

## Big Data Mining to Improve Decision Support Systems: A Review Paper

*Noor M. Al-Attar*

*Department of Information Systems*

*King Saud University, Riyadh, Saudi Arabia*

*Mehmet Sabih Aksoy*

*Department of Information Systems*

*King Saud University, Riyadh, Saudi Arabia*

**Published on: 6 Mar. 2023**



This work is licensed under a  
Creative Commons Attribution-  
NonCommercial 4.0  
International License.

### Abstract

With the great technological progress in the current era, especially in the areas of artificial intelligence and the facilities that these technologies can provide at the level of individuals, institutions and countries, the trend has become great toward developing artificial intelligence systems, including decision support systems that provide a study of many environment variables that experts can overlook. In the traditional study and analysis, and with the advent of the concept of big data resulting from the unprecedented increase in data and information that is

transmitted, especially with the spread of the Internet, the analysis of this data to extract knowledge, information and indicators from it has become a matter of great interest that affects the decision-making process at various levels and even super sensitive ones. The biggest problem that we may face is that big data has multiple formats (images - documents - HTML pages, etc.). In this research, we will study how to analyze big data and invest it optimally in decision support systems.

Keywords: Big Data, Decision Support Systems (DSS), Artificial Intelligence (AI), Big Data Mining

### \* **Introduction**

With the spread of information technology and its entry into various areas of life, health, commercial, economic, even military, etc., and with the advent of the Internet and the huge amounts of data it contains, in addition to the huge amount of data that is uploaded and received from the Internet (ZHIHAN LV, 2021), the need for mechanisms to study has been arising to analyze, coordinate and sort this data. This is characterized by different forms, sources and contents, so that information and knowledge can be extracted from it. This is known as Big Data Mining. This information and knowledge can be invested strongly by various parties to achieve their goals, for example, companies can by analyzing their customers' data determine their satisfaction with the products and services provided by this company at the present time, in addition to predicting their opinions regarding any new policies or products that will be offered through prospecting in a huge database that represents a set of features which characterizes the behavior of each

individual, it can be said about the process of benefiting from this data as decision support systems (Fyodor O. Fedin, 2019). It helps companies and stakeholders to choose what is the best and to report on various current and future policies, hence we find that decision support systems are auxiliary tools for producing decisions that may be correct according to the features taken into account.

### \* **Definitions**

**1- Big data:** It is a large group of data (Chaowei Yang, 2017). The most important thing that distinguishes it is the diversity of its forms and formats, as it can be images, numbers, audio recordings, web pages, etc. This data is huge, complex and difficult to analyze, process and store using traditional tools such as SQL.

**2- Decision support systems:** They are information systems based on artificial intelligence (Shaofeng Liu, 2009), as they represent the interaction between the human factor and computer systems, as they help to produce information and knowledge that benefit senior leaders and decision-makers in making informed decisions. Therefore the goal of these systems is not to make decisions on behalf of the human factor, but rather it

is a complementary tool in the decision-making process to develop proactive policies and anticipate customer behavior with changes that could occur.

**3- Big data mining:** A technique that aims to infer knowledge from huge amounts of data (Sowmya R, 2017). This technology is based on many sciences, such as statistics, logic science, computer science, and others. It aims to reach results, information, or knowledge that can be extracted from this data in a more effective and usable manner.

**\* Literature review**

Big data (Robert G. Aykroyd, 2019), relative to its name, is data of super sizes, and it is not possible to specify a maximum limit for these sizes due to the huge inflation in the world of the Internet and information. It can be said that data volumes of the rank of Exabyte EB or Zettabyte ZB are considered huge data, The most important characteristic of big data (Hai, 2022) are:-

- 1- Large size: The volume of huge data is very large, and therefore requires large processors and devices capable of handling this data.
- 2- Multiple types: As big data comes in many different shapes and formats

such as image, audio, video, and text (Reihaneh H. Hariri, 2019).

3- Multiple and credibility: Not all information and data received can be used in the service of the institution and decision-making. Therefore the useless part of it must be disposed, which is known as pre-processing.

4- Rapidly growing: As it swells and multiplies greatly as a result of active interaction with subjects by individuals, customers and beneficiaries, therefore the response must be rapid in order to use this data in serving the institution and achieving its goals.

5- Multiple manifestations: When using big data, it must be analyzed and presented in different forms commensurate with the nature of its use, which can take multiple forms such as statistics, numbers, geometric shapes, and others.

6- Data can be generally classified into three categories (Xindong Wu, 2014) :

7- Structured data, which is usually organized in a relational database, and therefore sorting and extracting knowledge from it is easier.

8- Semi-structured data, such as databases, which include separators between data elements, and the sorting process is of medium difficulty.

9- Unstructured data, which is random, such as texts, and the process of sorting and extracting information from it is difficult.

#### **\* Decision support systems in artificial intelligence**

They are systems that rely on analyzing and processing large amounts of data with the aim of extracting knowledge and results that can help leaders and opinion-makers in choosing the most appropriate option for the problem at hand. It is also useful in the comparison process between the options presented (Hasna El Alaoui El Abdallaoui, 2018), Steps of the decision support process (Anthony P. Sunjaya, 2022) are:-

1- Data is obtained from different data sources (internal within the company or external) according to the purpose of the system.

2- Since the collected data is heterogeneous, this data must first be cleaned (pre-processed) before storing it in the database. Entities are also specified.

3- Processing and analysis in order to identify the most useful characteristics and display the most relevant measures to be extracted, and then apply techniques to explore and classify the data on which the decision is based.

#### **\* Decision support systems stages**

The stage of collecting data on the problem and its decision: This is done by seeking to collect data and information on the subject for which the decision is taken. In addition, making an efficient decision depends on the amount of data and information available and their efficiency, relevance and representation of the topic (Wadii Boulila, 2017).

Data classification stage: This stage converts data into other forms such as reports, tables, and graphs so that it seeks to preserve appropriate information and data that add great value to the decision-making process through its close connection with the specified topic (Emrana Kabir Hashi, 2017).

Data validation stage: This stage will not have any importance or value if the data is not used in a timely manner for decision-making in the organization (Wellem Anselmus Teniwut, 2020).

The stage of measuring the resources needed by this decision (risk analysis): the success of the institution depends on the efficiency and effectiveness of its decision-making control, and according to the availability of quantities of data and the

comprehensiveness of information, and the accuracy of the timing in providing information that enhances that efficiency and effectiveness, as well as the availability of data and information in a large and effective manner, determines the ability of the institution to adapt and adapt to the external environment (Wellem Anselmus Teniwut, 2020).

Stage of differentiation between options (decision analysis): This stage is concerned with issuing the largest number of decisions, working to limit and analyze them, and then excluding decisions that do not achieve the goal and purpose of decision-making, leading to obtain implementable decisions that are compatible with the goals of the institution (Montserrat Acosta, 2017).

The decision-making stage: passing through the previous stages, comes the decisive stage of decision-making and declaring it to opinion leaders, and when it is applied and its results appear, the specialists evaluate the results to see the effectiveness of this decision and the extent of its success (Wellem Anselmus Teniwut, 2020).

## \* Discussion

From the view of decision makers, the importance of big data lies in its ability to provide important information and knowledge on which decisions can be based due to the detailed data it provides and from various sources such as scanners, mobile phones, loyalty cards, the web and social media platforms. The first stage of the AI decision-making process is to collect data that can be equipped to identify problems and opportunities from both internal and external data sources. During this stage, the sources of big data must be identified, as well as data from various sources must be collected, processed and stored. After identifying the sources of information and the types of information required for analysis, the selected data is obtained and stored in any of the big data storage and management tools. Afterward the data is cleaned and processed to move to the next stage, which is the design stage, where possible work paths are developed and analyzed through a conceptual visualization, or a representative model of the issue. This stage can be divided into three steps, model planning, data analysis, and conceptual analysis. The last stage in

the process of decision support is the implementation of the scheme that was established (Nikhil Madaan, 2020).

Big data analysis (Hossein Hassani, 2020) is an effective basic pillar in defining and crystallizing the problem, identifying alternatives, making a comparison between these alternatives, and then making the appropriate decision from among the available alternatives. We note that the value of the data does not stop at the decision-making only, but continues to know the consequences of the decision taken if it was applied and work to assess its efficiency. It serves the goals of the institution, as well as taking corrective measures when necessary.

Not all data is effective and of value in decision-making. There are a set of characteristics of data that decision-makers need for the system to be reliable, relevant to the subject, accurate, and timely (Marijn Janssen, 2016). When this data is available with these characteristics, data scientists analyze it and link it to the appropriate issues, and then present it in a simple and convincing template to decision-makers in order to make the appropriate logical decisions. Therefore, educational institutions and universities must qualify those who

have the ability to deal with big data and prepare scientists and specialists who have skills in designing databases and designing appropriate programs, having statistical skills, being able to express the data and information they deal with in order to process big data in all possible ways, and then present it to decision-makers.

The value of big data multiplies when the confidence and conviction of managers in the data used in decision-making doubles (Chunquan Li, 2021). The value of the data is also attributed to the extent of its impact on making sound decisions and the number of alternatives it proposes. The differentiation between these alternatives are based on the data of the institutions, and therefore the value of the data depends on the adequacy of the flowing data and its availability, which shows the difference in making a specific decision among several available alternatives. Relying on the data of each institution, which greatly affects this process, and the quality of data is considered the main factor in the process of changing decisions. Changing decisions is attributed to the availability of additional and clear data, and therefore this additional data is considered sufficient to distinguish

one decision from another. In contrast, the lack of data leads to make wrong decisions (Schildkamp, 2019), which affects the strategic planning process and thus the deviation of the institution from achieving its goals and lack of confidence in decision support systems.

### **\* Conclusion**

Many managers, people with authority, and decision-makers strive to make their decisions wise, logical, and creative, and therefore they seek decisions based on facts and evidence, and from here comes the role of data scientists to provide solutions and extract facts from various data sources to help making logical and high-quality administrative decisions.

Also, educational institutions must keep pace with these needs and requirements that must be met to prepare qualified data scientists to support the decision-making process, as at the present time data sources vary greatly, and the forms and formats of this data vary, and therefore the need appears to prepare qualified people to analyze this data and determine where it is available the most important data that can support effective decision-making, as well as the use of the latest technologies that have been reached in

the field of big data analysis to extract as much knowledge as possible. Evidence also indicates that decisions based on accurate and logical data are much better than decisions based on experience only without scientific evidence to support them. Therefore managers and decision-makers realize this fact and deal with it seriously to succeed.

The greater the volume of data (big data), the greater the problems associated with analyzing, storing, and extracting knowledge, but on the other hand, increasing the volume of data provides more features that can be used in the decision support process.

Given the importance of data mining to the formulation of decisions that are supported by evidence, it is recommended that future work include the development of large data analysis tools that are specifically intended to facilitate the learning process.

Utilizing analytics allows managers and decision-makers to mine large amounts of unstructured or semi-structured data for trends, patterns, and novel concepts. This can be done with either completely unstructured or partially structured data. These insights can help business executives better understand the requirements of their

customers and direct the strategic direction of their companies. Additionally, analytics can assist executives in making decisions that are both quicker and more cost-effective by reducing the amount of data that they are required to review. It is projected that fresh insights will be uncovered through the utilization of analytics as a result of there being more data available on more subjects.

A further benefit of conducting an analysis of big data is that it can help pinpoint areas in which judgments that have already been made can be enhanced.

#### \* References

- Anthony P. Sunjaya, S.A..C.R.J. (2022) 'A systematic review on the effectiveness and impact of clinical decision support systems for breathlessness'. *npj Primary Care Respiratory Medicine*, volume 32, p. 7.
- Chaowei Yang, Q.H..Z.L..K.L..F.H. (2017) 'Big Data and cloud computing: innovation opportunities and challenges'. *International Journal of Digital Earth*, Volume 10, 2017 - Issue 1, p. 42.
- Chunquan Li, Y.C..Y.S. (2021) 'A review of industrial big data for decision making in intelligent manufacturing'. *Engineering Science and Technology, an International Journal*, Volume 29, p. 16.
- Emrana Kabir Hashi, S.U.Z..R.H. (2017) 'An expert clinical decision support system to predict disease using classification techniques'. *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, p. 2017.
- Fyodor O. Fedin, O.V.T..S.V.C. (2019) 'Assessment of Intelligent Decision Support Systems Effectiveness in Technological Processes of big Data'. *International Russian Automation Conference (RusAutoCon)*, p. 6.
- Hai, A.A.M.A. (2022) 'Big data and its applications and its relationship to sustainable development'. <https://journals.ekb.eg/>, p. 21.
- Hasna El Alaoui El Abdallaoui, A.E.F..F.Z.E..M.S. (2018) 'Decision support system for the analysis of traffic accident Big Data'. 2018 14th International Conference on Signal-Image



- Technology & Internet-Based Systems (SITIS), p. 8.
- Hossein Hassani, C.B..S.U..M.T.M. (2020) 'Text Mining in Big Data Analytics'. *Big Data Cogn. Comput*, Volume 4 Issue 1 , p. 34.
- Marijn Janssen, H.v.d.V..A.W. (2016) 'Factors influencing big data decision-making quality'. *Journal of Business Research*, Volume 70, p. 8.
- Montserrat Acosta, S.C. (2017) 'Multicriteria Decision Analysis and Participatory Decision Support Systems in Forest Management'. *Forests*, p. 15.
- Nikhil Madaan, U.K..S.K.J. (2020) 'Big Data Analytics: A Literature Review Paper'. *International Journal of Engineering Research & Technology (IJERT)*, Volume 8, Issue 10, p. 9.
- Reihaneh H. Hariri, E.M.F..K.M.B. (2019) 'Uncertainty in big data analytics: survey, opportunities, and challenges'. *Hariri et al. J Big Data*, p. 16.
- Robert G. Aykroyd, V.L..F.R. (2019) 'Recent developments of control charts, identification of big data sources and future trends of current research'. *Technological Forecasting and Social Change*, Volume 144, p. 12.
- Schildkamp, K. (2019) 'Data-based decision-making for school improvement: Research insights and gaps'. *Educational Research*, Volume 61, p. 18.
- Shaofeng Liu, A.H.B.D..R.I.W..I.M.B. (2009) 'Integration of decision support systems to improve decision support performance'. *springer*, p. 26.
- Sowmya R, S.K.R. (2017) 'Data Mining with Big Data'. 2017 11th International Conference on Intelligent Systems and Control (ISCO) , p. 5.
- Wadii Boulila, I.R.F..A.H. (2017) 'A novel decision support system for the interpretation of remote sensing big data'. *Earth Science Informatics* , volume 11, p. 15.
- Wellem Anselmus Teniwut, C.L.H. (2020) 'Decision support system in supply chain: A systematic literature review'. *GrowingScience*, Volume 8 Issue 1, p. 18.
- Xindong Wu, X.Z..G.-Q.W..W.D. (2014) 'Data Mining with Big Data'. *IEEE TRANSACTIONS ON KNOWLEDGE AND*

DATA ENGINEERING, VOL.  
26, NO. 1, p. 11.

ZHIHAN LV, A.K.S. (2021) 'Big Data  
Analysis of Internet of Things  
System'. ACM Transactions on  
Internet Technology, Vol. 21,  
No. 2, p. 15.