Customer Churn Prediction

Mandar Vaidya
x17153409
MSc in Data Analytics

17th June 2018

Area Customer Churn Prediction topic is broad. This paper will discuss the research in Statistics in Search and Analysis, Machine Learning in Cognitive Systems and Prediction, and Visualization in Visualization and Interaction.

> *Analysis of Customer Churn Prediction in Telecom Industry Using Machine Learning and Statistical Testing Techniques?*

1 Value

Now-a-days, to retain existing customers has become more challenging than acquiring new ones. Churn is defined as the act of a customer leaving the network permanently and Churn Prediction represents a problem of identifying users who are likely to churn Milosevic et al. (2017). Churn prediction has become a common problem in many industries like - Telecommunication, Banking, Mobile Games, etc.

With a churn rate of 30 percentage, the telecommunication sector is most likely to notice existing customers switching to competitors' network. Thus, telecommunication firms have been paying more attention to the problem of identifying the customer churn behavior, as the cost of retaining the existing customer is far greater than acquiring the new ones. In various countries, especially the developed ones, the market is saturated to the extent that each new customer must be won over from competitors. At the same time, public policies and standardization of mobile communication now allow customers to easily switch over from one carrier to another, resulting in a fluid market Qureshi et al. (2013). Studies have demonstrated that more than 75 percentage of customers will consult a friend before deciding on the purchase of a certain product or adoption of a service. It is mandatory for operators to identify unhappy subscribers and prevent them from switching to another competitor or canceling their subscription. For this, the telecom industry needs to develop strategies and implement the best predictive model for their business management to retain pre-existing customers and obtain new ones. Thus, based on the history of the customers calling pattern, bill payments, data usage etc., there is a possibility to identify their mindset whether they will leave the existing firm or not.

Customer Churn is closely related to the customer retention and loyalty. This question is worth investigating because it will reduce churn and its connected huge acquisition costs. The telecommunication firms will know their customers better. By gauging why, the customer leaves their product or service, will allow them to adopt and develop new retention strategies. Further, using, visualizations, it will be easy to predict and target the customers who are mostly likely to leave, thus will save the time and money by targeting the wrong customer. Additionally, organizations can use Customer Relationship Management (CRM) tool to boost revenues by understanding the customer satisfaction information - By creating personalized programs to up-sell and cross-sell for better understanding. Hence, contributing to the field of Data Analytics in a huge manner.

2 Justification

Customer churn prediction problem is important and challenging at the same time. Telecommunication companies are largely investing in building accurate churn prediction models to help them in designing effective customer retention strategies. As, Churn Prediction is likely to increase further, it can be decreased by Machine Learning and Statistics. Machine Learning is a field in Data Science that focuses

on design, which will be used to make decisions and predictions based on the data. It is designed to learn and improve over time when exposed to new data.

This paper compared the predications of Traditional, Network and Combined attribute models and found that Network attributes greatly improve prediction accuracy. (Zhang et al.; 2010)

This paper proposed Multilayer Propagation (MLP) network and compared them with Multiple Regression Analysis, Logistic Regression Analysis with the help of Lerenberg Marqyardt (LM) learning algorithm and suggested that Neural Network techniques can be a good alternative to traditional predictive methods. (Ismail et al.; 2015)

Machine learning is used with the help of Linear, Logistic, Artificial Neural Networks, K-means Clustering, Decision Trees including Chi-square Automatic Interaction Detector (CHAID) exhaustive, Chi-square Automatic Interaction Detector, Classification And Regression Tree (CART) and Quick, Unbiased, Efficient Statistical Tree and found that exhaustive CHAID algorithm have the best result with a small dataset. (Qureshi et al.; 2013)

Monte-Carlo simulations was performed using Two-Layer Back Propagation network, Tree Classifier, Support Vector Machine, Logistic Regression, Naive Bayes, and found out Two-Layer Back Propagation and Tree Classifier have good accuracy and f-measure respectively. Secondly, Naïve Bayes and Logistic Regression cannot be boosted by AdaBoost.M1 algorithm due to lack of free parameters (Vafeiadis et al.; 2015)

This paper proposed a selection method based on Orientation Ordering Pruning Method (OOPM) algorithm and introduced a new feature based on Random Forest and Transduction and found out that OOPM with the help of Random Forrest and Transduction are effective for extraction and indicator system and customer churn prediction. (Yihui and Chiyu; 2016)

This paper proposed linear like Logistic Regression and non-linear techniques like Random Forrest, Deep Neural Networks etc. and result showed that non-linear techniques performed better than linear one (R et al.; 2017)

Machine learning and statistical techniques, with the help of visualization, could be utilized with the proposed machine learning methodology to reduce the churn rate. New technique in Deep Learning has brought breakthroughs in different areas of research like natural language processing, and data recognition. This will be helpful for data set analysis and will aid significantly in the main research proposed by the paper.

By using, Machine Learning tools like, Support Vector Machines (SVM), Decision Trees, k-nearest Neighbor (KNN), Naive Bayes, Linear Modeling, Artificial Neural Networks (ANN), organizations can predict the churn behavior of customers thus by significantly improving the productivity of the firms and reducing the churn customers.

Also by using, Statistical techniques with the help of R Programming, Logistic Regression, Correlation, Random Forest, Confusion Matrix, Receiver Operating Characteristic (ROC) can be calculated and accuracy can be obtained with the true positive and false positive rates. Lastly, with the help of Visualization and integrating it with Machine Learning, Predictive Analysis by using data recognition and data patterns and Business Intelligence queries by Tableau and Power BI can be performed.

As, churning is not restricted to the telecommunications market, it is of great concern for banking, finance, credit card firms, internet service providers, mobile game firms where competition is stiff and switching to another firm is easy. The aforementioned industries can expand and develop new technique to other industries suggests interesting directions for future research. This will impact greatly on the business of telecommunication companies as preventing churn will save the firm from customer loss which eventually means less resources spent on trying to acquire new customers. Thus, directly impacting into profits. Keeping track of customer behavior will help with customer segmentation and customer behavior that will help in predicting their future purchases, and thus by keeping those products in stock and deliver quickly to keep customers happy and satisfied. Churn prediction model can be leveraged by different industries as a baseline template and this model can be tailored as per the specific industry requirements. And, this is ethical as it does abide by the new GDPR/research ethics as the data used will be public and private data of customers provided for the research will be masked by the source.

References

Ismail, M., Awang, M. K., Nordin A Rahman, M. and Makhtar, M. (2015). A multi-layer perceptron approach for customer churn prediction, **10**: 213–222.

Milosevic, M., Živić, N. and Andjelković, I. (2017). Early churn prediction with personalized targeting in mobile social games, **83**.

Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A. and Rehman, A. (2013). Telecommunication subscribers' churn prediction model using machine learning, *Eighth International Conference on Digital Information Management (ICDIM 2013)*, pp. 131–136.

R, P., Deepak, K. and Kumar Meher, A. (2017). High accuracy predictive modelling for customer churn prediction in telecom industry, pp. 391–402.

Vafeiadis, T., Diamantaras, K., Sarigiannidis, G. and Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction, **55**.

Yihui, Q. and Chiyu, Z. (2016). Research of indicator system in customer churn prediction for telecom industry, *2016 11th International Conference on Computer Science Education (ICCSE)*, pp. 123–130.

Zhang, X., Liu, Z., Yang, X., Shi, W. and Wang, Q. (2010). Predicting customer churn by integrating the effect of the customer contact network, *Proceedings of 2010 IEEE International Conference on Service Operations and Logistics, and Informatics*, pp. 392–397.