Churn Prediction:

A Comparative Study Using KNN and Decision Trees

Mohammad A. Hassonah(1), Ali Rodan(1),(2) , Abdel-Karim Al-Tamimi(2),(3), and Jamal Alsakran(1),(2)

(1)King Abdullah II School for Information Technology
University of Jordan
Amman, Jordan

(2)CIS Department
Higher Colleges of Technology
UAE

(3)Computer Engineering Dept
Yarmouk University
Irbid, Jordan

***Abstract*—Churn prediction represents one of the most important components of Customer Relationship Management (CRM). In the purpose of retaining customers and maintaining their satisfaction, researchers of many fields including business intelligence, marketing and information technology were motivated to investigate the best methods that deliver the best services for customers. Many machine learning algorithms had been implemented in the purpose of optimally predicting the possible churning customers and making the right decisions at the right moments. Researchers had conducted several studies on various types of algorithms and results were found very promising. In this paper, we are conducting a comparison study of the performance towards churn prediction between two of the most powerful machine learning algorithms which are Decision Tree and K-Nearest Neighbor algorithms. Results were quite interesting showing a quite large dissimilarity in many areas between the two algorithms.**

Keywords—churn prediciton; churn rate; k-nearest neighbor; decision tree; customer relationship management

I. INTRODUCTION

Customer Relationship Management (CRM) represents the main focus of attention for many organizations around the world, due to the competitive nature among these organizations. Preserving customers for service providers forms a great issue for all companies, service providers and especially telecommunication companies. Customers are able to switch providers anytime they want when the supplied services do not meet their demands, or for any reason that leads to their dissatisfaction. For this purpose, churn rate was introduced in order to calculate the percentage of customers ending their subscriptions with their service providers or with certain services, and then, to determine the actions to be taken to improve the overall performance based on these historical statistics [1-3].

Mobile telecommunication companies normally suffer a loss of 20-40% every year [4]. Loosing such numbers of customers can significantly reduce companies profitability, and affect prices of offers made on their services because of the need to stabilize profitability by compensating the loss in the number of customers [5]. In addition, the process of gaining new customers costs much more than keeping existing ones due to marketing and advertising expenses related to this activity [6].

The prediction of churn customers is not the only requirement for the telecommunication companies, but the ability to also predict the remaining time for customers before they leave the company. This allows the organization to determine the time needed to react and initiate recovery programs for churn-risky customers [7].

In order to enforce such actions for churn customers and practically apply them in practice, churn prediction models have to be carefully and accurately applied. This is accomplished by using predictive modeling methods that use a set of data mining algorithms, which help organizations to classify and effectively identify possible churn-risky customers [8].

The remainder of this paper is organized as follows: Section two will present related work and literature review of churn prediction research utilizing Decision Tree and K-Nearest Neighbor algorithms. Section three of this paper discusses the methodology, experiment setup and results. Section four will conclude and summarize the findings of this paper.

II. LITERATURE REVIEW

There has been a significant amount of research related to churn prediction addressing the limitations of their datasets and proposing different algorithms and techniques to identify churn customers from certain datasets. The following subsections discuss the related work of the most important challenges that have faced the researchers with their respective proposed solutions. Moreover, we will focus our literature review on discussing related work that used Decision Tree and K-Nearest Neighbor classification techniques.

A. Overcoming The Limitations of Raw Datasets

Many challenges have faced researchers while testing the techniques and algorithms to extract the optimal churn rate and identify churn customers. One of the most complicated issues is dealing with incomplete datasets that have missing values [9]. In addition, large datasets can be hard for researchers to organize and handle because of the existence of noisy data records [10]. Another issue that was also considered by [9] and [11] is the mixed nature of the data because of the characteristics of inputs for data mining algorithms.

However, recent studies and approaches were made in order address these problems. For example, [12] and [13] had

978-1-7281-5061-1/19/$31.00 ©2019 IEEE

proposed two approaches for improving the quality of the data through performing several pre-processing steps including data cleaning and transformation.

B. *Churn Prediction using Decision Trees*

Decision tree technique is one of the most used classification algorithm [14-15], due to its ability to operate in an environment where there are large amounts of noisy data, and its ability to avoid over-fitting issue in classification. The authors in [16] suggested using decision trees for churn prediction. Their experiment showed improvements over previous models that were already used by phone system services. The precision of the classification results reached 0.95.

A comparative study was made by [17], comparing churn results against a set of machine learning algorithms including Artificial Neural Networks (ANN), K-Means Clustering and decision trees. The performance was measured through precision, recall and F-measure. Decision trees, and especially Chi-squared Automatic Interaction Detector (CHAID) type, which had the best results among all of them with respect to the previous measures.

The authors in [18] used a composite criterion to find the best classification technique in addition to accuracy. They compared decision trees with other techniques, and concluded that all methods have accepted values of accuracy, but the decision tree method had the highest values in terms of Evaluation Composite Indicator (ECI).

C. *Churn Prediction Using K-Nearest Neighbor*

There is a limited of research efforts that utilize K-Nearest Neighbor (K-NN) on churn prediction. However, authors in [19] used churn prediction to evaluate the implementation of K-NN on time-series classification. The results showed that their proposed approach achieve better performance than the traditional information-based approach.

In addition, the authors in [20] combined K-NN with Logistic Regression (LR) in order to improve the accuracy for LR. Results were compared to a churn model using C4.5, which is an algorithm used to generate decision trees [21]. Results showed that the K-NN/LR model was slightly better than the C4.5 model.

III. METHODOLOGY

Our study focuses on the performing a comprehensive comparison between Decision Trees and K-Nearest Neighbor classification algorithms. Both of the algorithms have the same input data with the same training and testing ratios. Input data for both algorithms is processed through filters to reduce noise and remove undesired data. Afterwards, the cleaned data is split into training and testing sets. Training sets are then modeled using algorithms to deliver the desired output. The following subsections will give more details on these processes.

A. *Experiment Setup*

K-NN algorithm setup used in this paper is configured with the standard number of neighbors k=5 (which we confirmed as the best k value after several trials), and with mixed Euclidean distance measures. As for the Decision Tree algorithm, we used the standard gini index criterion with maximal depth of a size equals to 20 levels.

The dataset that we used to compare both algorithms is a historical dataset which belongs to a telecommunications company, consisting of 3333 samples (customers), with 20 different variables, and a churn decision variable for each sample. A small percentage of 5% of the dataset was used for churn prediction, and the remaining 95% of the sample was allocated for training and testing the algorithms. The 95% (3166) samples are split into both training and testing sets with percentages of 60% and 40%, respectively.

B. *Performance Evaluation Criteria*

Different performance criteria were targeted to form the basis of our comparison of these algorithms. As follows, we mention these measures with a brief explanation for each one of them.

- *Accuracy*: Measures the percentage of true rate values of all classes. The higher percentage of accuracy, the better performance (higher is better).
- *Precision*: Calculates the percentage of true positive rate values of relevant elements to the irrelevant ones. The higher the percentage of precision, the more relevant results retrieved than the irrelevant ones (higher is better).
- *Recall*: Also known as Sensitivity Measure or True Positive Rate. It is the ratio of true positive rate of relevant values. Higher ratio means higher retrieved relevant elements (higher is better).
- *F-Measure*: Also known as F1 score, and it is a weighted average of both precision and recall. The highest value of F-Measure is 1, which means the best score. The lowest value is 0, which means the worst score (higher is better).
- *Area Under Receiver Operating Characteristic Curve* (AUROC) or (AUC): The area calculated under the performance curve, which contains the true positive rate representing the sensitivity measure as its y-axis, and the true negative rate representing specificity measure as its x-axis. The higher the percentage of AUC, the better performance of the algorithm (higher is better).
- *Lift Measure*: In our case, lift measure is important to calculate the ratio of the count of predicted churn customers against the count of confidence of the predicted values for both algorithms (higher is better).

C. Experiment Results

In this subsection, we share our results of comparing Decision Tree and K-NN algorithms performance in predicting the remaining 40% of samples.

1) Confusion Matrix

Confusion matrix results allow us to view the actual results for both algorithms, by viewing the number of truly predicted positive and negative results. Tables 2 and 3 demonstrate the confusion matrices for both decision tree and k-nearest neighbor algorithms, respectively.

TABLE 1: Confusion Matrix for Decision Tree (DT) Algorithm

	Actual -	Actual +	
Predicted -	TN= 1082	FP= 174	1256
Predicted +	FN= 1	TP= 8	9
	1083	182	1265

TABLE 2: Confusion Matrix for K-NN Algorithm

	Actual -	Actual +	
Predicted -	TN= 1063	FP= 150	1213
Predicted +	FN= 20	TP= 32	52
	1083	182	1265

The true prediction of churning (true positive) customers for the decision tree forms only 8 customers out of 182 which represents a low percentage, while the prediction for the same set regarding the k-nearest neighbor reached 32. This means that better results were obtained using K-NN algorithm. However, the results for non-churning (true negative) customers for the DT were 1082 out of 1083, and 1063 out of 1083 for the K-NN algorithm. Nonetheless, the prediction for churning customers, or true positive results, represents the most important aspect not only for this study, but for telecommunication companies.

2) Weights of Attributes

Weights are considered in this paper in order to indicate that all of the differences were according to the algorithms only. As depicted in Fig.1, we can notice that all the attributes have the same weights, for instance, both *Day Mins* and *Day Charge* attributes had the highest value of 1. Being nearly as important, both *CustServCalls* and *Int'l Plan* attributes have the weights of 85% and 65%, respectively. On the other hand, the attributes *Area Code*, *Account Length* and *Phone* were given the value 0 from both algorithms since they have no contribution to the classification process.

All of the previously mentioned points indicate that the weights given to the attributes are logically acceptable and the initial configuration values for both algorithms are correctly set.

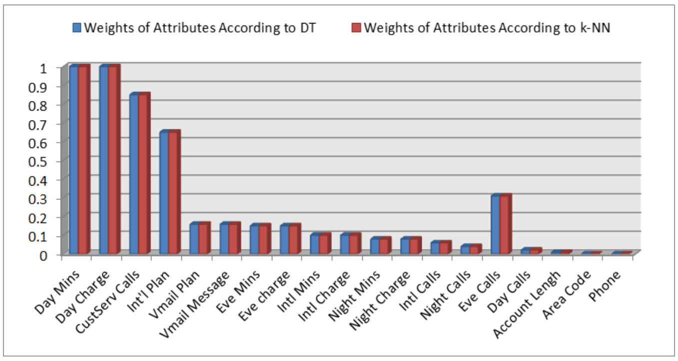

Fig.1: The Attributes Weights for K-NN and DT

TABLE 3: Performance Evaluation Measures for DT and K-NN

Criterion / **Algorithm**	**Accuracy**	**Precision**	**Recall**	**F-Measure**
K-Nearest Neighbor	0.868	0.615	0.220	0.326
Decision Tree	0.926	0.775	0.681	0.725

3) Accuracy, Precision, Recall and F-Measure

Table 3 summarizes the performance results of both algorithms. Both algorithms have similar accuracy results as the DT had a percentage of 93%, while the accuracy of K-NN reached 87%. The largest difference between the two algorithms was with the F1 score reaching around 33% and 73% for K-NN and DT respectively.

The cause of this difference is due to the recall (sensitivity) measure; as it appears to be very low for K-NN algorithm. Such a low value is caused by the number of K neighbors, and while trying to improve the recall by decreasing the number of neighbors, the precision was decreasing in around the same ratio with the increasing recall, resulting in the same F-Measure.

4) Area Under the Curve (AUC)

By examining Fig.2, we can note that the values of AUC for both algorithms are close. K-NN reached the value of around 82% and DT reached almost the value of 86%. This shows that the AUC value for DT is slightly better than the AUC value for K-NN. The closeness of this result is a direct result of the difference between the two algorithms in terms of their true negative rates, which is represented by the specificity as shown in the x-axis in Fig. 2. In other words, DT has higher true negative rates than K-NN, causing the proximity of results in the AUC outcomes.

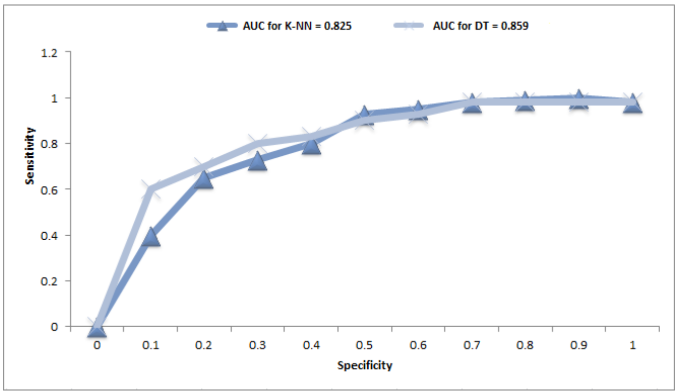

Fig.2: The AUC values for DT and K-NN algorithm

5) Lift Measure

The Lift measure was found to be higher in DT algorithm, reaching the value of 5.387, and 4.277 for the K-NN algorithm. This is the result of the DT average confidence-for-yes which reached the value of %97.5 with a deviation equal to only %7, and a K-NN average confidence-for-yes to be %13 with a deviation of %19. This indicates that the DT has much more responsiveness than K-NN algorithm.

IV. Conclusions and Future Work

Churn rate prediction remains to be one of the most interesting topics regarding Customer Relationship Management in the business world. Many machine learning techniques had been introduced, trying to find the optimal one for churn prediction. This paper introduced a comparison between two of the most powerful learning algorithms and compared their performance and efficiency towards the churn prediction problem. Statistical results were found to be in favor of the decision tree algorithm; accuracy, precision, recall, F-measure and the Lift measure were found to be better in DT than K-NN algorithm. The AUC was found to have nearly the same results for both algorithms which is caused by the low values of specificity for DT leading to the similarity of AUC values for both algorithms. However, results shown in the confusion matrix indicate that K-NN had more true positive rates than the DT algorithm telling us that the K-NN predicted better than the DT regarding real churning customers.

More future work should be done in order to measure other important values such as the Up-Selling and the Appetency predictions which also help in proving the efficiency either in favor of the Decision Tree or the K-Nearest Neighbor algorithms.

References

[1] Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, *34*(10), 2902-2917.

[2] Churn Rate Definition, . Investpedia. URL[http://www.investopedia.com/terms/c/churnrate.asp]

[3] RESCI. Customer Churn Rate: Definition, Measuring Churn and Increasing Revenue. URL[http://en.wikipedia.org/wiki/Churn_rate]

[4] Berson, A., & Smith, S. J. (2002). *Building data mining applications for CRM*. McGraw-Hill, Inc..

[5] Rechinhheld, F., & Sasser, W. (1990). Zero defections: Quality comes to service. *Harvard Business Review*, *68*(5), 105-111.

[6] Siber, R. (1997). Combating the Churn Phenomenon-As the problem of customer defection increases, carriers are having to find new strategies for keeping subscribers happy. *Telecommunications-International Edition*, *31*(10), 77-81.

[7] Lu, J. (2002). Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS. *SAS User Group International (SUGI27) Online Proceedings*, 114-27.

[8] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, *31*(3), 515-524.

[9] Huang, B. Q., Kechadi, T. M., Buckley, B., Kiernan, G., Keogh, E., & Rashid, T. (2010). A new feature set with new window techniques for customer churn prediction in land-line telecommunications. *Expert Systems with Applications*,*37*(5), 3657-3665.

[10] Idris, A., Rizwan, M., & Khan, A. (2012). Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, *38*(6), 1808-1819.

[11] Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*,*36*(3), 5445-5449.

[12] Galhardas, H., Florescu, D., Shasha, D., & Simon, E. (2000). Declaratively cleaning your data using AJAX. In *In Journees Bases de Donnees*.

[13] Raman, V., & Hellerstein, J. M. (2000). *An interactive framework for data cleaning*. Computer Science Division, University of California.

[14] Seol, H., Choi, J., Park, G., & Park, Y. (2007). A framework for benchmarking service process using data envelopment analysis and decision tree. *Expert Systems with Applications*, *32*(2), 432-440.

[15] Mendonça, L. F., Vieira, S. M., & Sousa, J. M. C. (2007). Decision tree search methods in fuzzy modeling and classification. *International Journal of Approximate Reasoning*, *44*(2), 106-123.

[16] Bin, L., Peiji, S., & Juan, L. (2007, June). Customer churn prediction based on the decision tree in personal handyphone system service. In *Service Systems and Service Management, 2007 International Conference on* (pp. 1-5). IEEE.

[17] Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A., & Rehman, A. (2013, September). Telecommunication subscribers' churn prediction model using machine learning. In *Digital Information Management (ICDIM), 2013 Eighth International Conference on* (pp. 131-136). IEEE.

[18] Clemente, M., Giner-Bosch, V., & San Matías, S. (2010). Assessing classification methods for churn prediction by composite indicators.*Manuscript, Dept. of Applied Statistics, OR & Quality, Universitat Politècnica de València, Camino de Vera s/n*, *46022*.

[19] Lee, Y. H., Wei, C. P., Cheng, T. H., & Yang, C. T. (2012). Nearest-neighbor-based approach to time-series classification. *Decision Support Systems*, *53*(1), 207-217.

[20] Zhang, Y., Qi, J., Shu, H., & Cao, J. (2007, October). A hybrid KNN-LR classifier and its application in customer churn prediction. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on* (pp. 3265-3269). IEEE.

[21] Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.