New Approach for Predictive Churn Analysis in Telecom

MELİKE GÜNAY[1], TOLGA ENSARİ[2]
[1]Computer Engineering, Istanbul Kultur University, Istanbul University-Cerrahpasa, Istanbul, Turkey,
m.gunay@iku.edu.tr
[2]Computer Engineering, Istanbul University-Cerrahpasa, Istanbul, Turkey,
ensari@istanbul.edu.tr

Abstract: - In this article, we propose a new approach for the churn analysis. Our target sector is Telecom industry, because most of the companies in the sector want to know which of the customers want to cancel the contract in the near future. Thus, they can propose new offers to the customers to convince them to continue using services from same company. For this purpose, churn analysis is getting more important. We analyze well-known machine learning methods that are logistic regression, Naïve Bayes, support vector machines, artificial neural networks and propose new prediction method. Our analysis consist of two parts which are success of predictions and speed measurements. Affect of the dimension reduction is also measured for the analysis. In addition, we test our new method with a second dataset. Artificial neural networks is the most successful as we expected but our new approach is better than artificial neural networks when we try it with data set 2. For both data sets, new method gives the better result than logistic regression and Naïve Bayes.

Key-Words: - artificial neural networks, churn analysis, logistic regression, Naïve Bayes, support vector machines.

1 Introduction

One of the recent studies about customer churn analysis was published in May, 2017 entitled ‘Customer Churn Analysis with Machine Learning Techniques’ [1]. In this study, the authors tried to generate three models wtih using support vector machines, Naive Bayes classifier and multi layer artificial neural networks algorithms. They worked on 4667 customer data with 21 features and two target label/class. Test data and training data are divided in to two parts with 25 and 75 percentage correspondingly. As a result of this study, the most successful method is artificial neural networks and second one is Naive Bayes classifier. They have also explained some features as a reason of the failure for support vector machines and the insufficient number of the customer data. Another study, about the customers in telecommunication sector has pointed to importance of the data pre-processing before implementing the learning techniques [2]. Choosing the customers and the features correctly makes the algorithms more efficient. In this study, the authors explain the six phase of pre-processing as understanding the aim of the work, understanding the data, processing the data, modelling and development. However, they worked about particularly on decreasing and preparing the data.

Decreasing the data is choosing relevant features from dataset and preparing the data is converting the data to suitable format for the learning algorithms. Value conversion is defined as converting the continues data to discrete data. Value conversion methods can analyse the missing values as a different category. This part of the study is well suited for the methods like logistic regression. In [3], the authors also use decision trees for the conversion of continuous values to discrete values. Their data is consisting of 30,104 customer data with 156 discrete and 800 continuous features. As a result, it is said that the data pre-processing is increasing the success of the predictions up to %34. In recent studies, deep learning techniques are tried to use to make predictions about customer churn [3]. They worked on unsupervised learning techniques and pointed to importance of decreasing the data size. As a result, they express that the deep learning techniques give almost the same result with unsupervised learning techniques, but it needs to improve.

In another study in the literature, recursive neural network is developed and according to the results the model has shown high performance for predicting the behaviour of the customers [4]. Another advantage of recursive neural networks is that it does not need any data pre-processing steps. It is said that the model had showed same or better results with other vector-based results like logistic regression. For similar methods such as logistic regression, neural networks and random forests, fixed length vectors should be used, however creating correct vectors is usually difficult and take much time. In this study, the proposed model is handling with the difficulty.

In an impressive study, data of online advertisement system is used and two important points for big data analysis is discussed [5]. One of them is insufficiency of data and another one is choosing the suitable machine learning techniques. The authors proposed architecture for managing the customer-advertisement relations. Online and offline advertisement model is improved. Logistic regression is selected to predict correct advertisements for customers. As a result, they decided that the data which is collected within one month is not sufficient. On the other hand, %85 of customers click the advertisement that is selected for themselves correctly and sale rate is increased by accessing the target customers.

Telecommunication sector has also need prediction techniques for customers. C.Cheng and X. Cheng studied on telecom customer segmentation in 2016 [6]. They compared MSQFLA-K, SFLA, BFO and PSO algorithms and stated that the proposed technique with MSQFLA-K shows better segmentation performance comparing to others.

2 Data Sets

2.1 Data Set-I

The first dataset that is used for this study is consisting of 3.333 telecom customers' data. There are 20 features in total that 4 of them are nominal and others are numeric and there is no missing value [7]. We convert nominal values from yes/no to 1/0. In the literature, choosing the features correctly has big effect on the predictions. To be able to select correct features we prefer to use attribute correlation and ranker methods. The technique is using the Pearson correlation for each feature and produces a coefficient and for non-numeric attributes it produces binary value by using weighted average. Pearson correlation computes the statistical relation between variables. Coefficients of the best 10 features can be seen in Table 1.

Table 1. Correlation Coefficients of Features

Correlation Coefficients	Feature Name
0.25985	International Plan
0.20875	Customer Service Calls
0.20515	Total Day Minutes
0.20515	Total Day Charge
0.10215	Voice Mail Plan
0.0928	Total Eve Minutes
0.09279	Total Eve Charge
0.08973	Number Vmail Messages
0.06826	Total Intl Charge
0.06824	Total Intl Minutes

The degrees of correlations are assigned according to interval of coefficients which is shown in Table 2.

Table 2. Degree of Correlation Coefficients

Degree	Value
Perfect	d = ~1
High Degree	0.5 < d < 1
Average Degree	0.3 < d < 0.49
Low Degree	d < 0.29
No Correlation	d = 0

We test machine learning algorithms with the dataset two times that are with and without dimension reduction. By looking degree of the coefficients, we decide ten most affective features from dataset. Thus, our 3.333x20 dataset became 3.333x10. Pearson correlation (r) coefficient is a coefficient that represents the coherence of two features as shown in below equation.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

2.2 Data Set-II

Our second dataset is used for testing of our new approach. However, we also try other well-known machine learning algorithms to find which method is the best one for the churn analysis. Second dataset consists of information about customers in telecom sector. There are 20 features of 7.043 customers [8]. We use all of the features in the data set but need to data conversion from nominal to numeric as shown in the Table 3.

Table 3. Data Conversion for Data Set 2

Feature	Type	Conversion
Gender	String	Female :0 Male:1
Partner	String	No: 0 Yes : 1
Dependents	String	No: 0 Yes : 1
Phone Service	String	No: 0 Yes : 1
Multiple Lines	String	No: 0 Yes : 1 No phone service : 2
Internet Service	String	No: 0 Fiber Optic: 1 DSL: 2

Online Security	String	No: 0 Yes: 1 No internet service: 2
Online Backup	String	No: 0 Yes: 1 No internet service: 2
Device Protection	String	No: 0 Yes: 1 No internet service: 2
Tech Support	String	No: 0 Yes: 1 No internet service: 2
Streaming TV	String	No: 0 Yes: 1 No internet service: 2
Streaming Movies	String	No: 0 Yes: 1 No internet service: 2
Contract	String	Month-to-month: 0 One year: 1 Two year : 2
Paperless Billing	String	No: 0 Yes: 1
Payment Method	String	Mailed Check: 0 Electronic Check: 1 Bank Transfer(Auto.): 2 Credit Card(Auto): 3
Churn	String	No: 0 Yes: 1

3 Methods

Our first aim is measuring the performance of the machine learning algorithms with comparing the original dataset and after dimension reduction. Our classification methods are logistic regression (LR), support vector machines (SVM), decision trees, Naïve Bayes (NB), artificial neural networks (ANN) which are mostly used in previous studies. Our data sets are divided into two parts as train set and test set with %45 and % 55 of the data set correspondingly. Time measurements can be seen for the algorithms when we compare original data set and reduced dataset in Fig. 1.

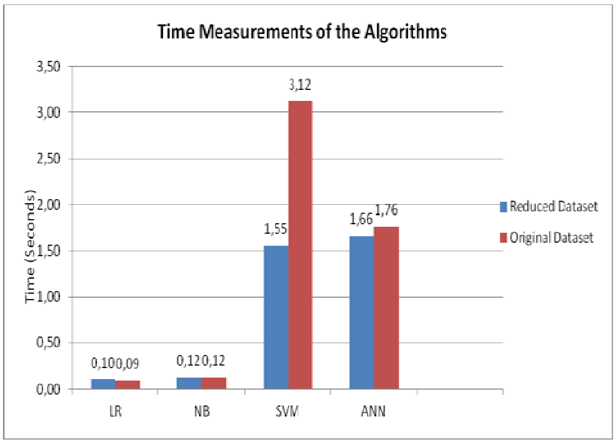

Fig. 1. Time Measurements of Algorithms with Dimension Reduction

From the Fig. 1, for SVM time difference is very important because it became 3.12 seconds from 1.55 seconds when we use original data without any dimension reduction. There are no dramatic results for other methods if we measure time only. Second step for the analysis is looking the correct prediction rates. When we look at the accuracy of the algorithms, ANN is the most successful method, but it takes more time. Detailed information about the algorithms can be seen from Fig. 2. From the graphics, dimension reduction is not a good idea for churn analysis.

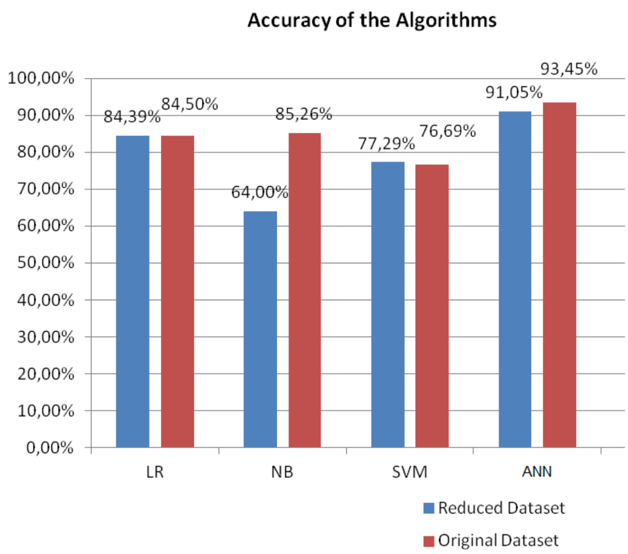

Fig. 2. Accuracy Rate of the Algorithms

Logistic regression is the second successful method after ANN. LR can be preferable if the constraint is time. Dimension reduction shows its affects when we work with Naïve Bayes. Accuracy of the Naïve Bayes in original dataset is %85 but after dimension reduction accuracy reduced to %64. On the other hand, time difference for Naïve Bayes algorithm is not much different, it is only 0.01 second.

4 New Approach

Result of our analysis, we saw that the LR and NB methods are giving the fastest reply, but their accuracy result is not as good as ANN. Thus, we wanted to improve their correct prediction rate with their high speed. For this purpose, we used the probabilities from NB and LR as shown in the Fig. 3.

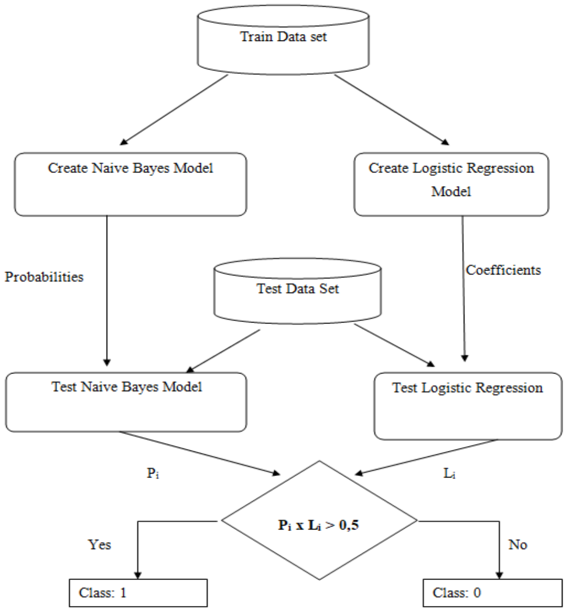

Fig. 3. Work Flow of New Approach

First of all, we use Naïve Bayes model and logistic regression model with coefficients for each feature by using same train data set. We use test data set with the models that is used with train data set for prediction. Multiplication of two probabilities Li and Pi from two models is used to determine class of each data. If multiplication of two probability is greater than 0,5 then class is written as 1 otherwise 0. Class value 1 means that there can be customer churn otherwise it means no churn. In the literature, there are some similar hybrid methods that the authors used average of the probabilities [9]. We also try the method, but our approach gives better results for churn analysis. Confusion matrix of our new methods can be seen from the Table 4.

Table 4. Confusion Matrix of New Approach for Data Set 1

		Predicted Class	
Actual Class		**1**	**0**
	1	TP: 32	FP: 261
	0	FN: 23	TN: 1517

Performance measurements of the method are calculated in the Table 5.

Table 5. Performance Measurements of New Approach for Data Set - I

Accuracy	0.8451
Fault Rate	0.1549
Precision	0.5818
Recall	0.1092
F-Score	0.1839

When we compare our method with NB and LR separately, accuracy is better than them, but we need to make better fault rate and precision. To be sure, the method is tested with a second dataset. Confusion matrix of the method by using second data set can be seen from Table 5. Accuracy with second data set is %79.73 for new method, on the other hand accuracy of ANN method with second dataset is %78.4 which is less than our new approach.

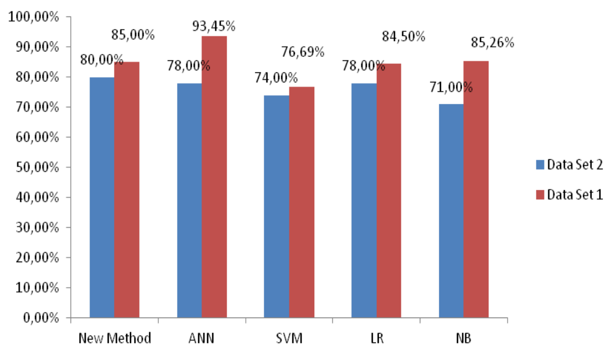

Fig. 4. Accuracy of New Method Comparing to Other Methods

5 Conclusions

In our article, firstly we measure the time for the well-known machine learning methods that are SVM, NB, ANN, and LR for the churn analysis. We compare the time to understand how dimension reduction in data set affects. ANN is the slowest method and LR is the fastest one. However, dimension reduction has the biggest effect on SVM, so it can be logical to reduce dimension of the data set if SVM will be used. On the other hand, when ANN, LR and NB is used, dimension reduction does not make any sense as the manner of time but it reduce the accuracy of the methods, thus dimension reduction is not recommended when we use the methods for churn analysis.

We propose a new approach by using LR and NB and tested it with two different datasets. For the data set 1, our new method shows better performance than LR and NB but not ANN. When we try it with data set 2, new method shows the best performance even more than ANN. As a result, even new approach has high prediction rate, there are another performance measurement that is needed to improve like precision, recall and fault rate. In general, ANN gives the highest result as we expected. Furthermore, from this study we understand that dimension reduction is not always good idea.

References:

[1] O. Kaynar, M. Tuna, Y.Görmez, M. Deveci,*"Makine öğrenmesi yöntemleriyle müşteri kaybı analizi"*, C.Ü. İktisadi ve İdari Bilimler Dergisi, Cilt 18, Sayı 1, 2017.

[2] K.Coussement, S. Lessmann, G. Verstraeten, "*A Comparative Analysis of Data Preparation Algorithms for Customer Churn Prediction: A Case Study in the Telecommunication Industry"*, Decision Support Systems, 95 (2017) 27–36 .

[3] P. Spanoudes, T. Nguyen, "*Deep Learning in Customer Churn Prediction: Unsupervised Feature Learning on Abstract Company Independent Feature Vectors"*,Cornell University, March, 2017.

[4] T. Lang, M. Rettenmeier, *"Understanding Consumer Behavior with Recurrent Neural Networks"*, Zalando SE Techblog, 2017.

[5] [5] T.Zhang, X.Cheng, M.Yuan, L.Xu, C.Cheng, K.Chao, *"Mining Target Users for Mobile Advertising Based on Telecom Big Data",* 6th International Symposium on Communications and Information Technologies (ISCIT), 26-28 Sept. 2016.

[6] C.Cheng, X.Cheng, *"Anovel Cluster Algorithm for Telecom Customer Segmentation"*, International Symposium on Communications and Information Technologies (ISCIT), 26-28 Sept. 2016.

[7] BIGML,https://bigml.com/user/bigml/gallery/dataset/4f89bff4155268645c000030, 10.01.2018.

[8] IBM, https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/, 12.03.2018.

[9] A. Chaudhary, S.Kolhe, R.Kamal, "*A Hybrid Ensemble for Classification in mutliclass datasets: An application to Oilseed Diasease Dataset"*, Computers and Electronics in Agriculture, April, 2016.