International Journal of Integrated Science and Technology (IJIST)
Vol. 4, No. 2 February 2026: 123-137

Customer Churn Prediction System Using Machine Learning: A Case Study ROSHAN Telecom-Afghanistan

Hafeezullah Shoja[1], Esmatullah Sabet[2*], Sayed Shafiullah Sadat[3]
[1]Assistant Professor, Department of Information Systems, Faculty of Computer Science, Kabul
[2]Assistant Professor, Department of Information Systems, Faculty of Computer Science, Parwan University, Charikar, Afghanistan
[3]Assistant Professor, Department of Information Technology, Faculty of Computer Science, Polytechnic University, Kabul, Afghanistan
Corresponding Author: Esmatullah Sabet: Esmatullah.sabet@parwan.edu.af

A R T I C L E I N F O

Keywords: Telecom, Customer Churn, Prediction, XGBoost Algorithm

Received : 20, December
Revised : 25, January
Accepted: 20, February

A B S T R A C T

The success of any business relies on its customers, so it is crucial for firms to prioritize customer satisfaction. Customer churn is a significant concern for companies due to increased competition, the growing importance of marketing strategies, and customers' awareness of their options. To address churn, organizations need to develop specific strategies that align with the services they offer. In the telecom sector, customer churn is especially important due to the high cost of acquiring new customers. This study aims to develop a customer churn prediction system in the telecom sector using machine learning, specifically the XGBoost algorithm. Real-time data from ROSHAN Telecom was used to predict risky customers likely to churn, enabling proactive management strategies for customer retention. The study's results show that the accuracy rate of churn prediction is 93.39 percent, demonstrating the effectiveness of machine learning techniques in predicting customer churn for the telecom company.

DOI: https://doi.org/10.59890/ijist.v4i2.287

URL: https://ntlmultitechpublisher.my.id/index.php/ijist

INTRODUCTION

The global telecommunications industry is rapidly growing and becoming a major sector, leading to increased competition due to advances in technology and the rising number of operators. To survive in this highly competitive market, telecom companies are implementing strategies to generate substantial revenue. This is very important for the company to retain its existing customer other than acquiring new customer. Existing customer can generate more revenue than new customer because they are already familiar with service and products of the company. Thus, there is need of a solution for the company to have a trained model data of the customers to predict the possible churner. The Telecom sector has many types of data sources related to customers such as CRM (Customer detail information), CDR (Call Detail Records), Customer Bundle activation, deactivation, and TOPUP. The solution can classify customers, detect the behavior of customer, train the model, and predict the churners using machine learning. The dataset for this model has been prepared from the mentioned data sources, and a one-year period of data has been selected. Then random data of 100K is further narrowed from that one-year data and passed to the prediction algorithm. The prediction model has predicted around 10K of customer with 80% train model and 20% of test model. The prediction uses machine learning technique and XGBOOST algorithm to classify customers and predict the churners.

Problem Statement

Customer churn prediction is a common problem in the telecom industry, where customers frequently switch between service providers (Ibrahim AlShourbaji, 2021). To overcome this problem, we have prepared dataset of customer information, including demographic data, usage patterns, and service attributes. As a result, the task is to predict which customers are at risk of churning or cancelling their subscriptions. The goal is to identify these customers early on so that the telecom company can take proactive measures to retain them, such as offering promotions or improving service quality. Machine learning algorithms such as XGBoost, can be used to build predictive models that analyse customer data and identify patterns that are indicative of churn risk. The XGBoost algorithm is particularly well-suited to this task, as it can handle large datasets with many features and can handle both categorical and numerical data. The dataset used for this problem is large enough that captures a representative sample of the customer population that listed both churned and non-churned customers. The dataset is pre-processed to clean and transform the data, as well as to handle missing values and outliers.

Aim & Objectives

The aim of this research is to explore the use of XGBoost for customer churn prediction in a telecom dataset. The objectives include:

- To collecting data from various sources.
- To classifying customers into different segments based on shared characteristics.
- To predicting which customers are likely to churn.
- To applying machine learning algorithms to derive results.

Research Questions

1. How to gather data from different sources?
2. How to get High accuracy using machine learning techniques?

LITERATURE REVIEW

According to Sudarshan Zimal (Sudarshan Zimal, 2023) Customer churn is a common problem for businesses operating in competitive markets. Churn occurs when customers stop using a company's product or service, leading to a loss of revenue and increased marketing expenses. Predicting customer churn is importance for any businesses, as it allows them to take pre-emptive action to retain customers and reduce the negative impact of churn. In recent years, there has been a growing interest in customer churn prediction, with many studies exploring different approaches to predicting churn. A study by (Sudarshan Zimal, 2023) stated that organizations all over the world have had to adjust because of globalization and digitalization, which have given rise to new business models. One outcome of the massive digitalization that has swept the world is subscription-based 12 services. This opens new opportunities and presents problems that call for fresh approaches. The way business is conducted has been transformed by digitalization, which has also boosted the availability of subscription-based services. As a result, businesses can have a harder time keeping customers.

Digitalization can reduce labour costs, increase productivity, and give a clearer picture of how a business is operating. This is essential to maintain competition and get an advantage over other businesses (Sudarshan Zimal, 2023). The growth of information technology has led to an increase in the volume of data and information in the last decade (Sudarshan Zimal, 2023). This quick increase has made it possible to store and processing of massive amounts of data has increased the demand for automatic knowledge creation and identification.

Reichheld (F. F. Reichheld, 1990) in the Ming Zhao's (Ming Zhao, 2021) research has been quoted "that the longer the business relationship between an enterprise and its customers lasts, the more profits the enterprise will make from its existing customers". The net present value of customers in the business environment will rise by 25% to 95% for every 5% improvement in customer retention rated by (F. F. Reichheld, 1990) which has been cited by (Ming Zhao, 2021).

Research shows that "when the customer churn rate of an enterprise decreases by 5%, the average profit rate of the enterprise will increase by 25%–85%". Therefore, predicting customer churn has practical value in that it will help businesses financially.

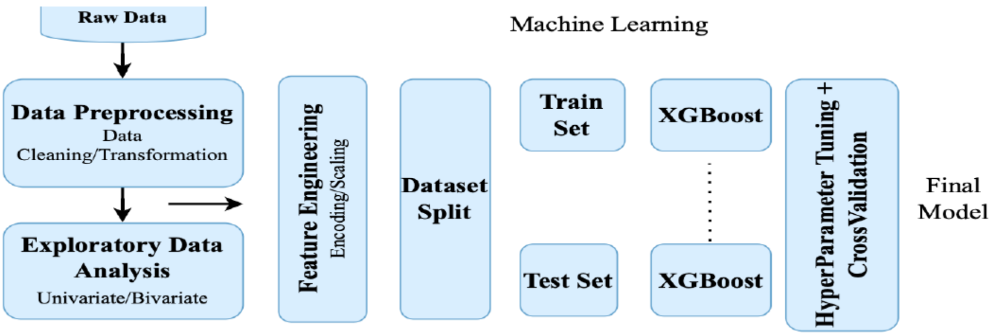
According to Parvatiyar (A. Parvatiyar and J. N. Sheth, 2001) in research done by Swetha (Swetha P a) is "Customer relationship management is one of the crucial and important perspectives for several markets, especially for a market like telecommunications where the firm invests a huge amount in designing and building the infrastructure" (A. Parvatiyar and J. N. Sheth, 2001). However, the architecture focus on retaining the old customers other than having new one, therefore retaining customer required to develop customer's loyalty that is build based on satisfaction and demand matching of customer (M. C. Mozer,2000). cited by (Swetha P a) that "Loyal and long-term customer generates high profit as they are not vulnerable and do not get easily attracted by other telecom services in comparison with the new customers". There are many reasons that a customer churn can happen, such as Service Rate, Lack of Support, Lack of offers, Location Change. Service Rate is the amount for SMS, call, and Data usage. Lack of Support is whenever customer have issue there is no quick support to solve his / her issue. Lack of offers means that there are not attractive offers on a timely manner compared to competitor, and Location change is that the customer goes from one province to another where the coverage of competitor is better. Moreover, by Swetha (Swetha P a), many reference has been stated by author that used aggregated function which combines different behavioral features such as call minutes and CDR (Call detail records). However, the aggregated mechanism has cons as this way it doesn't consider the frequently changes of customer behavior. Therefore, it has been recommended in the paper that needs to consider dynamic behavior. Further on dynamic behavior the paper followed hybrid model that include three sub-model that process time series in various ways proposed. A study by (Khan, 2015) and (Zhang, 2016) stated which has extracted data based on behavioral feature that compared with LSTM mechanism. Although as per the analysis this method focusses on model end to end but data aspects ignore such as data imbalance and data sparsity. Imbalance data is one of the major problems in churn prediction, which has found by researchers that XGBoost is the boosting mechanism that solve the imbalance issue.

According to Swetha (Swetha P a) XGBoost require fine tuning, that the research work on fine-tuning of XGBoost. Feature Selection, Feature Extraction, Data Processing, and Model Deployment are the three main phases that make up Fine Tuned XgBoost. When using a system framework, attributes are taken from the available data and pre-processing is carried out according to each one. As a result, evaluation of XGBoost algorithm as done through this research over other algorithms in terms of Accuracy, precision, and recall. Figure 1 stated the accuracy of XGBoost algorithms over others machine learning algorithms. This evaluation has been done considering the existing work on that provide comparison over other techniques.

Figure 1: Accuracy Comparison on Different Algorithms (Swetha P a)

According to (Sagar Maan Shrestha, 2022) Telecoms mostly try to maintain the growth and competition in the market. Therefore, they always try to have less churn rate. They mostly overcome this problem by offering new products and services, but sometime companies miss out to catch customers in not involving proactive campaign and not act on time to retain their loyal customers. Furthermore, predicting is such a big problem using machine learning technique can make a significant difference to company revenue and growth.
Moreover (Sagar Maan Shrestha, 2022) mentioned Telecommunications dataset for churn is always imbalance, means that numbers customer who churns are more less than those who are not going for churn. Therefore, XGBoost algorithm is more effective than any other algorithm to tackle the predict accurately on imbalanced data. This research showed the effectiveness of XGBoost Algorithm on both datasets publicly available and native dataset.

Figure 2: Methodology Framework for the research done by Sagar Maan (Sagar Maan Shrestha, 2022)



In research article of (Sagar Maan Shrestha, 2022) the Model building has done after data pre-processing, by splitting datasets into Train and Test ratio of 80:20. Training dataset trained using XGBoost and hyperparameters tuned using

GridSearchCV and further validated using 10-fold Stratified KFold Cross validation. The performance of the algorithm is evaluated by measuring the accuracy, precession, recall and F1 score. The XGBoost classified used in this research is the extended version of gradient tree boosting algorithm. Table 1 showed the results classification for Dataset (which is from one of telecom company in Nepal) and Table 2 is the test metrics for all measured point.

Table 1: Result classification of Dataset II from one of Telecom Company in Nepal (Sagar Maan Shrestha, 2022)

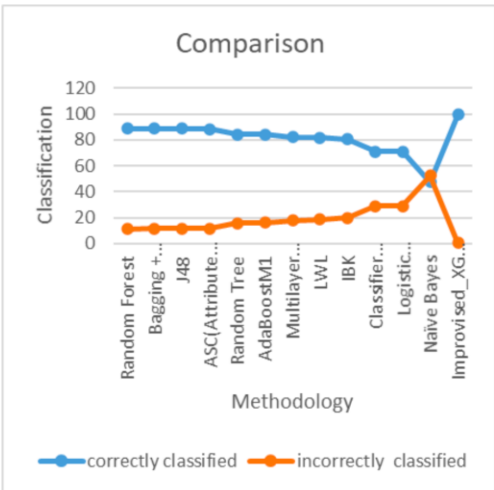
Class	Precision	Recall	F1-Score	Support
Non Churn	0.98	0.99	0.98	9241
Churn	0.91	0.84	0.87	1226

Table 2: Test Metrics for all measured point (Sagar Maan Shrestha, 2022)

Test_Metrics	Average	Minimum	Maximum
test_accuracy	0.9735	0.9657	0.9818
test_recall	0.8557	0.8107	0.8874
test_precision	0.9153	0.838	0.9555
test_f1	0.8837	0.8573	0.9196

Furthermore, in the research by (Swetha P, 2020) there is a graphic representation of many algorithms that is compared with Improvised_XGBoost prediction model. Figure 3 shows the comparison of many predictions model along with Improvised_XGBoost algorithm. This graph indicates the correctly classified and incorrectly classified compared with many other machine learning algorithms, where the Improvised_XGBoost has achieved the higher accuracy rate.

Figure3: Comparison of prediction model along with Improvised_XGBoost algorithm by (Swetha P, 2020)



Moreover, the research by (Swetha P, 2020) concluded that the proposed model (Improvised_XGBoost) has achieved the higher accuracy of 99.41%, the precision rate is 99.44% and the recall with 99.94%. It has been also indicated by

Author the model achieved the nearer to absolute accuracy, still few areas need to be focused and needs to be evaluated by considering different dataset which would be carried out as part of the future work.

The Author (Ammar A Ahmed, 2017) in the research presented a case study customer churn prediction using different techniques, such as meta-heuristic, machine learning, neural networks, and data mining. According to author most of churn prediction method used machine learning and meta-heuristic algorithms for most accurate prediction. Table 3 shows some studies that applied different methods along with its advantage and dis-advantage.

Table 3: Different studies that applied different method along with Advantage and dis-advantage by (Ammar A Ahmed, 2017)

Author	Method	Dataset	Advantages	Disadvantages
J. Burez et al, [16]	Random and advanced under sampling, Gradient boosting machine and WRF	Six real-life proprietary European churn modeling datasets	Increases churn prediction accuracy.	individual issues reduces the overall performance
Veronikha Effendy et al, [17]	Combined sampling with WRF	Categorical type churn data	High values of prediction accuracy and F-measure. Resolves imbalance data problem.	Most general under-sampling process is employed
Ning Lu et al, [18]	Logistic regression with Gentle AdaBoost	Telecom data (2010)	Accurate definition of high risk customer group	Unable to finalize the reasons for customers churn
Xiaojun Wu et al, [19]	Improved SMOTE & AdaBoost	Data from B2C e-commerce site	Higher accuracy with low cost processing. Flexible to utilize in different fields.	Does not consider class rarity
G. Ganesh Sundarkumar et al, [20]	One-class SVM under-sampling with	Insurance dataset	High accuracy and reduced system complexity	More applicable for fraud detection than churn prediction

	Decision tree			
Qiu Yihui et al, [21]	OOPM Feature selection and FE_RF&T feature extraction	Business analysis system of Chie Mobile communication	Highly accurate churn prediction Removes irrelevant information	Application needs are fulfilled only based on distribution information of test samples
Qiuhua Shen et at, [22]	Churn prediction based on compleme ntary fusion of multi-layer features	European telecommunica tions company data	High prediction with better dimensionality reduction	Unbalanced data problem re-occurs when feature selection is not withstanding
Sebastian Maldonado et al, [23]	Profile based SVM	UCI-Telecom, Operator 1, Cell2Cell	Increased accuracy with profile priority	Regulatory reasons are not satisfied in SVM

Therefore, a study by (Abdelrahim Kasem Ahmad, 2019) stated three main strategies have been proposed to generate more revenue for the company.
Acquire new customer
Upsell the existing customer
Increase the retention period of the customers

Also, the author (Abdelrahim Kasem Ahmad, 2019) tackled with unbalanced data, because unbalance dataset could cause negative impact on the final models. The dataset was unbalanced because the secondary class what 5% (churned customers) of whole dataset. The researcher dealt with this by rebalancing the sample training by taking the sample of data two make both classes balanced. Oversampling has done by duplicating the secondary class. Under sampling has done by reducing the sample size of first class. Table 4 shows the result of dataset with different classification algorithms.

Technique used for unbalanced dataset	XGBOOST(%)	GSM(B)(%)	Random Forest (%)	Decision Tree (%)
Oversampling	92	90.01	84.2	76.25
Under sampling	93.12	90.21	87.76	83
without balancing	93.3	90.89	78.47	72.2

Mostly in churn prediction balancing data is needed, that according to (Benjamin Ghaffari, 2021) that use the XGBoost classifier performs the best based on overall accuracy, precision, recall, and F1-score. The study also finds that balancing the training dataset using sampling methods has a positive influence on the studied problem.

Finally, to bring a light on the topic with digital governance much research has been done on efficiency of Machine learning on government services. The author (Charalampos Alexopoulos, 2019) agreed on identified the benefits and obstacles towards the adoption of Machine Learning (ML) in the public sector. The study finds that ML can be used to analyse Big Data and generate new knowledge, and the main approach for solving classification problems is ML techniques used. The benefits of ML usage in government include accuracy, efficiency, scalability, and flexibility. However, the nature of data and human intervention needed for interpretation may lead to misleading results, limiting the benefits that ML can provide. The study suggests that combining deep learning techniques with neural networks can solve complex problems and avoid computing power and time issues. Despite the barriers, ML is a promising technology that can provide potential benefits to the public sector.

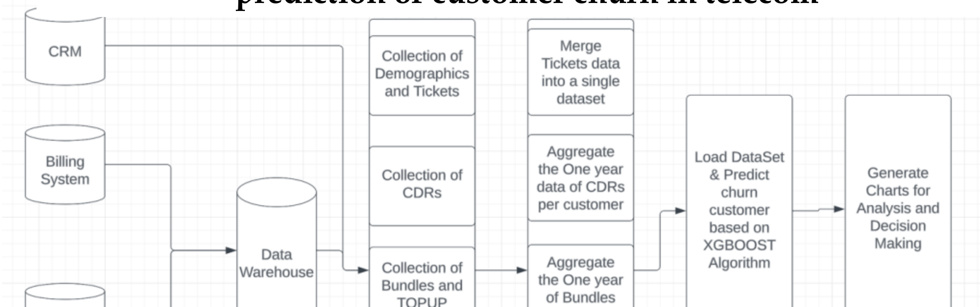
METHODOLOGY

Data Collection and Preparation

Data Collection and preparation is a critical step for data analysis and machine learning algorithm. The data collected from multiple sources exist in any telecom industry such as CDR from Billing System (Defined Cost), TOPUP, MSC CDR (Un-defined cost), Bundle Subscription and CRM. The Billing CDR consist of Voice Call, SMS and Data. In this research we have collected data of Voice Call both ON-NET and OFF-NET as well as SMS both ON-NET and OFNET and Data Usage per customer on daily basis. Beside this the TOPUP that customers credited their account on daily basis capture to show Total TOPUP per customer. Bundle Subscription we captured the data of each bundle that has been subscribed by customer on daily basis. Further we consolidated all daily base CDR into one table called Sub_Usage. Then we segregated one-year data from Sub_Usage table into Sub_Usage_CCP table and updated the gender, province and tickets from CRM data accordingly. Also we have updated the flag of churn and non-churn from all active base of billing system dump table. Those who existing in billing dump table means not churned yet and those who has been removed from dump table means already churned.

Finally, we randomly selected 100K of numbers from this segregated data for developing predictive model using Python language. The 100K numbers of customer exported into a CSV file then passed to python code for prediction of churners.

Figure 4: Process & Data flow diagram for data collection and prediction of customer churn in telecom



Feature Engineering

In this research we have developed a procedure collecting data from each network elements mentioned in data preparation section. The procedure run on daily

basis and collect 24 hours back data from CDR, TOPUP and MSC per day per MSISDN (number of customer). The data is 24 hours back because the Data warehouse team in ROSHAN dumps CDR data for yesterday each day. The data size in telecom is huge and can't be processed on real time basis. Secondly, we collected the tickets Information of the customer from CRM per day per MSISDN and merge it with CDR data. Then we took the aggregated data of one year as our final Dataset and updated the information of churn that shows a customer is churned or still active. The final Dataset looks like as in Table 5.

Table 4: Final dataset for predicting customer churn.

Variables	Type	Description
MSISDN	numeric	Number of Customer
GRNDER	numeric	1 is Mala 2 is Female
PROILE	string	Is the type of Sim Card
AON	numeric	Age of Network, this is the lifetime of customer, there number are in days
ONNET CALL	numeric	The total usage of call that comes from ROSHAN to ROSHAN
OFFNET_CALL	numeric	The total usage of call that comes from ROSHAN to other providers
INTERNTIONAL CALL	numeric	the total usage of sms come from ROSHAN to ROSHAN
ROAMING CALL	numeric	the total usage of sms come from Roaming where A party the customer is outside the country
ONNET SMS	numeric	The total usage of data
OFFNET SMS	numeric	The total money spend on bundle subscription (activation)

INTERNATIONAL SMS	numeric	The total usage of SMS coms from international party
ROAMING SMS	numeric	The total usage of SMS comes from roaming where a party the costumer is outside of country
DATA	numeric	The total usage data
BUNBLE AMOUNT	numeric	The total money spend on bundle subscription (activation)
BUNBLE COUNT	numeric	The total time that customer activation bundle
TOPUP AMOUNT	numeric	The total amount toped up by customer to their account
TOPUP COUNT	numeric	The total time that customer toped up their account
INCOMING	numeric	The total time that customer receive call
INCOMING DURATION	numeric	The total minutes that customer receive call
LAST RGE DATE	numeric	Last date of activity
Last top up date	numeric	Last date of top up
LAST USAGE DATE	numeric	Last date of usage
GHURN FLAG	numeric	Whether the customer churned or still active .1 is churned and 0 is still active
TOTAL COMPLAIN	numeric	Total complain that customer done
TOTAL INQUIRY	numeric	Total INQUIRY that customer done
TOTAL REQUSET	numeric	Total request that customer done
PRVINCE	string	What is the province of customer

Data Set Extraction and Implementation

The random data of around 100K customers extracted from the 7 million base that include churn and non-churn customer. The extracted dataset is in CSV file and the script to load CSV and apply the XGBOOST algorithm is prepared in Python language. The code related to python script. Figure 6 shows a sample of CSV file.

The Data set is then split into training and testing model, and finally we applied the XGBOOST algorithm, to predict the customers that are in a high risk of going for churn. Further, the script gets the predicted customers as a result of XGBOOST algorithm implementation and save it in an output file. The script further loads the output file and classify risky customer (churners) based on Gender, Profile, Province, and AON (Age on network) for the company management team. The charts help the management team in taking proactive steps to retain the customers.

Figure 5: Sample view of extracted dataset

MSISDN	GENDER	PROFILE	AON	ONNET CALL	OFFNET CALL	INTERNATIONAL CALL	ROAMING CALL	ONNET SMS	OFNET SMS	INTERNATIONAL SMS	ROAMING SMS	DATA	BUNDLE AMOUNT	BUNDLE COUNT	TOPUP AMOUNT	TOPUP COUNT	INCOMING	INCOMING DURATION	LAST RGE DATE	LAST TOPUP DATE	LAST USAGE DATE	CHURN FLAG	TOTAL COMPLAIN	TOTAL INQUIRY	TOTAL REQUEST	PROVINCE
93728304xxx	1	AALI	260	5.8	22.75	0	0	0	0	0	0	0	0	0	0	0	5	109	20220816	20220504	20220811	1	0	0	0	Bamyan
93793244xxx	1	SAADAT	593	10.6	32.8	0	0	12.5	10	10	0	0	5	1	75	2	28	674	20220909	20220822	20220909	0	0	1	0	Balkh
93793288xxx	2	SAADAT	809	0	6.6	0	0	0	0	0	0	0	1500	3	1660	3	118	0	20230407	20230124	20230407	0	0	4	0	Faryab
93793566xxx	1	SAADAT	853	38.95	113.8	10	0	0	0	0	0	0	0	0	100	2	22	414	20230118	20230117	20230118	0	0	0	0	Urozgan
93793613xxx	1	AALI	839	117.475	9.9	0	0	2.5	0	0	0	0	242.5	5	450	9	271	44639	20230514	20230321	20230509	0	0	1	0	Farah
93793632xxx	1	SAADAT	853	227.275	247.5	12	0	0	5	0	0	0	0	0	550	9	260	71223	20230514	20230506	20230514	0	0	0	0	Badghis
93791792xxx	1	YARAAN	1134	24.05	10.55	0	0	3	2.5	0	0	0	190	3	240	4	231	1772	20230514	20230504	20230512	0	0	0	0	Bamyan
93797341xxx	1	SAADAT	977	84.175	85.8	0	0	0	0	0	0	0	500	9	730	8	352	6327	20230514	20230420	20230420	0	0	0	0	Kapisa
93797331xxx	1	GOVT	958	40.5	0	0	0	0	0	0	0	0	447.5	7	295	3	224	2238	20230514	20230420	20230420	0	0	1	0	Balkh

Data Analysis

The data prepared for this study is real time data of ROSHAN telecom company. We have collected one-year data for each customer that shows behavior trend of customer. Then the churn flag for each customer updated with 0/1, 0 means non churn and 1 means churn. Finally, we exported randomly 100k customer in a file named export.csv and load the file in python program that has been prepared. The code split the dataset into train and test dataset 80:20 percentages. Then apply the XGBOOST algorithm to classify and predict the churn customer and gives us output file "out.csv" that include those customers that has higher risk to churn. According to the research question that is "How to classify customers and predict churners using machine learning XGBOOST algorithm." We have solved the question and the result shown in Table 6.

Table 5: The result of research question.

Evaluation Question	Data Sources	Data Analysis
RQ1	Data gathered from different sources.	The data has been classified into Gender Base, Profile Base, AON base, and churn / non-churn base
RQ2	100k records randomly selected and splitting into 80:20 percentage to train and test model.	Got the result of 4127 customers at high risk of going to churn, with the accuracy rate of 95.32%

Model Implementation

The XGBoost algorithm was selected for its efficiency and scalability in handling large datasets with complex relationships. The dataset was split into training (80%) and testing (20%) sets. Hyperparameter tuning was performed using grid search to optimize model performance, with evaluation metrics including accuracy, precision, recall, and the confusion matrix.

RESULTS

Model Performance

The XGBoost model achieved an accuracy of 95.32%, with a high precision and recall indicating robust predictive capability. The confusion matrix revealed a balanced classification of churners and non-churners, demonstrating the model's effectiveness in handling imbalanced data.

Model Evaluation

The Model evaluated using two machine learning technique, Accuracy and Confusion Matrix. Accuracy is a common metric used to evaluate how well a classification model predicts the correct labels of a dataset. In XGBoost, accuracy is calculated by dividing the number of correctly classified instances by the total number of instances in the validation dataset. In binary classification, accuracy is calculated using true positives, true negatives, false positives, and false negatives. In multi-class classification, accuracy is calculated by summing up the number of correctly classified instances across all classes and dividing by the total number of instances.

A confusion matrix is a table that compares the actual and predicted class labels of a set of data points to evaluate the performance of a classification model. It shows the number of true positives, true negatives, false positives, and false negatives for each class. The accuracy for the Dataset is 95.32% in Table 5 shows the confusion matrix table.

Table 6: Confusion Matrix (compares the actual and predicted class labels to evaluate the performance of model)

	Predicted False	Predicted True
Actual False	35604	1510
Actual True	797	12089

Confusion Matrix identified that 35604 customers are non-churners, and 1510 customer are the churners, therefore the number of non-churners is more than number of churners, which tells us that the classification accuracy is correctly tested.

Future Work

While the model shows strong predictive performance, its reliance on historical data may limit its ability to capture sudden shifts in customer behaviour. Future research could explore the integration of additional data sources, such as social media sentiment analysis, to enhance the model's adaptability. Moreover, expanding the scope to include other machine learning techniques like neural networks could provide further improvements in predictive accuracy.

CONCLUSION

This study developed a customer churn prediction system using XGBoost, tailored for the telecom industry. The model's high accuracy and practical insights provide a valuable resource for ROSHAN Telecom in Afghanistan to enhance customer retention efforts. Future work could explore the integration of additional data sources and machine learning techniques to further refine the predictive model.

In overall, the results indicate that the churn prediction system is effective in identifying at-risk customers and providing valuable insights for the telecom company to improve customer retention. The system can help the company take proactive measures to retain customers and reduce churn, which can lead to improved revenue and customer satisfaction. The analysis of customer segmentation based on gender, profile, province, and AON provides useful information for the company to tailor retention campaigns and offers to specific customer groups. The total churn per province, AON, and profile can identify areas where the company may need to improve its services or offer targeted promotions to retain customers. The predicted churn customer call graph and ticket percentage can help the company understand the behavior and preferences of at-risk customers, which can lead to more effective retention campaigns. Overall, the churn prediction system can provide significant benefits to the telecom company and help improve customer retention and satisfaction.

ACKNOWLEDGMENT

No acknowledgements to declare.

REFERENCES

Abdelrahim Kasem Ahmad, A. J. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 1 - 24.

Adnan Amin, F. A.-O. (2018). Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research , 1-12.

Ammar A Ahmed, D. D. (2017). A Review And Analysis Of Churn Prediction Methods For Customer Retention In Telecom Industries. International Conference on Advanced Computing and Communication Systems, (pp. 1 - 7). Coimbatore, INDIA.

Benjamin Ghaffari, Y. O. (2021). Customer churn prediction using machine learning: A study in the B2B subscription based service context. Sweden: Blekinge Institute of Technology.

Charalampos Alexopoulos, Z. L. (2019). How Machine Learning is Changing e-Government. ICEGOV, 1-10.

F. F. Reichheld, W. E. (1990). Zero defeofions: quoliiy comes to services. Harvard Business Review, 105-111.

Ibrahim AlShourbaji, N. H. (2021). Customer Churn Prediction in Telecom Sector: A Survey and way ahead. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 1 - 12.

Irfanullah, B. R. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. IEEE Access, 60134 - 60149.

Joolfoo, M. (2020). Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. University of Mauritius: ResearchGate.

Kristof Coussement, S. L. (2016). A Comparative Analysis of Data Preparation Algorithms for Customer Churn Prediction: A case study in the telecommunication industry, Decision Support Systems, 1-37.

Ming Zhao, Q. Z. (2021). A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China. Hindawi, 1 - 12.

N. Alboukaey, A. J. (2020). Dynamic behavior based churn prediction in mobile telecom. Expert Systems with Applications.

Pamina J, B. R. (2019). An Effective Classifier for Predicting Churn in Telecommunication. Jour of Adv Research in Dynamical & Control Systems, 1-9.

Praveen Lalwani, M. K. (2021, January 12). Customer churn prediction system: a machine learning approach. pp. 1 - 24.

Sagar Maan Shrestha, A. S. (2022). A Customer Churn Prediction Model using XGBoost for the Telecommunication Industry in Nepal Telecommunication Industry in Nepal., ScienceDirect, 652 - 661.

Samah Wael Fuju, S. S. (2022). Customer Churn Prediction in Telecommunication Industry using Deep Learning. Information Sciences Letters, 15.

Sudarshan Zimal, C. S. (2023). Customer Churn Prediction Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 14.

Swetha P a, D. R. (n.d.). Customer churn prediction in telecommunication industry through machine learning based Fine-tuned XGBoost algorithm. Bangalore, India: Department of CSE, KS School of Engineering and Management.

Swetha P, D. R. (2020). Improvised_XgBoost Machine Learning Algorithm for Customer Churn Prediction. EAI.EU, 1 - 7.

T. O. Jones, W. E. (1995). Why satisfied customers defect. 88.

Tianpei Xu, Y. M. (2021). Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping. MDPI, 1-12.

Vafeidis T, D. K. (2015). A comparison of machine learning techniques for customer churn prediction. CrossMark, 1-9.

Vishal Mahajan, R. M. (2017). Review on factors affecting customer churn in telecom sector. Inderscience Enterprises Ltd, 1-23.