**Impact of Data-mining and Data Warehousing on Business**

**[Name of the Writer]**

**[Name of the Institution]**

The process of studying big databases that already exist for generating new information is called Data-mining. This helps in discovering patterns, anomalies, and correlations in data by using software(s). It is also called knowledge discovery in databases (KDD). This process is used by companies so that raw data can be made useful by learning about their customers and their needs/ preferences and form better and effective marketing strategies targeted towards those customers and consequently generate higher profits at minimized costs. Data mining is also useful in finance and banking to minimize the risks of loans, mortgages, and detection of fraudulent transactions. Data mining is useful for understanding customers’ preferences and business objectives by taking into account the available resources and given constraints and other important factors. It is a complex and difficult task as the data from different sources usually do not coincide with each other and poses challenges in the process of data integration. So companies use “metadata” to omit the errors that occur in the process. The data is aggregated, generalized (e.g. replacing a city with a country) and normalized (create ranges) for easy analysis and obtaining the results.

Data mining is used in education, manufacturing, communications, insurance, retail, E-commerce, crime investigation, supermarkets, and bioinformatics.

Data mining can be obtained from Object-oriented and object-relational databases, Transactional and Spatial databases, Heterogeneous and legacy databases, Multimedia and streaming database, Text databases and Text mining and Web mining. Based on the objectives of the business, relevant mathematical techniques are used for generating the results and it’s important to make sure that the results are understood and studied by all stakeholders to meet data mining goals.

The main steps in data mining are classification, clustering, regression and association of data that then predict the trends, patterns and possible future events.

There are challenges for business for the implementation for data mining as skillful and expert people are needed to work on the data in question. Sometimes the data is too small for the test and large databases are difficult to handle. The nature of data also mattes as for accurate analysis, diverse data is required.

The tools used in data mining are **R-language (statistics and graphics) and** Oracle Data Mining (for predictions).

Data warehousing is the main component of business intelligence and influences the decision-making process of the business managers. It allows space for data to be stored that is required by the business to report and analyze and for other business intelligence functions. This helps in controlling the stock and inventory. This helps in providing the customers with the desired goods and services and to keep the products up to date for future use. In data warehousing, the data is used at many levels. For example, the user of data may want to see the total sales of a good/ service in a community, city, state and the country. Usually, it starts from a higher level; and goes down to lower levels.

The usage of data warehouse is called **Online Analytical Processing** (**OLAP**), which uses complex queries to analyze rather than process transactions. The main purpose of data warehousing is to see the connection b/w different data from various systems. The difference between database and data warehouse is that a data warehouse can store huge amount of historical data and uses OLAP to solve queries across the data while a database records the transactions and helps look at the details of the transactions if need be. It is called “Online Transaction Processing” (OLTP). A Data warehouse can be used to combine data from various places to form one database. It mitigates the difficulty of analysis in transaction process database and maintains the history of the data. It enables to see data from different sources from the same point. This is mainly useful when organizations merge. It also improves the data quality and even omits the errors in data. Data warehouse also helps in restructuring the data for use by the business doer as it gives better query fulfillment with no impact on the system of operations and adds value to CRMs. It enables easy writing of queries and removes the ambiguities by organizing recurring data.

There have been many pieces researches on the impact of data mining and data warehousing on different businesses. One such research suggests that businesses have gained important hints from the data gathered via such techniques/ systems by analyzing the results given by these systems. For many years, the use of such systems has led businesses to a new and interesting domain of business intelligence. This also enables better and focused market search, which further leads to a better understanding of customer behavior and preferences, consumer requirements, and to recognize and cease new investment and growth options. And this doesn’t stop here. New technologies and methods will further lead to even better and mature business intelligence and actionable news/ insights. This even points out that the next generation of employees will need even polished skills and abilities to equip themselves with the newest technologies and work in a more competitive business environment. It will have applications not only in business but also in healthcare, government, and security and there is a need for more research on this shift (Chen, Chiang, & Storey, 2012).

Another study clarifies that the technology of Data mining given by Cloud computing is an extremely important practice for any business to decide actively on the basis of actionable knowledge, so that they may have an idea on upcoming patterns and predict the outcomes of their decisions. The study shows that it is necessarily important and is of great benefits to use data mining in cloud computing. As the requirement for tools of data-mining tools is ever increasing, the need to integrate them in cloud computing is also becoming ever significant.

Another study revealed that the use of information systems like data warehousing has had a significant impact on the consumer satisfaction as they enable the business doers to look into the market and help them make important decisions and amendments in the way of doing business. This further helps the business grow and improve. The information systems greatly impact the performance of the business (Chen, Soliman, Mao, & Frolick, 2000).

Another study emphasizes on the designing of the systems as they are extremely valuable and helpful in decision making process. It suggests that the designing of the systems should be precise and thoughtful so that it may satisfy the needs to the best level. The systems are essential for business as nowadays no business can survive without business intelligence and market research. So the systems should be engineered I such a way that they provide accurate, precise authentic and actionable data trends and analysis for the business to grow and operate in a cost-effective mode (Golfarelli, Maio, & Rizzi, 1998).

Data warehousing has enabled companies to meet the need for better and accurate decision making by providing integrated, organized and readable data on the right time. This enables the business to survive in a highly competitive environment by identifying the opportunities and weaknesses and guiding them on the changes to be made to get better results. The information systems help in forming the strategies and their implementation and evaluating the results which further leads to innovation and beneficial customer relationships (March & Hevner, 2007).

The art of data mining has been studied by many researchers to better understand the process so that the possible mistakes be avoided and correct the ones already committed. The tasks that eat up most of the time in this process is the selection of data and its initial processing. However, much of the decisions are affected by the analysis of that data. The selection of data is its initial processing has a great influence on a company’s efficiency. The effective way of collecting and analyzing the data needs to involve the experts in the working, who have command over the business area to be studied. Although there are many technologies that enable automatic data collection and its analysis, there is still a need of experts who have the required knowledge of the process and are helpful in mitigating the risks associated with mistakes and errors (Feelders, Daniels, & Holsheimer, 2000).

There is a term “Churn Management” which refers to being able to identify valuable consumers and being able to retain them. This is noticeably important in the telecommunication industry. As mentioned earlier, the practice of data mining is essentially important in the telecommunication industry, while I looked at a study regarding churn management, it emphasized that as the economy is becoming digital day by day, churn management is even tougher for companies nowadays. This has led companies to formulate more consumer-focused plans to maximize profits and the developments in data mining have to play a key role in the process of decision making (Lejeune, 2001).

Information systems are playing a key role in hospitals and healthcare as well. A study on the use of information systems in hospitals in Greece has suggested that they’ll be beneficial for smooth and efficient outcomes for the hospitals if there is an encouragement for their efficient usage as well as the contentment level of the people using such systems. Many pieces of researche have been carried out to see the level of satisfaction of the users of hospital information systems and the factors that determine it. The research in this regard which I studied highlights that there is enough support for the usage of information systems and the EUCS model (which incorporates new factors) can gauge the satisfaction level of the users as it is robust and reasonable to use in different cultures and linguistic settings (Aggelidis & Chatzoglou, 2012). The research in this regards hints that more hospitals are using the information systems for better service provision.

As almost all the companies are using the systems for better performance, we also need to acknowledge that the security of information of various companies is of utmost importance. Companies are investing to secure their information systems. It is mainly concerned with the management of the organization. It has significance as some companies may sell their data of consumers to other companies in exchange for money which is not approved by the consumers and is ethically wrong as well. Some companies can hack the systems of other companies and obtain their data. A study on the topic of information security systems reveals that the company data cannot be protected merely by technical means. The study reveals that the users think that the security of information systems is essential but it makes it difficult to use the systems and the efficiency of the systems is compromised. Companies need to see the tradeoff between the benefits of securing the systems and the loss in productivity (Montesdioca & Maçada, 2015).

The process called “Volunteered Geographic Information” (VGI) is used to increase the capabilities of the analysts with the assumption that data is credible and precise. Research integrates the data of volunteered geographic information and spatial OLAP system and finds out that both the data sets resemble in certain ways and vary in other. The research provides a logical guideline for framing warehouse and OLAP data (VGI data). The purpose is to address the issues of accuracy and reliability of volunteer geographic information data. Using a Volunteer geographic information system, the researchers are also defining a newer spatio-multifaceted model which will help the business planners to describe the accuracy of the accumulated data (Bimonte, Boucelma, Machabert, & Sellami, 2014).

There is also a need to use good software for data mining. When I researched which is a good software to serve the purpose of data mining, I came up with the a list that includes the Sisense, Oracle Data Mining, Rapidminer, Microsoft SharePoint, IBM Cognos, KNIME, Dundas BI, Board, Orange, SAP Business Objects, Salesforce Analytics Cloud, DOMO, SPSS Modeler, Limestats and AtomLynx Insights Engine etc. All these software packages have distinct characteristics that the companies can use according to their needs. For example, the companies can use SQL developer by Oracle with a Data Miner GUI to amend the information readily in the database, which further helps in integrating the predictions to the apps so that the intelligence patterns can be used automatically.

The top companies that use data mining are the Amazon, American Express, BDO, Capital One, General Electric (GE), Miniclip, Netflix, Next Big Sound, Starbucks and, T-Mobile. All of these companies are highly competitive and efficient, the leaders in the industry in their domain and are doing great. The fact that all these big companies use data mining is proof that it has significant implications in decision-making process of the business. And data mining is essential for a business to operate efficiently and compete in the industry.

If we look for data warehouses in the market that are in high demand nowadays, we come up with the Teradata, Oracle, Amazon web services, Cloudera and the MarkLogic.

The leading companies like Apple Inc., eBay and Walmart have the biggest data warehouses one can see. The industries that use data warehouses are computer software, financial services, information technology and services, hospitals and healthcare, staffing and recruiting, banking, insurance, higher education, telecommunications and retail. This means that usage of data warehousing is huge and is ever increasing day by day.

The countries that top the list of data warehouse usage are the United States, United Kingdom, Canada, Australia, India, France, Germany, Netherlands, Italy and New Zealand. All these countries are the leaders in economy, science and technology. India is an emerging economy with the most favorable market for startups and entrepreneurships. This implies that the use of data warehouses is an undeniable reality and the businesses need to make arrangements for their needs of data and information if they aspire to operate efficiently with minimum costs and maximum profits.

All the above discussion proves that data mining and data warehousing is part and parcel of effective operation of a business and without this, the companies are surely unable to make efficient and beneficial decisions and are less likely to succeed in the respective industries. This also provides better opportunities to data miners and developers as well as poses a challenge as more and more companies are using the data mining and warehousing techniques, data miners need to keep themselves up to date and polish their skills so that they might be of help to businesses to grow and cater to the needs of the growing customer base.

**References**

Bimonte, S., Boucelma, O., Machabert, O., & Sellami, S. (2014). A new Spatial OLAP approach for the analysis of Volunteered Geographic Information. *Computers, Environment and Urban Systems*, *48*, 111–123. https://doi.org/10.1016/j.compenvurbsys.2014.07.006

Chen, Chiang, & Storey. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165. https://doi.org/10.2307/41703503

Aggelidis, V. P., & Chatzoglou, P. D. (2012). Hospital information systems: Measuring end user computing satisfaction (EUCS). *Journal of Biomedical Informatics*, *45*(3), 566–579. https://doi.org/10.1016/j.jbi.2012.02.009

March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, *43*(3), 1031–1043. https://doi.org/10.1016/j.dss.2005.05.029

Lejeune, M. A. P. M. (2001). Measuring the impact of data mining on churn management. *Internet Research*, *11*(5), 375–387. https://doi.org/10.1108/10662240110410183

Chen, L., Soliman, K. S., Mao, E., & Frolick, M. N. (2000). Measuring user satisfaction with data warehouses: an exploratory study. *Information & Management*, *37*(3), 103–110. https://doi.org/10.1016/S0378-7206(99)00042-7

Montesdioca, G. P. Z., & Maçada, A. C. G. (2015). Measuring user satisfaction with information security practices. *Computers & Security*, *48*, 267–280. https://doi.org/10.1016/j.cose.2014.10.015

Feelders, A., Daniels, H., & Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Information & Management*, *37*(5), 271–281. https://doi.org/10.1016/S0378-7206(99)00051-8

Golfarelli, M., Maio, D., & Rizzi, S. (1998). The dimensional fact model: a conceptual model for data warehouses. *International Journal of Cooperative Information Systems*, *07*(02n03), 215–247. https://doi.org/10.1142/S0218843098000118