Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

# Efficient Data Mining of Political Results Through the Apache Hadoop Structure Unveiling Insights and Patterns in the Era of Big Data

**Chander Shekar**
*Professor, Department of Computer Science Engineering DITM, Sonipat*

***Abstract*:** In this era of incessant data generation, the sheer volume of information produced daily is staggering, surpassing the capacity of traditional data storage methods. The exponential growth of big data, now reaching beyond two billion terabytes, dwarfs even the concept of Exabytes. Such a colossal amount of data necessitates storage solutions of unprecedented scale and efficiency to manage, process, and analyze it effectively. it's not just the sheer volume of data that defines the big data landscape; it's also the speed at which this data is generated and processed. Referred to as the "Velocity of Big Data," this pace is relentless, with data being created, transmitted, and analyzed in real-time or near real-time. Every second, an unfathomable amount of information floods the digital realm from various sources such as sensors, mobile devices, social media platforms, online servers, and more. The emergence of contemporary big data infrastructure and technologies has largely been a response to this rapid expansion. Advanced analytics tools, distributed computing frameworks, and scalable storage solutions have become indispensable for organizations striving to harness the potential insights hidden within this deluge of data.As our reliance on electronic devices and interconnected systems continues to grow, the velocity, volume, and variety of big data will only escalate, posing challenges and opportunities for industries, governments, and society as a whole. Adapting to this data-driven paradigm requires innovative approaches to data management, analysis, and utilization to unlock its full potential for innovation, efficiency, and progress.

***Keywords:*** Big Data, Hadoop-Structure, Analytics, Devices, Technological, Information

## I. Introduction

Presently, the global population stands at 7.2 billion, with approximately 2 billion of these individuals possessing an internet connection. Other than this, five billion individuals are reported to use cell phones. As a result of this technological revolution, a considerable number of individuals generate vast quantities of data through the increased utilization of these devices. This data in constant creation is referred to as "Big Data." Big Data is a term used to describe the enormous volume of structured and unstructured data that has an impact on businesses. In 2012, Gartner reemerged with a more precise definition: "Big Data" refers to resources for data that are large in volume, fast in speed, or possibly diverse in nature. These resources necessitate novel forms of preparation in order to enable enhanced knowledge sharing, engagement optimization, and dynamic information retrieval." Synthesized tools for data processing are necessary to generate enormous results from big data. Therefore, the primary emphasis should not be on the quantity of data, instead focusing on the potential for data to provide insightful insights and data that elevate the significance of public aspects and allies, enabling them to provide more advanced services for residents or clients. Big Data has altered the received approach. Google determined in 2010 that, similar to the world's mechanism, the volume of data delivered was equivalent to the total it generated prior to 2003. Despite the recent "Colossal Data Leader Study 2013" by New Vantage Partners, which asserts that "It is about combination, not volume," numerous individuals (including industry experts) continue to hold the belief that scale or volume is the primary concern associated with massive data. Undoubtedly, vast amounts of data consist of an inconceivable range of forms, including text, images, records, noises, and anything else that may be incorporated into the performance, as well as their affective

compositions.Big Data has distinguished itself in the past period from startup capitalists, government and industry executives, market experts, and data experts. The exponential growth of information in the past decade has surpassed Moore's law, and the exponential percentage of information is exacerbating the scourge of exploration and regulation. Nevertheless, this vast amount of knowledge possesses an extraordinary capacity, and it conceals information that is incomprehensibly vital. Intelligent exposing that is data-centric facilitates the identification of Big Data issues. Big Data challenges are present in diverse sectors and domains, including financial activities, public health, intelligent investigation, and exceptional arranging implementation. Numerous advancements across diverse domains have been facilitated by Big Data, and there is no uncertainty that forthcoming challenges in business improvements will require an analysis of Big Data. The challenges associated with Big Data are minimal and can be circumvented through data acquisition, data analysis, data integration, and data retrieval. Initiating data exploration generally entails providing a concise overview of the principal features of a given data set, such as its magnitude, precision, beginning structures, and additional qualities. It is frequently performed by data analysts utilizing visual analytics tools, although more sophisticated statistical software, for example R, can also be employed.

Data exploration serves as the foundational stage in data analysis, wherein users unstructured examine a vast dataset in order to identify initial instances, attributes, and focal points. This procedure is not designed to reveal every piece of information contained in a dataset; instead, it aids in constructing a comprehensive representation of noteworthy patterns and focal points that can be examined in greater depth. This procedure simplifies subsequent analysis by aiding in the concentration on future endeavors and initiating the process of prohibiting irrelevant data centers and search methods that may yield no results. Even more critically, it is not an expertise with the most recent data that significantly simplifies the process of finding solutions. Data exploration frequently employs perception because it provides greater clarity on data sets than simply examining a large number of distinct names or figures. When conducting research on data, the manual as well as automated components each consider distinct facets of the same concept. Manual analysis facilitates user acquaintance with the data and can identify overarching patterns. The focus of exploration is increasingly directed towards Big Data. Both "Top 10 Critical Invention Trends For 2013" and "Top 10 Fundamental Technology Trends for the Next Five Years" list Big Data. This illustrates how domains such as financial analysis, policy formulation, and business associations are focusing on the implementation of Big Data. Big Data is categorized according to its volume, velocity, and variety. Subsequent individuals started providing novel 'V's in accordance with their specific needs. In this vein, the era of Big Data has arrived. Ranging in potential from 3Vs to 4Vs. At present, the fourth 'V' is assigned diverse attributes in accordance with the requirements, such as value, virtual, or fluctuation. Traditional methods of data preparation and management struggle to exploit vast and diverse datasets. To enhance the dynamic method, the novel preparation methods are required. Thus, the immense volume, tremendous variety, and high velocity of Big Data have been characterized. Extensively augmented data are those obtained from telescopes, deductive inquiries, sensor businesses, and high quantities devices. It illustrates the rate of growth in the computing limit and the need for data storage expansion.

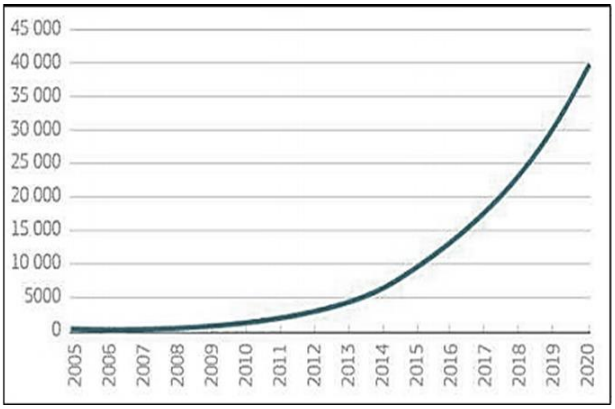


**Figure 1: A global explosion of data Source: "Big data as the bedrock of the future economy"**

The implementation of Big Data has resulted in significant transformations in traditional administration, planning strategies, and research methods. Data-focused processing is the domain in which the tools required to

manage Big Data concerns are examined. In addition to the 3 rational ideal models—observational science, hypothetical science, and computational science—data serious science is being introduced as a fourth viewpoint.

***Definition***

The phrase "big data" has been in circulation since the 1990s, although some attribute its popularity to John Mashey. Big data typically comprises datasets that are too large for frequently used software applications to efficiently capture, organize, control, and analyze in a reasonable amount of time. Although big data ideology accepts structured, semi-structured, and unstructured information, unorganized information is the primary emphasis. The "size" of big data is an ever-changing concept, with values ranging from several dozen terabytes to hundreds of zettabytes as of 2012. Big data necessitates a collection of methodologies and methods that incorporate novel forms of coupling in order to extract ideas from enormous, complicated and varied data sets. Others add the adjectives "variety" and "veracity," among others, to characterize it; a number of industrial experts has contested this revision. Frequently, the "three Vs," "four Vs," and "five Vs" were applied to the Vs of large data. They exemplified the volume, diversity, velocity, veracity, and worth of big data. Variability is frequently cited as an additional large data attribute. It is common to refer to Volume, Velocity, and Variety as the "3V's" of Big Data. The magnitude of big data surpasses two billion terabytes, and is significantly greater than exabytes, necessitating an enormous storage capacity. The pace or repetition of the information age is referred to as the "Velocity of Big Data." The contemporary iteration of Big Data has emerged exclusively in the past few years. As the proliferation of electronic devices that include sensors, smartphones, online media servers, web log files, clinical information, and predictive data continues, no moment passes without generating gigabytes of data. Thus, the rapid influx of data contributes to the overall volume of Big Data, which becomes problematic in and of itself. The term "variety" denotes the vast extent of information that is currently generated. There are three types of data structures: structured, semi-structured, and unstructured. Conventional database management software was designed to handle structured data, such as data that followed a specific structure, as opposed to data that could be stored in valid social data tables or databases. On certain occasions, the semi-organized data can also be considered rational. Nevertheless, it is highly implausible for conventional databases to manage and supervise Big Data to the extent that it could potentially be compromised. We have accessibility to a variety of data, including material arrangements, sound files, video files, and other types of data that cannot be confined to line and segment layouts. Analyzing such data will improve the efficiency of the organizations and generate greater returns. This Big Data must be managed and evaluated in order to maximize its potential. In order to anticipate much from the heaps, the organizations have initiated the implementation of "Big Data technologies" and are rethinking their foundations. Big Data technologies such as Spark, SAP-HANA, and High-Performance Cluster Computing (HPCC) are available in the marketplace. Some of these, Hadoop is the most widely utilized. It is not inaccurate to assert that the term "Hadoop" is being employed interchangeably. By utilizing big data. Hadoop was established by Doug Cutting as an Apache foundation. Hadoop makes use of the Map Reduce () framework created by Google, as well as the Hadoop Distributed File System, an additional file system.

## II. Related Work

**Sadia Zafar (2021),** Globally, massive amounts of data are produced in a variety of digital methods. The Big Data transformation anticipates that by 2020, a billion-plus gadgets will be linked to the high-speed internet, and the rapid prediction of vast information will attract the interest of academics, authorities, and industry professionals. Big Data is beneficial for increasing business productivity and facilitating evolving advances in numerous scientific disciplines. Nonetheless, it is indisputable that the management of Big Data presents numerous obstacles, including data visualization, preservation, evaluation, and the development of novel technologies to address Big Data issues. The purpose of the article is to illustrate the difficulties, novel tools, advantages, and uses of Big Data exploration so as to assist academics and users in selecting more suitable tools for their operations and requirements.

**Richard Millham (2021)**, As the implementation of data extraction becomes more pervasive in both academia and sector, there is an increasing demand for the development of necessary tools and reliance on their results for valuable insights. As the numerous components of vast amounts of information evolved, a multitude of instruments emerged to facilitate the application of these perspectives. Diverse clientele had distinct demands

for a given aspect of data extraction. Moreover, these tools originated as instruments for rudimentary data mining projects within software-oriented organizations before evolving into more specialized and specialized applications within smaller organizations or groups. In this discussion, we will examine the various types of thoughts that can influence the consumer base of a specific instrument. In conclusion, we shall describe various instruments in terms of the tasks they accomplish and the clientele for which they are designed.

**Zhaohao Sun (2020),** The appeal of big data and analytics intelligence to academia and industry has been substantial. As a newly formed area for big data analysis, business intelligence, and artificial intelligence (AI), analyze big data analytics intelligence and big data analytics reasoning. In particular, this course will cover the foundational principles of analytics thought and analytics intelligence in accordance with the most recent advancements in academic and industrial research as well as creation. The section then examines the interrelationships between big data analytics intelligence and data analytics intelligence, as well as big data analytics intelligence and data analytics reasoning. A united technical basis for AI-driven big data analytics intelligence is proposed in this article as a comprehensive strategy for big data analytics intelligence. Exploration and advancement in the fields of business intelligence, big data analytics reasoning, big data analytics intelligence, artificial intelligence, and data science could be facilitated by the method suggested in this article. Data science, big data intelligence, big data analytics, analytics reasoning, and artificial intelligence (AI) are all essential terms.

**Zhaohao Sun (2020)**, Big data and analytics intelligence have garnered considerable interest from both academics and industry. As an emergent area for big data analytics, business intelligence, and artificial intelligence (AI), this article will analyze big data analytics intelligence and big data analytics reasoning. In particular, the foundational principles of analytics reasoning and insights intelligence as they pertain to the most recent advancements in academic and industrial research and creation. The section then examines the connections between big data analytics expertise and data analytics intelligence, as well as big data analytics intelligence and data analytics reasoning. A united technical basis for big data analytics intelligence powered by AI; an all-encompassing approach to big data analytics intelligence. Development and investigation in the fields of company intelligence, big data analytics reasoning, big data analytics intelligence, artificial intelligence, and data science could be facilitated by the method suggested in this article.

**Ren et al. (2019),** Certain aspects of the aforementioned definitions are complementary, including the 5Vs definition of big data. Moreover, certain of them are in conflict with the 6 illustrative meanings that were utilized as a starting point. They characterized big data in terms of three "Vs" and ignored every other factor in favor of emphasizing data magnitude. When viewed through the lens of user comprehension, these definitions illustrate various facets of big data as it is implemented in companies and studies. Certain aspects of them prioritize features pertaining to volume, variety, and velocity. Conversely, others shift their attention towards their purpose and demands, including the business standards and the method of data storage. Nevertheless, the term utilized in this study encompasses every dimension, specifically the 5Vs. This is due to its perceived high density, rapidity, and diversity in framework, layout, and source material, all of which necessitate handling with high performance. It is evident from the various definitions of big data that scale predominates, notwithstanding the significance of other attributes. The three V's, which were proposed as the elements of the obstacle to data management, comprise a single structure. All three of these factors are not separate from each other; a rise in one dimension enhances the likelihood that the other two will also undergo a change.

**Philipp Reichenbach (2019)** is intrigued by the influence of the media on written and spoken language before the advent of the Internet and Web 2.0. Over 400 BC, Plato had already addressed the correlation between spoken and written language. Opportunities presented by the advent of digitalization include technical capabilities for language identification, assessment, and research. Scholarly fields including digital humanities and semantics are not solely concerned with the intersection of digitalization and language, but also delve into the possibilities and challenges that arise from this interplay. The application of Big Data to corpus linguistics research. Within the scope of this paper, I analyze the Big Data corpus GloWbE (Corpus of Global Web-based English) as a linguistic research instrument by spotlighting its benefits, placing emphasis on crucial elements, and proposing constructive methodologies and ideas. I offer some recommendations regarding how researchers might approach the aforementioned concerns pertaining to morality, geographical limitations, and non-native

speakers. diverse authors, technological obstacles, accessibility and involvement. I conclude with a summary of the findings in the form of opportunities and difficulties associated with the use of Big Data corpora as instruments for language studies, as well as suggestions for future research. I am consistently delighted to receive constructive criticism.

**Saurabh Garg (2018)** Current developments in big data methods, including NoSql databases, Hadoop, Storm, and Spark, have facilitated the development of software alternatives. Possessing such a wide variety of instruments for various data processing duties does, nevertheless, complicate the process. The development, execution, installation, and administration of the majority of big data systems necessitate specialized expertise in various platforms and tools; therefore, orchestration is a difficult task. The difficulty is further compounded by the standard of service demands of users, the unpredictability of big data, and the foundational nature of cloud infrastructure. The present section provides a review of the various operations involved in the collaboration of big data solutions, obstacles encountered in accompaniment, the current state of the art, and specific unresolved problems pertaining to big data collaboration.

**Rahul Khullar (2018)** Data is now an indispensable component of every organization, industry, country, and business capability. Diverse corporations, machinery, and institutions are exponentially increasing their research information in the developed globe. As this data set grows, the task of selecting the relevant information becomes increasingly challenging. As a result, the rapidity, volume, and variety of this sophisticated world contribute to the development of a data categorization known as "Big Data," which is characterized by its on-demand and command-based nature. Its primary function is to partition the various datasets whose quantities exceed the ability of the database management software to monitor, evaluate, and keep. It presents specific computing and logical challenges, including prediction errors, storing and flexibility constraints, and noise accumulation. Due to a specific characteristic of the Big Data, it is stored in the distributed document framework Hadoop (HDFS). Despite this, Hadoop is reasonably confounding. Given the novelty of Hadoop to users, this research paper investigates the substantial challenges and concerns encountered throughout the information extraction and storage structure organization processes.

Lee (2017) The term "Big Data" is not entirely novel; its origins can be traced back several decades prior to the current frenzy surrounding it. The utilization of data analysis and methods of analytics to aid in decision-making commenced several decades ago. However, over the past twenty years, the exponential growth of data has surpassed human capacity to comprehend, primarily due to the widespread adoption of social networking sites and the internet. Describe Big Data using the three V's: "A situation in which the quantity, velocity, and variety of data stored by a company exceed the computational capabilities required to make accurate and timely decisions." 2001's 3 V's of Big Data had expanded to 42 V's by the conclusion of 2017.

**K. G. Srinivasa (2017)** A multitude of alternatives are available for selection in our software for big data analyses. Team or organizational structures, seller tool types and tool characteristics, and user methods and approaches are among these alternatives. There are numerous complex elements on the list that we may not have given due consideration to. Regardless of the phase of a Big Data mining project one is in, understanding the accessible alternatives is fundamental for making informed choices regarding software or hardware items to assess and approaches to pursue.

## III. Proposed Methodology

Coming up next are the different strides in the calculation for AI:

**Stage 1**: Input the information into the Spark data outline and request time.

**Stage 2:** Randomly divide the data into test data using the Spark library's Split function.

**Stage 3:** Vector Assembler changes over the data into terms of vectors.

**Stage 4:** Vectors are transformed into foundational data outlines through transform operations.

**Stage 5:** Map the label col and feature col utilizing Machine Learning algorithms from MLlib.

- Linear Regression

- Decision Tree
- Random Forest
- Gradient Boosting Tree

**Stage 6**: Pipeline is composed of a transformer and evaluator in phases to manage the Machine Learning process.

**Stage 7:** Fit the preparation data into the model for forecast.

**Stage 8:** Regression Evaluator is utilized to assess the forecast on the included data.

**Stage 9:** RMSE is determined to track down the mean square error.

**Stage 10:** Accurate forecast value is seen from various learning techniques.

**Stage 11:** Time and Space are the boundaries experimental on Machine Learning algorithms.

### 3.1 Flowchart for the Proposed System

The flowchart illustrated in Figure 2. illustrates the complete utility of the proposed model. The dataset is predominantly utilized across a collection of four devices. The attended information is transformed into vector parts and subsequently modified. The process of the regression evaluator commences the forecast result after the change phase. The accuracy of each forecast extracted from learning algorithms is evaluated. The reality complexities of the two systems MapReduce and Spark are evaluated in relation to each learning strategy.

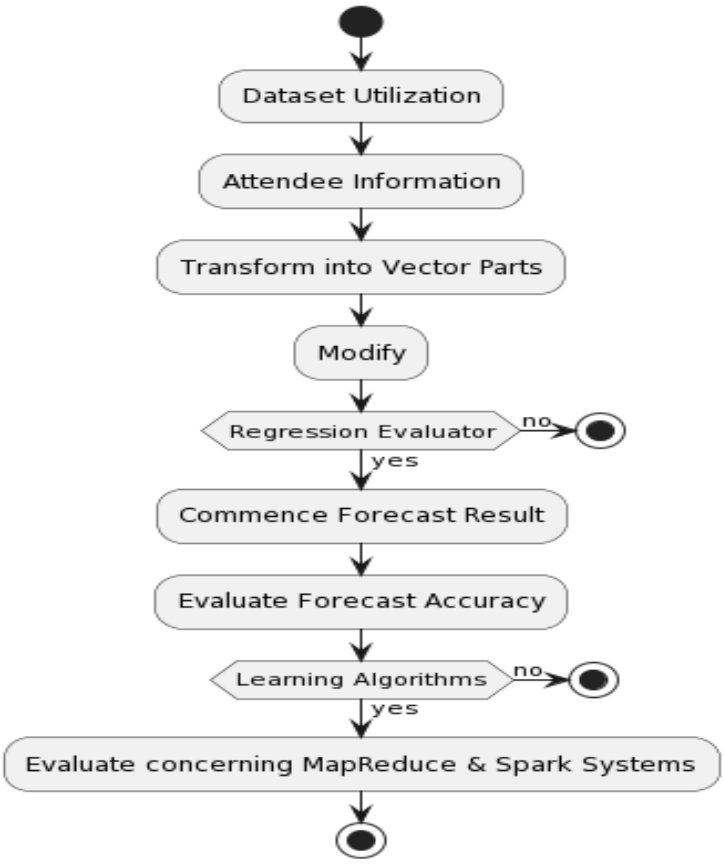


**Figure 2: Flowchart for the Functionality of the Proposed System**

## III.Result Analysis

The adopted model employs various artificial intelligence (AI) algorithms for forecasting purposes, including Linear Regression, Random Forest, Decision Tree, and the Gradient Boosting Tree algorithms. The approximate results of the various learning algorithms are presented in Table 1.

**Table 1: Prediction Analysis from Various Machine Learning Methods**

| Year | Linear Regression | Random Forest | Gradient Booster | Decision Tree |
|---|---|---|---|---|
| 1905 | 23.8106 | 19.045 | 12.697 | 15.825 |
| 1910 | 24.2519 | 19.045 | 19.2367 | 15.825 |
| 1915 | 24.4107 | 19.045 | 19.2367 | 15.825 |
| 1920 | 24.2019 | 18.9079 | 19.2367 | 24.08 |
| 1925 | 25.451 | 18.9079 | 19.0267 | 24.08 |
| 1930 | 23.4775 | 19.04 | 19.0267 | 24.08 |
| 1935 | 24.2686 | 19.04 | 18.967 | 24.169 |
| 1940 | 24.6349 | 19.103 | 19.28 | 24.169 |
| 1945 | 23.8023 | 18.62 | 19.28 | 23.963 |
| 1950 | 24.5434 | 19.317 | 19.28 | 23.963 |
| 1955 | 24.1936 | 19.317 | 29.33 | 24.0795 |
| 1960 | 23.8855 | 19.317 | 29.33 | 24.0795 |
| 1965 | 24.7349 | 19.317 | 19.225 | 24.2567 |
| 1970 | 23.719 | 19.988 | 19.3716 | 24.2567 |
| 1975 | 23.744 | 19.456 | 19.3716 | 24.276 |
| 1980 | 24.4518 | 19.456 | 19.3716 | 24.267 |
| 1985 | 24.2436 | 19.456 | 19.337 | 23.8818 |
| 1990 | 25.2511 | 19.2493 | 19.554 | 23.8818 |
| 1995 | 24.6349 | 19.08 | 19.554 | 24.1435 |
| 2000 | 24.3019 | 19.08 | 19.554 | 24.1435 |
| 2005 | 23.5327 | 19.457 | 19.475 | 24.517 |
| 2010 | 24.7825 | 19.251 | 19.147 | 23.486 |
| 2015 | 25.2378 | 19.249 | 19.357 | 24.143 |
| 2020 | 24.5673 | 19.364 | 19.264 | 24.0245 |

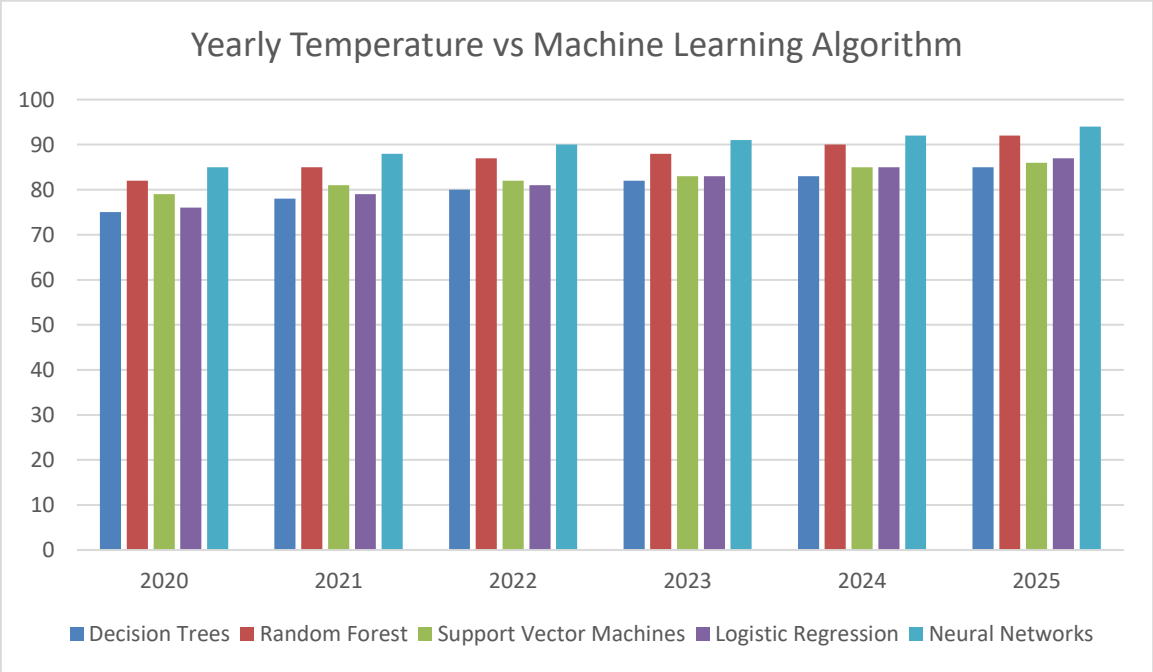


**Figure – 3: Predictive Values based on Various Machine Learning Techniques**

- ***Observation:*** Figure 3: The comparable outcomes of yearly temperatures and expectation values generated by machine learning algorithms. The saw value reveals the annual temperature accuracy of the random

forest. It demonstrates exceptional precision in comparison to alternative learning methodologies.

## IV. Conclusion

The study explores the comparative analysis and underscores the substantial performance advantages of the Spark framework over MapReduce in data mining tasks, particularly in processing political results. While Map Reduce exhibits commendable throughput rates during read and write operations, its performance on the processor is comparatively diminished. In contrast, Spark demonstrates superior execution times for both read and combine operations, indicating more optimized resource utilization and enhanced processing efficiency.The comprehensive evaluation, encompassing various metrics such as time, space, read, and create tasks, further solidifies the superiority of Spark, showcasing a remarkable 65.21 per cent improvement over MapReduce. These findings highlight the significance of leveraging advanced frameworks like Spark in handling big data analytics tasks, especially in domains like political analysis where timely and efficient processing of vast datasets is crucial for informed decision-making and insights generation.

## References

[1] Yassir Rochd “Performance Improvement of PrePost Algorithm Based on Hadoop for Big Dat” 2018
[2] YichuanWang, LeeAnn Kung, Terry Anthony Byrd (2018). ‘Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations’ Technological Forecasting and Social Change Volume 126, January 2018, Pages 3-13, Elsevier, Science Direct.
[3] Zaharia, Matei; Chowdhury, Mosharaf; Das, Tahagata; Dave, Ankur; Ma, Justin; McCauley, Murphy; Franklin, Michael; Shenker, Scott; Stoica, Ion;. Resilient Distributed Datasets: A Fault-Tolerant. Berkeley : University of California at Berkeley, Electrical Engineering and Computer Sciences, 2011.
[4] Zhaohao Sun “Big Data Analytics Thinking and Big Data Analytics Intelligence” 2020
[5] Cuzzocrea, A., Song, I.-Y., & Davis, K. C. (2011). Analytics over Large-scale Multidimensional Data: The Big Data Revolution! In Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP (pp. 101–104). New York, NY, USA: ACM. http://doi.org/10.1145/2064676.2064695
[6] D. Che, M. Safran, Z. Peng, "From big data to big data mining: challenges, issues, and opportunities." International conference on database systems for advanced applications. Springer, Berlin, Heidelberg, 2013.
[7] D. P. Acharjya, Kauser Ahmed P (2016). ‘A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools’, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016, pp 511-518.
[8] Daqian Wei, Bo Wang, Gang Lin, Dichen Liu, Zhaoyang Dong, Hesen Liu, and Yilu Liu. Research on unstructured text data mining and fault classification based on rnn-lstm with malfunction inspection report. Energies, 10(3), 2017.
[9] Davidov, D., Rappoport, (2006) ‘A. Efficient Unsupervised Discovery of Word Categories Using Symmetric Patterns and High Frequency Words’. In Proceedings of the International Conference on Computational Linguistics, Sydney, Australia, 17–21 July 2006; pp. 297–304.
[10] Citizen Science: The Law and Ethics of Public Access to Medical Big Data. Hoffman, Sharona. 3, Cleveland : Case Western Reserve University School of Law, 2014, Vol. 30.
[11] Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). The Integration of Business Intelligence and Knowledge Management. IBM Syst. J., 41(4), 697–713. http://doi.org/10.1147/sj.414.0697
[12] Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. (2009). MAD Skills: New Analysis Practices for Big Data. Proc. VLDB Endow., 2(2), 1481–1492. http://doi.org/10.14778/1687553.1687576
[13] Constantiou, I.D. and Kallinikos, J., 2015. New games, new rules: big data and the changing context of strategy. Journal of Information Technology, Volume 30, pp. 44-57.
[14] Galletta, A., & Cross, W. E. (2013). Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication. New York: NYU Press.
[15] Gantz J, Reinsel D. The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East. New York: IDC iView: IDC Analyse future; 2012
[16] Gartner. (2014). Gartner’s 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital

Business. Retrieved August 16, 2015, from http://www.gartner.com/newsroom/id/2819918

[17] H. Lee, R. Gross, R. Ranganath, and A. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations" In ICML, 2009

[18] Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, Volume 35, pp. 137-144.

[19] HannahSnyder "Literature review as a research methodology: An overview and guidelines" 2019

[20] Herodotou, H., Lim, H., Luo, G., Borisov, N., Dong, L., Cetin, F. B., & Babu, S. (2011). Starfish: A Self-tuning System for Big Data Analytics. In In CIDR (pp. 261–272).