Journal for Research in Applied Sciences and Biotechnology

www.jrasb.com

ISSN: 2583-4053

Volume-2 Issue-1 || February 2023 || PP. 96-98

https://doi.org/10.55544/jrasb.2.1.12

Data Mining Techniques in Telecommunication Company

Nazak Ahmadzai[1], Hameedullah Mohammadi[2] and Naqibullah Mangal[3]

[1]Computer Science Department, Education Faculty, Paktia University Gardiz City, Paktia, AFGHANISTAN.

[2]Paktia University Computer Science Department, Education Faculty, Gardiz City, Paktia AFGHANISTAN.

[3]Paktia University Computer Science Department, Education Faculty, Gardiz City, Paktia AFGHANISTAN.

[1]Corresponding Author: nazakmir.ahmadzai@gmail.com

www.jrasb.com || Vol. 2 No. 1 (2023): February Issue

Received: 06-01-2023	**Revised:** 27-01-2023	**Accepted:** 06-02-2023

ABSTRACT

Due to emerging of amalgam amount of data from variety sources, the data mining has become a hot trend in field of Computer Science. Data mining extracts useful pattern and information from huge amount of existing data with the help of machine learning algorithms that can be helpful in solving many sophisticated problems.

Telecommunication companies also generates big amount of data from providing services to their customers, besides that telecommunication companies suffers from many problems like fraud, Customer churn and …etc.

The generated amount of data from these companies can help them to address the solution for their problems such as Customer Churn. Customer churn indicates to the event when a customer stops using the service of a company and starts to use the service of another company.

Churning of a Customer plays a vital role in having a sustainable business development for a telecommunication company since attracting new customers do not profit a company without retaining the old ones.

Data mining can address the problem by predicting the occurrence of customer churn in Telecom Company, which helps the company to be proactive in this event and can have the chance to retain them before the churn occurs.

In this study, I have chosen two open Telecom Churn data sets and have applied Support Vector Machine, Logistic Regression and Decision Tree Machine Learning Algorithms on each data sets independently, which conclude my work to six experiments.

I have used k-fold cross validation as validation technique during my experiments and confusion matrix for calculating the accuracy of each algorithm, the result of experiments will provide the accuracy of each algorithm in churn prediction for each data set.

At the end we will have a general comparison table from all six experiments which will show the algorithms performance summary and will indicate which algorithm will outperform the others.

Keywords- Customer Churn, Fraud and Machine Learning Algorithms.

I. INTRODUCTION

Data mining used to remove non-trivial forms and valuable information out of historical data, which will be suitable in solving many real world problems, one those problems is the existence of Customer shake in Telecommunication Company.

By definition the Customer Churn happens when one customer shifts the using of the service provided by a Company to another Company, in Telecommunication Companies Avoiding the Customers from Churn play an important role on having a maintainable Business Development and has incredible effect on overall total income of the company.

The forecast of Churn helps a company to be upbeat against the existence of its Customer Churn; it means it can get the chance to hold its customers.

Many Data mining procedures such as Decision Tree (DT), Support Vector Machine (SVM), Neural Network (NN), Genetic Algorithms (GA) or Fuzzy logic (FL) are used to forecast churn.

Journal for Research in Applied Sciences and Biotechnology

www.jrasb.com

ISSN: 2583-4053

Volume-2 Issue-1 || February 2023 || PP. 96-98

https://doi.org/10.55544/jrasb.2.1.12

II. OVERVIEW

Due to arise of huge volume of Data, the need to remove useful information out of mixture data has improved much more than before, later there has been formed different methods and ways to remove the meaningful information out of vast quantity of historical data using different restraints where one of them is Data mining.

Data mining aim is to remove information from a dataset and convert it into a more comprehensible structure for additional use, we can say its techniques are an intersection of machine learning, statistics and database system.

There is another word known KDD or Knowledge Discovery in Database, which synonymously used with Data mining but in fact, data mining is really a part of KDD process.

The KDD involve of few iterative phases, which magnets new knowledge and information from a set of rare data. The steps containing in KDD showed in figure 1 are:

- Data cleaning:

  The stage where noise and inessential data are detached or corrected.
- Data integration

  The stage where several data sources are combined.
- Data selection

  The stage where related data are extracted for analysis.
- Data transformation

  The stage where data will be converted into suitable form for mining.
- Data mining

  The stage where distinct methods are applied for useful design extraction.
- Pattern evaluation

  The stage where valuable and motivating patterns are recognized based on given dimension.
- Knowledge representation

  The final stage in which the beneficial extracted knowledge are presented visually to the user.

Some of the above phases can be joint together such as data cleaning and data integration as a pre-processing segment to produce a data warehouse (Osmar R. Zaïane, 1999).

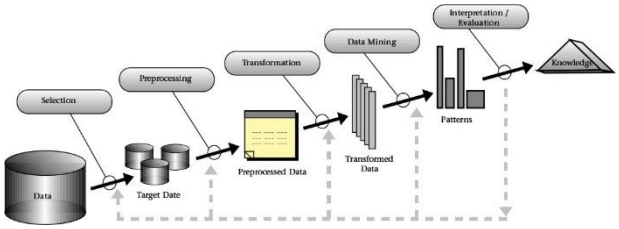

Figure 1: An overview of the steps of KDD process

2.1. Data Mining Techniques

There are several data mining techniques used for knowledge detection, here is a short explanation of them.

2.1.1 Classification:

It is the most normally used data mining technique, which includes a set of pre-classified modules to design a model that can categorize all the records completely. It includes the process of learning and classification, in learning course the data will be analyzed using specific algorithm and in classification process new data will be tested whether it gratifies the classification rule which was created throughout session or not. If it satisfied the rule then the classification can be applied on new tuples of data (Ramageri, 2010).

The arrangement can be different based on type of algorithms used, which can be:

- Classification by decision tree induction
- Bayesian classification
- Classification using Neural Networks
- Classification using Support vector machines
- Classification based on association

2.1.2 Clustering:

This technique is used for identifying similar classes and objects, by applying this method we can analyze the total distribution and diffusion among class objects and characteristics (Ramageri, 2010).

Here are different kinds of clustering techniques:

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

2.1.3 Prediction

As its name specifies, it is used for forecast. Mostly the reversion analysis is used in it.

Compare to other data mining techniques this one uses much more composite model as frequently the prediction of real world problems such as stock prices, product failure rates and churn occurrence are reliant on to many variables (Ramageri, 2010).

Here are some Prediction approaches of data mining.

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

2.1.4 Association rule

This method is used for identification of numerous items set occurrence among huge data set (Ramageri, 2010).

Some methods of Connotation rules are:

ISSN: 2583-4053

Volume-2 Issue-1 || February 2023 || PP. 96-98

https://doi.org/10.55544/jrasb.2.1.12

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

Data mining tools makes a representation to the user in the model form, which describes the pattern and association between the current data. Based on process orientation viewpoint data mining activities falls into three groups (SAS Institute & Federal Data Corporation, 1999):

1. Discovery

The processes of discovery hidden patterns without having any static idea or statement regarding the patterns will be about from the database.

2. Predictive modeling

The procedure of discovery hidden patterns from the database for prediction of the upcoming circumstances.

3. Forensic analysis

The process of applying the extracted patterns to find abnormal or rare data elements existing in data.

2.2 Customer Churn

In trade environment when a customer stop using the services or doing business with a service provider or company, we can say the customer has churned.

Advanced technologies revolutionized the business, big companies are presiding the world where their customers are not restricted to their city, country or even continent and also the competition of these companies for leading the market has been tougher than before.

These mega companies are putting their maximum efforts not even to gain new customers but to retain their old once too because gaining new customers won't have any positive effect in terms of income revenue to the company if on other side the company loses the old once. Therefor retaining old customer is much important for having a sustainable business development

2.3 Churn in Telecommunication Industry

In this competitive world, it is hard to develop and maintain the business specifically in Telecommunication, companies must put their maximum effort to attract and acquire as much as many customers in order to have a sustainable business development.

There won't be always the customer acquisition a big concern for the companies but also maintaining the existing one too. If a company acquires new customers monthly and in opposite loses its old customers than in fact it hasn't achieved anything.

In telecommunication business, customers must be satisfied from the services offered by the provider unless they will switch from service provider to another one that suits their needs.

Customer churn is a big issue for telecommunication companies due to fierce competition with the national and international companies, which always upgrades their services with new innovative technologies. The losing of customer of a telecom company is a big loss and serious problem since the cost of acquiring new customers is much higher than retaining the existing ones.

III. CONCLUSION

Data mining used to extract non-trivial patterns and useful information out of historical data, which will be useful in solving many real world problems, one those problems is the occurrence of Customer churn in Telecommunication Company.

By definition the Customer Churn occurs when one customer switch the using of the service provided by a Company to another Company, in Telecommunication Companies Preventing the Customers from Churn play a Significant role on having a sustainable Business Development and has tremendous effect on overall total revenue of the company.

The prediction of Churn helps a company to be proactive against the occurrence of its Customer Churn; it means it can get the chance to retain its customers.

REFERENCES

[1] Han, J., Altman, R. B., Kumar, V., Mannila, H., Pregibon, D. Emerging scientific applications in data mining. Communications of the ACM 2002; 36(7): 38-49.

[2] Cortes, C.Pregibom, D. Signatue-based methods for data streams. Data Minning and Knowledge Discvory 2001; 7(4):145-148.

[3] Ezawa, K., Norton, S. Knowledge discovery in telecommunication services data using Bayesian network models. Proceedings of the First International Conference on Knowledge Discovery and Data Mining; 1995 August 20-21. Montreal Canada. AAAI Press: Menlo Park, CA, 1995.

[4] Fawcett, T, Provost, F. Activity monitoring: Noticing interesting changes in behavior. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 53-62. San Diego. ACM Press: New York, NY, 1999.