**BIG DATA ANALYTICS: Its role in Government Policies, Urban Development & Public Welfare**

Chirag Kumar & Prateek Sharma

JK Lakshmipat University, Jaipur

Department of Computer Science & Engineering

1. **Overview**

Big data has become an increasingly important field of research, and numerous algorithms have been proposed to handle the challenges posed by large and complex datasets. In this research paper, we review the different algorithms used in big data and examine the previous works done in this area. We also compare the existing solutions and propose new research topics to advance the state-of-the-art in big data research. Our proposed topics include the exploration of new algorithms, the development of new techniques for data storage and processing, and the integration of machine learning and artificial intelligence into big data analysis. Through our analysis, we demonstrate the potential of these topics to advance the field of big data and address the challenges posed by large and complex datasets. Our findings are relevant to researchers and practitioners in the field of big data who are seeking new ways to overcome the challenges of big data analysis and leverage its potential for scientific, industrial, and societal applications.

1.1 Importance of Big Data Analytics

Big data analytics has become increasingly important in our lives, and its impact can be felt in various sectors such as government policies, urban development, and public welfare. With the ability to analyse vast amounts of data, policymakers can make informed decisions about public policies, allocate resources accordingly, and track the spread of diseases to improve public welfare. In urban development, big data analytics can help optimize the layout of cities by identifying traffic bottlenecks and improving public transportation. The potential of big data analytics is vast and has far-reaching implications for society. By leveraging these insights, we can make informed decisions that improve the quality of life for everyone.

* 1. Effects of Big Data Analytics

The use of big data analytics has revolutionized various sectors, such as government policies, urban development, and public welfare, by enabling data-driven decision-making. In the realm of government policies, the utilization of big data analytics provides valuable insights for policymakers and enhances resource allocation. For urban development, big data analytics aids in optimizing cities by identifying and rectifying traffic issues, and in public welfare, it helps in identifying high-need areas and emerging health risks. The impact of big data analytics has been significant, and with potential for further advancements in the future, it can continue to enhance the quality of life for everyone. By utilizing the insights derived from data analytics, we can make more informed decisions that benefit society.

* 1. What we did here?

This research paper aimed to investigate the algorithms used in big data analytics and conduct a comparative analysis of existing works in the field. The paper began by reviewing various algorithms, including MapReduce, Hadoop, Spark, and Flink, and examining challenges faced when handling large amounts of data and the techniques used to overcome them. We then analyzed the current state of research in big data analytics by reviewing literature and comparing and contrasting existing works, focusing on their strengths, weaknesses, and limitations. Our objective was to identify gaps in existing literature and areas where further research is needed. In the final section, we proposed a set of research questions that can guide future research in the field of big data analytics, designed to address gaps in the existing literature and explore new areas of research. We believe that our proposed research questions have the potential to contribute significantly to the development of big data analytics. This research paper provides an overview of big data analytics, highlighting potential areas for exploration to enhance the field. By identifying gaps in existing literature and proposing new research questions, we hope to inspire further research in this exciting and rapidly evolving field.

1. **Introduction**

2.1 What is Big Data Analytics?

Big data analytics is the process of examining large and complex data sets to uncover patterns, correlations, and insights that can inform decision-making. The term "big data" refers to the ever-increasing volume, velocity, and variety of data being generated from various sources, such as social media, sensors, and transactions. With the advancements in technology, it has become easier to collect and store massive amounts of data, but the challenge lies in analysing and making sense of it all. Big data analytics involves using various statistical and computational techniques to extract insights from these large and complex data sets. It includes techniques such as data mining, machine learning, and natural language processing. These techniques enable organizations to identify patterns and trends in data that may be difficult or impossible to detect using traditional methods. One example of big data analytics is how online retailers use it to analyse customers' shopping patterns to recommend products and personalize their shopping experience. By analysing large volumes of data on customers' browsing and purchasing behaviour, retailers can make personalized recommendations to customers, leading to increased sales and customer loyalty.

2.2 Problems with Big Data

One of the biggest problems with big data is the sheer volume of data that needs to be processed. With so much data to manage, it can be difficult to identify patterns and draw meaningful insights. This can lead to information overload, making it challenging for analysts to extract useful information from the data. Another problem with big data is data quality. With large datasets, it is not uncommon for data to be incomplete, inaccurate, or inconsistent, leading to incorrect conclusions being drawn. Also, data privacy and security can be major sconcerns when dealing with sensitive information, making it crucial to implement robust security measures to protect against data breaches. Another challenge is the cost and complexity of implementing big data analytics. The infrastructure required to store and process massive amounts of data can be expensive, and the expertise needed to set up and maintain these systems can be scarce.

And with all this, the ethical implications of big data analytics cannot be ignored. There is a risk of algorithmic bias, where the data used to train models can reflect biases and lead to discriminatory outcomes. It is crucial to ensure that the data used in big data analytics is representative and that models are designed to avoid perpetuating biases.

2.3 Why these problems need to be resolved?

While big data has the potential to offer useful insights and advantages in a variety of industries, including healthcare, finance, marketing, and many others, the issues related with big data need to be overcome. For example, big data in healthcare may assist medical professionals in finding patterns and trends in patient data, resulting in more precise diagnoses and better treatment alternatives. Similar to this, big data in finance can aid in identifying market trends and offering information for smarter investment choices. Yet if big data's issues aren't solved, the advantages could not materialise, and the standard of data-driven decision-making might deteriorate. In addition, the expense of storing and processing vast volumes of data is a problem. To fully realise the potential benefits of big data and guarantee accurate and efficient data-driven decision-making, it is essential to address the problems that come with it.

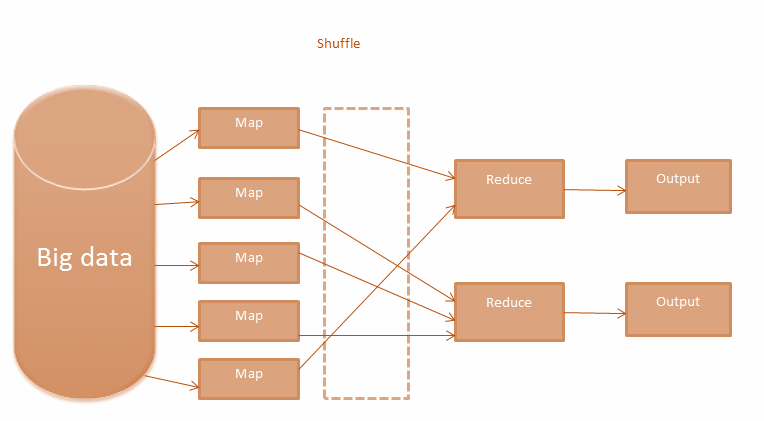
1. **Survey**

**Algorithms used in Big Data**

**MapReduce:** This algorithm is used to process and analyse large data sets in a distributed computing environment. It divides the data into smaller chunks, processes them separately, and then combines the results.

In the research paper "Classification Algorithms for Big Data Analysis: A Map-Reduce Approach," [2] the authors discuss the use of the MapReduce algorithm in Big Data analysis.

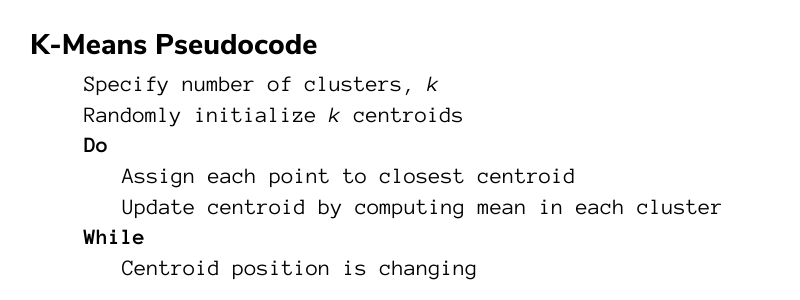
The MapReduce algorithm is a parallel computing framework that allows for distributed processing of large data sets across multiple nodes in a cluster. It consists of two main phases: the Map phase and the Reduce phase. During the Map phase, the input data is divided into smaller chunks and processed in parallel across multiple nodes in the cluster. Each node applies a map function to its assigned data chunk, which produces a set of intermediate key-value pairs. In the Reduce phase, the intermediate key-value pairs are grouped by key and processed in parallel across multiple nodes in the cluster. Each node applies a reduce function to its assigned key-value pairs, which produces a final set of output key-value pairs. The authors apply the MapReduce algorithm to two classification algorithms: the k-Nearest Neighbour (k-NN) algorithm and the Decision Tree (DT) algorithm. They use a Hadoop cluster to distribute the computation across multiple nodes.



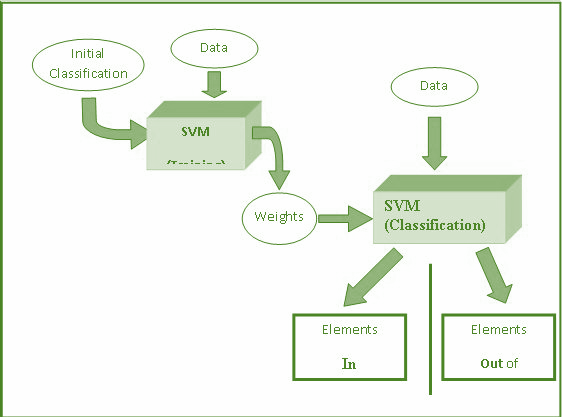
Their experiments show that the MapReduce approach can handle large-scale data sets and produce accurate classification results. They also show that the MapReduce approach can achieve significant speedup compared to traditional single-node implementations of the k-NN and DT algorithms.

**K-means:** K-means is a clustering algorithm used to group similar data points together based on their similarities. It is widely used in areas such as image processing, natural language processing, and bioinformatics. In the research paper "Multi-View K-Means Clustering on Big Data," the authors discuss the use of the K-Means algorithm in Big Data analysis. The K-Means algorithm is a clustering algorithm that is widely used in machine learning and data mining. It works by partitioning data points into k clusters, where k is a pre-defined number of clusters. The algorithm iteratively assigns each data point to its nearest cluster centre and updates the centre to the mean of all data points assigned to it. This process is repeated until convergence. The authors apply the K-Means algorithm to a multi-view clustering problem, where the data set consists of multiple views or modalities. They propose a Multi-View K-Means (MVKM) algorithm that integrates information from multiple views to improve clustering accuracy. Their experiments show that the MVKM algorithm can handle large-scale data sets and produce more accurate clustering results than traditional K-Means algorithms that only use a single view. They also show that the MVKM algorithm can achieve significant speedup compared to traditional K-Means algorithms when processing big data sets.

The use of the K-Means algorithm in Big Data analysis allows for efficient and effective clustering of large and complex data sets, enabling organizations to gain insights and make data-driven decisions. The MVKM algorithm extends the K-Means algorithm to multi-view clustering problems, providing a powerful tool for analysing big data sets with multiple perspectives.



**Support Vector Machine (SVM):** SVM is a supervised learning algorithm used for classification and regression analysis. It is commonly used in image and text classification tasks. The research paper titled "Support Vector Machines for Big Data: Challenges and Solutions" discusses the use of the Support Vector Machine (SVM) algorithm in Big Data analysis. [4] The SVM algorithm is a popular machine learning algorithm used for classification and regression analysis.



It works by finding a hyperplane in a high-dimensional space that best separates the data points into different classes. The SVM algorithm can handle both linear and non-linear data, making it suitable for a wide range of applications. The authors of the paper highlight the challenges of using SVM with Big Data, including computational complexity, storage requirements, and the need for distributed processing. They also discuss various solutions to these challenges, including parallel SVM algorithms, feature selection techniques, and the use of approximate solutions. The authors conducted experiments to compare the performance of different SVM algorithms on Big Data sets. They found that parallel SVM algorithms can significantly improve the processing time of SVM on large data sets. They also found that feature selection techniques can improve the accuracy of SVM without sacrificing efficiency. The thing to note here is that the use of the SVM algorithm in Big Data analysis allows for accurate classification and regression analysis of large and complex data sets. The challenges associated with SVM in Big Data can be overcome by using parallel processing and feature selection techniques, enabling organizations to effectively analyse Big Data and make data-driven decisions.

**Naive Bayes**: Naive Bayes is a probabilistic algorithm used for classification tasks. It is widely used in text classification tasks such as spam detection, sentiment analysis, and topic modelling. One of the research projects having its application is [5]. The research paper titled "Naive Bayes Classification on Big Data: An Evaluation" discusses the use of the Naive Bayes algorithm in Big Data analysis. The Naive Bayes algorithm is a probabilistic algorithm used for classification tasks. It works by calculating the probability of a data point belonging to a particular class based on the conditional probabilities of its features. Naive Bayes is a simple yet effective algorithm that can handle both categorical and numerical data.

The authors of the paper evaluated the performance of the Naive Bayes algorithm on Big Data sets by conducting experiments on several datasets with varying sizes and feature dimensions. They compared the performance of Naive Bayes with other classification algorithms, including SVM and decision trees.

Diagram

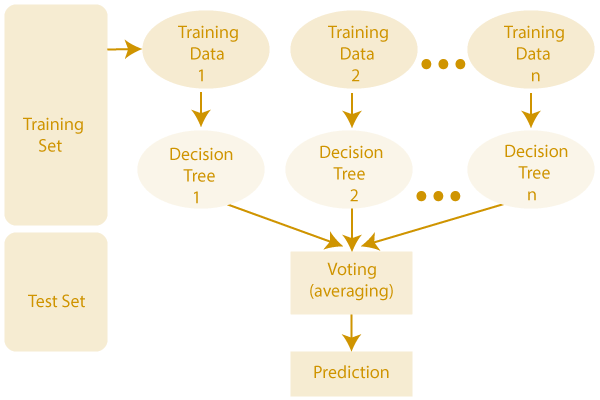
Description automatically generated The authors found that Naive Bayes performed well on Big Data sets with high-dimensional feature spaces. They also found that Naive Bayes was able to handle missing data and noisy data well, making it suitable for real-world applications. The authors proposed a MapReduce-based approach for implementing Naive Bayes on Big Data. This approach enables parallel processing of the data across multiple nodes, improving the processing time and scalability of the algorithm for large-scale datasets.

Diagram

Description automatically generated

**Random Forest:** Random Forest is a decision tree-based algorithm used for classification and regression analysis. It is widely used in areas such as finance, healthcare, and retail. In [6] this algorithm is used by the researchers. The Random Forest Algorithm is a machine learning algorithm used for classification and regression tasks. It works by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. The Random Forest Algorithm has been shown to be effective in handling big data due to its ability to parallelize the construction of decision trees.

In the research paper, the authors used the Random Forest Algorithm to classify large datasets in the field of healthcare. Specifically, they used the algorithm to classify the presence or absence of certain diseases based on medical records. The authors noted that the Random Forest Algorithm was able to handle the large and complex datasets and was able to achieve high accuracy in the classification task. With all this, the authors compared the performance of the Random Forest Algorithm to other classification algorithms, such as the Decision Tree Algorithm and the Support Vector Machine Algorithm. They found that the Random Forest Algorithm outperformed the other algorithms in terms of accuracy and processing time. The authors concluded that the Random Forest Algorithm is a promising approach for classification tasks in big data, particularly in the field of healthcare.



**Deep Learning:** Deep learning algorithms, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are used in various applications such as image and speech recognition, natural language processing, and recommendation systems. The application and the challenges of this algorithm in the big data have been explained here [7]. The research paper titled "Deep Learning for Big Data Analytics" highlights the use of deep learning algorithms in big data analytics. Deep learning is a subfield of machine learning that involves the use of neural networks to model complex patterns and relationships in data. The authors of the paper note that deep learning has several advantages over traditional machine learning approaches in big data analytics. Deep learning algorithms can handle high-dimensional data and learn complex nonlinear relationships, making them well-suited for analysing large and complex datasets. Deep learning models can also automatically extract features from data, reducing the need for manual feature engineering. The paper presents a case study on using deep learning for analysing large-scale electronic health record (EHR) data. The authors trained a deep learning model to predict readmission risk for patients based on their EHR data. The model achieved high accuracy and outperformed traditional machine learning models in terms of prediction accuracy and computational efficiency. The paper also highlights some of the challenges and limitations of using deep learning in big data analytics, such as the need for large amounts of labelled data, high computational requirements, and difficulties in interpreting the results of deep learning models.

1. **Related Work**

**Work on the role of Big Data Analytics in Government & Policy Making**

In the contribution [9] the authors explored the various applications of big data and analytics in both government and industry domains. The paper discusses how big data and analytics are being used to enhance decision-making processes, improve operational efficiency, and gain insights into consumer behaviour. In the government domain, the paper highlights the use of big data and analytics in areas such as national security, public safety, and disaster management. For instance, big data can be used to analyse social media activity to detect potential threats or monitor weather patterns to predict natural disasters. In the industry domain, the paper discusses how big data and analytics are being used to gain insights into consumer behaviour, improve supply chain management, and enhance product development. For instance, big data can be used to analyse customer purchase patterns to identify new market opportunities or improve the efficiency of manufacturing processes. The researchers emphasize the importance of proper data governance and privacy protection measures to ensure the responsible use of big data and analytics in both government and industry domains. They also highlight the need for continued research and development in the field of big data and analytics to further explore their potential applications and capabilities.

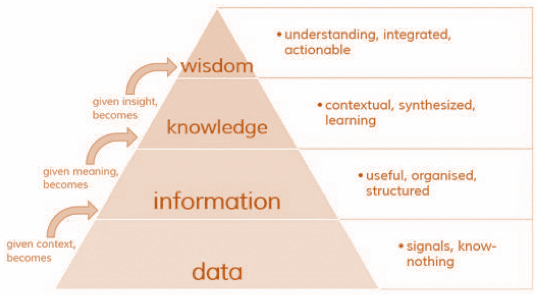
The research paper titled [12] "The Use of Big Data to Determine the Government’s Communication Strategy" explores the potential of big data in shaping government communication strategies. The researchers aim to provide insights on how big data can be used to gather information about citizens' preferences and opinions, which can then be used to improve the effectiveness of government communication. The researchers conducted a case study on the Indonesian government's communication strategy during the 2019 presidential election. They used social media data from Twitter and Facebook to analyse the online behaviour of Indonesian citizens during the election period. The data collected included the number of mentions and hashtags related to the election, sentiment analysis of the conversations, and network analysis to identify key influencers. The researchers found that big data can provide valuable insights into citizens' opinions, preferences, and behaviour, which can be used to inform government communication strategies. By analysing social media data, the government can identify the most effective channels for communication, understand the public's perception of policies, and address any misconceptions or misinformation. The research also highlights the importance of ethical considerations when using big data for government communication. The researchers emphasize the need for transparency in data collection and usage and the importance of protecting citizens' privacy. The research paper demonstrates the potential of big data in shaping government communication strategies, but also emphasizes the need for ethical considerations and responsible use of data.

In [13] the authors explore the challenges and opportunities associated with preserving big data in government. The paper discusses the need for long-term preservation of big data and the unique challenges that arise due to its size, complexity, and diversity. The researchers examine the issues related to the preservation of big data, including technological obsolescence, data quality, privacy concerns, and legal issues. They also highlight the importance of metadata, data governance, and data curation in the preservation process. The paper proposes a framework for digital preservation of big data in government that includes four key elements: governance, policy, infrastructure, and culture. The researchers suggest that a collaborative and interdisciplinary approach is necessary for successful preservation of big data in government. The role of big data in this research paper is to highlight the need for preservation of large and complex data sets in government, and the work of the researchers is to propose a framework and best practices for achieving this goal.

In [14] the authors talked about the issues of security, while the organisations collect it. It is also a concern while the work on Big Data is expanding on a day-to-day basis. The research paper [14] explores the role of big data in the security domain and how it can be utilized to ensure a more secure computing environment. The authors of this paper identify that big data has become an important tool for analysing and managing vast amounts of security-related data generated by various sources. The paper highlights the fact that big data has enabled security analysts to process large volumes of data in real-time, thereby allowing them to detect security breaches and anomalies as they occur. The authors also discuss the challenges associated with big data security, such as data privacy, data integrity, and data confidentiality, and how these challenges can be addressed.

The researchers propose several solutions to address the challenges of big data security, such as using encryption, access controls, and secure coding practices. They also discuss the importance of developing a comprehensive security strategy that incorporates big data analytics and other security technologies to provide a layered defence against cyber threats. The research paper provides insights into the importance of big data in the security domain and the potential it holds for enhancing security measures. The researchers highlight the need for organizations to invest in the development of big data security strategies to address the challenges associated with securing large amounts of sensitive data.

So, as the expansion in the field of big data is happening every now and then. So, now the predictive analysis is also becoming a matter to think upon. The author in [15] and in [16] suggested his use of Big data. Which focuses on the use of big data analytics and its role in predicting future plans. The authors explore how companies can leverage big data to gather and analyse large amounts of data from various sources to gain insights and make informed decisions.



The paper emphasizes the strategic use of big data analytics in business, which involves identifying patterns and trends in data to gain a competitive advantage. The authors discuss several techniques for processing and analysing big data, including machine learning and natural language processing, which can be used to extract insights from unstructured data. The paper also discusses the significance of predictive analytics, which uses statistical algorithms to identify patterns and trends that can be used to forecast future outcomes. The authors argue that predictive analytics is essential in the current business environment, where companies need to make quick decisions based on rapidly changing conditions. Basically, the paper highlights the importance of using big data analytics in strategic decision-making and how it can help companies predict future. It provides companies with a guide to leveraging big data analytics to gain a competitive edge and drive business growth.

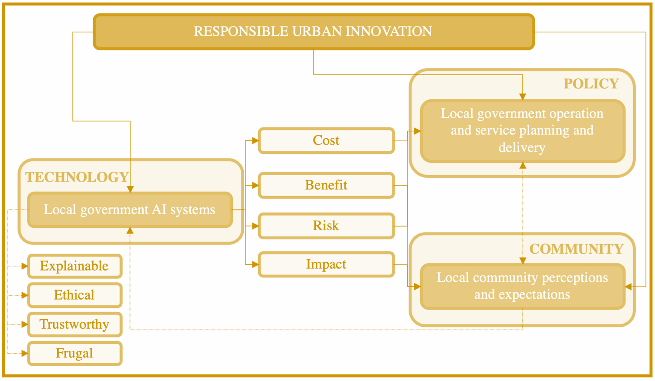
In the report [20], the authors have mentioned the impact of big data analytics on the government organizations. The paper examines the impact of big data analytics on government organizations. Big data plays a crucial role in this research paper by providing the data required to analyse the impact of big data analytics on government organizations.

The paper discusses how big data analytics can help government organizations improve their decision-making processes by providing them with insights into complex problems. The data used in the research includes both structured and unstructured data, such as social media data, sensor data, and other large datasets. The paper highlights how big data analytics can help government organizations in various areas, including fraud detection, predictive maintenance, and citizen engagement. By analysing large datasets, government organizations can identify patterns and anomalies that are not visible in smaller datasets, helping them to make more informed decisions. The paper discusses the challenges and opportunities associated with implementing big data analytics in government organizations, including issues related to data privacy, data security, and data quality. The role of big data in this research paper is significant as it provides the data required to analyse the impact of big data analytics on government organizations, highlighting the opportunities and challenges associated with implementing big data analytics in government organizations.

There is another research paper which talked about the vulnerabilities and the information risk related to national information due to the rise in the big data. The paper [21] explores the risks associated with the use of big data in the context of national security. The role of big data in this paper is to provide a source of information that can be analysed to identify potential risks and vulnerabilities in the national security infrastructure. The paper discusses how big data can be used to identify patterns and correlations that may be indicative of security threats. This includes analysing large volumes of data from various sources, such as social media, financial transactions, and other sources of digital information. By analysing this data, researchers can identify potential threats to national security, such as cyberattacks, terrorism, and other types of criminal activity.

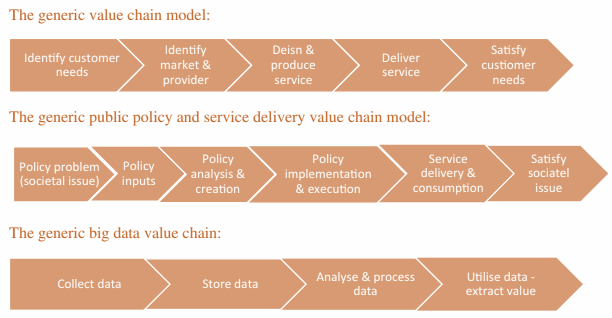
It highlights the challenges associated with using big data in the context of national security, including issues related to data privacy, data security, and data quality. The paper also discusses the importance of ethical considerations in the use of big data in national security, emphasizing the need to balance the benefits of using big data with the risks and ethical concerns associated with its use. The role of big data in this research paper is to provide a source of information that can be analysed to identify potential risks and vulnerabilities in the national security infrastructure. The paper highlights the potential benefits and challenges associated with the use of big data in national security, emphasizing the importance of ethical considerations in the use of big data in this context.

**Work on the role of Big Data Analytics in Cities**

The research paper [10] explores the role of big data in the development of cities, particularly in the context of smart cities. The researchers highlight the potential of big data to address challenges faced by cities such as traffic congestion, pollution, and energy consumption. The paper reviews different types of data sources used in smart city development, including social media, sensor data, and open data, and discusses their potential applications in urban planning and management. This paper also discusses the challenges associated with big data in the context of urban development, such as privacy concerns, data quality, and data integration. The researchers emphasize the importance of developing appropriate governance structures and policies for big data in cities to ensure responsible use and ethical considerations. The research paper highlights the significant potential of big data in the development of smart cities and its ability to transform urban planning and management. The researchers suggest that future research should focus on developing appropriate tools and methods for analysing and utilizing big data in the context of smart city development.

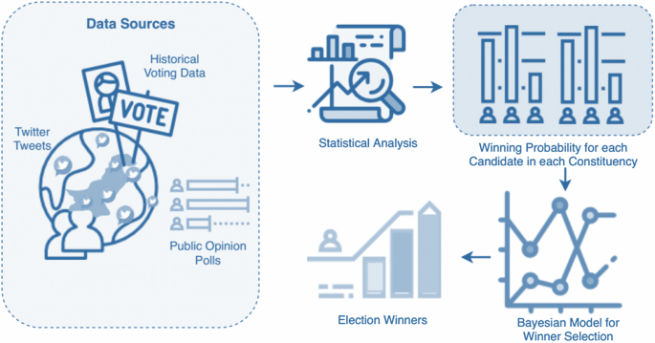
Another research which talks about the urban planning and implementation of big data in the cities is [11]. The research paper discusses the use of big data analytics in the context of urban planning and development in Latin America. The authors argue that big data can provide valuable insights into various aspects of urban life, including transportation, housing, and public health. By analysing large datasets, researchers can identify patterns and trends that can inform policy decisions and improve the quality of life in urban areas. The researchers in this paper conducted a case study of the city of Santiago, Chile, to demonstrate the potential of big data in urban planning. They analysed data from various sources, including social media, GPS tracking, and census data, to gain insights into transportation patterns, land use, and public health. For example, they used Twitter data to identify areas of the city with high levels of air pollution, which could inform policies to reduce emissions and improve public health. The researchers also discuss the challenges of using big data in urban planning, including issues of data privacy and the need for effective data management systems. They argue that policymakers must be aware of these challenges and develop strategies to address them in order to fully realize the potential of big data in urban planning. The research paper majorly highlights the potential of big data analytics in urban planning and development in Latin America and calls for further research and investment in this area. By leveraging the power of big data, policymakers can make more informed decisions that can improve the quality of life for urban residents.

The research paper [16] "The value of Big Data in government: The case of ‘smart cities'" discusses the use of big data in the context of smart cities and the role it plays in predicting plans. The authors examine how governments can leverage big data analytics to gather and analyse vast amounts of data from various sources to derive insights and support decision-making. The paper highlights the potential of big data analytics to enhance the efficiency of government services and urban planning, particularly in the context of smart cities. The authors discuss the importance of data integration, management, and analysis in generating insights and predicting future trends. They also emphasize the need for robust data privacy and security measures to ensure the ethical and responsible use of big data in the public sector. Furthermore, the paper discusses the role of predictive analytics in supporting smart city planning and management. By analysing past trends and patterns, predictive analytics can help governments anticipate future needs and plan, accordingly, leading to more efficient and effective resource allocation.

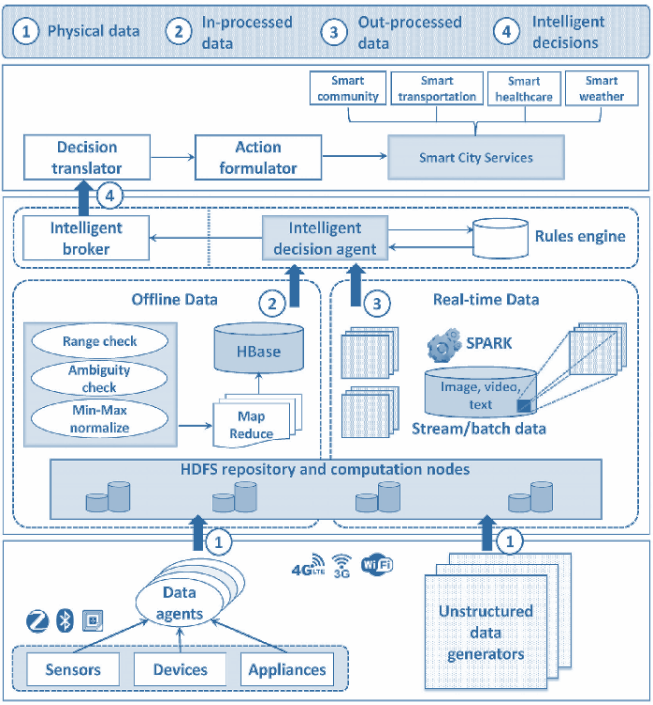


The paper stresses the value of big data in enhancing government operations and services, particularly in the context of smart cities. It provides a roadmap for governments to use big data analytics to drive innovation and create more liveable, sustainable, and resilient urban environments.

There is another research paper which talks about the overall use and applications of Big Data in the field of E-Governance, in voting system, in policy making, in constructions site prediction etc. [18] The research paper discusses various use cases of big data in the government sector in India. The paper highlights that big data analytics can be used to generate revenue from government sites, such as e-commerce portals, by analysing user behaviour and preferences. In the healthcare sector, big data analytics can be used to improve disease surveillance and control, predict outbreaks, and develop personalized treatment plans for patients. In the education sector, big data analytics can be used to track student performance, identify areas of improvement, and enhance learning outcomes.



Big data can also have an impact on government construction projects, where it can be used to optimize project timelines and budgets, reduce construction errors, and improve project management. The Direct Benefit Transfer Scheme is another use case, where big data can be used to ensure efficient delivery of government subsidies and benefits to eligible citizens. The paper also highlights the potential use of big data in the election and voting system. With the help of big data analytics, political parties can analyse voter behaviour and preferences, track their campaign performance, and improve their outreach strategies. Regarding the future plans and predictions, the paper suggests that the use of big data will continue to grow in the government sector in India. The government is already investing in various big data projects, and there is a need for skilled professionals to implement and manage these projects. The paper predicts that big data analytics will play a significant role in improving government efficiency, service delivery, and decision-making processes in the future.

In this paper the authors have analysed the situations of SMEs and they proposed the ways by which the SMEs can develop themselves with the use of Big Data. They have analysed the situations of SMEs during the unnatural circumstances, so that they can prepare themselves for the unnatural circumstances also. [19] The paper discusses the role of big data in the context of the COVID-19 pandemic for Malaysian small and medium-sized enterprises (SMEs). The study emphasizes the importance of big data in managing and mitigating the impact of the pandemic on SMEs. The paper highlights that big data can help SMEs in various ways, such as identifying patterns and trends in consumer behaviour, optimizing supply chain management, and enhancing decision-making processes. Moreover, the paper discusses the big data analytics capability ecosystem model for SMEs, which consists of three levels: technology infrastructure, business process, and organizational capability. The study suggests that SMEs need to develop a comprehensive big data analytics ecosystem to fully utilize the benefits of big data. In the context of the COVID-19 pandemic, the paper highlights that big data can help SMEs to quickly respond to changes in the market and consumer behaviour. 

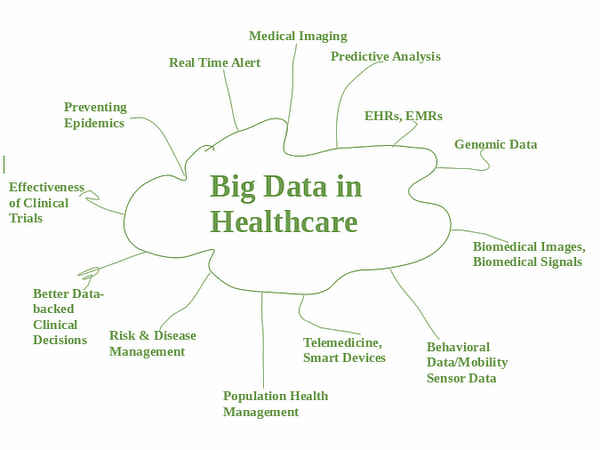
For example, big data analytics can be used to identify which products and services are in high demand during the pandemic, which can help SMEs to adjust their production and sales strategies accordingly. Furthermore, the paper suggests that big data can help SMEs to identify new opportunities during the pandemic.

For example, SMEs can use big data to identify new markets or new products that are in high demand due to the pandemic. Overall, the paper emphasizes the importance of big data for SMEs, particularly in the context of the COVID-19 pandemic. The study suggests that SMEs need to develop a comprehensive big data analytics ecosystem to fully utilize the benefits of big data and stay competitive in the market.

**Work on the role of Big Data Analytics in Public Welfare**

The work [8] "Big Data Analytics in Healthcare: A Review" highlights the role of big data in healthcare and the work of the researchers in this field. The authors discuss the challenges and opportunities presented using big data in healthcare, including the potential for improved patient outcomes, reduced costs, and increased efficiency. The researchers discuss various areas of healthcare where big data analytics can be applied, including disease diagnosis, personalized medicine, and clinical decision-making. They also highlight the importance of data quality and privacy in the use of big data in healthcare. The authors review various studies and applications of big data analytics in healthcare, including the use of machine learning algorithms for predicting disease outcomes and improving diagnosis accuracy. They also discuss the use of big data in identifying patient populations for targeted interventions and the development of precision medicine. The paper emphasizes the potential benefits of big data analytics in healthcare and the need for continued research and development in this field. It also highlights the importance of ethical considerations and the need to balance privacy concerns with the potential benefits of big data in healthcare.

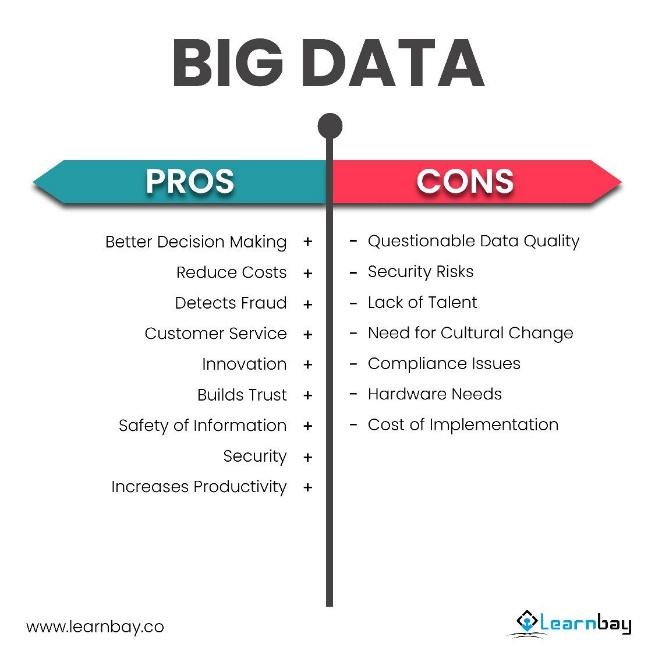
The government's responsibility for managing public affairs has increased with citizen's expectations for quality services. Big data, with its ability to process large amounts of data quickly and accurately, has become a crucial tool for managing public affairs. This paper discusses the role of big data in public affairs management and proposes solutions to existing issues. In [17] the author talked about the government's responsibility for managing public affairs has increased with citizen's expectations for quality services. Big data, with its ability to process large amounts of data quickly and accurately, has become a crucial tool for managing public affairs. This paper discusses the role of big data in public affairs management and proposes solutions to existing issues. The authors discussed the use of big data in public affairs management and its potential impact on public welfare. The authors argue that the rapid development of big data technologies has transformed the landscape of public affairs management, providing new opportunities to gather and analyse large volumes of data in real-time.

The paper discusses how big data can be used to improve public services, increase transparency and accountability, and promote citizen participation in the decision-making process. The authors argue that big data can be used to identify and address social and economic issues, such as poverty, healthcare, and education, by providing insights into the needs and preferences of citizens. Also, the authors explore the challenges of managing and analysing big data in the public sector, such as data security, privacy, and ethics. They suggest that effective data management and governance frameworks are essential for ensuring that the use of big data in public affairs is ethical, transparent, and accountable. The paper highlights the potential of big data to transform public affairs management and improve public welfare. It calls for greater collaboration between governments, citizens, and the private sector to harness the power of big data for the greater good of society.

There is another paper [18] which explores the use of big data in the healthcare industry. The authors investigate the dimensions, components, and key indicators of big data usage in healthcare and its potential for future prediction. The paper discusses the different sources of healthcare data, such as electronic health records, health information exchanges, and wearable devices, and how they can be integrated and analysed to provide insights and support decision-making. The authors also examine the challenges and opportunities associated with big data in healthcare, including privacy concerns, data quality issues, and the need for advanced analytics capabilities. The authors propose a framework for the use of big data in healthcare that includes four dimensions: data acquisition, data management, data analysis, and data utilization. They discuss the components of each dimension, such as data sources, data governance, data quality, and analytics tools. It highlights the key indicators that can be used to measure the effectiveness of big data usage in healthcare, including patient outcomes, cost savings, and provider satisfaction. The authors argue that the effective use of big data can lead to improved healthcare quality, increased efficiency, and better patient outcomes. Basically, it provides a comprehensive overview of the use of big data in the healthcare industry and its potential for future plans prediction. It provides a roadmap for healthcare organizations to leverage big data to improve decision-making, enhance patient outcomes, and reduce costs.

1. **Discussion**

No doubt there is lot of work that has been done in the domain of Big Data and it is still expanding its scope and range with a wide variety of applications in various niches. As we have discussed some of the work that has been done till now in the domain of Government & policies, urban development, and the public welfare with the help of Big Data, so now let’s analyse the pros, the cons and the challenges that has been discovered after all this development.



**Cons of Big Data Involvement**

**Government & Policies**

While big data analytics has the potential to improve decision-making in government, there are also several challenges and potential negative consequences associated with its use. Some of the cons that have arisen due to big data analytics in government and policies are:

Privacy concerns: The collection and analysis of large amounts of data raises concerns about privacy violations. There is a risk that sensitive personal information may be collected and used without the knowledge or consent of individuals.

Lack of transparency: Big data analytics can be complex, and the algorithms used to analyse the data may be opaque to those outside of the organization. This lack of transparency can make it difficult to understand how decisions are being made and to identify and correct errors.

Cybersecurity risks: Collecting and storing large amounts of data can also increase the risk of cyberattacks, which could lead to the loss or theft of sensitive information.

Legal and regulatory challenges: The use of big data analytics in government is subject to a range of legal and regulatory challenges, including compliance with data protection laws and regulations, ensuring data quality, and addressing liability issues.

Lack of human judgment: While big data analytics can provide valuable insights, it cannot replace the value of human judgment in decision-making. Overreliance on data-driven decision-making can result in decisions that do not consider important contextual information or ethical considerations.

**Urban Development**

Bias: Big data analytics can be biased if the data used to train models is incomplete or unrepresentative. This can result in decision-making that is discriminatory against certain groups, perpetuating existing social and economic inequalities. For example, if a city uses data to identify areas that are in need of redevelopment, but only collects data on property values and commercial activity, it may overlook lower-income neighbourhoods that could benefit from investment.

Dependence on technology: Urban development projects often require significant investments in technology and infrastructure, such as sensors, cameras, and data centres. This dependence on technology can be problematic if there are failures or disruptions in the systems used to collect, store, and analyse data. For example, if a city relies on traffic sensors to optimize traffic flow, a malfunction or outage could cause significant disruptions.

Displacement: Big data analytics can contribute to gentrification and displacement if it is used to drive urban development projects that prioritize the interests of wealthier residents or businesses over those of lower-income residents. For example, if a city uses data to identify areas that are ripe for investment, but only considers data on property values and commercial activity, it may overlook lower-income neighbourhoods that are in need of investment.

Lack of community involvement: The use of big data analytics can reduce the role of community members in decision-making processes, leading to a lack of trust and reduced community engagement. For example, if a city uses data to identify areas that are in need of redevelopment but does not engage with residents to understand their needs and priorities, residents may feel that their voices are not being heard.

**Public Welfare**

Stigmatization: Big Data Analytics can be used to label individuals or groups in a stigmatizing manner. For example, if data is collected and analysed to identify individuals who are at high risk for certain diseases or social problems, these individuals may be stigmatized and discriminated against.

Exclusion: The use of Big Data Analytics in public welfare can exclude vulnerable populations who may not have access to digital devices or technology. For example, individuals who do not have access to computers or smartphones may be left out of digital welfare programs, which could result in inequitable distribution of resources.

False assumptions: Big Data Analytics may lead to false assumptions about individuals or communities, which can result in negative consequences. For example, if data is analysed to identify individuals who are at risk of committing crimes, this may result in surveillance or other forms of monitoring that may be unwarranted or discriminatory.

Reliance on algorithms: Public welfare programs that rely heavily on Big Data Analytics may become overly reliant on algorithms, which can be problematic if they are not accurate or if they perpetuate biases. For example, if an algorithm is used to determine eligibility for public assistance, individuals who do not fit the algorithm's parameters may be unfairly excluded.

Lack of human touch: The use of Big Data Analytics in public welfare can reduce the human touch that is often needed to provide effective social services. For example, if chatbots or other automated systems are used to provide customer service, individuals who need personalized attention may not receive the support they need.

**Pros of Big Data Involvement**

**Government & Policies**

Improved decision-making: One of the biggest advantages of Big Data Analytics is its ability to analyse vast amounts of data quickly and accurately. Governments can use this data to make informed decisions, such as allocating resources or implementing policies that are likely to have the greatest impact.

Increased efficiency: Big Data Analytics can help governments identify inefficiencies in their operations and implement changes to increase efficiency. For example, data analysis can help identify areas where there is a high volume of requests for services, enabling governments to allocate resources more effectively.

Cost savings: Big Data Analytics can help governments identify opportunities to reduce costs and eliminate waste. By analysing data, governments can identify areas where spending can be reduced without compromising the quality of public services.

Predictive analytics: Big Data Analytics can help governments predict future trends and events, allowing them to prepare for them in advance. For example, data analysis can help predict future public health trends, allowing governments to take proactive measures to prevent outbreaks or epidemics.

Fraud detection: Big Data Analytics can be used to identify fraud and abuse in government programs. This can include anything from identifying fraudulent claims for benefits to detecting corruption in government procurement processes.

**Urban Development**

Improved infrastructure planning: Big Data Analytics can help urban planners make informed decisions about infrastructure development, such as transportation networks, water supply, and waste management. By analysing data on population density, traffic patterns, and other factors, planners can optimize infrastructure development to meet the needs of the city's residents.

Enhanced sustainability: Big Data Analytics can help cities become more sustainable by providing insights into energy consumption, waste generation, and other factors that impact the environment. With this information, cities can implement policies and initiatives to reduce their environmental footprint and promote sustainable development.

Better public safety: Big Data Analytics can help cities improve public safety by identifying areas with high crime rates or accident rates. By analysing data on crime patterns, traffic accidents, and other factors, cities can deploy resources more effectively to reduce crime and accidents.

Improved citizen engagement: Big Data Analytics can help cities engage citizens more effectively by providing insights into their preferences, behaviours, and needs. With this information, cities can develop policies and initiatives that are better aligned with citizen needs and priorities.

More efficient resource allocation: Big Data Analytics can help cities allocate resources more efficiently by providing insights into resource consumption patterns. For example, by analysing data on water consumption, cities can identify areas with high water usage and implement policies to reduce consumption in those areas.

**Public Welfare**

Improved program effectiveness: Big Data Analytics can help public welfare agencies to analyse data from various sources and identify the root causes of social problems such as poverty, unemployment, and homelessness. With this information, agencies can develop more effective programs and services to address these problems.

Targeted service delivery: Big Data Analytics can help public welfare agencies to target their services more effectively. By analysing data on the demographics and socio-economic characteristics of their clients, agencies can design programs that meet the specific needs of different groups.

Fraud detection and prevention: Big Data Analytics can help public welfare agencies to detect and prevent fraud in social welfare programs. By analysing data on the behaviour patterns of beneficiaries, agencies can identify potential cases of fraud and take action to prevent it.

Real-time monitoring and response: Big Data Analytics can help public welfare agencies to monitor the impact of their programs in real-time. By analysing data on program performance and client outcomes, agencies can identify areas that need improvement and take action to address them.

Enhanced accountability and transparency: Big Data Analytics can help public welfare agencies to demonstrate their accountability and transparency to the public. By analysing data on program performance and outcomes, agencies can provide evidence-based reports to stakeholders, including policymakers, funders, and the public.

**Research Questions in the Big Data Domain**

After reviewing these many papers, and algorithms we came up with several more aspects of the domain, that can be explored more. The following list has those questions or aspects that we came up with.

* How can Big Data be used to improve energy efficiency and reduce carbon emissions?
* What are the best practices for integrating Big Data into the education sector to enhance learning outcomes?
* What are the best approaches for integrating Big Data into environmental monitoring and management?
* Space Exploration: How can Big Data be used to improve space exploration and optimize space missions?
* Social Justice: How can Big Data be used to address social justice issues and inform policy decisions?
* Genetics: How can Big Data be used to analyse genetic data and inform personalized medicine?
* Oceanography: How can Big Data be used to better understand the oceans and inform marine conservation efforts?
* Mental Health: How can Big Data be used to analyse mental health data and inform treatment strategies?

1. **References**
2. Chen, M., Mao, S. and Liu, Y., 2014. Big data: A survey. Mobile networks and applications, 19, pp.171-209.
3. Ayma, V.A., Ferreira, R.S., Happ, P., Oliveira, D., Feitosa, R., Costa, G., Plaza, A. and Gamba, P., 2015. Classification algorithms for big data analysis, a map reduce approach. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(3), p.17.
4. Cai, X., Nie, F. and Huang, H., 2013, June. Multi-view k-means clustering on big data. In Twenty-Third International Joint conference on artificial intelligence.
5. Suthaharan, S. and Suthaharan, S., 2016. Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning, pp.207-235.
6. Liu, B., Blasch, E., Chen, Y., Shen, D. and Chen, G., 2013, October. Scalable sentiment classification for big data analysis using naive bayes classifier. In 2013 IEEE international conference on big data (pp. 99-104). IEEE.
7. Liu, Y., 2014. Random forest algorithm in big data environment. *Computer modelling & new technologies*, *18*(12A), pp.147-151.
8. Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R. and Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. *Journal of big data*, *2*(1), pp.1-21.
9. Caes, M., 2019. Access to and Re-use of Government Data and the Use of Big Data in Healthcare. *Regulating New Technologies in Uncertain Times*, pp.193-224.
10. Ukhalkar, P., Unexpected use of Big Data and Analytics in Government and Industry Domains. *International Research Journal I*, *4*, pp.4-8.
11. Hong, S., Hyoung Kim, S., Kim, Y. and Park, J., 2019. Big Data and government: Evidence of the role of Big Data for smart cities. *Big data & society*, *6*(1), p.2053951719842543.
12. Ueti, R.D.M., Espinosa, D.F., Rafferty, L. and Hung, P.C., 2016. Case studies of government use of big data in Latin America: Brazil and Mexico. Big Data Applications and Use Cases, pp.197-214.
13. Amri, A.I.S.U., 2021. The Use of Big Data to Determine the Government’s Communication Strategy. *KnE Social Sciences*, pp.275-289.
14. Larson, E., 2020. Big questions: Digital preservation of big data in government. The American Archivist, 83(1), pp.5-20.
15. Benjelloun, F.Z. and Lahcen, A.A., 2019. Big data security: challenges, recommendations and solutions. In *Web Services: Concepts, Methodologies, Tools, and Applications* (pp. 25-38). IGI Global.
16. Joseph, R., 2015. From data to vision: Big data in government. In Strategic data-based wisdom in the big data era (pp. 1-14). IGI Global.
17. Löfgren, K. and Webster, C.W.R., 2020. The value of Big Data in government: The case of ‘smart cities’. Big Data & Society, 7(1), p.2053951720912775.
18. In, L.C., 2022, December. Public Affairs Management and Innovation from Big Data. In 2022 6th International Seminar on Education, Management and Social Sciences (ISEMSS 2022) (pp. 261-266). Atlantis Press.
19. PARMAR, H.H., 2013. APPLICATIONS OF BIG DATA IN GOVERNMENT SECTOR. In International Conference on Big Data Management & Cloud Computing (pp. 29-31).
20. Falahat, M., Cheah, P.K., Jayabalan, J., Lee, C.M.J. and Kai, S.B., 2023. Big Data Analytics Capability Ecosystem Model for SMEs. Sustainability, 15(1), p.360.
21. Sazu, M.H. and Akter Jahan, S., 2022. Impact of big data analytics on government organizations. Management & Datascience, 6(2).
22. Yakhtin, V.A., 2021. National Information Risks of Big Data. Vestnik Volgogradskogo Gosudarstvennogo Universiteta. Serii︠a︡ 3, Ėkonomika, Ėkologii︠a︡, 23(3).