Exhaustive Measurement on Patients Portable Churn Dissection

Swaminathan B.[1*], Karthikeyan U.[2*], Kumar P.[3*], Franciska I.[4*]

[1, 2, 3]Rajalakshmi Engineering College, Thandalam, Chennai, India
[4]Karaikudi Institute of Technology and Karaikudi Institute of Management, Karaikudi, India
[1]swamikb@gmail.com, [2]kathikeyan.u@rajalakshmi.edu.in,
[3]kumar@rajalakshmi.edu.in, [4]franciska217@gmail.com

ABSTRACT

Patient portable churn dissection and prediction plays an important role to decide a suitable hospital for predicted diseases, patient's age group and health sustainability. Existing patients are grouped together based on various attributes such as admission details, cost for treatment, type of treatment, feedback given by the hospital and the patient for analysis and prediction. The prediction is based on identification of patients who are about to churn that can be analysed using Clustering algorithm. Healthcare centre will improve the quality of service and treatment methods which are based on prediction obtained from the clustering by using existing patient history. This will give confidence to the patient. As Healthcare centre may not know the exact reasons for patient churn, the patient retention is an important factor that boosts engagement, revenue and thereby decreases the churn rate. It will increase reputation of the hospital and surveillance of the patient. Patient satisfaction surveys are useful to listen patient's needs, provide regular meetings and check-ups to improve patient engagement with hospital that would certainly reduce churn rate.

Keywords
Churn Analysis, Prediction, Patient, Healthcare,

Introduction

The customer churn is a big deal in business enterprise sectors. The traditional solution is to predict high-tendency churners and address their needs via a custodian service, advertising campaigns or by applying special consideration. These approaches can vary from company to company or industry to industry and even from different sector from same organizations/ industry. Telecommunications, banking, insurance and healthcare are few examples. Most preventive measures are based on marketing and regular communication with their customers through e-mails, Short Message Service and voice synthesis calls. Customer Relationship Management (CRM) system acquires and retains customer relationships for an organizational growth. Relationship between duty staff and patient is an important factor in treating patients in healthcare centres.

Patient Relationship Management (PRM)

In [18, 24], analysis were based on PRM system also called CRM, in a public hospital based on patient satisfaction. PRM system would maintain patient details such as full body check-up, diet, exercise, genetic problem and allergy information. Using this information, health care centres can communicate through e-mail information or short message service to the patients about the latest available health care studies and treatments, which are suitable to the patients. Health care centres could suggest latest treatment plans that are relevant or suitable to patient profile. In [22], PRM system is connected with the subsequent systems.

Hospital Information System (HIS) – [20] make use of PRM to withstand relationship between patient and hospitals. It is the modern tool to overcome major challenges in hospitals by achieving operational efficiencies, adopting EMR technologies, delivering superior healthcare to the patients.

Electronic Health Record (EHR) – [23] reveal that EHR is the digital form of patient record and plays a vital role in prediction churn analysis. This contains medical history of the patient, diagnosis, treatment plans, test results, immunization dates. The data can be accessed through proper tools to make decisions on patient care and can stream line work flow of the organization.

Laboratory Information System (LIS) – [26] provides information about laboratory workflow systems. LIS maintains appointed laboratory tests, available results and the option that are prescribed by Laboratory Analysts. LIS and PRM systems are not integrated.

Attentive services – Attentive services are provided by PRM for sending reminder messages to the patients about appointments and date of check-ups. In [25] use reminder messaging systems to manage patient engagement with healthcare centre.

Level of Strategies

[9] analyse and review some of the key strategies that were identified to overcome barriers. Two major divisions of strategies are Organizational-level strategies and System-level strategies.

Organization-Level Strategies   This strategy focuses on patient cantered care at the organization level. Its components are mission of the centre, goals and autonomy, team work (that improves communication and coordination of care towards patients), work flows (that include Electronic Health Record, scheduling and timeframes), resources (time, money, space, work force).
System-Level Strategies
This centres on the frame work of hospital or health care centre which include policies and guidelines of the centre, Quality, certification, payment model, Legislation, education and licensing, culture of health care delivery.

Customer Churn

This is a measurement of how many patients are continued to approach the health care centre and how many were stopped to approach the health care centre. Hospitals are using many churn prediction models to maximize the profit. Cost of acquiring a new patient is more when compared to the cost of retaining the existing patient. There are two types of churn namely voluntary churn and involuntary churn. If the patient moves from one hospital to other voluntarily then it is termed as voluntary churn. If the patient moves from one hospital to other due to some factors namely death, migration then it is termed as an involuntary churn.

Churn prediction detects users who are likely to avoid a subscription to a service or better quality. Big data, clustering techniques are used in business for predicting the churn customers.
Simple attrition rate formula is the ratio between the numbers of customers who are quit from that service provider by the end of period total number of customers at starting of the same period in the service period.
Churn rate is the percentage of people/patients who discontinue their link/contact with the health care centre within a given time period.
Churn management   is the art of identifying the valuable customers who are likely to churn from a hospital and execute proactive steps to retain them.
Retention rate is the percentage of patients / people, a health care centre retains its contact over a given period of time. [12], measured patient churn based on organizational unit instead of an individual patient. Identification of specific patients that are likely to churn is analysed. Online Communities of Practice (CoPs) [13] are web based critical factors that are considered to identify the success in healthcare communities practice.

Related Work

Customer churn prediction is an important factor that can be discussed on different sectors such as Banking, Telecommunication, Internet access providers and Insurance companies. Predictive models are developed from artificial intelligence and further deployed to predict new inputs. Churn prediction models are developed based on service, customer relationship, profits and marketing strategies. Predictive modelling is concerned with forecasting on how the customer will behave in the future by analysing their past behaviour. Churn prediction is useful in

healthcare organizations to predict patients those who are willing to churn from healthcare centre for providing a better care to satisfy them.

Telecommunication Industry

In [8], the concept of probability is applied for customer churn prediction model. Data include calls to/from fixed line calls, International calls, On-net calls, VAS calls, Activity, Distinct callers and Contract-related features. In [15], data mining techniques were used to identify the customers who quit the service and rapid action has been taken to overcome the problem. [27, 3] have revealed the customer churn prediction in telecommunication where data is extracted from Business Support System (BSS). Data include call details, bill details, profile of customer and packages.

Banking System

In [2], analysis was based on banking sector in which one could always try to track customer communication with bank. Most of the studies that were reviewed about banking system was that bank customer would always try to communicate with bank or had relationship with the bank. [11], have considered 5000 clients of different ages for churn prediction in retail banking.

In [19], analysis was based on customer churn in bank customer churn prediction by using data mining techniques. Iranian Bank-e Tejarat in Tehran, customers are considered for churn analysis. Analysis further uses the questionnaires to get more information about the customer for churn prediction.

Insurance Companies

In [21], Iran Insurance Policy Company is considered for customer churn prediction. Insurance customer churn prediction models are created by using data mining concepts that have gender, age, selected policy, occupation, income, tenure, automatic renewal status, and credit score as attributes.

Analysis based on customer retention and prediction in health insurance industry which faces more competition than other industries. Costs of acquiring a new customer are at least 6-10 times more than the retaining an existing one and acquisition costs are more considerable portion of insurance companies. The competition provides customers with more choices, gaining cost has increased and customer faithfulness have decreased. The customer retention becomes one of the key aspects in the business growth for sustainability.

Internet Service Provider

[1] have revealed that life without Internet is difficult. Hence its fame is growing at impressive rate. ISP companies thus provide internet access services in various forms or packages. It is very difficult to add more and more customers to their service. At one point of time, every customer will decide an ISP of their own choices. It is more expensive in business when attempting to bring a new customer than to retain existing customers. Much more research has been conducted to maintain or discover new ways of identifying those customers who have a high risk of churning.

Literature Review

Various clustering algorithms are used for churn prediction and to discover the group of customers who are likely to churn. Clustering algorithms are used to group the similar characteristic objects into clusters for analysis, and prediction is useful to provide suggestions to improve the performance based on services. [7], focused on real world data are continuous in nature, having the following characteristics namely noisy and inconsistent due to human or computer error at the time of data entry.

Data Mining

Process of extracting useful knowledge from database, data warehouse and other repositories is called as data mining. It is a computerized process used to discover novel, suitable, helpful and potentially interesting knowledge from large data sources.
Process involved in Data mining:

- Extract, convert, and load data into the data warehouse system.
- Multidimensional database system stores and manages data.
- Data access is provided to analysts and IT professionals.
- Different application software is used to analyze the data.
- Result of maimed data in the form of graph or table.

Different types of data analysis are genetic algorithms, decision trees, nearest neighbor method, rule induction, data visualization.

Machine Learning Algorithm

Supervised and unsupervised learning algorithms are the classification of Machine Learning algorithm such as Decision tree, Naive Bayes, k-means, KNN, Linear regression, Logistic regression. These algorithms would deal with representation and generalization of data. Representation of data instances can be evaluated through functions.

Supervised Learning Algorithm/ Predictive models

In Supervised learning algorithm, data objects are classified into groups based on class label which is known already by the process of data training. Classification algorithms come under supervised learning algorithm and that assigns items into groups that are target categories or classes as low, medium or high credit risks. The purpose of classification is to perfectly predict the target class for each instance in the data.

Classification Regression Tree

CART is a type of decision tree model which involves both classification and regression. [5, 6] This processes specific type of attributes and ranks them using regression. The attribute which have a best score is chosen. A series of nested and pruning trees are optimized to give a better result [14, 21].

Improved Balanced Random Forests

This method involves selection of characteristics randomly and creates a classifier with a bootstrapped training data. A large number of trees or classifiers are assigned an unknown pixel to the class. Performance of the random forest classifier is better when compared with other classifiers.

Naive Bayesian Classification

In [13], Bayes theorem is applied for classification with minimum error rate when compared to other classifiers. It provides a theoretical justification for other classifiers that do not explicitly use 'Bayes' theorem.

Unsupervised Learning Algorithm/Descriptive models

Unsupervised learning algorithms have the process of grouping objects based on similar characteristics and each cluster does not posse the label.

K-Means and Canopy Clustering

In [21], Reza and Keyvan (2012) K- Means clustering algorithm have applied for grouping the objects that have

similar characteristics. Then the clusters are analyzed for identifying churners using RIIPER (JRip) [2] algorithm. In[4], k-means and Canopy clustering were compared and it has inferred that canopy would provide best cluster centroids.

Hybrid Clustering Algorithms

[10] proposed a hybrid clustering models, by combining two phases namely clustering phase and prediction phase and have been performed as a single model. Filters the data by using K-means clustering algorithm and prediction is by Multilayer Perceptron Artificial Neural Networks (MLP-ANN). In this model, the result is accurate and effective for churn prediction. clusiVAT technique determines the clusters in the given data set in [9].

Fuzzy C-Means Clustering

[11] Analyzed customer churn in retail banking by using fuzzy c-means method. This method has given a better performance than the traditional clustering algorithms. This method is best suitable for retrieving the hidden data and overlapped data. But this method requires an additional preprocessing steps and computation time would be more.

[16] Analyzed large volume of diverse social media data requires the computation of clustering mechanism of fuzzy adaptive resonance theory (Fuzzy ART),with complement coding and Data clustering was done through the vigilance parameter. Fuzzy Clustering in [17] was incomplete, missing and, attributes interval size variable are suitable.

Reinforcement Learning (RL)

Reinforcement learning is a machine learning algorithm where the machine is trained to take specific decisions with less involvement of human expertise. This software agent maximize the efficiency by applying any game theory, operations research, information theory, simulation based optimization, multi agent systems, statistics and genetic algorithms.

In data analysis process, machine learning algorithms plays a major role which have some steps to process data and building a model for prediction.

1. Collecting data: Raw data are gathered and kept in to excel forms, access and text files methods.

2. Preparing data: Quality of data is determined by considering missing data, outliers. Pre-processing step is used to preparing data by removing noise and unrelated data. Integration step is involved to group different individual data files into one data file and that can be stored in single database.

3. Training model: Appropriate training algorithms are selected, based on dataset to train the Training model. Cleaned data can be split into (i) Training Data which is used for deployment and (ii) Testing Data is used as a reference for a trained model. Techniques have been adopted for classification and predictions are decision tree induction, support vector machine, neural network, and Bayesian networks.

4. Evaluating model: To test the accuracy of a model, the model can be deployed using training data to test its accuracy. By the choice of suitable algorithm, the outcome will be more precise.

5. Improving the performance: To improve the performance, different models are grouped together and introducing more arguments. Spend additional amount of time for data collection and preparation provides better results.

Data Collection

Process of collecting relevant details about the patient in churn prediction analysis, is useful information. This information is called attributes. The various attributes are grouped to form data sets which are closely related with the relevant task.

Attributes

In the process of patient churn prediction analysis, considered data set is used to analyze and provide knowledge using pattern recognition. Every attribute has useful information for analyzing patient churn by using machine learning algorithms which may be k-means, decision tree and naive Bayes algorithm. Patient churn Data set contains 7044 instances and 10 attributes which include Patient ID, Gender, Age, and Date of Admission. It includes the patient's admitted date for the references.

Date of Discharge: Patient discharge date also important attribute in the patient churn dataset.
Treatment: Type of the treatment is represented by the duration which is further represented by character.
Glucose: At the time of admission or during treatment, patient needs to have glucose or not.
Payment: Mode of payment such as online credit card, cash, bank transfer or cheque.
Total charges: Total spend floating point values,
Feedback: Finally the discharging period feedback is collected from the patients.
Churn: One of the important attribute in the data set, it is represented for future references by character data type (YES or NO).

In data set, the important attributes are gender, age group, total charge, feedback and churn. Patient ID is a unique identity for every patient, who will be provided at the time of admission and other details about age, gender, and address are collected at the same time. Treatment type, payment details, feedback and churn details are gathered after admitting patients. Churn information is important attribute for building a churn prediction model where new patients are tested by this developed model.

Volume of Data

As many companies refuse to give their filtered data due to security reasons and at the same time gathered data from branches were not useful enough. Around 7044 instances and 13 attributes of data set are generated for the analysis. Important attributes are considered for the churn prediction which may be in character, integer and floating values. Data set is in the form of CSV for KNIME Analytics platform and arff form data set is used for java in eclipse. Missing values are removed during the conversion of arff form from CSV. Preprocessing steps arealso included to normalize the data set. Table 1 represents the view of sample dataset which is considered for churn prediction.

Table 1. Training Data

CustomerID	Gender	Age	Phone Service	Treatment	Monthly Charges	Total Charges	Feed back	Churn
7590-VHVEG	Female	15	No	M-to-M	29.85	29.85	Good	No
5575-GNVDE	Male	34	Yes	One year	56.95	1889.5	Good	No
3668-QPYBK	Male	2	Yes	M-to-M	53.85	108.15	Bad	Yes
7795-CFOCW	Male	45	No	One year	42.3	1840.75	Good	No
9237-HQITU	Female	2	Yes	M-to-M	70.7	151.65	Bad	Yes
9305-CDSKC	Female	8	Yes	M-to-M	99.65	820.5	Better	Yes
1452-KIOVK	Male	22	Yes	M-to-M	89.1	1949.4	Good	No
6713-OKOMC	Female	10	No	M-to-M	29.75	301.9	Good	No

Sample of training data is represented in Table 1. It contains the important attributes Gender, Age, Total charges, Feedback and Churn. Implementation with the real time data could provide a better solution for churn prediction model.

Table 2. Testing Data

CustomerID	Gender	Age	Phone Service	Treatment	Monthly Charges	Total Charges	Feed back

5590-VHVEG	Female	15	No	Month-to-month	5000	5050	Good
3575-GNVDE	Female	34	Yes	One year	590	1889	Good
3728-QPYBK	Male	50	Yes	Month-to-month	5300	10800	Bad

Trained model has been tested by testing data which is represented in Table 2. Training data have the detail about Patient ID, Gender, Age and Transaction but it does not included the churn detail of the patient and that can be predicted by the developed model.

Proposed System Design

Churn prediction model is designed and developed based on the steps and algorithms. Predicting the patient's behavior is important phenomenon to determine the reasons and needs of patients. Customization of CRM application is to enable patient relationship management within health organizations. Designed healthcare system provides relationship between patients and management system, by providing proactive care to ill patient, which establishing productive, long-term relationships with patients. Following characteristics is Patient relationship management systems.

Patient satisfaction

To increase patient satisfaction continuously, analyze the routine steps time to time such as process of entering, discharges, transfers from one place to another and more endorsements can be made.

Co-ordinate delivery of care

Customized workflows are deployed to automate care co-ordination activities between service renders and organizations would help to improve patient health. This improves outcomes, decrease operational cost efficiently and effectively.

Proactively manage chronic patient:

Clinicians have identified patients with specific chronic diseases and targeted for proactive communications. Through proactive communications, clinicians can educate the patient of upcoming syndromes and remind them of ways to manage their illnesses.

Following are the basic steps involved in model generation.
1. Read input data: Patient case history is being collected in the form of information and that information is kept in to database. Information can be kept in the form of either in xls or csv or JSON or SQL database. Dataset is read and loaded by using this step. Data can be loaded from the database or from the websites and can join two data into one data.

2. Pre-processing: After loading and joining data, some pre-processing steps are used on to the data in which the churn customer's visualization is considered and different colors are provided to the churn attribute.

3. Partitioning: Model is developed based on training and testing phases in which same data set is portioned in (i) training phase reads 80% of data from the input and model is trained (ii) Testing phase use remaining 20% of data to test the trained model. This partition process takes data randomly from the dataset.

4. Train model: Training the model was based on the partitioned data, in which some amount of data has taken to analyze and, generating rules to build a model, in which machine learning algorithms are applied to train a model.

5. Evaluation: Developed model is tested by using the test data and it is useful to determine the accuracy and performance of the developed model. Good performance is obtained by quality of data set.

Predicting Patient Churn System

Patient churn prediction system model developed to predict churners in the healthcare organization is represented in Figure 1.
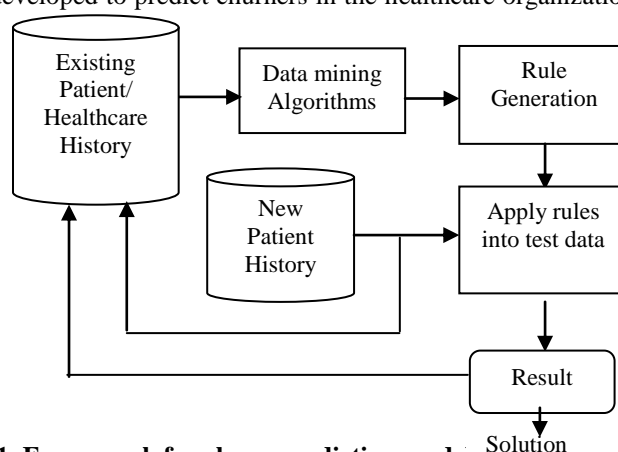

Figure 1. Framework for churn prediction model

Predictive modeling has concentrated on analyzing CRM data and customer-level models to describe the probability of a customer who take necessary action. For few cases, a continuous monitoring of the patient will improve the relation between patient and health care professional. Here data mining techniques and forecasting models have been adopted. Data mining system would prepare data models and these models have consequently helped for future predictions. Forecast models are decision tree, k-means and naïve bayes algorithms are used for prediction.

Churn prediction model is built by using old data which are partitioned into two groups namely training and testing. Algorithms are applied to train and develop a model by using training dataset and rules are generated for prediction. Test data is applied to test the performance and accuracy of a developed model. Prediction for new input data is obtained by using the developed model and managerial suggestions are made by organizations using the result.

Past data: Existing patient information is a collection of patient's data from the healthcare center. It can be partitioned into training and testing data.

Data mining algorithms: Data mining algorithms are applied to past dataset for clustering and develop a new model.

Rule generation: Model is built by generating the rules based on algorithms, with object similarity and probability between the objects in the training data.

Admitted patients information: New input (new patient) details are given as input to the trained model to predict whether the patient will churn or not and the new patient data is stored into the database.

Apply rules into test data: Generated rules are applied into the new patient history which does not have churn/not churn details and it provides the prediction result.

Result: This provides the prediction result of the new input. Then result of the prediction is updated into the existing patient history database.

Solution: From the result, healthcare center is able to view the problem of churn and overcome it to retain the patients. It provides marketing strategies to compare and improve the healthcare benefits.

Algorithms

Clustering and classification algorithms are implemented using java code. K-means clustering algorithm group the data objects into clusters based on similarity in the characteristics. J48 classifier is a type of decision tree classification algorithm which is used to classify the data objects into their groups based on distance functions.

K-Means clustering algorithm

K-means clustering algorithm is implemented using java that takes two input values namely number of clusters and seeds (iterations). It is mandatory to give input values for number of clusters to generate clusters. Based on given iterations, the cluster objects are reassigned into clusters when distance between mean and data object would be even more. Data objects are moved to the nearest clusters within the given iterations.

K-means pseudo code is given below where inputs are given by the user and algorithm would run based on it and provides the k-clusters. Initial clusters are generated by k number of random centroids. All objects are assigned into clusters based on shortest distance. Recalculate the distance between object and neighboring centroid. If mean distance is less when compared with previous cluster then reassign objects into their clusters based on distance. Iterative process ends until there is no change in cluster center.

Table 3. K-means Clustering pseudo code

```
INPUT: S=set of instances, k= number of clusters
OUTPUT: k Clusters
Require: S≠ø,k>0

Procedure Generate Cluster
    Initialize k random centroids
repeat
            for all Instance I in S do
            shortest ← 0
            membership ← null
            for all centroid c do
                    dist ← Distance(c)
                    ifdist< shortest then
                    shortest ← dist
                    membership ← c
                    end if
            end for
            end for
            RecalculateCentroids(c)
until convergence
```

K-means clustering algorithm results in Table 3shows patient churn analysis and Table 4 represents cluster with Customer ID.

Number of clusters: 3

Table 4. Result of k-means clustering

Cluster 0	Cluster 1	Cluster 2
6	10	3

Table 5. Cluster with Patient ID

Patient ID	Cluster Number
0280-XJGEX	Cluster1
1452-KIOVK	Cluster1
3655-SNQYZ	Cluster0
3668-QPYBK	Cluster1
4190-MFLUW	Cluster0
5129-JLPIS	Cluster0
5575-GNVDE	Cluster1
6388-TABGU	Cluster 1
6713-OKOMC	Cluster0
7469-LKBCI	Cluster1

7590-VMVEG	Cluster1
7795-CFOCW	Cluster1
7892-POOKP	Cluster2
8091-TTVAX	Cluster0
8191-XWSZG	Cluster1
9237-HQITU	Cluster2
9305-CDSKC	Cluster1
9763-GRSKD	Cluster2
9959-WOFKT	Cluster0

Decision Tree (J48):

One type of decision tree   J48 is implemented using java code in eclipse. This comes under the classification algorithm of supervised learning method which has class labels to classify data objects into classes. Decision tree is used to provide a decision for future prediction by using existing information. In the analysis, patient records are taken to build a churn prediction model by using J48 classifier. An algorithm takes two input values. One is training data and another one is testing data. The output is represented in the form of patient ID, actual and predicted information of testing data. It represents the accuracy of the model by representing instances which are correctly and incorrectly classified from the data. Pseudo code for decision tree is:

Table 6. Decision Tree pseudo code

```
INPUT: Training data, testing data
OUTPUT: Test data instances with predicted value
Procedure Training data
    Initialize Index attribute
        Run Classifier
    Classify data based on index value
    P: pruning (true/false)
    Store a current tree
    Repeat classifier until no binary split is possible
    Store current solution tree (trained model)
        Apply test data
        Test a data by using trained model
End
```

Decision tree is used to build a churn prediction model using existing patient's history by constructing rule. Then pruning is applied for tree reduction which is used to reduce the size of tree based on relatedness. Result of decision tree is represented in Table 7.

Table 7. Result of decision tree

Classification of Instances	No of instances	Accuracy
Correctly	19	100%
Incorrectly	0	0

From the classification result, test data is applied to the trained model and it provides the predictor value. Table 8 contains Patient ID, Actual value of churn and predictor value of churn. Patient churn data are considered for this churn prediction model and its results are represented in the following Table 8.

Table 8. Result with predictor value

Patient ID	Actual	Prediction
0280-XJGEX	No	No
1452-KIOVK	No	No

3655-SNQYZ	Yes	Yes
3668-QPYBK	No	No
4190-MFLUW	Yes	Yes
5129-JLPIS	Yes	Yes
5575-GNVDE	No	No
6388-TABGU	No	No
6713-OKOMC	Yes	Yes
7469-LKBCI	No	No
7590-VMVEG	No	No
7795-CFOCW	No	No
7892-POOKP	No	No
8091-TTVAX	Yes	Yes
8191-XWSZG	No	No
9237-HQITU	No	No
9305-CDSKC	No	No
9763-GRSKD	No	No
9959-WOFKT	Yes	Yes

Conclusion and Future Work

Health care organization, avoiding the patient churn and retaining them is phenomenal task, which has to be performed by all the stake holders of a hospital. Performance of the hospitals will be rated based on good future, way of communication with patients and the best performance will get more trustiness of the hospital. From this analysis, k-means cluster is used to get useful patterns from the data. Then decision tree algorithm takes training and testing dataset as an input files to process and it provides the results in the form of patient ID, Actual stage and predicted stage of patient for testing data. In future, this work will be implemented with real time data to obtain a better solution.

References

[1]   AfaqAlam Khan, Sanjay Jamwal, M.M.Sepehri, "Applying Data Mining to Customer Churn Prediction in an Internet Service Provider", International Journal of Computer Applications (0975 – 8887), Vol. 9, No.7, November 2010.

[2]   A.O.Oyeniyi & A.B.Adeyemo, "Customer Churn Analysis in Banking sector using data mining techniques" IEEE African Journal of Computing & ICT, Vol. 8, No. 3, 2015

[3]   AmjadHudaib, RehamDannoun, Osama Harfoushi, RubaObiedat, HossamFaris,"Hybrid Data Mining Models for Predicting Customer Churn", Int. J. Communications, Network and System Sciences, 2015

[4]   AniketGavhale, PunamPofale, Sushil Samarth, SayaliBaitule, "Study of K-Means Algorithm with Canopy Clustering Algorithm in Hadoop", IJCAT - International Journal of Computing and Technology, Vol. 2, Issue 3, 2015

[5]   Arunadevi and Nirmala, "Construction of Decision Tree : Attribute Selection Measures", International Journal of Advancements in Research & Technology, Volume 2, Issue 4, April-2013

[6]   L Breiman, JH Friedman, RA Olshen, and CJ Stone., "Classification and Regression Trees", Wadsworth Inc, 1984.

[7]   Chih-Fong Tsai and Yu-Hsin Lu, "Data Mining Techniques in Customer Churn Prediction", ASIAIR Recent Patents on Computer Science, Vol. 3, No. 1, 2010

[8]   Clement Kirui1, Li Hong, Wilson Cheruiyot and Hillary Kirui, "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, 2013

[9]   Dale Shaller, "Patient-Centered Care:What does it take?", The Picker Institute, 2007

[10]   Dheeraj Kumar, James C. Bezdek, MarimuthuPalaniswami, SutharshanRajasegarar, Christopher Leckie, Timothy Craig Havens, "A Hybrid Approach To Clustering In Big Data", IEEE Transactions on cybernetics, Vol. 46, No. 10, 2016

[11]   DžulijanaPopović, BojanaDalbeloBašić, "Churn Prediction Model in Retail Banking Using Fuzzy C-means Clustering", 2009.

[12]   Gregory W. Ramsey and Sanjay Bapna,"A Technique to Exploit Free-Form Notes to Predict Customer Churn", International Journal of Computational Models and Algorithms in Medicine, 4(1), 1-16, 2014

[13]   HaithamAlali, "Online Healthcare Communities of Practice Identifying the Critical Success Factors", International Journal of Computers in Clinical Practice Volume 1, Issue 1, January-June 2016

[14]   Jiawei Han, Micheline Kamber , Jian Pei, "Data Mining: Concepts and Techniques", Second Edition (The Morgan Kaufmann Series in Data Management Systems) 2012.

[15]   Kiran Dahiya and KanikaTalwar, "Customer Churn Prediction in Telecommunication Industries using Data Mining Techniques- A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015

[16]   Lei Meng, Ah-Hwee Tan, Donald C. WunschII,"Adaptive Scaling of Cluster Boundaries for Large-Scale Social Media Data Clustering", IEEE Transactions in Neural Networks And Learning Systems, 2015.

[17]   Li Zhang, Baoxing Li, Liyong Zhang, Dawei Li, "Fuzzy Clustering of Incomplete Data Based on Missing Attribute Interval Size", 2015.

[18]   MahsaGhaziAsgar, SomayehShahri, "Customer relationship management in healthcare", Cumhuriyet University Faculty of Science Science Journal (CSJ), Vol. 36, No: 3 Special Issue, ISSN: 1300-1949, 2015

[19]   Mirza Hasan Hosseini; Mohammad MahmoudiMaymand ,OjanKarimii; Mostafa Ahmadinejad, "Predicting the Bank Customer Switching Based on Data Mining Technique", Spectrum: A Journal Of Multidisciplinary Research, Vol.2 Issue 10, ISSN 2278-0637, 2013

[20]   Prodromos D. Chatzoglou, Leonidas L. Fragidis, TriantafylliaDoumpa, and Vassilios P. Aggelidis, "Hospital Information System Evaluation", 10th International Conference on Information Communication Technologies in Health, 12-14, Samos Island Greece, July 2012

[21]   Reza AllahyariSoeini, KeyvanVahidyRodpysh, "Applying Data Mining to Insurance Customer Churn Management", IACSIT Hong Kong Conferences, IPCSIT, Vol. 30, 2012

[22]   S.-Y. Hung, W.-H. Hung, C.-A.Tsai, and S.-C. Jiang, "Critical factors of hospital adoption on CRM system: Organizational and information system perspectives" Decision Support Systems, Vol. 48, pp. 592-603, 2010

[23]   Tom Seymour, Dean Frantsvog, and Tod Graeber, "Electronic Health Records (EHR)",American Journal of Health Sciences – Third Quarter, Volume 3,Number 3, 2012

[24]   T. Verbraken, V. Wouter, and B. Bart, "A novel profit maximizing metric for measuring classification performance of customer churn prediction models" IEEE Trans. Knowl. Data Eng., Vol. 25, No. 5, pp. 961–973, 2013

[25]   Tomé R. Vardasca, Henrique MG Martins, "Patient Relationship Management in public healthcare settings", IEEE ICC-2011.

[26]   Vera Lukić, "Laboratory Information System – Where are we Today?", J Med Biochem 36: 220 –224, 2017.

[27]   Wenjie Bi, MeiliCai, Mengqi Liu, and Guo Li, "A bigdata clustering algorithm for mitigating the risk of customer churn" IEEE Transactions on Industrial Informatics, Vol. 12, No. 3, 2016