

# Advanced Algorithm for Prediction of Churn in Various Industries in the Fast Growing World

Ajay Varma, Dr. Rajesh Deshmukh  
Kalinga University, Raipur

## Abstract

In the study it has been proven that the cost of acquisition of a new customer is much more than the cost of the retention of the existing customer. Also, it becomes very easy for the organizations if they come to know of the customers that are likely to get churn in advance studying the behavioral aspects so that they can take appropriate measures to keep the customer in their own territory. So, there has been a lot of study on the existing algorithms to understand which can provide the better accuracy in terms of the prediction analysis of their customers. This can be very useful for the service providers to in order to maintain trust and loyalty towards their customers and in good will against their competitors.

**Keywords:** Customer, Prediction Analysis, Behavioral aspects, Cost of acquisition, Cost of retention

## 1. INTRODUCTION

It is not much importance when you get a new customer, how it has been treated and the way in which he or she is satisfied by the service provider is being important. Also, the time for which the customer is associated with the service provider gives the impression of the goodwill of the service provider when it comes to research of the market to determine which is the best service provider in the market. So, the organizations have been very keen in identifying the customer that are prone to getting churn or discontinuing the services of the organization in the near future so that they can give them appropriate measures which can defy the customers from getting churn. Also, the word of mouth can also be very essential to get the new customers onboard based on the review of the existing customers. So, there has been a lot of research work already done in this regard to devise an algorithm which can help predict the churn probability of all the customers analyzing the data of the customers from the organizations, considering various key factors which can determine the chances of the customers getting discontinued from the service provider.

In the past times, many service providers have devised various strategies to get hold of the existing customers for the longer duration of time. Also understanding the competitor's strengths and weaknesses is another important factor out. Out of the currently existing domains, the telecommunication industry has been a talk of the town. Before 10 years, there was not much competition as there were not many service providers. Now, with the growth of the number of service providers, it has been a very competitive environment nowadays, as the customer has got a lot of options for the service providers, and the concept of porting to another service provider also provides the options to keep the mobile number same and switch to any other provider. Companies have been using

various strategies to keep the customers aligned with them for longer duration and various schemes to attract more customers. In fact when we talk about the service sectors, customer movement from one service provider to another is considered as a very big matter of concern. Various studies have shown that the machine learning techniques have been very efficient in predicting the churn probability of the customers.

## PROBLEM IDENTIFICATION

The main purpose of the paper is to suggest a new churn prediction algorithm where we can give the probability of the churning of the potential customers for any organization. Also can provide the main deciding factor which has been dominant in the identify features of attributes set which caused the churn possibility of that corresponding customer.

## OBJECTIVES

- To take the example of a dataset from a telecom industry and apply the existing as well as the proposed algorithm to get the churn probability of the customers. The data has been collected from last 3 years.
- To calculate the accuracy of the model with the existing techniques for comparison. Also, to get the attributes list where we can determine what is the possible reason for the churn of the respective customer.

## 2. RESEARCH WORK ALREADY DONE

A lot of research work has been done in the field of churn analysis, taking in account of the level of competition that has been increasing in the various domains. Several methods have in this scenario, which has been mainly focus on applying data mining techniques to extract features that can be used to determine the probability of the customer being

churn with the maximum chances. been applied Many approaches were applied to predict churn in telecom companies.

Customer churn prediction has been widely researched, specifically within the domain of B2C. According to Gordini & Veglio [9], customers get churn owing to a cheaper cost structure compared to the one provided by the current service provider. Also it can be seen that the new service provider is providing more services according to their needs than the existing provider. Most of the studies has been done in the case of the subscription based models, where the customers chose some kind of subscriptions from the service providers which they later discontinue and shift to the new service provider. Some of the prediction methods include time to churn prediction [1,2] and other use the machine learning methods along with the time-to-churn prediction [21, 19, 22, 16, 17, 23, 10]. Both the methods determine the churns, however the later method determines the time in which the particular customer is probable to churn. And the former determines whether the customer get churn in the current scenario.

Mostly the research work like Vafeiadis et al. [21] that has been done in the past has been focused on the machine learning classifiers that exists to determine the churn probability. The telecommunication sector has been the main area of interest for researchers. Some researchers like Brandusoiu et al. [22], used the data mining techniques for the improvement of the performance and accuracy of the prediction algorithm. They used churn and non-churn as the categorical variables using Support Vectors machines combined with Bayesian Network and Neural Networks. The calculation metrics were confusion matrix and a gain measure , which in turn gave the accuracy of 99 and 100 percent. They also used the dataset from the telecom industry, where the data of how much calls, texts and complaints given by each and every customer is being taken into account.

Also, the very important point that needs to be taken care when dealing with the churn prediction is the domain in which the prediction is being done, whether it is Business to Customer (B2C) or Business to Business (B2B).

Based on the study done by Figalist et al. [13], churn prediction gets more importance when we talk about the B2B than B2C, since the customer churn affects a large amount of sum when churn happens in Business2Business then Business2Customer.

Based on the results Figalist et al. [13], conclude that churn prediction within B2B is often of greater significance relative to B2C since companies within B2B stand to lose a greater deal when a customer churn, compared to a customer within B2C.

### 3. STEPS INVOLVED IN THE CHURN

### PREDICTION ALGORITHMS

The churn prediction algorithms has been divided into various steps which in turn gives us the detailsof each and every customer with the churn probability and then comparison of the data with the existingcustomer.

#### a. DATA COLLECTION

We gathered data from various source systems of the telecom provider. Also, we integrated data from heterogeneous sources into a common format, so that we can have a same structure to make the datacomparison etc. very easy.

#### b. DATA PREPROCESSING

Preprocessing of the data has been done to eliminatethe duplicate values, null values, outliers etc.

#### c. EXPLORATORY DATA ANALYSIS 2(EDA)

We determined the main feature set that will play a important role in the prediction of the churn probability of the customers. Also, weightage has been given to the each feature so that we can determine which has importance and which has less importance.

#### d. DATA SPLITTING

We have divided the data into training and testing database with 80%-20% so that the training dataset can be used to train the model for the datasets.

#### e. MODEL SELECTION

The model selection has been done on the existing dataset, with the existing all models along withour proposed algorithms also being applied on the same. So as to compare the accuracy measure of thealgorithms.

#### f. MODEL TRAINING

Train the selected model on the training dataset. The model learns the patterns and relationships between features andthetarget variable (churn) during this phase.

#### g. MODEL EVALUATION

Evaluate the model's performance on the testing dataset using appropriate metrics such as accuracy,precision, recall, F1 score, or area under the receiver operating characteristic (ROC) curve. This step helps assess how well the model generalizes to new, unseen data.

#### h. MONITORING AND UPDATING

Continuously monitor the model's performance in a production environment. If necessary, update the model to adapt to changing patterns and ensure its effectiveness over time.

### 4. CHALLENGES

- The main challenge with the designing of the new churn prediction model is to get the data of various sources of the

organization into a integral format. As there might be data available in chunks in different source systems due to which the data may be in different formats and structures. Also, the accuracy of the model has been determined with the churn data that we collected from the organization.

- Apart from the telecom industry, the IT companies has been very keen to keep the track of the churn probability of the employees, which in turn can create a lot of issues for the companies, like searching for the replacement of the employee, spending a new amount of money to the new employee, also training it on the new technologies etc. as the exact matching skill set is very difficult to find.
- Also, as in the current scenario, there has been a notice period of 90 days, due to which a organization faces quite issues with the onboarding of the new employee, since he/she has been associated with the existing company so it may need 30-60 days to get relieved from the services of the existing company. So the waiting time is also considered, meaning the company would need to wait till the new employee joins.

### 5 DISCUSSION OF RELATED WORKS

Churn prediction is a crucial task in customer relationship management, and various algorithms can be employed to predict and mitigate customer churn. Here's an overview of some commonly used algorithms for churn prediction:

- **Logistic Regression:**
  - A simple and interpretable algorithm that models the probability of churn based on input features.
  - Works well when the relationship between features and churn is linear.
- **Decision Trees:**
  - Decision trees are used to make decisions based on input features.
  - Can handle both numerical and categorical data.
  - Prone to overfitting, but techniques like pruning can be used to mitigate this.
- **Random Forest:**
  - An ensemble method that builds multiple decision trees and combines their predictions.
  - Reduces overfitting compared to individual decision trees and provides higher accuracy.
- **Gradient Boosting Machines (GBM):**
  - An ensemble technique that builds trees sequentially, with each tree correcting the errors of the previous ones.
  - XGBoost and LightGBM are popular implementations of GBM.
- **Support Vector Machines (SVM):**
  - SVM aims to find a hyperplane that best separates the churn and non-churn instances.
  - Effective when dealing with high-dimensional data.

- **Neural Networks:**
  - Deep learning models, such as feedforward neural networks or more complex architectures like recurrent neural networks (RNN) or long short-term memory networks (LSTM).
  - Effective for capturing complex relationships in large datasets but may require more computational resources.
- **K-Nearest Neighbors (KNN):**
  - Classifies a data point based on the majority class among its k-nearest neighbors in the feature space.
  - Simple and effective, but sensitive to the choice of distance metric.
- **Naive Bayes:**
  - Assumes independence between features and calculates the probability of churn based on Bayes' theorem.
  - Simple and computationally efficient, but may not capture complex dependencies.
- **Ensemble Methods:**
  - Combining predictions from multiple models, such as bagging or stacking, can often lead to better overall performance.
- **Time Series Analysis:**
  - For scenarios where historical data and temporal patterns are essential, time series models like ARIMA or exponential smoothing can be valuable.

### 6. RESULTS & ANALYSIS

We applied the existing algorithms and the proposed algorithm on the selected telecom domain data. And get the results of their accuracies:

Model	Accuracy
Proposed Algorithm	95.0%
Random Forest	78.75%
Decision Tree	72.36%
Support Vector	79.37%
Adaboost	73.64%
Advanced Adaboost	79.6%

### 7 CONCLUSION

This paper proposed a new prediction algorithm which can be very useful in the prediction of the churn probability of the customers for any organization. Also, we compare the different existing algorithms with the new algorithms, thereby improving its accuracy on the most difficult and trending domains data.

REFERENCES

- [1] M. Masarifoglu and A. H. Buyuklu , "Applying Survival Analysis to Telecom Churn Data," *American Journal of Theoretical and Applied Statistics*, vol. 8, no. 6, pp. 261- 275, 23 November 2019.
- [2] L. Junxiang, "Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS," *Sprint Communications Company*, 2002.
- [3] Agrawal, S., Das, A., Gaikwad, A., & Dhage, S.(2018, July). Customer churn prediction modelling based on behavioural patterns analysis using deep learning. In 2018International conference on smart computing and electronic enterprise(ICSCEE)(pp.1-6).IEEE. <https://doi.org/10.1109/ICSCEE.2018.8538420>
- [4] Ahmad, A. K., Jafar, A., &Aljoumaa, K. (2019).Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), 1-24.DOI: <https://doi.org/10.1186/s40537-019-0191-6>
- [5] Ahmed,U.,Khan,A.,Khan,S.H.,Basit,A.,Haq, I. U., & Lee, Y. S. (2019). Transfer learning and meta classification based deep churn prediction system for telecom industry .arXiv preprintarXiv:1901.06091.<https://doi.org/10.48550/arXiv.1901.06091>
- [6] Alboukaey, N., Joukhadar, A., &Ghneim, N.(2020). Dynamic behavior based churn prediction in mobile telecom. *ExpertSystemswithApplications*,162,113779. <https://doi.org/10.1016/j.eswa.2020.113779>
- [7] Amin, A., Al-Obeidat, F., Shah, B., Adnan, A.,Loo, J., & Anwar, S. (2019a). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94, 290-301. DOI:<https://doi.org/10.1016/j.jbusres.2018.03.003>
- [8] Amin, A., Anwar, S., Adnan, A., Nawaz, M.,Alawfi, K., Hussain, A., & Huang, K.(2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neuro computing*, 237,242-254. <https://doi.org/10.1016/j.neucom.2016.12.009>
- [9] Amin,A.,Anwar,S.,Adnan,A.,Nawaz,M., Howard, N., Qadir, J., ... & Hussain, A.(2016). Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study. *IEEE Access*, 4, 7940-7957.<https://doi.org/10.1109/ACCESS.2016.2619719>
- [10] Amin,A.,Khan,C.,Ali,I.,&Anwar,S.(2014). Customer churn prediction in telecommunication industry: with and without counter-example. In *Mexican international conference on artificial intelligence* (pp. 206-218). Springer, Cham. [https://doi.org/10.1007/978-3-319-13650-9\\_19](https://doi.org/10.1007/978-3-319-13650-9_19)
- [11] N. Gordini and V. Veglio, "Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter- selection technique in B2B e-commerce industry," *Industrial Marketing Management* , vol. 62, pp. 100-107, 2017.
- [12] Amin,A.,Shah,B.,Abbas,A.,Anwar,S.,Alfandi, O., & Moreira, F. (2019b).Features weight estimation using a genetic algorithm for customer churn prediction in the telecom sector. In *World conference on information systems and technologies* (pp. 483- 491). Springer,Cham. [https://doi.org/10.1007/978-3-030-16184-2\\_46](https://doi.org/10.1007/978-3-030-16184-2_46).
- [13] I. Figalist, C. Elsner, J. Bosch and H. H. Olsson, "Customer Churn Prediction in B2B Contexts," *Conference: International Conference on Software Business*, 2020.
- [14] Amin,A.,Shah,B.,Khattak,A.M.,Moreira,F.J.L., Ali, G., Rocha, A., & Anwar, S.(2019c). Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. *International Journal of Information Management*, 46, 304-319.DOI: <https://doi.org/10.1016/j.ijinfomgt.2018.08.015>
- [15] Amuda,K.A.,& Adeyemo,A.B.(2019). Customers churn prediction in financial institution using artificial neural network. *arXivpreprintarXiv:1912.11346*.
- [16] Banday, S. A., & Khan, S. (2021). Evaluation Study of Churn Prediction Models for Business Intelligence. *Big Data Analytics*(pp.201-213). Auerbach Publications. Britto,M.M.J.,&Gobinath,D.R.(2021).
- [17] P. Lalwani, M. M. Kumar, J. Singh Chadha and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, pp. 1-24, 2021.
- [18] Improved Churn Prediction Model In Banking Industry And Comparison Of Deep Learning Algorithms. *International Journal of AquaticScience*,12(2),2521-2529.
- [19] Chabumba, D. R., Jadhav, A., &Ajoodha, R.(2021). Predicting telecommunication customer churn using machine learning techniques. In *Interdisciplinary Research in Technology and Management* (pp. 625-636):CRCPress. <https://doi.org/10.1201/9781003202240-98>
- [20] Dalmia, H., Nikil, C. V., & Kumar, S. (2020).Churning of Bank Customers Using Supervised Learning Innovations in Electronics and Communication Engineering (pp. 681-691): Springer, Singapore. [https://doi.org/10.1007/978-981-15-3172-9\\_64](https://doi.org/10.1007/978-981-15-3172-9_64)
- [21] A. K. Ahmad, A. Jafar and K. Aljoumaa, "Customer

- churn prediction in telecom using machine learning in big data platform," *Journal of Big Data* volume 6, vol. 6, no. 28, 2019.
- [22] DeCaigny, A., Coussemment, K., & DeBock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760-772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- [23] Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. *Computation*, 9(3), 34. <https://doi.org/10.3390/computation9030034>.
- [24] P. Lalwani, M. M. Kumar, J. Singh Chadha and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, pp. 1-24, 2021.
- [25] I. Brandusoiu, G. Todorean and H. Beleiu, "Methods for Churn Prediction in the Prepaid Mobile Telecommunications Industry," *International conference on communications*, pp. 97-100, 2016.
- [26] T. Vafeiadis, K. Diamantaras and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1-9, June 2015.
- [27] Dwiyanti, E., & Ardiyanti, A. (2017). Handling imbalanced data in churn prediction using rusboost and feature selection (case study: Pt. telekomunikasi indonesia regional 7). In *Recent Advances on Soft Computing and Data Mining: The Second International Conference on Soft Computing and Data Mining (SCDM-2016)*, Bandung, Indonesia, August 18-20, 2016 Proceedings Second (pp. 376-385). Springer International Publishing. [https://doi.org/10.1007/978-3-319-51281-5\\_38](https://doi.org/10.1007/978-3-319-51281-5_38)
- [28] Ebrah, K., & Elnasir, S. (2019). Churn prediction using machine learning and recommendations plans for telecoms. *Journal of Computer and Communications*, 7(11), 33-53. <https://doi.org/10.4236/jcc.2019.711003>.
- [29] Faris, H. (2018). A hybrid swarm intelligent neural network model for customer churn prediction and identifying the influencing factors. *Information*, 9(11), 288. <https://doi.org/10.3390/info9110288>.
- [30] Faritha Banu, J., Neelakandan, S., Geetha, B., Selvalakshmi, V., Umadevi, A., & Martinson, E. O. (2022). Artificial Intelligence Based Customer Churn Prediction Model for Business Markets. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/1703696>
- [31] Ghobaei-Arani, M. (2021). A workload clustering based resource provisioning mechanism using Biogeography based optimization technique in the cloud-based systems. *Soft Computing*, 25(5), 3813-3830. DOI: <https://doi.org/10.1007/s00500-020-05409-2>
- [32] Ghobaei-Arani, M., & Shahidinejad, A. (2021). An efficient resource provisioning approach for analyzing cloud workloads: a metaheuristic-based clustering approach. *The Journal of Supercomputing*, 77(1), 711-750. DOI: <https://doi.org/10.1007/s11227-020-03296-w>
- [33] Ghobaei-Arani, M., Shamsi, M., & Rahmanian, A. A. (2017). An efficient approach for improving virtual machine placement in cloud computing environment. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(6), 1149-1171. DOI: <https://doi.org/10.1080/0952813X.2017.1310308>
- [34] Ghobaei-Arani, M., & Souri, A. (2019). LP-WSC: a linear programming approach for webservice composition in geographically distributed cloud environments. *The Journal of Supercomputing*, 75(5), 2603-2628. DOI: <https://doi.org/10.1007/s11227-018-2656-3>
- [35] Ghobaei-Arani, M., Souri, A., Baker, T., & Hussien, A. (2019). ControCity: an autonomous approach for controlling elasticity using buffer Management in Cloud Computing Environment. *IEEE Access*, 7, 106912-106924. <https://doi.org/10.1109/ACCESS.2019.2932462>
- [36] Gui, C. (2017). Analysis of imbalanced data set problem: The case of churn prediction for telecommunication. *Artificial Intelligence Research*, 6(2), 93-103. <https://doi.org/10.5430/air.v6n2p93>
- [37] Hartati, E. P., & Bijaksana, M. A. (2018). Handling imbalance data in churn prediction using combined SMOTE and RUS with bagging method. In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012007). IOP Publishing. <https://doi.org/10.1088/1742-6596/971/1/012007>
- [38] Xiong, Y., & Tsai, Y. (2020). Machine learning based approaches to predict customer churn for an insurance company. In *2020 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1-6). IEEE. <https://doi.org/10.1109/SIEDS49339.2020.9106691>
- [39] Ibrahim, K., Aborizka, M., & Maghraby, F. (2018). Prediction of users charging time in cloud environment using machine learning. *International Journal of Intelligent Computing and Information Sciences*, 18(2), 39-57. <https://doi.org/10.21608/ijicis.2018.30121>