

# Cluster Analysis as a Decision-Making Tool

**Bindu Rani**

*Sharda University, India*

**Shri Kant**

*Sharda University, India*

## INTRODUCTION

Informative value is a vital factor for the success of any organization's decision making process. Nowadays, evolution of new digital technologies, telecommunication 5G technologies and Internet of Things lead the way towards generation of massive amounts of data exemplified as big data. That data has been characterized by its varying characteristics, few of them are high volume, high dimensionality, diverse formats and rapid velocity. The data can be used to make growth in the financial, marketing and sales and customer services area with all types of competitive advantages. However, merely gathering data does not make a positive impact on organizations. Data must be analyzed and transformed into pieces of beneficial and valuable information (Thirathon *et al.*, 2017).

Data and analytics are playing an essential role to provide support for managers as well as decision makers in functioning their organizations and grasping benefits. As a matter of fact, research studies reveal that data centric organizations accompanied with data analytics make better expedient decisions in terms of high operational productiveness, better customer satisfaction and retention levels. Big data analytics play a key role in an organization to analyze large amounts of data to uncover hidden patterns, correlations and other insights in the data. It helps organizations as a decision driving force towards futuristic solutions (Jeble *et al.*, 2018).

The important research questions towards achieving informed decisions are “How can organizations incorporate big data analytics into the decision making process?” and “What data analytics methodologies can be applied to get smarter and informed decisions?”

Clustering, one of the most commonly known data analytical methodology is playing an indispensable role in big data analysis. It has been seen as a precious tool for marketing and business areas providing the capabilities to help in organization's decision making process (Jain, 2010). Hence, improvement and modification in cluster analysis techniques may be considered an effective approach for getting improved decisions. Moreover, important issues of the clustering process lie in assessing the clustering quality and finding optimal cluster numbers in a dataset of various domains. Cluster validation aims for quality assessment and quantitative estimation of true cluster number in datasets.

The implications from these situations highlight the need to consider suitable frameworks to incorporate data analytical techniques in decision making approaches. While considering clustering as a decision making tool, finding appropriate metrics for measuring cluster shape, cluster numbers and cluster quality has always been an exploratory research area across the sphere of data mining.

## BACKGROUND

### Big Data-A Term

The extensive use of emerging internet and mobile technologies is the genesis of big data age with huge capacity, complex and rapidly growing data in variety of forms. In 2020, around 2.5 quintillion bytes of data has been generated by internet users and around 1.7 megabytes data by every person in just a second. Google search statistics predicted for year 2021 that Google will get over 1.2 trillion searches per year, and more so 40,000 number of search queries per second.

Besides, big data is omnipresent but its origin is unknown yet. It was stated that in the mid 1990's, John Mashey coined this term at a lunch table conversation at Silicon Graphics Inc.(SGI) (Diebold, 2012). A repeated definition is given by Laney (2001) "Data generated very fast that contains an extensive volume of content". The term became a well-known term till 2011 (Gandomi & Haider, 2015).

Apart from mass of data, big data has been defined by several other characteristics. Doug Laney in 2011 pointed out challenges and opportunities accomplished by growing data into 3 Vs model (Increased Volume, Velocity and Variety). Later in 2011, IDC (International Data Corporation) most significant known name in big data stated that the evolution defines the development of tools and technologies in new form to extract value from big data. This results as the 4th V for big data –Value. The 4 Vs were widely accepted since it drew attention towards the usefulness of big data. It points the most significant problem in big data. If data is not utilized properly, it is only a bunch of data. Another term Veracity has been added as 5th V which means checks the provenance or reliability of the data source and verifies how meaningful the data is i.e truthfulness in data. This is not the end for expansion of Vs, big data is continuously characterizing with more Vs such as Validity, Variability, Vocabulary, Venue and Vague defining new parameters (Sun, 2018).

### Role of Big Data Analytics into Decision Making

The big data generation has reformed the perspectives towards the living approach, thinking modes and working styles. The transformation objective is to seize large amounts of data, perceive knowledge and make better decisions.

In an intelligence talk (2017), big data has been termed as fuel for digital world while engine used is analytics. Big data analytics is being exploited as the way to strengthen the effectiveness of the decision making process in each and every field of this revolutionary world. The potentiality to extract value from big data depends on data analytics and it has been considered as the core of the big data transformation age (Jagadish et al., 2014).

Labrinidis & Jagadish (2012) suggested to divide big data process into two main parts: Data management and analytic as shown in Figure 1. Data management concerns with the acquisition and retrieval, storage and preparation of data, although big data analytics is required to involve the techniques for extracting knowledge from big data and converting it into the intelligence.

Practically, big data analytics is concerned with the technologies to transform data gathered from different sources into shaped and smart decisions. These decisions can drive valuable insights and preventing risks and gain an edge over the competitors as a result of enhanced business performance. Few focussing parameters are faster decisions, act and react quickly, decisions making on high quality and complex data. Cluster analysis has been viewed as one of the most significant technique to find groups and patterns in support of managerial decisions.

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/cluster-analysis-as-a-decision-making-tool/317461](http://www.igi-global.com/chapter/cluster-analysis-as-a-decision-making-tool/317461)

## Related Content

---

### Real-Time Mental Health Monitoring: The Indian Scenario

Ritika Wason, Prashant Singh Rana and Vishal Jain (2020). *Applications of Deep Learning and Big IoT on Personalized Healthcare Services* (pp. 89-98).

[www.irma-international.org/chapter/real-time-mental-health-monitoring/251236](http://www.irma-international.org/chapter/real-time-mental-health-monitoring/251236)

### Review on Reliability and Energy-Efficiency Issues in Cloud Computing

Chitra A. Dhawale and Kritika A. Dhawale (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 790-802).

[www.irma-international.org/chapter/review-on-reliability-and-energy-efficiency-issues-in-cloud-computing/317486](http://www.irma-international.org/chapter/review-on-reliability-and-energy-efficiency-issues-in-cloud-computing/317486)

### Quorum Sensing Digital Simulations for the Emergence of Scalable and Cooperative Artificial Networks

Nedjma Djeddar, Iñaki Fernández Pérez, Noureddine Djedi and Yves Duthen (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 13-34).

[www.irma-international.org/article/quorum-sensing-digital-simulations-for-the-emergence-of-scalable-and-cooperative-artificial-networks/233888](http://www.irma-international.org/article/quorum-sensing-digital-simulations-for-the-emergence-of-scalable-and-cooperative-artificial-networks/233888)

### Palmpoint And Dorsal Hand Vein Multi-Modal Biometric Fusion Using Deep Learning

Norah Abdullah Al-johani and Lamiaa A. Elrefaei (2020). *International Journal of Artificial Intelligence and Machine Learning* (pp. 18-42).

[www.irma-international.org/article/palmpoint-and-dorsal-hand-vein-multi-modal-biometric-fusion-using-deep-learning/257270](http://www.irma-international.org/article/palmpoint-and-dorsal-hand-vein-multi-modal-biometric-fusion-using-deep-learning/257270)

### A Machine Learning Approach to Prevent Cancer

Ahan Chatterjee, Swagatam Roy and Rupali Shrivastava (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning* (pp. 112-141).

[www.irma-international.org/chapter/a-machine-learning-approach-to-prevent-cancer/263317](http://www.irma-international.org/chapter/a-machine-learning-approach-to-prevent-cancer/263317)