

# Mapping Big Data's 5V Constraints To Machine Learning Architecture And Scalability

Priyanshu Singh<sup>1</sup>, Nishtha Gupta<sup>2</sup>, Annu Gautam<sup>3</sup>

<sup>1,2,3</sup>Students, School of Management, Noida Institute of Engineering and Technology Greater Noida, India  
priyanshu24singh@gmail.com<sup>1</sup>, nishthakastwar08@gmail.com<sup>2</sup>, annugautamgautam79@gmail.com<sup>3</sup>

## Abstract

The speedily developing field of big data analytics uses machine learning to analyse large and diverse information to help companies make smart business decisions. Analytics of Big data is important for national intelligence, cybersecurity, biology, fraud detection, and medical informatics. The use of massive volumes of data powered the development of big data in the early 2000s. The term "Big Data" refers to data collections of high size, speed, or complexity, making normal processing methods unsatisfactory. Large datasets have important potential, and machine learning powers artificial intelligence to extract information to support informed decision-making. In this article, we discuss machine learning procedures, big data knowledges, and applications of machine learning. This article discusses noteworthy issues with applying machine learning on large datasets. This research will inspect the current advances and problems at this unique convergence, revealing the potential for transformation in big data processing using machine learning.

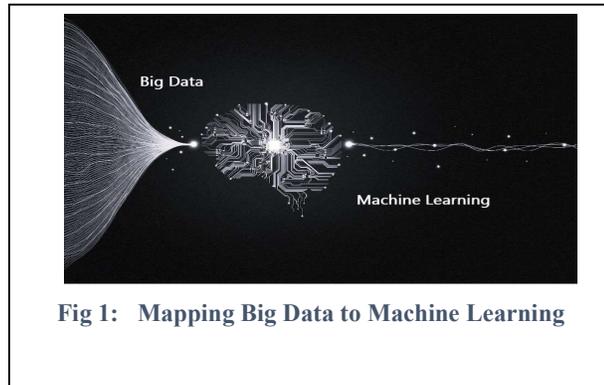


Fig 1: Mapping Big Data to Machine Learning

**Keywords**—Big Data, Data Analytics, Machine Learning, Big Data Analysis, Big Data Coordination ML

## I. INTRODUCTION

The path to machine learning and large-scale data processing is shaped by several past and current advances in big data, which are appealing, but their use still is challenging. When dealing with large and complex datasets, traditional machine learning algorithms designed for normal scale of datasets that fit within computer memory will become outdated. Machine learning has grown in data-concentrated fields, including multifield such as medicine, astronomy, biology, and others. These emerging spectacles may provide unique insights from large datasets across many disciplines. However, the scale and complexity of big data render conventional learning methods outdated. The current method for extracting data from traditional datasets is not capable of addressing these new challenges. Machine learning and big data collide to affect information discovery and decision-making. Predictive analytics or modelling aims to provide accurate forecasts. Big data sources collect and transform vast amounts of data; however, extracting meaningful insights requires different machine learning models. Machine learning has the potential to identify valuable patterns or the hidden patterns inside the datasets. Machine-learning algorithms benefit from larger training datasets because the use of machine learning and big data

analysis provides two distinct benefits. Algorithms management by processing a constant data stream which will help us to work with extensive and diverse data enhances algorithms. Examining several techniques for integration within large datasets, a machine learning system which can uncover hidden patterns and provide valuable insights for predictive modelling can be useful. These machine learning algorithms can mechanism manual labour in several industries. Organisations use algorithmic outcomes to guide their operational decisions. Advanced computer systems use data analytics and artificial intelligence to outperform human capabilities, yet they lack decision-making authority as all the computer systems. Algorithmic discoveries lack reliability without accurate data provided by specialists. An expert's lack of understanding of algorithmic outcomes may have detrimental effects on business decision-making. Ultimately, machine learning provides efficient, automated techniques for data collection, analysis, and integration and the use of machine learning, in combination with enhanced internet-based computing capacity, improves processing efficiency and which simplifies the integration of large volumes of data from multiple sources. Machine learning algorithms can be employed in a variety of Big Data processes, like including data segmentation, analytics, and simulation

and the integration of various processes produces a complete representation of Big Data, including insights and patterns, which are then organised and presented cohesively and intelligibly. The relationship between Machine Learning and Big Data is a never-ending loop while data is processed both within and outside the system, the algorithms created for specific goals are constantly monitored and refined. As machine learning plays a active role in addressing the challenges posed by large volumes of data and in leveraging its potential to generate new insights for commercial decision-making and scientific research which lead to the future path of machine learning analytics, entails a shift toward a more declarative approach, enabling enhanced interaction between domain experts and diverse data sources.

#### **Research Question:**

RQ1: How can a systematic, V-dimension-based analytical framework can be used to classify and evaluate the operational limitations of core Machine Learning methodologies?

RQ2: What is the necessary architectural shift required to support modern distributed computing platforms?

## **II. LITERATURE REVIEW**

While some researchers have focused on approaches, others have tackled the most general problems of machine learning in the context of Big Data. Researcher [1] noted several difficulties with machine learning when it comes to large data. The management of unstructured data formats, streaming data, multi-source input, noisy and low-quality data, high dimensionality, ensuring algorithm scalability, addressing imbalanced input distributions, handling unlabeled data, and managing limited labelled data are among the challenges faced in this field. Their work was mostly concerned with using deep learning methods. They stressed the need for flexible and extendable frameworks, a thorough comprehension of statistical data, and experience with bigger datasets before recommending algorithms. In their research on machine learning using Big Data, [2] were unable to pinpoint the precise reasons behind each issue they encountered. Furthermore, they talk about broad issues without offering any solutions. Their as professionals is focused on identifying, analyzing, and offering workable solutions to problems that arise, particularly those related to the aspects of big data. Qiu et al. [3] conducted a survey focusing on signal processing and the use of machine learning for huge data sets. The research determined the association between five crucial criteria and the Big Data dimensions, which are as follows: massive size, diverse data kinds, fast velocity, incomplete or uncertain data, and poor value density. The study clarifies the subtleties of obstacles and shows how they relate to the V factors. He provides a thorough explanation and characterization of the different learning strategies used in the

processing of Big Data signals. Even when issues and potential solutions are identified, choosing the best learning paradigm or approach for a certain use case or circumstance may be challenging. The lack of clear categorization and linkage between techniques is the cause of this. Another Researcher emphasized the importance of large-scale systems and suggested memory-efficient strategies for assessing Big Data machine learning. They recognized the difficulties that Big Data offered, but their summary lacked the depth that ours did. Al-Jarrah and colleagues focused on distributed system analysis and did not include computational complexity mitigation techniques [4]. This paper examines and highlights the differences between analytical and computational complexity in distributed environments [5]. Unlike previous studies, this one shows a relationship between the problems that the platforms now in use address and the solutions they provide. In addition, this article looks at other choices, such Big Data platforms [6]. The challenges of using Big Data for data mining are well documented in the literature. As Fan and Bifet [7] spoke on the difficulties data mining has when faced with large data sets, but they didn't provide any classifications or solutions. Level I (Big Data mining platforms), Level II (Semantics and implementation skills), and Level III (Big Data mining methods) are the three levels divided into difficulties. In contrast, this study uses the V dimensions to categorize data. While Wu et al. focused mostly on data mining, this paper clarifies a particular aspect of machine learning. This study demonstrates a relationship between big data solutions and the associated difficulties. This research uses big data to categorize and pinpoint the causes of machine learning problems. Furthermore, we evaluate and examine how well different machine learning methods work to solve certain issues. This makes it easier for researchers to choose the best Big Data solution or learning paradigm. This makes it easier to find gaps in the literature and possible directions for further study on the use of large data in machine learning. So, this study provides a solid basis for further investigation. We surveyed machine learning in big data analysis to identify these limitations.

## **III. FROM ALGORITHMS TO INSIGHTS:**

### **METHODOLOGICAL APPROACHES IN MACHINE LEARNING FOR BIG DATA PROCESSING**

Machine learning is considered a fundamental branch of artificial intelligence due to the utilization of a diverse range of learning algorithms and methodologies for the examination of extensive data sets. Unsupervised learning involves the categorization of cases into groups characterized by significant similarities, while supervised learning methods, such as classification and regression, are used to identify clusters exhibiting a substantial probability density. The categorization of clustering

methods may be organized into three distinct categories: supervised, unsupervised, and semi-supervised. The complexity associated with decision trees poses challenges in their implementation inside

applications that use large-scale datasets. The trees classify data by considering the values of the characteristics.

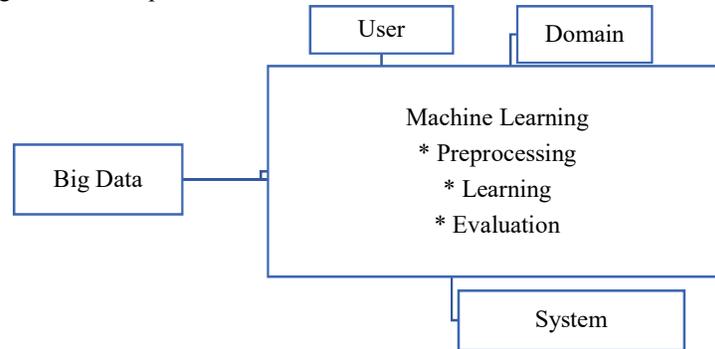


Fig 3: Implementation of Machine Learning in Big Data

Classification and regression are part of supervised learning. Regression is used for continuous class attributes, whereas classification is used for discrete ones. Decision tree learning, k-NN, and naive Bayes classifiers are used for classification. Regression techniques include logistic and linear regression. A similarity- and dissimilarity-metric-based unsupervised clustering approach enhances intra-cluster similarity and reduces inter-cluster similarity. A specific goal function is used. Segmentation employs unsupervised clustering algorithms such as K-means and hierarchical clustering. In addition to the similarity measure, semi-supervised clustering uses domain knowledge to guide and refine the clustering process. Examples of domain knowledge include paired restrictions on observations or objective variables. Trees classify data by feature values. Using greedy top-down feature selection, decision trees are repeatedly trained on the training data. Decision tree classifiers create three-structured diagrams using training data. Decision trees begin with a root node that contains the entire dataset. Data subsets partitioned by the division criteria are then assigned to leaf nodes. The entire dataset for each growing node must be managed, since information gain and other quality measures establish the splitting criteria. Thus, integrating decision trees into large data applications is difficult [8]. A binary classifier, a support vector

machine (SVM), maps the input data space to a higher-dimensional feature space to find a linear classifier. When applied to moderate-sized datasets, SVM excels [11]. Big data applications have restrictions [9]. Deep machine learning is a major area of AI research. It employs hierarchical frameworks to process information at multiple levels in machine learning. In hierarchical representations of observational data, the model deduces higher-level qualities from lower-level features. Hierarchical learning gives deep learning algorithms complex, high-level data representations. Deep learning can also be learned from labelled data, but its primary goal is to find patterns or important representations in massive unlabeled data sets. Deep learning is intriguing for pattern recognition and analysis due to the above characteristics. Deep learning-based architectural designs and methods handle Big Data analytics complexity better. In large data sets, deep machine learning may be used. Its extensive training time limits its usefulness to large data sets. Parallel Learner for Assembling Numerous Ensemble Trees (PLANET) regression tree implementation on Hadoop uses MapReduce jobs. Although this technique is not suitable for categorical variable datasets, it can manage large amounts of data. Multiple learners' discoveries are included in machine learning to increase accuracy

TABLE I. COMPARATIVE EVALUATION OF ML TECHNIQUES FOR BIG DATA

Algorithm	Task	Strengths for Big Data	Weaknesses for Big Data	Common Applications
<b>Distributed Linear Regression</b>	Regression	Highly scalable, efficient for dense data, interpretable	Sensitive to outliers, assumes linear relationships	Large-scale financial analysis, marketing campaign analysis

<b>Logistic Regression (Spark MLlib)</b>	Classification	Scale, handles binary classification well, efficient for linear relationships	Assumes linear relationships can be computationally expensive for very large datasets	Fraud detection, credit risk assessment, customer churn prediction
<b>Random Forest (Spark MLlib)</b>	Classification & Regression	Robust to overfitting, handles complex relationships, scalable	Interpretability can be challenging, computationally expensive	Large-scale fraud detection, customer segmentation, risk assessment
<b>Gradient Boosting Trees (XGBoost)</b>	Regression & Classification	Highly accurate, scalable, handles missing data well	Requires careful parameter tuning, not as interpretable as decision trees	Click-through rate prediction, product recommendation, anomaly detection
<b>Apache Spark ALS (Alternating Least Squares)</b>	Collaborative Filtering	Efficiently scales to large datasets, handles sparse data well	May not capture complex relationships, cold start problem	Recommender systems, personalized content delivery
<b>K-Means Clustering (Spark MLlib)</b>	Clustering	Efficient, works well with numerical data, scalable	Sensitive to outliers, choosing k can be challenging	Large-scale market research, customer segmentation, anomaly detection
<b>Dimensionality Reduction Techniques (PCA, SVD)</b>	Unsupervised	Efficiently reduces data size, preserves most variance, scalable	May not capture non-linear relationships, can be challenging to interpret	Feature engineering, data visualization, anomaly detection

**IV. MACHINE LEARNING HURDLES IN THE FACE OF EVOLVING BIG DATA DEFINITIONS**

Machine learning encounters several challenges while handling large datasets. Examples of interconnected concerns include developing scalable architectures, such as computational frameworks that can handle enormous datasets, which are essential for systems to scale effectively. Data characteristics that lead to a complete understanding of data qualities. Scaling with complexity which move to develop machine learning models that can process larger datasets, higher-dimensional features, and labels and quantities. While multiple machine learning techniques can process large quantities of data, scalability remains a big challenge in data analytics. Which include recommender systems, NLP, association rule learning, and ensemble learning. In bringing "Big Data analytics" platforms, such as Hadoop, did not include

machine learning capabilities, such as repetition and recursion. These systems almost depend on the MapReduce framework [13]. Developers also use other methodologies to handle iterative processes that do not conform to the MapReduce framework [14]. One of the biggest challenges in using machine learning to classify large datasets is the risk of accuracy deprivation when applying algorithms trained on one dataset to another. Traditional machine learning methods have challenges in cooperatively expanding datasets and class categories. Machine learning (ML) algorithms often focus on individual learning goals, which makes them insufficient for the various learning requirements of big data analytics. General information sharing and the use of online education are often considered effective remedies. None of these strategies is ideal for managing large data rivers due to the lengthy training process or the

efficiency loss resulting from the use of multiple processors [16]. Many machine learning algorithms struggle to handle high speed, large scale, and other characteristics of large datasets [15]. Deep learning presents numerous challenges in large-scale data analytics. The difficulties at hand include incremental learning for non-stationary data, managing huge models, and processing high-dimensional data. Inefficiencies arise due to the absence of optimisation in key data mining and machine learning techniques for high-level data parallel frameworks such as MapReduce [17]. The GraphLab abstraction allows for efficient parallel processing in shared-memory environments by providing data-consistent, asynchronous, and graph-parallel capabilities.

#### V. FROM THEORY TO REALITY: CHARTING THE TECHNOLOGICAL EVOLUTION OF MACHINE LEARNING IN BIG DATA PROCESSING

The massively parallel database processing community developed most scalable machine learning advancements, including Apache Mahout and Madlib. Machine learning algorithms with predictive skills for large-scale processes may improve workplace productivity in the Big Data age [12]. Several methods have been tested, including Deep learning algorithms that automate feature engineering by using data-driven feature generation and analysis; incremental learning algorithms in associative memory architectures are easily adaptable to new sources and datasets; faceted learning reveals hierarchical structure; and multi-task learning allows data to simultaneously learn about several tasks [2]. Big Data is a massive, growing data domain that needs multi-domain, representation-learning (MDRL) classification. Distance-metric learning, feature extraction, and feature variable acquisition are among MDRL components. Many representation-learning methods have been proposed in machine learning. The proposed network architecture and cross-domain representation-learning (CDRL) may improve Big Data classification. Deep learning's ability to understand and interpret massive volumes of unstructured data makes it essential to big data analytics. Deep learning might ease discriminative tasks, semantic indexing, data labeling, quick information retrieval, and difficult pattern extraction from vast datasets in Big Data analytics. Bayesian Networks are common machine learning methods for mimicking probabilistic variable interactions. Smart preparation of massive datasets and data quality scores to measure data dependability and quality were offered. A weight-based ensemble method for creating a Bayesian Network structure from local data was also given. The approach is easy to integrate with distributed data-parallelism (DDP) engines like Hadoop since the learning process was created using the Kepler scientific workflow. Some

studies also showed how Kepler may speed up Big Data BN learning application development and deployment. Scalable Bayesian Network Learning (SBNL) combines workflow, distributed computing, and machine learning [12]. Ensemble learning and distributed computing enable intelligent Big Data preparation and expert Bayesian Network (BN) learning from Big Data. Stream processing must simultaneously evaluate new and old input to provide usable results. Most stream processing uses pre-trained models to save computer costs and response time. A new framework allows machine learning on huge data streams. Architecture ensures data archiving consistency and durability between HBase and HDFS. The design centers on bulk and stream processing units [18]. We can say the platform offers machine learning algorithms and tools for clustering, prediction, recommendation, and classification. Regression problem decision rules were derived using distributed streaming. The method is used using Scalable Advanced enormous Online Analysis (SAMOA), an open-source huge data stream mining platform. Vertical and horizontal parallelism distributes adaptive model rules (AMRules) throughout a cluster. AMRules creates simple decision rule models. SAMOA simplifies developing and integrating new distributed machine learning algorithms with cutting-edge DSPEs [17]. It is a networked library of machine learning algorithms that allows users to modify pre-built algorithms. Data mining and machine learning emphasize feature selection. Feature selection selects key traits to create the best prediction models. FS strategies have been suggested several times [14]. These strategies may be categorized as filter, wrapper, or embedding based on selection criteria. Instead of learning, filter approaches use correlation, information, and distance. Wrapper techniques must be introduced to a learning system to assess attribute effectiveness. In embedded techniques, feature selection is part of model training [15]. They give enough information for learning, yet they are more efficient than wrapper strategies. We explored the challenge of feature selection in big data mining based on online learning and proposed online feature selection (OFS) [16]. OFS aimed at constructing online classifiers with limited features. The results show that the suggested web application feature selection techniques perform successfully. They also surpass sophisticated batch feature selection in efficacy and scalability. [10]

#### VI. EXPLORING PRACTICAL MACHINE LEARNING APPLICATIONS FOR BIG DATA

The use of machine learning in the context of big data has resulted in a multitude of applications that have significantly transformed several sectors.

TABLE II. MACHINE LEARNING APPLICATIONS FOR BIG DATA EXEMPLIFY

Application	Description	Machine Learning Techniques (Examples)	Benefits (Examples)
<b>Predictive Analytics</b>	Predict future trends and patterns from historical and real-time data.	Regression, Time series analysis, Forecasting models	Improved decision-making (stock price prediction), risk mitigation (equipment failure prediction).
<b>Recommendation Systems</b>	Personalise user experiences with tailored suggestions.	Collaborative filtering, Content-based filtering, Deep learning	Increased user engagement (Netflix recommendations), sales conversion (Amazon product suggestions).
<b>Natural Language Processing (NLP)</b>	Analyse large volumes of text data to derive insights, sentiment, and language-related tasks.	Text classification, Sentiment analysis, Machine translation, Chatbots	Automated document analysis (spam filtering), market research (social media sentiment analysis).
<b>Image and Video Analysis</b>	Classify and identify images and videos.	Computer vision, Deep learning, Image/video recognition	Automated content moderation (facial recognition), autonomous vehicles (object detection).
<b>Fraud Detection</b>	Identify fraudulent transactions from financial data.	Anomaly detection, Classification algorithms, Fraud scoring models	Reduced financial losses (credit card fraud) and improved security (insurance fraud detection).
<b>Healthcare and Genomics</b>	Analyse medical records and genetic data for disease prediction, drug discovery, and personalised medicine.	Deep learning, Support vector machines, Clustering algorithms	Early disease diagnosis (cancer risk prediction), personalised treatment plans, and improved drug development.
<b>Customer Care Prediction</b>	Identify customers at risk of churning to prevent their loss.	Customer segmentation, Churn prediction models, Customer lifetime value analysis	Improved customer retention (targeted incentives), reduced customer churn.

## VII. CONCLUSION

Conventional machine learning methods do not possess intrinsic efficiency or scalability to effectively address the attributes associated with substantial data volumes, diverse data kinds, rapid data generation rates, uncertainty, incomplete data, and poor value density. To surmount these hurdles, it is essential for machine learning to undergo a reformation tailored for the processing of large-scale datasets. To expand nodes, all data sets must be examined and quality indicators used to divide a decision tree. Decision trees in large-data applications are difficult to implement. Data sets of suitable size work well with SVM. The implementation of big data applications has restrictions. Deep learning effectively handles volume and diversity issues in huge datasets. Due to the long training time, it performs poorly with big volumes of data. PLANET handles large amounts of data but not category properties. Big data machine learning applications face memory constraints, MapReduce support issues, high velocity, volume, and diversity of data, and limited training on labeled datasets or class categories. Faceted learning, multi-task learning, online feature selection, streaming data processing, and multi-domain/cross-domain representation learning are technological breakthroughs. Large-scale machine learning qualities and obstacles might be studied further. Technological progress encompasses several developments such as faceted learning for hierarchical data structures, parallel multi-task learning, representation-learning across multiple domains, streaming data processing, high-dimensional data processing, and online feature selection. In forthcoming times, there will be a heightened emphasis on enhancing and evaluating the efficacy of machine learning approaches across diverse issue domains. One promising area for investigation is the extension of machine learning techniques that exhibit effectiveness and scalability in managing datasets with a large number of dimensions to the domain of big data.

## REFERENCES

- [1] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015, February 24). *Deep learning applications and challenges in big data analytics*. Journal of Big Data. <https://doi.org/10.1186/s40537-014-0007-7>
- [2] L'Heureux, A., Grolinger, K., ElYamany, H. F., & Capretz, M. A. M. (2017, January 1). *Machine Learning With Big Data: Challenges and Approaches*. IEEE Access. <https://doi.org/10.1109/access.2017.2696365>
- [3] Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020, May 1). *A survey on machine learning for data fusion*. Information Fusion. <https://doi.org/10.1016/j.inffus.2019.12.001>
- [4] Al-Jarrah, O. Y. (2015, March 18). *Efficient Machine Learning for Big Data: A Review*. arXiv.org. <https://arxiv.org/abs/1503.05296>
- [5] Gandomi, A. (2015). *Beyond the hype: Big data concepts, methods, and analytics*. <https://www.semanticscholar.org/paper/Beyond-the-hype%3A-Big-data-concepts%2C-methods%2C-and-Gandomi-Haider/1bc34cb22131554ba18f6ba9e6ede5be42939f1>
- [6] (PDF) *A Survey on Platforms for Big Data Analytics*. (2014, October 9). ACME. <https://research.amanote.com/publication/NKt-AnQBKQvf0Bhitijv/a-survey-on-platforms-for-big-data-analytics>
- [7] Wang, F., & Bifet, A. (2013, April 30). *Mining big data*. SIGKDD Explorations. <https://doi.org/10.1145/2481244.2481246>
- [8] Çetinkaya, Z., & Horasan, F. (2021, January 18). *Decision Trees in Large Data Sets*. Uluslararası Mühendislik Araştırma Ve Geliştirme Dergisi. <https://doi.org/10.29137/umagd.763490>
- [9] J. (n.d.). Lee, Y. et al. (2014). Scribd. <https://www.scribd.com/document/478922616/Lee-Y-et-al-2014>
- [10] Devi, D. R., & Sasikala, S. (2019, November 21). *Online Feature Selection (OFS) with Accelerated Bat Algorithm (ABA) and Ensemble Incremental Deep Multiple Layer Perceptron (EIDMLP) for big data streams*. Journal of Big Data. <https://doi.org/10.1186/s40537-019-0267-3>
- [11] G. Singhal and A. Singh, "The Large Action Model: Pioneering the Next Generation of Web and App Engagement," *2024 11th International Conference on Reliability, Infocom Technologies and Optimisation (Trends and Future Directions) (ICRITO)*, Noida, India, 2024, pp. 1-6, doi: <https://doi.org/10.1109/ICRITO61523.2024.10522132>
- [12] Singhal, G. (2024). The influence of artificial intelligence on the diverse financial sector. In *Sustainability-digital-transformation-era-driving-innovative-growth* (pp. 390–399). <https://doi.org/10.1201/9781003534136-66>
- [13] Jain, R., Upreti, K., Ekatpure, S. R., Gupta, V. K., Rai, A. K., & Gothania, J. (2025). The convergence of IoT, ML, and big data in Self-Service innovation. In *IGI Global eBooks* (pp. 241–270). <https://doi.org/10.4018/979-8-3373-1295-8.ch009>
- [14] Li, F., & Xu, J. (2025). Revolutionizing AI-Enabled information systems using integrated big data analytics and Multi-Modal Data fusion. *IEEE Access*, 13, 212316–212340. <https://doi.org/10.1109/access.2025.3552039>

- [15] Shahnawaz, M., & Kumar, M. (2025). A comprehensive survey on big data analytics: characteristics, tools and techniques. *ACM Computing Surveys*, 57(8), 1–33. <https://doi.org/10.1145/3718364>
- [16] Chaudhary, P.S., Khurana, M.R., Ayalasomayajula, M. (2024). Real-World Applications of Data Analytics, Big Data, and Machine Learning. In: Singh, P., Mishra, A.R., Garg, P. (eds) *Data Analytics and Machine Learning. Studies in Big Data*, vol 145. Springer, Singapore. [https://doi.org/10.1007/978-981-97-0448-4\\_12](https://doi.org/10.1007/978-981-97-0448-4_12)
- [17] Rane, N. L., Paramesha, M., Choudhary, S. P., & Rane, J. (2024). Machine Learning and Deep Learning for Big Data Analytics: A Review of Methods and applications. *Zenodo (CERN European Organization for Nuclear Research)*. <https://doi.org/10.5281/zenodo.12271006>
- [18] Theodorakopoulos, L., & Theodoropoulou, A. (2024). Leveraging big data analytics for Understanding Consumer Behavior in Digital Marketing: A Systematic review. *Human Behavior and Emerging Technologies*, 2024(1). <https://doi.org/10.1155/2024/3641502>