Enhancing Customer Churn Prediction using Machine Learning and Deep Learning Approaches with Principal Component Analysis

Md Saidul Islam
Department of Computer Science and Engineering
Port City International University
Chattogram, Bangladesh

Taofica Amrine
Department of Computer Science and Engineering
Port City International University
Chattogram, Bangladesh

Tahmina Akter
Department of Computer Science and Engineering
Port City International University
Chattogram, Bangladesh

Muhammad Anwarul Azim
Department of Computer Science and Engineering
University of Chittagong
Chittagong-4331, Bangladesh

ABSTRACT
In this research effort, we present a comprehensive approach for predicting customer churn using a combination of traditional Ma- chine Learning and Deep Learning methodologies. The primary focus of this investigation centers on the crucial phase of Data Pre-Processing, involving fundamental tasks such as the handling of missing data, removal of duplicates, and the elimination of outliers. To enhance data quality and representation, techniques such as Data Transformation, Normalization, and Principal Com- ponent Analysis (PCA) have been employed. To tackle class im- balance, the method of Random Over-Sampling has been implemented. The process of Feature Extraction encompasses One- Hot Encoding and PCA, further enhancing data representation. Subsequently, a diverse set of predictive models has been evaluated, including Random Forest (RF), Support Vector Classifier (SVC), Gaussian Naive Bayes (GNB), Decision Tree (DT), XG- Boost (XGB), Logistic Regression (LR), Artificial Neural Net- work (ANN), Convolutional Neural Network (CNN), Long Short- Term Memory (LSTM), and Recurrent Neural Network (RNN). The results indicate that XGBoost surpasses other models, achieving an exceptional accuracy of 98.26%. Furthermore, a hybrid CNN & XGB model demonstrates an impressive accuracy of 97.53%.

General Terms
Customer Retention Analytics, Predictive Analytics for Customer Behavior

Keywords
Customer Churn Prediction, Principal Com- ponent Analysis, Data Pre-Processing, XGBoost, CNN, Customer Retention.

1. INTRODUCTION
In the contemporary business environment, customer retention stands as a paramount concern for enterprises striving to maintain sustainable growth and lasting success. Customer churn, referring to the loss of clientele, presents a significant obstacle across diverse industries. It not only exerts a detrimental influence on revenue streams but also prompts inquiries into service quality, customer contentment, and market competitiveness. In response to this challenge, the field of predictive analytics has emerged as a formidable tool, enabling organizations to proactively discern and mitigate the risk of customer attrition.

The process of customer churn prediction involves the adept application of advanced data analysis methodologies, including Machine Learning and Deep Learning, to extensive datasets comprising customer information. By gleaning invaluable insights from these data reservoirs, businesses can cultivate a more profound comprehension of customer behavioral trends and prefigure instances where customers are inclined to terminate their association with the company.

This prescient capability bestows companies with the power to strategize and take premeditated measures, whether through targeted marketing tasks, the personalization of customer interactions, or the proactive resolution of issues, in order to retain cherished clientele and heighten overall customer gratification. In this era characterized by data-informed decision-making, the capacity to forecast and manage customer attrition emerges as an essential competitive edge, and the following study delves into the method- ologies and approaches that serve as the bedrock of this indispensable practice.

2. LITERATURE REVIEW
The article Customer Churn Prediction Using Convolutional Neural Networks presents an intriguing approach to addressing the prob- lem of customer churn. The reported accuracy of 81.16 percent on Telecom's dataset is commendable, but the authors' decision to use a single algorithm raises questions about the possibility of further improving predictive performance. Future research in this domain should explore a wider range of algorithms to determine the optimal approach for enhancing customer churn prediction accuracy (Cenggoro et al.,2021).

To develop classification methods for the problem of customer churn in the telecommunications sector, the algorithms used in the study include Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), Light GBM, and XGBoost. They achieved 68.61% accuracy using the Light GBM model and utilized the Churn in Telecom dataset (Wang et al.,2020).

A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor

Identification in the Telecom Sector proposes a churn prediction model that initially classifies churn customer data using classification algorithms. The Random Forest (RF) algorithm performed well, achieving an 88.63% correct classification rate. The proposed churn prediction model is evaluated using metrics such as accuracy, precision, recall, f-measure, and receiver operating characteristics (ROC). The author used the South Asia GSM telecom service dataset as a private dataset(Ullah et al.,2019).

Determining the intervening effects of exploratory data analysis and feature engineering in telecoms customer churn modeling dis- cusses the effects of exploratory data analysis and feature engineering in telecoms customer churn modeling. The authors ap- plied seven classification techniques namely, Na¨ıve Bayes, Generalized Linear Model, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Artificial Neural Net- work and they achieved 86% accuracy. The authors used a public domain Telecoms dataset. The authors claimed that their proposed model outperformed other models by using machine learning techniques (Halibas et al.,2019).

The study analyzed customer behavior information from an actual water purifier rental company to develop and verify a churn prediction model. The primary contribution of this study lies in the development and validation of a churn prediction model, which can assist companies in retaining their customers and increasing revenue. The author utilized the ’Accounts of Water Purifiers’ dataset, which included approximately 84,000 customers for both training and testing. The author employed the RF and LGBM algorithms, achieving the highest accuracy of 88% with the LGBM algorithm (Suh and Youngjung, 2023).

The study aimed to predict customer churn in influencer commerce using the Decision Trees (DT) algorithm. This research contributes to the field of customer churn prediction in e-commerce from the perspective of influencers. The dataset used in this study was col- lected by an influencer marketing agency in Korea from August. The Decision Trees (DT) algorithm was applied to the dataset for predicting churning customers, achieving a maximum prediction accuracy of 82% (Kim et al.,2022).

To examine the prediction of customer churn in the banking sec- tor, a unique customer-level dataset from a large Brazilian bank was used. The author’s main contribution lies in exploring this rich dataset, which contains prior client behavior traits, enabling new insights into the main determinants predicting future client churn. The dataset from the large Brazilian bank included prior client be- havior traits, such as account balances, credit lines, loan applications, and other customer data. Several supervised machine learning algorithms were employed, including decision trees, k-nearest neighbors, logistic regression, support vector machine models, and the random forests technique, for churn prediction. Notably, the random forests technique outperformed other algorithms with an accuracy of 80.2% (de Lima Lemos et al.,2022).

The study focuses on predicting customer churn in the competitive telecom industry, particularly in the case of China Telecom. It employs various algorithm models, with a decision tree yielding the best results. Data preprocessing and feature selection were vital for accurate predictions. The research dataset included over 3,000 records with key variables like total day charge and international plan standing out. Additionally, the study highlights the significance of market segmentation as a strategic tool for tele- com companies to retain customers and optimize services (Xu et al.,2022).

3. METHODOLOGY

This chapter contains information on how workflow is being maintained. Fig 1 shows the methodology for customer churn prediction encompasses a series of vital steps designed to harness the power of machine learning and deep learning techniques. It will give a clear concept of the research. This will give knowledge about preprocessing, feature extraction, and system architecture and this chapter will end with giving the best model that we found for our system.

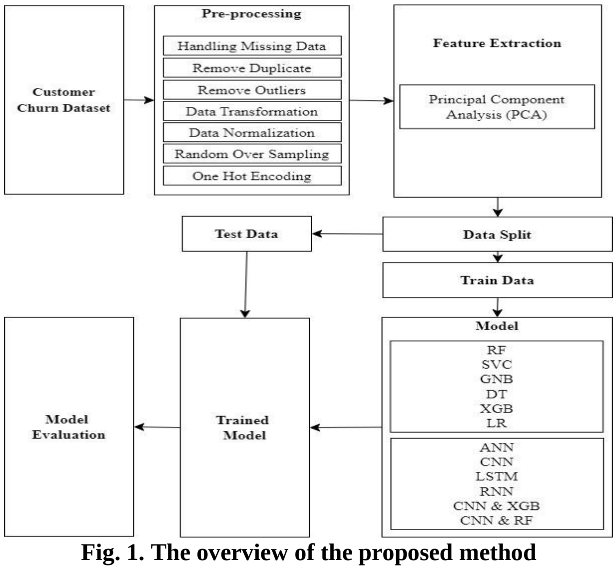

Fig. 1. The overview of the proposed method

3.1 Data Collection and Dataset

The ”Orange Telecommunication Dataset” is a publicly available dataset with 3,333 telecommunications customers’ data, including churn information. It originates from a prominent telecommunications company, offering insights into customer behavior and industry interactions. The dataset is rich, covering demographics, communication patterns, subscriptions, payments, and customer service interactions. It serves as a valuable resource for tasks like churn prediction, segmentation, marketing, and service optimization. Researchers and businesses can leverage it to enhance customer satisfaction and foster growth in the telecom sector.

3.2 Data Preprocessing

In the data preprocessing section, The Orange Telecommunication dataset is subjected to a thorough preparation process. This involves several key steps to enhance its quality and suitability for subsequent analyses. Firstly, missing data is addressed using advanced imputation techniques such as mean or median imputation, ensuring the integrity of the dataset for accurate insights into customer churn prediction and behavior analysis. Duplicate entries are re- moved to eliminate redundancy and improve data accuracy, con- tributing to more representative customer interaction patterns. Outliers are meticulously removed to enhance analytical accuracy by eliminating extreme values that could skew analyses. Data trans- formation optimizes usability by scaling, log-transforming, or standardizing features, aligning with modeling assumptions, and enhancing interpretability. Data normalization ensures that attributes with varying scales don't bias analyses, contributing to more robust insights for customer experience enhancement and service optimization. Random oversampling effectively addresses class imbalance by augmenting minority class instances, enhancing predictive model training, and preventing bias towards the majority class in customer churn prediction. Finally, one hot encoding is employed to convert categorical variables into a numerical format suitable for machine learning algorithms, ensuring the accuracy of analyses in the context of telecommunications data, and retaining relevant information from categorical attributes. These preprocessing steps collectively prepare the Orange Telecommunication dataset for insightful and reliable outcomes in the realm of telecommunications data analysis.

3.3 Feature Extraction

Feature extraction is conducted on the Orange Telecommunication dataset to distill relevant insights from complex data. Advanced techniques like Principal Component Analysis (PCA) are applied to reduce the dimensionality of the dataset while preserving its inherent patterns. By transforming the dataset into a more compact representation, feature extraction enhances computational efficiency and reduces noise, facilitating accurate analyses such as customer churn prediction and behavior modeling. This process aids in identifying key drivers of customer engagement and churn, enabling telecom companies to tailor strategies that foster customer loyalty and optimize services.

3.4 Split Dataset

The dataset is split into training and testing subsets to evaluate model performance accurately. Typically, around 70-80% of the data is allocated for training, enabling the model to learn patterns. The remaining 20-30% is used for testing, assessing the model's ability to generalize to new, unseen data. This division ensures that the model's effectiveness in customer churn prediction and behavior analysis is properly assessed, helping telecom companies make informed decisions based on reliable insights derived from their data.

3.5 Machine Learning and Deep Learning

Machine learning employs algorithms and data to predict customer churn by analyzing historical customer behavior. It identifies pat- terns and factors contributing to customer attrition, enabling businesses to proactively retain customers and minimize revenue loss. Deep learning, a subset of machine learning, employs neural net- works with multiple layers to detect intricate patterns in customer data for churn prediction. Its ability to handle complex, unstructured data, such as text and images, enhances accuracy in identifying at-risk customers, empowering businesses to take targeted retention actions.

3.6 Random Forest

In the domain of Customer Churn Prediction, we employed a Random Forest (Breiman et al., 2001) algorithm to enhance predictive accuracy. We then selected a subset of crucial features for input into the RF model. These essential features include account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and the number of customer service calls. This meticulous feature selection process ensured that the RF model focused on the most influential variables for making accurate predictions in our customer churn analysis.

3.7 Support Vector Classifier

In the process of Customer Churn Prediction, we employed the Support Vector Classifier (Pearl et al., 2000) algorithm as a means to enhance our predictive accuracy. Following this, we embarked on a meticulous curation of a subset of crucial features, specifically tailored for the SVC model's input. This curated selection encompassed pivotal variables such as account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and the count of customer service calls. This rigorous feature selection process was instrumental in channeling the focus of the SVC model toward the most influential variables, ensuring the precision of pre- dictions in our analysis of customer churn.

3.8 Gaussian Na¨ıve Bayes

In the sphere of Customer Churn Prediction, we employed the Gaussian Naive Bayes (GNB) (Duda et al., 2006) model, esteemed for its simplicity and efficacy. To enhance the model's predictive prowess, we thoughtfully selected and incorporated crucial features as input. These pivotal factors, spanning elements like ac- count length, international plan status, voice mail plan availability, total day calls, total day charge, and the number of customer service calls, empowered the GNB model to make insightful pre- dictions regarding customer churn. GNB's proficiency in handling probabilistic data and categorical attributes makes it an advantageous choice for this predictive venture.

3.9 Decision Tree

Decision Tree (Breiman et al., 1984) is a versatile and interpretable machine learning technique frequently harnessed in customer churn prediction. In this approach, data is recursively split based on key features, giving rise to a tree-like structure that provides intuitive insights into customer behavior and churn factors. This method excels in its ability to uncover decision paths and identify significant predictors, making it valuable for both analysis and actionable decision-making. Decision Trees not only offer transparency in understanding the logic behind churn predictions but also facilitate the discovery of critical patterns and drivers of customer attrition. Their adaptability and comprehensibility make them a noteworthy tool for organizations seeking to enhance customer retention strategies.

3.10 XGBoost

In our pursuit of Customer Churn Prediction, we harnessed the formidable capabilities of XGBoost (Chen et al., 2016)—an advanced gradient boosting algorithm. Renowned for its prowess, XGBoost excels in unraveling intricate patterns and elevating churn predictions. This is achieved through the iterative refinement of decision trees, culminating in a high-performance ensemble model. Beyond improving predictive accuracy, XGBoost offers invaluable insights by ranking feature importance, shedding light on the core drivers of churn. Our strategic selection of critical features, including account length, international plan status, voice mail plan avail- ability, total day calls, total day charge, and the count of customer service calls, fueled XGBoost's success. Its adaptability, efficiency, and reliability make XGBoost the favored choice among data scientists, propelling proactive customer retention strategies and delivering profound insights.

3.11 Logistic Regression

In our journey to predict Customer Churn, we turned to Logis- tic Regression (Cox et al., 1958)—a reliable algorithm renowned for its predictive prowess. With precision, LR deciphers patterns and amplifies churn predictions through statistical modeling. This involves meticulous analysis of the connections between the de- pendent variable (churn) and an array of independent features. Our strategic selection of vital features, spanning variables like account length, international plan status, voice mail plan availability, total day calls, total day charge, and customer service call counts, enriched the LR model's capabilities. Its transparency and robustness have solidified Logistic Regression as the go-to choice among data scientists, steering impactful customer retention strategies and in- formed decision-making.

3.12 ANN

In the landscape of Customer Churn Prediction, our approach lever- aged an Artificial Neural Network (Rumelhart et al., 1986) model. The ANN incorporated a Dense layer with 256 units and a dropout rate of 0.3 to enhance its generalization capabilities. Employing the binary crossentropy loss function for training, we supplied the ANN with pivotal features, including account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and customer service calls. This architecture empowered our ANN to discern intricate churn dynamics, facilitating precise predictions for effective customer retention strategies.

3.13 CNN

In our strategy for Customer Churn Prediction, we embraced an ID CNN model—a Convolutional Neural Network (CNN) variant. This CNN configuration featured a convolutional layer with 64 filters and a kernel size of 3, coupled with max-pooling using a window size of 2. To bolster its generalization capabilities, we introduced a dropout rate of 0.25. For training, we opted for the binary cross entropy loss function. The CNN was equipped with essential features, including account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and customer service calls. This tailored CNN architecture enabled us to effectively capture churn dynamics, facilitating accurate predictions.

3.14 LSTM

In the domain of Customer Churn Prediction, we employed an LSTM (Long Short-Term Memory) (Hochreiter et al., 1997) model known for its sequence processing capabilities. Our LSTM architecture featured 64 filters to capture intricate temporal patterns. For training, we utilized the binary cross entropy loss function and equipped the model with critical features, including account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and customer service calls. This LSTM framework enabled us to effectively analyze sequential data and make accurate churn predictions.

3.15 RNN

In our approach to Customer Churn Prediction, we implemented a Simple RNN (Jordan et al., 1997)model equipped with 64 filters, specifically designed for sequential data analysis. Training was optimized using the binary cross entropy loss function. The model's input consisted of crucial features, encompassing account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and customer service calls. This Simple RNN framework enabled us to uncover temporal dynamics, leading to precise churn predictions.

3.16 CNN & XGB Combined Model

In our approach to Customer Churn Prediction, we devised an innovative solution by combining the capabilities of CNN and XGB models. CNN was employed for classification, leveraging its pat- tern recognition prowess, while XGB was used for prediction. This hybrid model featured a seamless connection between the dense layer of CNN and XGB, ensuring efficient data transfer. We sup- plied the model with critical customer churn prediction features, encompassing account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and customer service calls. This pioneering fusion of deep learning and gradient-boosting techniques enabled accurate churn predictions. The following equation expressed the combination of CNN and XGBoost as follows:

$$\hat{y} = \mathrm{XGBoost}(f_{\mathrm{CNN}}(X))$$

where:

—$\hat{y}$ is the final prediction,

—XGBoost$(\cdot)$ represents the XGBoost model,

—$f_{\mathrm{CNN}}(X)$ represents the features extracted by the CNN from in- put X.

3.17 CNN & RF Combined Model

In our quest to predict Customer Churn, we pioneered a ground-breaking approach by merging the capabilities of CNN and RF models. CNN, renowned for its pattern recognition prowess, took charge of classification tasks, while RF excelled in prediction accuracy. This hybrid model seamlessly combined CNN's dense layer with RF, ensuring seamless data exchange. We armed the model with vital customer churn prediction features, including account length, international plan status, voice mail plan availability, total day calls, total day charge, total evening calls, total evening charge, total night calls, total night charge, total international calls, total international charge, and customer service calls. This innovative fusion of deep learning and ensemble methods delivered highly ac- curate and reliable churn predictions. The following equation rep- resents the combination of CNN and Rf:

$$\hat{y} = \mathrm{RF}\,(f_{\mathrm{CNN}}(X))$$

at 94.04%, and Recurrent Neural Network (RNN) at 95.41%. Notably, the Support Vector Classifier (SVC) and Random Forest

(RF) displayed significant enhancements after fine-tuning, underscoring the importance of hyperparameter optimization.

Additionally, we explored hybrid models, combining CNN with XGB and RF, which demonstrated impressive accuracy scores of 97.53% and 97.43%, respectively. These findings highlight the potential synergies between traditional and deep learning techniques.

It is important to acknowledge that our best result was achieved using an 80% training and 20% testing split, a commonly employed practice in machine learning experimentation. The choice of this split ratio can have a notable impact on model performance and is a significant consideration for future research efforts .

where:

—$\hat{y}$ is the final prediction,

—RF$(\cdot)$ represents the Random Forest model,

—$f_{\text{CNN}}(X)$ represents the features extracted by the CNN from in- put X.

3.18 Model Hyper Parameter Tuning

In the process of model hyperparameter tuning, key settings were meticulously adjusted to optimize performance, including the utilization of the Adam optimizer, a dense layer featuring 256 neurons to capture intricate data patterns, a small batch size of 4 for efficient model training, and a low learning rate of 0.0001. These hyperparameter adjustments were carefully implemented to fine-tune the model's convergence and enhance its predictive accuracy. The Adam optimizer, known for its efficiency and adaptability, aids in optimizing weight updates during training. The inclusion of a dense layer with 256 neurons enables the model to comprehend complex relationships within data. The choice of a small batch size and a low learning rate promotes stability in training and precise convergence, ultimately resulting in a well-optimized predictive system.

4 RESULT AND ANALYSIS

Table 4.1 presents the performance of various models in the context of customer churn prediction. In this study, we conducted a comprehensive evaluation, drawing inspiration from the work of Xu et al. (2022). Our primary focus was to propose and assess the efficacy of a Gaussian Naive Bayes (GNB) model with an accuracy of 87.86%. Additionally, we examined a range of other well-established classifiers, including Logistic Regression (LR) at 85.16%, Support Vec- tor Classifier (SVC) at 86.66%, Decision Tree (DT) at 88.91%, and Artificial Neural Network (ANN) at 85.46%.

However, our most notable contribution was the introduction of our proposed ensemble model, Extreme Gradient Boosting (XGB), which achieved an outstanding accuracy score of 98.26%. This highlighted the potential of ensemble techniques to substantially enhance predictive accuracy, setting a new benchmark in our study.

Furthermore, we fine-tuned our classifiers to optimize their performance, resulting in improved accuracy rates as follows: GNB at 91.66%, LR at 91.93%, SVC at 96.33%, DT at 96.15%, Random Forest (RF) at 98.17%, ANN at 92.48%, Convolutional Neural Network (CNN) at 96.97%, Long Short-Term Memory (LSTM)

Table 1 Comparative analysis

Model	(Xu et al., 2022) [?]	Our Approach
XGB	-	**98.26**
GNB	87.86	91.66
LR	85.16	91.93
SVC	86.66	96.33
DT	88.91	96.15
RF	-	98.17
ANN	85.46	92.48
CNN	-	96.97
LSTM	-	94.04
RNN	-	95.41
CNN & XGB	-	97.53
CNN & RF	-	97.43

4.1 Confusion Matrix of XGB

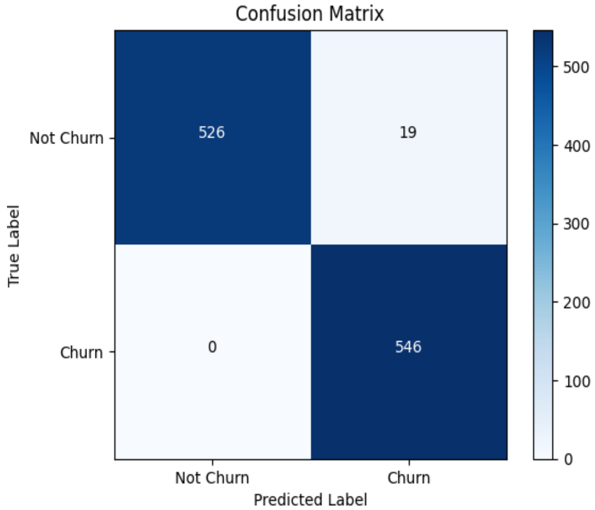

Fig. 2. Confusion matrix of XGB model.

Fig. 2 shows the XGBoost (XGB) confusion matrix for customer churn prediction, demonstrating its exceptional performance. In the 'not churn' class, it correctly classified 526 cases and had only 19 misclassifications. This indicates a very low error rate of approximately 2% for the 'not churn' class. For the 'churn' class, XGBoost correctly classified 546 cases and impressively had no misclassifications, resulting in a 0% error rate for the 'churn' class. These results showcase the high accuracy and predictive capability of the XGBoost model, especially in accurately identifying cases of customer churn while maintaining an extremely low error rate in the 'not churn' class.

4.2 Classification Report of XGB

```
              precision    recall  f1-score   support

           0       1.00      0.97      0.98       545
           1       0.97      1.00      0.98       546

    accuracy                           0.98      1091
   macro avg       0.98      0.98      0.98      1091
weighted avg       0.98      0.98      0.98      1091
```

Fig. 3. Classification Report of XGB model.

In Fig 3, the XGBoost (XGB) model excelled with 98.26% accuracy in Customer Churn Prediction. The precision for non-churn customers was perfect at 1.00, accompanied by a recall of 0.97 and an F1 score of 98, demonstrating accurate identification.

5 CONCLUSION

In this research effort, we have undertaken a comprehensive approach to address the intricate task of Customer Churn Prediction. Our study places a primary emphasis on the pivotal phase of Data Pre-Processing, encompassing essential procedures such as managing missing data, eliminating redundant entries, and detecting and rectifying outliers. These foundational steps lay the groundwork for reliable and precise predictive modeling. To elevate the quality and representativeness of our dataset, we systematically employed a suite of techniques, including Data Transformation, Data Normalization, and Principal Component Analysis (PCA). These preprocessing methodologies are vital in ensuring that our predictive models are fed with high-quality input, thereby facilitating improved predictive accuracy and generalization. Confronting the challenge of class imbalance, we implemented the Random Over-Sampling method, which aids in achieving a more balanced dataset, promoting fair and equitable model training. Furthermore, we delved into Feature Extraction techniques, such as One-Hot En- coding and PCA, to enrich data representation, rendering it more amenable for modeling. Our research rigorously evaluated a di- verse ensemble of predictive models, encompassing Random Forest (RF), Support Vector Classifier (SVC), Gaussian Naive Bayes (GNB), Decision Tree (DT), XGBoost (XGB), Logistic Regression (LR), Artificial Neural Network (ANN), Convolutional Neural Net- work (CNN), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN). The comprehensive analysis yielded compelling results, with XGBoost emerging as the standout performer, boasting an exceptional accuracy rate of 98.26%. Moreover, our exploration extended to hybrid models, notably the CNN+XGB architecture, which exhibited a remarkable accuracy of 97.53%. This innovative amalgamation harnesses the synergies of Convolutional Neural Networks and XGBoost, affording a potent framework that can significantly enhance predictive capabilities and offers great potential for practical applications in Customer Churn Prediction.

6 REFERENCES

[1] Cenggoro, T.W., Wirastari, R.A., Rudianto, E., Mohadi, M.I., Ratj, D. and Pardamean, B., 2021. Deep learning as a vec- tor embedding model for customer churn. Procedia Computer Science, 179, pp.624-631.

[2] Wang, X., Nguyen, K. and Nguyen, B.P., 2020, January. Churn prediction using ensemble learning. In Proceedings of the 4th international conference on machine learning and soft computing (pp. 56-60).

[3] Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S.U. and Kim, S.W., 2019. A churn prediction model using random for- est: analysis of machine learning techniques for churn predic- tion and factor identification in telecom sector. IEEE access, 7, pp.60134-60149.

[4] Halibas, A.S., Matthew, A.C., Pillai, I.G., Reazol, J.H., Delvo, E.G. and Reazol, L.B., 2019, January. Determining the inter- vening effects of exploratory data analysis and feature engi- neering in telecoms customer churn modelling. In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-7). IEEE.

[5] Suh, Y., 2023. Machine learning based customer churn pre- diction in home appliance rental business. Journal of big Data, 10(1), p.41.

[6] Kim, S. and Lee, H., 2022. Customer churn prediction in in- fluencer commerce: An application of decision trees. Procedia Computer Science, 199, pp.1332-1339.

[7] de Lima Lemos, R.A., Silva, T.C. and Tabak, B.M., 2022. Propension to customer churn in a financial institution: A machine learning approach. Neural Computing and Applica- tions, 34(14), pp.11751-11768.

[8] Xu, J., Li, X., He, Z. and Zhou, J., 2022. Early Warning of Telecom Customer Churn Based on Multialgorithm Model Optimization. Frontiers in Energy Research, 10, p.946933.

[9] Umayaparvathi, V. and Iyakutti, K., 2016. A survey on cus- tomer churn prediction in telecom industry: Datasets, meth- ods and metrics. International Research Journal of Engineer- ing and Technology (IRJET), 3(04).

[10] Gerpott, T.J., Rams, W. and Schindler, A., 2001. Customer re- tention, loyalty, and satisfaction in the German mobile cellu- lar telecommunications market. Telecommunications policy, 25(4), pp.249-269.

[11] Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C., 2015. A comparison of machine learning techniques for customer churn prediction. Simulation Mod- elling Practice and Theory, 55, pp.1-9.

[12] Lu, N., Lin, H., Lu, J. and Zhang, G., 2012. A customer churn prediction model in telecom industry using boosting. IEEE Transactions on Industrial Informatics, 10(2), pp.1659-1665.

[13] Lee, Y.H., Wei, C.P., Cheng, T.H. and Yang, C.T., 2012. Nearest-neighbor-based approach to time-series classifica- tion. Decision Support Systems, 53(1), pp.207-217.

[14] Zabkowski, T.S. and Szczesny, W., 2012. Insolvency model- ing in the cellular telecommunication industry. Expert Sys- tems with Applications, 39(8), pp.6879-6886.

[15] Saravanan, M. and Vijay Raajaa, G.S., 2012. A graph-based churn prediction model for mobile telecom networks. In Ad- vanced Data Mining and Applications: 8th International Con- ference, ADMA 2012, Nanjing, China, December 15-18, 2012. Proceedings 8 (pp. 367-382). Springer Berlin Heidel- berg.

[16] Keramati, A. and Ardabili, S.M., 2011. Churn analysis for an Iranian mobile operator. Telecommunications Policy, 35(4), pp.344-356.

[17] Farquad, M.A.H., Ravi, V. and Raju, S.B., 2014. Churn pre- diction using comprehensible support vector machine: An analytical CRM application. Applied Soft Computing, 19, pp.31-40.

[18] Sundarkumar, G.G., Ravi, V. and Siddeshwar, V., 2015, De- cember. One-class support vector machine based undersam- pling: Application to churn prediction and insurance fraud detection. In 2015 IEEE International Conference on Compu- tational Intelligence and Computing Research (ICCIC) (pp. 1-7). IEEE.

[19] Maldonado, S., Flores, A ́ ., Verbraken, T., Baesens, B. and We- ber, R., 2015. Profit-based feature selection using support vec- tor machines–General framework and an application for cus- tomer retention. Applied Soft Computing, 35, pp.740-748.

[20] Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C., 2015. A comparison of machine learning techniques for customer churn prediction. Simulation Mod- elling Practice and Theory, 55, pp.1-9.

[21] Chagas, B.N.R., Viana, J.A.N., Reinhold, O., Lobato, F., Ja- cob, A.F. and Alt, R., 2018, December. Current applications of machine learning techniques in CRM: a literature review and practical implications. In 2018

IEEE/WIC/ACM Interna- tional Conference on Web Intelligence (WI) (pp. 452-458). IEEE.

[22] Sailaja, G., Jayasree, K., Keerthana, M., Kavya, D., Mohan, C.R. and Reddy, M.N.K., Dynamic Churn Prediction Using Machine Learning Algorithms on Telecommunication.

[23] VLN, R.K. and Deeplakshmi, P., 2021, January. Dynamic churn prediction using machine learning algorithms-predict your customer through customer behaviour. In 2021 Interna- tional Conference on Computer Communication and Infor- matics (ICCCI) (pp. 1-6). IEEE.

[24] Domingos, E., Ojeme, B. and Daramola, O., 2021. Experi-mental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. Computation, 9(3), p.34.

[25] Katelaris, L. and Themistocleous, M., 2017. Predicting cus- tomer churn: Customer behavior forecasting for subscription- based organizations. In Information Systems: 14th Euro- pean, Mediterranean, and Middle Eastern Conference, EM- CIS 2017, Coimbra, Portugal, September 7-8, 2017, Proceed- ings 14 (pp. 128-135). Springer International Publishing.

[26] Kumar, S. and Kumar, M., 2019. Predicting customer churn using artificial neural network. In Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24-26, 2019, Proceed- ings 20 (pp. 299-306). Springer International Publishing.

[27] Yontar, M., Dag˘, O¨ .H.N. and Yanık, S., 2020. Using sup- port vector machine for the prediction of unpaid credit card debts. In Intelligent and Fuzzy Techniques in Big Data Ana- lytics and Decision Making: Proceedings of the INFUS 2019 Conference, Istanbul, Turkey, July 23-25, 2019 (pp. 377-385). Springer International Publishing.

[28] Yazdani, R., Taghipourian, M.J., Pourpasha, M.M. and Hosseini, S.S., 2022. Attracting potential customers in E-commerce environments: a comparative study of metaheuris- tic algorithms. Processes, 10(2), p.369.

[29] Breiman, L., 2001. Random forests. Machine learning, 45, pp.5-32.

[30] Pearl, J., 2000. Models, reasoning and inference. Cambridge, UK: CambridgeUniversityPress, 19(2), p.3.

[31] Duda, R.O. and Hart, P.E., 2006. Pattern classification. John Wiley & Sons.

[32] Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A., 1984. Classification and regression trees. CRC press.

[33] Chen, T. and Guestrin, C., 2016, August. Xgboost: A scal-able tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[34] Cox, D.R., 1958. The regression analysis of binary sequences. Journal of the Royal Statistical Society Series B: Statistical Methodology, 20(2), pp.215-232.

[35] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning representations by back-propagating errors. nature, 323(6088), pp.533-536.

[36] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324.

[37] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

[38] Jordan, M.I., 1997. Serial order: A parallel distributed pro-cessing approach. In Advances in psychology (Vol. 121, pp. 471-495). North-Holland.