Chapter 7

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

Pranali Dhawas

b https://orcid.org/0009-0003-4276-2310 G.H. Raisoni College of Engineering, Nagpur, India

Minakshi Ashok Ramteke

G.H. Raisoni College of Engineering, Nagpur, India

Aarti Thakur

G.H. Raisoni College of Engineering, Nagpur, India

ABSTRACT

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. Raisoni College of Engineering, Nagpur, India

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 Figure 1.



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

Related Content

Evolutionary Intelligence and Quality-Of-Information: A Specific Case Modelling

Eliana Pereira, Eva Silva, Bruno Fernandesand José Neves (2016). Applying Business Intelligence to Clinical and Healthcare Organizations (pp. 284-313).

www.irma-international.org/chapter/evolutionary-intelligence-and-quality-of-information/146074

Identification of User Preference for Multi-Criteria Reporting in Business Intelligence: A Case Study on Banking Business

Martin Aruldoss, Miranda Lakshmi Travisand Prasanna Venkatesan Venkatasamy (2018). *Improving E-Commerce Web Applications Through Business Intelligence Techniques (pp. 16-47).* www.irma-international.org/chapter/identification-of-user-preference-for-multi-criteria-reporting-in-businessintelligence/197188

Alleviation of Delay in Tele-Surgical Operations Using Markov Approach-Based Smith Predictor

Ratish Kumar, Rajiv Kumarand Madhav Ji Nigam (2022). *International Journal of Business Analytics (pp. 1-14).*

www.irma-international.org/article/alleviation-of-delay-in-tele-surgical-operations-using-markov-approach-based-smithpredictor/292057

Radial Efficiency Measures in Data Envelopment Analysis

Hirofumi Fukuyama (2014). *Encyclopedia of Business Analytics and Optimization (pp. 1967-1976).* www.irma-international.org/chapter/radial-efficiency-measures-in-data-envelopment-analysis/107385

Business Intelligence in the Music Industry Value Chain: Ensuring Sustainability in a Turbulent Business Environment

Hanne Russ, Jean-Pierre Kuilboerand Noushin Ashrafi (2014). International Journal of Business Intelligence Research (pp. 50-63).

www.irma-international.org/article/business-intelligence-in-the-music-industry-value-chain/108012