To Predict Customer Churn By Using Different Algorithms

1st MD Sahrial Alam
Department of Computer Science and Engineering
Bangladesh University Of Business And Technology - BUBT
Dhaka-1216, Bangladesh
sahrialalam@gmail.com

2nd MD Sayedur Rahman
Department of Computer Science and Engineering
Bangladesh University Of Business And Technology - BUBT
Dhaka-1216, Bangladesh
srsayed2677@gmail.com

3rd MD Ikbal Hosen
Department of Computer Science and Engineering
Bangladesh University Of Business And Technology - BUBT
Dhaka-1216, Bangladesh
mdiqbalhossain6463@gmail.com

***Abstract*—Customer churn prediction is the supreme thing for every company. Every company is facing that issue. A company can easily predict how many customers will stay and take their service and how many customers will omit their service. This paper predicts if their customers will take their service in the future or not. In our paper, we have used the " Impact Learning" algorithm to predict customer churn. The data is trained by the impacts of features from the intrinsic rate of natural increase in the impact learning algorithm. Here, the training set of data is numerical data. Our proposed model is implemented by using three stages namely data collection, identifying null value, and data preprocessing. This paper has also shown the performance comparison between Logistic regression, Impact learning, and Artificial Neural Network algorithm. From the comparison, we can see that the Impact learning algorithm gives the best and higher accuracy than the other two algorithms.**

***Index Terms*—Customer Churn, Impact Learning, Rate of Natural Increase(RNI), Label Encoder, Polynomial Regression.**

I. Introduction

At present in this competitive marketplace, the profit of a company depends on the customer. The customer has become a valuable asset of a business. So keeping an existing customer is an important concern for a business. Customer churn prediction refers to the percentage of customers or users who are unsubscribing for their service. Customer churn has become one of the most important parts of business today for growing a business.

Nowadays the internet is the most important thing in our day-to-day life. Internet-related business like telecommunication business is an important part of the economy. In telecommunication, the customer is the main resource for business. So, it's very important to keep customers and predict how many customers churn and why. Using accurate churn prediction, any company will be able to understand the revenue of the business and will take the necessary steps to reduce the churn rate and analysis the reason for churn. For any business, an organization that predicts the churn rate is important so that they can understand why customers stop taking their services and how it will impact the business. There are many approaches for predicting customer churn. Machine learning approaches are one of them.

Recently Machine learning procedures are being very significant in lots of fields. Impact learning is one the most unique in recent. This algorithm is enriched for being capable of learning from a competition, which is the impact of independent features [1]. "Impact Learning" is a parametric statistical learning system and observes a different learning methodology. In demography, the rate of natural increase (RNI) [2] is a statistical concept and used in environmental science more. As data comes from reality, to make it freestanding, the RNI also can be used for building ML or data science applications.

The rest of the paper is summarized as follows: A brief review of some existing research work is provided in Sect. 2. In Sect. 3, a detailed description of our proposed framework is presented. In Sect. 4., our experimental result is shown. In Sect. 5., we discuss performance analysis. Finally, a conclusion section is provided in Sect. 6.

II. Related Work

In [3], to forecast churn in the banking industry, they employ the IBRF approach. IBRF is a new way of learning. It incorporates sampling and cost-sensitive learning techniques.

According to [4], the widespread use of the Internet of Things (IoT) and cloud computing environments makes it possible to collect customer data for churn prediction and CCP.In a cloud computing system, the CCP model employs a machine learning algorithm. Data selection, preprocessing, and Adaptive Gain with Back Propagation Neural Networks are all used in the CCP model (AGBPNN). But this model needs to improve by using feature selection and clustering techniques.

In [5], Relational learning algorithms are used to predict customer churn using social network knowledge. To monitor large scale networks, a time-based class mark, and a distorted class distribution, this algorithm was designed to integrate social network effects within a customer churn prediction environment. However, this model requires relational classifiers to integrate non-Markovian network effects.

In [6], a novel framework called Group-First Churn Prediction is used to predict consumer churn in mobile networks by analyzing social groups. It does away with the need to know who has recently churned a priori. They would have achieved better results if they used individual-based models.

In [7], the Data Mining (DM) model will correctly identify potential churners. The six steps in this model are: identifying the problem domain, data collection, investigating the data set, classification, clustering, and knowledge application. The Data Mining model was created to assist a CRM department in keeping track of its customers and their actions to prevent churn. However, the model was unable to predict and cluster data to allocate appropriate retention strategies to each churner class.

According to [8], in customer churn prediction, decision trees and logistic regression are two common algorithms with high predictive efficiency. The logit leaf model (LLM), a new hybrid algorithm, is proposed to better classify data. The LLM states that different models built on parts of the data rather than the entire dataset perform better in terms of prediction while preserving the comprehensibility of the models built in the leaves. It's conceivable that the model could be improved further by imposing more complex rules for the number of leaves and leaf sizes.

The average customer churn rate in the mobile telecommunications industry is 2.2 percent per month, with nearly 27 percent of customers canceling their service [9]. The internet is one of the most critical aspects of our lives. Our internet service provider (ISP) provides us with internet access. Every month, a large number of customers discontinue their service for a variety of reasons. In ISP [10], data mining is used to estimate customer churn. The data mining method KDD (Knowledge Discovery in Database) is used to predict the rate of customer churn.

III. Proposed Methodology

A. Stage I: Data Collection:

The consumer churn prediction dataset can be found on the kaggle website. There are 7043 distinct values in the dataset. There are 21 columns in the dataset with 7043 rows. In that data, 5174 customers were pleased with their service, while 1869 customers were not. This dataset contains 7043 instances, each of which has 20 features. Sex, Senior Citizen, Partner, Dependents, tenure, Phone, Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing Payment Method, Payment Method, Monthly Charges, Total Charge, and a label called Churn are among the features.

TABLE I
Details Of Dataset

Description	Dataset
Instance Count	7073
Feature	20
Class Count	2

B. Stage II: Data Preprocessing:

Before training a dataset, it must be preprocessed to detect and handle null values and other errors. After selecting the dataset, we must first determine if it contains any null values. If a null value is found, use the mean formula to fill in the blanks.

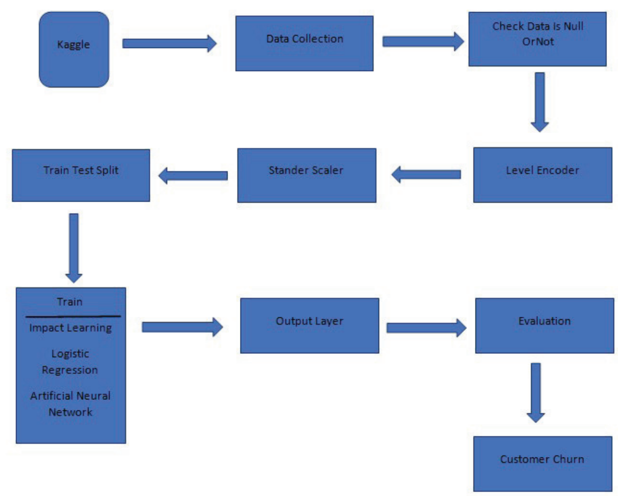

Fig. 1. Developing a Framework for Customer Churn Prediction

We can see that the majority of the values in our dataset are categorical, so we can convert them to numerical after finding Null Values. Since the majority of data values are ordinal, we prefer Label Encoder. The SciKit Learn documentation for Label Encoder can be found here.

Algorithm of Label Encoder...
1: Take the features values as input
2: for column in df.columns
3: if df[column].dtype == np.number
4: continue
5: df[column] = LabelEncoder().fit transform(df[column])
6: end if
7: end for

For loop is used for checking if column values are numerical or not. If Column values are numerical then the continue function is used. After that the LabelEncoder().fit transform() method is used for converting categorical values into numerical values.

After that, the dataset was Label Encoded and Standard Scaler was used to render a low range of numerical distance. The SciKit Learn documentation for the Standard Scaler function can be found here.

feature x = StandardScaler().fit transform(x)

Here x is the number of features. All features come to a low range of numerical distance by using Standard Scaler.

C. Stage III: Data Train:

Impact learning is a supervised machine learning algorithm for resolving classification and linear or polynomial regression knowledge from examples. It also adds value to applications for analyzing vast amounts of data in a competitive environment [1].

Here, we are considering r as RNI Normally, the RNI's trend gets obstructed for every element by a limitation, and we address that as the term carrying capacity(K). This concept can be better realized from the logistic growth model [11] of environmental science and statistics.

The target features cannot walk in the direction of the RNI's curve because of Back Impact on Target (BIT). If x is the back-impact variable, then x and y both keep their impact on each other's, and y keeps its impact on itself.

Algorithm of Impact learning...

1: Taken input of all the features
2: sum=0
3: If(index¡= number of features)
4: for(index =1;index¡= number of features ;index++)
5: Sum=Sum+ coefficient(index)*feature(index)
6: break
7: end for
8: else
9: A=carrying capacity*sum
10: A=A/(RNI-target coefficient*carrying capacity)
11: Y=A+ bias

IV. Result Analysis

In our dataset using impact learning algorithm used default optimizer 'GD' and loss function="Categorical Crossentropy" we used 6000 epochs and per progress 300 then get the result is 81.06957 % accuracy.

Before Train, the Customer Churn dataset has 26.537 % of customers were left the company and 73.463 % of customers have satisfy the company service and stay.

TABLE II
Before Train, The Customer Churn Dataset Percentage

Customer Churn Percentage	Customer Churn Percentage
26.537 %	73.463 %

After Train, the Customer Churn dataset has 27.165 % of customers have left the company and 72.834 % ofcustomers were satisfy the company service and stay.

TABLE III
After Train, The Customer Churn Dataset Percentage

Customer Churn Percentage	Customer Churn Percentage
27.165 %	72.834 %

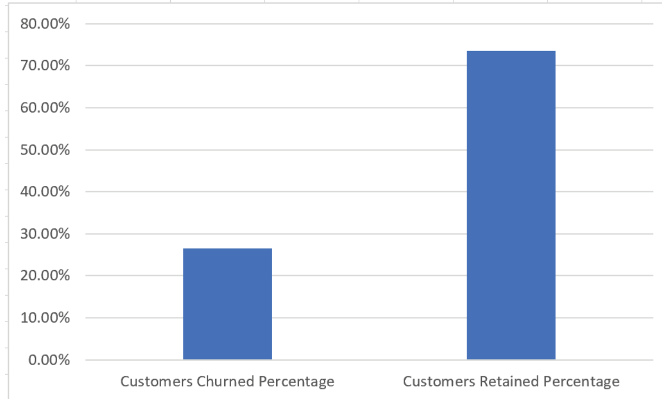

Fig. 2. Before training the Customer Churn dataset Percentage

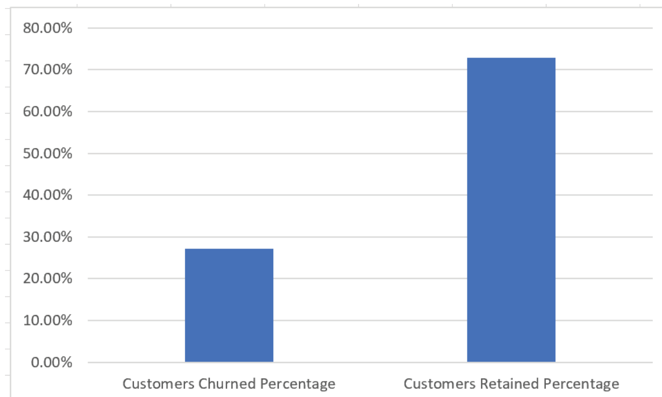

Fig. 3. After training the Customer Churn dataset Percentage

V. Performance Analysis

Logistic regression is a part of supervised learning. It is a classification algorithm. It works both continuous or discrete but always gives discrete value output. Logistic Regression is an effective analytic method that is used to derive and test hypotheses about relationships between a categorical variable and one or more categorical or continuous variables [12]. An

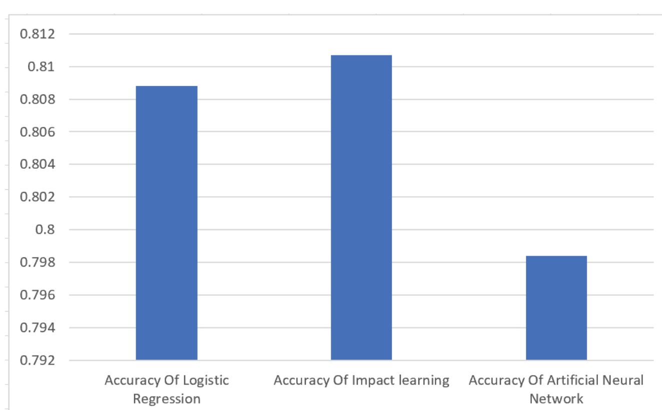

Fig. 4. Analysis between different algorithms

artificial neural network is the pieces of a computing system inspired by biological neurons. Artificial Neural Network has three-layer which are input, hidden, and output layer. Nowadays, ANN is used for image processing and data prediction and analysis data. In ANN, we used optimizer='adam' loss='binary crossentropy' for compiler. It is feed-forward multilayer process.

We have compared our proposed model with the above algorithms and get the following the output

TABLE IV
DETAILS OF DATASET

Methods	Accuracy
Logistic Regression	80.88 %
Impact Learning	81.06 %
Artificial Neural	79.84 %

From the above table, we can see that our proposed model "Impact Learning" algorithm gives the best accuracy among these two algorithms which is 81.06% where as logistic regression gives 80.88% and ANN gives 79.84%.

VI. CONCLUSION

This research paper is an experimental investigation of customer churn prediction based on real data sets. It has presented the effectiveness of the Impact learning model in the prediction environment. We can see that Impact learning has 81.06% accuracy. This is much better than Artificial Neural Network and Logistic regression models. We could get a better result if we tune the data first and then use the CNN algorithm.

REFERENCES

[1] M. Kowsher, A. Tahabilder, and S. A. Murad, "Impact-learning: a robust machine learning algorithm," in Proceedings of the 8th International Conference on Computer and Communications Management, 2020, pp. 9–13.

[2] A. Taberner, P. Castaera, E. Silvestre, and J. Dopazo, "Estimation of the intrinsic rate of natural increase and its error by both algebraic and resampling approaches," Bioinformatics, vol. 9, no. 5, pp. 535–540, 1993.

[3] Y. Xie, X. Li, E. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," Expert Systems with Applications, vol. 36, no. 3, pp. 5445–5449, 2009.

[4] M. Jeyakarthic and S. Venkatesh, "An effective customer churn prediction model using adaptive gain with back propagation neural network in cloud computing environment," Journal of Research on the Lepidoptera, vol. 51, no. 1, pp. 386–399, 2020.

[5] W. Verbeke, D. Martens, and B. Baesens, "Social network analysis for customer churn prediction," Applied Soft Computing, vol. 14, pp. 431–446, 2014.

[6] Y. Richter, E. Yom-Tov, and N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups," in Proceedings of the 2010 SIAM international conference on data mining. SIAM, 2010, pp. 732–741.

[7] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, "A proposed churn prediction model," International Journal of Engineering Research and Applications, vol. 2, no. 4, pp. 693–697, 2012.

[8] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," European Journal of Operational Research, vol. 269, no. 2, pp. 760–772, 2018.

[9] L. Yan, R. H. Wolniewicz, and R. Dodier, "Predicting customer behavior in telecommunications," IEEE Intelligent Systems, vol. 19, no. 2, pp. 50–58, 2004.

[10] A. A. Khan, S. Jamwal, and M. M. Sepehri, "Applying data mining to customer churn prediction in an internet service provider," International Journal of Computer Applications, vol. 9, no. 7, pp. 8–14, 2010.

[11] X. Chen, X. Zhang, J. A. Church, C. S. Watson, M. A. King, D. Monselesan, B. Legresy, and C. Harig, "The increasing rate of global mean sea-level rise during 1993–2014," Nature Climate Change, vol. 7, no. 7, pp. 492–495, 2017.

[12] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1659–1665, 2012.