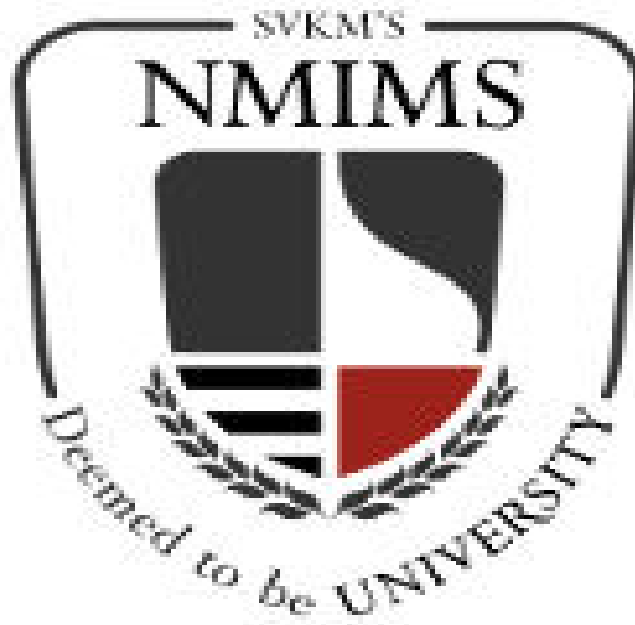


# Applications of Operations Research and Data Mining Techniques in the Healthcare Sector



Written by: Students of Anil Surendra Modi School of Commerce, NMIMS, Mumbai

Anirudh Jindal E006

Anusha Phadke E007

Anushka Sharma E008

Ashwin Sharma E009

Asmeet Kaur E010

Under the Guidance of Professor Richa Saxena

## **ABSTRACT**

Data mining has been extensively by many sectors across the globe. In the Healthcare Industry it is not only beneficial to patients but also aids owners, employers, employees, insurance companies and multiple other parties involved in the same. From our research we observed that Data mining applications can help in managing patient files/records, making customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services. Furthermore, it was seen that huge amounts of data generated by healthcare transactions are too complex and voluminous to *be* processed and analysed by traditional methods. Here is where Operations Research techniques come in the picture. Operations Research provides the methodology and technology to transform these mounds of data into useful information for decision making. Hence, the combination of using Operations Research and Data Mining in the Healthcare Industry has become exceedingly popular if not necessary. We thus wanted to do an in-depth study of this combination in the Healthcare Industry.

## **OBJECTIVE**

To understand the synergies between Operations Research and Data Mining in the Healthcare Industry

## INTRODUCTION

Operations Research is a subject/concept that deals with the application of analytical methods to help make versatile and accurate decisions. It is used in a variety of fields be it education, hospitality, airlines, management, manufacturing etc.

When it comes to Healthcare, Operations Research can be used in assisting patient care(primary function of a hospital) but also in deciding on the locations of medical clinics and emergency vehicles for giving adequate medical coverage to whosoever needs it, making estimating the number,of medical ambulances needed if the total distance from the locations to the hospitals must be less than a given number,planning chemotherapyof a cancer patient so as to minimise the treatment time of him/ her,maintaining a balanced staff-rooster to endure proper functioning of the clinic/centre/hospital and a variety of life-impacting things amongst other things.

Optimisation of such problems in Healthcare have received considerable attention during latter half of the 20<sup>th</sup> century but the combination of Data Mining and Operations Research reached its peak in the mid-2000s with increasing number of pressing and relevant issues like a growing ageing population, decreasing birth rates in most of the developed countries and increasing longevity globally. Moreover, enormous public and private funds required to cover the rapidly escalating Healthcare costs also necessitate a much closer scrutiny for cost-saving measures.

Data mining (DM), that is the process of observing patterns in large amount data involving methods of machine learning, statistics, and database systems ,has seen a boom in interest in many fields of applications including Healthcare, Defence, Government Policy making etc. The primary reasons for the same can be cited as increasing amount of data available, the growing understanding that deeper analyses are far more valuable than simple summary statistics and advent of technology. Data mining is characterized by the inference of general laws from particular instances. Data mining problems raise interesting challenges for several research domains, such as statistics, information theory, databases, and also for Operations Research (OR), since very large search spaces of solutions and questions that need to be answered.

The human-eye or manual observation and analysis of patterns from data has occurred for centuries. Early methods of identifying patterns in data which includes Bayes' theorem (18<sup>th</sup> century) and regression analysis. Data Mining came into existence in the middle of 1990's and appeared as a powerful tool that is suitable for discovering unknown pattern and useful information from large amount of data.

For example, earlier without data mining it was almost impossible to find important statistics or data about the patient such as estimating the length of stay of patients in the hospital, effective medical diagnosis finding costs and bills etc. but after data mining came into existence, it proved a great tool for the hospitals and healthcare sector in general to estimate certain stats and handling large amount of data.

The amount of data to be analysed has been increasing day by day. Hence, for several years, numerous research efforts using Operational Research methods to solve data mining problems have been reported. The synergy between operations research (OR) and Data mining (DM) is not a one-way street and three kinds of synergies can be observed:

- Operations Research quicken to the proper implementation of Data mining techniques
- A variety of avenues can come under Operations Research by using Data Mining
- Increased system performance can result from complementary uses of these two research domains.

There is an upward trend observed in the amount of electronic health records or EHRs being collected by healthcare facilities. It has been the norm for medical staff to take responsibility in handling patient data input that were traditionally recorded in paper-based forms. Accuracy and speed is very vital because it involves monetary resources and more importantly, lives of beings. But its humanly impossible to swiftly go about this data without any errors or omissions. Earlier, roughly 100-150 years ago, a lot of deaths, injuries and unnecessary monetary expenditure occurred because of data redundancy.

This is the area where data mining has turned out to be to a great tool. It has been utilized to reveal different patterns from stored data and after that it is used to build predictive models. Since the mid 90s, this training has been utilized to help with fraud recognition, credit scoring and maintenance scheduling but now its importance is finally being recognized all across the globe in the field of healthcare.

Enhancing the nature of patient care and decreasing healthcare costs are the main objectives of many projects. Data mining has helped these projects succeed. Primary benefits of the combination of Data Mining and Operations Research in the Healthcare Sector:

- Patients receive more affordable and better healthcare services because of proper resource allocation.
- This combination also helps in finding out new, more efficient practices and treatments. These methods run comparison amongst symptoms, causes, treatments and negative effects and then judge and analyse which course will provide the best facility for both the patient and the medical practitioner.

- Insurance companies are able to easily detect medical insurance abuse because of data. Out of the ordinary claims patterns are easier to find out with this synergy and it can identify incorrect referrals and fraudulent claims. Thus, the cost of healthcare services also reduces.
- Patient satisfaction and respect for healthcare practitioners is improved because data mining provides information that will help staff with patient interactions by identifying different usage patterns, present and future needs, and there preferences.

## **APPLICATION OF OPERATIONS RESEARCH IN HEALTHCARE SECTOR**

The health care sector has to make a lot of decisions for its operation like locations for clinics, the number of ambulances required for emergency cases, the number of patients per room and so on.

### **A. All These Issues Can Be Addressed With the Application of Operation Research.**

The growth in ageing population from decreasing birth rates and increase in longevity globally has made it very important to optimise healthcare sector issues. The economic problem of scarce resources has made it essential for health care sector to allot the right resources at the right time in various scenarios. For example: the increase in requirements of funds has made it necessary to ensure that cost has been allocated to the right area and in the right quantity. Service planning, resource scheduling, logistics, medical therapeutics, disease diagnosis and preventive care are the key healthcare optimisation issues.

To begin with, health care planning is very essential. The application of Operation Research enables demands forecasting, selection of location of clinics, hospitals as well as emergency vehicles and capacity planning. Such planning gives estimates for the unavoidable hospital costs per unit of measurable output. How these costs can be decreased can be found by identifying and eliminating inefficiencies in the planning and execution process. These inefficiencies may arise in simplest of situations such as allocations of beds.

Without Operations Research, it may be almost impossible to implement the right management technique. The most extensively referenced management problems are the scheduling of patients and resources along with logistics in healthcare sector.

To address such management problems there is a study of resources requirement and inventory assessment. The inventory needs are based upon the importance of the product or service provided.

### **B. Healthcare Practices is Another Field Developed Using Operation Research**

Drug treatment planning, infectious disease prevention and infectious disease control, epidemic, emergency response and organ donation are areas of healthcare in which advancements have been made. Asthma policy model for example is a system developed using Operations Research to forecast asthma related symptoms.

Prevention of a lot of diseases can be done with the application of OR techniques. For example, for the prevention of HIV, funds are not divided appropriately the costs after implementation may shoot up.

## TECHNIQUES OF DATA MINING AND OPERATIONS RESEARCH IN HEALTHCARE SECTOR

### A. Classification Techniques

Classification techniques are one of the most popularly used methods of data mining in healthcare sector. It mainly divides samples into target classes of each kind. With the help of this method, various risk factors can be associated with the patient by analyzing their pattern of disease.

The two methods of classification are 1. Binary 2. Multilevel. In the binary method, the patient can be divided into only two classes i.e. 'high' risk and 'low' risk patient whereas in the multilevel model, the patient can be classified into multiple classes such as 'high medium' and 'high low' risk patient. Further, the data is divided into training and testing dataset. This involves predicting a certain outcome based on the given input. For example, Breast cancer is one of the most dangerous diseases in women. Potter et al. performed an experiment on the breast cancer data set using Weka tool and then analyze the performance of different classifier using 10-fold

cross validation method. The research work revealed that there is no single best algorithm which yields a better result for every dataset.

Classification techniques are mainly used for predicting the treatment cost of healthcare services which increases significantly along with rapid growth of the humankind.

**Following are the various classification algorithms which are used in the healthcare sector –**

- K-Nearest Neighbor - K-Nearest Neighbour (K-NN) classifier is one of the simplest classifier that discovers the unrecognized data point using the previous known points (nearest neighbour) and then it classifies data points according to a voting system. K-NN has a number of uses in many areas like health datasets, image field, cluster analysis, pattern recognition, online marketing etc.
- Decision Tree - DT is similar to the flowchart in which every non-leaf nodes denotes a test on a particular attribute and every branch denotes an outcome of that test. The node at the top most labels in the tree is called root node. For example, we have a financial institution decision tree which is used to

decide that a person must grant the loan or not. Building a decision for any problem doesn't need any type of domain knowledge.

- Support Vector Machine - SVMs were initially developed for binary classification but it could be effectively and efficiently extended for multiclass problems. The support vector machine classifier creates a hyper plane or multiple hyper planes in a high dimensional space that is useful for regression and other efficient tasks like classification technique.
- Neural Network - It is an system or an algorithm for classification that uses gradient descent method and based on biological nervous system having multiple interrelated processing elements known as neurons, functioning in unity to solve specific problem. A neural network model is adaptive in nature because it customises its structure and adjusts its weight in order to minimize the error.
- Bayesian Methods - The classification based on bayes theory is known as Bayesian classification. It is a simple classifier which is achieved by using classification algorithm.

## **B. Regression**

It is used to find out different sets that explain the correlation among different sets of variables. A mathematical model is constructed using the dataset. In statistical modelling two kinds of variables are used. The first one is called dependent variable namely 'X' and other one is called independent variable namely 'Y'. There is only one dependent variable while independent variable may be one or more than one. Regression is a statistical method which investigates relationships between variables.

Based on number of independent variables regression is of two types, one is Linear and another one is Non-linear. Linear regression identifies relation of a dependent variable and one or more independent variables. Logistic regression, a type of non-linear regression can accept categorical data and predicts the probability of occurrence using logic function. Logistic regression is of two types, one is Binomial and other is multinomial.



### **C. Clustering**

Clustering is another method that is completely different from classification as unlike classification it has no predefined classes. In clustering large database are separated into the form of small different subgroups or clusters. Clustering approach is used to identify similarities between different data points. Each data points within the same cluster are having greater similarity as compare to the data points belongs to other cluster. Clustering needs very minimal information for analyzing the data. So, it is mainly used for analyzing small data because very little details are available for genes.

There are 3 types of clustering:

- i. Partitioned Clustering
- ii. Hierarchical Clustering
- iii. Density Based Clustering

### **D. Association**

Association is one of the most important approach of data mining that is used to find out relationships among a set of data items in the data repository. It is also known as market basket analysis due to its capability of discovering the association among unknown patterns of sales of customers in a transaction database. For example, if a customer is buying a computer then the chance of buying a computer protection software is high. This pattern helps the storekeeper to further enhance their sales. Association also has great impact in the healthcare field to detect the relationships among diseases, health state and symptoms.

Healthcare organization widely used Association approach for discovering patterns between various diseases and drugs. It is also used for detecting fraud in health insurance. Association is also used along with the classification method to increase the analysis capability of data mining. This approach is also helpful for identifying the improper prescriptions, irregular patterns in medical claims made by physicians, patients, hospitals etc.

There are 2 types of association:

- i. Apriori Algorithm
- ii. Frequent Pattern Tree Algorithm

## LIMITATIONS OF DATA MINING

- Data mining and operations research applications can be greatly beneficial but at the same time suffer from limitations as well. Since, raw inputs for data mining often exist in different setting and systems, such as administration, clinic, laboratories and more, they have to be collected and integrated, thus making accessibility of the data the biggest limitation. This is costly and time consuming as well. This limitation can be overcome through data warehouses, as then by 'Intermountain Healthcare'. They built a warehouse from five sources – a clinical data repository, acute care case mix system, laboratory information system, ambulatory case mix system and health plans database. Another organization is 'Maccabi Healthcare Services' who used existing databases to guide subsequent data mining.
- Another limitation is the about the content of data, which may be corrupted, non consistent or non-relevant data, such as information recorded in different formats in different data sources. These issues arise as a result of the volume, complexity and heterogeneity of medical data. Further there may be ethical, legal and social issues such as data ownership and privacy issues related to healthcare data. In the end, the quality of data mining results and successful applications depend on the quality of data.
- Drawing inferences from mining of data will certainly yield some patterns which are a product of random fluctuations. This is especially true for large data sets with many variables. Hence not all patterns and relationships found are useful. This is true for all cases where data mining is involved. The successful application of data mining requires knowledge of the domain area as well as the data mining techniques and tools. Without sufficient knowledge in both aspects, the user may not be aware of or be able to avoid the pitfalls of data mining.
- Finally, creating data mining and operations inquire about applications utilize generous investment of resources in terms of time, money and effort. It requires escalated arranging and innovative arrangement work. Also, doctors and administrators must be persuaded of the handiness of information mining and change work forms. Further, all gatherings associated with the information mining need to team up and coordinate.

## **FUTURE DIRECTIONS**

Data mining techniques in healthcare services can have a lot of potential and helpfulness. Accessibility of clean healthcare services information is essential in opening this potential. In this regard, it is important that these healthcare services industry consider how information can be better stored and captured, put away, arranged and mined. In this manner, one conceivable course might be the standardization of clinical vocabulary and the sharing of information crosswise over associations to upgrade the advantages of medicinal services data mining applications.

Further, healthcare data also includes qualitative data such as physician's remarks. To make the results more accurate, the use of text mining needs to be explored so that we can integrate data and text mining. Also, it will be useful if digital diagnostic images can be brought into healthcare data mining applications.

A few specialists believe the chances to enhance care and decrease costs simultaneously could apply to as much as 30% of general healthcare spending. This could be a win/win in general. In any case, because of the complex nature of medicinal services and a slower rate of innovation selection, our industry lags behind these others in executing powerful data mining and analytic strategies.

## CONCLUSION

Data mining though has great importance in the area of medicine and a lot of scope , for now it is just an academic exercise with only a few success stories . Approaches like trees , clusters and neural networks and time series are being used by academicians to publish research papers but no practical implementation. However, the field of Healthcare has always been slow to incorporate latest researches into daily practises.

The field of medicine will reach great heights with proper implementation of data mining and researches. Knowledge gained with the use of data mining techniques can be used to make successful decisions and can lead to improvement and success of healthcare sector. The possibility for research of hidden patterns in the data sets is where the primary potential of these techniques lies. These patterns can be used for clinical diagnosis. However, the issue is that the available raw material data is widely distributed voluminous by nature. The data must be collected and stored in data warehouses which can then be used for integration with the hospital information systems. Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments.

Healthcare institutions get a customer oriented approach towards new and hidden pattern in data. Through this knowledge is produced. This knowledge is then used to provide help in medical and others services to the patients. Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments.

The prediction of diseases using Data Mining applications is a challenging task but it drastically reduces the human effort and it also increases the diagnostic accuracy. With the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in the medical data.

## REFERENCES

- [1]. Benko, A. & Wilson, B. (2003). Online decision support gives plans an edge. *Managed Healthcare Executive*, 13(5), 20.
- [2]. Biafore, S. (1999). Predictive solutions bring more power to decision makers. *Health Management Technology*, 20(10), 12-14.
- [3]. Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC, Health Insights; 2012.
- [4]. Christy, T. (1997). Analytical tools help health firms fight fraud. *Insurance & Technology*, 22(3), 22-26.
- [5]. Dembosky A: “Data Prescription for Better Healthcare.” *Financial Times*, December 12, 2012, p. 19; 2012. Available from: <http://www.ft.com/intl/cms/s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz2W9cuwajK>. Feldman B, Martin EM, Skotnes T: “Big Data in Healthcare Hype and Hope.”
- [6]. Fernandes L, O’Connor M, Weaver V: Big data, bigger outcomes. *J AHIMA*, 2012:38–42.
- [7]. Gillespie, G. (2000). There’s gold in them thar’ databases. *Health Data Management*, 8(11), 40-52.
- [8]. IHTT: Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry; 2013. <http://ihealthtran.com/>
- [9]. Kincade, K. (1998). Data mining: digging for healthcare gold. *Insurance & Technology*, 23(2), IM2-IM7.
- [10]. L Delesie Operations research and knowledge discovery: a data mining method applied to health care management *International Transactions in Operational Research*, Volume 7, Issue 2, Pages 159-170
- [11]. Mark A. Lawley, Departments of Industrial & Systems Engineering, Texas A&M University, US
- [12]. Milley, A. (2000). Healthcare and data mining. *Health Management Technology*, 21(8), 44-47. October 2012. Dr. Bonnie 360; 2012. <http://www.west-info.eu/files/big-data-inhealthcare.pdf>.
- [13]. S. Ólafsson Department of Industrial Engineering, Iowa State University, Ames, IA 50010, USA  
Silver, M. Sakata, T. Su, H.C. Herman, C. Dolins, S.B. & O’Shea, M.J. (2001). Case study: how to apply data mining techniques in a healthcare data Warehouse. *Journal of Healthcare Information Management*, 15(2), 155-164.
- [14]. Xiaolan Xie, Centre for Biomedical & Healthcare Engineering & CNRS UMR 6158 LIMOS, Ecole Nationale Supérieure des Mines de Saint Etienne, France