A Survey Paper on Data Mining Techniques in Drug Industry

Nithya Jojen
St Joseph's College,
Irinjalakuda

Abstract--Data mining helps to transform data into meaningful knowledge. Data Mining offers the potential for much deeper analysis and predictions in the field of medicines and health. Modern medicine generates a great deal of information stored in the medical database. Taking useful knowledge and giving scientific decision-making about the use of medicines and side effects from the database increasingly becomes necessary. Data mining in medicine will be helpful in this problem. In this paper we discuss about different data mining techniques used in pharmaceutical field to collect the relevant information. For detecting underlying undetected information about drugs data mining techniques are an effective tool.

Keywords: Data mining, KNN, decision trees, clustering, Neural network, classification.

I. INTRODUCTION

Pharmaceutical industry is growing day by day; this vast frowning industry also contains tremendous data which can be converted to useful knowledge. Data mining always provides a way to uncover the underlying data. Data mining contains several techniques and tools so that we can effectively collect data. Data mining helps to collect valuable information about drug development testing and selling. This paper discuss about how data mining helpful in collection of useful data and how it can be applied. This paper also shows how data mining improves the quality of decision making process in medical world. Production of drugs or medicines includes a huge detail of data so it can only be extracted using several data mining techniques.

II. LITERATURE SURVEY

Literature survey describes the existing and established theory and research in your report area by providing a context for your work. This survey shows where you are filling a perceived gap in the existing theory or knowledge [0]

A. Applications of data mining techniques in pharmaceutical industry

This paper includes several data mining techniques used for collecting data in pharmaceutical field. This paper also provides data mining applications used in pharmaceutical field. this paper is successful to provide information about the importance of data mining in pharmaceutical industry, but this paper explains only few data mining techniques and also applications described in drug field is limited.

This paper is efficient to show that data mining provides way to improve drug development. It covers the basic information about the concept of data mining and its various techniques applied to various fields so there by implementing data mining. [1]

B. A survey on data mining approaches for health care

This paper tries to provide different approaches for data mining in health care. This paper explains several data mining techniques such as KNN, decision trees, clustering which can be effectively applied for collecting health care information. This survey aims to search out several approaches for data mining. This paper classifies data mining techniques with its advantages and disadvantages. This survey also highlights applications, challenges and future issues of Data Mining in healthcare. Recommendation regarding the suitable choice of available Data Mining technique is also discussed in this paper. This paper is effective in showing challenges in health care, but this paper does not provide data about drugs. This paper fails to suggest a single data mining techniques which give consistent results for all types of healthcare data. [2]

C. A review on data mining approach for adverse drug reaction research

This paper provides a survey on a vast variety of computational methods on ADRs. In particular, this paper mainly discusses sources of data and their characteristics, as well as a classification of computational methods based on data and techniques used, including their advantages and limitations. The challenges and further opportunities in this area are also provided. This paper fails to explain data mining techniques that can be used for collection of information about adverse drug events. This paper only presents a survey of existing solution for ADR analysis. [3]

D. Top 10 algorithms in data mining

This paper provides several data mining algorithms that can be effectively used in data mining. This paper presents the top 10 data mining algorithms identified by the IEE. These algorithms can be added in any data mining techniques, but it does not provide any knowledge about drugs, and its adverse effects. This paper is efficient to describe several data mining algorithms such as: kNN (k-nearest neighbor classification), CART, the EM algorithm, The Apriori algorithm, and the k-means algorithms and also provide limitations and relevance of these algorithms. [4]

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTPPTDM-2015 Conference Proceedings

E. Data mining in medicines production system

This paper outlines the application of data mining in production of medicines. Pharmaceutical industry is a huge industry which manufactures drugs in various formats like pills tablets, and Ointment. Production of drugs or medicines includes a huge deal with various kinds of data. This paper provides only information about data mining in production of drugs alone. It does not provide applications used in pharmaceutical industry This paper gives several data mining techniques in drugs but it does not say anything about adverse effects of drugs

The data set to be changed continuously as the changes occurs in market and generations of the medicines. So the data provided by this paper may be upgraded and invalidated. [5]

F. Scope of Data Mining in Medicine

This paper gives a brief review of scope of data mining in medicines this paper also submits the related works and included discussions on this topic; this paper describes features of data mining but fails to explain data mining techniques that can be used in medical field. And this paper does not contain the applications of data Ming in medical field.[6]

III. DATA MINING TECHNIQUES

Pharma industries need to depend on decision oriented and systematic models for the improvement of drug development. Many of the models produces generate a bulk of data about drugs, its adverse effects and so on. Making use of this data we can enhance the quality of drugs. Mainly steps involved in data mining are giving by [teeldors, Daniels and hdsheimer, 2000]: problem definition, knowledge acquiring, data selection, data pre processing, analysis and interpretation, and reporting the use.

The techniques used mainly in field of drug industry are:

A. Neural network based classification schemes

These are very useful in data mining. classifications are made on unorganized data this method helps to easily identify complex relationships between dependent and independent variables and able to collect data from bulk database.

B. Clustering:

In clustering grouping of similar record are done. This technique helps to bring data into a single field and helps to draw easy conclusions from data.

C. classification and prediction:

Helps in explaining about data classes and also helps to declassify data and to evaluate whether the given data is efficient or not. prediction helps for predicting the future behavior of data, it helps to know that whether a product will be relevant or not.

d) Decision trees: helps to minimize the confusions in decision making process and provide correct values to results of various events.

It helps to manage large information of data. Easy to construct and helps to draw really relevant information.

IV. RELEVANCE OF DATA MINING IN DRUG INDUSTRY

The data from past successes and failures are collected and try to verify whether it is relevant in future or not. in drug industry if we collect data about drugs we can see that there will be only a few drugs is efficient, we can analysis the data and check how many drugs is needed more in market. Data mining in drug industry helps to know the most efficient and demanding drugs in market, new drug combinations, drug effectiveness and its adverse effect, how to efficiently use drugs, mainly demanding drug companies, innovations of new drugs, most top shelled drugs, frequently used drugs and so on

V. APPLICATIONS OF DATA MINING IN THE DRUG INDUSTRY

The pharmacy industry needs to have a relevant and meaningful knowledge to improve the quality of drugs and for essaying task for innovations of drugs but unfortunately they lack these data's ,so information's are collected from patients and from other reports we evaluate the data. A data can be collected from patients by creating user interfaces, questionnaires and interviews. data mining is helpful in field of drug creation, testing and in drug test on humans.

A. Drug creation:

Using classification, neural networking and clustering drug creation will be easy. Using clustering we can draw a conclusion that which all molecules help for the effectiveness of drugs, we can as Also know which part of these molecules are reacting adversely and helps to avoid such molecules. Data mining also helps to know which the real molecules efficient in medicines are
.

B. Drug testing:

By verifying the reports we can know that whether the drug has adverse effect or not, at warehousing helps to achieve this. The data mining technique association is also helpful for identifying improper prescriptions, irregular or fake patterns in medical fields made by doctors and patients. Classifications are also useful to examine the effect of drugs.

C. Clinical trial testing of drugs in humans

By collecting data from pharmaceutical companies and patients we can derive conclusions. When we find out many people reporting negative effects about drug then we can conclude that drug is poisonous and need to eliminate from industries**.**

VI. PROPOSED SYSTEM

The proposed system is helpful to know data mining techniques in drug field and it also provides the

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTPPTDM-2015 Conference Proceedings

applications and relevance of data mining in drug field. Using K-NN, association, SVM and decision trees data mining techniques we can collect more data in area of pharmacy industry. Using association technique we can collect the data of wrong prescriptions and claims of doctors and patients. The proposed system also gives sufficient details about different techniques used in data mining and shows some of the difficulties that can be happened in data mining and how to be resolved. Pharmacy industry is large growing area and using these data mining techniques it can have a comparative study on drugs and convert this knowledge to meaningful results or outcomes. The proposed system shows how data mining is helpful in each stages of drug development, testing and using. The proposed system collects adverse drug information from various sources. It includes all possible data mining techniques that can be used in data mining in pharmacy industry. The proposed system tries to combine all data mining technique into a powerful technique so that we can find a general result which may helpful to achieve great heights in drug development

A. Predicting the nature of drug behavior

The reaction of drugs can be predicted befor development of drugs, using k-nearest neighbor(K-NN)we will get some data about previously created drugs which contains some of the molecules which is including in new drugs. So data mining will help to determine whether to add a molecule predicting its behavior. Support vector machine (SVM) is also helpful to analyze data based on statistical learning.

B. To check validity of prescriptions and claims

Using association data mining technique proposed system helps to find whether the prescription of doctor is correct or not, this technique also helps to fake claims of patients.

VII. FUTURE WORK

The data from drug industry grows large day by day and it gets updated too, so large data bases are required for storing these information and more powerful data mining techniques are to be introduced for effective data collection. A method need to be introduced for bringing all the data under a single field from which we can generate a favorable result. We can use hybrid or integrated Data Mining technique such as fusion of different classifiers, fusion of clustering with classification or association with clustering or classification etc. for achieving better data mining in this field.

VIII. CONCLUSION

Data mining is always helpful for extracting information and this paper describes several data mining techniques which are helpful for discovering and try to gain data about medicines. the data techniques such as :Clustering, classifications, neural networks all these provide an effective technique for collecting data from huge drug data bases. Several other methods are also proposed that could increase efficiency of data mining on drugs information. Data mining is important in each phase of drug development, testing, and in marketing. This paper summarized overall content and several methods for implementing data mining in pharmaceutical field. This paper gives several applications of data mining in medicines. Production of drugs or medicines includes a huge detail of data so data mining techniques need to be so powerful to categorize this information's in to a single unit and once these are grouped then we can found out great meaningful knowledge. The technique of data mining depends on the type of data set that we have taken for doing experiment. Using drug data mining it becomes possible to improve the quality of drugs and increase the efficiency of drugs.

ACKNOWLEDGEMENT

REFERENCES

[1] http://mycourse.solent.ac.uk/mod/book/view.php?id=2744&chapterid=1295
[2] "Ajanta Rajang"-Applications of data mining techniques in pharmaceutical industry
[3] "Diva Tamar and Somali Agawam"- A survey on data mining approaches for health care International Journal of Bio-Science and Bio-Technology, Vol.5, No.5 (2013).
[4] A review on data mining approach for adverse drug reaction research Siriwon Taewijit-Tu Bao Ho-Thanaruk Theeramunkong
[5] "XindongWu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang ·Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu ·Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg"Top 10 algorithms in data mining
[6] " Moral Kothari , Priti Sadaria "Data mining in medicines production system – tablet As a product
[7] "Divdeep Singh Sukhpreet Kaur"- Scope of Data Mining in Medicine