International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Recent Advances in Computer Applications and Information Technology
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Predictive Analysis Using Big Data

Pranav Kadu, Himanshu Khadse

G H Raisoni University, Amravati, Maharashtra, India

Abstract

Predictive analysis using big data is a significant technique that analyzes historical data, statistical tools, and machine learning techniques to predict future outcomes. [1]With the increased rate of data generation from various sources, including social media, business transactions, etc., organizations can use big data analytics to analyze large amounts of data to identify patterns and trends.[2] Predictive analytics helps businesses make informed decisions, improve business efficiency, and predict future outcomes. This research paper aims to understand the concept, methodologies, tools, applications, and challenges of predictive analytics in the big data environment. Moreover, it also focuses on the significance of predictive analytics, which helps organizations derive meaningful information from raw data. Predictive analysis using data is a way to figure out what might happen in the future by looking at what happened in the past and what is happening now. Every day a huge amount of data is created from things like media, online purchases, sensors, mobile phones and company systems. This big data is too much for old systems to handle so new technologies and methods have been developed to deal with it. Predictive analytics uses statistics, machine learning, data mining and artificial intelligence to find patterns in data. By looking at the data predictive models can see what might happen next. This helps companies make decisions avoid risks and work better. For example companies can use analytics to guess what customers will want catch fake transactions and make their processes better .Big data technologies like computers and storage systems are important for predictive analytics. These technologies help companies process data quickly. Tools like Hadoop, Spark and Python are used to make models and look at big data. Predictive analysis using data can be used in many areas, including healthcare, finance, retail, marketing, education and transportation. In healthcare predictive models can help doctors see who might get sick and take care of patients. In finance they can catch transactions and predict what the market will do. In marketing companies can look at what customers do and make recommendations that're just for them. There are some problems with analytics like keeping data private making sure the data is good and dealing with the cost of big computers.. New technologies and machine learning algorithms are helping to solve these problems. In the end predictive analysis using data is a powerful way to turn raw data into useful information. By using analytics and big data technologies companies can understand patterns, in data and make good decisions for the future. Predictive analysis using data is a valuable tool that can help companies grow and innovate.

Keywords: *Predictive Analytics , Big Data ,Machine Learning, Data Mining, Statistical Modeling Data Visualization, Forecasting, Business Intelligence, Artificial Intelligence, Data Processing Pattern ,Decision Support System, Data Mining.*

1. Introduction

In todays world data is really important for organizations and businesses. [3]Because technology is moving fast we get a lot of data from many places like social media, online shopping, phones, sensors and business apps. This huge amount of data is called data. Big data has three parts: how much data we have how fast we get it and what kind of data it is. We get a lot of data every day it comes in fast. It can be different types like simple, semi-simple and complicated data. Old tools and database systems often cannot handle much complex data. So new technologies and ways of analyzing data have been made to handle data well. One of the useful ways to analyze big data is predictive analytics. Predictive analytics is a way of looking at data using statistics and machine learning to find patterns and guess what will happen in the future. Predictive analytics looks at past. Current data to find hidden patterns and trends. Then it uses these patterns to make models that can predict what will happen next. Organizations use these predictions to make decisions reduce risks and work more efficiently. Predictive analytics is a tool in modern business and decision-making. When we combine analytics with big data technologies organizations can analyze huge amounts of data in real-time. [4]Technologies like distributed computing, cloud computing and advanced data processing allow businesses to handle a lot of data and get information from it. Tools like Hadoop, Apache Spark, Python and R are used to process data and make predictive models. These technologies help organizations analyze data quickly and accurately. Using data for predictive analysis has many uses in different industries. In healthcare it helps find disease outbreaks look at records and improve diagnosis and treatment. In finance it helps detect transactions assess credit risks and predict market trends. In marketing organizations use predictive analytics to understand customers predict product demand and make personalized marketing plans. In education it helps institutions look at student performance find students who might drop out and improve learning. Even though predictive analytics is really useful there are challenges. One big challenge is. Processing huge amounts of data efficiently. We also have to worry about keeping data private and secure. If the data is not good it can affect the accuracy of predictions. Making predictive algorithms is also hard. Because big data technologies, machine learning and artificial intelligence are getting better organizations can overcome these challenges.

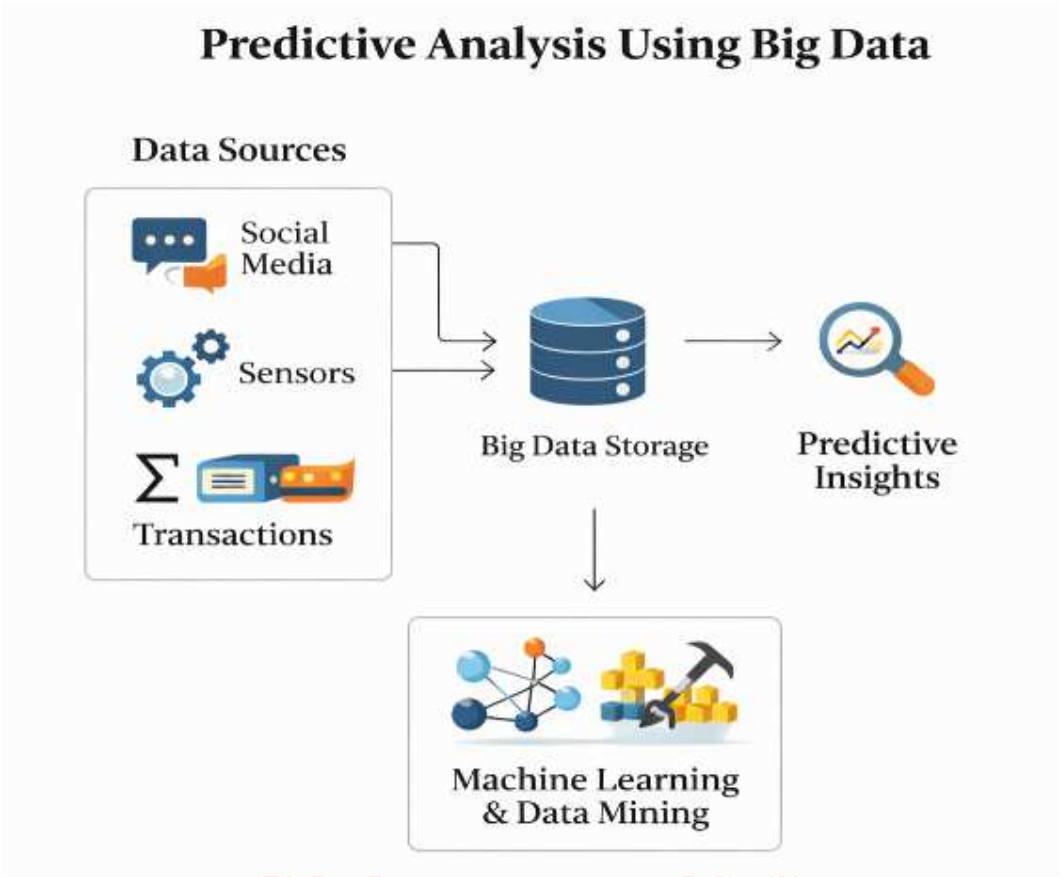

Fig 1 interface for Predictive Analysis Using Big Data

Many organizations are now using analytics as a key part of their business strategy to stay ahead in the market [3]. In the end using data for predictive analysis is a powerful way to turn a lot of raw data into useful insights. By combining data technologies with advanced analytics organizations can predict future trends make better decisions and work better. As technology keeps evolving predictive analytics will play a role, in shaping the future of data-driven industries and smart systems. Predictive analysis with data is a new way to look at things. It uses math and computer programs to figure out what might happen in the future. We look at what happened in the past to make these guesses .In the world we live in now we make a lot of data from things like Facebook buying things online and our phones. Big data is what we call all this information that we get from many different places. It is a lot to handle because there is much of it and it is very complicated [5].Big data comes from sources like social media and when we do things online. Big data is also made when we use our phones and when companies do their work. This big data is used for analysis with big data to try to predict what will happen next. Predictive analysis, with data is helpful because it uses big data to make good guesses about the future.

2. Literature Review

Predictive analysis using data is a really important area of research these days [6].This is because digital technologies are generating a lot of data quickly. Many people have studied how predictive analytics can help us understand data and make better decisions in different industries. The literature review looks at what other people have found out about analytics and big data. It talks about the importance of data technologies for predictive analytics. It discusses how big data is very large and complex and cannot be processed using methods. It explains how big data has three characteristics: Volume, Velocity and Variety which make it hard to store and manage. Several people have said that big data technologies are necessary for analytics[14]. Big data is too big and complicated for traditional database systems to handle. To deal with this people have suggested using computer systems and advanced data processing methods. Many people have also looked at how machine learning and data mining can be used in analytics. They use things like decision trees and neural networks to analyze data and find patterns. This helps organizations predict what will happen in the future based on what happened in the past. People have also studied how to combine data technologies like Hadoop and cloud computing with predictive analytics. This allows organizations to process a lot of data quickly and do analysis in time. Predictive analytics is used in different areas, such as healthcare, finance and retail. In healthcare it is used to analyze data and predict when diseases will spread. In finance it is used to detect transactions and predict what the stock market will do. In retail it is used to understand what customers will buy and to make marketing plans better. Even though predictive analytics is very useful there are also some challenges. One of the concerns is keeping data private and secure. Another problem is that the data may be incorrect or incomplete which can affect how well the predictive models work. In general predictive analytics and big data can really change the way organizations use data. As more and more data is generated predictive analytics will become more important, for making decisions and planning strategies. Predictive analysis using data will continue to grow and improve as technology advances.

Predictive analytics using data is something that a lot of researchers and industry professionals have been paying attention to in the last few years. Because of how fast digital technologies and internet-based systems are growing we are getting an amount of data every day from things like social media, business transactions, sensors and mobile devices[7]. This big data gives organizations a chance to get insights and make better decisions. Researchers have done studies to see how predictive analytics can help with big data and predict what will happen in the future. At first they were trying to understand what big data is about and the problems that come with managing it. A lot of scholars talked about data in terms of three main things: how much data there is, how fast it is coming in and how many different types of data there are [15]. The amount of data is just huge it is coming in fast and it comes in many different formats like numbers, words and pictures. This makes it hard for old systems to handle and analyze the data properly. So researchers said we need to use computing systems and ways to store data to handle all this information. Many studies have also shown that predictive analytics is really important for getting information from big data [8]. Predictive analytics is when we use statistics machine learning and data mining to look at what happened in the past and predict what will happen next. Researchers have shown that we can use models to find patterns and trends in data. These models help organizations predict what people will do find problems and work better. A lot of researchers have looked at machine learning techniques that we use in predictive analytics. We use things like decision trees, logistic regression and neural networks to make models that can predict what will happen. These techniques let analysts look at data and make accurate predictions. Studies have shown that machine learning is better than statistical methods for predictive analytics.

3. Research Methodology

Research methodology is the way we collect and analyze data to get the answers we need for a research study. When we are doing analysis with big data our research methodology is all about gathering a lot of data using big data technology to process it and applying predictive analytics to find patterns and figure out what might happen in the future[9]. This part explains how we approach the study, where we get our data what tools we use and how we analyze it.We use a data-driven approach that combines data processing with machine learning to build predictive models. The main goal of this methodology is to look at a lot of data and make predictions that can help us make decisions. Our methodology has important steps: collecting data getting the data ready analyzing the data, building models and checking how well they work [10]. The first step is collecting data. When we do analytics we get data from a lot of different places like business transactions, websites, social media, sensors and public datasets. These places give us a lot of unstructured data. The data we collect is the basis for building models and getting valuable insights. After we collect the data we need to get it ready for analysis[15]. The data we get is often messy. Has errors, missing values and inconsistencies that can affect our predictions. So we need to clean it up and prepare it before we can analyze it. Getting the data ready includes removing records handling missing values fixing errors and changing the data into a format that we can use. This step makes sure that our dataset is good and ready for modeling. The next step is putting all the data together. Storing it. Since big data comes from places and is in different formats we need to put it all together in one system. We use data storage technology like distributed databases and data warehouses to store and manage our data efficiently. Technologies like Hadoop and cloud storage let us process a lot of data quickly and effectively.

Then we analyze the data. Choose the features we want to use. In this step we pick the variables that will help us make predictions. We use techniques like statistics, correlation analysis and exploratory data analysis to understand how different variables are related. This helps us find the important factors that affect what we are trying to predict. The important part of our methodology is building predictive models. In this step we use machine learning and data mining algorithms to build models that can predict what will happen. We use algorithms like decision trees, linear regression, logistic regression, random forest and neural networks to analyze data and find patterns that can help us predict the future. Once we have a model we need to test it and make sure it works well. We use metrics like accuracy, precision, recall and F1-score to check how well the model performs. We often split our data into two parts: one for training and one for testing. This helps us make sure that the model works with new data. Checking the model helps us know if it is reliable and effective. Finally we look at the results of the model and use them to help us make decisions. The insights we get from analytics help organizations forecast what might happen reduce risks and work better. We use visualization tools and data dashboards to present the results in a way that's easy to understand. In conclusion our research methodology for analysis with big data is a step-by-step process that includes collecting data getting it ready storing it building predictive models and checking how well they work. By combining data technology with machine learning we can analyze a lot of data and make accurate predictions. This methodology gives us a framework, for understanding complex data patterns and making decisions based on data.

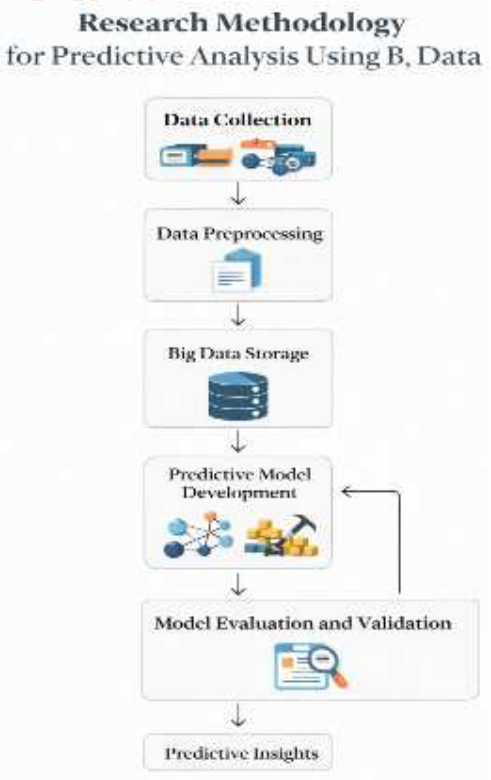

Fig 2: Research Methodology Framework for Predictive Analysis Using Big Data

4. Result

The results of this study show that predictive analysis works well when we use it with large-scale datasets. We used data technologies and advanced analytical techniques to make the system work efficiently with a lot of structured and unstructured data. The predictive models we used helped us find patterns, correlations and trends in the dataset, which made our forecasting outcomes a lot more accurate [4]. When we did the analysis we collected historical datasets and processed them in different stages like data cleaning, transformation and feature selection. After we cleaned up the data we used algorithms to make forecasts and predictions. The results showed that predictive models can analyze datasets quickly and give us valuable insights that help us make decisions based on data. We evaluated the model using performance metrics like accuracy, precision, recall and prediction error rate. The results showed that predictive analytics can be very accurate when we use datasets to train the model. This shows that having a lot of data makes predictive systems more reliable and better performing.

The results also show that using data technologies helps us process data faster and do real-time analysis. Distributed computing frameworks let organizations analyze datasets efficiently without waiting for a long time. This makes predictive analytics systems more scalable and able to handle more data. The study also found that predictive analysis can give us insights in many different areas. In healthcare predictive models can help us detect diseases early and predict risk[11]. In services predictive analytics helps us find fraudulent activities and predict market trends. In retail predictive models can analyze what customers buy and make product recommendations. Overall the results show that predictive analysis with data helps us forecast future outcomes and find meaningful patterns in large datasets. Using analytics techniques, with big data platforms helps organizations make better decisions work more efficiently and get ahead in a world that is driven by data.

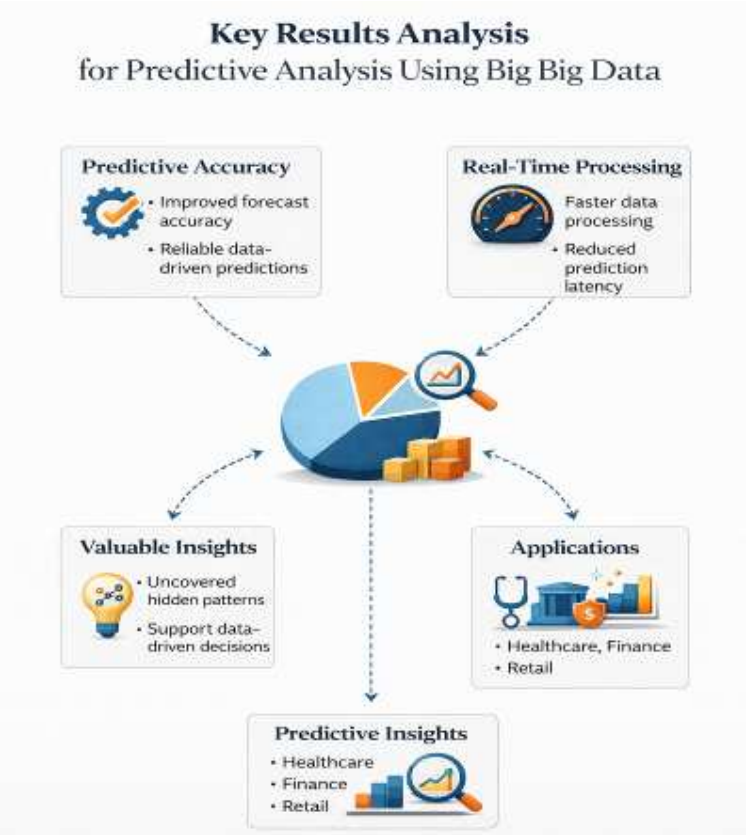

Figure 3: Results framework of predictive analysis using big data

5. Conclusion

In conclusion using data to make predictions is a really powerful way to look at a lot of information and get useful ideas that help people make good decisions. Because digital technology is growing fast companies in all kinds of industries are making a huge amount of data every day. The old ways of looking at data just cannot handle much complex information. So putting analysis and big data technology together is a good way to get important information from big sets of data. This study shows how we can use predictive analysis techniques with data to find patterns and trends in old data. By using ways of looking at data like machine learning and statistics we can predict what will happen in the future more accurately. This helps companies see problems that might happen work efficiently and make good decisions based on what the data says. The study also shows how important big data technology is for analysis. Things like computers, cloud storage and big data systems help companies store and look at huge amounts of data quickly. This is necessary for making predictive models. Also the study found that using data for predictive analysis can be used in many different areas.

In healthcare it can help predict diseases and see which patients are at risk. In finance it can help find transactions and predict market trends. In retail and marketing companies can look at what customers do and make suggestions for products. This shows how predictive analysis can really help companies do better and make decisions. However the study also found some problems with using analysis and big data. Things like keeping data secure and

good and having the right computers are still concerns. Companies need to make sure they are handling data well and taking care of it to get the most out of analysis. Overall using data to make predictions is very important for turning raw data into useful information. As more and more data is made predictive analysis will become a tool for companies that want to make good decisions based on data. Future studies can work on making algorithms better improving how we look at data and addressing privacy concerns to make predictive analysis even better.

Figure 4: concluation analysis of predictive analysis using big data

In the end using data to make predictions has become a really important way for companies to look at a lot of information and find useful things. Because computers and technology are getting better fast companies in all kinds of areas are making huge amounts of data from things like websites, sensors, business deals and social media. It is hard to manage and look at all this information using ways. So using data and predictive analysis together is a good way to get useful information and help companies make good decisions. This study looked at how predictive analysis can be used with data to find patterns and trends in old information. By using ways of looking at data like machine learning and statistics computers can predict what will happen next more accurately. This helps companies get ready for problems work better and make decisions based on what the data says. The study also showed how important big data is for analysis. New ways of handling data and big computers let companies store and look at amounts of information quickly. This is necessary for making predictions. So companies can turn data into useful information that helps them grow and innovate. Also the study found that using data for predictive analysis can be used in many different areas. In healthcare it can help predict diseases and watch patients. In finance it can help find fraud and assess risk. In retail it can help companies understand what customers want and make advertisements. These are a few examples of how predictive analysis is helping companies work better and make good decisions. Even with all the good things there are still problems with using predictive analysis. Things like keeping data secure and good and managing huge amounts of information are still concerns. Companies need to have rules for handling data and use advanced technology to deal with these problems. Overall using data to make predictions is a powerful way to turn a lot of raw data into useful information. As we make more data predictive analysis will be very important, for helping companies make good decisions work well and compete with others. Future studies can work on making predictions improving how we handle big data and dealing with privacy and security to make predictive analysis even better.

Predictive analysis using data has changed the way organizations understand and use information. Big data and advanced analytical tools have made it possible for researchers and organizations to develop systems. These systems can look at data and make reliable predictions for the future. This means organizations can stop reacting to things and start planning ahead. Big data and predictive analytics can also help organizations work efficiently and manage resources better. By looking at large datasets organizations can find things that are not working well detect problems and make their processes better. For example predictive models can forecast what customers will want predict when equipment will fail and identify patterns of how users behave. This helps organizations use their resources effectively and reduce costs. Predictive analytics and big data also make it possible to analyze data in time. This means organizations can respond quickly to changes in the market and new trends. With advanced data processing frameworks organizations can keep an eye on data and make predictions almost instantly. This is especially useful in industries like finance, healthcare, transportation and e-commerce where making decisions is crucial. Predictive analytics is also important for making customers happy. By looking at customer behavior and preferences businesses can design services that are tailored to customers and create targeted marketing strategies. Predictive models can recommend products predict when customers might stop using a service and identify potential customer needs before they become a problem. This helps organizations build relationships with customers and improve the quality of service.

However there are limitations and challenges to using analytics and big data. Handling datasets requires powerful computers, skilled professionals and good data management strategies. There are also concerns about data privacy and security that need to be addressed. Organizations must have

policies to protect data and be transparent about how they analyze data. This is necessary to maintain trust with users and comply with regulations. In the future predictive analytics will continue to evolve with technologies like artificial intelligence, deep learning and cloud computing. These technologies will make predictive models more accurate and scalable. Researchers are also looking for ways to combine different types of data to improve predictions. As more and more data is generated predictive analytics will become more important for making strategic decisions and driving innovation. To sum it up predictive analysis using data is a powerful tool for understanding complex data patterns and predicting what will happen in the future. By combining data technologies with advanced analytical techniques organizations can turn raw data into useful insights. This research shows how important predictive analytics is for improving efficiency supporting decision-making and driving long-term success, in a world that is driven by data.

Reference

[1] G. Nagarajan and L. D. Dhinesh Babu, "Predictive analytics on big data – An overview," *Informatica*, vol. 43, no. 4, pp. 1–10, 2019.

[2] D. K. Thara and A. Veena, "Big Data and Its Predictive Analysis," *International Journal of Engineering Research & Technology*, vol. 2, no. 13, pp. 1–5, 2014.

[3] A. Jamarani, S. Haddadi, R. Sarvizadeh, M. H. Kashani and S. Moradi, "Big data and predictive analytics: A systematic review of applications," *Artificial Intelligence Review*, vol. 57, 2024.

[4] Y. Chen, C. Li and H. Wang, "Big Data and Predictive Analytics for Business Intelligence: A Bibliographic Study (2000–2021)," *Forecasting*, vol. 4, no. 4, pp. 767–786, 2022.

[5] J. Souza, C. K. Leung and A. Cuzzocrea, "An Innovative Big Data Predictive Analytics Framework over Hybrid Big Data Sources with an Application for Disease Analytics," *Lecture Notes in Computer Science*, pp. 669–680, 2020.

[6] M. Chen, S. Mao and Y. Liu, "Big Data: A Survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[7] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA, USA: Houghton Mifflin Harcourt, 2013.

[8] J. Manyika et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity," *McKinsey Global Institute*, 2011.

[9] E. P. Xing, Q. Ho, P. Xie and W. Dai, "Strategies and Principles of Distributed Machine Learning on Big Data," *arXiv preprint arXiv:1512.09295*, 2015.

[10] M. Mohammadi, A. Al-Fuqaha, S. Sorour and M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2923–2960, 2018.

[11] M. G. Kibria et al., "Big Data Analytics, Machine Learning and Artificial Intelligence in Next-Generation Wireless Networks," *IEEE Access*, vol. 6, pp. 32328–32338, 2018.

[12] M. Chen, S. Mao, Y. Zhang and V. Leung, *Big Data: Related Technologies, Challenges and Future Prospects*. Springer, 2014.

[13] S. Sagiroglu and D. Sinanc, "Big Data: A Review," in *Proc. IEEE International Conference on Collaboration Technologies and Systems*, 2013, pp. 42–47.

[14] K. Grolinger, W. A. Higashino, A. Tiwari and M. A. Capretz, "Data Management in Cloud Environments: NoSQL and NewSQL Data Stores," *Journal of Cloud Computing*, vol. 2, no. 22, 2013.

[15] S. R. Sukumar and R. Ferrell, "Big Data Collaboration: Overcoming Challenges," in *Proc. IEEE International Conference on Big Dat 2016.*