The International Journal of Engineering and Science (IJES)
|| Volume || 8 || Issue || 6 Series I || Pages || PP 48-52|| 2019 ||
ISSN (e): 2319 – 1813 ISSN (p): 23-19 – 1805

Designing a framework for Real time Churn Prediction in Mobile Telecommunication Industry

Turiho Jean Claude[1, 2], Kipruto W. Cheriuyot[1], Ann Muthoni Kibe[1]
[1]*Jomo Kenyatta University of Agriculture and Technology, Kigali Campus, Rwanda*
[2]*University of Lay Adventists of Kigali, Kigali, Rwanda*
Corresponding Author: Turiho Jean Claude

ABSTRACT
In Telecommunication Industry, the customer data are kept in structured data base. Customer behavior is extracted from this customer database. Customer behavior reveals two types of customers: customer who is about to move out to the new competitors and customer who is loyal. The data mining algorithms are based on past data for predicting churn cases. This improving Customer Churn Prediction Using Hybrid Techniques does not reject the historical data but adds the real data for improvement. By this new technique in Mobile Industry we monitor, in real time, customers' behaviors and perform satisfactorily, and it is effective to detect for predicting the Mobile industry customer churn. The researcher presents the new customer churn prediction model based on a combination analysis of historical and live data in order to capture, monitor churning and non-churning status. The proposed model thereby generates optimal customers who will leave the service provider which is beneficial for any enterprise in the current scenario for effective decision making and perform appropriate steps to retain those customers. The new model is divided into the following phases: Call Data Record acquisition, converting Call Data Record into text data by API to be loaded into data warehouse by ETL, applying data mining classification technique, and live capturing and monitoring of customer churn.
KEYWORDS:- *ETL, Call Data Record, K-nearest neighbor, Data mining, Decision tree*

DATE OF SUBMISSION: 21-06-2019 DATE OF ACCEPTANCE: 05-07-2019

I. INTRODUCTION

Customer churn is perhaps the biggest challenge in telecom industry. The prepaid customers has a significantly higher churn rate (on average 9.4%) than the postpaid customers (on average 5.2%), because the prepaid customers are not bound to contract and can quit easily without recharging (Huang, et al., 2015). Since the cost of attracting new customers may be substantially higher than that of keeping the existing ones (e.g., around 3 times higher), it is urgent to build in real time customer churn prediction systems to predict the most likely churners for proper retention campaigns.

Building a churn prediction model will ease the customer retention process, and in this way the telecom companies will succeed in this constantly increasing competitive market. The churn prediction modeling process is strongly depending on the data mining process and techniques due to an increased performance generated by machine learning algorithms compared to the statistical techniques for non-parametric data [15]. Further to the profiling of the customers a real time environment is useful to constantly monitor the customer's interactions with the telecommunication company and dynamically shift a customer to a more appropriate category if needed.

II. RELATED WORK

In Telecom industry, data-driven churn predictive techniques generally includes designing useful features (predictor variables) and designing good classifiers (predictors) or classifier ensembles with these features [1]. Binary classifier is used from instances of training data, for which the class labels: churners or non-churners are known. By the said classifier, researchers want to predict the class labels of instances of test data in the near future, where training and test data have no overlap in time intervals. Each example commonly will be depicted toward a vector for features which will be utilized for prediction. Following this research line, many classifiers have been adopted for churn prediction, including logistic regression [2], decision trees [3], boosting algorithms [2], boosted trees (gradient boosted decision trees) or random forest [4], neural networks [5], evolutionary computation (e.g., genetic algorithm and ant colony optimization) [6], ensemble of support vector machines [7], and ensemble of hybrid methods [8]. According to [13] it is very difficult to determine what is a good or bad user experience at every single moment of the day and without asking the users for direct feedback.

2.1 Churn Prediction Methods

All classifiers listed above have the same aim to classify these customers into two classes: churners and non-churners. Within the context of churn management, predictive modeling uses past transactions and characteristics of a subscriber to predict future subscriber behavior. Churning customers are partitioned under two fundamental groups, voluntary and non-voluntary churners. Non-voluntary churners are the easiest to detect, as these are the customers who have had their service withdrawn by the company. Voluntary churn is most difficult to identify, because this type of churn happens when a subscriber makes a personal and conscious decision to terminate his/her service with the provider [11]. Right now, in literature there two types of decision tree mostly used in churn prediction. Classification trees are when the predicted outcome is the class to which the data belongs. Regression tree is when the predicted outcome can be considered as real number [12].

2.2 Customer Behavior Features

According to [9] customers are ranked based on behavior features qualified as predictor of churn as following: percent of inactive days, total incoming calls, total outgoing calls, outbound network degree, incoming text messages received from competitor's network, average number of calls to information portal per active day, unique weekend contacts per active day, average daily text messages received from competitor's network, total outgoing degree. All previous customer behavior features are grouped by [10] in three groups such as Demographic attributes, Contract attributes, and Customer behavior attributes.

III. CUSTOMER CHURN PREDICTION MODEL

The researcher has designed a new customer churn prediction model based on a combination analysis of historical and live data in order to monitor customer churning and non-churning status. The new model is divided into the following phases: Call Data Record acquisition, converting Call Data Record into text data by API to be loaded into data warehouse by ETL, applying data mining decision tree and k-nearest neighbor techniques, and live monitoring of customer churn.

The proposed framework for customer churn prediction for real time using decision tree and K-nearest neighbor is organized in four phases: (1) Capturing raw data, (2) building data warehouse, (3) analyzing trained data by applying decision tree and k-nearest algorithms, and (4) view of churners and non-churners in real time on dash board as presented below:

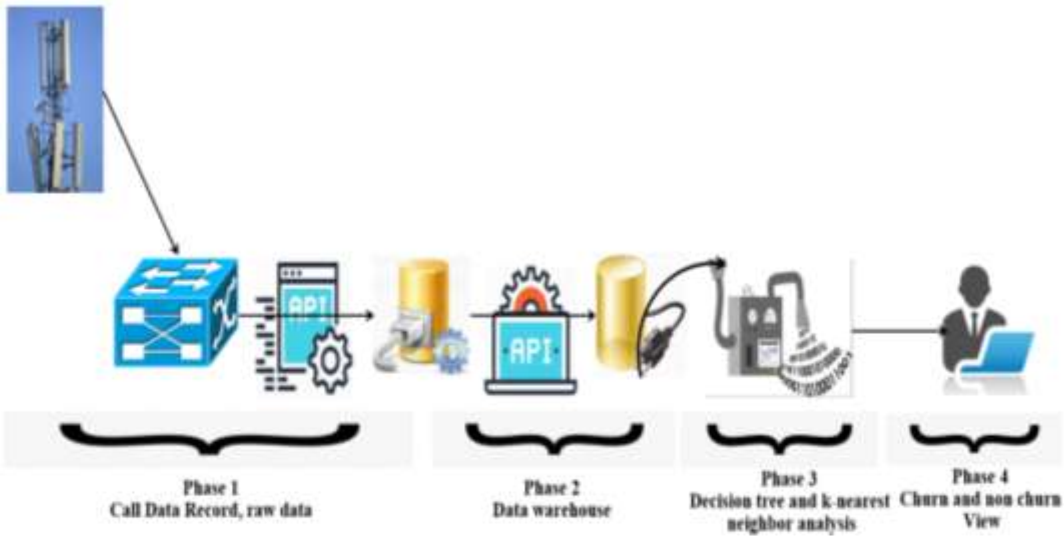

Figure 1. Proposed framework for Customer churn prediction in real time

The Data Mining Process is an iterative process which does not stop when a particular solution is deployed. A GSM system is basically designed as a combination of three major subsystems: the Network Switching System (NSS), the Base Station System (BSS) and the Operation Support System (OSS). In general, data is collected from the Base Station Controller (BSC), which is a part of the radio subsystem. When a user makes a call, a mobile phone connects to the closest Base Transceiver Station (BTS). In the telecommunications sector, huge amount of data are produced and stored, including: Call data describing the calls that traverse the telecommunication networks, Network data concerning the state of both hardware and software components and customer related data. Within such huge amount of critical business data, valuable knowledge can be hidden said [14]. The new Customer churn prediction model concerns acquiring Call Data Record from switch, converting them into text data to be pulled in data warehouse. After data preprocessing data are loaded by our API into new data source data mining software where data mining techniques: decision tree and k-nearest classification algorithms were applied. All Customer status is monitored on dashboard in real time.

Acquiring data is the first and most important phase in our model. Data are directly captured from telecommunication company switch. Call Data Record are raw data for this model. Second phase is Data preprocessing phase. In this phase data are transformed into required format and data cleansing is performed, which is the process of detecting and correcting, or removing corrupt, inaccurate or irrelevant records. Data preprocessing tasks are likely to be performed multiple times, and not in any prescribed order.

Then, there is Model building and evaluation phase. In this third phase, two data mining classification algorithms are used: Decision tree predictor and K-nearest neighbor are selected, applied and parameters are calibrated to optimal values. In this model, decision tree and k-nearest neighbor were used for retrieving the churners and non-churners in real time because are most popular type data mining predictive models.

Finally, the last phase is reporting customer churn and non-churn status. This report is done using Microsoft Visual Studio C#. The discovered knowledge is visually presented to different partners. Visualization techniques are more effective in understanding the output for end users. In this model the knowledge is displayed on dashboard.

3.1 Features selection

In fact this analysis is especially based on some specific fields of telecommunication database. Among other fields we have chosen the following according to their importance in demonstrating customer activities. The following are selected fields: Competitors Calls, Monthly Revenue, Monthly Minutes, Total Recurring Charge, Director Assisted Calls, Average Minutes, Roaming Calls, Perc Change Minutes, Perc Change Revenues, Dropped Calls, Blocked Calls, Unanswered Calls, Customer Care Calls, Three way Calls, Received Calls, Outbound Calls, Inbound Calls, Peak Calls In Out, Off Peak Calls In Out, Dropped Blocked Calls, Call Forwarding Calls, Call Waiting Calls, Months In Service, Home ownership, Occupation, Marital Status. In this database, each row represents a customer, each column contains customer's attributes described on the column Metadata. The raw data contains 51048 rows (customers) and 26 features. This selection was also motivated by the question asked by [13].

3.2 Customer Churn prediction framework

Among selected variables, we have variables describing customer profiles and their past behavior, and usually include demographical characteristics and transactions. Two phases are used in this framework: learner phase and predictive phase.

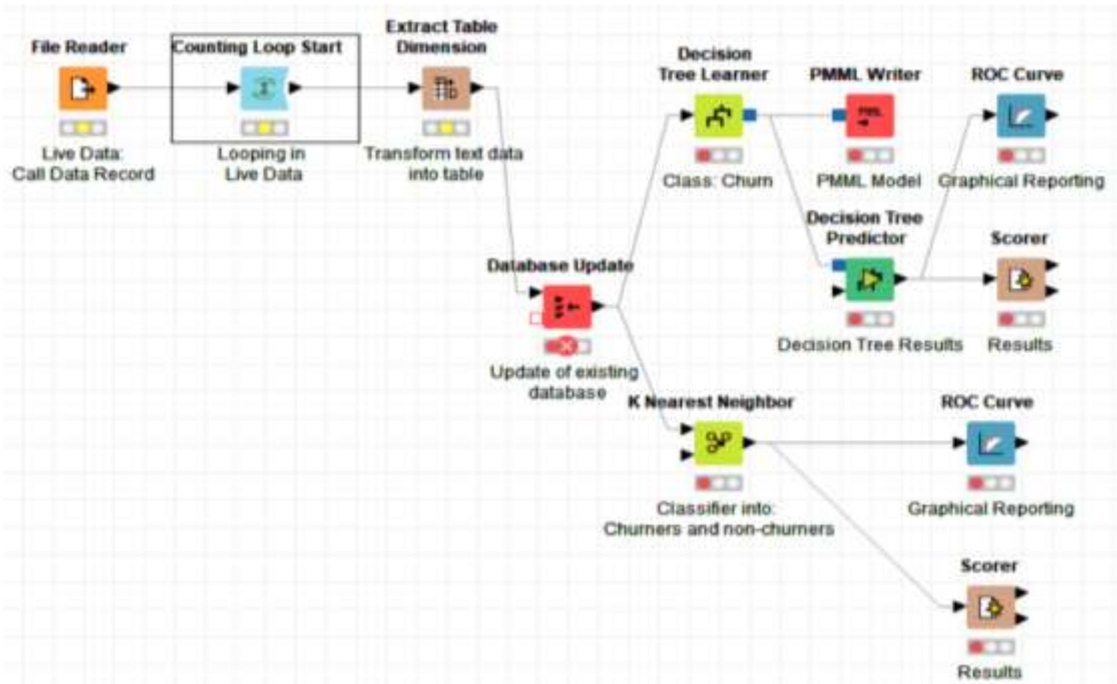

Figure 2. Learner phase of Customer Churn Prediction model

In this first phase called learner, Call Data Record considered as raw data are transformed into file text and stored in databases. In proposed framework customer code and churn status are used to differentiate the customer current status. Dataset is simultaneously treated using two different data mining classification algorithms: decision tree and k-nearest neighbor. In decision tree, dataset is split into two partitions with partition node to train decision tree learner and predictor nodes. ROC curve and Scorer nodes are used test prediction accuracy with user defined ranges. If prediction accuracy of model is satisfied, trained model will be saved in PMML (Predictive Model Markup Language) model node to use in second phase.

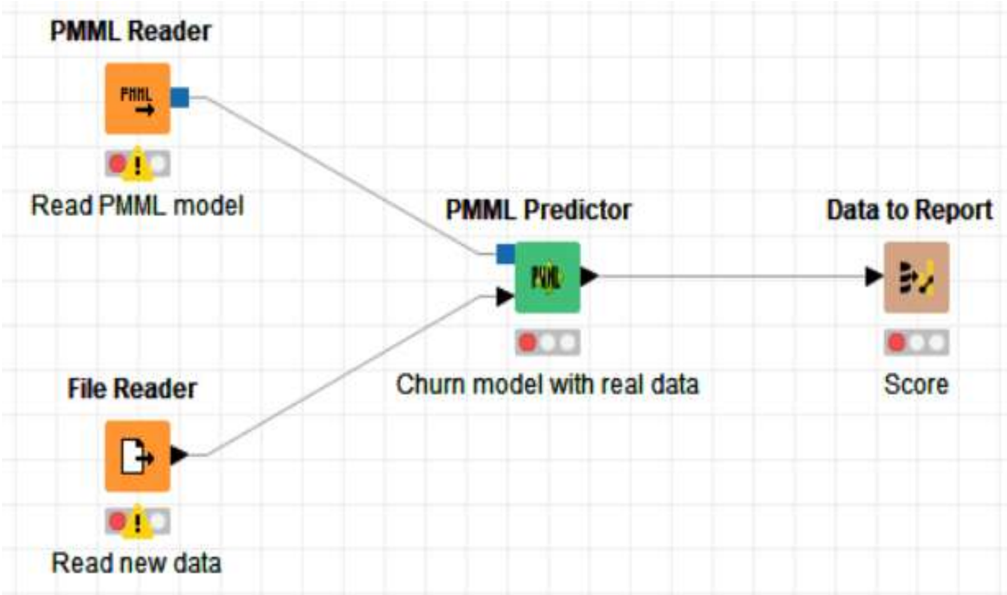

Figure 3. Prediction phase of Customer Churn Prediction

This second phase of customer churn prediction in real time, real data are used to predict customer decision according to trained model. The PMML model is applied on real data integrated into existing data stored in data warehouse. In order to develop a customer specific offering, prediction phase model will be used to improve decision making process. Further to the profiling of the customers a real time environment will be developed using any programming language to constantly monitor the customer's interactions with the telecommunication company.

IV. PERFORMANCE EVALUATION

With this new model, all customer behavior is in real time be known especially based on competitors' calls among others transactions.

V. CONCLUSION

After designing the customer churn prediction framework, we will develop different APIs and user dashboard for converting, loading data and presenting customer behavior to different users in the company.

REFERENCE

[1]. H.-F. L. H.-P. H. J.-K. L. T. G. M. J.-W. C. P.-H. C. C.-H. H. C.-F. C. Y.-H. W. e. a. H.-F. Yu, "Feature engineering and classifier ensemble for kdd cup 2010," LMLR W & CP, pp. 1-16, 2010.
[2]. H. L. J. L. a. G. Z. N. Lu, "A Customer Churn prediction model in telecom industry using boosting," IEEE Trans, on Instrustry Informatics, pp. 1659-1665, 2014.
[3]. C.-P. W. a. I. Chiu, "Turning Telecommunications Call details to churn prediction: a data mining, approach," Expert Systems with Applications, vol. 23, no. 2, pp. 103-112, 2002.
[4]. H. C. J. B. G. M. N. N. C. Phua, "Predicting near feature churners and win-backs in telecommunication industry," arXiv preprint arXiv, p. 1210.6891, 2012.
[5]. D. Y. a. H. W. S. Hung, "Applying Data Mining to telecom churn management," Expert Systems with Applications, vol. 31, no. 3, pp. 451-524, 2006.
[6]. D. M. C. M. a. B. B. W. Verbeke, "Building oomprehensive customer churn prediction models with advanced rule induction echniques," Expert Systems with Applications, pp. 2354-2364, 2011.
[7]. K.-H. J. Y. S. K. a. J. L. N. Kim, "Uniformly subsampled ensemble (use) for churn management: Theory and implementation," Expert Systems with Applications, vol. 39, no. 15, pp. 11839-11845, 2012.
[8]. V. L. M. B. G. D. a. D. V. I. Guyon, "Analysis of KDD Cup 2009: Fast scoring on a large orange customer database," pp. 1-22, 2009.

[9]. J. M. A. S. a. J. B. Muhammad Raza Khan, "Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty," arxiv, 2015.
[10]. B. L. R. S. K. V. T. a. S. A. K. Aleksandar J. Petkovski, "Analysis of Churn Prediction: A Case Study on Telecommunication Services in Macedonia," IEEE, pp. 1-4, 2016.
[11]. A. T. R. R. a. D. R. John Hadden, "Churn prediction: Does technology matter," International Journal of Industrial and Manufacturing Engineering, pp. 524-530, 2008.
[12]. M. K. Kemal Yayla, "A framework for customer churn prediction in air cargo industry," INTERNATIONAL LOGISTICS AND SUPPLY CHAIN CONGRESS, 2016.
[13]. F. P. K. L. Z. N. Y. G. E. B. a. F. C. E. C. a. P. H. J. S. Ernesto Diaz-Aviles, "Towards Real-time Customer Experience Prediction for Telecommunication Operators," arXiv 1508.02884v2 , 2015.
[14]. A. T. Saumya Saraswat, "A New Approach for Customer Churn Prediction in Telecom Industry," International Journal of Computer Applications, pp. 40-46, 2018.
[15]. T. G. BRANDUSOIU Ionut, "Churn prediction modeling in mobile telecommunications industry using decision trees," Journal of Computer Science and Control Systems, vol. 6, no. 1, pp. 14-19, 2013.

Turiho Jean Claude. A computer science engineer, a data analyst and PhD student in the Department of Computing at the Jomo Kenyatta University of Agriculture and Technology, MSc. in Computer Science and Information Technology (Hunan University-China), BSc. In Information Management (AUCA-Rwanda). His principle research domains are machine learning, business intelligence. He currently teachs at UNILAK and has then years' experience in Data analysis. Published papers: An optimal class association rule algorithm.
Dr. Kipruto Wilson Cheruiyot is the current director of Jomo Kenyatta University of Agriculture and Technology Kigali Campus-Rwanda. He has the following academic qualifications Bsc (Hons); PGD-E (Egerton University, Kenya); Msc in Computer Application and Technology (central south university of technology Hunan, china); PhD in Computer Application and Technology (Central South University, China).
Dr. Ann Muthoni Kibe, Lecturer in School of Computing and Information Technology at Jomo Kenyatta University of Agriculture and Technology.

Turiho Jean Claude" Designing a framework for Real time Churn Prediction in Mobile Telecommunication Industry" The International Journal of Engineering and Science (IJES), 8.6 (2019): 48-52