Implementing Data Mining Techniques for Marketing of Pharmaceutical Products

Ms. Priti Sadaria
Saurastra University
Virani Science College,
Yogidham Gurukul,
Kalawad Road, Rajkot.

Ms. Miral Kothari
Gujarat Technological
University
AITS, Yogidham Gurukul,
Kalawad Road, Rajkot.

Ms. Jalpa Gondaliya
Saurastra University
Virani Science College,
Yogidham Gurukul,
Kalawad Road, Rajkot.

ABSTRACT

In the era of 21st century as the people become more health conscious as compare to earlier time but even though selling of any pharmaceutical products become more difficult job for the marketer. By doing market research and taking feedback from different group of customer inform of some questionnaire, we can collect data. With the use of particular Data Mining technique these data can be analyzed and information can be extracted. Such kind of information can be used according to the requirement to take meaningful decision for particular product. Pharma industries use such information to produce new products or to decide quantity for manufacturing of existing products. We have used Rule-Based Classification techniques to analyze data.

Keywords

Marketing, Data Mining, KDD, Decision Tree, Rule-Based Classification

1. INTRODUCTION

Marketing is a process which deals with "meeting needs profitability" of customer. Marketing is also identify as a set of various processes for creating, communicating and delivering values to customers [1]. It also deals with managing customer relationships for long term retention of customer for specific brand. In nut shell, marketing leads to managing target markets to enhance benefits of the company. Following figure (see Figure 1) illustrate simple marketing system.
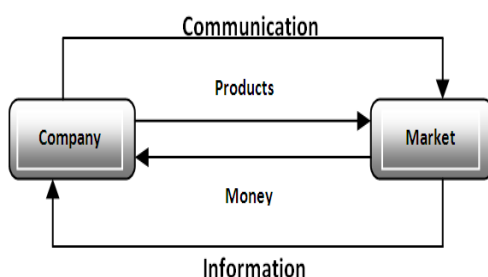

Fig 1 : Marketing System

Product, Price, Place and Promotion are basic factor which directly affects the market. Marketer must have to identify about needs, wants and demands of customer [1]. Marketing of pharmaceutical products include consumer oriented promotion as well as doctor

oriented promotion [2]. End user means patients and prescriber means doctors play vital role in pharma industry. This concept is shown in Figure (see Figure 2)[3].

Interaction between physicians and companies is vital part for marketing of pharmaceutical products. Worthwhile conversation between companies and physician also improve relationship of physician and patients [4]. Many cross promotional and other marketing strategies are used to enhance relationship with physician [5]. Free samples are also provided by pharma companies to the physicians.

For marketing purpose of product, usually companies are appointing Medical Representative. Medical Representative meets personally to doctor and gives lots of free samples per year with branded promotional materials to the doctors[6].

Now a days as people become more health conscious and they are more aware about specific products of particular brand. People also surfing web for health information and sometime they also fill prescription form [7]. Such kind of data is stored in the form of databases. Thus data can be gathered from physicians as well from patients and database is created.

By implementing different data mining techniques, database can be analyzed and group of relevant information can be developed. Marketing can be expanded by using such relevant information. Company can also take meaningful decision for future plan [8].

2. DATA MINING

Since last few years "Data Mining" terms has been widely used in various fields like banking, marketing, engineering etc[13]. No fields remain untouched by Data Mining. Data Mining plays an important role in designing different strategies of any company.

Data Mining refers to extracting or "mining" knowledge from large amount of data [9].

Data Mining have following five main aspects :

- – Extract, transform and load transaction data onto the data warehouse system.
- – Store the data in a multidimensional database system.
- – Provide data access to business analysts.
- – Analyze the data by application software.
- – Represent data in form of graph or table.

Various data mining techniques are used for mining information from large databases. Frequently used data mining techniques are as follow:

www.ijcait.com

International Journal of Computer Applications & Information Technology
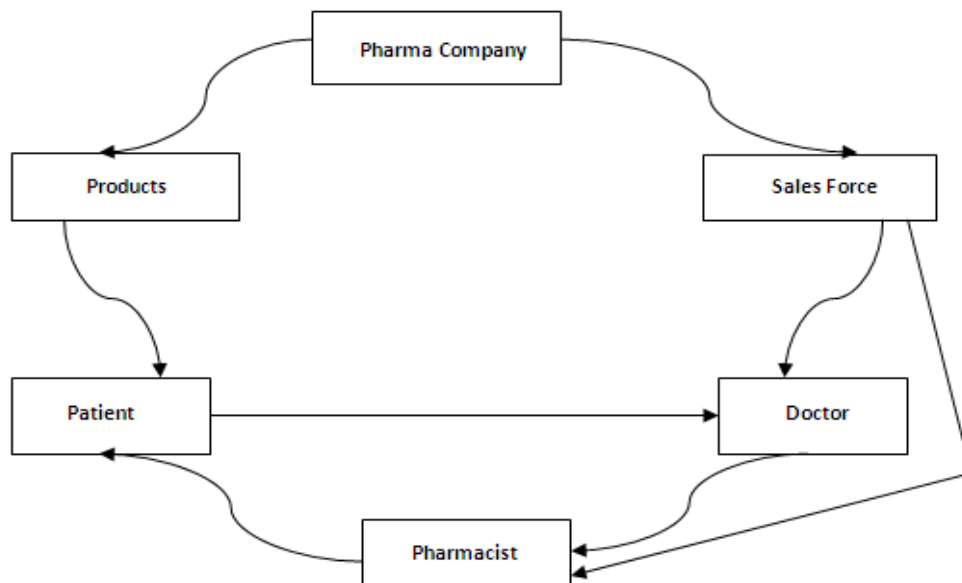Vol. II, Issue I, January 2013 (ISSN: 2278-7720)

Fig. 2 : Relationship between Pharmacist, Patient and Doctor

- Association modeling
- Classification
- Clustering
- Forecasting
- Regression
- Sequence discovery
- Visualization

Association modeling is used to identify different items which are available together depending on the market basket analysis and in cross selling of products. Algorithms like apriori and statistics are use for association modeling. Classification includes different classes of the information depending on the prediction of future customer behavior depending on certain criteria. Generally neural networks and decision trees are used for classification. The clustering technique of data mining model is used for segmenting heterogeneous population into a number of more homogeneous clusters [10, 11]. Discrimination analysis and neural networks tools are used for clustering. Future planning for the company is designed with help of forecasting technique.

Forecasting is based on continuously valued outcomes. Tools used for forecasting is survival analysis and neural networks.

One example of fore casting model is demand forecast.Regression technique is one kind of statistical estimation technique which is used to map each data object to real value. Modeling of causal relationship, prediction based on forecasting are the example of usage of regression. Tools used for regression are logistic regression and linear regression. Sequence discovery technique of data mining concentrate on the states of the process generating the sequence over some particular time period. Tools used for sequence discovery are set theory and statistics. The last technique of data mining model is visualization. Complex pattern can be viewed by customer by presentation of data through visualization. Most common examples of visualization are Hygraphs,3D graphs and SeeNet.

Combination of any two or more data mining technique can be used according to the requirement of the company for particular product's production. To understands and predict preferences of customer, data mining techniques are used. The aim of implementing data mining techniques is to identify potentially useful information from large data set[12]. Various data mining techniques are used to identify correlated patterns. This process is also known as Knowledge Discovery in Database (KDD). This concept is illustrated in Figure (see Figure 3). Deployment into the market is the main purpose of identifying such patterns. As a result of such implementation ultimately company can increase profit margin as well as can gain proper markets for selling the products.

Different data mining software are used traditional statistical methods, and non-traditional statistical methods. Traditional statistical method includes cluster analysis, discriminant analysis and regression analysis while non-traditional statistical method includes decision trees, neural networks, association analysis and link analysis. Hidden patterns can be detected from data [10].

We have used classification techniques of data mining to mine data. Decision tree induction is most commonly used to classify data. Specifically here Rule-Based Classification is

www.ijcait.com

International Journal of Computer Applications & Information Technology
Vol. II, Issue I, January 2013 (ISSN: 2278-7720)

used to analyze data and find different kind of patterns and relations in data[13].

3. DECISION TREE

Decision trees are simple classifier techniques of data mining. This technique is most popular due to graphical representation of data. Data can be graphically represented as tree structure
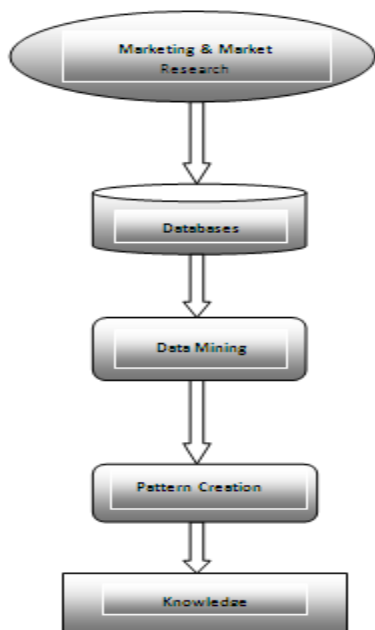

Fig 3 : Knowledge Discovery in Database

and sometimes also in form of rules[14]. The construction of decision tree classifiers does not require any domain knowledge and parameter setting so classification become easier. Decision tree can handle multidimensional data. Simple structure of decision tree is illustrated in Figure(see Figure 4).
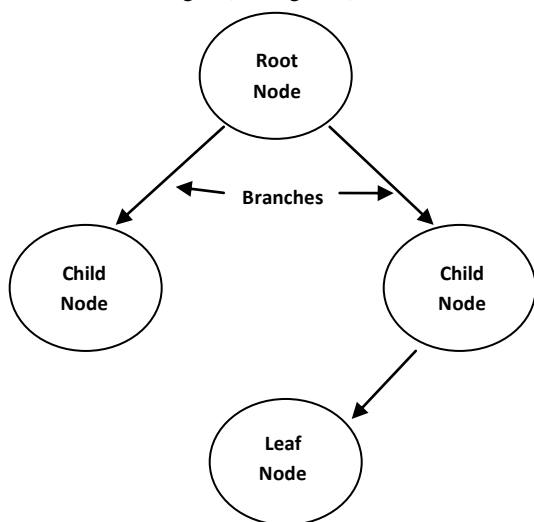

Fig 4 : Simple Decision Tree Structure

Classification and learning steps of decision tree are very simple and fast. Each non leaf node specifies a test on attribute and each leaf node specifies a class label. Each branch specifies an outcome of the test [9]. Decision tree induction algorithm is widely used for classification in various fields like finance analysis, manufacturing and production and medicine. The accuracy of decision tree classifier is high so it is frequently used to analyze data.
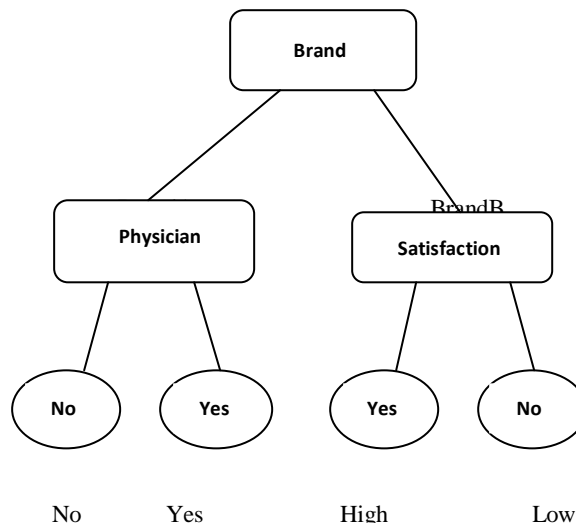

Fig 5 : Decision Tree for Comparison of Brands

The type of decision tree can be either binary or ternary depending on the data set. In binary tree structure, each node has two leaf which is also known as child nodes and it specify "yes" or "no" depending on the test of the attribute. The outcome specifies the way of data to go into next level. In ternary tree structure it is not necessary to have two child nodes for each node. In either or tree structure, process starts from the root node.

Figure(see Figure 5) illustrate simple decision tree for comparison of two different brand of medicine. Concept of get_commission and satisfaction level of patient is shown in decision tree. Each internal node indicates test on an attribute. Each leaf node indicates a class which indicates either physician get_ commission = yes or get_commission = no as well as it also indicates whether patient prefer_brand = yes or prefer_brand = no.

CHAID (Chisquared Automatic Induction) algorithm and CART (Classification and Regression Tree) algorithm are most commonly used algorithms in decision tree techniques.

3.1 Rule-Based Classification

In rule-based classification, data set is represented as a set of IF_THEN rules. The rules can be generated either from a decision tree or from a sequential covering algorithm. Rule-based classifier uses a set of IF_THEN rules for classification. General form of rule is as follow:

IF condition THEN conclusion

The IF part is known as the rule antecedent and THEN part is known as rule consequent. The condition part contains one or more attribute test conditions which are logically ANDed. If

www.ijcait.com

International Journal of Computer Applications & Information Technology
Vol. II, Issue I, January 2013 (ISSN: 2278-7720)

the condition specified with IF is true then the conclusion specified with THEN is evaluated and it shows that rule covers the tuple for which IF part contain the rule[9 Jiawei]. The fields or columns that are used for that are used to create the rules are known as inputs[9].

 Coverage and accuracy can be defined as follow for tuple X and a class labeled data set D. Now n_{covers} be the number of tuples covered by rule R and $n_{correct}$ be the number of tuples correctly classified by R and $|D|$ be the number of tuples in D then we can define coverage and accuracy as follow for rule R[9] :

$$Coverage\ (R) = n_{covers}\ /\ |D|$$

$$Accuracy\ (R) = n_{correct}\ /\ n_{covers}$$

3.2 Rules Extraction from a Decision Tree

By using decision tree IF-THEN rules can be extracted. IF-THEN rules are easy to understand as compare to decision tree. One rule is designed for each path from root to leaf node to extract rule for decision tree. More than one splitting criteria are logically ANDed which are exist into one path which firm IF part of the rule. The leaf node includes the prediction criteria and firm THEN part of the rule.

Following rules are extracted from decision tree shown in Fig 5.

● Rule 1 : IF Brand = BrandA and Physician = No THEN get_ commission = No.

● Rule 2 : IF Brand = BrandA and Physician = Yes THEN get_ commission = Yes.

● Rule 3 : IF Brand = BrandB and Satisfaction = High THEN prefer_brand = Yes.

● Rule 4 : IF Brand = BrandB and Satisfaction = Low THEN prefer_brand = No.

4. CONCLUSION

Since last few years as pharmaceutical industry is facing difficulty for marketing of pharma products. To satisfy the requirements of physicians and patients, pharma industries have to spend more money for giving different types of promotional products as well as some other schemes. Here in this paper we have shown that by collecting data from different sources and analyzing it with decision tree and rule-based classification technique, some rules are extracted which are useful for taking specific decision about particular medicines or brands.

5. REFERENCES

[1] Kotler, P. and Keller, K. L., 2006 Marketing Management

[2] Buckley, J., Pharmacuetical Marketing, Time for change

[3] Petel, H., 2003 Analyze – Focus – Optimise Roes Seminar

[4] closerlook.com/thinking/make_a_better_impression

[5] prwatch.org/node/7026, 2008

[6] Jayanthi, R.,   2005-2007 Application of Data Mining Techniques in Pharmaceutical Industry

[7] Freundlich, N., 2010 "DM" of Prescription and Patient Records

[8] Cohen, J. J., Data Mining of Market Knowledge in the Pharmaceutical Industry

[9] Jiawei, H. and Micheline, K.,   2011 Data Mining Concepts and Techniques

[10] Ahmed, S. R., 2004 Application of data mining in retail business Information Technology : Coding and Computing

[11] Berry, M. J. & Linoff, G. S., 2004 Data Mining techniques second edition – for marketing, sales and customer relationship management

[12] Koh, H. C. and Tan, G.,   2005 Data Mining Applications in Healthcare

[13] Bellazzi, R. , Zupan, B., 2008 Predictive data mining in clinical medicine : Current issues

[14] Mirjana, P. B., Dijana, C., 2008 Data mining usage in health care management : literature survey and decision tree application