


# Chapter 7

## Big Data Analysis Techniques: Data Preprocessing Techniques, Data Mining Techniques, Machine Learning Algorithm, Visualization

**Pranali Dhawas**

 <https://orcid.org/0009-0003-4276-2310>  
*G.H. Rasoni College of Engineering, Nagpur,  
India*

**Minakshi Ashok Ramteke**

*G.H. Rasoni College of Engineering, Nagpur,  
India*

**Aarti Thakur**

*G.H. Rasoni College of Engineering, Nagpur,  
India*


**Poonam Vijay Polshetwar**

*CSMSS College of Polytechnic, India*

**Ramadevi Vitthal Salunkhe**

*Rajarambapu Institute of Technology, India*

**Dhananjay Bhagat**

 <https://orcid.org/0009-0009-1100-3219>  
*G.H. Rasoni College of Engineering, Nagpur,  
India*

### ABSTRACT

*Big data analysis techniques are the methods and tools utilized for extracting insights and knowledge from vast and intricate datasets. Due to the increasing velocity, volume, and variety of data being produced, conventional data analysis methods have become inadequate. Therefore, big data analysis techniques employ advanced computational and statistical methods to extract treasured information from big data. There are several big data analysis techniques, including data mining, natural language processing, machine learning, predictive analytics, and deep learning. For example, data mining involves identifying patterns and relationships within data sets, while machine learning enables systems to learn from data without explicit programming. Additionally, natural language processing focuses on analyzing human language, and predictive analytics utilizes statistical modeling techniques to predict future outcomes. Deep learning, which uses neural networks to model complex data patterns, is also a common big data analysis technique.*

DOI: 10.4018/979-8-3693-0413-6.ch007

# 1. INTRODUCTION

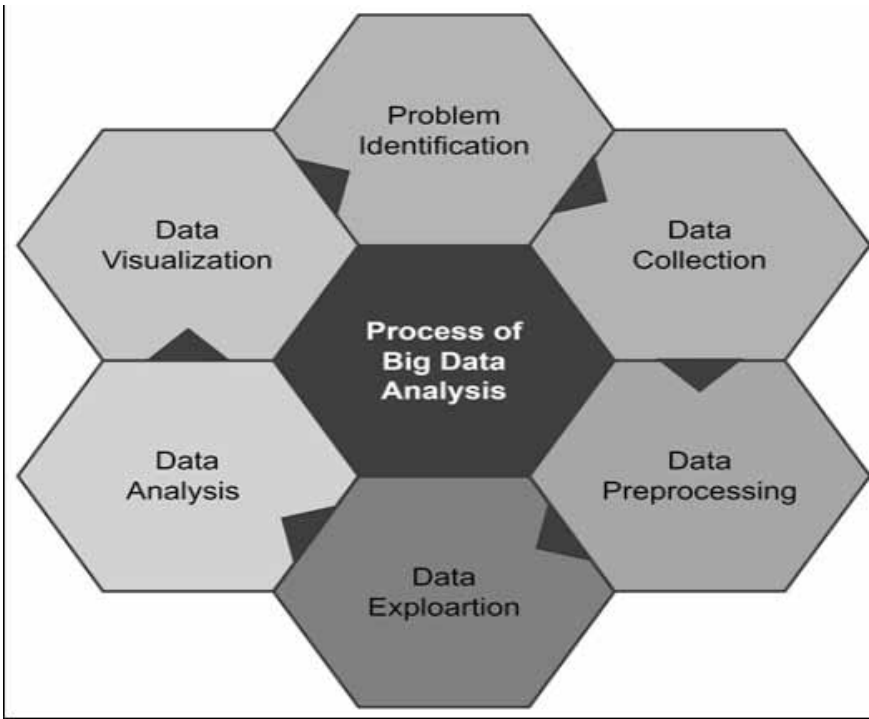
Big data analysis is a process of examining and interpreting massive and complex datasets to extract valuable insights, enabling decision-makers to make informed decisions. “Big data” describes datasets too large or complicated for conventional data processing tools to handle. Advanced software and hardware tools are used in big data analysis to collect, store, process, and analyze large datasets using various analytical techniques, which includes statistical analysis, machine learning, data mining, and natural language processing.

The insights obtained through big data analysis which can be used to inform decision-making in multiple industries and applications, such as healthcare, finance, marketing, and manufacturing. For example, healthcare providers can identify patterns in patient data, leading to more effective diagnosis and treatment. Retailers can analyze customer purchase patterns and create better marketing strategies.

Big data analysis has become prevalent as a result of the substantial increase in data volume originating from digital technologies like mobile devices, social media and Internet of Things (IoT). By utilizing these massive datasets, organizations can obtain a deeper understanding of their customers, operations, and markets, resulting in better decision-making and improved business outcomes.

Big data analytics involves identifying problem, gathering relevant data from diverse sources, storing it in scalable infrastructures, cleaning and preprocessing the data, integrating multiple datasets, exploring and analyzing the data using statistical and machine learning techniques, evaluating and validating models, interpreting and visualizing insights, and utilizing those insights for decision making and value creation as shown in below Figure1.

Figure 1. Process of big data analytics



24 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/big-data-analysis-techniques/336350](http://www.igi-global.com/chapter/big-data-analysis-techniques/336350)

## Related Content

---

### A Systematic Approach for Business Data Analytics with a Real Case Study

Kaibo Liu and Jianjun Shi (2015). *International Journal of Business Analytics* (pp. 23-44).

[www.irma-international.org/article/a-systematic-approach-for-business-data-analytics-with-a-real-case-study/132800](http://www.irma-international.org/article/a-systematic-approach-for-business-data-analytics-with-a-real-case-study/132800)

### Champion for Business Intelligence: SMART Goals for Business Focused and Financially Backed Results

Irina Dymarsky (2011). *International Journal of Business Intelligence Research* (pp. 22-36).

[www.irma-international.org/article/champion-business-intelligence/53866](http://www.irma-international.org/article/champion-business-intelligence/53866)

### Co-Engineering Business Need and IT Services

Jay Ramanathan and Rajiv Ramnath (2009). *Co-Engineering Applications and Adaptive Business Technologies in Practice: Enterprise Service Ontologies, Models, and Frameworks* (pp. 246-270).

[www.irma-international.org/chapter/engineering-business-need-services/6596](http://www.irma-international.org/chapter/engineering-business-need-services/6596)

### Consumption Characteristics of the Millennial Generation and Strategies Implemented to Satisfy Them

José G. Vargas-Hernández (2020). *Handbook of Research on IT Applications for Strategic Competitive Advantage and Decision Making* (pp. 361-376).

[www.irma-international.org/chapter/consumption-characteristics-of-the-millennial-generation-and-strategies-implemented-to-satisfy-them/262485](http://www.irma-international.org/chapter/consumption-characteristics-of-the-millennial-generation-and-strategies-implemented-to-satisfy-them/262485)

### Enterprise Information System and Data Mining

Kenneth D. Lawrence, Dinesh R. Pai, Ronald Klimberg and Sheila M. Lawrence (2010). *International Journal of Business Intelligence Research* (pp. 34-41).

[www.irma-international.org/article/enterprise-information-system-data-mining/45725](http://www.irma-international.org/article/enterprise-information-system-data-mining/45725)