**DWARAKA DOSS GOVERDHAN DOSS VAISHNAV COLLEGE**

(Autonomous)

Re-accredited with "A++" Grade by NAAC

College with Potential for Excellence

(Linguistic Minority Institution)

Affiliated to the University of Madras,

Arumbakkam, Chennai- 600 106, Tamil Nadu, India.

**Department of Computer Science (UG & PG)**


*Organizes*

Proceedings of the International Conference on

# CYBERNETIC NEXUS

## ICCN - 2025

On: 14th February 2025 Venue: SRG Auditorium Time: 9:00 AM to 5:00PM


VR1 PUBLICATIONS

EMPOWERMENT THROUGH EDUCATION

| | | |
|---|---|---|
| ***Title of the Book*** | : | International Conference on Cybernetic Nexus (ICCN -2025) |
| ***Editors*** | : | Dr. S. Santhosh Baboo, Dr. P. Suganya |
| ***Published by*** | : | VR1 Publications |
| ***Publisher's Address*** | : | #23/25, Thambusamy Nagar, Thiruverkadu, Chennai - 77.<br>98848 47857, info@vr1publications.com, www.vr1publications.com |
| ***Printer's Details*** | : | Magizh Prints, Chennai. |
| ***First Edition*** | : | February 2025 |
| ***ISBN*** | : | 978-93-48498-56-4 |



VR1 PUBLICATIONS
EMPOWERMENT THROUGH EDUCATION

*Typeset*: **Goodwill Infotech, Chennai-73.**

# TABLE OF CONTENTS

# THE ROLE OF BIG DATA ANALYTICS IN ADVANCING INFORMATION SYSTEMS RESEARCH

**Jeevitha R[1] and Ramesh L[2]**

[1]Research Scholar, Department of Computer Applications, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, TN, India

[2]Assistant Professor, Department of Computer Applications, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, TN, India

*jeevi4rajee@gmail.com / lramesh.scs@vistas.ac.in*

**Abstract**

Big Data Analytics has emerged as a critical method for data collection and analysis, enabling the examination of vast sets of structured and unstructured data. Supported by a range of advanced technologies, it plays a pivotal role in the modern big data ecosystem. The literature reviewed in this study highlights various scenarios where big data analytics has been successfully implemented or holds potential for future application. By examining these cases, the paper provides a comprehensive understanding of how big data analytics can be leveraged across different domains and contexts.

The study further maps the existing literature to the six phases of Design Science, offering a structured approach to understanding its application in research. A straightforward yet effective criterion is introduced to classify the literature into the most relevant phase of Design Science it contributes to. Based on this mapping, the paper generates recommendations for future researchers, encouraging them to explore underutilized phases of Design Science and improve the application of big data analytics in these areas. This approach aims to guide researchers in addressing gaps and advancing the use of big data analytics within the framework of Design Science. Keywords: Design Science Process, Big Data Analytics, Information System.

## INTRODUCTION

This paper discusses the role of Big Data Analytics in Information Systems research. Big Data Analytics is used to analyze large sets of structured or unstructured data to extract useful insights. Many companies collect customer data through various platforms, and Big Data Analytics helps them identify patterns and trends that boost revenue. It is also a key tool in Information Systems research, where its applications are explored across different areas. This study examines how and why researchers currently use Big Data Analytics in their work. The collected literature is categorized into the six phases of the design cycle to understand its usage patterns. The analysis explains which phases use this method more frequently and why. Finally, recommendations are provided to encourage its application in other phases of design science (Peffers et al., 2007).

The six phases of design science which will be used as a basis here are:

1. **Problem Identification and Motivation**: This phase involves identifying a relevant problem and explaining why it is important to find a solution. It focuses on understanding the issue and justifying the need to address it.

2. **Objectives of a Solution**: In this phase, the key goals and requirements for creating an effective solution are defined. It outlines what the solution should achieve to solve the identified problem.

3. **Design and Development**: This phase focuses on creating and developing an artifact, such as a model, method, or framework, that addresses the solution's objectives. The artifact is designed to meet the needs of the problem and provide a practical way to implement the solution.

4. **Demonstration**: This phase involves showcasing the developed artifact to all stakeholders connected to the solution. It demonstrates how the artifact functions and its potential to address the identified problem.

5. **Evaluation**: In this phase, the artifact is assessed to determine how effectively it solves the problem. The evaluation measures the artifact's performance and ensures it meets the objectives of the solution.

Information Systems research plays a crucial role in driving industry growth and fostering innovation. Meaningful research, which relies on effective data gathering and analysis, is essential for producing accurate and impactful results. Selecting the right methodology not only enhances the accuracy of the research but also saves time by minimizing the need for rework. This project aims to identify areas where big data analytics can be effectively utilized and provide recommendations for improving this method to generate more efficient and reliable outcomes. By doing so, it seeks to contribute to the advancement of research practices in Information Systems.

The paper offers insights into how each phase of design science contributes to Information Systems research. In the industry context, major corporations are already leveraging big data analytics to transform unstructured data into actionable insights. This study also serves as a guide for future researchers, helping them classify their work according to the Design Science Research Methodology (DSRM). By aligning their research with DSRM, researchers can ensure a structured and systematic approach, ultimately enhancing the quality and relevance of their findings in the field of Information Systems.

## METHODOLOGICAL APPROACH: DESIGN AND EXECUTION

The research methodology provides an overview of how this project was conducted to contribute meaningfully to research and development. This project is a literature review that aims to analyze over 35 papers from the past 12 years to extract high- quality data for analysis. A structured and formal approach will be followed to ensure the project's objectives are met effectively.

The type of literature review chosen for this study is a **Narrative Literature Review**, as it involves examining previously conducted research, identifying gaps, and formulating recommendations based on the findings. The following steps will be meticulously carried out to achieve the proposed deliverables, as outlined by Pare et al. (2015):

1. **Formulate the Problem Statement**: Clearly define the problem statement before initiating the research to establish a focused direction.

2. **Search for Appropriate Literature**: Use strategic search techniques to gather relevant journal articles and papers. Keywords will be combined using Boolean operators (AND, OR, NOT) to create effective search queries.

3. **Conduct a Preliminary Screening**: Identify the most relevant articles related to the research topic. Apply filters to prioritize recent publications and ensure the literature is up-to-date.

4. **Assess the Quality of Extracted Material**: Review the selected literature with the immediate supervisor to ensure its relevance and credibility.

5. **Gather and Summarize Data**: Extract key insights from each piece of literature, summarizing the findings to form a foundation for analysis.

This systematic approach ensures a thorough and well-organized review, enabling the identification of gaps and the formulation of actionable recommendations for future research.

### Paper 1: Building A Big Data Analytical Pipeline with Hadoop For Processing Enterprise XML Data

Viktor et al. (2017) explore the processing of high-volume XML-based big data using the Hadoop ecosystem, addressing challenges through the Extract, Load, and Transform (ELT) cycle. The study introduces a novel analytical pipeline that leverages Hadoop functionalities to handle 10–15 GB of daily data comprising millions of XML files. Data is fed into the pipeline via a File Transfer Protocol (FTP) server, where unstructured XML data is first loaded into the Hadoop Distributed File System (HDFS) using an Agent of XML Transportation (AXT) (Borthakur, 2008). The AXT preprocesses the data into a hierarchical structure, preparing it for transformation and storing it in a separate folder.

The transformed data is then parsed using a Sax Parser on Python (SaPPy), which converts it into structured CSV files. SaPPy utilizes configuration files to extract meaningful information, and the process can be executed using either Apache Spark (Zaharia et al., 2012) or Apache Hadoop Streaming, both integral components of the Hadoop framework. Finally, the CSV files are imported into a Hive Database using a custom tool called HImp for further processing. The paper also references similar work, such as the ELTA method (Dmitriyev et al., 2015), which effectively processes semi-structured enterprise data. To optimize the ELT cycle, the study emphasizes the importance of incorporating self-developed components like AXT, SaPPy, and HImp.

### Paper 2: Adaptive Big Data Analytics for Deceptive Review Detection in Online Social Media

Zhang et al. (2014) propose a framework to detect deceptive reviews in social media data, a critical tool for e-commerce and social media platforms that rely on user reviews. The framework helps organizations filter out fake reviews, enhancing their business intelligence and strategy development. Unlike other methods, this approach provides detailed computational and evaluation results.

The framework consists of seven components, supported by open-source tools like Apache HBase and Hadoop Distributed File System (HDFS). Data flows through a Data Stream Processor, which filters product-related messages. A Data Pre-Processor

then organizes these messages by separating syntactic, lexical, and stylistic elements. A Feature Miner extracts key features to identify deceptive reviews, which are further refined by an Ensemble of Classifiers. A Learning and Adaptation tool optimizes feature selection, while a Communication Manager presents results using infographics like graphs and tables.

The system's novelty lies in filtering near-duplicate deceptive reviews, improving efficiency over existing methods. Tested on Twitter and Amazon data, the framework showed better accuracy and reduced computational time for large data streams.

### Paper 3: Managing Big Data for Firm Performance: A Configurationally Approach

Kung et al. (2015) propose a framework to help organizations improve performance by integrating Big Data with other capabilities. Davenport et al. (2010) highlight that leveraging Big Data requires changes across departments, not just IT. This study focuses on combining Big Data with organizational capabilities like IT and Organizational Improvisational Capability (OIC) to maximize firm performance.

The framework emphasizes the 3Vs of Big Data: Volume, Velocity, and Variety. It integrates Data Lifecycle Management (DLCM)—comprising Data Collection, Repository, Processing, and Dissemination—with organizational factors to enhance decision-making. Using fuzzy-set Qualitative Comparative Analysis (fsQCA), the study identifies optimal configurations for high market growth strategies, considering factors like IT capabilities and market turbulence.

### Paper 4: Automated Competitor Analysis Using Big Data Analytics

Guo et al. (2015) present a method to automate competitor analysis using Big Data,

helping businesses gather intelligence on competitors' strengths and weaknesses. Unlike manual methods, this approach extracts data from diverse sources like customer reviews, surveys, and social media.

The process begins with data collection using web crawlers, gathering information from platforms like Google Play. A Naïve Bayes classifier categorizes the data, such as classifying games into combat, racing, or kids' categories. A Data Analyzer then evaluates metrics like downloads, reviews, and pricing to identify market leaders and competitors. The results help firms refine their products and strategies to stay competitive.

### Paper 5: Advanced Computing Model for Geosocial Media Using Big Data Analytics

Rathore et al. (2017) propose a framework to analyze geosocial media data, which includes location-based posts on platforms like Twitter and Facebook. This data can help governments and organizations respond to emergencies, disasters, and other real- time events.

The system collects data in a 4-tuple format (Location, Time, User, Content) and stores it in HDFS. Streaming and State Transfer APIs harvest real-time data, with

Apache Spark handling Twitter data. The processing layer filters unnecessary data and performs computations, while the application layer communicates results for specific needs, such as disaster management or traffic control. This framework offers improved data quality and computational efficiency over existing methods.

The Design Science Research Method (Peffers et al., 2007) consists of six phases, each playing a significant role in Information Systems research. However, some phases are not directly relevant to data collection and analysis using big data analytics. These phases are:

- Identify Problem and Motivation – This phase focuses on problem identification, which does not require data collection methods.
- Demonstration – Since this step involves demonstrating the artefact rather than data collection, other methods are more relevant.
- Communication – This phase involves sharing research findings with stakeholders, making data collection methods irrelevant.

*ISBN* : 978-93-48498-56-4

The key phases where data collection methodologies apply are:

1. Defining Objectives for a Solution
2. Design and Development
3. Evaluation (Formative evaluation occurs before artefact completion, while summative evaluation takes place afterward).

A framework was implemented to map the identified literature, ensuring alignment with key criteria from the literature summary. The benchmarking criteria are outlined below.

**Table 1: Criteria for classifying the papers.**

| Phase | Main Criteria for classifying the paper |
|---|---|
| Defining Objectives for a solution | Requires collection of Data (DC) |
| | Gives Prerequisites of designing a solution (PRQ) |
| Design and Development | Requires collection of Data (DC) |
| Evaluation | Requires collection of Data (DC) |

## CONCLUSION

This paper studies the use of Big Data Analytics in Information Systems research. It reviews research papers that use this method for data collection. Big Data Analytics is a new concept, so there is limited research available. Finding high-quality journal papers was difficult. Most researchers use Apache Hadoop or Spark for big data projects. The Hadoop File System is the most common storage method. These findings will help future researchers choose the best big data technology.

### References

1. Bodnar T, Dering ML, Tucker C, Hopkinson KM (2016) Using large-scale social media networks as a scalable sensing system for modeling real-time energy utilization patterns. IEEE Trans Syst Man, Cybern Syst PP(99):1–14
2. Borthakur, D. (2008). HDFS architecture guide. In Hadoop Apache Project http://hadoop.apache.org/common/docs/current/hdfsdesign.pdf (p. 39).
3. Chen, H., Chiang, R.H. and Storey, V.C. (2012), “Business intelligence and analytics: from big data to big impact”, MIS Quarterly, Vol. 36 No. 4, pp. 1165-1188.
4. Davenport, T. and Patil, D. (2012), “Data scientist: the sexiest job of the 21st century”, Harvard Business Review, Vol. 90 No. 10, pp. 70-76.
5. Davenport, T. H., Harris, J. G., and Morison, R. 2010. “Analytics at Work: Smarter Decisions, Better Results,” Boston, Massachusetts: Harvard Business School Publishing Corporation.
6. Davenport, T.H. (2013), “Analytics 3.0”, Harvard Business Review, Vol. 91 No. 12, pp. 64-72.
7. Decker, R. and Trusov, M. (2010), “Estimating aggregate consumer preferences from online product reviews”, International Journal of Research in Marketing, Vol. 27 No. 4, pp. 293-307.

8. Dmitriyev, V., Kruse, F., Precht, H., Becker, S., Solsbach, A., & Marx Gómez, J. (2017). Building A Big Data Analytical Pipeline With Hadoop For Processing Enterprise XML Data.

9. Dmitriyev, V., Mahmoud, T., & Marin-Ortega, P. M. (2015). SOA enabled ELTA: approach in designing business intelligence solutions in Era of Big Data. In International Journal of Information Systems and Project Management (Vol. 3, pp. 49–63). https://doi.org/10.12821/ijispm030303

10. Eriksson B, Barford P, Sommers J, and Nowak R (2010) A learning-based approach for IP geolocation. In Krishnamurthy A and Plattner B (eds) Passive and active measurement. Berlin, Springer Lecture Notes in Computer Science 6032: 171–80

11. Feldman, R., Govindaraj, S., Livnat, J. and Segal, B. (2010), "Management's tone change, post earnings announcement drift and accruals", Revenue Accounting Studies Journal, Vol. 15 No. 4, pp. 915-953.

12. Fielding RT (2000) "Architectural styles and the design of network-based software architectures",^ Thesis (Ph.D.). University of California, Irvine

13. George, G., Haas, M.R. and Pentland, A. (2014), "Big data and management", Academy of Management Journal, Vol. 57 No. 2, pp. 321-326.

14. Guo, L., Sharma, R., Yin, L., Lu, R., & Rong, K. (2017). Automated competitor analysis using big data analytics: Evidence from the fitness mobile app business. Business Process Management Journal, 23(3), 735-762.

15. Hevner, A., March, S., Park, J., Ram, S. 2004. "Design Science in Information Systems Research," MIS Quarterly (28:1), pp. 75-105.

16. Jarle Gressgård, L., Amundsen, O., Merethe Aasen, T. and Hansen, K. (2014), "Use of information and communication technology to support employee- driven innovation in organizations: a knowledge management perspective", Journal of Knowledge Management, Vol. 18 No. 4, pp. 633-650.

17. Khan, Z., & Vorley, T. (2017). Big data text analytics: an enabler of knowledge management. Journal of Knowledge Management, 21(1), 18-34.