Analysis of Customer Churn prediction in Logistic Industry using Machine Learning

Pradeep B[‡], Sushmitha Vishwanath Rao* and Swati M Puranik[†]

Department of Computer Science

NMAMIT, Nitte ,

Karkala taluk, Udupi district,

Karnataka, India – 574 110

Akshay Hegde[§]

Department of Computer Science

BVB College of Engineering and Technology

Vidyanagar, Hubli

Karnataka, India - 580034

Email: [‡]pradeep.b2505@gmail.com, *itsmesushmitharao@gmail.com, [†]swatimp15@gmail.com, [§]akshaykumarhs@gmail.com

Abstract: **Customer churn prediction in logistics industry is one of the most prominent research topics in recent years. It consists of detecting customers who are likely to cancel a subscription to a service. Recently, logistics market has changed from a rapidly growing market into a state of saturation and fierce competition. The focus of the logistic companies has therefore shifted from building a large customer base into keeping customers in house. For that reason, it is valuable to know which customers are likely to switch to a competitor in the near future. The data extracted from the industry can help analyse the reasons of customer churn and use that information to retain the customers. We have proposed to build a model for churn prediction for a company using data mining and machine learning techniques namely logistic regression and decision trees. A comparison is made based on efficiency of these algorithms on the available dataset.**

Keywords: Customer churn, logistic regression, linear regression, predictive analysis, data mining, machine learning.

I. INTRODUCTION

In simple words, customer attrition occurs whenever a customer stops doing the business with a service provider or a company. Churners have always been a big issue for any service providing company. Churning increases cost of the company as well as decreases their rate of profit. Considering machine learning perspective, the churn prediction is supervised and that can be defined as: Given a predefined forecast horizon, the goal is to predict the future churners over that horizon, given the data associated with each subscriber in the network. Churn prediction aims to identify subscribers who are about to transfer their business to a competitor. Since the cost associated with customer acquisition is much greater than the cost of customer retention, churn prediction has emerged as a crucial Business Intelligence (BI) application for modern service providing companies. The flow of the raw materials or other goods to end customers is called as logistics. The essential success factor for any logistic industry lies in delivering items to the correct place and at appropriate time with a reasonable cost. Customer dissatisfaction at any of the stages of this process leads to a huge loss in the business and that is where the exact concern lies. A long-term relationship with the customers is a very crucial factor in the logistics

industry because of the innumerable aspects of service encounters which can easily be imitated by the competitors. One of the gauging successes in the logistics industry is customer churn. Therefore, there is a huge need for a defensive marketing strategy which prevents the customers from switching the service providers. Customer churn causes revenue loss and other negative effects on corporate operations. Therefore, our idea mainly focuses on customer churn prediction model for identifying the key factors which are crucial and which cause the churn. The set of techniques that we use to do the same, include Logistic regression(LGR), decision tree analysis and artificial neural network.

II. LITERATURE SURVEY

K B Oseman [8] in technical paper, explains about the predictive modelling for churners which are based on data mining methods. The paper discusses about the usage of decision tree analysis model in detail.

The paper mainly discussed about the customer churning from a business perspective. However, at the end of the paper they also discussed case studies along with process flows and
modelling techniques.

Teemu Mutanen[9] in the technical paper, described a case study on customer attrition. The paper described in detail the methods used for the prediction, data used and the result that was achieved. The author described two methods for churn analysis. The first one is logistic regression.
Logistic regression is used to predict a discrete outcome based on continuous and/or categorical variables. In this method only one dependent variable can exist. This method applies maximum likelihood estimation after transforming the dependent variable into a logistic variable.
The second method analyses the estimation results of the logistic regression. It is known as the lift curve. This curve is related to ROC curve of signal detection theory and precision-recall curve. The lift is a measure of predictive model calculated as the ratio between the results obtained with and without the predictive model.

In [10] Shyam V. Nath describes a case study in which an Oracle based database of fifty thousand customers of wireless telecommunication industry was analysed to predict churners. The study used JDevelpoer tools and the analysis was done using Naïve Bayes algorithm with supervised learning.

Marco Richeldi and Alessandro Perucci [11] wrote a paper on case study of churn analysis. This paper discusses the use of Mining Mart, a churn analysis tool. It mainly discusses the pre-processing of data to analyse with Mining Mart.

III. PROBLEM AND ITS IMPACT

Churn rate is the number of customers or subscribers who cut ties with your service or company during a given period. These customers have "churned."

(A) Understanding the problem:
- The flow of the raw materials or other goods to end customers is called as logistics. The essential success factor for any logistic industry lies in delivering items to the correct place and at appropriate time with a reasonable cost.
- Customer dissatisfaction at any of the stages of this process leads to a huge loss in the business and that is where the exact concern lies.
- A long-term relationship with the customers is a very crucial factor in the logistics industry because of the innumerable aspects of service encounters which can easily be imitated by the competitors.
- One of the gauging successes in the logistics industry is customer churn. Therefore, there is a huge need for a defensive marketing strategy which prevents the customers from switching the service providers.
- Customer churn causes revenue loss and other negative effects on corporate operations.

(B) Churn impact:
- The aftermath of churn is Long-Reaching
  -The company loses the chance to upsell them with other products or services in its portfolio as the brand is now likely tarnished in the minds of the spurned customer.
- A small improvement in churn will have a big impact on value over time
  -As gaining new customers is difficult and if your company continues to lose existing customers it will effect on company's revenue and turnover.
- Churn helps competitors
  -Negative customer reviews are a gift to the competitor companies as these act as powerful ammunition for the competitors to position against your product.
- Higher customer churn may indicate bigger problems.
  -High churn indicates that something is failing in your customer relationships. It reflects poorly on your company and your product.

(C)Objective:
To develop an AI prototype to analyse and predict customer churn and the probability to win an opportunity at an early stage of the customer engagement using a variety of structured or unstructured data sources. The solution should show quantifiable measurable metrics displaying thorough analytics on customers to help shape the future strategies and decision making.

IV. METHODOLOGY

(A)Methods of churn analysis:
There are many ways to do churn analysis. However, these methods can be primarily categorized in two sections. They are:
- Supervised methods
- Unsupervised methods

1) Supervised methods: In supervised methods [7], the method basically learns to classify the data based on what type of data it learns from given training data. The training data, for example is a pair consisting of an input object and a desired output value. In our case the input object would be the

rows of our data and the output value would be the probability depicting whether the customer has churned or not.
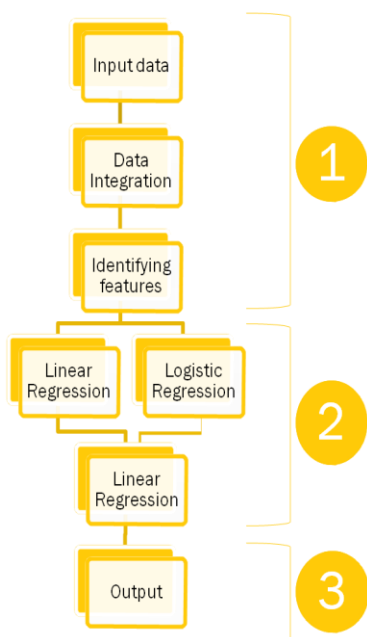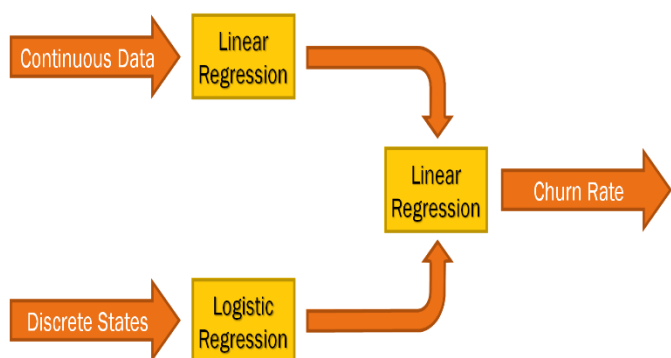However, our dataset does not have any attribute that will be saying who will churn. Therefore, we cannot use supervised approach with our dataset.

2) Unsupervised methods: On the other hand, unsupervised methods refer to the problem by trying to find hidden structure in unlabelled data. We can use unsupervised methods to cluster our data set. This way the churners could become part of a separate cluster. Again, we can use methods which self-learn the data.

(B)Our approach:
We begin by conducting a series of data pre-processing tasks which include merging the customer, shipping and delivery tables, removing records with missing values, deleting duplicate records and also aggregating the records for each business customer. The groups of active and lost customers are defined. The lost customers are further classified into groups to determine the reason for not engaging in transactions. We mainly designed three models. The first one was the linear regression model which gave us the predicted probability of 76% of all the features that were considered. This Linear Regression model considered all the KPI values and the milestones. Without stopping at that point, we moved on to design another model considering the status and that too yielded us the predicted probability of around 93%. We combined the predicted probabilities and constructed a last model using Linear Regression and ended up with the results of final predicted probability of around 94%. The case company (that is, the company under

consideration) gives a total of N variables divided into general categories to build the model. The KPI weights were multiplied with the values and because valuable customers contribute more to the profitability of the firm than do less valuable customers we apply random sampling technique for over M times to generate multiple datasets and for each of the generated datasets, a tenfold cross-validation is applied to evaluate sample quality.

V. CONCLUSION

To conclude, purpose of customer value analysis is to identify valuable customers that potentially contribute to the profitability of the company. A customer churn prediction model can be used as an early warning tool for businesses and extracting the critical factors related to the customer churn that provides additional useful knowledge which supports decision making.

VI. FUTURE WORK

The method can be used for campaign management, by modelling the best groups to be approached with a specific marketing campaign and by pinpointing individuals who are the most influential over their peers. Our method can also be used in other domains where the links between people can be measured or inferred, such as social networking sites on the Internet, in order to predict customer behaviour.

VII. REFERENCES

[1] Customer attrition. Retrieved from http://en.wikipedia.org/wiki/Customer_attrition, on, February 25, 2011.

[2] Predictive analytics, Retrieved from http://en.wikipedia.org/wiki/Predictive_analytics on April 14, 2011.

[3] Predictive modeling, Retrieved from http://en.wikipedia.org/wiki/Predictive_modelling, on March 17, 2011.

[4] R. Mattison, The Telco Churn Management Handbook, 2001.

[5] Churn Analysis. (n.d.). Retrieved from http://www.ambarasoft.com/researchservices/churnanalysis.html

[6] I. H. Witten, E. Frank. Data Mining Practical Learning Tools and Techniques, Morgan Kaufmann Publishers, 2005.

[7] Supervised learning Retrieved from http://en.wikipedia.org/wiki/Supervised_learning, retrieved on April 16, 2011.

[8] K. B. Oseman, S.B.M. Shukor, N. A. Haris, F. Bakar, "Data Mining in Churn Analysis Model for Telecommunication Industry", Journal of Statistical Modeling and Analytics, Vol. 1 No. 19-27, 2010.

[9] T. Mutanen, Customer churn analysis – a case study, Technical Report, Retrieved from, http://www.vtt.fi/inf/julkaisut/muut/2006/customer_churn_case_study.pdf, April 12, 2014.

[10] S.V. Nath, Customer Churn Analysis in the Wireless Industry: A Data Mining Approach, Technical Report, retrieved from http://download.oracle.com/owsf_2003/40332.pdf, April 14, 2014.

[11] M. Richeldi and A. Perrucci, "Churn Analysis Case Study", Technical Report, Telecom Italia Lab, Italy, retrieved from http://sfb876.tudortmund.de/PublicPublicationFiles/richeldi_perrucci_2002b.pdf, April 12, 2014.