

RESEARCH

Open Access



The use of knowledge extraction in predicting customer churn in B2B

Arwa A. Jamjoom* 

*Correspondence:
ajamjoom@kau.edu.sa
Department
of Information System,
Faculty of Computing
and Information Technology,
King Abdulaziz University,
Jeddah, Saudi Arabia

Abstract

Data mining techniques were used to investigate the use of knowledge extraction in predicting customer churn in insurance companies. Data were included from a health insurance company for providing insight into churn behaviour based on a design and application of a prediction model. Additionally, three promising data mining techniques were identified for the prediction of modeling, including logistic regression, neural network, and K-means. The decision tree method was used in the modeling phase of CRISP-DM for identifying the attributes of churned customers. The predictive analysis task is undertaken through classification and regression techniques. K-means clustering variation is selected for exploring if the clustering algorithms categorize the customers in churning and non-churning groups with homogeneous profiles. The findings of the study show that data mining procedures can be very successful in extracting hidden information and get to know customer's information. The 50:50 training set distribution resulted in effective outcomes when the logistic regression technique was used throughout this study. A 70:30 distribution worked effectively for the neural network technique. In this regard, it is concluded that each technique works effectively with a different training set distribution. The predicted findings can have direct implications for the marketing department of the selected insurance company, whereas the models are anticipated to be readily applicable in other environments via this data mining approach. This study has shown that the prediction models can be utilized throughout a health insurance company's marketing strategy and in a general academic context with a combination of a research-based emphasis with a business problem-solving approach.

Keywords: Customer Churn, Data mining techniques, Extraction knowledge, Marketing department

Introduction

A customer offers valuable insights on prospective decisions through customer behaviour data. For instance, churn prediction models determine customers who stop utilizing a service or product [12]. This is of significant interest for product providers because a large number of churning customers not merely drives to attrition of revenue but can also harm the reputation of a company [19]. According to Engel, Blackwell, and Mansard, consumer behaviour is the action and the decision process of people who buy goods and services for their consumption. It is important to forecast consumer behaviour as

it helps to differentiate consumers from each other which helps to set a target market for the company, it also helps in retaining potential consumers, helps in re-designing marketing programs, assist in predicting market trends, create competition, bring innovation to the existing products and develop new ones, stay relevant to the market and improve customer services [24].

The field of customer churn prediction receives much less emphasis in business-to-business contexts while it is well-researched in the customer-to-business context. The frequency of customers is often substantially lower in B2B businesses, but their transactional values are usually a lot higher [9]. Thereby, single customers are of great interest to a company and the effect of attrition can be much greater. This supports the appropriateness of customer churn prediction in B2B domains. On the contrary, measures established for B2C systems can usually not be integrated into B2B environments because of their multifaceted setups [27].

Customer churn occurs when customer expectations are not fulfilled. It is the loss of a retained customer to a competitor [20]. A competitor is a different brand in this study, which can result in churning customers even though the customer remains in the same company [1]. According to Rohini and Devaki, [21] inspecting customer churn for huge data in terms of customer retention is open research in machine learning technology. They further explained that customer churn means a loss of customers who switches from one sector to another. When customer churn is misclassified utilizing clustering, it can yield huge financial losses and even hurt the development of the association.

Firstly, the churning customers must be identified and then such customers must be encouraged to stay for managing customer churn. The marketing cost to attract new customers is three to five times higher retaining customers, allowing customer retention an interesting topic for all businesses [23]. For instance, insurance companies are specifically associated with customer retention and satisfaction since the needed fundamental insurance package is generally the same for each company [11]. This can create a highly competitive and dynamic environment in which customers are competent enough to instantly switching between insurance companies. The majority of the firms usually serve millions of customers, which makes it complicated for extracting beneficial data on customer switching behaviour and for predicting modifications in customer retention [22]. A study conducted by Amjad et al. [2], suggested a hybrid data mining learning approach for predicting customers' churn. Their 3 models were carried by stages of clustering and predicting performance. The information about the customers was filtered and utilized through the K-means algorithm and a Multilayer Perceptron Artificial Neural Systems (MLP-ANN) for prediction. The use of clustering with MLP-ANN, their model, used self-sorting out maps (SOM) with MLP-ANN on the data. The churn rates and precision values were compared and determined with other state-of-art. To be precise, their work reflected that the 3 crossover models outperformed single normal and common models.

Another study by Wenjie et al. [28] was carried out in which a clustering algorithm called semantic driven subtractive clustering technique (SDSCM) using a Hadoop map reduced structure was suggested. This model suggested in the study proved to be fast as compared to different techniques as well as recommended few showcasing procedures according to clustering algorithm to assure benefit implications.

Furthermore, a study by Fathian et al. [8] showed a comparison of single standard classifiers and ensemble classifiers to predict customer churn. Their study built an aggregate of 14 prediction models which were grouped into four categories such as; fundamental classifier, (Decision Tree, K-Nearest Neighbour, SVM, and Artificial Neural Network), Classifier with SOM and basic classifier, Classifier with SOM, and reducing features with PCA and basic classifier, and lastly Classifier with SOM and reducing features with PCA and bagging and boosting ensemble classifier.

A dynamic competitive environment is evident in such a strictly regulated market. The interference of government is not observed with the additional insurance policies and this combination develops a competitive and dynamic environment [25]. There is a reduction in customer churn from 6.9% to 5.3% in 2018 in health insurance firms, but this still covers 1.2 million customers due to the stagnant price level of health insurance. In that year, the high churn percentage is reflected from the outflow of 2018 comprising of switches in group insurances [26].

Arowolo et al. [4] have utilized PCA feature extraction algorithm for acquiring the latent constituents that can assist improving the classification of a mosquito anopheles gambiae data through SVM polynomial kernel and Gaussian kernel on a reduced dimensional data, integrating PCA algorithm. The study has comparatively showed that SVM-Gaussian Kernel was outperformed with 99.68% through SVM-Polynomial kernel. Olaolu et al. [18] have applied dimensionality reduction methods for obtaining the minimal set of genes that contributes to the efficient performance of classification algorithms in microarray data. The findings have shown high accuracies and, thus, compared the performances of the dimension reduction techniques. Significant accuracy was observed through the PLS-based method as compared to other dimension reduction methods such as PCA and One-way ANOVA).

Arowolo et al. [3, 6] have combined feature extraction and selection into a generalized model to obtain an efficient and robust dimensional space. The study has employed One-Way ANOVA for obtaining an optimal number of genes, partial least squares, and PCA as feature extraction methods, independently. In this regard, irrelevant and redundant attributes were removed to present an efficient and accurate performance of almost 98% over the state-of-art.

Arowolo et al. [4, 5] have utilized PCA feature extraction algorithm for obtaining latent constituents and assesses its classification performance via decision tree classification and KNN algorithms. The effectiveness of this experiment was validated via RNA-Seq dataset on a mosquito anopheles gambiae. The findings have indicated an accurate performance metric with a classification accuracy of 86.7% and 83.3%, respectively. Arowolo et al. [4, 5] have proposed a hybrid dimensionality reduction technique for fetching pertinent subset attributes from the data. Features selected were passed into PCA and independent component analysis methods based on the class variants, for helping transform the chosen attributes into a lower dimension independently. The reduced malaria vector dataset was used within SVM kernel classifiers for evaluating the classification performance of the experiment. Arowolo et al. [3, 6] demonstrated the effectiveness of feature extraction and investigated the most efficient approach that can be utilized for improving microarray classification. This study has undertaken PCA and PLS as a supervised technique for the dataset.

The overall findings have indicated that PLS algorithm offers an enhanced performance of 95.2% accuracy as compared to PCA algorithms.

Not many studies have been conducted to find out how customer churn can be predicted in a health insurance company; therefore, the study aims to forecast which customers will switch and comprehend why such customers switch. Moreover, the study suggests a prediction model which is authentic and applicable for the marketing department. The study contributes in a way that it provides a prediction model which can be adopted by the marketing department to implement predict customer churn in a B2B business. This prediction model can help the managers at health insurance companies to forecast consumer behavior to identify potential customers, re-design marketing strategies, predict market trends, create a competition, bring innovation in existing products and services as well as develop new ones, stay relevant to the market and improve customer services. The majority of the studies have investigated a business-to-customer relationship. This research contributes to filling the literature gap by conducting a study on the use of knowledge extraction in predicting customer churn in a B2B environment. This study suggests a customer churn prediction model using CRISP-DM, decision tree method, and data mining techniques are also taken into consideration while designing the model for health insurance companies. The data collection for predicting customer churn in health insurance is the novel aspect of this research.

Besides, this prediction model can be utilized throughout the marketing strategy of a health insurance company and in a general academic context, combining a research-based emphasis with a business problem-solving approach. The objective of this study is to investigate the use of knowledge extraction in predicting customer churn in insurance companies in this regard, the following questions were addressed:

Question 1 What are the possibilities to create highly accurate prediction models for predicting the number of customer churn?

Question 2 Which customer behaviour and attributes are essential in predicting customer churn behaviour?

Question 3 Which techniques can be used for generating effective churn prediction models?

The remainder of this paper is structured as follows: “[Introduction](#)” Section provides an overview of related work in this area, before elaborating the research method and the research context in “[Methods](#)” Section. The findings and discussion of the study were described in “[Churn prediction model generation](#)” Section, followed by “[conclusion and recommendations](#)” Section.

Methods

There are different clustering algorithms such as K-Means Clustering, Mean-shift clustering, Density-Based Spatial Clustering of Applications with Noise, Expectation–Maximization clustering using Gaussian Mixture Models, Agglomerative Hierarchical Clustering and Decision Tree Method. This study adopted, K-means and Decision tree method to built a customer churn prediction model using CRISP-DM.

CRISP-DM

A customer churn prediction model is built using CRISP-DM based on a six-phase approach including (1) business understanding, (2) data understanding, (3) data pre-processing, (4) modeling, (5) evaluation, and (6) deployment [14]. The decision tree method was used in the modeling phase of CRISP-DM for identifying the attributes of churned customers. According to Nadali et al. [15], the Cross-Industry Standard Process for data mining (CRISP-DM) is a processing model which consists of 6 phases that describes the life cycle of data science. They help to plan, organize and then implement the data science project. The six phases of this model are as follow:

Business understanding This phase helps to understand the objectives and the requirements of the project; data mining and the problem definition.

Data understanding This phase helps in familiarization and collection of data, identifies the issues related to data quality and initial obvious results.

Data preparation In this phase recording and attribution selection takes place in short, data cleansing takes place.

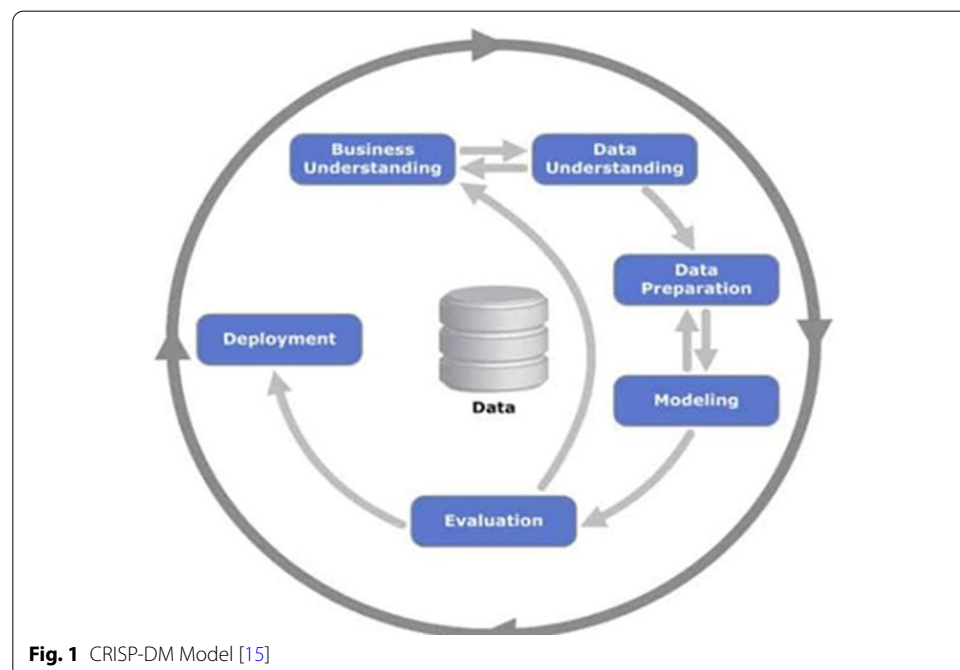
Modeling In this phase, data mining tools are run.

Evaluation This phase helps to determine whether the results met the objectives of the business and identifies business issues that should have been addressed earlier.

Deployment This phase puts the resulting models into practice and sets continuous data mining. Figure 1 sums up all the six phases of the CRISP-DM model.

Data mining techniques

In data mining, the predictive analysis task is undertaken through classification and regression techniques. Regression is a statistical method that is used to estimate relationships between dependent variables to one or more independent variables. It can also be used to assess the strength of the relationship between variables as well as to



model future relationships between them, whereas classification is a predictive modeling problem where a class label is predicted for input data. They instigated classification as a procedure to find a model that demonstrates and identifies data concepts or classes. Afterward, the model has been used for predicting class labels of objects with unidentified labels.

K-means

K-means is a famous and important clustering technique that was introduced by Mc Queen in the year 1967. The important stride in K-mean clustering consists of the initial step that chooses k objects that have their mean. The relationships are assessed for determining the distance between cluster means and the objects after the computation, whereas the new focal point is identified for the facts. The above steps are repeated until the required function is achieved. In k-means clustering, the most crucial point is to explore the numbers that are optimum as the separation between the means of clusters and objects. The algorithm works in a way that no new cluster component leaves a cluster and goes into the other group as well as no new focal point is set for any cluster. When the objective is accomplished at this point, the algorithm is ceased [16].

K-means clustering intends to divide the cases into K clusters. Each case or customer is a segment of that cluster, which has its core nearer to the case centroid. The centroid refers to the average value for all the variables [10]. The Euclidean distance and the Manhattan distance were utilized for calculating the centroid concerning K-means. The number of clusters varies from $K=1$ to $K=8$. This variation is selected for exploring if the clustering algorithms categorize the customers in churning and non-churning groups with homogeneous profiles. It is essential for normalizing the variables before the implementation of the algorithm. The weight of the variable with the largest variation is larger when the variables are not normalized [13].

Table 1 showed that young customers churn more often even though they consume less health insurance as compared to the average ones and pay the premium themselves. On the other hand, young consumers are also encompassed, which do not pay the premium themselves and have group insurance. Older customers are majorly non-churning customers who have no voluntary deductible excess and consume additional health insurance as compared to the average ones. The k-means can be calculated through the following equation.

Table 1 Profiles defined with K-means clustering in health insurance firm

| Clusters ($K=3$) | Churn rate (%) |
|--------------------|----------------|
| 1 | 2.5 |
| 2 | 3.5 |
| 3 | 10 |
| Clusters ($K=5$) | |
| 1 | 2.7 |
| 2 | 3.2 |
| 3 | 3.7 |
| 4 | 7.1 |
| 5 | 13.1 |

$$J = \sum_{j=1}^k \sum_{i=1}^n \left\| x_i^{(j)} - c_j \right\|^2$$

Decision tree

The decision tree sorts it via the tree to the appropriate leaf node when an instance is categorized by a DT model [29]. A classification is showed by each leaf node. The DT further produces outcomes, which are easy to comprehend, and it further has the competence for developing models via categorical and numerical datasets. DT techniques were integrated for building a prediction model for consumer churn from services. One reason is associated to explore the attributes of churners and the need for understanding if–then rules for this objective [7]. DT technique was chosen for the modeling phase since DT offers an easy understanding of rules. The type of data is the other reason. Numerical and categorical data types are included and, thus, DT was appropriate for such data types. Thereby, DT was integrated for the modeling phase.

The most effective split is applied at each node. The criteria of splitting rely on information extraction. This information extraction refers to the appropriateness of a feature. The Minimal Description Length pruning method was integrated for identifying the least justifiable branches of the tree after the decision tree is constructed. Nodes replace the least authentic branches throughout the pruning process, which were initially split nodes. The application of the pruning method took place throughout the training set on the decision tree generated. The models were generated for various parameter settings where the minimum number of records will be 0.1 and 1% of the training set for 5 and 50 cases, respectively. In addition, the different training set distributions are integrated into this study. The findings of the five models are presented with these two different parameter settings (Tables 2 and 3).

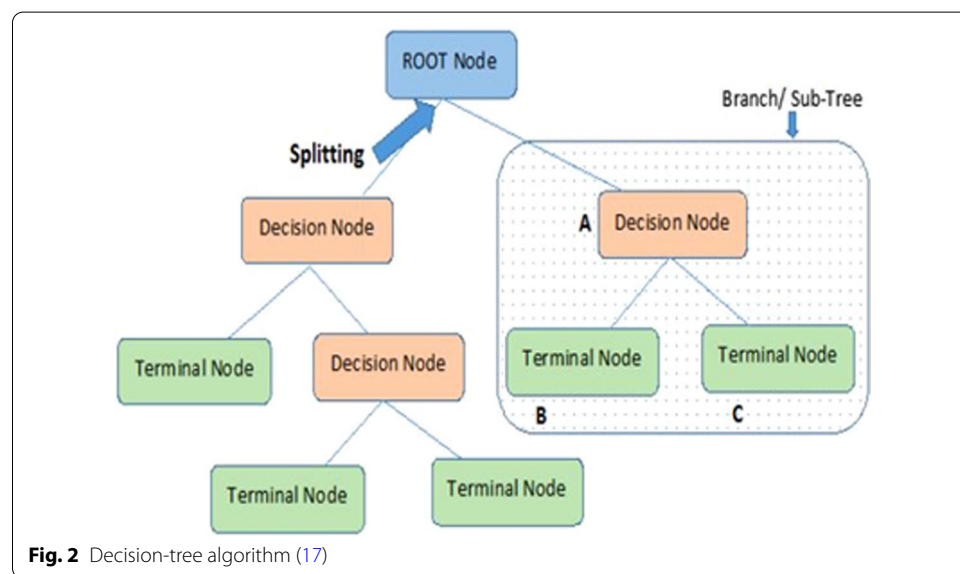
The largest AUK value is generated with a training set distribution of 60:40 when the minimum number of records is adjusted to 0.1%. However, the value of AUC is low and higher for model 5 with the minimal number of records adjusted to 1%. Moreover, the AUK value undertakes that the data is imbalanced, resulting in the conclusion that model 3 is the best model with a minimum of 0.1% records per node when the AUK and

Table 2 Decision tree models performance parameters

| Minimum number of records set to 0.1% | | | | | |
|---------------------------------------|---------|---------|---------|---------|---------|
| Performance parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| AUC | 0.030 | 0.015 | 0.075 | 0.085 | 0.025 |
| AUK | 0.562 | 0.520 | 0.565 | 0.683 | 0.612 |
| Minimum number of records set to 0.1% | | | | | |
| Performance parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| AUC | 0.023 | 0.035 | 0.046 | 0.061 | 0.062 |
| AUK | 0.521 | 0.565 | 0.730 | 0.607 | 0.781 |

Table 3 Performance parameters, sensitivity, and precision of the decision trees

| Decision tree 0.1% | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|--------------------|---------|---------|---------|---------|---------|
| Threshold | 0.4855 | 0.089 | 0.591 | 0.852 | 0.237 |
| Precision | 0.827 | 1.839 | 0.326 | 1.660 | 0.517 |
| Sensitivity | 0.033 | 1.780 | 2.159 | 0.557 | 0.271 |
| Kappa | 0.088 | 0.145 | 0.283 | 0.922 | 0.048 |
| Decision tree 1% | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Threshold | 0.428 | 0.190 | 0.284 | 0.808 | 0.938 |
| Precision | 0.226 | 0.142 | 2.526 | 1.114 | 0.198 |
| Sensitivity | 0.751 | 1.183 | 1.292 | 0.759 | 0.973 |
| Kappa | 0.257 | 0.499 | 0.704 | 0.148 | 1.587 |

**Fig. 2** Decision-tree algorithm (17)

the AUC were utilized as performance measures. It can be observed that the precision is the second-highest for all models when model 3 was examined with the performance measures precision and sensitivity with a minimum of 0.1% records per node.

Model 1 is the highest precision model, but it can be concluded that this model does not do effectively in the presence of low sensitivity. On the contrary, a high sensitivity level is not indicated by all the models generated with the decision tree techniques. For these performance parameters, it is concluded that model 3 is the best performing model of the decision tree technique with a minimum of 0.1% records per node. Figure 2 explains the decision-tree algorithm.

Churn prediction model generation

Five training sets were created for generating the prediction models through decision-tree, varying in the contribution between churners and non-churners. The training set with the original distribution was represented in Model 1 where model 2 presents 70:30 (non-churning: churning) distribution, and model 3 to 5 presents 80:20, 65:35, and 50:50 distributions, respectively. These training set distributions are usually associated with

the same model number for all techniques. It should be noted that the first three data sets were gathered from the database of the selected company. The original distribution between churners and non-churners is represented through one data set. A set was extracted with merely churning customers and one set with non-churning customers from the selected company database.

Afterward, the test set and the training sets are constructed. The test set was used for testing all models. The training sets were utilized in four different techniques for predicting which customer will be going to churn. Different results were witnessed based on different technique settings. During the model generation, the settings were also discussed per technique.

Performance measurement

Accuracy, sensitivity, and specificity are considered quality measurements. The distributions are utilized as a training set for generating the model. The testing set will usually have a distribution equal to the distribution of the original data set. Some problems can be raised through an imbalanced data set when it is evaluated on accuracy. For instance, every customer with negative accuracy of 90% is reached if the population comprises a positive to the negative class ratio of 1:9. Thereby, this quality indicator is straightforward. It is more surprising for exploring how mostly churners are accurately predicting and represent the true positive rate.

The accuracy over the cases is given by precision which is predicted to be positive. The accuracy measurement indicates how effectively the cases are predicted through the model, which is labelled as positive. The positive cases are referred to as churners in this study. The optimal threshold value is identified through Cohen's Kappa value. The correctly classified minority cases were favoured by the Kappa value over the majority of cases, which is further considered in this study. The optimal threshold value is detected through the largest Kappa value. The confusion matrix is used for the computation of sensitivity and precision that belongs to the optimal threshold value. The sensitivity and precision are computed for each model generated. Confusion matrices were used to calculate the receiver operating characteristic curve and the area under the Kappa curve on different threshold values. The ROC curve is used in a wide range of research areas for measuring model performance when a model has an AUC value. On the contrary, a major drawback of the AUC value is that it does not undertake the weight of a false positive and false negative prediction. This is an essential aspect since false-negative predictions are more potentially to take place as compared to the false-positive predictions with an uneven data set. Thereby, the Area under the Kappa curve is measured by undertaking the class skewness in the data. The x-axis refers to the representation of the rate of false positives and the y-axis refers to the Cohen's Kappa value for the Kappa curve. The Kappa value can be observed as a non-linear transformation between the true negative rate and the true positive rate.

Precision and sensitivity can further be interpreted that the models perform worst with the original churning rate and the performance parameters. The sensitivity level is also the highest for the best-performing model indicated by the AUC and AUK. On the contrary, the precision is higher for model 2 with 24 hidden neurons, which indicates the fact that fewer non-churning customers are classified as churning as compared to the

other models. In this regard, it is essential for predicting additional churning customers, which is the minority group. Thereby, it is concluded that model 3 with 24 hidden neurons performs effectively as compared to the models with one hidden layer. The findings indicated in Table 4 revealed that the training set distribution performs effectively for the models with two hidden layers.

The model with 12 hidden neurons resulted effectively in the highest AUK and AUC. The model with 12 neurons is the worst model predicting with two hidden layers and the training set with the original churners versus non-churners rate, which was further the case with one hidden layer. None of the models generated two hidden layers when the sensitivity and precision are checked collectively for performing effectively on both parameters. In this regard, the AUK value results in 12 hidden neurons with two hidden layers model.

Selection of the model

It was observed that the majority of the papers utilize the decision tree technique; however, neural networks were majorly used for predicting customer churn. It has been observed that neural networks perform effectively when undertaking all performance parameters. On the contrary, the decision tree technique performs effectively when only the AUK value was undertaken, but this technique has low sensitivity and AUC levels. This shows that it becomes essential for making an informed decision with multiple performance indicators. The maximum of the ROC curve and the Cohen's Kappa curve was the same based on the decision tree technique. This is not the case for support vector machines and neural networks and shows different corresponding threshold values. The point closest to the upper left corner cannot identify because of time issues and was not identified what the best threshold value would be if the ROC curve is utilized.

Due to visual inspection, it can merely be concluded that the optimal threshold value is sometimes equal between the two measurements and sometimes it is not. This stimulates the conclusion that it is essential for using multiple performance measures in the presence of an imbalanced data set. However, it was observed that the worse performing technique was the decision tree, and thus all cases were predicted with the same churn possibility with this technique. Table 5 displayed the bad performance parameters where no lift chart was represented because of the bad performance.

Neural networks and logistic regression perform effectively on the test set of 2019, as can be observed in the performance indicators of Table 5. The AUK value of the support vector machine is even higher as compared to the generated results with a

Table 4 Performance parameters, sensitivity, and precision of the decision trees

| Performance Parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|------------------------------------|---------|---------|---------|---------|---------|
| Number of hidden neurons set to 12 | | | | | |
| AUC | — 0.570 | — 0.749 | 0.885 | — 0.862 | 3.611 |
| AUK | — 0.755 | — 0.827 | 1.328 | — 1.685 | 1.089 |
| Number of hidden neurons set to 24 | | | | | |
| AUC | 0.257 | 0.499 | 0.704 | 0.148 | 1.587 |
| AUK | 0.751 | 1.183 | 1.292 | 0.759 | 0.973 |

Table 5 Performance parameters for the models tested

| Performance parameter | Logistic regression | Decision tree | Neural networks |
|-----------------------|---------------------|---------------|-----------------|
| AUC | 0.036 | 0 | 0.075 |
| AUK | 0.769 | 0.5 | 0.629 |
| Kappa | 0.291 | 0 | 0.181 |
| Threshold | 0.8 | 0.1 | 0.5 |
| Precision | 0.220 | – | 0.117 |
| Sensitivity | 0.103 | 1 | 0.591 |

Table 6 Cost–benefit analysis

| Performance parameter | Logistic regression E9 | Logistic regression E10 | Logistic regression M9 | Logistic regression M10 |
|-----------------------|------------------------|-------------------------|------------------------|-------------------------|
| AUC | 0.045 | 0.081 | 0.066 | – |
| AUK | 0.527 | 0.552 | 0.816 | – |
| Performance parameter | Neural Network E9 | Neural Network E10 | Neural Network M9 | Neural Network M10 |
| AUC | 0.061 | 0.043 | 0.069 | 0.071 |
| AUK | 0.215 | 0.531 | 0.926 | 0.052 |

test set generated with data from 2018. None of the techniques showed better outcomes when the values of Table 5 were compared with each other. When 20% of the population is contacted, almost 50% of the churners are reached with all techniques. These findings are comparable with the findings of the literature. It can be interpreted those neural networks and logistic regression are the most important techniques for predicting customer churn. Both techniques perform effectively on test data from 2018 and 2019, which shows that the generated models apply to data of several years.

Lastly, a cost–benefit analysis was conducted on test data from 2018 to explore which technique results in the highest advantage. The Cost–benefit analysis is implemented on the best models generated with neural networks and logistic regression (Tables 6 and 7). The training and test sets are created through the K-means clustering technique. The Euclidean distance and Manhattan distance are utilized since these settings indicated the best outcomes when the number of clusters is set to 9 and 10. These clusters comprise young customers, who consume little insurance as compared to average and pay the premium. The findings of Table 6 indicated that the Manhattan distance performs effectively for all models based on the performance parameters AUC and AUK. In particular, the model generated with $K=9$ and neural networks.

Only the fit clusters were selected in the churning profile from these nine clusters. This indicates that there were 2 clusters chosen, which result in 344 cases. Surprisingly, the logistic regression did not perform well for the clusters chosen with the Manhattan distance with 10 clusters the accurate reason for this is unapparent, but it is mostly possible that it did not run because of the inadequate information. The small sample size might be the reason for this. The best findings generated through the logistic regression with Euclidean distance are presented in Table 7. On

Table 7 Performance parameters, precision, and sensitivity of the decision tree models

| Performance parameter | Logistic regression E9 | Logistic regression E10 | Logistic regression M9 | Logistic regression M10 |
|-----------------------|------------------------|-------------------------|------------------------|-------------------------|
| Kappa | 0.257 | − 0.500 | − 0.705 | 0.148 |
| Threshold | − 0.198 | 0.751 | 1.183 | − 1.293 |
| Precision | − 0.938 | − 0.226 | 0.142 | − 2.526 |
| Sensitivity | 0.429 | 0.190 | − 0.285 | 0.809 |
| Performance parameter | Neural Network E9 | Neural Network E10 | Neural Network M9 | Neural Network M10 |
| Kappa | 0.157 | 0.600 | 0.805 | 0.048 |
| Threshold | 0.298 | 0.651 | 1.083 | 1.393 |
| Precision | 1.038 | 0.326 | 0.042 | 2.626 |
| Sensitivity | 0.329 | 0.090 | 0.385 | 0.709 |

the contrary, the differences are small. In addition, the clusters are selected with an average churning profile, which results in 150 cases.

A high accuracy level is generated through the techniques that are utilized for churn prediction. The churning customers are the minority group in the data set in this study, resulting in a prediction that none of the customers is going to churn, and an accuracy level should be 90% or greater. The training sets were corrected in this regard where five different churnings: non-churning distributions were utilized. Homogeneous profiles were identified through two profiling techniques including decision trees and K-means. This is conducted since the literature indicated that this results in better prediction estimates. Additional observations are predicted and cluster in homogenous groups in the profiles. The essential findings are revealed with the K-means clustering technique, which shows that four different types of profiles must be examined independently. The average profile of the population is represented in the first profile, non-churning customers are represented in the second and third profile, and a churning profile is indicated in the last profile. AUK, precision, AUC, and sensitivity were selected to compare the accuracy of the models. However, the best performance was shown by neural networks and regression analysis. Almost 50% of the churners are captured when 20% of the population was contacted.

The findings from this test set are relatable to the findings of the models tested on data from 2018. The models of 2018 revealed slightly better outcomes. The advantages of these techniques were computed with a cost–benefit analysis for the year 2018, resulting in the highest advantage for the neural network. The use of different training set distributions is effective for each technique. The 50:50 training set distribution resulted in effective outcomes when the logistic regression technique was used throughout this study. A 70:30 distribution worked effectively for the neural network technique. In this regard, it is concluded that each technique works effectively with a different training set distribution from this distribution.

Conclusion and recommendations

The data mining procedures can be very successful in extracting hidden information and get to know customer's information. This learning system can be used by the directors to process decision-making. Customer clustering and the analyzing conducts of each cluster can be an important step in actualizing customer relationship executives' framework. This study proposed a data mining method to help build a framework that contains collectives and classifications of churned customers. K-means algorithm was used for customer segmentation. The churn prediction was carried out by using customer attractiveness and churn behaviour. The suggested system can enable the health insurance companies to predict the likelihood of investigating past conduct, current, and future clients. With a better understanding of the churning features, health insurance managers can prevent churn by considering some strategies. These strategies include providing the facilities required, improving the quality of services and products, identifying the different needs of the consumers, and increasing customer responsiveness.

Acknowledgements

The author is very thankful to all the associated personnel in any reference that contributed in/for the purpose of this research.

Authors' contributions

AAJ is solely responsible for the conception, drafting, methods, analysis and correspondence of this research. All authors have read and approved the final manuscript.

Funding

This study is self-funded.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not required.

Competing interests

The author declares no competing interest.

Received: 12 February 2021 Accepted: 8 August 2021

Published online: 17 August 2021

References

1. Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Huang K. Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*. 2017;237:242–54. <https://doi.org/10.1016/j.neucom.2016.12.009>.
2. Amjad H, Reham D, Osama H, Ruba O, Hossam F. Hybrid data mining models for predicting customer churn. *Int J Commun Netw Syst Sci*. 2015;8:91–6.
3. Arowolo MO, Abdulsalam SO, Isiaka RM, Gbolagade KA. A hybrid dimensionality reduction model for classification of microarray dataset. *Int J Inf Technol Comput Sci*. 2017;9(11):57–63.
4. Arowolo MO, Adebisi MO, Adebisi AA, Okesola OJ. A hybrid heuristic dimensionality reduction method for classifying malaria vector gene expression data. *IEEE Access*. 2020;8:182422–30.
5. Arowolo MO, Adebisi M, Adebisi AA, Okesola JO. PCA model for RNA-Seq malaria vector data classification using KNN and decision tree algorithm. *Int Conf Math Comput Eng Comput Sci*. 2020. <https://doi.org/10.1109/ICMCECS47690.2020.240881>.
6. Arowolo MO, Isiaka RM, Abdulsalam SO, Saheed YK, Gbolagade KA. A comparative analysis of feature extraction methods for classifying colon cancer microarray data. *EAI Endorsed Trans Scalable Inf Syst*. 2017. <https://doi.org/10.4108/eai.25-9-2017.153147>.

7. Dolatabadi SH, Keynia F. Designing of customer and employee churn prediction model based on data mining method and neural predictor. In: 2017 2nd International Conference on Computer and Communication Systems (ICCCS). New York: IEEE; 2017. p. 74–7.
8. Fathian M, Hoseinpoor Y, Minaei-Bidgoli B. Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes*. 2016;45(5):732–43.
9. Figalst I, Elsner C, Bosch J, Olsson HH. Customer Churn Prediction in B2B Contexts. In: Hyrynsalmi S, Suoranta M, Nguyen-Duc A, Tyrväinen P, Abrahamsson P, editors. International Conference on Software Business. Cham: Springer; 2019. p. 378–86.
10. Franciska I, Swaminathan B (2017) Churn prediction analysis using various clustering algorithms in KNIME analytics platform. In: 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS). New York: IEEE; 2017. p. 166–70. <https://doi.org/10.1109/ssps.2017.8071585>.
11. He Y, Xiong Y, Tsai Y. Machine learning based approaches to predict customer churn for an insurance company. In: 2020 Systems and Information Engineering Design Symposium (SIEDS). New York: IEEE; 2020. p. 1–6. <https://doi.org/10.1109/sieds49339.2020.9106691>.
12. Keramati A, Ghaneai H, Mirmohammadi SM. Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innov*. 2016;2:2–13. <https://doi.org/10.1186/s40854-016-0029-6>.
13. Liu RQ, Lee YC, Mu HL. Customer classification and market basket analysis using K-means clustering and association rules: evidence from distribution big data of korean retailing company. *Knowl Manag Res*. 2018;19:59–76.
14. Martínez-Plumed F, Contreras-Ochando L, Ferri C, et al. CRISP-DM twenty years later: from data mining processes to data science trajectories. *IEEE Trans Knowl Data Eng*. 2019. <https://doi.org/10.1109/TKDE.2019.2962680>.
15. Nadali A, Kakhky EN, Nosratabadi HE. Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system. In: 2011 3rd International Conference on Electronics Computer Technology. New York: IEEE; 2011. p. 161–5.
16. Nadeem AN, Umar S, Shahzad MS. A review on customer churn prediction data mining modeling techniques. *Indian J Sci Technol*. 2018;11(27):1–7. <https://doi.org/10.17485/ijst/2018/v11i27/121478>.
17. Nagesh Singh Chauhan (Decision Tree Algorithm) Decision Tree Algorithm. 2021. <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>. Accessed April 2021.
18. Olaolu AM, Abdulsalam SO, Mope IR, Kazeem GA. A comparative analysis of feature selection and feature extraction models for classifying microarray dataset. *Comput Inf Syst J*. 2018;29:1.
19. Renjith S. B2C E-commerce customer churn management: churn detection using support vector machine and personalized retention using hybrid recommendations. *Int J Future Revolut Comput Sci Commun Eng*. 2017;3:34–9.
20. Rodpysh KV. Model to predict the behavior of customers churn at the industry. *Int J Comput Appl*. 2012. <https://doi.org/10.5120/7702-1059>.
21. Rohini M, Devaki P. Analysis of customer churn by big data clustering. *Int J Innov Res Comput Commun Eng*. 2017;5(3):7.
22. Santharam A, Krishnan SB. Survey on customer churn prediction techniques. *Int Res J Eng Technol*. 2018;5:131–7.
23. Scriney M, Nie D, Roantree M. Predicting customer churn for insurance data. In: Song M, Song Y, Kotsis G, Tjoa M, Khalil I, editors. Big data analytics and knowledge discovery. Cham: Springer; 2020. p. 256–65.
24. Singh SD. Consumer behaviour. Anmol. Mumbai: Himalaya publishing house; 2012.
25. Spiteri M. Customer churn prediction for an insurance company. Master's thesis, University of Malta. 2018.
26. Stucki O (2019) Predicting the Customer Churn with Machine Learning Methods: Case: Private Insurance Customer Data. Master's Thesis. LUT University. 2019.
27. Tamaddon A, Stakhovych S, Ewing M. The impact of personalized incentives on the profitability of customer retention campaigns. *J Mark Manag*. 2017;33:327–47. <https://doi.org/10.1080/0267257x.2017.1295094>.
28. Wenjie B, Meili C, Mengqi L, Guo L. A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn. *IEEE Trans Industr Inf*. 2016;12(3):1270–2181. <https://doi.org/10.1109/TII.2016.2547584>.
29. Zhang W, Mo T, Li W, Huang H, Tian X. The comparison of decision tree-based insurance churn prediction between spark ML and SPSS. In: 2016 9th International Conference on Service Science (ICSS). New York: IEEE; 2016. p. 134–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)