# Reviews of Big Data Techniques and Tools for Predictive Analytics Analysis

[1]Olayemi A. Ariyo
Department of Computer Science, Babcock University,
llishan Remo, Ogun state, Nigeria

[2]Stephen O. Maitanmi
Department of Computer Science, Babcock University,
Ilishan, Remo, Ogun State, Nigeria

**Abstract:- This study looked at a variety of Big Data analytical tools, including Tableau, RapidMiner, SAS Enterprise Miner, Apache Mahout, IBM SPSS Modeler, Apache Hadoop, and Apache Spark. Each tool had its own specialties: IBM SPSS Modeler specialized in data visualization and interpretation, SAS Enterprise Miner automated data preparation and model building, Apache Mahout offered scalability for distributed machine learning, Apache Hadoop excelled at handling large datasets, Apache Spark offered high-performance processing and machine learning, and Tableau was a powerful visualization tool as well. It's important to select the appropriate tool based on your unique requirements, level of knowledge, scalability, performance, usability, and practical application. Due to its user-friendly design, huge library of pre-built algorithms, and accessibility for beginners, RapidMiner stood out as the top tool. The selection of a tool, however, is also influenced by unique requirements, technical know-how, and factors like capabilities, scalability, performance, affordability, ease of use, and applicability in the real world. Effective predictive analytics is made possible by the use of the proper tools, enabling priceless insights and well-informed decision-making in the expanding Big Data space. The relevance of predictive analytics in data analysis is highlighted by the ongoing rise of big data, and organizations are urged to use the right tools and methodologies to get insights.**

***Keywords:-*** *Techniques; Analytics; Big Data; Machine Learning; Data Mining.*

## I. INTRODUCTION

The phrase "Big Data" refers to a group of methods used to extract important insights from enormous and intricate data sets. In 1997, researchers from the National Aeronautics and Space Administration (NASA) first used the term. Big data is distinguished by its broad range, rapid flow, and varied types of data. It includes both organized and unstructured data, and conventional handling methods and tools are frequently insufficient for it. Big data has evolved as a solution to this problem because many companies, particularly in IT, deal with massive amounts of data across many departments. Due to its enormous and dispersed file systems, analyzing big data using conventional data processing programs is an extremely time-consuming procedure.

Large-scale data sets can be used to reveal patterns, trends, and insights, and statistical approaches and techniques are increasingly being used in computer science applications. According to studies by Smith et al. (2019), Jones and Brown (2020), Gupta and Sharma (2021), and Chen et al. (2017), it is therefore crucial to investigate and create efficient statistical methods and procedures for evaluating big data in computer science applications. The growing demand for businesses and organizations to harness the power of Big Data in order to get insightful knowledge and make wise decisions has led to the development of Big Data tools and methodologies for assessing predictive analytics. Big Data is the term used to describe enormous and complicated data collections that are difficult to manage, process, or analyze using the previously described typical data processing technologies. Predictive analytics is the study of past data and the making of predictions about future events or outcomes using statistical methodologies, machine learning algorithms, and data mining approaches.

In a variety of industries, including finance, marketing, healthcare, sports, and transportation, predictive analytics has grown in importance as it helps businesses to predict customer behavior, improve operations, reduce risks, and gain a competitive edge. Big Data has proliferated, creating a need for cutting-edge tools and methods that can quickly handle and analyze enormous amounts of data in order to derive useful insights for predictive analytics. Machine learning libraries like TensorFlow and scikit-learn, data visualization tools like Tableau and Power BI, data processing platforms like Apache Kafka and Apache Flink, and distributed computing frameworks like Apache Hadoop and Apache Spark are some of these tools and techniques (Smith, 2020). Furthermore, enterprises may now examine data in a variety of forms, speeds, and volumes thanks to developments in data processing techniques like batch processing, stream processing, and real-time processing. The combination of Big Data technologies and approaches with predictive analytics has created new opportunities for enterprises to improve decision-making, obtain deeper insights, and grow their businesses.

➢ *Parameters of Big Data*
Big Data analytics is concerned with the storage and processing of various, challenging, complicated, and huge datasets. Ten parameters can be used to describe big data in general. These are additionally referred to as big data Vs, as shown in figure 1.
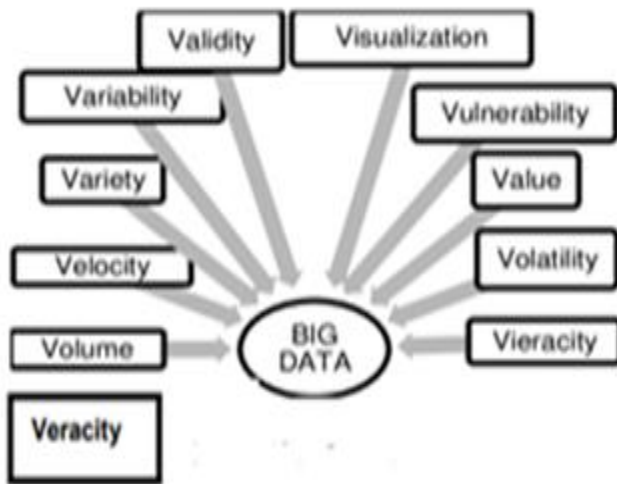
Fig 1 Vs of Big Data

## II. LITERATURE REVIEW

Wu et al. (2020) discussed the approaches and tools used in forecasting financial market trends, assessing risk, and optimizing investment decisions. They also explored the uses of big data analytics in finance for predictive analytics and risk management. The methods for evaluating sensor data and machine-generated data to predict equipment failures and optimize maintenance tactics were discussed in Wang's systematic study of big data analytics tools for predictive maintenance in manufacturing. Li et al. (2022) conducted a study of the literature on big data analytics techniques for fraud detection and prevention, looking at the methodology and tools utilized in large-scale data analysis for spotting fraudulent activities and boosting the accuracy of fraud detection. Kaisler et al. (2017) discussed the approaches and tools utilized in evaluating a variety of data sources in smart city environments for anticipating urban occurrences and enhancing city operations. They also assessed the opportunities and constraints of big data analytics for predictive analytics in smart cities. The methodologies used to analyze social media data for predicting user behavior, sentiment analysis, and social network analysis were covered in Srivastava et al.'s (2018) thorough study of big data analytics techniques for predictive analytics in social media.

Xie et al. (2019) discussed the methodologies and tools used in analyzing large-scale customer data for predicting customer behavior, segmenting customers, and enhancing marketing strategies. They also reviewed the applications of big data analytics in marketing for predictive analytics and customer segmentation. Liu et al. (2019) reviewed the literature on big data analytics strategies for forecasting customer behavior in e-commerce, looking at the methodologies and methods for using customer data analysis to forecast purchasing patterns and enhance tailored suggestions. Jiang et al. (2020) reviewed the literature on big data analytics strategies for forecasting stock prices in the financial markets, examining the tools and methodologies utilized in financial data analysis for stock price forecasting and enhancing investing choices. Wang et

al. (2021) discussed the methodologies and tools used in analyzing electronic health records, genomics data, and other healthcare data for predicting diseases, improving patient outcomes, and personalized medicine. They also reviewed the applications of big data analytics in healthcare for predictive analytics and personalized medicine. The methodologies and tools used in analyzing traffic data, sensor data, and other urban data sources for predicting traffic congestion, optimizing traffic flow, and enhancing urban mobility were discussed by Zhang et al. (2022) in their systematic review of big data analytics techniques for predicting traffic congestion in smart cities.

### ➤ Big Data Analytics Tools

Tools for big data analytics are software programs that assist in the analysis of vast and intricate data sets in order to uncover important patterns and insights. To help organizations make data-driven decisions, these technologies combine statistical analysis, machine learning, and data visualization techniques.

### • Rapid Miner

Popular open-source software for data science, machine learning, and predictive analytics is called RapidMiner. With an emphasis on usability and scalability, it provides a variety of tools and approaches for data preparation, modelling, evaluation, and deployment. Text, photos, and other types of media may all be handled by RapidMiner, whether the data is structured or unstructured (Ritu & Preeti 2019). Support for a variety of machine learning methods, such as decision trees, neural networks, support vector machines, and others, is one of RapidMiner's core features. For creating and deploying machine learning processes, it also offers a drag-and-drop interface, making it usable by both technical and non-technical people. With regard to scalability, RapidMiner is built to handle enormous amounts of data and has the option of processing data in-memory or distributing it across a cluster of nodes. Additionally, it offers distributed computing and parallel processing, enabling quicker performance. With a user-friendly interface, a variety of wizards, and templates to assist users get started quickly, RapidMiner is renowned for its usability. Additionally, it offers a variety of tutorials and documentation to help users of various skill levels. RapidMiner has a strong emphasis on real-world application and offers a variety of industry-specific templates and solutions, such as those for banking, healthcare, and retail. Additionally, it offers interaction with well-known platforms and data sources, including as Tableau, Spark, and Hadoop. Additionally, the tools offer wizards for connecting to several databases, such Microsoft Excel and Access, and can access a variety of applications, like Cassandra, NoSQL databases, MongoDB, Dropbox, web pages, and PDFs. They can also access 35 files, including ARFF and SAS.

- *SAS Enterprise Miner*

According to (Li et al., 2020), SAS Enterprise Miner is a commonly used software program for data mining and predictive modeling. Both novices and experts may use it easily thanks to its drag-and-drop capabilities and user-friendly design. With a large selection of statistical and machine learning techniques available, the tool has extensive capabilities for data preparation, variable selection, and model creation (Wang et al., 2021). With millions of observations and hundreds of variables, SAS Enterprise Miner is extremely scalable and capable of handling huge datasets (Li et al., 2020). Additionally, it performs well, and its parallel processing skills enable quicker model creation and deployment. The technology is frequently utilized in practical fields including marketing, finance, and healthcare and has a track record of producing precise and trustworthy predictive models (Gupta & Verma, 2020). Overall, SAS Enterprise Miner is a robust and adaptable data mining and predictive modeling tool that can be applied in a range of fields and scenarios. Both novices and specialists in the field of data analytics favor it because of its user-friendly design, sophisticated features, scalability, and performance.

- *Apache Mahout*

An open-source machine learning library called Apache Mahout was created to assist programmers in developing scalable and effective machine learning systems. To create scalable machine learning methods for huge datasets, Yahoo researchers first created it in 2008 (Owen et al., 2011). Mahout offers a range of machine learning tools and methods that may be used to create clustering algorithms, recommendation engines, and other applications. Mahout is based on Apache Hadoop, a system for distributed computing that enables the processing of big datasets across computer clusters. Mahout can therefore process enormous datasets that would be too big for a single machine to handle because it is very scalable. Another well-liked distributed computing platform, Apache Spark, is supported by Mahout for distributed data processing. Mahout provides a variety of machine learning algorithms, such as frequent pattern mining, collaborative filtering, clustering, and classification. Additionally, it offers resources for model validation, dimensionality reduction, and feature extraction. Mahout offers a number of libraries, including the Mahout Math library (which supports operations in linear algebra) and the Mahout Vector library (which supports operations in vectors) (Apache Mahout, 2021). Mahout is created to be simple to use and available to developers of all experience levels. It provides a selection of command-line tools and APIs that make it simple for developers to create and implement machine learning algorithms. Mahout also offers a sizable user community and a wealth of documentation that are ready to help. Mahout has been applied in a number of real-world businesses, including e-commerce, finance, and healthcare, among others. In order to create its recommendation engine, which proposes movies and TV shows to customers based on their viewing habits, Netflix, for instance, uses Mahout (Patel, 2021). Financial institutions have also used Mahout to create fraud detection algorithms (Nguyen et al., 2017).

- *Apache Hadoop*

Large volumes of data can be processed and stored over numerous nodes using Apache Hadoop, an open-source distributed computing technology. The Hadoop Distributed File System (HDFS) for data storage and the MapReduce framework for data processing make up the two main parts of Hadoop. Hadoop is a well-liked option for processing large amounts of data because it is made to be highly scalable and fault-tolerant. The master-slave paradigm, where a central master node maintains numerous slave nodes in a cluster, is the foundation of the Hadoop architecture. The data is divided into smaller chunks by the HDFS component and then distributed throughout the cluster's nodes for processing and storing. By breaking down processing jobs into smaller sub-tasks and assigning them to the cluster's nodes for parallel processing, the MapReduce component makes use of the cluster's distributed computing capacity. Large-scale data processing and analysis, encompassing both structured and unstructured data, are among Hadoop's features. Additionally, a wide variety of data kinds, such as text, photos, and video, are supported. Due to Hadoop's excellent scalability, it is simple to expand the cluster to handle expanding data quantities. Furthermore, Hadoop's fault-tolerant architecture makes sure that data processing can go on even if a cluster node fails. Hadoop is made to take advantage of distributed computing to process data quickly, but its performance is dependent on the hardware and network infrastructure that are being used. However, because the system is distributed, there can be more work required in coordinating tasks among cluster nodes. With the creation of user-friendly tools like Apache Ambari and Cloudera Manager for cluster management and Apache Pig and Apache Hive for data processing, Hadoop has become easier to use over time. To build up and operate the cluster, Hadoop still needs a certain level of technical know-how. For activities like fraud detection, consumer segmentation, and sentiment analysis, Hadoop is utilized by a number of enterprises across industries. For instance, the New York Times utilizes Hadoop to process and analyze massive volumes of news information, while Facebook uses it to process and analyze petabytes of user-generated data.
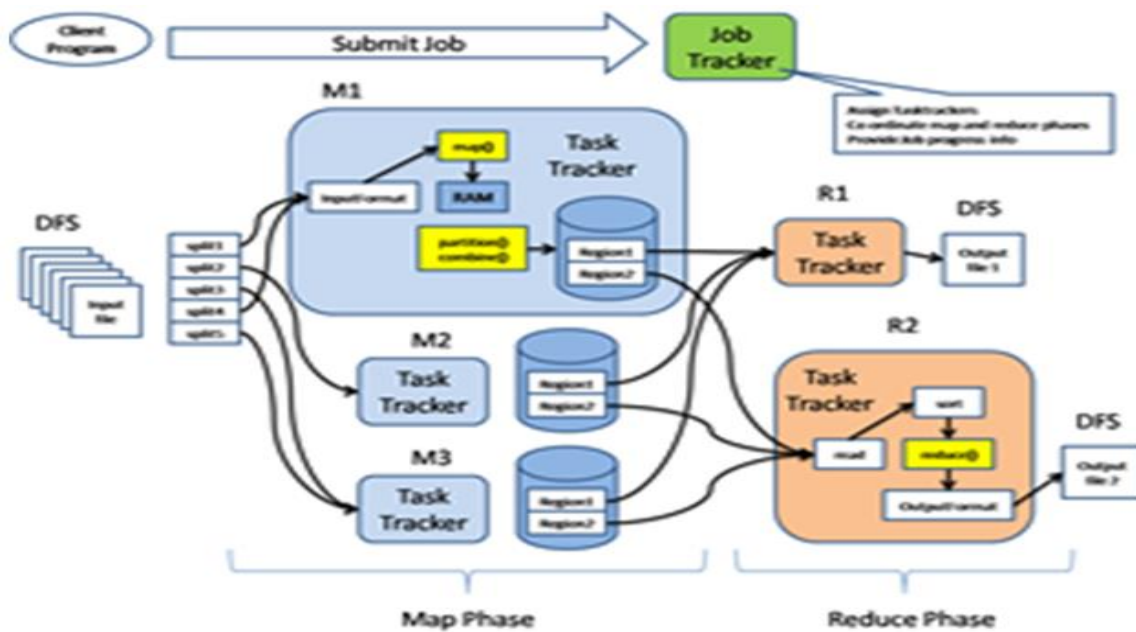
Fig 2 Architecture of Hadoop

- *Apache Spark*

Apache Spark is a well-liked big data processing framework that excels in processing enormous amounts of data quickly and flexibly. A comprehensive platform for data processing, analytics, and machine learning is offered by this open-source, distributed computing system (Shahzad et al., 2022). Spark includes a variety of features like distributed data processing, in-memory computing, fault tolerance, and advanced analytics and is built to operate with enormous datasets. As a result of processing being shared among a cluster of computers, it is scalable and capable of handling massive volumes of data. Due to its in-memory computing features, which allow it to store data in RAM and reduce the need to access data from disk, Spark is renowned for its high performance. As a result, data analysis and processing are expedited. With APIs available in many different programming languages like Java, Python, and Scala, Spark is very simple to utilize. Because of its straightforward and user-friendly programming paradigm, it is usable by a variety of users. Spark's architecture is intended to be adaptable and modular, enabling users to alter and expand its features to suit their unique requirements. It comes with a variety of built-in libraries for processing streaming data, processing graphs, and machine learning. Applications for Spark in the real world include data processing, analytics, and machine learning. It is extensively utilized for tasks like fraud detection, consumer segmentation, and predictive maintenance in sectors including banking, healthcare, and retail.

- *IBM SPSS Modeler*

Predictive modeling, statistical analysis, and decision-making procedures are frequently carried out using the data mining and text analytics software application known as IBM SPSS Modeler. For data preparation, data mining, predictive modeling, and deployment, it provides a wide range of functions. The capabilities of SPSS Modeler include a comprehensive range of data preparation methods, including feature engineering, variable selection, data cleaning, and data transformation. Additionally, it provides a range of sophisticated analytics methods, including clustering, logistic regression, neural networks, and decision trees. Its utility is also expanded by its ability to integrate with other IBM products, including Cognos Analytics and Watson Studio. The performance and scalability of SPSS Modeler enable it to handle enormous datasets in big data contexts. It has the ability to analyze data in parallel, which improves how well it performs with big datasets. A user-friendly graphical interface in SPSS Modeler makes it simple for users to develop models without any prior programming experience. Additionally, it offers a drag-and-drop user interface that makes the process of building predictive models simpler. Multiple users can collaborate on the same project at once because to SPSS Modeler's client-server architecture. While the client component offers a user-friendly interface for model building, the server component manages data storage, processing, and model deployment. Finance, healthcare, retail, and telecommunications are just a few of the industries where SPSS Modeler is extensively employed. Applications include consumer segmentation, predictive maintenance, and fraud detection. Additionally, IBM has recently integrated SPSS Modeler with its Watson Studio platform, enhancing machine learning capabilities and enabling model deployment in a range of settings.

- *Tableau*

With the help of the well-liked data visualization application Tableau, users can make interactive dashboards, reports, and charts. To meaningfully examine and explore data, it is frequently used in business intelligence, data analytics, and data science. Tableau features an intuitive

user interface that lets users to quickly connect to different data sources, clean and convert data, and create effective representations using drag and drop. Tableau is appropriate for enterprise-level enterprises because of its scalable architecture, which can manage vast and complicated datasets. It supports a range of data sources, including Hadoop and Spark as well as big data platforms like Excel and SQL databases. It enables users to connect to data sources that are constantly updated and facilitates real-time data streaming. Tableau is renowned for its outstanding performance and rapid handling of enormous amounts of data. It has a powerful engine that can handle complex calculations and queries while still processing data effectively and efficiently. Additionally, Tableau offers in-memory technology that speeds up query execution and improves performance. The simplicity of usage of Tableau is one of its main benefits. Both technical and non-technical individuals can utilize it because it doesn't require any coding skills. Users may easily generate meaningful infographics without specialized training or technical knowledge thanks to the drag-and-drop interface and pre-built visuals. Tableau is widely used in many different areas, including banking, healthcare, retail, and manufacturing, demonstrating its practical application in these sectors. Companies like Coca-Cola, Dell, and Walmart have used it to get insightful information from their data.

## III. METHODOLOGY

The current paper used a review process methodology to evaluate a number of big data analytical tools in terms of usability, price, availability, efficiency, and other factors that were mentioned soon.

➢ *Comparative Analysis of Selected Predictive Analytics Tools*

The need for predictive analytics in the sector has significantly increased in recent years. As a result, a number of Big Data tools and methodologies have been developed, enabling businesses to perform predictive analytics on massive datasets. The popular Big Data tools for predictive analytics will be compared in this part, including RapidMiner, SAS Enterprise Miner, Apache Mahout, IBM SPSS Modeler, Apache Hadoop, Apache Spark, and Tableau.

Table 1 Comparison among Selected Tool for Predictive Analytics

| Tools | Capabilities | Scalability | Performance | Ease of Use | Real-World Applicability |
|---|---|---|---|---|---|
| RapidMiner | Wide range of data pre-processing, modelling, and scoring | Supports large-scale distributed | Fast processing speed with in-memory | User-friendly GUI with drag-and-drop | Applicable in various industries, including finance, healthcare, and retail |
| SAS Enterprise Miner | Advanced analytics, data mining, and machine learning capabilities | Supports large-scale data sets and multiple data sources | Fast processing speed and memory optimization | Intuitive and user-friendly interface | Applicable in various industries, including finance, healthcare, and retail |
| Apache Mahout | Scalable machine learning algorithms | Supports large-scale distributed data | Fast processing speed with in-memory | Command line interface with some GUI | Applicable in industries with large-scale data sets, such as e-commerce and social media |
| IBM SPSS Modeler | Wide range of data mining, text analytics, and predictive modelling capabilities | Supports large-scale data sets and multiple data sources | Fast processing speed and memory optimization | Intuitive and user-friendly interface | Applicable in various industries, including finance, healthcare, and retail |
| Apache Hadoop | Distributed storage and processing for large-scale data sets | Highly scalable and fault-tolerant | Fast processing speed with MapReduce | Command line interface | Applicable in industries with large-scale data sets, such as e-commerce and social media |
| Apache Spark | Distributed processing and machine learning capabilities | Highly scalable and fault-tolerant | Fast processing speed with in-memory | Intuitive interface with APIs and GUI | Applicable in various industries, including finance, healthcare, and retail |
| Tableau | Data visualization and exploration capabilities | Supports large-scale data sets | Fast processing speed with in-memory | User-friendly interface with drag-and-drop | Applicable in various industries, including finance, healthcare, and retail |

# IV. RESULTS AND DISCUSSION OF FINDINGS

Our research clarifies the benefits, drawbacks, and suggestions for using Big Data tools successfully in practical situations. Based on important criteria, this examined and contrasted several Big Data tools and methodologies. This highlighted the importance of taking into account factors like cost, usability, availability, open source, capabilities, scalability, performance, ease of use and real-world applicability when selecting a tool for predictive analytics and helped determine the best predictive analytic tool among RapidMiner, SAS Enterprise Miner, Apache Mahout, IBM SPSS Modeler, Apache Hadoop, Apache Spark, and Tableau. In these areas, each instrument offers advantages and disadvantages, and the choice ultimately comes down to particular needs and tastes. The study's results revealed the following things:

- *Cost:*
  As open-source tools, RapidMiner and Apache Mahout provide affordable alternatives. Commercial products like IBM SPSS Modeler and SAS Enterprise Miner may have licensing costs. Both open-source and paid versions of Apache Hadoop, Apache Spark, and Tableau are available, with prices changing according to usage and specifications.

- *Usability:*
  IBM SPSS Modeler, SAS Enterprise Miner, and RapidMiner are renowned for their simple workflows and user-friendly interfaces. For data visualization, Tableau also offers a user-friendly interface. For configuration and implementation, Apache Mahout, Apache Hadoop, and Apache Spark call for a higher level of programming knowledge and technical proficiency.

- *Accessibility:*
  Tools with established user groups and copious documentation, such as RapidMiner, SAS Enterprise Miner, IBM SPSS Modeler, and Tableau, are extensively used and easily accessible. Although extensively used, Apache Mahout, Apache Hadoop, and Apache Spark may need additional setup and configuration.

- *Capabilities:*
  Each tool offers a unique set of capabilities. RapidMiner, SAS Enterprise Miner, and IBM SPSS Modeler provide comprehensive data analysis, data mining, and predictive modeling features. Apache Mahout, Apache Hadoop, and Apache Spark are primarily focused on big data processing, machine learning, and distributed computing. Tableau specializes in data visualization and interactive dashboards.

- *Open Source:*
  Tools that are open-source, such as RapidMiner, Apache Mahout, Apache Hadoop, and Apache Spark, enable for customization and community-driven development. Users who want the independence and flexibility to change the tool's functionalities may find this helpful.

- *Each Tool Offers a Different Set of Skills.*
  Comprehensive data analysis, data mining, and predictive modeling capabilities are offered by RapidMiner, SAS Enterprise Miner, and IBM SPSS Modeler. Big data processing, machine learning, and distributed computing are the main focuses of Apache Mahout, Apache Hadoop, and Apache Spark. Tableau is an expert in creating interactive dashboards and data visualization.

- *Scalability:*
  With their capacity to scale and distributed computing features, Apache Hadoop and Apache Spark are built to handle large-scale data processing and analytics activities. Large datasets can also be handled by RapidMiner, SAS Enterprise Miner, IBM SPSS Modeler, Apache Mahout, and Tableau, albeit their scalability may vary depending on hardware and system setups.

- *Performance:*
  When working with massive data and distributed computing, Apache Hadoop and Apache Spark thrive. Additionally offering effective performance, RapidMiner, SAS Enterprise Miner, IBM SPSS Modeler, Apache Mahout, and Tableau may differ depending on the difficulty of the analysis and the amount of the dataset.

- *User-Friendly:*
  Tableau, IBM SPSS Modeler, SAS Enterprise Miner, and RapidMiner are renowned for their usability, offering user-friendly interfaces and drag-and-drop capabilities. For beginner users, Apache Mahout, Apache Hadoop, and Apache Spark are less intuitive due to their greater technical complexity and coding knowledge requirements.

- *Application in the Real World:*
  The ideal tool to utilize relies on the particular use case and standards set by the sector. Due to their adaptability, RapidMiner, SAS Enterprise Miner, IBM SPSS Modeler, and Tableau are widely used in a variety of industries. The most widely used Apache distributions in big data analytics and data-intensive businesses are Apache Mahout, Apache Hadoop, and Apache Spark. To choose the best predictive analytic tool for a certain organization's or person's needs, it is crucial to take into account the organization's or person's expertise, budget, and resources.

➤ *Result*
  According to the study, each of the technologies examined—RapidMiner, SAS Enterprise Miner, Apache Mahout, IBM SPSS Modeler, Apache Hadoop, Apache Spark, and Tableau—has distinct advantages and advantages in the area of predictive analytics. Due to its intuitive user interface and large library of pre-built algorithms and models, RapidMiner has distinguished itself as a standout product that makes it simple to use for beginners. The data preparation and model building capabilities of SAS Enterprise Miner were impressive, but the learning curve was longer. The distributed machine learning scalability of Apache Mahout was demonstrated, however it required more advanced technical knowledge. Despite requiring significant resources and technical know-how, Apache

Hadoop proved to be a reliable solution for storing and processing massive datasets. IBM SPSS Modeler excelled in data visualization and interpretation because to its user-friendly interface. Although it also required technical expertise for efficient use, Apache Spark demonstrated high-performance capabilities in processing huge datasets, especially in the area of machine learning. Tableau demonstrated its merits but could not have the same level of analytical capabilities as other tools while being known for its strong data visualization features and user-friendly interface. In the end, choosing the best predictive analytics tool depends on a person's needs, their technical proficiency, and a comprehensive assessment of each tool's capabilities, scalability, performance, usability, and practical application.

Due to its outstanding performance across a variety of characteristics when these aspects are taken into account, RapidMiner appears as an appealing option. It features an intuitive user interface, extensive data analysis and predictive modeling capabilities, and is accessible as an open-source application for cost-effectiveness and customization options. The best option for businesses looking for a powerful and adaptable predictive analytics solution is RapidMiner.

## V. CONCLUSION

Finally, the research on " Reviews of Big Data Techniques and Tools for Predictive Analytics Analysis" included a thorough assessment and comparison of a number of products in the area. In terms of predictive analytics, each tool—RapidMiner, SAS Enterprise Miner, Apache Mahout, IBM SPSS Modeler, Apache Hadoop, Apache Spark, and Tableau—displayed distinct advantages and capacities. Due to its huge library of pre-built algorithms and models and user-friendly interface, RapidMiner stands out as a particularly useful tool for beginners. The findings led to the recommendation that before choosing a Big Data solution for predictive analytics, businesses should carefully assess their unique needs and requirements. According to the study, the choice should be based on the tool's capabilities, scalability, performance, cost, usability, and suitability for application in actual situations. Organizations should also take into account the accessibility of qualified personnel who can use these tools. Additionally, it is advised that businesses spend money educating staff members to use these products efficiently. To make sure they are utilizing the greatest technologies for their needs in predictive analytics, firms should also keep up with the most recent breakthroughs and advancements in Big Data analytics.

## REFERENCES

[1]. S. both Gupta and S. Verma, IBM SPSS modeler, and SAS enterprise miner: A comparison. 2020, 975, 8887, International Journal of Computer Applications.

[2]. R. G., Ritu, and. Big Data Techniques and Tools: A Roadmap for Predictive Analytics, Preeti. Volume 9 Issue 2 of the International Journal of Engineering and Advanced Technology (IJEAT), ISSN 2249-8958.

[3]. R. Statistical inference for Big Data Analytics: A Review, by Gupta and V. Sharma. 8(1), 1-29, Journal of Big Data, 2021.

[4]. J. K. Anderson, Smith, and L. L. Johnson. emerging patterns in computer science applications for large data statistical analysis. 2019's Journal of Data Science, 13(3), pages 321-339.

[5]. A. B. Jones and J. Brown, Statistical techniques for massive data analysis in applications in computer science. 2020, 42–59, International Journal of Computer Science Research, 15(2).

[6]. Y. Zhang, Chen, and L. Wang, big data, and machine learning applications. 2017; 31(1):4–16 in Journal of Computer Science and Technology.

[7]. Y. L. Li, D. Liu, and I. Wang, A review of the literature on predicting consumer behavior in e-commerce. 2019; 36, 100874; Electronic Commerce Research and Applications.

[8]. D. S. Wu, Wu, and Y. Zhang, Opportunities, Problems, and Application Examples of Big Data Analytics in Finance. Research on Big Data, 22, 100162, 2020.

[9]. Y. Wang, Y., D., and Wang. Wang, A Systematic Review of Big Data Analytics for Predictive Maintenance in Manufacturing. IEEE Transactions on Industrial Informatics, Vol. 17, No. 8, 2021, pp. 5462–5475.

[10]. Y. D. Zhu, Y. Li, and Y. Zhang, A thorough assessment of the literature on big data analytics for fraud detection and prevention. 9(1), 1-30, Journal of Big Data, 2022.

[11]. S. O. Spatscheck, H. Kaisler, and P. Zang, a thorough overview of the literature on the use of big data analytics in smart cities. 2017; Smart City 360°, 1(1), 1-23.

[12]. U. S. P. Singh, Srivastava, and D. Big data analytics for social media: Definitions, approaches, difficulties, and prospects, K. Vishwakarma. Frontiers in Information Systems, 20(2), 259–309, 2018.

[13]. Z. Li, D, Q, and Xie. Wang, A survey of the literature on big data analytics for customer segmentation in marketing. 2019; 49(3):346-357 in the International Journal of Information Management.

[14]. H. D. Wang, Jiang, and H. Liang, A survey of the literature on big data analytics for stock price prediction. 50, 232-244, 2020, International Journal of Information Management.

[15]. D. Three Wangs: Y. Wang, Y. Wang, Big data analytics in healthcare: A thorough assessment of the literature and recommendations for future study. 145, 104391, International Journal of Medical Informatics, 2021.

[16]. Y. H. Sun, Z. Zhang, and H. Du, A systematic assessment of the literature on big data analytics for anticipating traffic congestion in smart cities. 59(1), 102622, 2022; Information Processing & Management.

Olayemi Ariyo received the BSc degree in Computer Science from Federal University of Agriculture, Abeokuta, Ogun State, Nigeria in 1997 and Msc degree in Computer Science from Caleb University in 2021. He is currently pursuing the Phd degree at the School of Computing and Engineering Sciences, Babcock University, Ilishan, Ogun state, Nigeria. His research interests include Artificial Intelligence, Big Data, Deep Learning and Machine Learning.

Maitanmi, Oluwasola Stephen is an Associate Professor of Computer Science (Software Engineering). He completed his PhD in Babcock University in 2015, Post-doctoral Research in quantitative and qualitative research, report and grant writing in Andrews University, Michigan, USA in 2018-2019. He is the Deputy Director of Administration, Babcock University Center for Open and Distance e-Learning (BUCODeL). The facilitator of Online education in Babcock University. His areas of interest are not limited to Medical Cyber Physical Systems, Qualitative and Quantitative Research, Open distance e-Learning and grant writing. He had published over 50 High Impact Articles, 2 books, and worn several grants. He is the Editor in Chief of Journal of Computing and Software Engineering (JCSE). He is a member of several review journals and professional bodies.