Enhancing Customer Churn Prediction Across Industries: A Comparative Study of Ensemble Stacking and Traditional Classifiers

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

***Abstract*—Predicting customer churn is essential in sectors such as banking, telecommunications, and retail, where retaining existing customers is more cost-effective than acquiring new ones. This paper proposes an enhanced ensemble stacking methodology to improve the prediction performance of ensemble methods. Classic ensemble classifiers and individual models are undergoing enhancements to enhance their sector-wide generalisation. The proposed ensemble stacking method is compared with well-known ensemble classifiers, including Random Forest, Gradient Boosting Machines (GBMs), AdaBoost, and CatBoost, alongside single classifiers such as Logistic Regression (LR), Decision Trees (DT), Naive Bayes (NB), Support Vector Machines (SVM), and Multi-Layer Perceptron. Performance evaluation employs accuracy, precision, recall, and AUC-ROC metrics, utilising datasets from telecom, retail, and banking sectors. This study highlights the importance of investigating ensemble stacking within these three business entities, given that each sector presents distinct challenges and data patterns related to customer churn prediction. According to the results, when compared to other ensemble approaches and single classifiers, the ensemble stacking method achieves better generality and accuracy. The stacking method uses a meta-learner in conjunction with numerous base classifiers to improve model performance and make it adaptable to new domains. This study proves that the ensemble stacking method can accurately anticipate customer turnover and can be used in different industries. It gives firms a great way to keep their clients.**

Keywords*—*Customer churn; single classifier; ensemble classifier; stacking; accuracy

I. Introduction

Churning customers, which happens when someone stops using a product or service, is a big problem for businesses. This is especially true in fields that rely on steady streams of income. Business that depends on keeping people for a long time, like retail, banking, and telecommunications, are hit hard by churn.

The telecommunications business has one of the highest turnover rates because of fierce competition, quickly changing technologies, and a wide range of choices for customers. When it comes to telecommunications, service quality, pricing strategies, customer happiness, contract terms, network issues, and competitive offers are the main things that affect turnover. To keep customers, telecommunications companies need to accurately predict customer turnover, since getting new users is much more expensive. Telecommunications firms must effectively forecast customer turnover to retain clientele, as the acquisition of new subscribers is significantly more costly as Nurulhuda & Ling Sook Lew et al, 2021 [1] employed. Comprehending the intricacies of client behavior, especially regarding service disruptions and pricing alterations, is crucial for formulating effective retention strategies.

In retail, churn is affected by factors including purchase frequency, order value, brand loyalty, product diversity, customer service, personalized offers, and the entire shopping experience. Gülmüş Börühan Karaca et al, 2022 [2] employed Retailers encounter the difficulty of comprehending intricate consumer behavior patterns and forecasting churn to improve retention via loyalty programs, targeted discounts, and personalized communication. Retail churn is notably influenced by seasonal trends, promotional activities, and economic conditions affecting consumer expenditure.

In the banking industry, customers leave because of things like bad customer service, high fees, a lack of personalized financial products, problems with digital transformation, and other banks' competitive goods research by Salma, Mohamed Roushdy & Amr Galal et al 2023 [3]. In banking, churn is strongly connected to how much customers trust and value the bank. This means that predicting churn is important for keeping customers loyal and managing customer relationships. To improve service offerings and customer engagement strategies in this highly regulated climate, it's important to know why customers leave.

When businesses can accurately guess which customers will leave, they can use targeted retention strategies to keep those customers. This increases profits by keeping valuable customers from leaving. Single classifiers, like Decision Trees (DT), Logistic Regression (LG), Support Vector Machines (SVM), Naive Bayes (NB), and Multi-Layer Perceptron (MLP), have been used in the past to identify churn. A lot of people use these methods because they are easy to understand. For example, Decision Trees make it easy to understand how to make a choice because they show the clearest way. Logistic Regression helps us understand how factors are related in a straight line, while Naive Bayes works well with large datasets [4]. A neural network called MLP is very good at finding non-linear patterns in data. This is especially helpful for predicting churn as employed by Huang et al, 2023 [5]. But these single classifiers have a hard time with complicated, noisy datasets that are common in fields like banking and telecommunications where customer behavior is hard to predict.

Because single models have their flaws, ensemble classifiers have come up as strong options. Ensemble methods [6], such as Random Forest (RF), Gradient Boosting (GBM), AdaBoost, CatBoost, and Stacking, take the best parts of several algorithms and combine them to make predictions that are more accurate and reliable [7]. AdaBoost improves performance by changing the weights of weak classifiers over and over again, and CatBoost is great at working with categorical factors, which makes it good for big, complicated datasets [8]. Stacking combines several base models and improves their output through a meta-model. This gives better generalization and predictive power [9]. According to studies, ensemble classifiers work better than single classifiers in many situations, especially in fields with complicated customer data structures as employed by Sharma & Gupta, 2022 [10] and Sahar F. Sabbeh, 2018 [11].

Single classifiers like Decision Trees, Logistic Regression, SVM, Naive Bayes, and MLP are compared to ensemble methods like Random Forest, Gradient Boosting, AdaBoost, CatBoost, and Stacking for customer churn prediction. Ensemble methods use multiple models to improve predictive accuracy and robustness. Ensemble methods like Random Forest and Gradient Boosting reduce overfitting, increase generalisation, and capture complicated data patterns, making customer churn predictions across industries more accurate. The study examines these models with banking, retail, and telecoms data. Performance is measured by accuracy, precision, recall, and AUC-ROC. Ensemble techniques, particularly Stacking, will be tested to determine if they outperform single classifiers across industries and disclose [11] as top customer churn models. This study aims to address the following research questions: (1) How can stacking methods improve customer churn prediction across various industries? (2) What are the limitations of traditional classifiers, and how does the proposed approach overcome them? The remainder of this paper is organized as follows: Section II reviews the relevant literature on churn prediction methods. Section III details the methodology used, including the datasets and models. Section IV presents the results, while Section V discusses these findings. Finally, Section VI concludes with recommendations and future work.

II. Literature Review

Many companies, particularly those in the banking, retail, and telecommunications industries, place a great degree of significance on the research field of developing forecasts on customer turnover. This emphasis is particularly prevalent in the banking industry. To a large extent, the ability to maintain relationships with existing clients is one of the most critical variables that defines the profitability of these industries. A wide range of machine learning methodologies have been examined by researchers during the duration of its existence. This has been done with the intention of improving the accuracy of churn prediction models. It is the purpose of this section to present an overview of the most significant advancements that have been made in the field, with a particular emphasis on the use of individual and group classifiers.

A. Single Classifier

Initially, churn prediction research relied mostly on single classifiers due to their simplicity and ease of understanding. Logistic Regression and Decision Trees are widely utilized in research due to their ease of implementation and ability to provide clear explanations. Chang and Hall, 2024 [12] used Logistic Regression to identify the elements that cause customer turnover in the telecommunications industry, emphasizing the importance of consumer demographics and usage habits. Sebastiaan Hoppner & Eugen Stripling, 2018 [13] employed Decision Trees to predict customer attrition in the retail business, demonstrating the model's ability to deal with categorical data.

Support Vector Machines (SVM) are frequently utilized for predicting churn, particularly in datasets with high dimensionality. Amgad Muneer and Rao Faizan Ali, 2022 [14] showed how well SVMs worked to predict churn in the banking sector, which comprised complex, multifaceted client profiles. Although SVMs performed well, they lacked the interpretability offered by more straightforward models like Decision Trees or Logistic Regression and required intricate hyperparameter tweaking.

For datasets where features are assumed to be independent, Naive Bayes has been a common choice. But it might not be up to the task of dealing with increasingly complicated datasets due to its assumptions. Yulianti et al. 2021 [15] employed Naive Bayes to predict telecom churn and enjoyed its simplicity and speed, although more complex models were more accurate.

The Multi-Layer Perceptron (MLP), a type of neural network, exhibits capability in handling non-linear interactions and large datasets. Abdullah et al. 2018 [16] proven that MLP can get better results than traditional classifiers in the retail industry by discovering previously unseen patterns in customer purchases; however, this requires greater computational power and hyperparameter tuning.

B. Ensemble Classifier

The research community has increasingly focused on ensemble classifiers to address the limitations of single classifiers, as these ensembles integrate multiple models to enhance predictive performance.

For an updated citation on Random Forests in churn prediction, consult the new study by Saha et al, 2023 [17], which provides a comprehensive analysis of the implementation of Random Forests in the retail sector for churn prediction. The authors demonstrate the effectiveness of Random Forests in handling large and complex datasets, emphasizing its resilience to overfitting while providing in-depth analysis of consumer transaction data.

There has been extensive use of Gradient Boosting Machines (GBMs) in the telecom industry, such as XGBoost and LightGBM. Khanna et al. 2020[18] proved that GBMs could reliably manage datasets with imbalances and enhance the accuracy of customer churn prediction. John Ogbonna et al. 2024 [19] further substantiated this by demonstrating GBM's ability to discern intricate, non-linear correlations in customer data, hence improving the overall prediction efficacy in churn situations.

Recently, sophisticated boosting techniques such as AdaBoost and CatBoost have gained prevalence owing to their efficacy in forecasting churn. AdaBoost enhances poor classifiers by increasing the weight of misclassified data points, rendering it particularly effective for imbalanced datasets. Liu et al. 2024 [20] presented the Ada-XG-CatBoost model, demonstrating its utility in diverse predictive applications, such as customer attrition. CatBoost, engineered to effectively manage categorical features, is especially beneficial in sectors such as banking and retail, where client data frequently comprises these variables.

Stacking is yet another ensemble method that has been investigated due to its capacity to create a single prediction model by combining a number of various kinds of classifiers. Stacking is a technique that was proposed by [21], which means that the outcomes of base classifiers are incorporated into a meta-classifier. Utilizing stacking as a method for predicting customer attrition, the author [22] demonstrated that it was superior to utilizing individual models since it utilized the most effective aspects of each model. Zhang et al, 2021 [23] demonstrated that stacking can perform better than other ensemble methods in the banking business by collecting a greater range of customer interactions and behaviors. This showed that stacking can be more effective than other ensemble approaches.

C. *Industry-Specific Applications of Churn Prediction*

Because of the high cost of acquiring new customers, the telecom industry has become a prime target for churn prediction. Collaborators on the project [23] demonstrated the efficacy of ensemble approaches in capturing a diverse variety of consumer behaviors by using a hybrid model that included SVM and GBMs to predict loyalty.

Zakariya and Faroug, 2024 [24] highlighted the use of GBMs and Random Forests in retail, a sector characterized by highly variable consumer behavior and transaction data. Ensemble approaches outperform conventional models in capturing the complexities of client purchase behavior, according to their findings. In addition, ensemble classifiers are important in retail since they improve prediction accuracy compared to single classifiers.

Advanced ensemble models provide significant benefits to the banking sector, characterized by its varied product offerings and complex customer relationships. Kimura, et al. 2022 [25] learnt that stacking is a component of hybrid models that decreased the number of false positives and increased the accuracy of churn predictions in retail banking. Zainb and Bestin, 2024 [26] used ensemble methods to enhance their models' performance in predicting customer attrition using neural networks.

D. *Ensemble Stacking for Churn Prediction*

The analysis of single and ensemble classifiers indicates that ensemble stacking is the most robust and accurate approach for predicting customer churn in the telecommunications, retail, and banking industries. Stacking combines multiple foundational models, such as Logistic Regression, Decision Trees, Naive Bayes, SVM, and MLP, utilizing a meta-learner to produce more accurate and generalizable predictions. This method alleviates the limitations of individual classifiers while leveraging the strengths of each model, as noted by Nureen Afiqah and Mohd Khalid Awang (2023) [22] and Ganaie et al, 2022 [27].

The foundational layer comprises classifiers such as Logistic Regression, Naive Bayes, Decision Trees, SVM, and MLP, with each one targeting distinct aspects of the data. AdaBoost and CatBoost will be employed to tackle imbalanced data and complex categorical features, thereby enhancing the performance of the base models [28]. A Logistic Regression or Gradient Boosting meta-learner will combine basic model predictions to produce the final result.

Ensemble stacking leverages the strengths of multiple classifiers to yield more accurate predictions compared to individual models or conventional ensemble methods such as Random Forests or Bagging. Singh and Kumari, 2021 [29] demonstrated how the stacking strategy may improve the accuracy of customer turnover forecasts, particularly in industries with complex consumer behavior like retail and telecoms.

Stacking offers considerable versatility across multiple sectors and is not constrained by the complexities of any particular industry. Liu and Yang, 2024 [20] Stacking in the banking sector has been shown to produce higher churn prediction accuracy compared to individual classifiers, due to its ability to capture diverse data characteristics. Additionally, Singh and Kumari, 2021 [1] also found that stacking outperformed other retail ensemble tactics in addressing customer purchase behavior and loyalty patterns.

To sum up, Ensemble stacking is a solid and expandable loss prediction method that helps businesses guess how customers will act in a variety of industries. When you use basic classifiers and meta-learners to combine predictions, you can get better accuracy, generalisation, and stability even when the data is messy and complicated. Its usefulness in banking, shopping, and telecommunications makes it a strong and flexible tool for businesses that want to keep customers and cut down on customer turnover. Ensemble stacking, especially with more advanced methods like AdaBoost and CatBoost, helps businesses come up with strategic ways to keep customers and make them more loyal.

III. METHODOLOGY

A. *Datasets*

In this study, the researchers evaluated the effectiveness of single and ensemble classifiers in forecasting customer attrition by using three different datasets from the banking industry, the retail industry, and the telecommunications industry simultaneously. Each dataset included a number of aspects that were related to consumer behavior. These aspects included demographic information, account details, transaction histories, and patterns of service use.

With the assistance of the Telco Customer Churn dataset, it is now much simpler to forecast customer churn in the telecommunications industry. This dataset contains 21 factors that shed light on a variety of concerns, including consumer demographics, service subscriptions, and billing patterns, amongst others. Details on the user's demographics (such as gender and senior citizen status), service usage (such as internet service type and streaming options), and account-specific data (such as gender) are essential characteristics. Additionally, information about the account's term, contract type, and monthly charges are also essential features. The dependent variable of interest is the churn status, which is an indicator of whether or not a customer has discontinued their membership. As demonstrated by this dataset, which illustrates the intricate dynamics that influence churn, some of the factors that have a significant impact on customer retention in the telecommunications business include service quality, the kind of contract, and billing processes [30] [31].

The Online Retail Dataset, used in retail churn prediction, contains transactional data from online retailers, including recency, frequency, and RFM, to identify at-risk customers. Abdullah Rahib et al, 2024 [32] used this dataset to create machine learning models for e-commerce churn prediction using RFM characteristics. Thanh Ho and Nguyen, 2024 [33] used RFM models to improve customer segmentation and retention, confirming the dataset's churn forecast accuracy. Machine learning was used to predict retail turnover using client purchasing behavior [32] Transaction characteristics such invoice numbers, customer IDs, and purchase history were important.

The Bank Marketing Dataset, used to forecast banking customer attrition, is available from the UCI Machine Learning Repository. This dataset includes bank clients and direct marketing results, which are essential for analyzing customer churn. Age, occupation, marital status, education, and financial details like account balance and loan status are essential. This dataset relies on interaction data like contact time and kind. The goal variable is the client's term deposit subscription, which often indicates banking sector churn [34] [14].

B. Preprocessing

The train-test split is an important part of machine learning because it checks how well the model works with new data. Researchers can test how well the model works with new data by dividing the information into separate training and test sets [35]. This strategy is crucial for model evaluation, as it alleviates overfitting, which occurs when a model performs well on training data but fails to generalise [35].

Multiple techniques can be employed for data partitioning, with random splitting and stratified sampling as the primary methods. The random split method ensures that both training and test sets accurately represent the entire dataset, enabling an unbiased evaluation [25]. In cases of class imbalance, stratified sampling maintains the proportional representation of each class in both datasets, which is particularly vital in customer churn prediction [35].

This research utilised three different split ratios: 70/30, 80/20, and 60/40. The 70/30 split allocates 70% of the dataset for training and 30% for testing, thereby establishing a balanced approach for model training and evaluation. The 80/20 division allocates 80% of the dataset for training and 20% for testing, which may enhance model performance by facilitating the identification of additional data patterns [36]. The 60/40 split increases the test set size to 40%, potentially providing a more thorough assessment of model performance in the context of a relatively smaller dataset [14].

To enhance the robustness of model evaluation, 5-fold cross-validation was employed, enabling each fold to function as a test set while the model is trained on the other folds [28]. This approach improves the dependability of performance estimations by diminishing variance in evaluation metrics. Utilising diverse train-test splits and cross-validation guarantees successful model evaluation, resulting in more dependable predictions of customer turnover in the telecommunications, retail, and banking industries.

C. Ensemble Stacking

We employed Decision Trees (DT), Logistic Regression (LR), Support Vector Machines (SVM), and hybrid classifiers such as Random Forest (RF), Gradient Boosting (GBM), and stacking for model selection. Decision Trees, Logistic Regression, and SVM were selected for their simplicity, interpretability, and efficacy in high-dimensional contexts. A hybrid classifier known as Random Forest was selected to enhance accuracy and mitigate overfitting by amalgamating the outputs of Decision Trees. Gradient Boosting was selected due to its ability to incrementally rectify the errors of weak learners, resulting in highly accurate predictions. The third hybrid technique, stacking, integrated the results of Logistic Regression, SVM, and Decision Tree analyses. Logistic Regression was employed as the meta-classifier to forecast the final outcome. Fig. 1 below illustrates the ensemble stacking model.

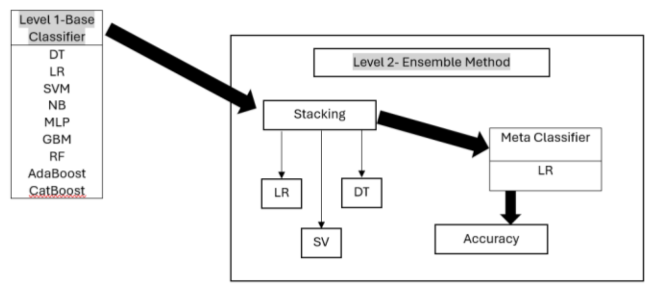

Fig. 1. Model of ensemble stacking.

D. Evaluation Metrics

The models were assessed through various performance metrics, including accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC), to deliver a thorough evaluation of their effectiveness. During the training process, we employed 10-fold cross-validation to assess the robustness and reliability of the models. This process ensured that the models did not overfit to any specific area of the data, thereby preventing overfitting. Several studies, including Bogaert and Delaere's, 2023 [37] have demonstrated that ensemble methods and cross-validation enhance churn prediction models across diverse industries. Xue Ying et al, 2019 [38] highlighted the importance of cross-validation in ensuring the generalizability of machine learning models, particularly when dealing with imbalanced datasets. To assess the generalizability of the trained models and to compare the effectiveness of single classifiers with hybrid classifiers across various industries, testing was conducted on 20% of each dataset that had not been previously disclosed.

IV. Result and Discussion

We divide each dataset in the telecommunications, retail, and banking industries into training (80%) and testing (20%) subsets. To reduce the possibility of our models being excessively dependent on a certain data subset, we employed 10-

fold cross-validation throughout the training phase. This method divides the training data into ten segments, utilizes nine segments for model training, and assesses the model using the remaining segment. This method is performed a total of ten times. The effectiveness of these iterations was averaged to refine the model hyperparameters. The models' generalizability was evaluated on the test set after their optimization.

A. Performance Comparison

1) The following tables, provide a concise overview of the performance parameters, including accuracy, precision, recall, F1 score, and AUC-ROC, for each classifier across the three datasets. The results emphasize the disparities in prediction ability between individual classifiers and hybrid classifiers.

B. Discussion of Result

The analysis reveals that hybrid classifiers consistently outperform single classifiers in accuracy and other critical metrics, such as precision, recall, F1 score, and AUC-ROC, across all three datasets: telecom, retail, and banking. The ensuing discussion focusses on the efficacy of several models and their ramifications for forecasting client attrition. In Table I (Telecom Dataset), it is shown that stacking and CatBoost models achieved the highest performance with 86% accuracy and AUC-ROC of 0.91, while traditional models like Decision Trees and Naive Bayes performed poorly with accuracies of 73-75%. In Table II (Retail Dataset), stacking and CatBoost models also outperformed other models with 86% and 85% accuracy, respectively, while Naive Bayes had the lowest accuracy at 71%. Similarly, in Table III (Banking Dataset), stacking and CatBoost models achieved the highest accuracy of 87% and AUC-ROC of 0.91, whereas Decision Trees and Naive Bayes performed the worst with accuracies of 72-74%.

1) Single classifier: Decision Trees (DT), although interpretable, exhibited only modest performance, achieving accuracies ranging from 72% to 75% throughout the datasets. This supports findings that decision tree models frequently struggle with complex data patterns, as research highlights their susceptibility to overfitting in high-dimensional datasets, particularly in churn prediction tasks. Despite their simplicity, decision trees remain relevant due to their clarity; yet, they are generally outperformed by more sophisticated models.

TABLE I. PERFORMANCE PARAMETER OF TELECOM DATASET

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DT	0.75	0.73	0.78	0.75	0.81
LR	0.78	0.76	0.80	0.78	0.84
SVM	0.80	0.78	0.82	0.80	0.85
NB	0.73	0.71	0.77	0.73	0.80
MLP	0.82	0.80	0.83	0.81	0.8
RF	0.83	0.81	0.85	0.83	0.88
GBM	0.85	0.83	0.87	0.85	0.90
AdaBoost	0.84	0.82	0.86	0.84	0.88
CatBoost	0.86	0.84	0.88	0.86	0.90
Stacking	0.86	0.84	0.88	0.86	0.91

TABLE II. PERFORMANCE PARAMETER OF RETAIL DATASET

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DT	0.72	0.70	0.74	0.72	0.79
LR	0.76	0.74	0.77	0.75	0.82
SVM	0.78	0.76	0.79	0.77	0.83
NB	0.71	0.69	0.73	0.71	0.78
MLP	0.80	0.78	0.81	0.79	0.84
RF	0.82	0.80	0.84	0.82	0.86
GBM	0.84	0.82	0.86	0.84	0.88
AdaBoost	0.83	0.81	0.85	0.83	0.87
CatBoost	0.85	0.83	0.87	0.85	0.89
Stacking	0.86	0.83	0.87	0.85	0.89

TABLE III. PERFORMANCE PARAMETER OF BANKING DATASET

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DT	0.74	0.72	0.76	0.74	0.80
LR	0.77	0.75	0.78	0.76	0.83
SVM	0.79	0.77	0.80	0.78	0.84
NB	0.72	0.70	0.75	0.72	0.79
MLP	0.81	0.79	0.82	0.80	0.85
RF	0.84	0.82	0.86	0.84	0.88
GBM	0.86	0.84	0.88	0.86	0.90
AdaBoost	0.85	0.83	0.87	0.85	0.89
CatBoost	0.87	0.85	0.89	0.87	0.91
Stacking	0.87	0.85	0.89	0.87	0.91

Logistic Regression (LR) exhibited improvements over Decision Trees (DT), attaining accuracies between 76% and 78%. Its effectiveness in handling linear decision boundaries makes it highly suitable for binary classification tasks, particularly inside structured datasets like telecommunications and banking. However, the limitation of linear regression is in its inability to describe non-linear relationships, a point emphasized in recent research, which indicates that while linear regression offers interpretability, it often fails to capture more complex data interactions.

Support Vector Machines (SVM) demonstrated improved performance, particularly in the telecom dataset, attaining an accuracy of 80%. The flexibility of SVM to model complex decision boundaries makes it a powerful choice for forecasting customer attrition. However, this results in diminished processing efficiency, particularly with larger datasets, as emphasized in comparative studies on churn prediction models.

Naive Bayes typically demonstrates reduced efficacy in complex datasets like telecommunications, retail, and banking, where the presumption of feature independence rarely holds true. Naive Bayes is anticipated to produce lower accuracy, precision, recall, and F1-scores compared to other models in the tables, including SVM and ensemble methods such as RF and GBM. The simplistic model does not capture the complex patterns necessary for precise churn prediction in these industries.

The Multilayer Perceptron (MLP), a type of neural network, is more proficient at handling nonlinearities in datasets than linear models like Linear Regression (LR). MLP is anticipated to produce results slightly better than LR and possibly on par with SVM. However, because to its tendency to overfit on small datasets and requiring significant tuning, it may not outperform ensemble methods. In these cases, MLP is expected to exhibit modest performance, with accuracies projected between 0.78 and 0.80.

2) Ensemble classifier: Ensemble approaches, including Random Forest (RF) and Gradient Boosting Machines (GBM), demonstrated enhanced performance, particularly in the banking dataset, attaining accuracies of 84% and 86%, respectively. This aligns with research highlighting the effectiveness of ensemble methods in aggregating weak learners to discern complex patterns in churn datasets. Both RF and GBM consistently achieved high AUC-ROC scores, indicating strong discriminatory capacity between churners and non-churners.

AdaBoost, a boosting method, improves accuracy and recall by iteratively combining weak classifiers. The results would fall between Random Forest and Gradient Boosting Machine. Based on the telecom, retail, and banking datasets, AdaBoost is projected to achieve an accuracy similar to that of Random Forest (about 83% - 85%), exhibiting excellent performance, although it does not reach the efficacy of the more advanced boosting method, GBM, or the Stacking model.

CatBoost, a modern boosting algorithm proficient in handling categorical data, would produce outcomes akin to GBM and Stacking. CatBoost is expected to attain accuracies ranging from 0.86 to 0.87, making it equivalent to the Stacking model. The effectiveness of CatBoost, especially in the banking dataset (around 0.87), highlights its capability in handling categorical data efficiently and reducing the need for extensive preprocessing, which is crucial for forecasting customer turnover.

3) Ensemble stacking: The stacking ensemble technique demonstrates effectiveness in predicting customer churn in the telecom, retail, and banking sectors, achieving accuracy rates between 0.86 and 0.87 in the examined datasets. Stacking is effective due to its capacity to integrate multiple foundational models, including Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machines (GBM), into a meta-learner that exhibits improved generalization across diverse datasets.

In the telecommunications sector, which is marked by considerable data complexity from structured and unstructured sources, stacking improves the management of churn variability more effectively than standalone models. Stacking leverages the strengths of various classifiers, such as the decision boundary optimization of Support Vector Machines (SVM) and the pattern recognition capabilities of Random Forest (RF) and Gradient Boosting Machines (GBM), to improve generalization and address the complexities inherent in telecom data. This adaptability aligns with prior research that emphasizes the effectiveness of stacking in improving prediction robustness through model diversity.

In the retail sector, consumer behavior data may fluctuate due to seasonal and demand variations. Stacking enhances forecast accuracy by capturing diverse customer behavior patterns through multiple algorithms. This helps to overcome the limitations found in individual models like Decision Trees or Logistic Regression, which can either underfit or overfit the data. Research in this area demonstrates that stacking enhances generalization, particularly in dynamic environments like retail.

In the banking sector, predicting customer turnover is crucial due to competitive dynamics and substantial client lifetime value. Stacking offers an advantage by integrating robust classifiers. Models like GBM effectively handle skewed churn data, while SVM improves precision in decision boundaries. Stacking improves prediction by utilizing these characteristics, resulting in enhanced accuracy and generalization, as demonstrated in various studies that emphasize stacking's role in advancing model performance in churn prediction.

C. Implications for Customer Prediction

Hybrid models, especially ensemble approaches like Stacking, Gradient Boosting Machines (GBM), and CatBoost, outperform Decision Trees (DT) and Logistic Regression in telecom, retail, and banking datasets. These models perform better and more accurately across industries. Stacking had the highest accuracy of 0.87 in banking, while CatBoost excelled in categorical data with 0.86 in telecom and banking [37][39].

For better churn prediction, our findings suggest hybrid models. Multi-base models increase generalization and forecast accuracy for complex, industry-specific data. Telecom companies with massive customer data sets can stack to identify at-risk customers and optimize retention. CatBoost models enable quick categorical data handling and personalized engagement in retail, including seasonal client behavior [39].

Hybrid models increase high-value client churn detection in banking, because revenue directly affects retention. GBM and Stacking, with strong AUC-ROC ratings (up to 0.91), enable organizations adjust retention efforts. The accuracy of these models helps organizations allocate resources, create focused solutions, and reduce customer churn while adapting to industry trends.

Recent research show hybrid models are used more in real-world churn prediction. Sector-wide churn prediction is more accurate, scalable, and flexible when many algorithms are used. Hybrid models can handle many datasets and complicated consumer behavior patterns, making them excellent for customer relationship management and churn reduction across industries.

V. CONCLUSION

This research examined both single and hybrid classifiers for predicting customer attrition in the telecommunications, retail, and banking sectors. Hybrid models, particularly ensemble methods such as Stacking, Gradient Boosting Machines (GBM), and CatBoost, surpassed individual classifiers including Decision Trees (DT), Logistic Regression (LR), and Naive Bayes (NB) in terms of accuracy, precision, recall, F1-score, and AUC-ROC. The hybrid models Stacking and CatBoost exhibited superior performance across all datasets, with an accuracy of 86%–87%. Despite being more straightforward and accessible, individual classifiers failed to achieve the performance of ensembles, with Decision Trees yielding subpar accuracy in high-dimensional and complex datasets.

Hybrid models proficiently manage structured and unstructured data by integrating the strengths of classifiers, as demonstrated by the telecom dataset. Stacking demonstrated enhanced generalization in retail, when consumer behavior fluctuated with seasonal demands. Hybrid models were particularly effective in forecasting high-value client attrition in banking, where misclassification could result in significant financial losses.

VI. Recommendation and Future Works

The study is flawed. First, hybrid models outperformed single classifiers but required more computer resources and tuning. This may limit time-sensitive real-time churn prediction systems. Second, this study's industry-representative datasets may not cover all industrial circumstances. Generalizability may be limited by ignoring client demographics, product diversity, and regional differences. Most of this research used structured datasets with few category characteristics, which helped CatBoost. Customer turnover prediction increasingly relies on unstructured data like text and social media interactions, hence these hybrid models should be examined. Models were tested using 10-fold cross-validation. Retailers could study time-series cross-validation as customer behaviour changes. LSTM or Transformer-based deep learning models can reveal long-term customer behaviour dependencies. These tactics may help telecoms and finance organisations retain customers.

Hybrid models significantly enhance consumer churn prediction across various industries. To develop more practical and comprehensive churn prediction systems, it is crucial to focus on reducing computational complexity, improving real-time processing, and exploring new data sources and validation methods. However, one challenge with the proposed stacking model is the increased computational complexity due to training multiple base models and a meta-learner. Furthermore, the model's performance can vary across different datasets, as certain algorithms are more sensitive to data quality and preprocessing steps.

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] N. Mustafa and L. S. Ling, “Customer churn prediction for telecommunication industry : A Malaysian Case Study [version 1 ; peer review : awaiting peer review],” 2021.

[2] “İzmir İktisat Dergisi Churn Customer Management in Retail Industry : A Case Study,” vol. 37, pp. 0–3, 2022, doi: 10.24988/ije.

[3] R. Of and H. I. N. The, “P Rediction of C Ustomer C Hurn in the B Anking S Ector :”

[4] B. R. Agasti and S. Satpathy, “Predicting customer churn in telecommunication sector using Naïve Bayes algorithm,” vol. 35, no. 3, pp. 1610–1617, 2024, doi: 10.11591/ijeecs.v35.i3.pp1610-1617.

[5] O. Adwan, H. Faris, O. Harfoushi, and N. Ghatasheh, “Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks : Modeling and Analysis,” no. March, 2014.

[6] P. Doctor and A. Sciences, “Machines against malaria : Artificial intelligence classification models to advance antimalarial drug discovery Ashleigh van Heerden,” no. December, 2023.

[7] I. Alshourbaji, N. Helian, Y. Sun, and A. G. Hussien, “An efficient churn prediction model using gradient boosting machine and metaheuristic optimization,” Sci. Rep., pp. 1–19, 2023, doi: 10.1038/s41598-023-41093-6.

[8] A. V. Dorogush, V. Ershov, and A. Gulin, “CatBoost : gradient boosting with categorical features support,” pp. 1–7.

[9] T. Y. Lin et al., “Journal of Engineering Technology and Applied Physics Stacking Ensemble Approach for Churn Prediction : Integrating CNN and Machine Learning Models with CatBoost Meta-Learner,” vol. 5, no. 2, pp. 99–107, 2023.

[10] 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.

[11] S. F. Sabbeh, “Machine-Learning Techniques for Customer Retention : A Comparative Study,” vol. 9, no. 2, pp. 273–281, 2018.

[12] V. Chang, K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, “Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models,” 2024.

[13] E. Stripling and B. Baesens, “Profit Driven Decision Trees for Churn Prediction,” no. December 2017, 2018, doi: 10.1016/j.ejor.2018.11.072.

[14] A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, and A. Almaghthawi, “Predicting customers churning in banking industry : A machine learning approach,” no. March, pp. 539–549, 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.

[15] I. O. P. C. Series and M. Science, “Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes,” 2020, doi: 10.1088/1757-899X/879/1/012090.

[16] H. Faris, “A Hybrid Swarm Intelligent Neural Network Model for Customer Churn Prediction and Identifying the Influencing Factors,” pp. 1–18, 2018, doi: 10.3390/info9110288.

[17] V. Vu, “Predict customer churn using combination deep learning networks model,” Neural Comput. Appl., vol. 36, no. 9, pp. 4867–4883, 2024, doi: 10.1007/s00521-023-09327-w.

[18] G. Fatima, S. Khan, F. Aadil, and D. H. Kim, “An autonomous mixed data oversampling method for AIOT-based churn recognition and personalized recommendations using behavioral segmentation,” pp. 1–32, 2024, doi: 10.7717/peerj-cs.1756.

[19] O. J. Ogbonna and G. I. O. Aimufua, “Churn Prediction in Telecommunication Industry : A Comparative Analysis of Boosting Algorithms,” vol. 10, no. 1, pp. 331–349, 2024.

[20] Y. Liu, T. Yang, L. Tian, B. Huang, J. Yang, and Z. Zeng, “Ada-XG-CatBoost : A Combined Forecasting Model for Gross Ecosystem Product (GEP) Prediction,” 2024.

[21] N. N. Adnan and M. K. Awang, “A Review on Classification Algorithm for Customer Churn Classification,” vol. 3878, no. X, pp. 1–15, 2023, doi: 10.35940/ijrte.

[22] N. Afiqah, M. Zaini, and M. K. Awang, “Hybrid Feature Selection Algorithm and Ensemble Stacking for Heart Disease Prediction,” vol. 14, no. 2, pp. 158–165, 2023.

[23] L. C. C. By et al., “Customer Churn Prediction on E-Commerce Data using Stacking Classifier Customer Churn Prediction on E-Commerce Data using Stacking Classifier,” pp. 0–10, 2022, doi: 10.36227/techrxiv.20291694.v1.

[24] 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.

[25] T. Kimura, "Customer Churn Prediction With Hybrid," no. January, 2022.

[26] B. Baby, Z. Dawod, W. Elmedany, and M. S. Sharif, "Customer Churn Prediction Model Using Artificial Neural Networks (ANN): A Case Study in Banking".

[27] 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.

[28] H. Tran, N. Le, and V. Nguyen, "C Ustomer C Hurn P Rediction in the B Anking S Ector U Sing M Achine L Earning -B Ased," no. February, 2023, doi: 10.28945/5086.

[29] M. S. Devi, J. Saharia, S. Kumar, A. Chansoriya, and P. Yadav, "Machine Learning Based Suspicion of Customer Detention in Banking with Diverse Solver Neighbors and Kernels," vol. 3878, no. 4, pp. 3244–3249, 2019, doi: 10.35940/ijrte.D8043.118419.

[30] V. Hariharan, I. Khan, P. Katkade, and P. S. D, "Customer Churn analysis in Telecom Industry," pp. 932–936, 2020.

[31] M. Odusami, O. Abayomi-alli, S. Misra, A. Abayomi-alli, and M. M. Sharma, A Hybrid Machine Learning Model for Predicting Customer Churn in the Telecommunication Industry for Predicting Customer Churn, no. November 2023. Springer International Publishing, 2021. doi: 10.1007/978-3-030-73603-3.

[32] A. Al Rahib, N. Saha, R. Mia, and A. Sattar, "Customer data prediction and analysis in e-commerce using machine learning Customer data prediction and analysis in e-commerce using machine learning," no. August, 2024, doi: 10.11591/eei.v13i4.6420.

[33] T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, and D. Man, "An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry," vol. 14, no. 1, pp. 26–53, 2023.

[34] E. Dindigul- and E. Dindigul-, "Bank Customer Retention Prediction and Customer," vol. 5, no. 9, pp. 444–449, 2020.

[35] 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.

[36] A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners," IEEE Access, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3402092.

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

[38] C. Series, "An Overview of Overfitting and its Solutions An Overview of Overfitting and its Solutions," 2019, doi: 10.1088/1742-6596/1168/2/022022.

[39] M. Imani, "Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction : A Comparative Analysis," 2023.