

Implementation of Data Mining Techniques in CRM of Pharmaceutical Industry

Durga Sadanand Tembhurne¹, Prof. Jayant Adhikari², Prof. Rajesh Babu³

¹M.Tech Scholar, Department of Computer Science and Engineering Tulsiramji Gaikwad-Patil College of Engineering and Technology Nagpur, Maharashtra, India

^{2,3} Dept. of Computer Science and Engineering Tulsiramji Gaikwad-Patil College of Engineering and Technology Nagpur, Maharashtra, India

Abstract: Customer relationship management (CRM) is a data mining technology, CRM is utilized in enterprise of medical industry and its applications for that it utilizes classification tree algorithm. The data flow in the pharmaceutical industry was simple and the application of technology was very quiet in last two decades. However, now a day's technology has become most vital part of the business processes, the process of transfer of information becomes more complicated. Today enhanced technology is being utilized to help the pharmaceutical firms manage their inventories and to develop new product and services. Here we present role of data mining (DM) in pharmaceutical industry. DM is utilized to extract and select essential features from large dataset because pharma data is growing very vastly. So general techniques are not capable to process that large data and it requires more time so utilized classification technique selection tree model and FP tree model to classify large dataset, select tree and FP tree classifier places each record in class. We also compare selection tree model with FP tree the comparison results show that FP tree is better than selection tree because it requires less time to classify and less memory.

Index Terms: Data mining; CRM, Selection tree algorithm, Frequent pattern algorithm.

I. Introduction

Data Mining is the process of getting information from large data sets through the utilization of algorithms and techniques drawn from the field of Statistics, Machine Learning (ML) and Data Base Management Systems (DBMS). Eventually data analysis methods often involve interpretation and manual work of information that is slow, costly.

Recently, number of professionals have guessed that revenue growth for the pharmaceutical company will slow from the healthy 12% rate to 5-6% rate and it has number of implications for the drug identification technologies companies. Most of the drug discovery technology industries are attempting to face this challenge through developing solutions that will force new drugs to "fail faster and safer." Although this is a vital goal and if realized, would definitely create value for the industry, these solutions often over-promise and unestimated the obstacles that stand in the path to lower clinical failure rates. Marketing systems focused on expanding incomes will be more persuading than those that address lessening cost. Exhibiting that advances will empower pharmaceutical organizations to all the more likely target and market to certain client portions will expand appropriation of that innovation, and will open the entryway for undertakings went for diminishing expenses and expanding clinical preliminaries throughput. The vitality of decision support in the delivery of managed healthcare can hardly be overemphasized. An assortment of choice help capacities will be vital to enhance the profitability of medicinal work force, examine care results, and ceaselessly refine care conveyance procedures to stay beneficial while holding the line on expenses and keeping up nature of consideration. Healthcare decision support is faced with the challenges of complex and diverse data and knowledge forms and tasks the absence of institutionalized phrasing contrasted with essential sciences, the stringent execution and precision prerequisites and the commonness of inheritance frameworks.

Data Mining, popularly known as knowledge discovery (KD) in huge data, enables firms and organizations to make proper decisions by assembling, accumulating, analyzing and accessing corporate data. It utilizes various tools like analytical processing tools, query and reporting tools, and Decision Support System (DSS) tools.

A. Data Mining Techniques

Pharma industries rely on decision-oriented, systemic selection models that enable the decision maker to evaluate the payoff that is expected to result from the implementation of a proposed selection tree model. Such models go beyond an examination of the size of the validity coefficient and take a host of problems such as capital budgeting and strategic results at the group and organizational levels. Many organizations generate millions of records about their new drugs discovered and its performance reports, etc. This data is a strategic resource. Now, making utilization of most of these strategic resources will lead to enhance the quality of pharma industries. The important steps in DM are Problem Definition, Knowledge acquisition, Data selection, Data Preprocessing, Analysis and Interpretation, Reporting. The techniques and methods in DM need brief mention to have better understanding.

B. Applications of Data Mining in The Pharmaceutical Industry

Most healthcare systems lack the appropriate information systems to generate reliable reports with respect to other information than purely financial and volume related statements. The management of pharma industry starts to recognize the relevance of the definition of drugs and products in relation to management information. In the turmoil between costs, care-results and patient satisfaction the right balance is required and can be found in upcoming information and communication technology. The delivery of healthcare has always been information intensive, and there are signs that the industry is recognizing the enhancing importance of information processing in the new managed care environment. Most mechanized frameworks are used as an apparatus for day by day work: they are centered on 'creation'. Every one of the information, which are utilized to keep the association running, operational information, are in these mechanized frameworks. These frameworks are likewise called heritage frameworks. There is a developing need to accomplish more with the information of an association than to utilize them for organization as it were. A great deal of data is covered up in the inheritance frameworks. This data can without much of a stretch be extricated. The majority of the occasions this is impossible straightforwardly from the inheritance frameworks, on the grounds that these are not work to address addresses that are eccentric. Research demonstrates that that fruitful choice frameworks enhanced with scientific arrangements are essential for social insurance data frameworks. Given the span of the databases being questioned, there is probably going to be an exchange off in exactness of data and handling time. Examining strategies and trial of noteworthiness might be attractive to recognize a portion of the more typical connections; be that as it may, extraordinary connections may require generous inquiry time. The thoroughness of the search depends on the importance of the query, the indexing structures used, and the level of detail supplied in the query. Of course, the real data mining challenge comes when the user supplies only a minimal amount of information. For example: find possible serious side effects involving food and any type or brand of antacid. A user-interface may be designed to accept all kinds of information from the user (e.g., weight, sex, age, foods consumed, reactions reported, dosage, length of usage). Then, based upon the information in the databases and the relevant data entered by the user, a list of warnings or known reactions should be reported. Note that user profiles can contain large amounts of information, and efficient and effective data mining tools need to be developed to probe the databases for relevant information. Secondly, the patient's profile should be recorded along with any adverse reactions reported by the patient, so that future correlations can be reported. Over time, the databases will become much larger, and interaction data for existing medicines will become more complete. The amount of existing pharmaceutical information (pharmacological properties, dosages, contraindications, warnings, etc.) is enormous; however, this fact reflects the number of medicines on the market, rather than an abundance of detailed information about each product. One of the major problems with pharmaceutical data is actually a lack of information. For example, a food and drug administration department estimated that only about 1% of serious events are reported to the food and drug administration department. Fear of litigation may be a contributing factor; however, most health care providers simply don't have the time to fill out reports of possible adverse drug reactions. Furthermore, it is expensive and time-consuming for pharmaceutical companies to perform a thorough job of data collection, especially when most of the information is not required by law. Finally, one should note that the food and drug administration department does not require manufacturers to test new medicines for potential interactions. There are in general three stages of drug development namely finding of new drugs, development tests and predicts drug behavior, clinical trials test the drug in humans and commercialization takes drug and sells it to likely consumers.

DM applications in health can have huge potential what's more, convenience. In any case, the accomplishment of healthcare services DM depends on the accessibility of clean social insurance information. In this regard, it is important that the human services industry investigate how information can be better caught, put away, readied what's more, mined. Conceivable bearings incorporate the institutionalization of clinical vocabulary and the sharing of information crosswise over associations to upgrade the advantages of social insurance DM applications

C. Data Mining Process Steps

- 1) Identification of business problems: This is most important step of DM also it needs to find the business situation and also needs to know every data related to industry. Otherwise DM could not able to asses result.
- 2) Establishment of DM database: Database is not necessary in data mining but by maintain it differently can help for administrative purpose. And can help to study the old data warehouse.
- 3) Data analysis: This task is helpful in identifying the results.
- 4) Preparation data modeling: It consist of four steps i. variable selection: data variable choosing is vital because you cannot enter all variables at a time by doing so require large time to process. ii. Data selection: it is also inefficient to select data and use the whole data. iii. New variable construction: is also necessary iv. Adjust variable: according to need of data miner.
- 5) Model development: acquire correct analytical method and model according to issues.
- 6) Evaluation of model: For getting result apply the model.

D. Data Mining Algorithms and Techniques

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are utilized for knowledge discovery from databases.

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Selection tree model is utilized as classifier, is an extension of decision tree model. Its basic principle and analytical methods are similar to decision tree rather selection tree provides more choices of branches. Selection tree gives more accuracy than other models.

The FP-Tree is a compressed representation of the input. While reading the data source each transaction t is mapped to a path in the FP-Tree. As different transaction can have several items in common, their path may overlap. With this it is possible to compress the structure.

The Frequent Pattern (FP)-Growth method is used with databases and not with streams. The Apriori algorithm needs $n+1$ scans if a database is used, where n is the length of the longest pattern. By using the FP-Growth method, the number of scans of the entire database can be reduced to two.

The section I explains the Introduction of data mining and data mining techniques and its application in pharma. Section II presents the literature review of existing systems and Section III present proposed system implementation details which includes preprocessing and feature extraction, and Graph Evaluation Section IV presents experimental analysis, results and discussion of proposed system. Section V concludes our proposed system. While at the end list of references paper are presented.

II. Literature Review

The development of IT has generated huge amount of databases and huge data in numerous areas. The research in databases and IT has given rise to an approach to store and manipulate this precious data for further decision making. DM is a process of extraction of useful data and patterns from huge data [1]. It is also known as knowledge discovery (KD) process, knowledge mining from data, knowledge extraction or pattern analysis. To generate information, it requires massive collection of data. The data can be simple numerical figures and text documents, to more complex information such as spatial data, multimedia data, and hypertext documents.

Jiawei Han [2] Discussed some key term regarding DM like what is DM, what kind of data is used in it, its functionalities, what kind of patterns are mined in DM, system classification, issues present in it, evaluation structure of database, author says that DM is referred as knowledge Discovery in Databases (KDD). Present DM classification based on kinds of databases mined, he kinds of knowledge mined, and the kinds of techniques utilized to mine it.

U.M.Fayyad [3] presented fundamental issues of knowledge discovery, clustering and classification, trend and analysis, dependency derivation, integrated discovery systems, augmented database systems, and application, nowa day we have seen large data is generated called as big data and for managing that data needs new tools and techniques that can guide us to transfer this data in useful knowledge. Advancement in Knowledge Discovery and Data Mining carries the latest research in databases, machine learning, and artificial intelligence that are part of the exciting and rapidly growing field of knowledge discovery and data mining.

The healthcare data are not limited to simply quantitative information, for instance, doctors' notes or clinical records, it is vital to additionally investigate the utilization of content mining to grow the extension and nature of what social insurance information mining can right now do. In specific, it is helpful to probably coordinate information and content mining [5]. It is additionally valuable to investigate how advanced indicative pictures can be brought into human services information mining applications. Some advancement has been made in these regions [6][7].

The delivery of health care has dependably been data serious, also, there are signs that the business is perceiving the expanding significance of data handling in the new overseen care condition [8].

DM applications likewise can advantage medicinal healthcare suppliers, for instance emergency clinics, facilities and doctors, and patients, for instance, by distinguishing powerful medications and best practices [9],[10] The Centers for Medicare also, Medicaid Services has utilized DM to build up an imminent installment framework for inpatient restoration [11] The human services industry can profit extraordinarily from information mining applications.

Pharmaceutical industries can profit by social insurance CRM furthermore, information mining, as well. By following which doctors endorse which drugs and for what purposes, pharmaceutical organizations can choose whom to target, show what is the most economical or on the other hand best treatment plan for an infirmity, help recognize doctors whose rehearses are fit to explicit clinical preliminaries [12]. Pharmaceutical companies can utilize data mining techniques to colossal masses of genomic information to anticipate how a patient's hereditary cosmetics decides his or her reaction to a medication treatment [13].

DM applications can be developed to evaluate the effectiveness of medical treatments. By comparing the causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective [14]. Now days, Sierra Health Services has utilized DM broadly to classify zones for quality upgrades, counting treatment rules, disease management groups, and cost administration [15].

A UI might be intended to acknowledge a wide range of data from the client (e.g., weight, sex, age, foods consumed, reactions occurred, dosage, length of use). At that point, in light of the data in the databases and the significant information entered by the client, a rundown of admonitions or known reactions ought to be accounted for. Note that client profiles can contain a lot of data, and productive and successful DM devices should be created to test the databases for applicable data. Second, the patient's profile ought to be recorded along with any unfavorable responses detailed by the patient, with the goal that future connections can be accounted for. After some time, the databases will turn out to be a lot bigger, and cooperation information for existing prescriptions will turn out to be increasingly finished [16].

III. System Architecture

Here in Fig.1 dataset of drug or pharma is taken as an input after that applied preprocessing on that to extract features. Among the extracted features vital features are selected for further processing after that selection tree algorithm is applied and then Frequent Pattern algorithm utilized for tree modeling. After that evaluation graph are generated.

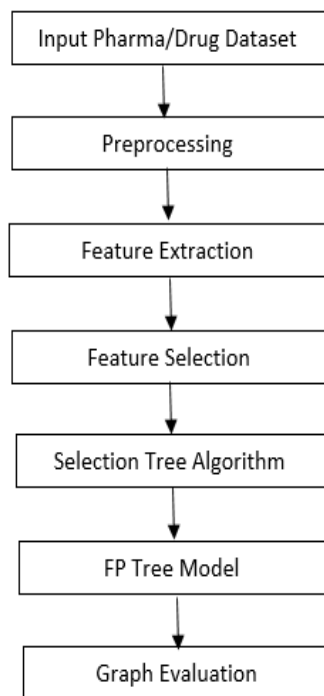


Fig 1. System Architecture

IV. Result And Discussions

E. Experimental Setup

I. All the experimental cases are implemented in Java in conjunction with Netbeans tools and MySQL as backend, algorithms and strategies, and the competing classification approach along with various feature extraction technique, and run in environment with System having configuration of Intel Core i5-6200U, 2.30 GHz Windows 10 (64 bit) machine with 8GB of RAM

II. Dataset

Drug or pharma dataset downloaded from koggle website consist the product code, size, product name, unit and number of prescriptions.

F. Comparison Results

This section presents the performance of the existing system that is selection tree model and proposed system FP tree model algorithms in terms of time required and memory. Fig 2 Shows Memory Comparison of selection tree and FP tree model algorithms for various Threshold. X-axis shows existing and proposed system that is Algorithm & Y-axis shows memory in bytes. Proposed system take less memory than existing system.

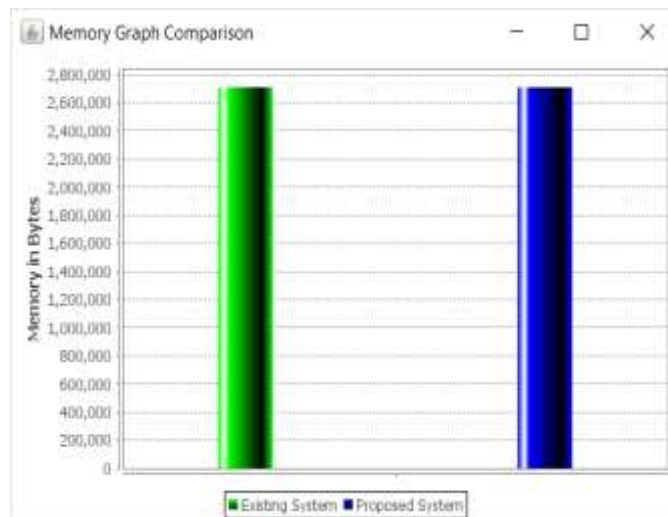


Fig. 2: Memory Comparison Graph

Fig 3 shows the Time comparison of existing and proposed system for various size. The X-axis shows algorithms and Y- axis shows Time in ms. The proposed system takes less time than existing system for classifying large dataset.

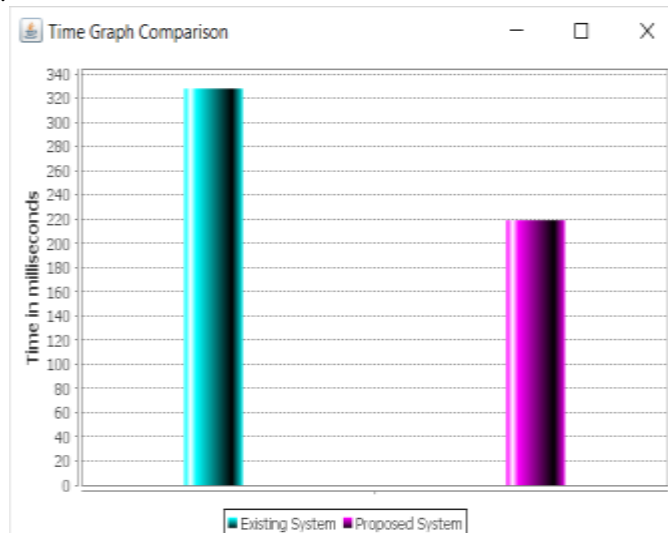


Fig. 3: Time Comparison Graph

V. Conclusion

Here we discussed how data mining techniques utilized in pharmaceutical industry and why it is utilized also we present some application where DM utilized. Basically DM is used in field of healthcare system to identify revenue of medicine also by using data mining we can gather huge data and separate useful information among them. It helps to gather drug data like which drug is generally more prescribed in which area or location. It can help to detect symptoms causes and analyze which course is required to cure the patient. So pharmaceutical company uses CRM and data mining for making profit by identifying which doctors prescribe which drug and for what purpose according to that they target them to generate revenue.

References

- [1]. song guijuan "Application of Data Mining Techniques in the CRM of Pharmaceutical Industry" 2018 *International Conference on Intelligent Transportation, Big Data & Smart City*
- [2]. JiaweiHan, "Data Mining: Concept and Technology Beijing: Machinery Industry" Press, 2012.8.(in Chinese)
- [3]. U.M.Fayyad, G.Piatetsky,P.Smyth and R.Uthurusamy, "Advances in Knowledge Discovery and Data Mining", *AAAI/MT Press*,1996.
- [4]. He rongqin, "CRM Principle, Design, Implementation, Beijing: Machinery Industry", Press 2002(in Chinese).
- [5]. Cody, W.F. Kreulen, J.T. Krishna, V. & Spangler, "The integration of business intelligence and knowledge management." *IBM Systems Journal*, 41(4), 697-713 W.S. (2002).
- [6]. Ceusters, "Medical natural language understanding as a supporting technology for data mining in healthcare". In *MedicalData Mining and Knowledge Discovery*, Cios, K. J. (Ed.), Physica-Verlag Heidelberg, New York, 41-69. W. (2001).
- [7]. Megalooikonomou, V. & Herskovits, "Mining structurefunction associations in a brain image database". In *MedicalData Mining and Knowledge Discovery*, Cios, K. J. (Ed.), Physica-Verlag Heidelberg, New York, 153-180 E.H. (2001).
- [8]. Morrissey, "Managed care steers info systems", *Modern Healthcare*, Vol. 25 No. 8 J. (1995).
- [9]. Gillespie, "There's gold in them thar' databases." *Health Data Management*, 8(11), 40-52. G. (2000).
- [10]. Kolar, "Caring for healthcare". *Health Management Technology*, 22(4), 46-47 H.R. (2001).
- [11]. Relles, D. Ridgeway, G. & Carter, "Data mining and the implementation of a prospective payment system for inpatientRehabilitation". *Health Services & Outcomes Research Methodology*,3(3-4), 247-266 G. (2002).
- [12]. Brannigan, "Quintiles seeks mother lode in health data mining." *Wall Street Journal*, March 2, 1 M. (1999).
- [13]. Thompson, W. & Roberson, "Making predictive medicine possible". *R & D*, 42(6), E4-E6, M. (2000).
- [14]. Kincade.K, "Data mining: digging for healthcare gold". *Insurance & Technology*, 23(2), IM2-IM7, (1998).
- [15]. Schuerenberg, "An information excavation. *Health Data Management*", 11(6), 80-82, B.K. (2003).
- [16]. JayanthiRanjan, "Data mining in pharma sector: Benefits", *International Journal of health care quality assurance*, 22(1), 2009, pp. 82-92.