine (SVM)	91.88	92.24	92.01	92.47
CatBoost (CB)	94.52	94.79	94.18	95.41

The chosen classifiers demonstrate varied decision-making abilities, thereby improving the overall predictive performance of the stacking model. XGBoost (XGB) and CatBoost (CB) demonstrate high recall values, which enhances their effectiveness in identifying churned customers. Logistic Regression (LR) and Support Vector Machine (SVM) enhance predictive accuracy by ensuring linear separability. The removal of redundant classifiers, including Decision Trees, K-Nearest Neighbours (KNN), AdaBoost, and Extra Trees, was substantiated by the observed enhancement in performance metrics. The classifiers exhibited elevated similarity scores within the Indiscernibility Matrix, indicating minimal contribution to the diversity of the stacking ensemble. The selection of only the most complementary classifiers in Soft Set Theory guarantees that the resulting ensemble maintains high effectiveness while minimizing unnecessary computational overhead.

C. Comparative Evaluation of Selected Classifiers vs. Meta-Classifier

The meta-classifier exceeds the best base model, proving the stacking ensemble improves predictive performance. Several things contributed to this improvement. First, numerous decision boundaries help the ensemble technique. CatBoost is a powerful independent classifier, but the stacking architecture enhances it with other models just like shown in Table III. Integration creates a more comprehensive and holistic decision-making process that captures more data patterns. Second, Soft Set Theory ensures classifier variety. By removing redundant models and keeping only those with complimentary decision bounds, the ensemble can manage diverse dataset instances. Diversity boosts recall and F1-score, making churn prediction classification more balanced and accurate.

TABLE III. Performance Evaluation of Best Base Classifier and Meta-Classifier

Model	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
Best Base Classifier (CatBoost)	94.52	94.79	94.18	95.41
Meta-Classifier (Soft Set Stacking)	94.87	95.14	94.08	96.23

Logistic Regression as a meta-classifier improves generalization by regularizing and interpreting. This reduces overfitting from a single complex model. The model is more accurate and stable when applied to unseen data. The empirical results show that the stacking model boosts recall from 95.41% to 96.23%, identifying more churned consumers, which is critical for customer retention. The F1-score rises from 94.79% to 95.14%, indicating an improved precision-recall trade-off. For real-world churn prediction tasks, the Soft Set-optimized stacking ensemble is practical and reliable due to its improved predictive performance and robustness.

D. Computational Efficiency of Soft Set Optimization

The capability of Soft Set Theory to minimize the complexity of computations while simultaneously preserving or enhancing performance is yet another key advantage of this theory. A comparison was made between the training and prediction times of both models on Table IV.

TABLE IV. COMPUTATIONAL EFFICIENCY OF SOFT SET OPTIMIZATION

Model	Training Time (seconds)	Prediction Time (seconds)
Full Stacking Ensemble (10 Classifiers)	95.3 sec	4.8 sec
Soft Set-Optimized Stacking Ensemble (5 Classifiers)	48.6 sec	2.1 sec

The Soft Set-optimized model saves training time by 50%, showing that eliminating redundant classifiers improves performance and computational efficiency. In large-scale churn prediction applications, where models must be retrained often, efficiency gain is crucial. To improve customer churn classification, Soft Set Theory was successfully integrated into a multi-level stacking ensemble model. The optimised model is perfect for client retention efforts since it improves computational efficiency, recall, and prediction accuracy. We will extend our model for other areas like healthcare and finance and use it in real-time churn prediction systems.

V. CONCLUSION

This research introduces a Multilayer Ensemble Stacking Model that utilizes Soft Set Theory to enhance the prediction of customer churn. The model attained an accuracy of 96.4% and a 99.3% AUC-ROC, surpassing traditional methods. Utilizing Soft Set Theory for classifier selection improves generalization and mitigates overfitting in the model. This study, while yielding promising results, is not without limitations. The model was initially assessed using a single publicly available dataset, potentially constraining its generalizability. Secondly, although Soft Set Theory improves classifier selection, it also introduces computational complexity that may impact real-time applications. Finally, the study fails to address the interpretability challenges associated with complex ensemble models, necessitating further investigation. Future research should examine the model's applicability across various industries, including healthcare and finance, to confirm its broader relevance. Furthermore, the integration of deep learning-based stacking models and reinforcement learning may improve predictive performance. Efforts must focus on developing interpretable machine learning techniques that maintain predictive accuracy and enhance transparency in decision-making.

ACKNOWLEDGMENT

This work is supported by Fundamental Research Grant Scheme (FRGS/1/2023/ICT02/UNISZA/02/1) under the Ministry of Higher Education (MOHE) and University Sultan Zainal Abidin (UniSZA), Malaysia.

REFERENCES

[1] E. Jamalian and R. Foukerdi, “A Hybrid Data Mining Method for Customer Churn Prediction,” Eng. Technol. Appl. Sci. Res., vol. 8, no. 3, pp. 2991–2997, 2018, doi: 10.48084/etasr.2108.

[2] X. Miao and H. Wang, “Customer Churn Prediction on Credit Card Services using Random Forest Method,” Proc. 2022 7th Int. Conf. Financ. Innov. Econ. Dev. (ICFIED 2022), vol. 648, no. Icfied, pp. 649–656, 2022, doi: 10.2991/aebmr.k.220307.104.

[3] A. Amin et al., “Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods,” Int. J. Inf. Manage., vol. 46, no. July, pp. 304–319, 2019, doi: 10.1016/j.ijinfomgt.2018.08.015.

[4] Y. Liu, J. Fan, J. Zhang, X. Yin, and Z. Song, “Research on telecom customer churn prediction based on ensemble learning,” J. Intell. Inf. Syst., vol. 60, no. 3, pp. 759–775, 2023, doi: 10.1007/s10844-022-00739-z.

[5] M. J. Shabankareh, M. A. Shabankareh, A. Nazarian, A. Ranjbaran, and N. Seyyedamiri, “A Stacking-Based Data Mining Solution to Customer Churn Prediction,” J. Relatsh. Mark., vol. 21, no. 2, pp. 124–147, 2022, doi: 10.1080/15332667.2021.1889743.

[6] Z. M. S. Mohammed, F. A. Abdalla, M. Salih, A. F. A. Mahmoud, and A. Satty, “A Comprehensive Bibliometric Analysis of Churn Prediction Research : an Essay on Trends , Key Contributors and Global Participation in The Field .,” vol. 36, no. 4, pp. 219–228, 2024.

[7] Z. Liu, J. C. R. Alcantud, K. Qin, and L. Xiong, “The Soft Sets and Fuzzy Sets-Based Neural Networks and Application,” IEEE Access, vol. 8, pp. 41615–41625, 2020, doi: 10.1109/ACCESS.2020.2976731.

[8] H. K. Thakkar, A. Desai, S. Ghosh, P. Singh, and G. Sharma, “Clairvoyant: AdaBoost with Cost-Enabled Cost-Sensitive Classifier for Customer Churn Prediction,” Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/9028580.

[9] M. Ganaie, M. Hu, A. K. Malik, and M. Tanveer, “Ensemble deep learning : A review,” no. October, 2022, doi: 10.1016/j.engappai.2022.105151.

[10] P. Chen, N. Liu, and B. Wang, “Evaluation of Customer Behaviour with Machine Learning for Churn Prediction: The Case of Bank Customer Churn in Europe,” 2023, doi: 10.4108/eai.28-10-2022.2328450.

[11] M. Z. Alotaibi, “Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods,” no. June, pp. 1–8, 2024, doi: 10.48084/etasr.7480.

[12] M. Bogaert, “Ensemble Methods in Customer Churn Prediction : A Comparative Analysis of the State-of-the-Art,” 2023.

[13] M. Z. Alotaibi and M. A. Haq, “Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods,” Eng. Technol. Appl. Sci. Res., vol. 14, no. 3, pp. 14572–14578, 2024, doi: 10.48084/etasr.7480.

[14] R. Mareay, M. Ali, and T. Medhat, “Soft Sets and Soft Topological Spaces via its Applications,” Kafrelsheikh J. Inf. Sci., vol. 3, no. 1, pp. 1–7, 2022, doi: 10.21608/kjis.2022.246103.

[15] M. Rai, B. Ghimire, N. Uprety, and R. Shrestha, “A Comparative Analysis of Data Balancing Techniques : SMOTE and ADASYN in Machine Learning for Enhancing Bank Loan Default Risk Predictions,” vol. 155, no. 1, pp. 159–168, 2024.

[16] M. K. Awang, M. Makhtar, N. Udin, and N. F. Mansor, “Improving Customer Churn Classification with Ensemble Stacking Method,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 11, pp. 277–285, 2021, doi: 10.14569/IJACSA.2021.0121132.