Journal of Business Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Customer churn prediction in telecommunication industry using data certainty

Adnan Amin[a,*], Feras Al-Obeidat[b], Babar Shah[b], Awais Adnan[a], Jonathan Loo[c], Sajid Anwar[a]

[a] Center for Excellence in Information Technology, Institute of Management Sciences, Peshawar 25000, Pakistan
[b] College of Technological Innovation, Zayed University, Abu Dhabi 144534, United Arab Emirates
[c] Computing and Communication Engineering, University of West London, United Kingdom

ARTICLE INFO

Keywords:
Churn prediction
Uncertain samples
Classification
Telecommunication
Customer churn

ABSTRACT

Customer Churn Prediction (CCP) is a challenging activity for decision makers and machine learning community because most of the time, churn and non-churn customers have resembling features. From different experiments on customer churn and related data, it can be seen that a classifier shows different accuracy levels for different zones of a dataset. In such situations, a correlation can easily be observed in the level of classifier's accuracy and certainty of its prediction. If a mechanism can be defined to estimate the classifier's certainty for different zones within the data, then the expected classifier's accuracy can be estimated even before the classification. In this paper, a novel CCP approach is presented based on the above concept of classifier's certainty estimation using distance factor. The dataset is grouped into different zones based on the distance factor which are then divided into two categories as; (i) data with high certainty, and (ii) data with low certainty, for predicting customers exhibiting Churn and Non-churn behavior. Using different state-of-the-art evaluation measures (e.g., accuracy, f-measure, precision and recall) on different publicly available the Telecommunication Industry (TCI) datasets show that (i) the distance factor is strongly co-related with the certainty of the classifier, and (ii) the classifier obtained high accuracy in the zone with greater distance factor's value (i.e., customer churn and non-churn with high certainty) than those placed in the zone with smaller distance factor's value (i.e., customer churn and non-churn with low certainty).

1. Introduction

The customers are considered one of the most important asset for a business in numerous dynamic and competitive companies within a marketplace (Coussement, Lessmann, & Verstraeten, 2017). In competitive market, companies in which the customers have numerous choice of service providers they can easily switch a service or even the provider. Such customers are referred to as churned customer (Óskarsdóttir et al., 2017). The causes of customer churn can be due to dissatisfaction, higher cost, low quality, lack of features, and privacy concerns (Sharma & Rajan, 2017). Many organizations e.g., financial service (Farquad, Ravi, & Raju, 2014; He, Shi, Wan, & Zhao, 2014; Lin, Tzeng, & Chin, 2011), airline ticketing services (Burges, Efimov, & Darrow, 1998), social network analysis (Maria, Verbeke, Sarraute, Baesens, & Vanthienen, 2016; Óskarsdóttir et al., 2017), online gaming (Suznjevic, Stupar, & Matijasevic, 2011), banking sector (Chitra & Subashini, 2011; Oyeniyi & Adeyemo, 2015; Sangar & Rastari, 2015), and telecommunication sector (Amin, Anwar et al., 2017; Mahajan, Richa, & Renuka, 2015; Amin, Shehzad, Khan, Ali, & Anwar, 2015; Amin, Faisal, Muhammad, & Sajid, 2015; Amin et al., 2016; Qureshi, Rehman, Qamar, Kamal, & Rehman, 2013), are ever more focusing on establishing and maintaining the long-term relationships with their existing customers (Amin, Al-Obeidat et al., 2017). Loyal customers can be considered long-term customers that are not only profitable for the company but also are great ambassadors in the market (Ganesh, Arnold, & Reynolds, 2000). One of the industry wherein this phenomenon is observed is the Telecommunication Industry (TCI). CCP in TCI is an increasingly well-known domain and popular research problem in the literature in recent years (Amin, Anwar et al., 2017; Garcia, Angela, & Vellido, 2017; Coussement et al., 2017; Katelaris & Themistocleous, 2017; Óskarsdóttir et al., 2017; Subramanya & Somani, 2017; Zhao, Gao, Dong, Dong, & Dong, 2017; Zhu, Baesens, & vanden Broucke, 2017). It is reported that TCI is suffering from the substantial problem of customer churn due to fierce competition, saturated markets, dynamic condition, and launching new attractive offers (Óskarsdóttir et al., 2017). It is observed that acquiring new customer can be more expensive for companies as compared to retention of the existing customer (Athanassopoulos, 2000). Also, the researchers have confirmed

* Corresponding author.
E-mail address: adnan.amin@imsciences.edu.pk (A. Amin).

https://doi.org/10.1016/j.jbusres.2018.03.003
Received 9 October 2017; Received in revised form 28 February 2018; Accepted 2 March 2018
0148-2963/

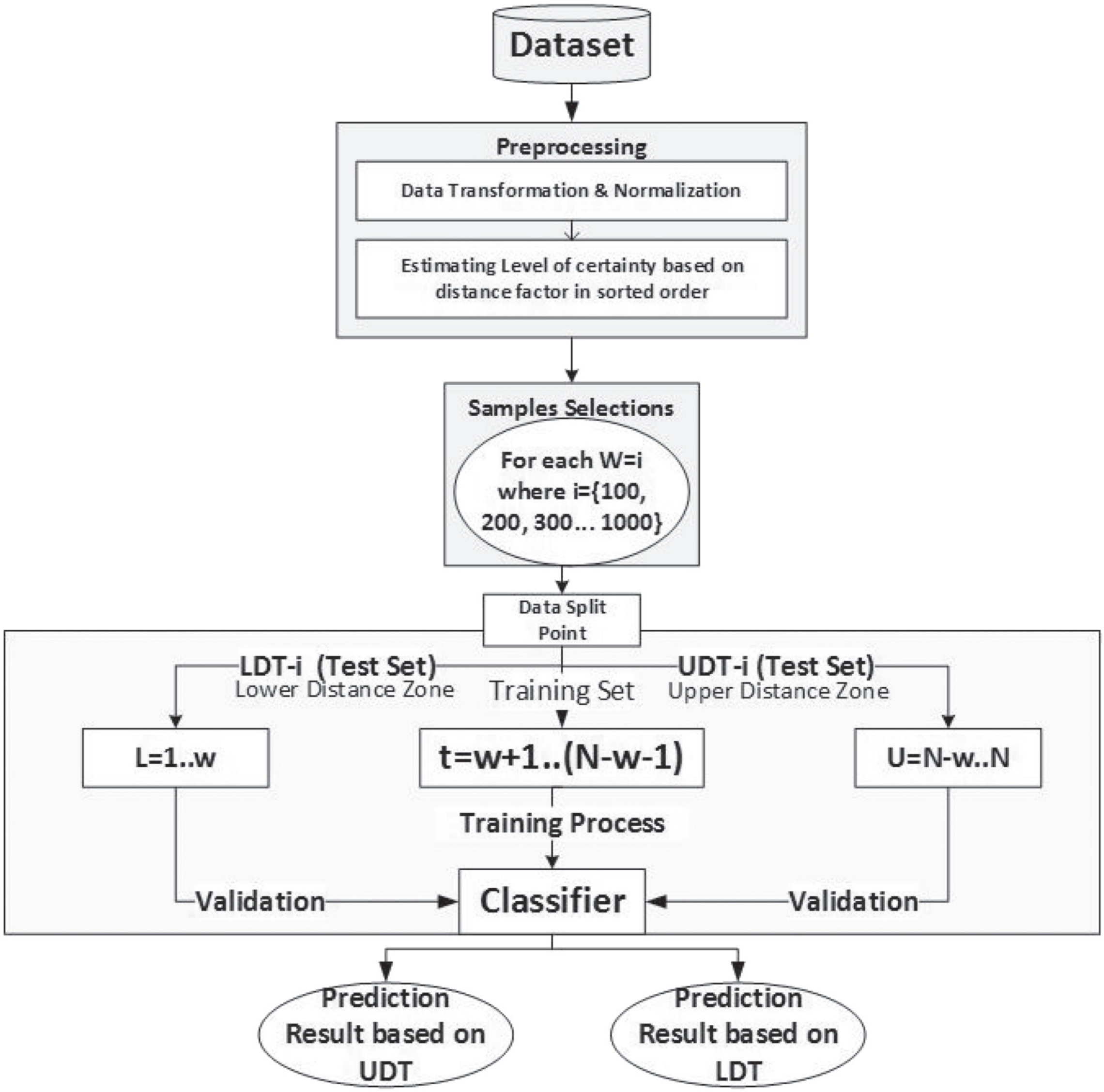


Fig. 1. Visualization of the benchmark framework. Where *L*,*U* are referring to lower and upper side of the distance based sorted samples, respectively. Where *W* is representing the window size (sample of test set i.e., 100, 200,…1000) and *N* is the total number of samples in the training set. Where UDT and LDT stands for Upper and Lower Distance Testset samples, respectively.

that CCP approaches can improve a company's revenue and good reputation in market (Maria et al., 2016). Currently, companies in the TCI have abundant information about their customers including local/international call records, short massages, voice mail, demographics, financial detail, and other usages behavior of the customers. This has created an opportunity for the machine learning (ML) community to develop predictive modeling techniques to handle the CCP in TCI. Therefore, a wide range of approaches based on ensemble techniques (Idris & Asifullah, 2014), probabilistic methods (Kirui, Hong, Cheruiyot, & Kirui, 2013), Support Vector Machine (SVM) (Farquad et al., 2014), K-nearest neighbor (KNN) (Ahmed & Maheswari, 2017), Rough Set Theory (RST) (Amin, Al-Obeidat et al., 2017), Fuzzy Logic Systems (Abbasimehr, 2011), Neural Networks (Kasiran, Ibrahim, Syahir, & Ribuan, 2014) etc., have been developed to identify customers with the highest tendency to churn. These CCP approaches help in deciding measures from the TCI data to prevent their customers from churning by offering them promotions and better deals (Jamil & Khan, 2016). However, these ML approaches lack the required effectiveness for the CCP model due to problem complexity (Idris, Khan, & Soo, 2013). Hence, there is greater challenge of which ML or data mining technique can be chosen for building CCP model. This is partly because most of the time, customer churn and non-churn have resembling features and behavior which increases the classification error rate. In other words, we can say that the classifier is uncertain about the decision and the level of certainty varies from case to case in the TCI. In this paper, we introduce a novel CCP approach using distance factor focusing on different distance zones (e.g., upper zone (greater distance factor's value) and lower zone (small distance factor's value)) pertaining to estimate the certainty of the classifier. Furthermore, the proposed CCP approach will not only predict the customer churns but can also calculate the level of the certainty of the prediction by evaluating the classifier's decision into the following categories, (i) customer churn and non-churn with high certainty, (ii) customer churn and non-churn with low certainty. The low certainty can be considered as uncertain classification for predicting the customer churns. The distance factors in term of upper and lower zones has not been considered for CCP in TCI yet. The proposed approach towards the target industry, exploring the discussed unexplored factors, can play a pivotal role in CCP models.

1.1. Paper organization

The rest of the paper is organized as follows: the next section presents literature review of CCP approaches and critical discussion on existing techniques; the propose methodology of this study is explored in Section 3. Section 4 presents the results, comparisons and discussion on our findings; the paper is concluded in Section 5.

2. Related work

The review in this section is primarily related to exploring the state-of-the-art techniques for CCP that have been adopted for CCP.

C.F. Tsai and Lu (2009) presented a hybrid neural networks approach for CCP in a CRM dataset of the American telecommunication company. They used an approach in which they have combined artificial neural network (ANN) and self-organized map (SOM) for CCP model. The ANN is used for data reduction in which unrepresentative data was filtered out from the training set. Then, the output of the first step is put into the SOM to build prediction model. The results indicate that combination of ANN + SOM outperform the single neural network with respect to accuracy. However, it can be observed that data reduction and filtering in first method (i.e., ANN) leads to loss of samples from the training set.

Wouter Verbeke, Martens, Mues, and Baesens (2011) explored the application of Ant-Miner+ which is based on Ant Colony Optimization. Ant-Miner + can allow to add domain knowledge through monotonicity constraints on the final decision rules. As a result, it produces highly accurate prediction model. Further, they have also incorporated the results to SVM, RIPPER, logistic regression and decision tree; wherein, combined with RIPPER reported the highest accuracy, while higher sensitivity obtained through decision tree. However, it is also investigated that Ant-Miner+ produces low sensitive decision rules and also require domain knowledge to be incorporated in the final rules-set for obtaining higher accuracy in CCP. On the other hand, RIPPER also produces good results in small rule-sets as well as it leads to unintuitive approach that violate the domain knowledge.

Benlan He et al. (2014) proposed CCP methodology based on the SVM classifier and random sampling technique. The random sampling method is used to change the data distribution in order to minimize the imbalance class problem which is caused due to the lack of availability of data. However, the class imbalance issue does not improve the predictive performance of their CCP model (Burez & Van den Poel, 2009). In connection to this, Burez and Van den Poel (2009) suggested to use weighted random forests method. However, random forests are often criticized for being very difficult to interpret and understand, particularly to identify reasons of customer churns which are important to explore for preventing the risky customers from churning (Richter & Slonim, 2010). Thus, it is not the appropriate methods to address CCP (Lu, Lin, Lu, & Zhang, 2014).

In order to improve the predictive performance of the CCP model, researchers have also proposed ensemble techniques. An ensemble method is the combination of several member of classifiers into one aggregated model. Koen W. De Bock and den Poel (2011) proposed ensemble technique based on rotation forest and Rotboost as two modeling methods for CCP. The rotation forests are used for features extraction while Rotboost method is applied in combination with rotation forest and AdaBoost to improve the performance of the CCP model using ensemble technique. The results shown that Rotboost outperform than rotation forests in term of accuracy while rotation forests obtained higher value of area under the curve (AUC) and lift measures. However, again there is problem of interpretability and understandability of the factors of customer churns. Therefore, they suggested another study (De Bock & Van den Poel, 2012) based on generalized additive models (GAM) concepts and incorporated this concept into ensemble classifier (e.g., Bagging and random subspace, and semi-parametric GAM). As a result, it obtained comparatively good predictive performance as compared to training classifier individually with logistic regression and GAM method.

Similarly, Pendharkar (2009) suggested to use genetic algorithm (GA) and neural network (NN) for CCP. Where the GA was used to search features space while NN applied to predict the customer churn. The GA based NN CCP model increase the prediction accuracy of the customer churn. Furthermore, they have compared the performance of the GA-based NN model with z-score model and evaluated using receiver operating characteristics (ROC) curve and lift measures. It was found that the GA-based NN model performed better than z-score statistical model. On the other hand, area under the ROC curve are often not recommended for profit maximization in the business (Maldonado, Flores, Verbraken, Baesens, & Weber, 2015).

Oyeniyi and Adeyemo (2015) presented a model for CCP to identify most likely customers exhibiting churn behavior. They developed a model based on simple k-means for clustering; and applied JRip for rules extractions. However, this model was limited to help in banking sector to recognize likely customer churns and hence provide modalities for customer retention. Muhammad et al. (2017) proposed a customer classification model based on rough set theory to efficiently classify the customer churn. They have claimed that rough set based classification model can outperform when compared to linear regression, J48, voted perception of neural network. On the other hand, Haenlein (2013) discussed the dynamics of social interactions for customer churn within the telecommunication networks. He associates the customer's churn behavior with incoming and outgoing calls. According to his study, a customer is more likely to churn if this customer has relationship with such customers who already churned from the network.

The researcher in the subject area has manifested several approaches as elaborated above. However, we cannot consider which prediction approach can be set as the standard approach to address CCP in more appropriate fashion. It remains an open challenge for research community of ML. Although the existing work (Brandusoiu & Toderean, 2013) argues that SVM is the best classification technique due to its ability to efficiently handle the arbitrary nonlinearities but M.A.H. Farquad et al. (2014) reported about SVM that it also generates black-box illusion which can be considered its main disadvantage. Also, in literature (He et al., 2014) researchers have argued that the random sampling method is good approach before applying the classification technique. Random sampling minimizes the imbalance data distribution which is caused due to unavailability of the target class data in the dataset (i.e., customer churn); however, it is also reported in another study (Burez & Van den Poel, 2009) that class imbalance does not significantly improve the performance of prediction model. On the other hand, Burez and Van den Poel (2009) also suggested to apply weighted random forests and report its obtained results, which showed significance, in CCP model. Their technique, however, has also been criticized for its complexity in understanding and interpretation (Richter & Slonim, 2010).

Apart from the above mentioned discussion and to the best of our knowledge, there is no study which has focused and considered the role

Table 1
Summary of datasets.

Description	Dataset-1	Dataset-2	Dataset-3	Dataset-4
No. of samples	3333	7043	18,000	100,000
No. of attributes	21	21	251	100
No. of class labels	2	2	2	2
Percentage of positive samples	14.49%	26.54%	12.44%	49.56%
Percentage of negative samples	85.51%	73.46%	87.55%	50.43%
Data sources	URL[1]	URL[2]	URL[3]	URL[4]

[1] http://www.sgi.com/tech/mlc/db/ (Last Access: November 30, 2017 02:00 PM).

[2] https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_Fn-UseC-Telco-Customer-Churn.xlsx (Last Access: November 30, 2017).

[3] http://pakdd.org/archive/pakdd2006/competition/overview.htm (Last Access: November 30, 2017).

[4] https://www.kaggle.com/abhinav89/telecom-customer/data (Last Access: December 13, 2017).

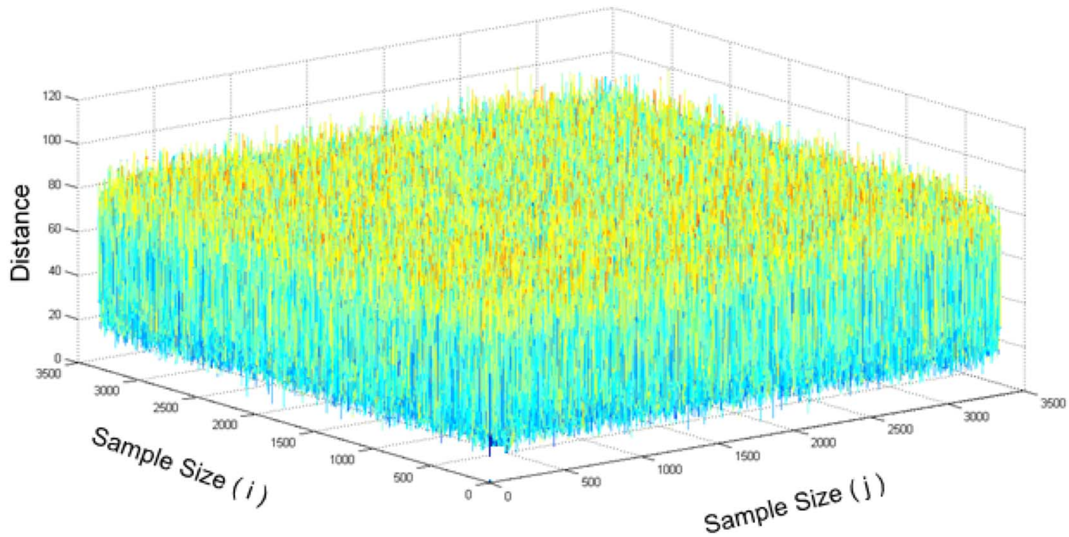


Fig. 2. Shows the distances between one sample against the rest of instances.

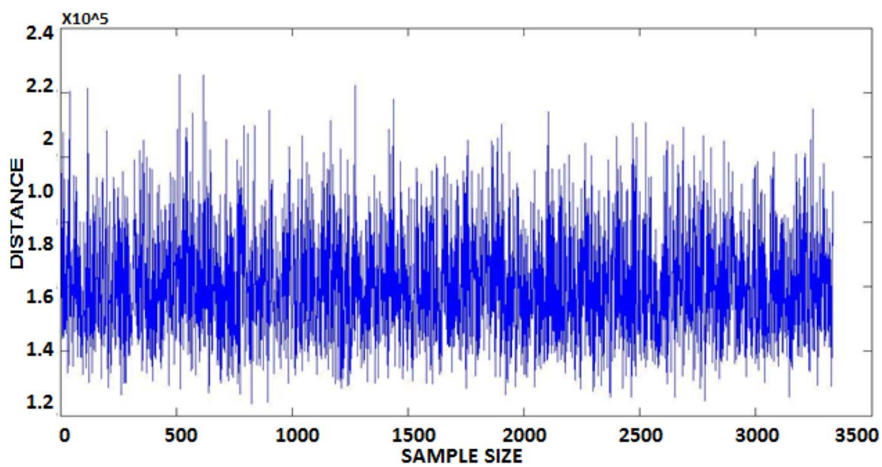


Fig. 3. Distances of each instance from all the instances, where x-axis represents sample size while the y-axis reflects the distance between samples.

of the distance factors in developing the CCP model for TCI. Therefore, this paper presents a novel CCP model based on distance factors to efficiently predict customer churns and also estimate the level of certainty of the classifier's decision in a given TCI dataset. The next section introduces the propose methodology and empirical setup of this study.

3. Methodology

In this section, we provide detailed descriptions of the proposed empirical study. Section 3.1 explains the problem statement. Sections 3.2 and 3.3 provide details of the empirical setup, and evaluation setup, respectively.

3.1. The problem statement

The CCP is binary classification problem where all the customers are divided into two possible behaviors: (i) Churn, and (ii) Non-Churn. Further, the churn behavior can be classified into the following subcategories: (a) voluntary customer churn, in which a customer decides to leave the service or even company, and (b) involuntary customer churn, in which the company or service provider decides to terminate a contract with the customer (Lu et al., 2014; Amin, Anwar et al., 2017). This study addresses the voluntary customer churns due to difficulty in predicting this type of customer churn while it is easier to filter out the involuntary customer churn by simple queries. On the other hand, the literature revealed that existing studies have been published but still there is no agreement on choosing the best approach to handle CCP problem. This might be because of the indifferent results shown by different classification techniques in different datasets. Therefore, there is dire need of assessing the validity of the classifier, in terms of its certainty or uncertainty, in predicting the customer churn. To the best of our knowledge, there is no state-of-the-art study which have considered the issue of checking the validity of the classifier. We propose to do that by considering distance factor using different distance zones within a dataset (i.e., Upper and Lower zones of the distance between the samples) for developing CCP model.

3.2. Empirical setup

We designed an empirical study to evaluate the proposed CCP model where we have focused on distance factor using different distance zones (i.e., Upper and Lower zones) in the given TCI datasets. Fig. 1 visualize the overall process of the framework.

For this study, we have selected arbitrary four publicly available datasets. The dataset-1 consists of 3333 samples and each sample represent individual customer; whereas, the ratio of churn and non-churn customers is 85.5% and 14.49%, respectively. Similarly, datasets-2, 3 and 4 contain 7043, 18,000 and 100,000 samples, respectively. Further, detail about these datasets is provided in Table 1.

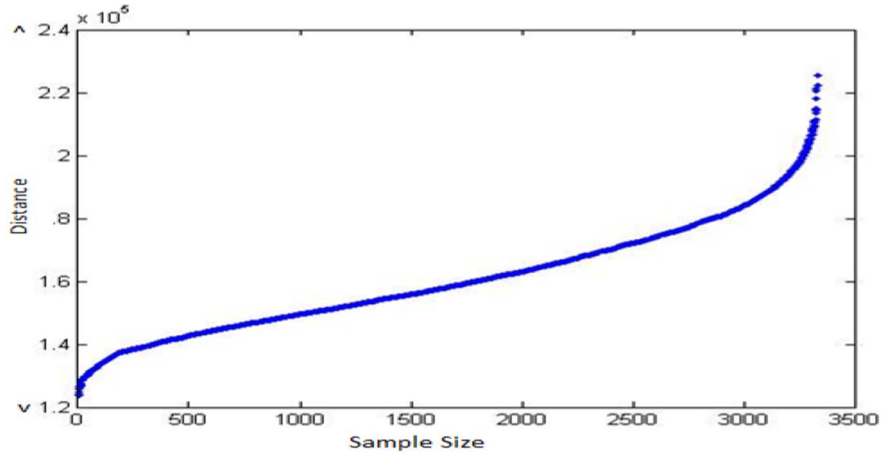


Fig. 4. Sorted sum of the distances of each instance against all instances, where x-axis represents the samples size and y-axis reflects the sum of the distance. The values of both axes are in increasing order. Where sign ∧ and v representing upper and lower zones of the distances, respectively.

3.2.1. Data preprocessing

In the preparation step, we have discretize by size, the values that exists in each attribute of the dataset, and then assigned certain labels e.g., Zero to Nine (0–9) possible values, to each discretized group. The discretizing by size leads to selecting the numerical attributes to nominal attributes and grouped them into specific size of bins. We then divide the total number of values in an attribute by size of bin. Ultimately, it produced specific list of values in different number of groups of an attribute. The step by step procedure for data preprocessing and discretization is elaborated as following:

1. Ignore the attributes consisting unique values which represents identity of the sample or descriptive text that serves for informational purposes and does not effect on the models training process.
2. Normalize the categorical values (such as 'yes' or 'no') into 0s and 1s where each value represents the corresponding category, then transformed the 0 and 1 into the same range and assign the same labels which applied for the rest of the attributes.
3. Find the distinct count of each value in every attribute, and also calculate the frequencies of these values in corresponding attributes.
4. Divide the range of values into 10 possible groups and assigned 0–9 label to each group in all the attributes.

3.3. Evaluation setup

In this section, a benchmarking framework is setup to present and evaluate the performance of the proposed study. These experiments were carried out using MATLAB toolkit to fulfill the objectives of the proposed study by addressing the following research questions:

- RQ1: Is it possible to derive a mechanism that can estimate the expected certainty level of the classifier results?
- RQ2: Can the level of certainty be estimated before the classification?
- RQ3: To what extent the estimated level of certainty is related to the accuracy obtained after classification?

To address the aforementioned important research questions (RQ1, RQ2 and RQ3), the following procedure is followed:

3.3.1. Distance factor

Suppose A is matrix with $m \times n$ elements.

$$A = \begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \vdots & \vdots \\ v_{m,1} & \cdots & v_{m,n} \end{pmatrix} \rightarrow \{(v_{i,j}) \in R^{m\times n}\}. \quad (1)$$

Arepresent the original dataset of size **mxn** where **m** is the number of samples and **n** is the number of attribute. $\mathbf{V_{i,j}}$ represent **j**th feature of **i**th case sample. The presentation of the dataset D in $R^{m\times n}$ form can be denoted as:

To find the different part in the dataset e.g., Upper and Lower distance zones, in the given dataset (i.e., D), first we calculated the distance of each instance from every other instance of the dataset using Manhattan distance formula. It can be generally expressed as: $P = [p_1,p_2,\ldots p_n]$ and $Q = [q_1,q_2,\ldots q_n]$.

$$D(p, q) = \sum_{i=1}^{i=n} |q_i - p_i| \quad (2)$$

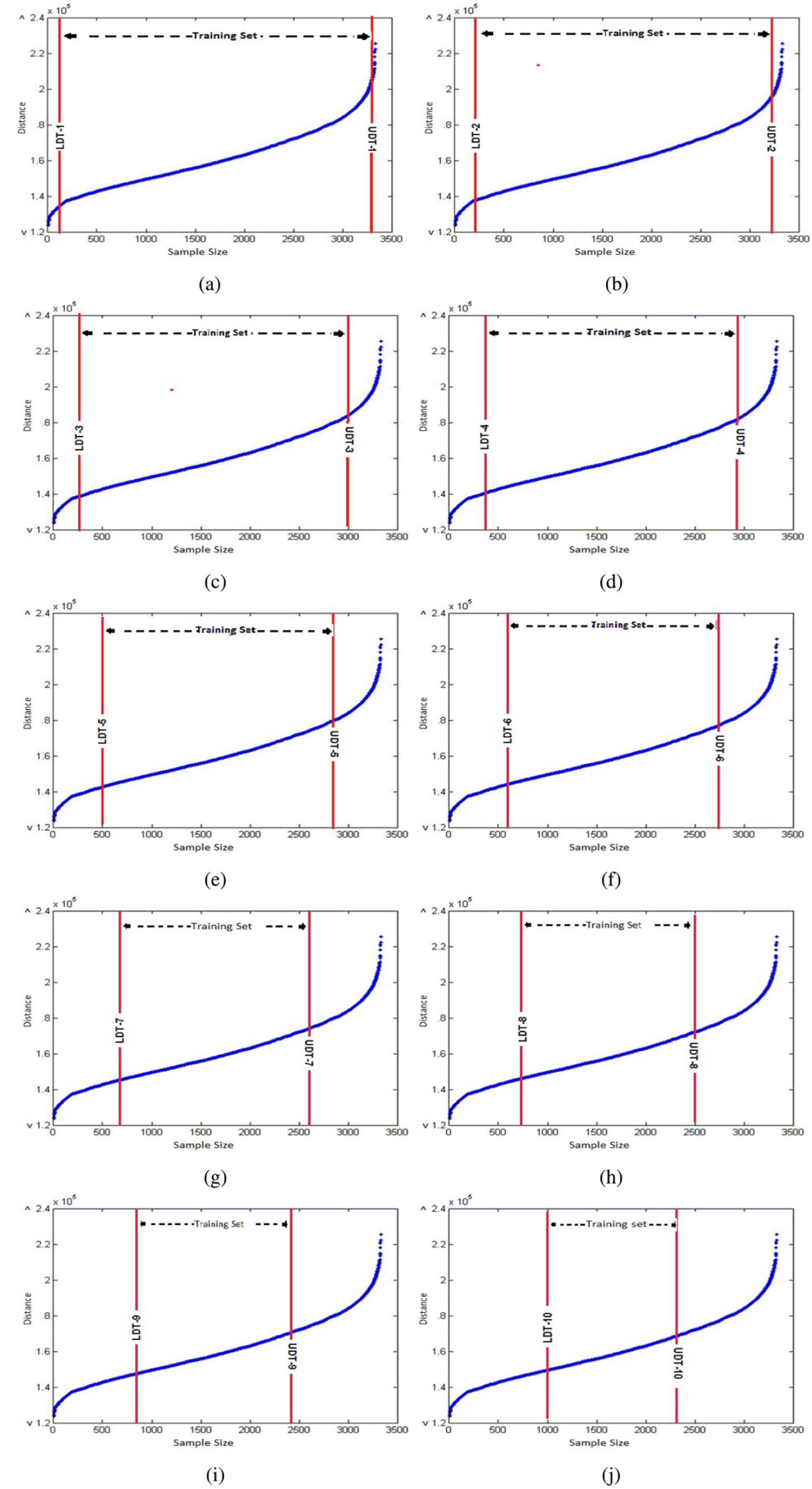
where i = 0, 1, 2, 3, …n and n is the total number of samples in given dataset. The Manhattan distance is the sum of absolute differences between points. Similarly, we have applied the Eq. (2) on the given dataset where p is considered one instance and q was considered as another instance from the same dataset. We calculated the distance $d_{i,j}$ between one instance (*e.g.*,S_i) with the rest of the instances and kept track of the distances. As a result, the following matrix is produced.

Fig. 2 illustrates the track of the obtained pair-wise distances of all samples. Wherein sample size of (i) and (j) are representing the instances of the same dataset while vertical dimension reflects the pair-wise (p_i and q_j) distances between all the samples of the dataset. The representation in Fig. 2 does not help in providing significant evidence of certainty and uncertainty based on distance factors. It is, therefore, difficult to identify the lower and upper zones of distances between the samples in a given dataset using the said figure. Therefore, we have calculated sum of the distance between individual instances and remaining samples (one vs all samples distances) using Eq. (3) in order to clearly investigate and visualize the upper and lower distance zones samples in the dataset depicted using Fig. 2.

$$DD_k = \sum_{i=1}^{n} \sum_{j=1}^{m} v_{i,j}. \quad (3)$$

All the calculated one vs all sample distances (d_k) for $k = 1,2,3\ldots m$ matrix DD is generated. Fig. 3 reflects the sum of the distances (one vs all sample distances) of each instance from all the instances. As we can see in Fig. 3 that the distances of each instance vary greatly from each other and can be viewed as three visible groups of unordered samples which are; (i) samples at lower distance side (lower zone), (ii) samples at upper distance side (upper zone), and (iii) samples that are in the middle (dark color) and can be seen as the major part of the samples.

Therefore, we have added distance vector (d_k) into the given dataset D (i.e., D = $(v_{i,j}) \in R^{m\times(n+1)}$) and then sorted all the instances based on the obtained distances d_k (One vs all sample distances) in ascending



(caption on next page)

Fig. 5. Visualization of the training and test samples. Where (a) to (j) represent the training and test sets on each iteration while LDT-i and UDT-i (i = 1, 2, 3,…10) represents the lower distance test set samples and upper distance test set samples, respectively.

Table 2
The performance CCP model based on LDT/UDT samples using Dataset 1.

Iteration sequence	Sample sizes	Precision %		Recall %		F-measure %		Accuracy %	
		UDT	LDT	UDT	LDT	UDT	LDT	UDT	LDT
LDT/UDT-1	100	52.38	33.33	64.71	35.29	57.89	34.29	84.00	77.00
LDT/UDT-2	200	48.72	30.56	61.29	42.31	54.29	35.48	84.00	80.00
LDT/UDT-3	300	50.00	30.19	59.09	43.24	54.17	35.56	85.33	80.67
LDT/UDT-4	400	52.78	36.00	58.46	49.09	55.47	41.54	84.75	81.00
LDT/UDT-5	500	52.81	34.00	60.26	50.00	56.29	40.48	85.40	80.00
LDT/UDT-6	600	51.38	37.93	58.95	52.38	54.90	44.00	84.67	81.33
LDT/UDT-7	700	52.94	37.04	63.16	53.19	57.60	43.67	84.86	81.57
LDT/UDT-8	800	53.46	37.58	63.43	56.19	58.02	45.04	84.63	82.00
LDT/UDT-9	900	51.76	36.57	61.54	56.14	56.23	44.29	84.78	82.11
LDT/UDT-10	1000	53.19	38.95	59.17	57.36	56.02	46.39	84.30	82.90

Table 3
The performance CCP model based on LDT/UDT samples using Dataset 2.

Iteration sequence	Sample sizes	Precision %		Recall %		F-measure %		Accuracy %	
		UDT	LDT	UDT	LDT	UDT	LDT	UDT	LDT
LDT/UDT-1	100	47.73	45.45	87.50	83.33	61.76	58.82	74.00	65.00
LDT/UDT-2	200	53.09	47.66	87.76	86.44	66.15	61.45	78.00	68.00
LDT/UDT-3	300	53.03	48.00	90.91	80.90	66.99	60.25	77.00	68.33
LDT/UDT-4	400	50.56	45.55	88.24	82.08	64.29	58.59	75.00	69.25
LDT/UDT-5	500	49.77	47.44	85.94	81.02	63.04	59.84	74.20	70.20
LDT/UDT-6	600	49.41	48.24	85.14	81.07	62.53	60.49	74.83	70.17
LDT/UDT-7	700	49.49	47.65	82.95	78.76	62.00	59.38	74.43	70.29
LDT/UDT-8	800	48.52	46.69	83.25	78.24	61.31	58.48	74.13	70.00
LDT/UDT-9	900	50.13	47.82	81.66	79.44	62.13	59.70	74.67	70.44
LDT/UDT-10	1000	50.48	47.38	82.42	80.37	62.61	59.62	74.80	70.60

Table 4
The performance CCP model based on LDT/UDT samples using Dataset 3.

Iteration sequence	Sample sizes	Precision %		Recall %		F-measure %		Accuracy %	
		UDT	LDT	UDT	LDT	UDT	LDT	UDT	LDT
LDT/UDT-1	100	43.75	21.52	06.03	72.65	10.61	33.20	88.20	65.80
LDT/UDT-2	200	46.67	19.72	06.67	72.16	11.67	30.97	88.22	65.33
LDT/UDT-3	300	42.86	19.56	06.45	73.81	11.21	30.92	88.13	65.38
LDT/UDT-4	400	33.32	19.06	05.00	71.62	08.70	30.11	88.00	64.86
LDT/UDT-5	500	27.27	20.66	04.55	73.53	07.79	32.26	88.17	65.00
LDT/UDT-6	600	12.50	22.63	01.92	72.88	03.33	34.54	88.40	67.40
LDT/UDT-7	700	16.67	21.15	00.22	70.21	03.92	32.51	87.75	65.75
LDT/UDT-8	800	25.00	22.03	03.03	78.97	05.41	34.44	88.33	67.00
LDT/UDT-9	900	25.00	22.78	05.00	81.82	08.33	35.64	89.00	67.50
LDT/UDT-10	1000	04.76	19.44	01.10	87.50	01.79	31.82	89.01	70.00

Table 5
The performance CCP model based on LDT/UDT samples using Dataset 4.

Iteration sequence	Sample sizes	Precision %		Recall %		F-measure %		Accuracy %	
		UDT	LDT	UDT	LDT	UDT	LDT	UDT	LDT
LDT/UDT-1	100	56.60	45.45	57.69	44.44	57.14	44.94	55.00	51.00
LDT/UDT-2	200	53.47	45.83	56.25	36.26	54.82	40.49	55.50	51.50
LDT/UDT-3	300	52.90	57.80	56.94	38.65	54.85	46.32	55.00	51.33
LDT/UDT-4	400	55.96	57.97	61.00	37.74	58.37	45.71	56.00	52.50
LDT/UDT-5	500	55.48	59.39	62.55	37.40	58.80	45.90	56.00	53.80
LDT/UDT-6	600	56.23	62.37	61.67	38.78	58.82	47.83	56.83	56.00
LDT/UDT-7	700	55.70	61.54	61.40	40.34	58.41	48.73	57.29	56.71
LDT/UDT-8	800	56.54	52.64	61.42	40.39	58.88	49.11	57.75	57.00
LDT/UDT-9	900	56.73	61.87	61.85	40.48	59.18	48.94	58.00	57.11
LDT/UDT-10	1000	54.14	60.00	61.30	40.96	57.50	48.69	56.80	57.00

Table 6
Overall performance of the base-classifier on 100 to 1000 samples of UDT and LDT.

Datasets	Test Sets	Accuracy on first iteration	Accuracy on last iteration	Difference
Dataset 1	LDT	77.00%	82.91%	5.91%
	UDT	84.00%	84.30%	0.30%
Dataset 2	LDT	65.00%	70.60%	5.60%
	UDT	74.00%	74.80%	0.80%
Dataset 3	LDT	65.80%	70.00%	4.20%
	UDT	88.20%	89.01%	0.81%
Dataset 4	LDT	50.00%	57.00%	7.00%
	UDT	55.00%	56.00%	1.00%

order, e.g., $d_1 \leq d_2 \leq d_3 \ldots \leq d_k$. Fig. 4 shows the sorted sum of distance of all instances from each other.

3.3.2. Selection of samples for building the CCP model

The splitting of dataset into training and test sets is a common process of ML for building predictive model (Witten, Frank, & Hall, 2011). The training set is usually used to train the model while test set is used in order to estimate how well the proposed model has been trained (performance evaluation of the model). In this study, we have introduced a novel procedure for training and validation process for finding the expected certainty (i.e., high and low) of the classifier's decision based on the distance factors as well as its impact on the classification performance. We assume the test set is new data where the value of the class label is obtained from the proposed predictive model. We then collected the predictions result from the trained classifier on the inputs from the test sets of both upper and lower zone samples and then compared them to obtain the empirical results of these test sets. This process allows us to evaluate the performance of the proposed model on the given test sets for both UDT and LDT highlighting the level of certainty of the classifier for both. For this purpose, initially, both zones sizes are set to 100 samples where first 100 and last 100 samples are selected for lower and upper zones of the distance between the samples, and then with each iteration the zone's size is increased by adding next 100 samples. As a result, on each iteration we obtained two prepared test set; (i) test set from lower zone of distance samples (i.e., LDT), and (ii) test set from upper zone of distance samples (i.e., UDT). While the rest of the samples are used as training set. Then we have used Naive Bayes (NB) as base-classifier which is based on the simplest statistical Bayesian theorem (Burez & Poel, 2012), studied since 1950. It assumes that the entire input variables are mutually correlated and contribute in the binary classification problems. The base-classifier is used to classify both non-churn and churn customers in the publicly available TCI dataset. For this purpose, then the base-classifier "Naive Bayes" is trained on the training sets to validate the performance of the classifier on both test sets (LDT-i and UDT-i, where i = 1) separately. After first iteration, 100 samples more increased in both sides test sets (LDT-i and UDT-i, where i = 2) and so on. The whole process of the training set and test set are visualized in Fig. 5.

4. Results and discussion

In this section, we have explored the results of the proposed empirical study and evaluated these results through the state-of-the-art evaluations measures (i.e., precision, recall, f-measure, and accuracy). The mathematical equations of these evaluation measure given in Eqs. (3), (4), (5), and (6) while further detail can be obtained from the studies referred as Amin, Anwar et al. (2017); Amin et al. (2016). The Eq. (3) mathematically expresses the precision measure, used to evaluate the correct degree of prediction power of the proposed model. On the other hand, Eq. (4) represents the recall measure, which is very important because prediction models intend to predict true churn customers as much as possible. However, there exists trade-off between

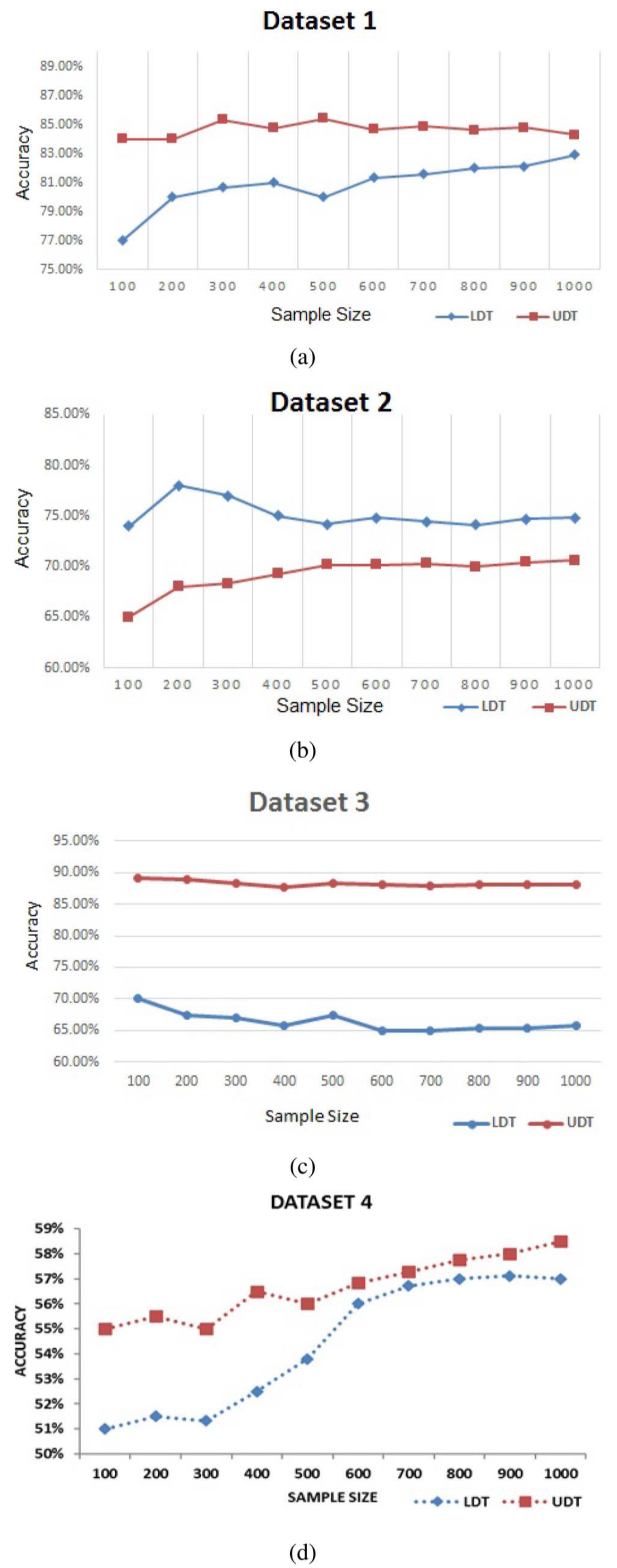


Fig. 6. The performance of CCP Models in term of accuracies on different size of UDT and LDT samples. Where (a), (b), and (c) represent the accuracies versus samples sizes of datasets 1, 2, 3 and 4 respectively.

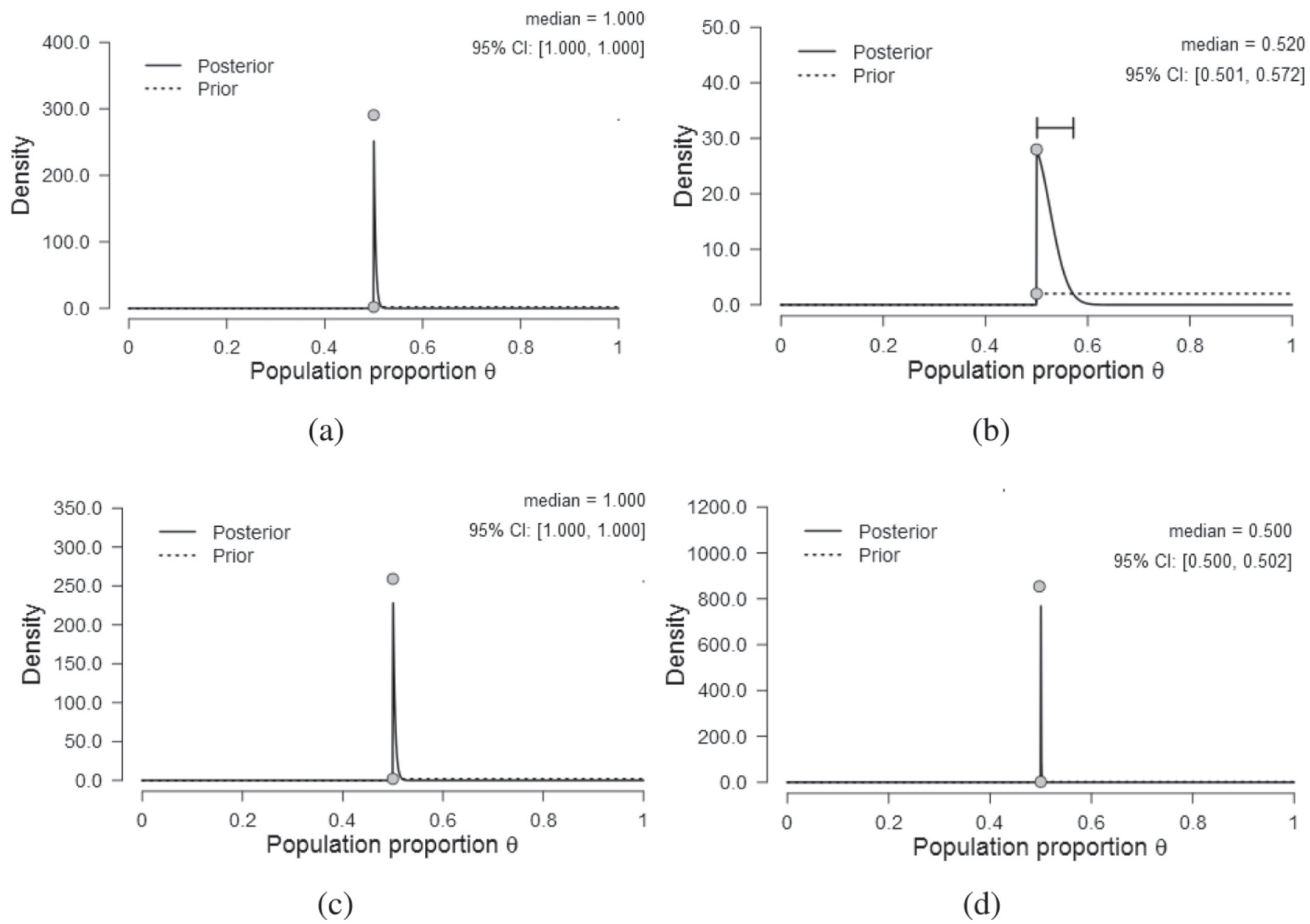


Fig. 7. Plots of prior and posterior analysis where (a), (b) and (c) represent the pictorial representation of obtained analysis after applying Bayesian binomial test on datasets-1, 2 and 3 respectively. The prior distribution is the dotted lines and the posterior distribution is the solid lines.

precision and recall. Therefore, a comprehensive measure is required for precision and recall. Here, we use Eq. (6) (F-measure) that calculates the harmonic mean of these two measures (e.g., precision and recall) resulting in achieving the balance between the said trade-off. Tables 2, 3, 4 and 5 show the CCP model performance on UDT and LDT samples on datasets 1, 2, 3 and 4, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{7}$$

Further, we have provided aggregate summary that obtained from the series of experiments, performed during the empirical evaluation setup. Table 6 reflects the overall performances of the base-classifier on four TCI datasets. The Table 6 consists of five columns; whereas, first column represents the datasets used, second column describe the labels for both test sets samples, while third and fourth columns refers to the performance of the base-classifiers. The fifth column “difference” reflects the obtained accuracy when the size of samples increases in both distance zones. This is on average 8.55% accuracy of proposed CCP model based UDT is higher than respective competitive LDT.

It is investigated that the proposed CCP model exhibit different performance on different distance zones (UDT and LDT) in datasets; whereas, in upper zone or UDT, the CCP model achieved overall best performance in term of correctly predicted true churn and non-churn customers as compared to lower zone or LDT. It is also observed from empirical results (i.e., given in Tables 2, 3, 4 and 5) that initially the classifier on UDT obtained higher performance in term of prediction of true customer churn and non-churn. But when the number of samples (e.g., customers) increases in both distance zones on each iteration, the classifier on LDT always obtained drastic improvement in prediction performance in term of accuracy as compared to previous classifier's results on LDT. While the classifier on UDT shown smooth and almost similar results, no big improvement in prediction performance in term of accuracy as compared to previous classifier's results on UDT. Which shows that the proposed model efficiently estimates the expected certainty of classifier decision in term of high and low certainty in the form of upper and lower distance zones of datasets, respectively which also report to RQ1.

Interestingly, the upper distance zone has not shown more effect on the performance of CCP model in TCI datasets because it has obtained the performance in term of differences in the accuracy is 0.30%, 0.80%, 0.81% and 1.00% in datasets 1, 2, 3, and 4, respectively. On the other hand, lower distance zone achieved dramatic changes in the performance when the zone size increases, it obtained differences in the accuracy such as 5.91%, 5.60%, 4.20% and 7.00% in datasets 1, 2, 3 and 4, respectively. Therefore, it is concluded that the upper zone is highly certain for classification because the classifier provided no big change in the results while the lower zone is highly uncertain for the classification due to drastic change in the classifier's results. Furthermore,

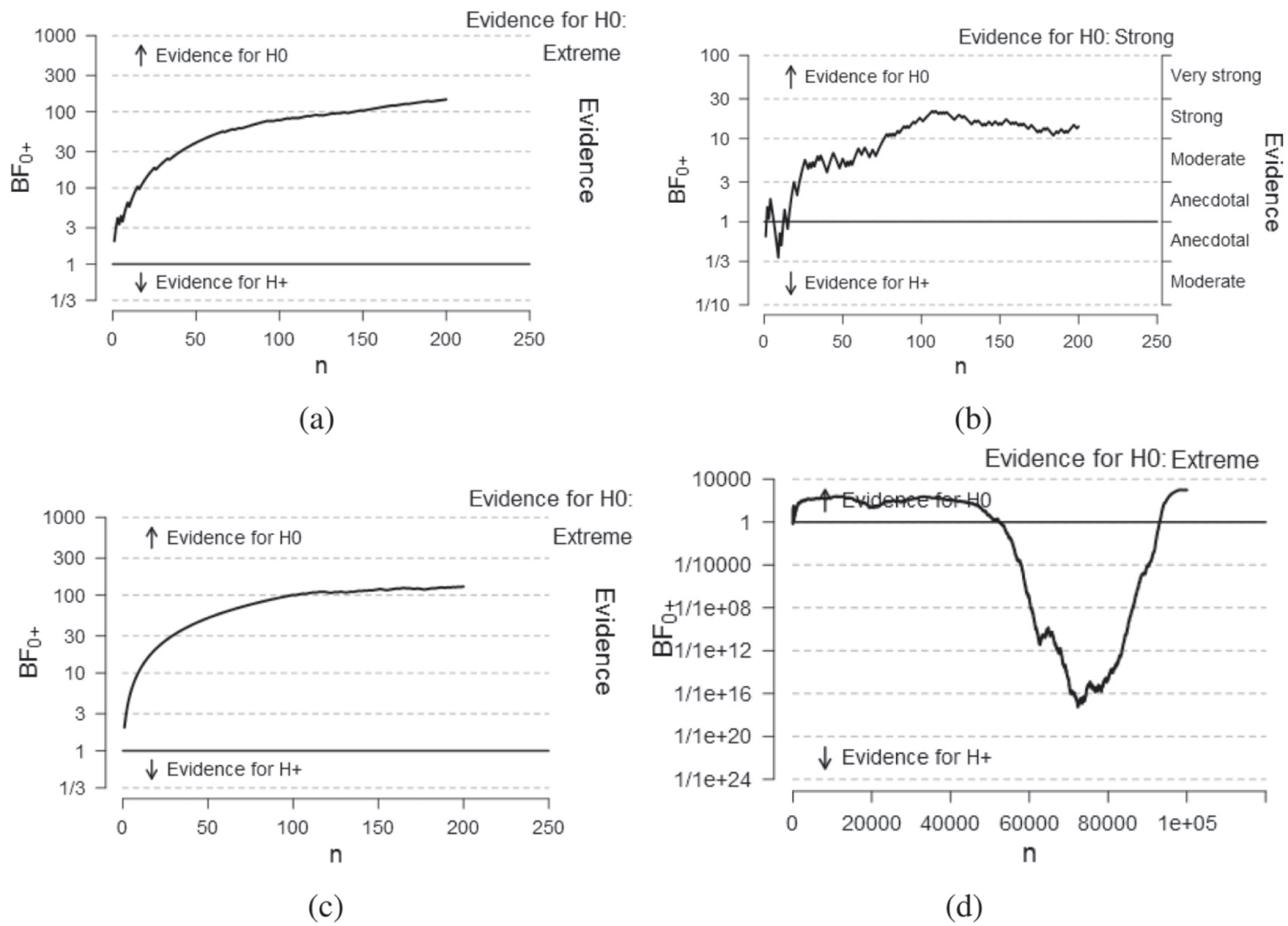


Fig. 8. The sequential analysis plot represents the performance on both UDT and LDT samples on 1st iteration of base-classifier. Where (a), (b) and (c) reflect the datasets 1, 2 and 3 respectively.

based on these level of certainty which we have identified based on different accuracy level for different parts (upper and lower distance zones in datasets), can help in to estimate the certainty of base-classifier before the classification process. These finding also addressed the RQ2. Finally, we have calculated the average accuracy of CCP model by considering different parts of the dataset that are grouped into upper and lower zones; whereas, upper zone on average accuracy 8.55% higher than the respective competitive lower zone. This shows that the proposed level of certainty is related to the expected level of accuracy on average 8.55% in the TCI datasets used, which also report to RQ3.

The proposed study further demonstrated the benefit of incorporating the distance factor and creating different data zones. Often, the researchers focused on a model which achieved higher accuracy rather than focusing on the level of certainty of the classifier predicting the Churn behavior. This study revealed that the decision maker should equally focus to propose level of certainty to effectively predict the customer churn and non-churn with high certainty as well as with low certainty. Customers identified by classifier with low certainty creates uncertain situation in the TCI because such customers may change their mind and may become churn from non-churn customer and vice versa. We can say that the performance of CCP model under uncertain situation in TCI is inversely proportional to accuracy as shown in Fig. 6.

4.1. Bayesian binomial test

To know whether our resulting classification model is correct or giving results purely by chance Bayesian Binomial Test (BBT) is used (Zhu & Lu, 2004). BBT is used to analyze all the empirical results of the proposed study performed on the publicly available TCI datasets. The BBT are performed on the results reported in Tables 2, 3, 4, 5 and the detailed statistical information provided in Table 1. These results are evaluating the performance of classifier, trained on different samples sizes in term of state-of-the-art evaluation measures (e.g., precision, recall, f-measure and accuracy) (Amin, Anwar et al., 2017; Amin et al., 2016). The Bayesian binomial test has only single parameter and an easy test to understand the distribution for the data. The Bayesian approach starts the estimation with prior distribution on the p parameter of interest, p could be a value between 0 and 1 with equal chance. Then, the posterior distribution of p is given through Bayes theorem given in Eq. (8) (Zhu & Lu, 2004).

$$\pi(P|x_1 \quad \dots \quad x_n) = \frac{f(x_1, \dots x_n|p)\pi(p)}{\int_0^1 f(x_1, \dots x_n|p)\pi(p)dp}. \tag{8}$$

The following configuration is set for applying the Bayesian binomial test; (i) first of all we set the test value to 0.5 which correspond to the traditional null hypothesis (H_0). In the proposed study, since there are two possible responses (i.e., churn and non-churn) and we want to know if the samples label is random guessing or is otherwise. Therefore, the test value is set to 0.5, then (ii) we have specified the alternative hypothesis (H_1) greater than test value because we are interested to know whether the predicted label values through base-classifier is performing better than the label predicted by chance, and (iii) finally, plots the prior and posterior analysis which is illustrated in Fig. 7.

In Fig. 7, it is observed that the most of the posterior distribution falls in between 0.4 and 0.5. It is also noticed that there are two small circles on the plots. These small circles represent the height of the curve at the test value. Where the first small circle on the prior distribution is lower than on the small circle of the posterior distribution. This mean the Bayes factor supports the H_0. If the small circle on the posterior distribution was lower than the small circle of prior distribution, then the alternative hypothesis would be supported. It is also noticed that 95% confidence interval (CI) is considered for this test.

Finally, Fig. 8 shows the sequential analysis plotting where we can see graphical representation of Bayes factor by following the CI = 95%. The x-axis is the samples and y-axis is the Bayes factor and through Fig. 8 it can be tracked the Bayes factor as it changes after every data point. If the Bayes factor is the above 1 represents the evidence in the favor of the H_0 and it is below 1 then it shows the evidence in favor of the alternative hypothesis. So it can be observed from Fig. 8 (a) to (d) that the obtained Bayes factor of above 100 showing the statistical evidence as extremely in the favor of H_0. On the other hand, Fig. 8 (b) obtained the Bayes factor as strong in favor of the H_0 and rejects the H_1.

4.2. Threats to validity

- Data preprocessing: applying different discretizing technique instead of the proposed steps for data preprocessing may leads to variance in evaluation and classification process which may also impact on the predictive performance. Additionally, we have used the original dataset without considering the class imbalance factor wherein class imbalance handling techniques may also lead to different results. Further, we have fixed the window size for training set to 50% for such cases when the number of test sets samples increases from the training set samples. If different size of training set window is fixed it may provide variance in the final results.
- Base-classifier: every classification algorithm has different mechanisms to classify the instances. We have arbitrarily chosen the Naive Bayes classification algorithm as base-classifier and self-implemented it in MATLAB for the propose study. All the experiments are carried out using the same base-classifier. However, applying different classification algorithm may produce different results.

5. Conclusion

With the terrific growth of digital data and associated technologies, there is an emerging trend, where industries become rapidly digitized. These technologies are providing great opportunities to identify and resolve diffuse problem of customer churn, particularly in TCI. Through a novel CCP approach, we have extracted insightful level of certainty of classifier decision based on the distance factor and also categories the customers into different customer groups based on lower zone and the upper zone of distance. Further, we empirically evaluated the impact of the level of certainty of classifier before the classification customers churn and non-churn. Moreover, we have also addressed (see Section 4) three important research questions (given in Section 3.3) with scientific evidence. Overall, the proposed study offers two main contributions to the existing literature such as: (i) introduced a novel approach for CCP in TCI based on distance factor, and (ii) revealed the effects of the distance factor in different distance zones (upper and lower zones) to estimate the expected certainty of the classifier decision. It is also investigated that distance factor is strongly co-related with the certainty of the classifier because the customers in lower zone shown uncertain behavior as compared to upper zone (certain behavior). Additionally, the performance of the resulting models was evaluated with four state-of-the-art evaluation measures (see Section 4) which gave consistent and robust results. A benchmarking of the propose model for CCP in TCI using uncertain samples on this scale, does not have precedence in the current studies to the best of our knowledge. While an overview of the empirical results can be seen in Tables 2, 3, 4 and 5 followed by Fig. 6. To show the effectiveness of our study in terms of precision, recall, accuracy, and f-measure. Future studies might be able to provide empirical results on the balanced dataset with multiple base-classifier. It would be interesting to see what will be the effect on the CCP model if we apply the feature selection method by assigning weights to the features. Another future direction can be to test more comprehensive study with other types of models would offer the possibility to compare our results and eventually help to evaluate this effect statistically. Furthermore, since our proposed model predicts level of certainty that leads to expected level of accuracy. This can be used to select good cases for training the classifier efficiently and more accurately. This can also be used to predict outliers in training data that can have negative effect on the classification. This technique can also be used on priority sampling. With minor modifications in this techniques it can be applied in social media for critical nodes identification.

References

Abbasimehr, H. (2011). A Neuro-Fuzzy Classifier for Customer Churn Prediction. *International Journal of Computer Applications, 19*(8), 35–41.

Ahmed, A. A., & Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly based classification. *Egyptian Informatics Journal, 18*(3), 215–220. http://www.sciencedirect.com/science/article/pii/S1110866517300403http://dx.doi.org/10.1016/j.eij.2017.02.002.

Amin, A., Al-Obeidat, F., Shah, B., Tae, M., Khan, C., Durrani, H., & Anwar, S. (2017). Just-in-time customer churn prediction in the telecommunication sector. *Journal of Supercomputing*. http://dx.doi.org/10.1007/s11227-017-2149-9.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing, 237*. http://dx.doi.org/10.1016/j.neucom.2016.12.009.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., ... Hussain, A. (2016). Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study. *IEEE Access, 4*. http://dx.doi.org/10.1109/ACCESS.2016.2619719.

Amin, A., Faisal, R., Muhammad, R., & Sajid, A. (2015). A Prudent Based Approach for Customer Churn Prediction. *11th international conference, BDAS 2015, Ustroń, Poland* (pp. 320–332). .

Amin, A., Shehzad, S., Khan, C., Ali, I., & Anwar, S. (2015). *Churn prediction in telecommunication industry using rough set approachVol.~572*.

Athanassopoulos, A. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research, 47*(3), 191–207.

Brandusoiu, I., & Toderean, G. (2013). Churn prediction in the telecommunications sector using support vector machines. *Annals Of The Oradea University Fascicle of Management and Technological Engineering*(1).

Burez, J., & Poel, D. V. D. (2012). Data mining concepts and techniques. *Pearson Education Asia Inc.*.

Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications, 36*(3), 4626–4636.

Burges, C. J., Efimov, D., & Darrow, R. (1998). Airline new customer tier level forecasting for real-time resource allocation of a miles program. *Data Mining and Knowledge Discovery, 2*(2), 121–167. http://link.springer.com/10.1023/A:1009715923555http://dx.doi.org/10.1023/A:1009715923555.

Chitra, K., & Subashini, B. (2011). Customer retention in banking sector using predictive data mining technique. *ICIT 2011 the 5th International Conference on Information Technology* (pp. 1–4). .

Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems, 95*, 27–36. http://dx.doi.org/10.1016/j.dss.2016.11.007.

C.Pendharkar, P.. (2009). Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Systems with Applications, 36*(3), 6714–6720. http://www.sciencedirect.com/science/article/pii/S0957417408005757http://dx.doi.org/10.1016/J.ESWA.2008.08.050.

De Bock, K. W., & den Poel, D. V. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications, 38*(10), 12293–12301. http://linkinghub.elsevier.com/retrieve/pii/S0957417411005239http://dx.doi.org/10.1016/j.eswa.2011.04.007.

De Bock, K. W., & Van den Poel, D. (2012). Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications, 39*(8), 6816–6826. http://linkinghub.elsevier.com/retrieve/pii/S0957417412000164http://dx.doi.org/10.1016/j.eswa.2012.01.014.

Farquad, M., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing, 19*, 31–40.

Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stayers. *Journal of Marketing, 64*(3), 65–87.

Garcia, D. L., Angela, N., & Vellido, A. (2017). Intelligent data analysis approaches to churn as a business problem: A survey. *Knowledge and Information Systems, 51*(3), 719–774. http://dx.doi.org/10.1007/s10115-016-0995-z.

Haenlein, M. (2013). Social interactions in customer churn decisions: The impact of relationship directionality. *International Journal of Research in Marketing, 30*(3), 236–248. http://www.sciencedirect.com/science/article/pii/S0167811613000402?via%3Dihubhttp://dx.doi.org/10.1016/J.IJRESMAR.2013.03.003.

He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of customer attrition of commercial banks based on SVM model. *Procedia computer science. Vol.~31. Procedia computer science* (pp. 423–430). Elsevier Masson SAS. http://dx.doi.org/10.1016/j.procs.2014.05.286.

Idris, A., & Asifullah, K. (2014). Ensemble based efficent churn prediction model for telecom. *12th International Conference on Frontiers of Information Technology (FIT),* (pp. 1–7). .

Idris, A., Khan, A., & Soo, Y. (2013). Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost based ensemble classification. *Springer Science Business Media New York* (pp. 659–672). .

Jamil, S., & Khan, A. (2016). Churn comprehension analysis for telecommunication industry using ALBA. *ICET 2016–2016 International Conference on Emerging Technologies* (pp. 1–5). . http://dx.doi.org/10.1109/ICET.2016.7813259.

Kasiran, Z., Ibrahim, Z., Syahir, M., & Ribuan, M. (2014). Customer churn prediction using recurrent neural network with reinforcement learning algorithm in mobile phone users. *International Journal of Intelligent Information Processing(IJIIP), 5*(March), 1–11.

Katelaris, L., & Themistocleous, M. (2017). Predicting customer churn: Customer behavior forecasting for subscription-based organizations. *European, Mediterranean, and Middle Eastern (EMCIS) on Information Systems, 299*, 128–135. http://dx.doi.org/10.1007/978-3-319-65930-5.

Kirui, C., Hong, L., Cheruiyot, W., & Kirui, H. (2013). Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. *IJCSI International Journal of Computer Science Issues, 10*(2), 165–172.

Lin, C.-S., Tzeng, G.-H., & Chin, Y.-C. (2011). Combined rough set theory and flow network graph to predict customer churn in credit card accounts. *Expert Systems with Applications, 38*(1), 8–15.

Lu, N., Lin, H., Lu, J., & Zhang, G. (2014). A customer churn prediction model in telecom industry using boosting. *Journal of Industrial Informatics, 10*(2), 1–7.

Mahajan, V., Richa, M., & Renuka, M. (2015). Review of data mining techniques for churn prediction in telecom. *Journal of Information and Organization Sciences (JIOS), 39*(2), 183–197.

Maldonado, S., Flores, Á., Verbraken, T., Baesens, B., & Weber, R. (2015). Profit-based feature selection using support vector machines — General framework and an application for customer retention. *Applied Soft Computing Journal, 35*, 740–748. http://dx.doi.org/10.1016/j.asoc.2015.05.058.

Maria, O., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2016). A comparative study of social network classifiers for predicting churn in the telecommunication industry. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1151–1158). .

Muhammad, M., Nafis, S., Mohammad, M., Awang, M., Rahman, M., & Deris, M. (2017). Churn classification model for local telecommunication company base don rough set theory. *Journal of Fundamental and Applied Sciences, 9*(68), 854–868.

Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2017). Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications, 85*, 204–220. http://dx.doi.org/10.1016/j.eswa.2017.05.028.

Oyeniyi, A. O., & Adeyemo, A. B. (2015). Customer churn analysis In banking sector using data mining techniques. *African Journal of Computing & ICT, 8*(3), 165–174.

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

Richter, Y., & Slonim, N. (2010). Predicting customer churn in mobile networks through analysis of social groups. *Proceedings of the 2010 SIAM International Conference on Data Mining* (pp. 732–741). .

Sangar, A. B., & Rastari, S. (2015). A model for increasing usability of mobile banking apps on smart phones. *Indian Journal of Science and Technology, 8*(November), 1–9. http://dx.doi.org/10.17485/ijst/2015/v8i30/8.

Sharma, R. R., & Rajan, S. (2017). Evaluating prediction of customer churn behavior based on artificial bee colony algorithm. *International Journal Of Engineering And Computer Science, 6*(1), 20017–20021. http://ijecs.in/issue/v6-i1/32ijecs.pdfhttp://dx.doi.org/10.18535/ijecs/v6i1.32.

Subramanya, K. B., & Somani, A. (2017). Enhanced feature mining and classifier models to predict customer churn for an E-retailer. *2017 7th International Conference on Cloud Computing, Data Science & Engineering — Confluence* (pp. 531–536). IEEE. http://ieeexplore.ieee.org/document/7943208/http://dx.doi.org/10.1109/CONFLUENCE.2017.7943208.

Suznjevic, M., Stupar, I., & Matijasevic, M. (2011). MMORPG Player behavior model based on player action categories. *Proceedings of the 10th Annual Workshop on Network and Systems Support for Games* (pp. 1–6). IEEE Press.

Tsai, C.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications, 36*(10), 12547–12553. http://linkinghub.elsevier.com/retrieve/pii/S0957417409004758http://dx.doi.org/10.1016/j.eswa.2009.05.032.

Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications, 38*(3), 2354–2364.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*http://books.google.com/books?id = bDtLM8CODsQC&pgis = 1.

Zhao, L., Gao, Q., Dong, X., Dong, A., & Dong, X. (2017). K-local maximum margin feature extraction algorithm for churn prediction in telecom. *Cluster Computing, 20*(2), 1401–1409. http://dx.doi.org/10.1007/s10586-017-0843-2.

Zhu, B., Baesens, B., & vanden Broucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences, 408*(June), 84–99. http://dx.doi.org/10.1016/j.ins.2017.04.015.

Zhu, M., & Lu, A. Y. (2004). The counter-intuitive non-informative prior for the Bernoulli family. *Journal of Statistics Education, 12*(2), 1–10. http://www.amstat.org/publications/jse/v12n2/zhu.pdf.