Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Importance of Big Data in the Pharma Industry

Syed Shoeb Peer[1], K. Samunnisa[2], A. David Donald[3]

[1,2,,3]Asst. Professor, Ashoka Women's Engineering College, Kurnool

Article Info
Page Number: 8289 - 8294
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received*: 15 September 2022*
Revised*: 25 October 2022*
Accepted*: 14 November 2022*
Publication*: 21 December 2022*

Abstract

The term "big data" was first used to describe information sets that were too enormous to be captured, stored, managed, and analysed by conventional database systems. The concept of big data analytics is being implemented in many different industries today Analytics for large datasets are used to glean insights for businesses. Using big data analytics, one can gain insight into topics like market movements, consumer preferences, and other similar phenomena. Not only that, but this sort of analytics also makes important contributions to the decision-making procedure in business contexts. Step-by-step approaches exist for analysing huge data, and there are various approaches to choose from. The purpose of this essay is to introduce the concept of data analytics, explore its many manifestations, and investigate how it may be useful in the selection of generic medications.

Index Terms: Big Data, Pharma.

1. Introduction

The pharmaceutical industry is a data-intensive sector, with data on generic medications being generated by every facet of the business at an exponential rate. This generic medicine information was produced utilising data from numerous sources, such as wholesalers, hospitals, retailers, medical reps, doctors, and the orange book. Insightful use and analysis of this data can help pharmaceutical firms find the most suitable generic drug with the desired therapeutic effect, speeding up the time it takes to develop these compounds into safe and legally marketable medicines. The pharmaceutical industry[1] will surely gain from making excellent use of big data analysis tools[2] in order to make decisions on the correct drug selection, as the process of generating new drugs is costly and requires a substantial amount of time and money.

2. Big Data

Analyzing massive amounts of heterogeneous data sets using cutting-edge analytic techniques is what is meant by "big data analytics." With the help of big data analytics, valuable information can be gleaned from massive databases including records about pharmaceutical drugs. The size of such data sets can be on the order of a few hundred gigabytes to a few thousand petabytes, or even more. Data mining [3] refers to the process of collecting, modifying, and analysing large amounts of data[4] "Diverse data" means that the information can be in any of several different formats, including structured, semi-structured, or unstructured data. In order to do an analysis that calls for the use of certain analytical procedures, it is possible to use big data analytics on medication data because it is available in so many different formats.

Fig.1 types of Big Data Analytics[5]

Big data analytics can be broken down into four distinct subcategories, due to the fact that different kinds of data necessitate unique methods. These subcategories are as follows:

1. Descriptive Analytics

2. Diagnostic Analytics

3. Predictive Analytics

4. Prescriptive Analytics

2.1 Descriptive Analytics: It sheds light on what has gone down in the past and reveals patterns that can be explored further. It simplifies the material while also condensing it into an easier-to-read

format. Descriptive analytics, such as clustering and association algorithms and summary statistics, find utility in applications such as market basket analysis. One of the most important aims of the pharmaceutical medication selection process is the analysis of sales trends, which may be uncovered with the help of descriptive analytics and used to create predictions about the medicine's future sales before it is manufactured.

2.2 Diagnostic Analytics: The purpose of this research is to diagnose the source of the issue, as the name implies. It's an investigation into the root causes of a problem. The process entails digging deep into your data to unearth insights. In order to help determine the root cause of a problem, diagnostic analytics looks at historical data on the issue at hand. It is widely employed to aid in spotting patterns among clients. Diagnostic analytics may take the form of, for instance, a customer churn reason analysis, a customer health score analysis, a data mining and data recovery analytics, or a drill-down analytics[6].

2.3 Predictive Analytics: Predictive analytics is the practise of making predictions about future outcomes and performance using statistical and modelling techniques. In order to predict what will happen in the future, this analysis method considers past and present information. Artificial intelligence (AI), also known as machine learning, data mining, and algorithmic analysis, accomplishes this by applying these concepts to the task of reviewing existing data and generating predictions about the future. Examples of predictive analysis include weather forecasting, finding out what customers like and don't like, predicting market trends, and so on. The ability to anticipate a client's needs is another. In order to make the best decision on which generic drug to use, we will employ this predictive method here.

Data analytics makes use of a wide variety of prediction models, including but not limited to the following:

i. **Forecast Model:** In this model, we use the available past data to make educated guesses about the present and future. It makes a prediction based on information provided by a number of variables. For instance, it factors in historical purchase data and uses it and other information to make projections about the near future[7].

ii. **Classification Model:** The data used in these models is organised into categories based on historical information. The best way to respond to questions with only yes or no answers, as well as providing wide analyses that can help guide decisions, is with the help of classification models. Questions like, "What industry, retail or hospital, is most suited for the drug?" can be answered by these models.

iii. **Clustering Model:** This approach classifies the information into groups with similar features. To work, things need to be put into categories based on the characteristics or actions they share, and from there, more comprehensive plans may be made for each category. E.g. By integrating these features, we are able to draw certain conclusions from the input and make a prediction about the cost and timing of the generic drug.

iv. **Outliers Model:** When applied to out-of-the-ordinary data points, the outliers' model yields useful results. It can identify odd numbers independently or in relation to other numerical or categorical data.

v. **Time Series Model:** Developing a numerical measure that may predict trends within a specific time frame requires time as an input component. This method can be useful when trying to establish a release date for the product.

2.4 Prescriptive Analytics: It combines the strengths of both descriptive and predictive methods of analysis. Using the available information, it determines which of several possible approaches to a problem has the greatest chance of success and implements that strategy. It's more than just a monitoring system because it gives you actionable insights. In a perfect world, it would provide a model that incorporates various possible solutions and enable organisations to make decisions based on optimising the consequences of future events or hazards. Businesses have to make a choice, and they need to make it based on how well it fits their needs and how effective it is. Statistical models with roots in mathematics and computing are crucial to this type of research. Algorithms and models created from AI and ML are used in this study to help businesses make decisions based on the connections and patterns found in their data.

3. Data Analysis

The focus of this paper will be on predictive analytics, which was determined after considering several types of big data analytics. Now that those foundational steps have been taken, we may proceed with further data analysis while bearing the aforementioned factors in mind.

Fig.2 Data analysis

i. **Define Problem:** Pharmaceutical firms are huge producers of data about generic medicines. These generic drugs are rushed into circulation without first doing any kind of market study. The cost of research and development for new pharmaceuticals is likewise high. Due to inaccurate drug testing and inaccurate market research, the company has sustained a variety of financial losses. Businesses can prevent these problems by using data analytical processes, which will enable them to choose the most suitable generic product.

ii. **Data Collection:** To collect, quantify, and assess data using tried-and-true methods. New ways for collecting and analysing massive volumes of data have emerged in recent years due to the challenges inherent in loading enormous datasets into a conventional relational database. The primary data utilised in this study came from a number of different distributors, while the secondary data came from a wide range of resources, including the internet, an orange book (a reference book containing information on all pharmaceuticals), historical data or patterns related to drugs, and so on.

iii. **Data Cleaning:** Before diving into the analysis of the data to draw conclusions, this data cleaning phase is essential. It is a method used to rectify a dataset by eliminating any and all instances of inaccurate, redundant, or otherwise unusable information. The results and error percentage may suffer if the data are of low quality. Before moving on to additional data analysis, it is important to clean the data so that it is as accurate and consistent as possible. Deleting null or duplicate entries, removing unnecessary columns, renaming columns, handling missing values, and finding outliers are all essential steps in the data cleaning process. ([6]).

iv. **Data Analysis:** This method involves manipulating the data in various ways to extract actionable intelligence. Several methods can be used to analyse the data. With the help of these analysis methods, it is much simpler to comprehend the data, make sense of it, and draw meaningful conclusions from it. Different data visualisation approaches are used, each of which presents the data graphically, to aid with comprehension. Statistical approaches such as correlation and regression are only two examples of the many available for establishing connections between data points. Here, we'll look at some of the open-source and structured data analysis tools that can help us analyse huge data in a structured fashion.

4. Conclusion

The goal of this work is to examine the methods used in big data analytics. Furthermore, it highlights the various data analytics methods. Furthermore, it stresses the numerous processes that must be carried out when conducting analysis. It discovers a variety of predictive models and provides a selection of predictive models that a pharmaceutical firm may find useful when deciding which generic medications to produce.

References:

[1] "how-is-big-data-impacting-the-pharmaceutical-industry@ www.iqviamedicalsalescareers.com." .
[2] "2ad04b0a6b4ee305859b1fd4cbd720584d224638 @ careerfoundry.com." .
[3] "data-mining @ www.ibm.com." .
[4] "What-Is-Big-Data @ Www.Sas.Com." .
[5] Q. Rida, "A Roadmap Towards Big Data Opportunities, Emerging Issues and Hadoop as a Solution," *Int. J. Educ. Manag. Eng.*, vol. 10, no. 4, pp. 8–17, 2020.
[6] "353d6773bd0143d07262c073881139c7300fd46b @ www.doxee.com." .
[7] S. Snapp, "Supply Chain Forecasting Software," no. July, 2012.