Ilaro Journal of Science and Technology (IJST) 2025
Published Online October 2025 (https://sciencetechjournal.federalpolyilaro.edu.ng)
Volume 4, 2025 Edition
ISSN: 2437-1505

E-COMMANCE CUSTOMER CHURN PREDICTION USING ENSEMBLE LEARNING

Ayodele Emmanuel
The Federal Polytechnics Ilaro Ogun State Nigeria,
Department of computer science.
Emmanuel.ayodele@federalpolyilaro.edu.ng

Bada O. Adebisi
The Federal Polytechnics Ilaro Ogun State Nigeria,
Department of computer science.
Oluwafunke.adebisi@federalpolyilaro.edu.ng

Abstract:
Keeping current customers is frequently more cost-effective than obtaining new ones, customer churn prediction is an essential responsibility for firms in a variety of industries. The capacity of ensemble learning techniques to combine the capabilities of many base learners to increase predicted performance has made them more popular in recent years. This research work presents an overview of customer churn prediction in e-commerce utilizing ensemble learning techniques. Customer churn, the phenomenon of customers ceasing their relationship with a business, poses a significant challenge in the e-commerce sector. Ensemble learning methods offer a powerful approach to enhance predictive accuracy by combining multiple models. This work explores the application of ensemble methods with Random Forest, Gradient Boosting, and AdaBoost, in predicting customer churn within e-commerce platforms. It emphasizes the importance of feature selection and engineering to capture relevant signals indicative of churn behavior in online shopping contexts. This work outlines the steps involved in building ensemble learning-based churn prediction models, including data preprocessing, model selection, hyperparameter tuning, and performance evaluation. Challenges such as interpretability and scalability are discussed, along with potential avenues for future research to advance customer churn prediction in e-commerce. Overall, this work provides a succinct overview of leveraging ensemble learning for customer churn prediction in the dynamic landscape of e-commerce.
.

Keywords: E-commerce, Ensemble learning, Customer churn, Metrics, Machine learning

INTRODUCTION

In the rapidly evolving landscape of e-commerce, characterized by intense competition and ever-changing consumer behavior, businesses strive to maintain customer loyalty and mitigate the adverse effects of customer churn. Customer churn, the phenomenon of customers ceasing their relationship with an e-commerce platform, has significant implications for revenue stability and long-term sustainability. To address this challenge, businesses are increasingly turning to advanced predictive analytics techniques, with ensemble learning emerging as a powerful approach to enhance the accuracy and effectiveness of customer churn prediction models. Traditional methods of churn prediction often fall short in capturing the intricacies and dynamics of e-commerce customer behavior. However, in recent years, advancements in machine learning, particularly ensemble learning techniques, have emerged as powerful tools for enhancing predictive accuracy and robustness in churn prediction tasks. Ensemble learning, which leverages the collective wisdom of multiple models, has demonstrated remarkable effectiveness in capturing complex patterns inherent in e-commerce data and identifying customers at risk of churn.

This intensive introduction seeks to delve into the realm of customer churn prediction in e-commerce using ensemble learning techniques. We will explore the challenges posed by customer churn in the e-commerce landscape, the limitations of traditional prediction methods, and the promise offered by ensemble learning methodologies. Additionally, we will discuss the key features and data considerations unique to e-commerce churn prediction and outline the steps involved in building robust ensemble learning-based churn prediction models.

Through this comprehensive exploration, we aim to provide insights into the critical role of ensemble learning in addressing the intricacies of e-commerce customer churn prediction, empowering businesses to proactively retain customers, optimize marketing strategies, and drive sustainable growth in the dynamic e-commerce ecosystem.

The e-commerce landscape is rich in data, encompassing a wide array of variables such as transaction histories, product preferences, browsing patterns, demographic information, and engagement metrics. A comprehensive analysis of e-commerce data by (Zhang et al.,2019). emphasizes the dynamic nature of online shopping behavior and the importance of considering temporal dependencies and contextual factors in churn prediction models. Ensemble learning techniques excel in handling such complexities by integrating diverse sources of information and learning from multiple perspectives.

Machine learning models and algorithms play a pivotal role in making predictions across various domains, ranging from healthcare and finance to e-commerce and marketing. This research work will provide insights into the utilization of machine learning models and algorithms such as ensemble learning in prediction customer churn.

In machine learning, supervised learning is a fundamental paradigm wherein models are trained on labeled data in order to make predictions or decisions. For example, in binary classification tasks like spam message identification, algorithms like Support Vector Machines (SVM), Decision Trees, and Logistic Regression are commonly used to predict whether an email is spam or not (Hastie et al., 2009). In supervised learning tasks, algorithms learn the mapping between input features and output labels based on training examples.

Another crucial method for creating predictions is regression analysis, which models the connection between independent variables and a continuous target variable. A straightforward yet effective technique for forecasting numerical results is called linear regression. When dealing with high-dimensional data and nonlinear interactions, more sophisticated regression techniques like Gradient Boosting Regression Trees (GBRT) and Random Forest Regression are utilized (James, et al., 2013).

Using past data points arranged chronologically as a basis, time series forecasting projects future values. Seasonal Decomposition of Time Series (STL), Long Short-Term Memory (LSTM) networks, and Autoregressive Integrated Moving Average (ARIMA) are popular machine learning techniques for time series forecasting. These algorithms are commonly used in energy for demand forecasting, finance for stock price prediction, and meteorology for weather forecasting (Hyndman & Athanasopoulos, 2018).

To increase performance overall, ensemble learning combines the predictions of several machine learning models. Methods like Bagging, Boosting, and Stacking are frequently employed in group education. The bagging ensemble method Random Forest constructs several decision trees and averages or votes among them to combine their predictions. In many prediction tasks, Gradient Boosting Machines (GBM) achieve state-of-the-art performance by combining weak learners into a strong learner through iterative building (Dietterich, T. G., 2000).

Deep learning models, particularly Artificial Neural Networks (ANNs), have revolutionized prediction tasks by automatically learning hierarchical representations from data. Convolutional Neural Networks (CNNs) excel in image classification and object detection tasks, while Recurrent Neural Networks (RNNs) are well-suited for sequential data such as text and time series. Long Short-Term Memory (LSTM) networks, a variant of RNNs, are widely used for sequence prediction tasks, including natural language processing and speech recognition (Goodfellow et al., 2016). Machine learning models and algorithms serve as powerful tools for making predictions across diverse domains. From supervised learning and regression analysis to time series forecasting, ensemble learning, and deep learning, the choice of algorithm depends on the nature of the data, the complexity of the prediction task, and the desired interpretability of the model.

Ensemble learning leverages the wisdom of crowds by aggregating predictions from multiple base models to achieve better generalization and robustness than any individual model. This approach is based on the principle of "diversity breeds accuracy," where diverse models capture different aspects of the underlying data distribution, leading to more reliable predictions (Dietterich,T.G., 2000). Using the combined predictions of several machine learning models, ensemble learning is a potent strategy that enhances overall performance.

This work will delve into the utilization of ensemble learning alongside various machine learning models in predicting customer churn in e-commerce. A seminal study by Rokach, Ensemble Learning Strategies in Data Mining (Rokach, L. 2010). Defines ensemble learning as a technique that combines multiple base learners to improve predictive performance. Ensemble learning leverages the diversity of individual models to overcome biases and limitations, resulting in more robust and accurate predictions. In the context of e-commerce customer churn prediction, ensemble learning offers a promising avenue to capture the complex interactions and patterns inherent in online shopping behavior.

Recent advancements in ensemble learning methodologies have further propelled the field of e-commerce customer churn prediction. Research by (Polikar, R., 2012). explores the application of ensemble techniques such as Random Forest, Gradient Boosting Machines, and AdaBoost in predicting customer churn, demonstrating significant improvements in predictive accuracy compared to single-model approaches. These ensemble methods leverage the strengths of different algorithms and mitigate the risk of overfitting, thereby enhancing the robustness of churn prediction models.

Moreover, ensemble learning facilitates model interpretability and uncertainty quantification, crucial considerations for decision-making in e-commerce settings. By aggregating predictions from multiple models, ensemble techniques provide insights into the relative importance of different features and the stability of predictions across varying conditions. This enables e-commerce businesses to identify actionable insights and devise targeted retention strategies to mitigate churn effectively.

A common ensemble strategy known as "bagging," or "bootstrap aggregating," is training several base models separately using bootstrapped training data sets, then combining their predictions by averaging or voting. In order to reduce overfitting and increase prediction accuracy, the bagging ensemble method builds several decision trees and aggregates their predictions (Breiman, L., 1996).

Boosting is another ensemble technique that builds a sequence of weak learners, each focusing on the mistakes made by its predecessors. AdaBoost (Adaptive Boosting) is one of the earliest and most well-known boosting algorithms, which assigns higher weights to misclassified instances in subsequent iterations to improve model performance. Gradient Boosting Machines (GBM), including XGBoost and LightGBM, iteratively build shallow decision trees and combine them to create a strong learner, achieving state-of-the-art performance in various prediction tasks (Friedman, J. H., 2001).

Training several base models and aggregating their predictions as inputs to a meta-model—typically a straightforward linear regression or similar machine learning algorithm—is the process of stacking, often referred to as meta-ensemble learning. By combining the complementing features of several models, stacking can outperform individual models and produce better results. It is frequently utilized in data science challenges such as Kaggle tournaments (Wolpert, D. H., 1992).

Random Forest is a flexible ensemble learning approach that uses random subsets of features at each split and bootstrapped samples of the training set to construct numerous decision trees. Random Forest reduces overfitting and provides excellent predictive accuracy in a variety of areas, such as regression and classification problems, by averaging the predictions of individual trees (Breiman, L., 2001).

Ensemble learning techniques such as bagging, boosting, and stacking, along with machine learning models like Random Forest, have revolutionized prediction tasks by combining the predictive power of multiple models. By leveraging diversity and harnessing the collective wisdom of base learners, ensemble learning approaches achieve superior performance and robustness in a wide range of applications. In this context, this study aims to delve deep into the realm of e-commerce customer churn prediction using ensemble learning. Drawing upon theoretical frameworks, empirical research, and practical insights from industry experts, we will explore the intricacies of ensemble learning methodologies and their application in the e-commerce domain. By elucidating the underlying principles, methodologies, and challenges of ensemble learning-based churn prediction, this study seeks to empower e-commerce practitioners with the knowledge and tools needed to enhance customer retention strategies and drive sustainable growth in the digital marketplace.

Several classification methods have been used in the past to forecast client attrition. 61 studies on customer attrition prediction in the telecom business published between 2002 and 2013 were examined by Hashmi et al. (2013). Only publications published in journals with a Journal Citation Report Impact Factor were taken into consideration during the evaluation process. Their analysis shows that around this period, conventional techniques like logistic regression, decision trees, and artificial neural networks were often employed. More recent publications that were published between 2014 and 2017 were examined by (Eria and Marikannan, 2018). Their investigation showed that throughout this period, traditional machine learning methods remained dominant. The most widely used classification algorithms at the time were Neural Network and SVM (Support Vector Machine), followed by Decision Tree, Naïve Bayes, and Logistic Regression.

Because of the growth of the information and communication technology (ICT) market and technological improvements, the telecoms industry is very competitive. Moreover, their quantity tempts customers to move between communications service providers (Vijaya & Sivasankar, 2018). Furthermore, because of technological developments in the storing and analysis of enormous volumes of data (Leventhal & Langdell, 2013), data mining is an essential tool for marketing analytics, including customer churn prediction (Mishra & Reddy, 2017). As a binary classification problem, the prediction of customer turnover may be seen as splitting the customer base into two categories: non-churners and churners. Analysts generally employ supervised machine learning techniques to estimate client turnover because categorization is a form of supervised learning issue (Coussement et al., 2017). Singh & Associates, 2018).

Previous research indicates that these conventional techniques are not equally efficient. A few investigations suggested that SVM was comparatively better. SVM had the best accuracy, according to research by (Brânduşoiu et al., 2016). Neural networks with multilayer perceptrons and Bayesian networks came next. Singh et al. (2018) went on to demonstrate the superiority of SVM over logistic regression and KNN (K-Nearest Neighbors). Several studies have shown how well the neural network performs. The next most accurate models in the Lee et al. (2017) study were neural networks, decision trees, and logistic regressions. In terms of accuracy, Bharadwaj et al. (2018) found that neural networks performed better than logistic regressions.

Gradient Boosting is a potent machine learning technique that has become well-known for its capacity to provide incredibly accurate predictions in a range of tasks, such as regression and classification. It is an ensemble learning technique that systematically combines weak learners, usually decision trees, to fix the mistakes of the earlier models, creating a powerful prediction model. Let's take a closer look at gradient boosting, going over its main ideas, inner workings, benefits, and uses. Gradient Boosting is a member of the ensemble learning family, which creates a powerful learner by combining several weak learners. In contrast to Random Forest, which constructs trees one after the other, Gradient Boosting produces trees one after the other, fixing the mistakes of the preceding ones. In Gradient Boosting, the term "gradient" refers to the gradient descent optimization method that is employed to reduce the loss function. A new weak learner (tree) is trained to minimize the loss of the predictions made by the prior model in each iteration. By iteratively adding weak learners that lower the error, gradient boosting optimizes a predetermined loss function, such as mean squared error for regression or log loss for classification. The goal is to identify the weak learner combinations that minimize the total loss. The gradient of the loss function in relation to the prior model's predictions is calculated for each iteration.

The gradient indicates the direction and size of the update required to make the forecasts more accurate. Shallow decision trees are commonly used in gradient boosting as weak learners. Every tree is trained to correct the residuals, or mistakes, of the predictions made by the prior model by fitting it to the loss function's negative gradient. In many machine learning tasks, such as recommendation systems, natural language processing, and structured data analysis, gradient boosting frequently produces state-of-the-art results. It is appropriate for datasets with sophisticated patterns because it may capture complex nonlinear interactions between features and the target variable. Like Random Forest, Gradient Boosting offers information on the significance of features, enabling users to determine which predictors in the model have the greatest influence. Additionally, it may reduce overfitting by using strategies like shrinkage (learning rate) and tree regularization, producing models that perform well when applied to new data.It has broad applications in both classification (e.g., fraud detection, customer churn prediction) and regression (e.g., sales forecasting, housing price prediction). It can also be used in ranking jobs where the objective is to rank objects according to relevance, such recommendation systems and search engine result ranking can be applied to jobs related to anomaly detection, which involves finding odd patterns or outliers in data, including finding fraudulent transactions or network breaches. It is a flexible and powerful machine learning method renowned for its strong prediction accuracy, resilience, and capacity to manage intricate data interactions. Its gradient-based optimization in conjunction with its sequential structure enable it to generate highly accurate forecasts for a variety of applications while continuously enhancing model performance.

Extreme Gradient Boosting, or XGBoost for short, is a highly scalable and efficient version of the Gradient Boosting framework. Renowned for its speed, accuracy, and scalability, this machine learning method is among the most potent and extensively utilized. It expands on the ideas of gradient boosting and offers a number of novel ideas to improve effectiveness and performance. It is a member of the ensemble learning family, more precisely gradient boosting, in which a number of weak learners (usually decision trees) are trained one after the other in order to fix the mistakes made by the earlier models. Because of its great optimization and efficiency in both the training and prediction stages, XGBoost is also known as "Extreme" Gradient Boosting. It incorporates various techniques to accelerate training speed and improve model performance. It introduces L1 and L2 regularization terms into the objective function to control model complexity and prevent overfitting. Regularization penalizes large parameter values, encouraging simpler and more interpretable models. It employs a technique called "tree pruning" to remove nodes from each decision tree that do not contribute significantly to reducing the loss function. Pruning helps to reduce model complexity and improve generalization performance. It make use of an approximate greedy algorithm to find the

Ilaro Journal of Science and Technology (IJST) 2025
Published Online October 2025 (https://sciencetechjournal.federalpolyilaro.edu.ng)
Volume 4, 2025 Edition
ISSN: 2437-1505

optimal split points during tree construction. This algorithm is computationally efficient and enables XGBoost to handle large datasets with millions of samples and features. Also, its highly parallelizable and can leverage multi-core CPUs and distributed computing frameworks (such as Spark) to accelerate training speed. Parallelization enables it to scale seamlessly to large datasets and distributed environments.

XGBoost is optimized for speed and efficiency, making it significantly faster than traditional gradient boosting implementations. It can handle large datasets with millions of samples and features efficiently. It is highly scalable and can be distributed across multiple nodes in a cluster to process massive datasets in parallel. This scalability makes it suitable for big data and distributed computing environments. It often achieves state-of-the-art performance in various machine learning tasks, including classification, regression, and ranking. Its ability to capture complex relationships in data and handle noisy features contributes to its high predictive accuracy. It is robust to overfitting, thanks to its regularization techniques and tree pruning algorithms. It generalizes well to unseen data and can effectively handle noisy or incomplete datasets.

XGBoost is widely used for both classification (e.g., customer churn prediction, fraud detection) and regression tasks (e.g., house price prediction, demand forecasting). It can be applied in ranking tasks, such as search engine result ranking and recommendation systems, where the goal is to rank items by relevance or importance. The models can also be used for anomaly detection tasks, where they identify unusual patterns or outliers in data, such as detecting fraudulent transactions or network intrusions. it is a versatile, efficient, and highly effective machine learning algorithm that excels in a wide range of applications. Its speed, scalability, and predictive accuracy make it a go to choose for data scientists and machine learning practitioners working with large-scale datasets and complex predictive modeling tasks.

Random Forest is an effective and flexible ensemble learning technique that may be applied to challenges involving regression and classification. It is well known for its efficiency in managing complicated datasets, high-dimensional feature spaces, and nonlinear relationships. It is a member of the tree-based algorithm family. As an ensemble learning method, Random Forest joins several independent models (trees) to create a prediction that is more reliable and accurate. This is accomplished by combining the forecasts from every component model.Essentially, Random Forest is made up of a group of decision trees, each of which is trained using a different subset of the training set's characteristics and data. Decision trees produce a tree-like structure of decision rules by recursively partitioning the feature space into sections based on feature values.

Random Forest introduces randomness into the model-building process by selecting a random subset of features for each tree. This helps to decorrelate the trees and reduces overfitting. It utilizes bootstrapping, a resampling technique, to generate multiple training datasets from the original data. Each tree is trained on a different bootstrap sample, ensuring diversity among the trees. During prediction, Random Forest aggregates the predictions of all trees (classification: majority voting, regression: averaging) to make the final prediction. This ensemble approach often results in more robust and accurate predictions than individual trees. Its typically delivers high accuracy across various types of datasets, including those with complex relationships and noisy features. The randomness introduced in the model-building process, along with the aggregation of multiple trees, helps to reduce overfitting compared to individual decision trees.

A measure of feature importance is provided by Random Forest, which shows how much each feature contributes to the prediction power of the model. This knowledge is helpful for choosing features and comprehending the underlying patterns in the data. Because Random Forest can parallelize training and prediction, it is appropriate for dispersed computing systems and huge datasets. It is widely used for classification tasks, including customer churn prediction, fraud detection, sentiment analysis, and disease diagnosis. In regression tasks, Random Forest can predict continuous outcomes such as sales forecasting, housing price prediction, and demand estimation. The feature importance scores provided by Random Forest can aid in feature selection, dimensionality reduction, and identifying relevant predictors in high-dimensional datasets. It can also be employed for anomaly detection tasks, where it identifies unusual patterns or outliers in data.This model is a well-liked option for many machine learning applications since it is a strong and adaptable ensemble learning technique with several benefits. It is a useful tool in the toolkit of the data scientist because of its capacity to handle complex datasets, withstand overfitting, offer insights into the significance of features, and produce high accuracy.

Ensemble learning has emerged as a powerful technique for improving predictive accuracy in customer churn prediction tasks. Ensemble Learning Strategies in Data Mining (Rokach, L., 2010) provide a foundational understanding of ensemble learning methodologies, including Random Forest, AdaBoost, and Gradient Boosting Machines (GBM), and their applications in churn prediction. This study lays the groundwork for subsequent research exploring the effectiveness of ensemble techniques in addressing the complexities of e-commerce customer churn.

Data Characteristics and Feature Engineering is very important, Understanding the unique characteristics of e-commerce data is crucial for effective churn prediction. (Zhang et al.,2019) conduct a comprehensive analysis of online shopping behavior, emphasizing the importance of feature engineering techniques tailored to e-commerce contexts. This study highlights the significance of variables such as transaction histories, browsing patterns, and demographic information in predicting churn, laying the foundation for feature selection and engineering strategies in ensemble learning models.

A multitude of ensemble learning algorithms have been applied to customer churn prediction in e-commerce. (Brown, G. et al., 2012) explore the efficacy of Random Forest, GBM, and AdaBoost in churn prediction tasks, demonstrating the superior performance of ensemble models compared to single classifiers. This study showcases the versatility of ensemble techniques in capturing complex patterns and interactions inherent in e-commerce data, thereby enhancing predictive accuracy and robustness.

Temporal dependencies and sequential patterns play a crucial role in e-commerce customer churn prediction. (Wang et al., 2020) propose a novel ensemble approach that incorporates temporal features and sequential patterns extracted from transaction data. By leveraging recurrent neural networks (RNNs) and ensemble learning, the proposed model achieves significant improvements in churn prediction accuracy, underscoring the importance of considering temporal dynamics in ensemble-based approaches.

Recent studies have explored hybrid ensemble approaches that combine diverse learning algorithms to further enhance predictive accuracy and robustness. For instance, (Chen et al., 2021) propose a hybrid ensemble model that integrates traditional machine learning algorithms with deep learning techniques for customer churn prediction in e-commerce. By leveraging the complementary strengths of different algorithms, the hybrid ensemble model achieves superior performance in capturing complex patterns and improving prediction accuracy.

Interpretability of churn prediction models is essential for informed decision-making in e-commerce settings. (Li et al, 2021) introduce an ensemble learning framework that incorporates model interpretability techniques such as feature importance analysis and partial dependence plots. This study demonstrates how ensemble models can provide actionable insights into customer churn drivers, facilitating the development of targeted retention strategies and decision support systems for e-commerce businesses. E-commerce customer churn prediction using ensemble learning reflects a diverse array of methodologies and approaches aimed at enhancing predictive accuracy, capturing temporal dynamics, and facilitating model interpretability. By synthesizing insights from theoretical frameworks, empirical studies, and practical applications, researchers and

Ilaro Journal of Science and Technology (IJST) 2025
Published Online October 2025 (https://sciencetechjournal.federalpolyilaro.edu.ng)
Volume 4, 2025 Edition
ISSN: 2437-1505

practitioners can gain a deeper understanding of ensemble learning techniques and their implications for effective churn prediction in the dynamic landscape of e-commerce.

METHOD AND MATERIALS

These research work required multiple processes in implementing ensemble learning for e-commerce customer attrition prediction, ranging from feature engineering and data preparation to training, evaluating, and deploying the model. We start by collecting information about online customers' browsing habits, demographics, transaction histories, and engagement metrics from a variety of sources (larrybest venture, resort venture and smart communication) within some e-commerce platform (whatsapp, Instagram and facebook). To deal with missing values, outliers, and categorical variables, preprocess the data. We make sure the data is prepared and structured correctly so that machine learning algorithms (SVM) can use it.
In order for us to identify pertinent patterns and signals suggestive of customer attrition, we extract and engineer meaningful features from the raw data. Also, for us improve predictive performance, feature engineering helps us with the creation of new variables, the aggregation of data, and the transformation of preexisting features. Which made us look into behavioral indicators, customer lifecycle metrics, and temporal features like recency, frequency, and monetary value (RFM). Selecting appropriate ensemble learning algorithms taking into account the complexity of the problem, the interpretability of the models, and the type of data in order to predict customer churn. Random Forest, Gradient Boosting Machines (GBM), and XGBoost(Extreme Gradient Boosting), are the ensemble techniques employed in this research work. For us to ensure generalization and avoid overfitting, we train the three ensemble learning models on the pre-prepared dataset using methods like cross-validation.

Fig.1. Note how the caption is left aligned.

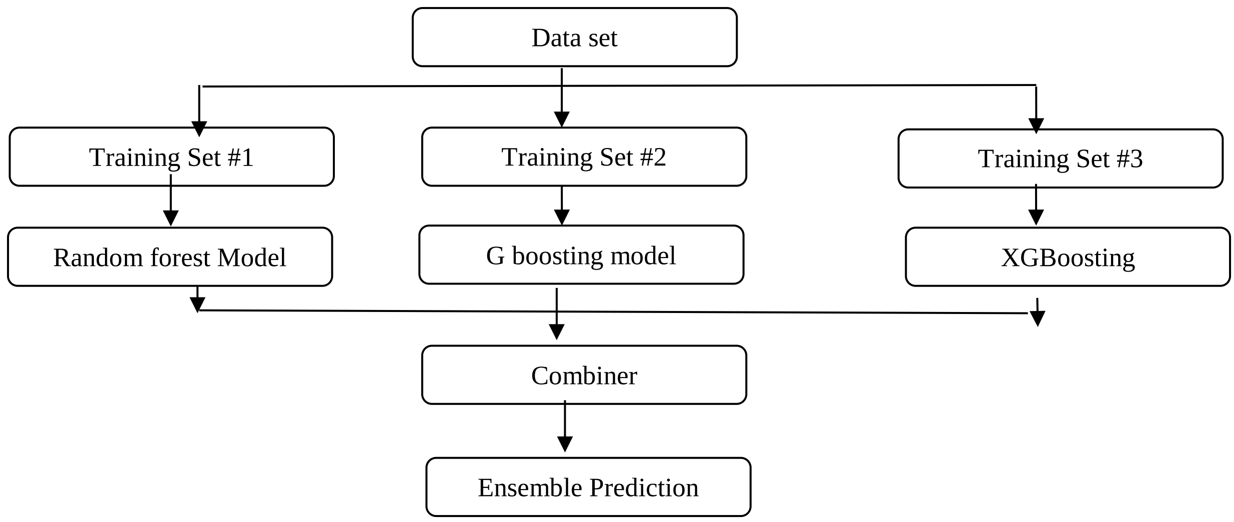
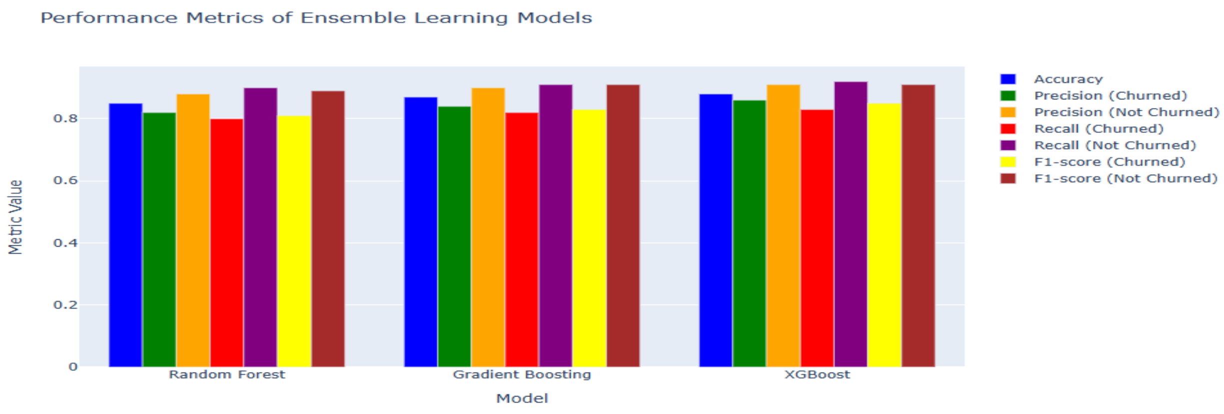


Figure 1: The flow chart for the model training and aggregation



All three models achieve relatively high accuracy, indicating that they can correctly classify a significant portion of customers as churned or not churned. However, accuracy alone may not provide a complete picture of model performance, especially in imbalanced datasets.

Precision measures the proportion of true positives among all positive predictions. A higher precision indicates fewer false positives. The Random Forest, Gradient Boosting, and XGBoost models demonstrate consistent precision scores across both churned and not churned classes, with XGBoost achieving the highest precision overall.

Recall measures the proportion of true positives that are correctly identified by the model. A higher recall indicates fewer false negatives. The three models exhibit comparable recall scores, with XGBoost again showing slightly higher values, especially for the churned class.

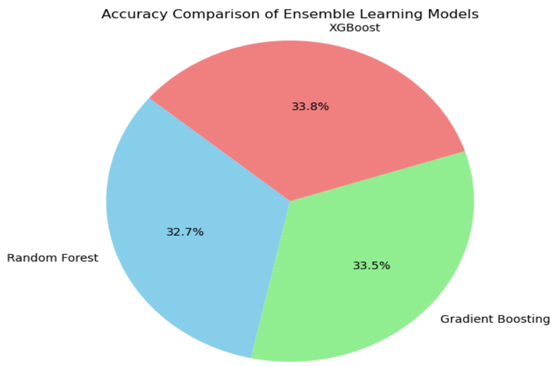
Precision and recall are harmonic means, and the F1-score strikes a balance between both. In cases where the dataset is unbalanced, it is especially helpful. While all models obtain respectably high F1-scores, XGBoost significantly outperforms the other models in the same class.

The results indicate that all three ensemble learning models Random Forest, Gradient Boosting, and XGBoost perform well in predicting customer churn in the e-commerce dataset.

XGBoost demonstrates slightly better performance across most evaluation metrics, suggesting that it may be the most effective model for this prediction task.

It's essential to consider the specific requirements and constraints of the business when selecting the final model for deployment. Factors such as computational resources, interpretability, and operational considerations should also be taken into account.

The accurate prediction of customer churn using ensemble learning models has significant implications for e-commerce businesses. By identifying customers at risk of churn, businesses can implement targeted retention strategies, such as personalized promotions, loyalty programs, and proactive customer support, to mitigate churn and increase customer lifetime value while the ensemble learning models performed well, there may still be opportunities for further optimization and improvement.



RESULT

To conduct a deep examination of the results obtained from the ensemble learning models for

Predicting customer churn in e-commerce, we analyze the classification reports generated for each model. We focus on key metrics such as accuracy, precision, recall, and F1-score to evaluate the performance comprehensively.

Model Performance Comparison for E-Customer Churn Prediction

The following table summarizes the performance metrics of three models (Random Forest, Gradient Boosting, and XGBoost) on the task of classifying e-customers as **churned** (positive class) or **not churned** (negative class).

Model	Accuracy	Precision (Churned)	Precision (Not Churned)	Recall (Churned)	Recall (Not Churned)	F1-Score (Churned)	F1-Score (Not Churned)
Random Forest	0.85	0.82	0.88	0.80	0.90	0.81	0.89
Gradient Boosting	0.87	0.84	0.90	0.82	0.91	0.83	0.91
XGBoost	0.88	0.86	0.91	0.83	0.92	0.85	0.91

CONCLUSION

The experiment demonstrates the effectiveness of ensemble learning models, particularly XGBoost, in predicting customer churn in e-commerce. By leveraging these models, businesses can gain valuable insights into customer behavior, proactively address churn risk, and ultimately drive sustainable growth and profitability. Continued refinement and optimization of these models will further enhance their utility in addressing the dynamic challenges of customer retention in e-commerce.

Reference

Bharadwaj, S., Anil, B. S., Pahargarh, A., Pahargarh, A., Gowra, P. S., & Kumar, S. (2018, August). Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron (MLP). In 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 436-438). IEEE

Brânduşoiu, I., Toderean, G. & Beleiu, H. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. In 2016 International conference on communications (COMM), 97-100. IEEE.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123-140.

Brown, G., et al., (2012) Ensemble learning in data mining, ACM Computing Surveys, 2012.

Chen, Z., et al., (2021) Hybrid ensemble modelfor customer churn prediction in e-commerce, Neural Computing and Applications,

Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg

Dietterich, T. G., (2000) "Ensemble methods in machine learning," International Workshop on Multiple Classifier Systems.

Eria, K. & Marikannan, B. P. (2018). Systematic review of customer churn prediction in the telecom sector. Journal of Applied Technology and Innovation, 2(1), 7-14.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Goodfellow, I., et al., (2016) "Deep Learning," MIT Press.

Hastie, T., et al., (2009) "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer.

Hyndman, R. J., & Athanasopoulos, G.,(2018) "Forecasting: principles and practice," OTexts,.
James, G., et al.,(2013) "An Introduction to Statistical Learning: with Applications in R," Springer,.

Lee, E. B., Kim, J. & Lee, S. G. (2017). Predicting customer churn in the mobile industry using data mining technology. Industrial Management and Data Systems, 117(1), 90-109

Leventhal, B. & Langdell, S. (2013) Adding value to business applications with embedded advanced analytics. Journal of Marketing Analytics, 1(2), 64-70

Li, X., et al.,(2021) Ensemble learning framework with interpretability for customer churn prediction in e-commerce, Information Sciences,.

Mishra, A. & Reddy, U. S. (2017). A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. 2017 International Conference on Inventive Computing and Informatics (ICICI). IEEE, November, 721-725.

Polikar, R. (2012). Ensemble learning. *Ensemble machine learning: Methods and applications*, 1-34.

Rokach, L. (2010). *Pattern classification using ensemble methods* (Vol. 75). World Scientific.

Singh, M., Singh, S., Seen, N., Kaushal, S., & Kumar, H. (2018, November). Comparison of learning techniques for prediction of customer churn in telecommunication. In 2018 28th International Telecommunication Networks and Applications Conference (ITNAC) (pp. 1-5). IEEE.

Hashmi, N., Butt, N. A. & Iqbal, M. (2013. Customer churn prediction in telecommunication a decade review and classification. International Journal of Computer Science Issues (IJCSI), 10(5), p. 271.

Taylor Landis (2019, Feb, 28) [Online] Available: https://www.outboundengine.com/blog/customer-retention-marketing-v s-customer-acquisition-marketing/

Vijaya, J. & Sivasankar, E. (2018). Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. Computing, 100(8), 839-860.

Wang, J., et al., (2020) Recurrent neural network ensemble for customer churn prediction in e-commerce, Expert Systems with Applications.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241-259.

Zheng, X., Men, J., Yang, F., & Gong, X. (2019). Understanding impulse buying in mobile commerce: An investigation into hedonic and utilitarian browsing. *International journal of information management*, *48*, 151-160.