e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.10, Oct 2022

Evaluating the Data Mining Technology in big data in the Retail Industry

Mrs Prajakta Joshi
Assistant Professor
prajakta.joshi@lsraheja.org

Abstract:

Big data is not just a large data in amount but it is also growing exponentially with time. Big data is based on the concept of 3 Vs i.e. Volume, Velocity & Variety. As specified in definition it is huge in terms of Volume & new data is received at fast rate. The big data is also available in various format. This data need not be structured, but at times it is semi structured or unstructured. This type of data also needs processing to make it useful. They are so voluminous that traditional data processing software just can't manage such huge data. To put simply, Big data contains greater variety, arrives in increasing volumes and is ever-changing frequently.

The world's data collection is reaching at threshold point, due to the arrival of companies like Facebook, YouTube & other such organisation providing online services that generates humongous data. We can presume that the foundation stone for modern Big Data was laid already. This data can be accessed in several ways which is really interesting. In today's era, data is known as the new oil. These huge loads of data are rich in information, but they need to be analysed to extract correct information, which can answer a lot of questions.

In this paper I would discuss applications of data mining in retail business. I would also evaluate, its pros and cons. I would also elaborate concept of Big data technologies in data Mining & its types. The Paper would also throw light on New Emerging technologies of Big Data.

Key Words: Big Data, Volume, Data Storage, Data Mining

1. Introduction:

1.1Data mining - concept & Definition

Data Mining is the process of sorting through large data sets to identify patterns and relationships that can help to solve business problems through data analysis. It extracts previously unknown, valid and accountable information from large databases and then using the information to make crucial business decisions. The various data mining techniques and tools allow us to predict future trends and make more informed decisions.

Data mining is a crucial component of data analytics, the information it generates is truly more valuable than oil. Data mining (DM) and knowledge discovery are intelligent tools that can help to accumulate and process data and make use of it.[6]

Think of it in this way, let's say you want to build a car.

- You need to build a chassis.
- This will be made predominantly using aluminium.

Now if you don't mine this metal from its source, you don't gather and process it, then none of the above is possible.

This was a small example to try and give you a gist of what data mining is and why we need it.

1.2 Big Data: Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently.[3]

Volume, Variety & Velocity are 3 Vs of Big Data. But 2 more Vs are introduced now that is Value & Veracity.

Data Mining Process

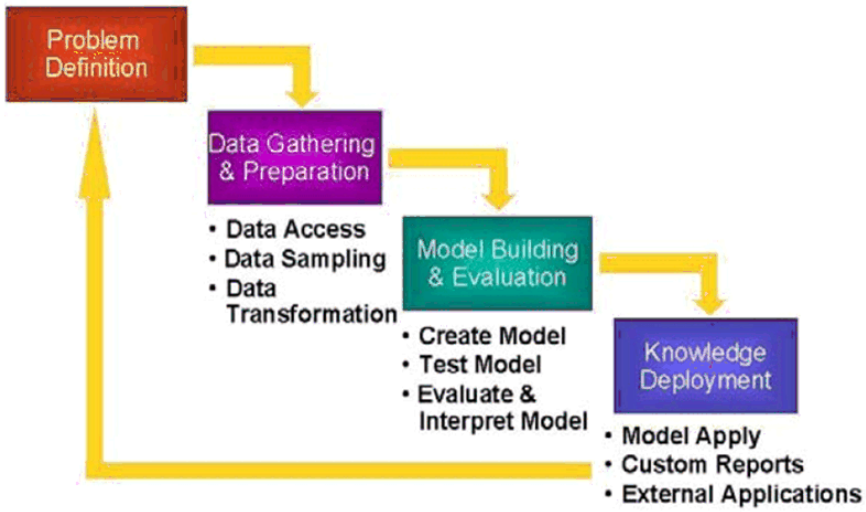

Fig1 Stages of Data Mining (https://docs.oracle.com/cd/E18283_01/datamine.112/e16808/process.htm)

Problem Definition[7]

This preliminary phase of a data mining project emphases on understanding the project objectives and requirements. After finalising the project from a business perspective, one can frame it as a data mining problem and develop a preliminary implementation plan.

Data Gathering and Preparation[7]

The data understanding phase involves data assemblage and exploration. If one can observe the data closely, can define how perfectly it points to the business problem. In this phase you can also identify data quality issues and to scan for patterns in the data.

Model Building and Evaluation[7]

After completing the two phases, now one can select and apply various modelling techniques and regulate the parameters to optimum values. If the algorithm requires data transformations, you will need to step back to the previous phase to implement them.

Knowledge Deployment[7]

Knowledge deployment is the practice of data mining within a target environment. In this phase, perception and actionable information can be derived from data.

Data Mining in Retail Industry

The retail industry witnesses a massive application of data mining. The data this industry deals with is very diverse, from the user's personal information i.e., location, shopping history toinventory of products, marketing activities to the sales figures on YTD basis etc. holds a lot of information.

Retail data mining helps determine shopping patterns, buying behaviours and current trends which are crucial in making data driven decisions, such as targeted ads, Promotional schemes, discount offering, inventory management, incentive schemes etc.

Now let's see this from a different perspective, how does it help the retailer?

In a dynamic industry which grows manifold each year, the consumption increases proportionally and so does the amount of data. This data is like a complex network of caves, with useful pieces of information creeped into the deepest crevices, data mining helps to find and analyse them for further decision-making. Data mining depends on effective data collection, warehousing, and computer processing.[2]

Advantages of Data Mining in Retail Industry[11]

- **It helps gather reliable information –** Data mining allows businesses, industries.\, organisations, and governments to gather reliable information. It can be used in marketing research to determine what products customers might be interested in and then make those products available to them.
- **Helps to make informed decisions –** It is often used for business purposes to improve decision making. As more data is collected, the accuracy of data mining becomes greater.
- **Helps to analyse very large quantities of data quickly –** Data mining can be used to analyse data that was previously too difficult to understand due to the sheer volume or type of information. Moreover, it is an important part of the modern world and most companies use it on a regular basis because it helps them to make more informed decisions about marketing and other business activities.
- **It helps detect risks and fraud –** Data mining helps to recognise risks and fraud that may not be noticeable through traditional tools of data analysis. Data Mining can discover patterns in data whichmight be difficult to expose, particularly when the data is not organised in a way that makes it easy to know what type of information to look for.

A few applications are:

Customer Relationship Management

- CRM focuses on acquiring and retaining customers, increasing customer loyalty, obtaining customer knowledge, and putting customer-centred initiatives into action. Adopting a business model that is truly centred on the needs of the client may help your company achieve operational excellence, new growth, and competitive agility.
- Customer relationship management (CRM) is a business idea that focuses on identifying, comprehending, and better serving your customers while building a relationship with each one in order to boost customer satisfaction and maximise profits. It entails identifying, anticipating, and satisfying customer demands.

- To manage the relationship with the client, a business must collect the pertinent information about its customers and organise that information for efficient analysis and action.

Market Basket Analysis[10]

- It is used to look into the inherent affinities between different things. One of the most well-known applications of market basket analysis is the beer-diaper affinity, which claims that men who buy diapers are likely to also buy beer.
- Market basket analysis can, however, become highly complex in practice and point out previously undetected affinities between a range of commodities. This analysis can be applied in many different ways by the retail organisation.
- Product placement in stores is one extremely common use. Another typical use is product bundling, which is the process of grouping things to be offered as a single package deal. Additional uses include creating the company's web store and product catalogues.

Customer Acquisition and Retention

- Data mining might also help retailers retain and draw in customers. The fiercely competitive retail industry can benefit from data mining by better understanding the needs of its customers. By looking at customers' past shopping behaviours, retailers may target them with the relevant promotions and incentives.
- Data mining can also be used to look at past purchasing habits of customers who have left a business and switched to competitors, using this information to deter new customers from doing the same.

Design and construction of Data Warehouses

- Due to the wide range of issues covered by retail data, there are several ways to build a data warehouse for the industry (such as sales, customers, staff, products transit, consumption, and services).
- The level of detail that should be provided can also differ substantially. The results of the initial data mining operations can be used as a guide when developing and building a data warehouse architecture. Effective data mining will be made possible by choosing the dimensions, levels, and pre-processing to use.

Multidimensional Analysis

- The retail sector needed current knowledge of customer demands, product sales, fashion trends, and styles, as well as the quality, cost, profit margins, and level of service of commodities.
- It is essential to provide flexible tools for multidimensional analysis and visualisation, such as the ability to build complicated data cubes in response to the demands of data analysis.

Sales Campaigns[10]

- Advertisements, coupons, and various discounts and incentives are used in retail market sales campaigns to promote products and attract customers. By carefully analysing the effectiveness of your sales campaigns, you can increase your business' revenue.

- Multi-dimensional analysis can be used to achieve these goals by comparing sales volume and various transactions of the sales item during the sales period with those containing the same before and after the sales campaign.

Minimize the Risk

- One of the roles of retail data mining is retail risk management. In contrast to other areas of retail, this requires minimal learning. Retailers use data mining techniques to identify items that are vulnerable to market risks and changes in consumer buying patterns.
- A customer's past buying behaviour is taken into account when assessing brand loyalty. By using data mining techniques to understand consumer behaviour, retailers can adjust tactics to stay competitive in the market and reduce the risk of loss. Using data mining techniques, retailers can target customers who are likely to purchase a particular brand of merchandise, and decide when and where to advertise as needed.

Minimize Fraud Cases

- Detecting retail fraud is essential to maintaining business success. Retailers are often concerned about fraud at the point of sale, and data mining can help prevent this.
- Retailers work hard to expose dishonest employees. In addition to POS data mining, multiple supermarkets use his CCTV camera system. These strategies eliminate the need for store managers to gather evidence to convince employees of the theft. Data mining tools can be used to extract suspicious transactions and CCTV video can be used to determine exactly what happened during the transaction. You don't have to be physically at the store to do this. Everything can be done from the office.

Top Big Data Technologies:[8][9]

Big data technology is defined as software-utility[9]. This technology is mainly crafted to analyse, process and mine information from a large data set and a huge set of extremely complex structures. This is very difficult for traditional data processing software to deal with.

Among the larger concepts of rage in technology, big data technologies are widely associated with many other technologies such as deep learning, machine learning, artificial intelligence (AI), and Internet of Things (IoT) that are massively augmented.[9]

The technologies are categorised into main four sections:

- Data Storage
- Data Mining
- Data Analytics
- Data virtualization

In this paper we will be discussing on Data Mining Technologies.

Technologies used for Data Mining in Big Data[8][9]

Data Mining is the process of extracting required data from a very huge source of data which is increasing very rapidly. To extract the information from this data is nearly impossible without special techniques like:[8]

- Presto
- RapidMiner
- ElasticSearch etc.

Presto

It is developed by Facebook, Presto is an open-source SQL query engine which allows interactive query analyses on massive quantities of data. This distributed search engine supports fast analytics queries on data sources of various sizes, from gigabytes to petabytes. With this technology, it is possible to query data right where it lives, without moving the data into separate analytics systems. It is possible even to query data from multiple sources within a single query.[8] Presto is a Java-based query engine that was developed in 2013 by the **Apache Software Foundation.**[9]

RapidMiner

RapidMiner is an advanced open-source data mining tool for predictive analytics. It's a powerful data science platform that lets data scientists and big data analysts analyze their data quickly. In addition to data mining, it enables model deployment and model operation. RapidMiner is a Java-based centralized solution developed in 2001 by **Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer** at the Technical University of Dortmund's AI unit.[9]

Elasticsearch

Built on Apache Lucene, Elasticsearch is an open-source, distributed, modern search and analytics engine that allows you to search, index, and analyze data of all types. Some of its most common use cases include log analytics, security intelligence, operational intelligence, full-text search, and business analytics. Unstructured data from various sources is retrieved and stored in a format that is highly optimized for language-based searches.ElasticSearch is primarily written in a Java programming language and was developed in 2010 by **Shay Banon**[9]

Limitations of Data Mining[11]

- **Data Mining tools are complex and require training to use –** Data analytics is a complicated process and often requires people with training to use the tools. The barrier to entry for data analytics can discourage small businesses from using this technology. It can also be difficult to find adequate data that isn't already private or proprietary in nature.
- **Rising privacy concerns –** One of the major disadvantages of data mining are data and privacy concerns. Traditionally, companies would only share personal data with other companies in order to provide a service. Nowadays, many people are worried that their personal information is being sold to third-parties without their knowledge.
- **Expensive –** Data mining can be a very expensive process. For example, companies have to hire additional employees and technology specialists to ensure that the data mining is done correctly.

Conclusion

An era of big data technology has given rise to various new innovations that are likely to gain popularity since the industry's demand has increased. These innovations will serve as a catalyst for business development. Overall, the future of Big Data looks promising.

In this article, we have seen a whole host of big data technologies including Apache Hadoop, Apache Spark, MongoDB, Cassandra, Plotly, and more. These technologies help with storing, mining, analyzing, and visualizing big data. Nevertheless, before settling on a big data tool or technique, it's important to conduct thorough research because each tool or technique has its own unique features and can be applied to specific businesses. In order to

make the most out of Big Data technologies that are available on the market, it is essential to identify the type of problems your organization faces. Here's your chance to make the move you want, based on your requirements so that you would get customized solution for your organization. Hopefully, this article will assist the reader in navigating Big Data technologies without getting lost.

References:

1. https://www.oracle.com/in/big-data/what-is-big-data/
2. https://www.investopedia.com/terms/d/datamining.asp
3. https://www.guru99.com/what-is-big-data.html
4. Raju, P. S., Bai, D. V. R., & Chaitanya, G. K. (2014). Data mining: Techniques for enhancing customer relationship management in banking and retail industries. International Journal of Innovative Research in Computer and Communication Engineering, 2(1), 2650-2657.
5. Pal, J. K. (2011). Usefulness and applications of data mining in extracting information from different perspectives. Annals of Library and Information Studies (ISSN: 0972-5423), 58(1), 7-16.
6. Samson, G., &Hammawa, M. B. (2011). Applying data mining research methodologies on information systems. *Oriental Journal of Computer Science and Technology*, *4*(2), 241-251.
7. https://docs.oracle.com/cd/E18283_01/datamine.112/e16808/process.htm
8. https://www.interviewbit.com/blog/big-data-technologies/
9. https://www.javatpoint.com/big-data-technologies
10. https://analyticssteps.com/blogs/8-applications-data-mining-retail
11. https://www.aplustopper.com/data-mining-advantages-and-disadvantages/