Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 120 (2017) 23–30

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 24-25 August 2017, Budapest, Hungary

A data mining approach for modeling churn behavior via RFM model in specialized clinics Case study: A public sector hospital in Tehran

Mehdi Mohammadzadeh[a*], Zeinab Zare Hoseini[b], Hamid Derafshi[c]

[a]*Department of Pharmacoeconomy&Administrative Pharmacy, Shahid Beheshti University of Medical Sciences,POBOX 14155-6153,Tehran Iran*
[b]*Department of Engineering&Technology, Payame Noor University, , Tehran, PO BOX 19395-3697, Iran*
[c]*Department of Ophthalmology, School of medicine, Alborz University of Medical Sciences, Karaj, POBOX 314977-9453, Iran*

Abstract

Nowadays Health care industry has a significant growth in using data mining techniques to discover hidden information for effective decision making. Huge amount of healthcare data is suitable to mine hidden patterns and knowledge. In this paper we traced behavior of patients during the period of 3 years in three clinics of a big public sector hospital and tried to detect special groups and their tendencies by RFML model as a customer life time value (CLV). The main goal was to detect 'potential for loyal' customers for strengthen relationships and 'potential to churn' customers for recovery of the efficiency of customer retention campaigns and reduce the costs associated with churn. This strategy helps hospital administrators to increase profit and reduce costs of customers' loss. At first, K-means clustering algorithm was applied for identification of target customers and groups and then, decision tree classifier as churn prediction was used. We compared performance of three clinics based on the number of loyal and churn customers. Our results showed that Pediatric Hematology clinic had a better performance than that of other clinics, because of more number of loyal customers.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception.

Keywords: Hospital information system (HIS); data mining; clustering; classification; RFM model; CLV.

* Corresponding author
E-mail address: m_mohammadzadeh@sbmu.ac.ir

1877-0509
Peer-review under responsibility of the scientific committee of the 9th International Conference on Theory and application of Soft Computing, Computing with Words and Perception.
10.1016/j.procs.2017.11.206

1. Main text

Business and marketing organizations may be ahead of healthcare in applying data mining to derive knowledge from data. This is quickly changing. Successful mining applications have been implemented in the healthcare arena. In recent years, data mining techniques are widely used in healthcare system. Three important mining applications include, Hospital Infection Control, Hospitals Ranking and High-Risk Patients' identification (Obenshain and MAT 2004). Most studies have investigated prediction of different types of disease and sickness and they were less focused on the relationship between patients as customer and hospitals as organization especially patients churn and loyalty. Customer churn are more attended in other sectors such as different companies, retails and shops (Khajvand et al. (2011) and telecommunication sectors Verbeke et al. (2012) and Amin et al. (2017). Ramanan et al. have determined theory of mind performance in Alzheimer disease using data mining study. This study was done on 48 Alzheimer's patients and 44 behavioral-variant Front temporal Dementia patients. They have highlighted the relevancy of data mining statistical approaches in clinical and cognitive neurosciences Ramanan et al. (2017). For neonatal jaundice in newborns, predictive models using Naive Bayes, multilayer perceptron, and simple logistic were applied Ferreira et al. (2012). The dataset consisted of 227 healthy newborns. Also in another study, Naive Bayes classifier and J48 decision tree algorithm were used for building predictive models for MERS-CoV infections Al-Turaiki et al. (2016). The dataset used consists of 1082 records. In this paper we have used other applications of data mining on patient's data records provided in Hospital Information System (HIS). We have supposed patients as customers and tried to detect special types of patients and their behavioral tendencies with transactional data. We have used RFM (Recency Frequency Monetary) model as customer lifetime value (CLV) Khajvand et al. (2011) and marketing analysis methods for segmentation and prediction of model. In 1995, RFM defined as Recency, Frequency and Monetary. Recency is the period since the last purchase. A lower value corresponds to a higher probability of the customers to making a repeat purchase. Frequency is a number of purchases made within a certain period and higher frequency indicates greater loyalty. Monetary defines as the money spent during a certain period; a higher value indicates that the company should focus more on that customer Bult and Wansbeek, (1995). In the healthcare system, higher frequency means referring of patients for all/more health needs to our hospital and this occurs when patients are satisfied with hospital services. Also the length of time that the patient has been in contact with us is a major factor that shows whether the patient has discontinued its relationship with us or not. We have added this factor to RFM model and used RFML model to evaluate patient's loyalty. The ultimate goal in the healthcare system is to have hospital and staffs who work well so that patients as customers refer again to this hospital if needed. In this way, patients will be satisfied as our loyal customers.

1. Methodology

2.1 materials and preprocessing data

In this research a big public sector hospital (Shohadaye Tajrish Educational Hospital) has been studied. We selected three specialize clinics based on highest number of patients; the Cardiovascular, Neurology and Pediatric Hematology clinics. Three year Patients' data records were extracted from Hospital Information System (HIS). The numbers of transactions in these clinics were 19036, 41257, 3425 respectively that belonged to 8979, 22210, and 677 patients. Due to the nature of treatment industry and contrast between customer loyalty and patient loyalty, we selected outpatient data records. Outpatients select their preferred hospital with no referral by the doctor treating.
The raw dataset consisted of socio demographic characteristics of patients, services that used in visit date, admission type, insurance type and amount paid. As regards, we wanted to use RFML model for evaluation of patients' categories, so R (recency), F (frequency), M (monetary) and L (length) attributes must be calculated. Then based on clustering model, weighted Life Time Value (LTV) was calculated for each cluster. We believe that greater amount of LTV represents more faithful customers and lower amount shows churn customers. So with the help of LTV amount, four clusters have been assigned into four class label of patient (loyal, potential for loyal, potential for churn and churn) to compare patients of three clinics. Also we used R, F, M and L attributes for classification of the model to predict churn behaviors of new customers and behavioral analysis of special current patients.