

Forecasting and Shun Mobility of Patient in Hospital

I. Franciska, B. Swaminathan

Abstract— Patient mobility prediction is one of the most important problems in patient relationship management (PRM), and its aim is to retain valuable patients to maximize the profit of a healthcare organization. Highly competitive organizations have understood the marketing strategies are important to retain existing and expensive patients to survive in market. In this analysis machine learning techniques are used to build better prediction model, which are k-means, decision tree and naïve bayes algorithm. In which clustering algorithm groups the patients to identify most likely churners group and predictive models are find the reason for mobility. Tracking patients details are used to provide better quality of service. Regular communication services Email, SMS, voice synthesis calls about their appointment, checkups, and treatment details will help to engage patients with the healthcare centers.

Index Terms— PRM customer mobility, Prediction, Clustering

I. INTRODUCTION

All the companies are based on marketing as a business promotion, The company believes that they invest more into the customer accusation, and customer will stay for long periods, the revenue can be obtain for every month/frequently with minimal cost or effort. The marketing is taking to the customer level. Mobility may be based on excessive and/or influence policies. The traditional solution is to predicting valuable customers and their needs to reduce the mobility rate.

CRM is used to analyze the customer activities in order to raise the value of their customer selection. It gives a technological solution that helps in improving the targeting efforts by preparing databases and applying the automation tools to link sales and marketing functions and CRM applications have complete information of each individual customer.

A. PATIENT RELATIONSHIP MANAGEMENT

A Patient Relationship Management (PRM) system, more commonly known as customer relationship management (CRM), can be adopted by a public hospital, and it is an important leverage of patient satisfaction.

Manuscript received October, 2016.

I. FRANCISKA, P. G. Student, Department of Compute Science and Engineering, Rajalakshmi Engineering College, Thandalam, Chennai, India.

B. SWAMINATHAN, Professor, Department of Compute Science and Engineering, Rajalakshmi Engineering College, Thandalam, Chennai, India.

PRM involves tracking patient information from last diagnosis, diet, exercise, family history and allergy information which can be given by the patient. By storing all of this information, health care providers will be able to send e-information to patients about newly published health care studies and treatments that may be relevant to the patient, or offer specialized suggestions that fit the patient profile.

This analysis is based on machine learning algorithms which are classified into supervised and unsupervised learning algorithms. In which k-means, Decision tree and Naive Bayes algorithms are analyzed for customer mobility prediction. K-means clustering is an unsupervised learning algorithm which is used to cluster similar characteristic objects by initializing n center or means. Decision Tree and Naive Bayes are supervised learning algorithms which are mainly used for prediction process.

B. DATA MINING

Generally, data mining is the process of analyzing data from various perspectives and shortening it into useful information from large amounts of data. Data mining is primarily used today by various companies with a strong consumer focus - retail, financial, communication, and marketing organizations. Data mining techniques are based on clustering and classification, here clustering is used to group the data points based on its similarity and classification algorithms are applied to predict the reason for mobility.

C. MACHINE LEARNING ALGORITHM:

Machine Learning algorithms are classified into supervised, unsupervised and semi supervised learning algorithms some of the examples form machine learning algorithms are Decision tree, k-means clustering, hierarchical clustering, k-nearest neighbor classifier, Naive Bayes classifier, Linear regression and Logistic regression, etc.,

- a) **Supervised Learning Algorithm:** In supervised learning algorithm class label of each training tuple is provided so classification is based on that class labels.
- b) **Unsupervised Learning Algorithm:** In which the class label of each training tuple is not known and the number of classes may not be known at advance.
- c) **Semi supervised Learning Algorithm:** Semi supervised learning is a class of supervised learning tasks and techniques that also make use of unlabeled data for training- typically a small amount of labeled data with a large amount of unlabeled data.

II. RELEATED WORK

Customer mobility prediction is an important factor to sustain in the market level with competitors, that can be applied various sectors such as telecommunication, banking, insurance and healthcare organizations. Mobility prediction models are developed based on the industries because it varies due to customer relationship, quality of services, revenue, marketing strategies. In healthcare organizations, mobility prediction models are helpful to predict patients who are about to churn.

D. TELECOMMUNICATION DEPARTMENT

Analysis china telecommunication industries and discussed about various parameters used in analysis. Why the customers are moving from one service provider to other service provider [22]. The analysis is based on existing customers behavior, the behavioral information is extracted from Business Support System (BSS). In BSS contains customer call details (minutes of international calls, counts of national calls), billing details, and complete profile about the customer, scheme, and customer level of important very important person (VIP) or not are considered.

Amjad, analyse is also based on the customer churn in telecommunication industry. In this analysis three hybrid data mining models are anticipated for predicting customer mobility in a telecommunication company. Models are developed to validate using real world data provided by Jordanian Telecommunication Company [1].

E. BANKING SYSTEM

Analysis in banking sector because banking sector has always tried to track customer interaction with company [14]. In almost all studies reviewed, bank customers are those customers who had relationship with the bank. Consequently churn in such condition could be defined as the terminating of contract from the customers side or no reactivating an account after going into dormancy.

Analyse the customer churn in retail banking. In this analysis mobility prediction model is developed for retail bank which has 5000 clients, those who are taken randomly from the age 18 to 80 years in the total population [7]. In which variable selection process is based on more transactional details that would lead to better model performance.

Most of the analysis on the customer mobility has been based on the data mining techniques. The statistical population of this analysis includes all customers of Iranian Bank-e Tejarat in Tehran [12]. In this analysis questionnaires are used to get more details about customer because prediction phase need to know the reasons of customer mobility, on the other hand bank data base is not enough for prediction

F. INSURANCE SYSTEM

Analyse the customer churn in Iran Insurance Policy Company. Churn customers are important aspect for insurance companies in competitive marketing environment of today [15]. Insurance customer mobility prediction model is developed by data mining concepts, which includes Understanding the business, Data Collect, Data preparation, normalization, Modeling and Evaluation.

G. PATIENT RELATIONSHIP MANAGEMENT

Analyse Patient Relationship Management or for healthcare which have the potential to increase patient satisfaction and reduce healthcare costs [16]. Developed architecture reduces implementation, maintenance, evolution effort and cost. This PRM system provides the capability to send customized alerts to patients about consultation appointments, exams, and treatments. Alerts should be defined by the patient-specified channels (an Email, SMS, call using voice synthesis, etc.), and schedule they prefer. And alert services are used which have the ability to cancel certain appointments. PRM is integrated with the Hospital Information System (HIS), Electronic Health Record (HER), and Laboratory Information System (LIS).

III. LITERATURE REVIEW

A. SEMANTIC DRIVEN SUBTRACTIVE CLUSTERING METHOD (SDSCM)

Propose a Semantic Driven Subtractive Clustering Method (SDSCM) which is a combination of Axiomatic Fuzzy Set (AFS) and Subtractive Clustering Method (SCM) [22]. AFS is used to quantize a membership function of semantic concept of SDSCM, and the quantized user semantic concept is used to automatically determine the density radius τ_1 , and semi automatically determine weight τ_2 . From the Experimental results SDSCM provides stronger clustering semantic strength than the other two methods subtractive clustering method (SCM) and fuzzy c-means (FCM).

B. K-MEANS AND CANOPY CLUSTERING

[15] Discussed about k-means clustering algorithm and its steps. Initially k numbers of objects are randomly selected that are represented as a cluster mean or center [15]. Then other objects are assigned to the cluster which is the most similar to the cluster center, based on the distance between the object and the cluster mean.

Aniket discussed about canopy cluster and k-means clustering. There are very big bottlenecks when traditional data mining algorithms deal with large data sets [2]. Canopy Clustering is an unsupervised pre-clustering algorithm related to the K-means algorithm, Canopy clustering is a very simple, fast and surprisingly accurate method for grouping objects into clusters by using two threshold T1 and T2 [2]. This canopy clustering method is used to speed up the clustering operations on large data sets.

C. HYBRID CLUSTERING ALGORITHM

Amjad proposed hybrid models which are (i) k-means with Multilayer Perceptron Artificial Neural Networks (MLP-ANN), (ii) Hierarchical clustering with MLP-ANN, (iii) Self Organizing Maps (SOM) with MLP-ANN. Here K-means clustering algorithm filters the data, then Multilayer Perceptron Artificial Neural Networks (MLP-ANN) is applied for prediction [1]. Similarly hierarchical clustering and SOM are used for filter the data then MLP-ANN is used for prediction. From this study hybrid models are outperformed a single model. And the model provides the accuracy and efficiency in mobility prediction by the prediction of future behavior of users[1].

D. FUZZY C MEANS CLUSTERING

Fuzzy method was used for mobility prediction in retail banking because it performed better than the classical clustering algorithms but this work need some Preprocessing steps needed. Model should also include costs of positive and negative misclassifications. This algorithm Provides more hidden information and Best result for overlapped data but Computational time is more than k-means also Sensitive for out layers.

Use the techniques neural networks to cluster the customers into loyal customers, educated customers, youths, and women [12]. In prediction phase decision tree is used to obtain the customer's mobility behavior. Analysis provides some suggestions for future work is to use unstructured and unorganized.

IV. PROPOSED SYSTEM

There are many techniques are proposed for customer mobility prediction to various industries. In this analysis we used [8] k-means for clustering and two probabilistic data mining algorithms Naïve Bayes and Decision tree are used to evaluate reason for the mobility. Following algorithms are widely used in many clustering, classification and prediction tasks, and that are used to develop a mobility prediction system:

- i. K-means clustering
- ii. Decision Tree
- iii. Naive Bayes

A. K-MEANS CLUSTERING:

- This clustering algorithm was developed by MacQueen. The k-means clustering algorithm is known to be efficient in clustering large data sets.
- The K-Means algorithm aims to partition a set of objects, based on their attributes/features, into k clusters, where k is a user-defined constant.
- The algorithm of k -means starts with initializing k number of cluster centroids. Then, data are iteratively assigned to the nearest cluster and the new centroids of k clusters are recalculated until the termination conditions are reached [8]. Sum of squared error function(E) is calculated by equation (1),

$$E = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - C_j\|^2 \quad (1)$$

k – Number of clusters

C_j – mean of cluster j

n – Number of data points

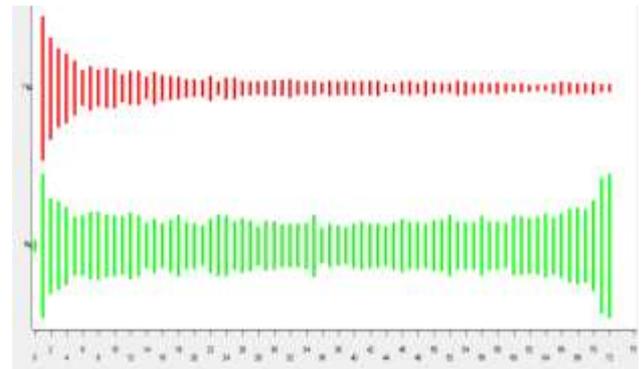


Fig 4.1: Age group clustering for mobility prediction

B. DECISION TREE ALGORITHM:

Decision tree is the most popular predictive model, which represents relationship between variables in the way of top-down order. Tree will start from root node by representing feature to be classified, which attribute have the highest information gain is chosen as the splitting attribute.

Information Gain:

The highest information gain is chosen as the splitting attribute for the node [8]. The expected information needed to classify a tuple in D is given by the equation (2),

$$Info(D) = -\sum_{i=1}^m p_i \log_2(p) \quad (2)$$

$Info(D)$ – is the average amount of information needed to identify the class label

P_i – probability of an arbitrary tuple in D belongs to class C_i

The lower level nodes are constructed in similar way by its feature. Decision tree is used for easy visualization and understanding, and it can process numerical and categorical data. But it suffers from the lack of robustness and over-sensitivity to training data sets.

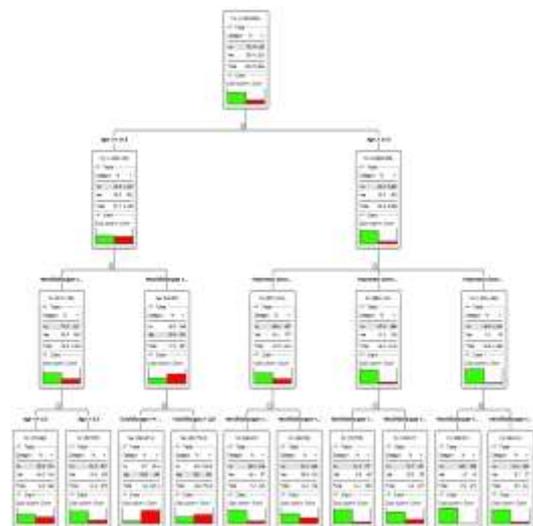


Fig 4.2: Decision tree representation

- Confusion matrix for decision tree is represented in Table 4.1 it consists the accuracy of the model and also it represents correctly classified and wrongly classified objects for decision tree algorithm. Here decision tree algorithm classifies 5267 objects correctly and it provides 93.486% of accuracy.

Table 4.1: Confusion matrix for decision tree

Mobility\prediction	No	Yes
No	3975	158
Yes	209	1292
Correctly classified: 5,267		
Wrongly classified: 367		
Accuracy:93.486%		
Error: 6.514%		

C. NAIVE BAYES ALGORITHM

Naïve Bayes classifiers associated with probabilistic classifiers based on applying Bayes’ theorem with strong independence assumptions between the features [11]. Training set of instances, are represented and characterized as a vector of features $[x_1, x_2, \dots, x_d]$, the assignment is learning from data to be capable to predict most likely class $y_j \in \mathbb{C}$ of the new object whose class is unknown. Naïve Bayes use the Bayes’s theorem given in Equation (3) to estimate the probabilities of the classes.

$$P(y_j | x_1, x_2, \dots, x_d) = \frac{P(y_j)P(x_1, x_2, \dots, x_d / y_j)}{P(x_1, x_2, \dots, x_d)} \quad (3)$$

- $P(y_j | x_1, x_2, \dots, x_d)$ - is posterior probability,
- $P(x_1, x_2, \dots, x_d / y_j)$ - is likelihood,
- $P(y_j)$ - is class prior probability,
- $P(x_1, x_2, \dots, x_d)$ - is predictor prior probability

Table 4.2: Confusion matrix for decision tree

Mobility\prediction	No	Yes
No	918	123
Yes	150	218
Correctly classified: 1,136		
Wrongly classified: 273		
Accuracy : 80.625%		
Error : 19.375%		

- Confusion matrix for naïve bayes algorithm is represented in Table 4.2 it consists accuracy of the model and also it represents correctly classified and wrongly classified objects for decision tree algorithm. Decision tree algorithm classifies 1136 objects correctly and it provides 80.625% of accuracy.

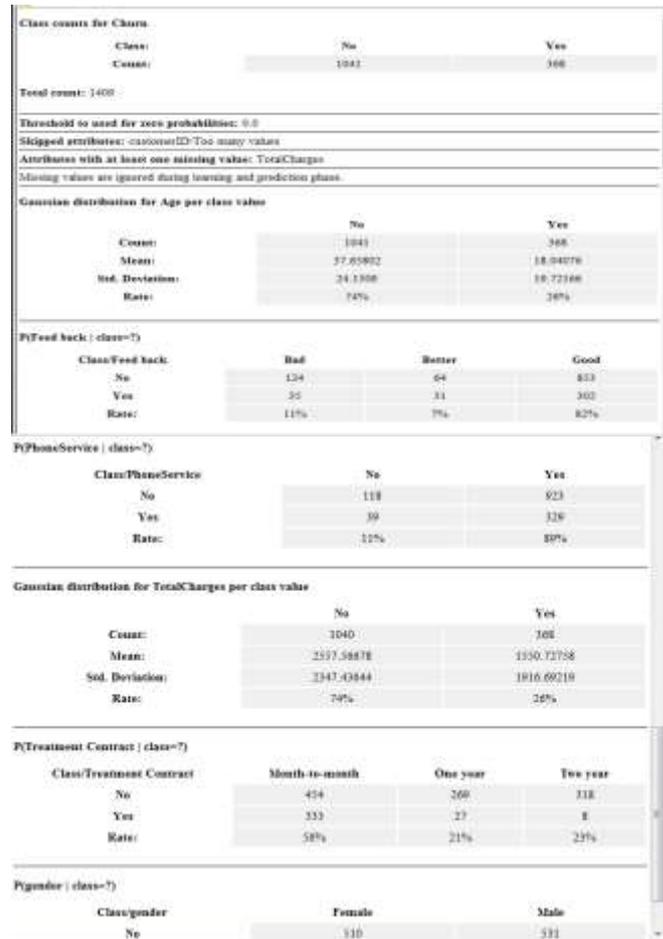


Figure 4.3 Naïve bayes representation

Naïve bayes algorithm result expressed in Figure 4.3 it provides the probability ratio and percentage for the every class. Gaussian distribution calculated for the considered attributes and that are compared with churn class and it provides churn rate based on gender, treatment, total charges and phone services.

V. CONCLUSION

In the health care organization, patient mobility and retention are most phenomenal task, and the performance of the hospitals will be rated base more customer. From this analysis k-means cluster is performed based on patient’s age group and total cost. Then decision tree algorithm and naïve bayes algorithm are used to prediction which customers are most likely to churn. From this result decision tree model provides 93% and naïve bayes model provides 81% of accuracy for the given input data, so the performs of decision tree is better than naïve bayes model.

REFERENCES

- [1] [Reza, Keyvan-2012] Reza Allahyari Soeini, Keyvan Vahidy Rodpysh, “Applying Data Mining to Insurance Customer Churn Management”, 2012 IACSIT Hong Kong Conferences, IACSIT vol. 30 (2012)
- [2] Amjad Hudaib, Reham Dannoun, Osama Harfoushi, Ruba Obiedat, Hossam Faris” Hybrid Data Mining Models for Predicting Customer Churn”, Int. J. Communications, Network and System Sciences, 2015.
- [3] Aniket Gavhale, Punam Pofale, Sushil Samarth, Sayali Baitule “Study of K-Means Algorithm with Canopy Clustering Algorithm in Hadoop”, IJCAT - International Journal of Computing and Technology, Volume 2, Issue 3, March 2015
- [4] 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.
- [5] 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, March 2013
- [6] Dale Shaller, “Patient-Centered Care: What does it take?” The Picker Institute, October 2007.
- [7] Dheeraj Kumar, James C. Bezdek, Marimuthu Palaniswami, Sutharshan Rajasegarar, Christopher Leckie, Timothy Craig Havens, “A Hybrid Approach To Clustering In Big Data”, IEEE Transactions on cybernetics, Vol. 46, No. 10, October 2016
- [8] Džulijana Popović, Bojana Dalbelo Bašić, “Churn Prediction Model in Retail Banking Using Fuzzy C-means Clustering”, 2009
- [9] Fred R. Beyette, Charlotte A. Gaydos, Gerald J. Kost and Bernhard H. Weigl, “Point-of-Care Technologies for Health Care”, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 58, NO. 3, MARCH 2011.
- [10] 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, January-March 2014
- [11] Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining: Concepts and Techniques”, Second Edition (The Morgan Kaufmann Series in Data Management Systems), 2012.
- [12] Kiran Dahiya and Kanika Talwar “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.
- [13] Lei Meng, Ah-Hwee Tan, Donald C. Wunsch II, “Adaptive Scaling of Cluster Boundaries for Large-Scale Social Media Data Clustering”, IEEE Transactions on Neural Networks and Learning Systems.
- [14] Li Zhang, Baoxing Li, Liyong Zhang, Dawei Li, “Fuzzy Clustering of Incomplete Data Based on Missing Attribute Interval Size” IEEE-2015.
- [15] Mirza Hasan Hosseini; Mohammad Mahmoudi Maymand, Ojan Karimii; Mostafa Ahmadinejad, “Predicting the Bank Customer Switching Based on Data Mining Technique”, Spectrum: A Journal Of Multidisciplinary Research Vol.2 Issue 10, October 2013, Issn 2278-0637
- [16] Narendra Sharma, Aman Bajpai, Mr. Ratnesh Litoriya, “Comparison the various clustering algorithms of weka Tools”, International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 5, May 2012.
- [17] Oyeniyi & A.B. Adeyemo, “Customer churn analysis in banking sector using data mining techniques” IEEE African Journal of Computing & ICT, Vol 8. No. 3 – september, 2015.
- [18] Tome R. Vardasca, Henrique MG Martins, “Patient Relationship Management in public healthcare settings”, IEEE ICC-2011.
- [19] Wei Shao, Flora D. Salim, Andy Song, and Athman Bouguettaya, “Clustering Big Spatiotemporal-Interval Data”, IEEE Transactions on Big Data, 2016
- [20] Weiguo Sheng, Shengyong Chen, Mengmeng Sheng, Gang Xiao, Jiafa Mao and Yujun Zheng “Adaptive Multi-Subpopulation Competition and Multi-Niche Crowding based Memetic Algorithm for Automatic Data Clustering”, IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION 2015.
- [21] Weiyun Ying, Xiu Li, Yaya Xie, Ellis Johnson, “Preventing Customer Churn by Using Random Forests Modeling”, IEEE-2008
- [22] Wenjie Bi, Meili Cai, 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, June 2016.
- [23] Zhiwen Yu, Peinan Luo, Jane You, Hau-San Wong, Hareton Leung, Si Wu, Jun Zhang, and Guoqiang Han, “Incremental Semi-Supervised Clustering Ensemble for High Dimensional Data Clustering”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 3, MARCH 2016