International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

AN EFFICIENT SYSTEM FOR CUSTOMER RELATIONSHIP MANAGEMENT ON CHURN PREDICTION USING MACHINE LEARNING TECHNIQUE

Davinder Pal Singh
Technical Architect
Salesforce
Toronto, Canada
Davinder.sfdc@gmail.com

Abstract

The paper offers a comprehensive review of the area, highlighting gaps and suggesting possible avenues for future study, in light of the recent advancements in customer relationship management (CRM) technology solutions using AI and big data. Many industries have significant challenges with customer acquisition and retention, but highly competitive and quickly expanding organisations face even greater difficulties. Large businesses are very concerned about customer turnover since keeping a loyal customer is much more important than getting a new one. Determining the reasons behind customer attrition is essential to putting the right measures in place to stop and lessen churn. This research looks at the telecom industry's customer churn prediction using Kaggle datasets and advanced ML algorithms. Methods include data preprocessing, model training with SVM, Stochastic Gradient Boosting, KNN, and GA-XGBoost, and evaluation using F1-score, recall, accuracy, and precision metrics. GA-XGBoost emerges as a most effective model, achieving 96.71% accuracy and outperforming others in precision, recall, and F1-score. Results highlight its potential for enhancing customer retention strategies, optimising resource allocation, and informing targeted marketing efforts, thereby contributing valuable insights for telecom operators aiming to mitigate customer churn and sustain business growth.

Index Terms – Customer Churn Dataset, CRM, Churn Prediction, GA-XGBoost

I. INTRODUCTION

Customers are a business's most important source of income and success. Companies today seek to satisfy their consumers because they recognise the important role that they play. CRM is becoming more and more important for businesses to succeed and get a competitive edge. The ability to access more data that businesses collect about their clients has made ML a cornerstone component of CRM in recent years [1]. The purpose of CRM is to leverage technology to collect, organise, and use data sensibly in order to develop long-lasting customer relationships and an exceptional customer experience [2].

When a client chooses to cease utilising a service or switches suppliers, this is known as customer churn. Loss of consumers may significantly cut into a service provider's profit margin. A churn prediction system is used to forestall the issue of client attrition. The telecom industry would do well to investigate customer turnover and identify its causes in order to forestall such practices in

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

the future[3]. They do this by utilising behaviour data about the consumer from the time of their service subscription to the present[4]. The primary causes of customer churn include inadequate network infrastructure, costly services, appropriate plans, and poor customer service [5].

Global telecoms service providers have recognised the growing importance of reducing customer churn, despite the difficulty of identifying churners due to oversaturation in the industry. As the market develops, the issue is becoming worse[5]. They used a simple but workable churn marketing programme in addition to presenting a prediction model that can reliably forecast the behaviour of consumers who are about to leave. There are several benefits for businesses of all sizes from using ML methods in CRM. CRM is aided and enhanced by ML methods including ensemble learning, instance-based learning, regression analysis, and DT-based learning, among others[6][7].

ML algorithms are the most accurate way to anticipate customer churn. Businesses may use ML pipelines to start retention campaigns for consumers who are identified as potential churn targets. To determine the best accurate algorithm for predicting customer churn in firms, this study uses a number of classification algorithms and compares their output [8][9]. The several methods that are used include SVM, XGB Classifier, AdaBoost Classifier, GNB, LR, SGD Classifier, and ETC. Additionally, a number of assessment parameters are utilised to identify a top-performing model overall for churn forecast in the initial phases[10][11].

In the telecom sector, retaining the customers is always a measure that would help the companies sustain their business and hence get profitable. The research is considered necessary due to the inability to accurately estimate customer attrition in the telecoms sector to map the vast customer attrition dataset sourced from Kaggle. The advanced machine learning methodologies including SVM, SGB, KNN, and GA-XGBoost should be used in this study to establish strong models that can be used to predict potential churners based on the behaviour of the users and the interaction of the services. The objective of this research is to yield concrete recommendations that can assist telecom operators to effectively manage customer touch points, manage churn and thus optimally retain customers and ensure customer satisfaction thereby increasing business viability. Proper data cleaning, selection of proper features, and assessing the churn prediction models using performance measures guarantee the credibility of the churn prediction models developed in this research. The key contributions of the study on forecasting sales using the Customer churn dataset:

- Therefore, the study will seek to establish a predictive model that best predicts customer churn through a comparison of several sophisticated models used in the analysis, including SVM, SGB, KNN, and GA-XGBoost.
- This information can assist telecom companies in creating Churn Management programs with special emphasis on how to retain the potential attrition customers by analysing those customers' behaviour and overall purchase trend.
- Telecom operators may more effectively manage resources by concentrating efforts on keeping customers who are likely to churn and cutting down on needless spending on already loyal customers thanks to accurate churn prediction.
- By predicting churn in advance, telecom operators can take proactive measures such as personalised offers, improved customer service, and targeted marketing campaigns to retain customers.
- The study adds to our knowledge of the mechanics of customer churn in the telecom industry by offering insightful information that will help guide future research and

industry standards for customer retention and management tactics.

A. Structure of the paper

Here follows the arrangement of the remaining portions of the paper. Current churn prediction practices, research gaps, and a comparison table were discussed in Section 2. Section 3 provides a flowchart-based description of the system's approach. The findings, data visualisation analysis, and discussion of the existing models are offered in Section 4. The study's limitations, potential future research, and conclusions are summarised in Section 5.

II. LITERATURE REVIEW

Innovative technologies like deep learning and ML have been used by a number of researchers to build automated systems. These use very precise techniques to determine CRM based on Churn prediction. Researchers have also put out a number of deep learning-based approaches for Churn prediction and categorisation, which are covered below:

This paper, Galal, Rady and Aref, (2022) uses customer profile data to generate a model that uses classifiers to forecast customer churn behaviour. This study applies and compares several supervised classification techniques, including LR, KNN, AdaBoost, Gb, and RF. By using the RF model, they can identify and prioritise the most crucial traits. In the provided experimental case study, the model got 87% accuracy with 10,000 observations[12].

This paper, Latheef and Vineetha, (2021) suggests a reliable method for predicting customer churn utilising an LSTM model as well as the SMOTE methodology for data preprocessing. The assessment results have shown this to be true; the suggested systems outperform the system without the SMOTE approach in terms of churn prediction accuracy (88%)[13].

In this study, Sagala and Permai, (2021) gives an evaluation of three boosting-based models—Cat Boost, XGBoost, and LightGBM—to classify customer turnover data on two situations. Hyperparameters and default values were both subjected to these procedures. Using LightGBM to run the model yields the best results, with accuracy and AUC values of 91.4%, precision scores of 94.8%, and recall scores of 87.7%[14].

We have compiled a list of the methods, datasets, limitations and future work that were used in our assessment of current developments in this subject, which may be seen in Table 1.

TABLE I. COMPARATIVE STUDY ON CUSTOMER RELATIONSHIP MANAGEMENT ON CHURN PREDICTION USING MACHINE LEARNING

References	Methodology	Dataset	Performance	Limitations & Future Work
[12]	Classifier-based model using KNN, LR, AdaBoost, Gradient Boosting, RF, and unified voting	Customer profile data (10K records)	87% accuracy	Future work could include larger datasets and other feature selection methods.

[13]	Customer churn prediction using LSTM and SMOTE	Customer churn data	88% accuracy	Further improvements could involve more advanced preprocessing techniques and exploring other DL models.
[15]	Feature selection combining Information Gain and Ranker methods	Telecom data	95.02% accuracy with feature selection, 92.92% without	Comparison with more feature selection techniques and larger datasets.
[14]	Boosting-based models: XGBoost, LightGBM, Cat Boost	Customer turnover data	91.4% accuracy, 94.8% AUC, 87.7% PRC and RE for LightGBM	Further investigation into hyperparameter tuning and comparison with other boosting techniques.

Despite significant advancements in customer churn prediction using various machine learning models, several research gaps remain. Current studies predominantly focus on improving prediction accuracy through techniques like SMOTE, hyperparameter optimisation, and feature selection. However, there is a lack of comprehensive comparison across diverse industries, such as telecom, banking, and retail, to understand the generalizability of these models. Additionally, most research emphasises traditional ML classifiers and boosting algorithms, with limited exploration of deep learning models like LSTM in varying contexts. Furthermore, while many studies highlight the importance of feature selection, there is insufficient investigation into the integration of novel feature engineering techniques and their impact on model performance. Lastly, the effect of real-time data and evolving customer behaviours on churn prediction accuracy remains underexplored, indicating a need for dynamic models that can adapt to changing patterns over time. The gaps might contribute to the enhancement of the flexibility and accuracy of churn prediction models in different industries as well as with shifting data.

III. METHODOLOGY

We provide a method for churn prediction using large-scale data; the system starts by handling a collection of data on customers who have churned. Data on customer behaviours and churn status has been collected from Kaggle's customer churn dataset as part of the process. Imputation was used to handle missing data, statistical techniques were used to manage outliers, numerical features were scaled for standardisation, and one-hot encoding was used to encode categorical variables. The dataset was split into a training set and a testing set with a ratio of 70:30 to facilitate model training and evaluation. SVM for hyperplane-based separation, GA-XGBoost for genetic algorithm and gradient boosting hyperparameter optimisation, KNN for closest neighbour classification, and stochastic gradient boosting for ensemble learning were the four classification models that were assessed. The most successful method for forecasting customer attrition was determined by comparing the outcomes among models utilising a variety of assessment approaches. The whole proposed methodology shows in Figure 1, with phases discussed below briefly:

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

A. Data Collection and Preprocessing

This study makes use of the customer churn dataset that was received from Kaggle. A dataset is used as input to generate a training set and a testing set. The pre-processing of raw data is an important first step in making it suitable for ML models. It entails a number of critical tasks, including imputation or deletion of missing values, statistical approaches for handling outliers, scaling numerical features to a standard range, encoding categorical variables into numerical representations (e.g., one-hot encoding), and potentially dimensionality reduction using PCA techniques. Enhancing data quality, ensuring algorithm compatibility, and enhancing model performance during testing are all outcomes of implementing all of these measures together.

- **Data Encoding:** Machine learning models may sometimes use data encoded whenever in relation to transforming categorical variables to numerical forms. Some of the examples of this technique include label encoding, which involves converting categories to numbers and one-hot encoding, which involves producing new columns, each of which represents a category.
- **Handling missing values:** Several techniques are available for handling missing values: Imputation where missing values are replaced with other mathematical measures such as mean, median; removal of part records; and the predictive models which use AI to estimate the missing values.
- **Data transformation:** This stage involves the preparation of numerical data through methods such as principal component analysis to decrease complexity or logarithmic transformation if the data distribution is skewed. Otherwise, characteristics can be brought to the same scale (normalisation). These changes positively affect the model's performance as well as the quality of the data it produces.

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

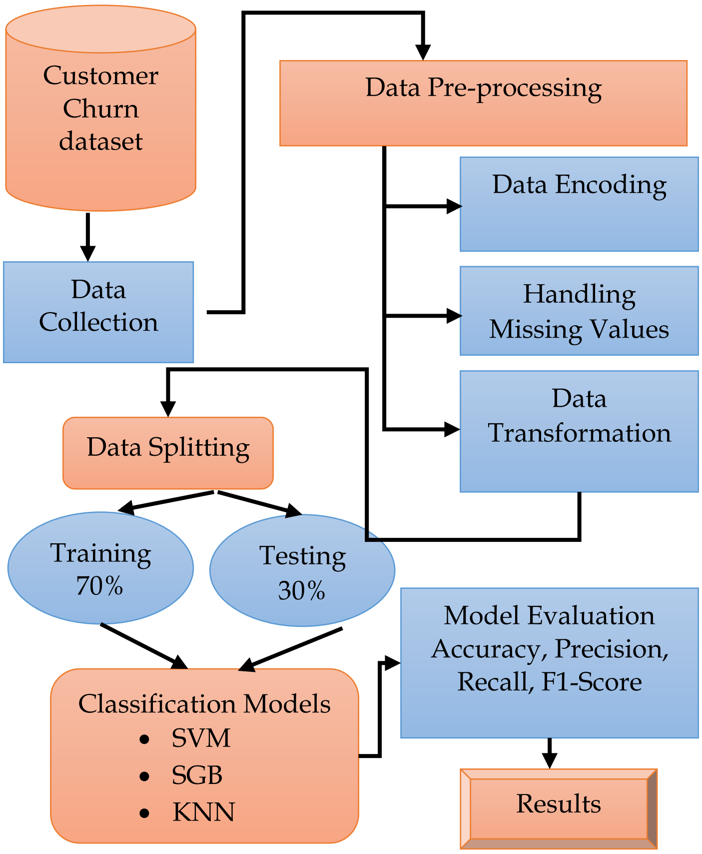

Fig. 1. Proposed flowchart for Customer churn system

B. Classification of Machine Learning Models

An algorithm for predicting customer churn was built using several ML approaches. It was determined that these methods would provide the most accurate predictions of customer churn, thus they were implemented. The SVM, SGB, KNN, and the suggested GA-XGBoost classifiers were the three methods used. The descriptions of each model and its purposes are provided below.

1). SVM

Cortes and Rapnik (1995) introduced supervised classification methods, including the SVM. ML algorithms like SVM are used to create predictions and generalisations about new data by learning from data whose distribution is unknown. The existence of a hyperplane that optimally separates the data into two groups is the foundation of the SVM. Based on the linear and nonlinear separations of the data set, two SVMs are created.

2). Stochastic Gradient Boosting (SGB)

SGB is an example of an ensemble algorithm. Each of the underlying models that make up an ensemble algorithm contributes something special to finding a solution to the issue. Combining all the responses from the fundamental models, often via weighted voting or average, generally results in a more consistent and accurate prediction for the final model output. A combination of boosting and bagging techniques, SGB is based on a model of tiny trees with L-terminal nodes [16].

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

3). KNN

The kNN classifier uses comparisons between the new structure and the instances of the training method to determine the classification of new standards or intrusive processes. Then, based on the k-nearest classes, it forecasts the new process class. It is assumed by the technique that processes belonging to the same class are grouped together in the vector space. The number of neighbours needed to define the data class is determined by the value of k. A majority vote determines who the nearest neighbours are. A simple application of the Euclidean distance formula yields the desired distance.

4). GA-XGBoost

GA-XGBoost optimises hyperparameters and performance using GA and XGBoost. The efficient gradient-boosting implementation XGBoost uses an ensemble of decision trees to correct prediction mistakes. Its objective function contains a loss function $l(y_i, \hat{y}_i)$ and a regularisation term $\Omega(f_k)$:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{1}$$

The Genetic Algorithm simulates natural selection in order to optimise the hyperparameters of XGBoost. This includes initialisation, selection, crossover, and mutation. The objective of this iterative approach is to identify the best hyperparameter combination that maximises the model's performance measures, such accuracy or F1-score. Because the model incorporates both GA and XGBoost, it is able to capitalise on the advantages of both approaches, which ultimately results in a predictive model that is more accurate and efficient.

IV. RESULTS AND DISCUSSION

A Jupyter Notebook running Python on an Intel Core i7 CPU with 4.20 GHz and 48 GB of RAM is utilised in this experiment. A telecom customer churn information is used. In this part, we will go over the data that was utilised for exploratory visual analysis, the experimental methodology that was followed, and then, with each set of findings, we will analyse them thoroughly and provide a suitable interpretation.

A. Dataset Description

In this work used Telco Customer Churn that collected from the Kaggle. The customer qualities indicated in the Metadata column are included in each column, and every row represents a customer. Details about are included in the data set.

- The Churn column displays the customers who left during the last 30 days.
- The customers' signed-up-for services, which may include phone, internet, device protection, multiple lines, security, tech assistance, backup, and streaming media.
- Information on a client's account, including the duration of their agreement, payment method, paperless billing options, monthly fees, and total charges.
- Information on the customers, such as their gender, age range, and marital status

Figure 2 to 4 depicts a visual representation of the correlation matrix, histograms of input dataset, and bar graph.

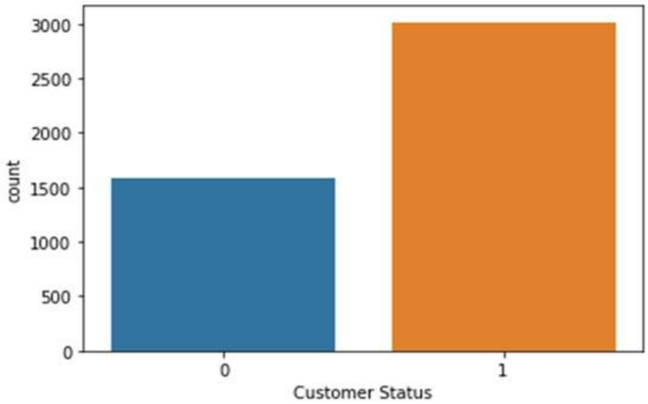

Fig. 2. Bar plot for customer churn distribution.

Below is a bar chart that was created to look at how the target variable was distributed over the customer service call variable. Figure 2 demonstrates that there are almost twice as many churning samples as non-churning samples, indicating that the dataset is imbalanced.

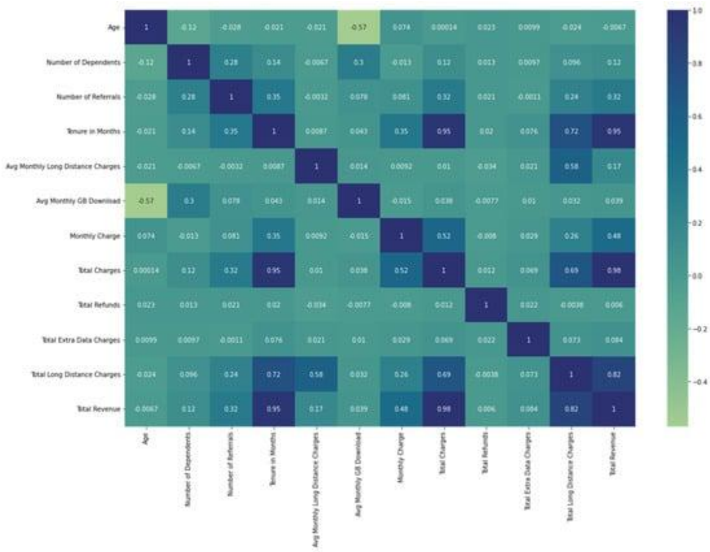

Fig. 3. Heatmap of Correlation Matrix.

Figure 3 shows the results of a correlation matrix that was used to investigate the inter-correlation among the characteristics. The correlation matrix was then shown as a heatmap to show the pairwise association among the telecom customer's call activity and features.

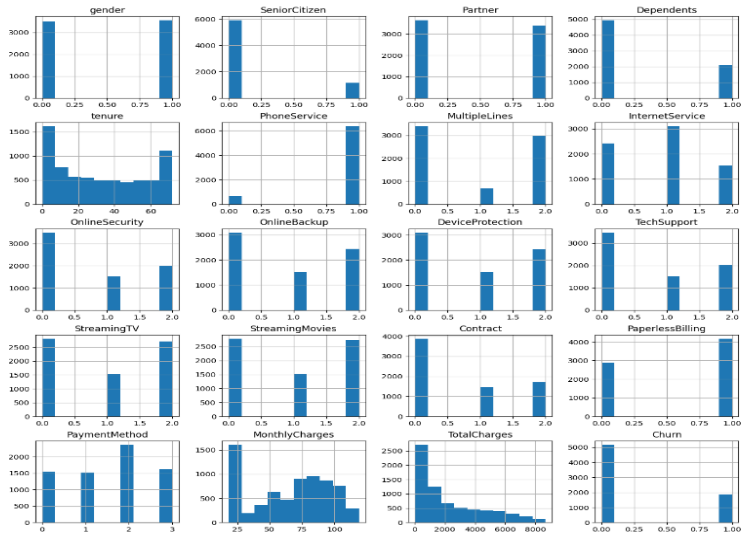

Fig. 4. Plotted Histogram of Dataset

The telecom dataset's plotted histogram is seen in Figure 4. A histogram is a graphical representation of the distribution of values for a numerical quantity, often shown as a series of bars. Before making a histogram, you need to divide the complete range of values into intervals and then make a bin of the ranges.

B. Performance measures

Evaluation is necessary to decide a model's performance since ML models aren't always ideal for the provided data. Metrics like precision, accuracy, recall, and F1-score are often used to assess binary classifiers. Figure 5 shows a confusion matrix that may be used to compute these measures.

Class	Y	N
Y	True positive (VP)	False negative (FN)
N	False positive (FP)	True negative (TN)

Fig. 5. Representation class of confusion matrix

A ML concept called a confusion matrix is used to characterise the classifier's performance. It contains data on both actual and expected classifications. Test cases that are successfully categorised are represented by True Positives (TP) and True Negatives (TN), while test instances that are wrongly classified are represented by False Negatives (FN) and False Positives (FP).

1). Accuracy

An indicator of the classifier's overall efficacy is its accuracy [48]. It is a measure that displays the percentage of cases that were properly categorised overall. Equation 2 defines accuracy.

$$Acuracy = \frac{TP+TN}{TP+TN+FN+FP} \tag{2}$$

2). Precision

The proportion of positive events that were correctly predicted is shown by the precision measure. In this instance, the metric denotes how well the model predicted the churner class. Equivalent to eq.3, precision is:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

3). Recall

A classifier's recall is a gauge of how effectively it can detect positive examples. This metric demonstrates how well the binary classifier can detect instances belonging to a given class. The formula for recall is given by eq.4:

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

4). F1-score

To measure how well a classifier does its job, many people turn to the F-score. Traditional definitions of the F-score as the harmonic mean of recall and precision indicate that it is a metric that accounts for both variables. Equation 5 shows that as the F-score approaches 1, the combined recall and precision improve.

$$F1-score = 2 * \frac{precision*recall}{precision+recall} \tag{5}$$

- **True positives (TP):** The quantity of churn that consumers correctly predicted as actual churn.
- **True negatives (TN):** The number of customers that are correctly predicted as non-churn.
- **False positives (FP):** The number of consumers who churn was incorrectly calculated as non-churn.
- **False negatives (FN):** Calculating the amount of non-churn customers as churn with accuracy.

ROC-AUC curve: ROC is a graphical depiction of the accuracy of classification. This graph shows the FPR on the x-axis and the TPR on the y-axis. One way to estimate the AUC is by looking at the ROC area. It is one approach to getting a more precise performance assessment of the model. An ideal model has an AUC of 1, and as the AUC value drops, the model's quality declines.

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

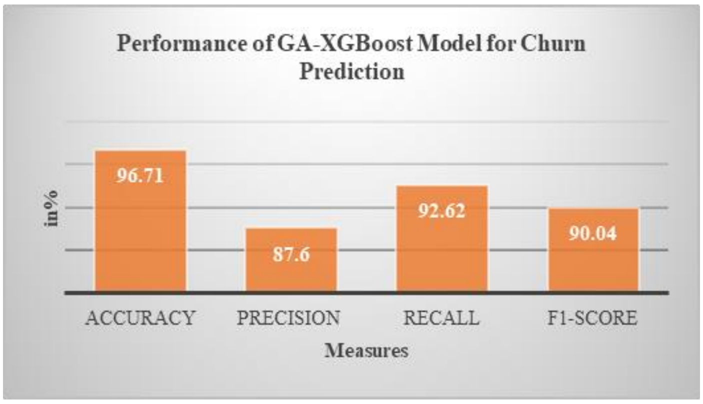

Fig. 6. Parameters performance of GA-XGBoost Model for Churn Prediction

A Figure 6 shows the bar graph of parameters performance of GA-XGBoost model for churn prediction according to accuracy, precision, recall and f1-score measures with 96.71%, 87.6%, 92.62% and 90.04% performance.

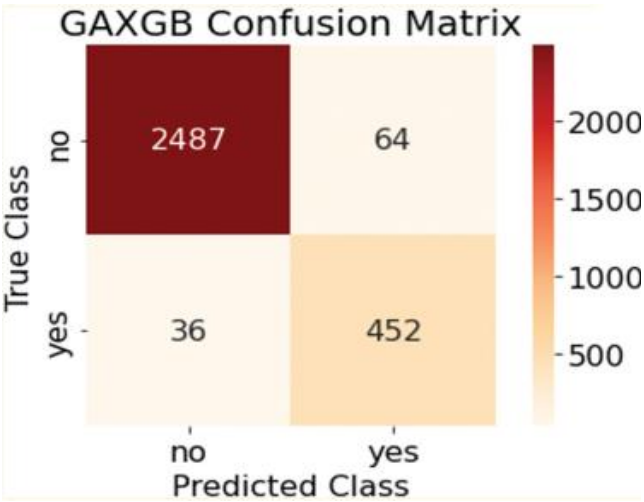

Fig. 7. GA-XGBoost confusion matrix.

Figure 7 depicts a confusion matrix for a GA-XGBoost model, that it correctly predicted 2487 instances as 'no' and 452 instances as 'yes'. There were 64 FP (predicted 'yes' but actually 'no') and 36 FN (predicted 'no' but actually 'yes'). This indicates a model's accuracy in distinguishing between a 'no' and 'yes' classes, with higher precision for the 'no' class.

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

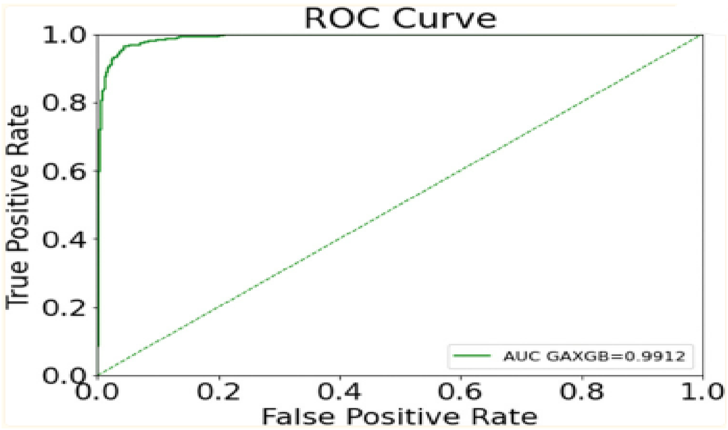

Fig. 8. ROC curve of GA-XGBoost

Figure 8 represents the ROC curve for the GA-XGBoost model, the trade-off among the TPR (sensitivity) and the FPR. The model's AUC (Area Under the Curve) is 0.9912, indicating excellent performance in distinguishing between classes. There is a low FPR and a high TPR when the curve is close to the top-left corner.

C. Comparative analysis

A comparative analysis for Customer Churn forecast based on machine learning. A following table shows the comparison between various machine learning models for comparative analysis and Customer Churn in terms of performance metrics.

TABLE II. COMPARISON BETWEEN VARIOUS MODEL CUSTOMER CHURN PREDICTION

Models	Accuracy	Precision	Recall	F1-score
SVM	79	48	69	57
SGB	78	81	92	86
KNN	77.01	59.95	40.82	48.57
GA-XGBoost	**96.71**	**87.60**	**92.62**	**90.04**

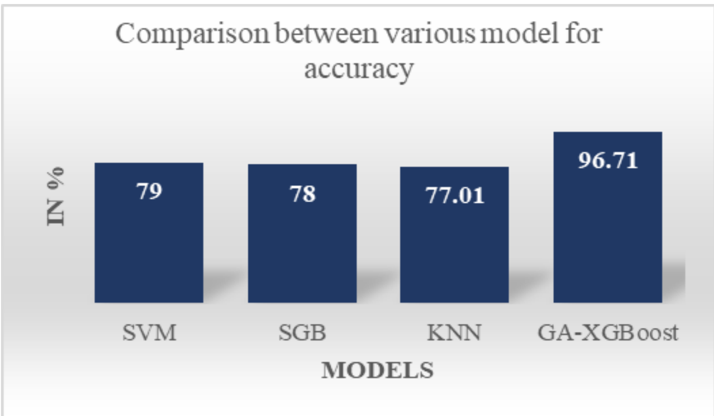

Fig. 9. Accuracy comparison of ML models for churn prediction

Figure 9 shows an accuracy comparison of ML models for churn forecast. The GA-XGBoost has a

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

highest accuracy score 96.71%, and KNN has the lowest accuracy with 77.01% score. While SVM and SGB model get 79% and 78% accuracy for churn prediction.

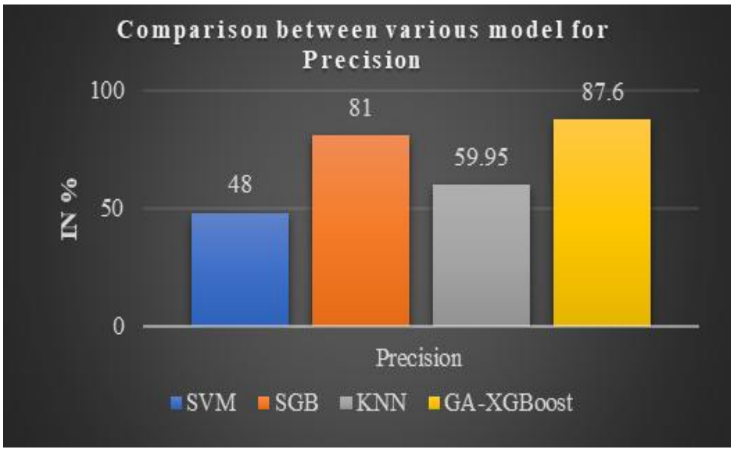

Fig. 10.Precision comparison of ML models for churn prediction

Figure 10 shows a precision comparison of ML models for churn forecast. The GA-XGBoost has a highest precision score 87.6%, and KNN precision of 81% score. While SGB and SVM model get 59.95% and 48% precision for churn prediction.

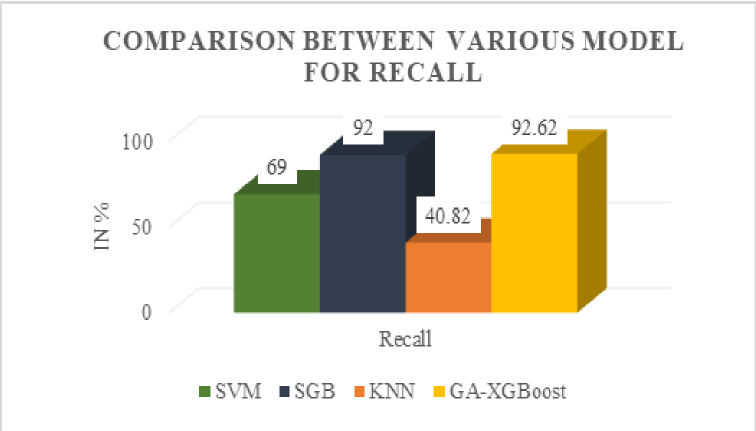

Fig. 11.Recall comparison of ML models for churn prediction

Figure 11 shows a recall comparison of ML models for churn forecast. The GA-XGBoost has a highest recall score 92.62%, and KNN recall of 40.82% score. While SGB and SVM model get 69% and 92% recall for churn prediction.

International Journal of Core Engineering & Management
Volume-7, Issue-04, 2022 ISSN No: 2348-9510

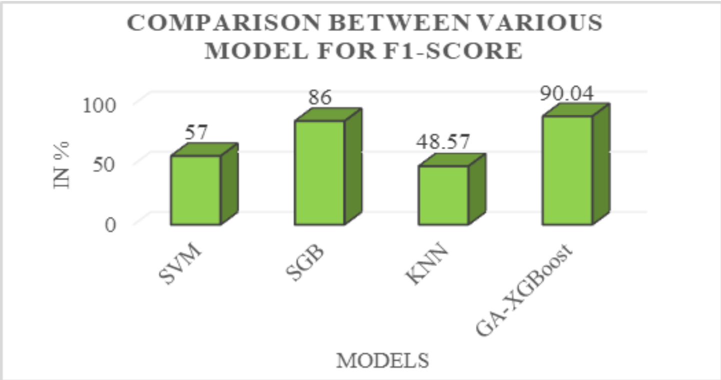

Fig. 12.F1-Score comparison of ML models for churn prediction

Figure 12 shows a F1-Score comparison of ML models for churn forecast. The GA-XGBoost has a highest F1-Score score 90.04%, and KNN F1-Score of 40.82% scores. While SGB and SVM model get 86% and 57% F1-Score for churn prediction.
The analysis demonstrates that a GA-XGBoost model outperforms other ML models in predicting customer churn across various performance metrics. As seen in the confusion matrix (Figure 7), GA-XGBoost achieves high accuracy in correctly predicting both 'no' and 'yes' classes, with a particularly high precision for the 'no' class. The ROC curve (Figure 8) further emphasises its robust performance with an AUC of 0.9912, indicative of excellent class distinction. Comparative metrics in Table 3 and subsequent Figures (9-12) show GA-XGBoost leading in accuracy (96.71%), precision (87.60%), recall (92.62%), and F1-score (90.04%), significantly surpassing the SVM, SGB, and KNN models. This comprehensive evaluation confirms GA-XGBoost as the superior model for customer churn prediction, highlighting its ability to balance true positive and false positive rates effectively, and thus, offering reliable and accurate predictions.

V. CONCLUSION AND FUTURE SCOPE

This research proved that GA-XGBoost and other ML models can accurately forecast telecom customer churn. After examining a large dataset and using state-of-the-art ML approaches, the best performing model was GA-XGBoost, which achieved impressive results in terms of accuracy (96.71%), precision (87.6%), recall (92.62%), and F1-score (90.04%). The findings of this research are significant for telecom players who wish to bring changes in their customer churn rate, explore the most profitable ways of utilising resources and work for better profitability of the business through retention strategies.
In the future, there is even more potential for optimisation implemented in several directions. First of all, using a larger set of customers' interaction data could enhance the model's stability and predictive capability. Further, extending the options of hyperparameters' optimisation different from genetic algorithms and also implementing the method of real-time predictions will open wider perspectives for proactive churn management. Finally, incorporating churn models right into current CRM platforms, where these tools can be used on a daily basis, and investigating the possibilities of developing more sophisticated AI-based methods with which to work on the churn prediction concept could provoke significant improvement in the outcomes of the undertaken actions in telecommunications business. These innovations shall help in crafting better customer

retention strategies as well as maintaining long-term profitable business growth in the context of the somewhat saturated telecommunications marketplace.

REFERENCES

[1] P. Zerbino, D. Aloini, R. Dulmin, and V. Mininno, "Big Data-enabled Customer Relationship Management: A holistic approach," *Inf. Process. Manag.*, 2018, doi: 10.1016/j.ipm.2017.10.005.

[2] C. Ledro, A. Nosella, and A. Vinelli, "Artificial intelligence in customer relationship management: literature review and future research directions," *Journal of Business and Industrial Marketing*. 2022. doi: 10.1108/JBIM-07-2021-0332.

[3] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection," *PLoS One*, 2022, doi: 10.1371/journal.pone.0278095.

[4] J. Thomas, "The Effect and Challenges of the Internet of Things (IoT) on the Management of Supply Chains," *Int. J. Res. Anal. Rev.*, vol. 8, no. 3, pp. 874–878, 2021.

[5] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2914999.

[6] V. Rohilla, M. Kaur, and S. Chakraborty, "An Empirical Framework for Recommendation-based Location Services Using Deep Learning," *Eng. Technol. Appl. Sci. Res.*, 2022, doi: 10.48084/etasr.5126.

[7] T.Kamalakannan, Dr.P.MAYILVAHANAN and PROFESSOR, "Efficient Customer Churn Prediction Model Using Support Vector Machine with Particle Swarm Optimization," *Int. J. Pure Appl. Math.*, vol. 119, no. 10, pp. 247–254, 2018.

[8] V. Rohilla, S. Chakraborty, and R. Kumar, "Car Auomation Simulator Using Machine Learning," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3566915.

[9] S. G. Kumar *et al.*, "Chronic Reductive Stress Modifies Ribosomal Proteins in Nrf2 Transgenic Mouse Hearts," *Free Radic. Biol. Med.*, 2022, doi: 10.1016/j.freeradbiomed.2022.10.125.

[10] R. K. Vinita Rohilla, Sudeshna Chakraborty, "Random Forest with harmony search optimisation for location based advertising," *Int J Innov Technol Explor Eng*, vol. 8, no. 9, pp. 1092–1097, 2019.

[11] S. R. Labhsetwar, "PREDICTIVE ANALYSIS OF CUSTOMER CHURN IN TELECOM INDUSTRY USING SUPERVISED LEARNING," *ICTACT J. Soft Comput.*, 2020, doi: 10.21917/ijsc.2020.0291.

[12] M. Galal, S. Rady, and M. Aref, "Enhancing Customer Churn Prediction in Digital Banking using Ensemble Modeling," in *NILES 2022 - 4th Novel Intelligent and Leading Emerging Sciences Conference, Proceedings*, 2022. doi: 10.1109/NILES56402.2022.9942408.

[13] J. Latheef and S. Vineetha, "LSTM Model to Predict Customer Churn in Banking Sector with SMOTE Data Preprocessing," in *ACCESS 2021 - Proceedings of 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems*, 2021. doi: 10.1109/ACCESS51619.2021.9563347.

[14] N. T. M. Sagala and S. D. Permai, "Enhanced Churn Prediction Model with Boosted Trees Algorithms in the Banking Sector," in *2021 International Conference on Data Science and Its Applications, ICoDSA 2021*, 2021. doi: 10.1109/ICoDSA53588.2021.9617503.

[15] Y. K. Saheed and M. A. Hambali, "Customer Churn Prediction in Telecom Sector with Machine Learning and Information Gain Filter Feature Selection Algorithms," in *2021 International Conference on Data Analytics for Business and Industry, ICDABI 2021*, 2021. doi: 10.1109/ICDABI53623.2021.9655792.

[16] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*. 2008. doi: 10.1111/j.1365-2656.2008.01390.x.