International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 4, Jul-Aug 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Big Data in Transportation

Matthew N. O. Sadiku[1], Paul A. Adekunte[2], Janet O. Sadiku[3]

[1]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]International Institute of Professional Security, Lagos, Nigeria
[3]Juliana King University, Houston, TX, USA

ABSTRACT

Data is transformed into information helping us to understand buying trends, social behaviors, and various predictive analytics. Big data blends together the collection of large volumes of high-velocity, heterogeneous, evolving domain data and the use of advanced techniques and models to store, retrieve, manage, process, and analyze the captured information. Big data analytics has changed many sectors, including transportation. This technology helps us understand and predict transportation systems. It makes managing transportation data easier. In transportation, big data is able to exploit information and solve transportation problems at unprecedented scales. In this paper, we focus on the applications of big data and big data analytics in the transportation industry.

KEYWORDS: *big data, big data analytics, big data applications, transportation, transportation data analytics.*

How to cite this paper: Matthew N. O. Sadiku | Paul A. Adekunte | Janet O. Sadiku "Big Data in Transportation" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-4, August 2025, pp.1049-1058, URL: www.ijtsrd.com/papers/ijtsrd97385.pdf

IJTSRD97385

INTRODUCTION

Transportation, as a means for moving goods and people between different locations, is a vital element of modern society. The transportation industry ensures the efficient and safe transportation of people and goods from one place to another. In an era where technology continues to reshape our world, the transportation sector stands out as one of the most significantly impacted. In recent years, the transportation industry has been disrupted by multiple forces, including the COVID-19 pandemic, an ongoing road safety crisis, and a growing push for decarbonization. The transportation sector accounts for one-fifth of global CO_2 emissions and is one of the most challenging sectors to decarbonize due to its heavy reliance on fossil fuels. The planning of a city faces the transportation dilemma. On one hand, transportation is considered fundamental for the operation of the region. On another hand, it is considered one of the most important causes of environmental deterioration. Figure 1 shows a typical urban traffic [1].

Organizations have started to analyze and explore how to examine a vast array of information in innovative and new ways to derive the best possible business outcomes. We have seen significant benefits of big data in education, healthcare, telecommunications, media and entertainment, finance, and marketing. And the transportation industry is not outside of the race. From optimizing traffic flow to improving safety and reducing costs, big data analytics has radically transformed transportation management and operations. Organizations across various transportation and travel segments like airports, airlines, freight logistics, railways, hospitality, and others are enjoying the benefits of big data in managing a large amount of data that they handle. The era of big data generates massive volumes of transportation data daily, with storage exceeding petabytes and transmission rates reaching hundreds of terabytes per day [2].

WHAT IS BIG DATA?

Big data applies to data sets of extreme size (e.g. exabytes, zettabytes) which are beyond the capability of the commonly used software tools. It involves situation where very large data sets are big in volume, velocity, veracity, and variability [3]. The data is too

big, too fast, or does not fit the regular database architecture. It may require different strategies and tools for profiling, measurement, assessment, and processing. Different components of big data are shown in Figure 2 [4]. The cloud word for big data is shown in Figure 3 [5].

Big Data is essentially classified into three types [6]:

- *Structured Data:* This is highly organized and is the easiest to work with. Any data that can be stored, accessed, and processed in the form of fixed format is known as a structured data. It may be stored in tabular format. Due to their nature, it is easy for programs to sort through and collect data. Structured data has quantitative data such as age, contact, address, billing, expenses, credit card numbers, etc. Data that is stored in a relational database management system is an example of structured data.
- *Unstructured Data:* This refers to unorganized data such as video files, log files, audio files, and image files. Any data with unknown form or the structure is classified as unstructured data. Almost everything generated by a computer is unstructured data. It takes a lot of time and effort required to make unstructured data readable. Examples of unstructured data include Metadata, Twitter tweets, and other social media posts.
- *Semi-structured Data:* This falls somewhere between structured data and unstructured data, i.e., both forms of data are present. Semi-structured data can be inherited such as location, time, email address, or device ID stamp.

The different types of big data are depicted in Figure 4 [7].

The process of examining big data is often referred to big data analytics. It is an emerging field since massive computing capabilities have been made available by e-infrastructures [8]. Big data analytics is the application of advanced analytic techniques to large, heterogeneous data sets that comprise structured, semi-structured, and unstructured data from many sources with sizes ranging from terabytes to zettabytes.

It enables predictive analytics, which involves using historical data to forecast future outcomes. Analytics include statistical models and other methods that are aimed at creating empirical predictions. Data-driven organizations use analytics to guide decisions at all levels. Several techniques have been proposed for analyzing big data. These include the HACE theorem, cloud computing, Hadoop, and MapReduce [9]. Figure 5 shows big data analytics [10].

CHARACTERISTICS OF BIG DATA

Big data is growing rapidly and expanding in all science and engineering, including physical, biological, and medical services. Different companies use different means to maintain their big data. As shown in Figure 6 [11], big data is characterized by 42 Vs. The first five Vs are volume, velocity, variety, veracity, and value .

- *Volume*: This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as data over one petabyte in volume.
- *Velocity*: This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- *Variety*: This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- *Veracity*: By this, we mean the truthfulness of data, i.e. weather the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- *Value*: This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

On this basis, small data can be regarded as having low volume, low velocity, low variety, low veracity, and low value. Additional five Vs has been added [11]:

- *Validity:* This refers to the accuracy and correctness of data. It also indicates how up to date it is.
- *Viability:* This identifies the relevancy of data for each use case. Relevancy of data is required to maintain the desired and accurate outcome through analytical and predictive measures.
- *Volatility:* Since data are generated and change at a rapid rate, volatility determines how quickly data change.

- *Vulnerability:* The vulnerability of data is essential because privacy and security are of utmost importance for personal data.
- *Visualization:* Data needs to be presented unambiguously and attractively to the user. Proper visualization of large and complex clinical reports helps in finding valuable insights.

Instead of the 10V's above, some suggest the following 5V's: Venue, Variability, Vocabulary, Vagueness, and Validity) [12].

Industries that benefit from big data include the healthcare, financial, airline, travel, restaurants, automobile, sports, agriculture, and hospitality industries. Big data technologies are playing an essential role in farming: machines are equipped with sensors that measure data in their environment. The analysis of both structured and unstructured data is crucial in the shipping industry to gain insights into customer behavior, improve operational efficiency, and make informed business decisions.

TRANSPORTATION DATA ANALYTICS

Transportation data should be evaluated against an existing data set with confirmed accuracy. This is usually data from road sensors or counters. A powerful data set is not a one-size-fits-all solution to every question or problem that transportation planners and managers confront, but it is a versatile multi-tool in the transportation toolbox. Big data analytics has emerged as a powerful tool for businesses and organizations to gain insights and make informed decisions. It can potentially revolutionize the transportation industry. It provides valuable insights into customer behavior, traffic patterns, and operational efficiency. Big data analytics in transportation planning enables effective navigation through the complexities of urban mobility. Figure 7 shows big data in transportation [13].

Transportation big data analytics refers to the process of collecting, analyzing, and interpreting large, complex datasets related to transportation systems to gain insights and improve various aspects of transportation planning and management. This involves utilizing data from diverse sources like connected vehicles, IoT devices, and public transportation systems to understand travel patterns, predict future trends, and optimize transportation networks. Transportation data analytics increasingly power mobility information and insights – transforming transportation planning and operations by making it easier, faster, cheaper, and safer to collect and understand critical information. It can provide complete end-to-end trip information, including trip origins and destinations, routes, trip distances, travel time, and even real-time data on how vehicles are moving. Using transportation data analytics, transportation professionals can quickly access accurate data for every road in the country, every day of the year — even in real time. More and more cities, transit organizations, departments of transportation, and other localities are using transportation data analytics to solve problems, prioritize investments, and win stakeholder support [1]. Figure 8 is a representation of big data analytics in transportation [14].

APPLICATIONS OF BIG DATA IN TRANSPORTATION

Some real-world applications of big data in transportation are depicted in Figure 9 [15]. Common areas of application of big data and big data analytics in transportation include the following [14,16,17]:

- *Smart Cities:* The term "smart city" denotes a city with the technology for instrumentation, interconnection, and smart systems. Instrumented means that the city is in the ability to capture and integrate data through the use of sensors, applications, personal devices, etc. Interconnected means that the data is on a computer platform that allows the communication of different sources of information for multiple services to the city. And the term smart means that analytics models, optimizations, and visualizations can be built as a result of complex decision-making applications. Every day, urban transport creates over 500 petabytes of data. This number grows with smart city tech. This rich data, from commuters, vehicles, and infrastructure, boosts transportation planning. Smart city transportation analytics inform sustainable urban development and transport policy-making. Cities must now use data-driven approaches to improve mobility, safety, and efficiency. For city dwellers, heavy traffic and a lack of parking are just part of urban living. But as cities get "smarter," both residents and commuters may reap benefits like less congestion and more readily identifiable parking spaces.
- *Smart Transportation:* Smart transportation refers to the integration of advanced technologies in transportation systems to improve efficiency, safety, and sustainability. It encompasses a wide range of applications, from traffic management and public transportation to logistics and freight operations. The core of smart transportation lies in real-time data collection and analysis, enabling better decision-making and improved service delivery. In the context of transportation, big data is generated from various sources, including GPS

devices, traffic cameras, social media, and IoT sensors embedded in vehicles and infrastructure. The role of big data analytics and AI also plays a crucial role in predicting transportation demand.

- *Intelligent Transport Systems:* The systems that support transport in smart cities are called Intelligent Transport Systems (ITS). ITS are applications that use the synergy between technologies and engineering concepts to develop and improve multiple types of transportation systems. As far as urban population continuously grows, the need of ITS becomes more and more relevant to deal with transport needs for both people and cargo. Automated devices for data capture requires the implementation of analytics techniques in order to provide insights for actual decision-making and policy design. Intelligent transportation systems and machine learning for transportation data are changing how we move. In the near future, the big data paradigm is expected to become a core building block for ITS.
- *Traffic Management:* Traffic jams are the bane of many city dwellers and a locus of profound pressure for city managers. No city is completely free of traffic jams. Traffic management proves to be one of the most essential applications of big data analytics in transportation. It allows public transportation companies to monitor real-time patterns and optimize traffic flow. They can significantly reduce traffic congestion, improve travel times, and enhance safety. Big data analytics improves public transportation scheduling. It helps match bus or train times to when people need them. This makes travel better for everyone while cutting costs. By analyzing traffic flow and congestion patterns, transportation agencies can implement strategies to optimize traffic flow, reduce congestion, and improve overall travel times. By analyzing data from various sources, traffic management systems can optimize traffic flow, reduce congestion, and minimize travel times.
- *Transportation Planning:* We are at the brink of a major breakthrough in city design and connectivity. The use of big data analytics is key for advanced transportation infrastructure planning. As cities grow quickly, we face many new challenges. Data-driven planning has become essential because it builds a system where choices are guided by up-to-date data and forecasts. In a changing transportation landscape, it is more important than ever for planners to have access to detailed data and analytics. Given our increasingly complex transportation networks, planners must consider where to implement road diets, add EV charging infrastructure, establish tollways, expand access to multimodal transport, and so much more. Big data analytics offers tools to understand vast amounts of data from different sources. It helps make better decisions in transportation planning and optimization. Having access to detailed real-time and historical transportation data empowers planners and transportation departments to develop better strategies, reduce costs, prioritize initiatives, and measure the effectiveness of each change or improvement.
- *Predictive Analytics:* Predictive modeling considers various factors, like weather and events, to avoid traffic problems. Using predictive analytics enables smoother traffic flow across all transport modes. Predictive analytics are a game-changer for railways. They make trains more timely and safe. By predicting needs, railways can plan better and avoid. With predictive analytics, agencies can answer the question of "what's the best possible result?" instead of using prior history information.
- *Predictive Maintenance:* When it comes to maintenance prediction, big data analytics do a great job of monitoring vehicle health and forecasting necessary preservation. To achieve this, you have to use tools to collect data from sensors embedded in cars, maintenance logs, and historical data on vehicle performance. The data, later on, will be studied thoroughly. Employing data analytics will contribute to reducing downtime while strengthening safety and saving maintenance costs. Data from sensors and operational logs help predict equipment failures before they occur, minimizing downtime and repair costs.
- *Road Infrastructure:* Cities around the world have one problem in common: road infrastructure. Road repair and development is a frustrating, and probably one of the most delayed processes. To address this problem, many cities that have resorted to use big data to address road infrastructure by venturing into the development of a big data app. For example, Boston is one of the first users whose residents are experimenting with an app known as Street Bump. It enables residents to identify jolts and potholes on roads. The app uses signal data from a cell phone's accelerometer for the purpose of detection.

BENEFITS

Big data is revolutionizing the transportation industry by offering transformative benefits. The benefits of

big data and analytics help the transportation companies to accurately optimize and model capacity, demand, pricing, revenue, customer sentiments, cost, and lot more. Transportation industry can make more informed decisions, improve safety, reduce costs, and gain positive customer feedback by leveraging the big data generated by sensors, cameras, and other devices. The new benefits that big data analytics brings to the table are speed and efficiency. Other benefits of big data in transportation include the following [18]:

- *Connectivity:* Connectivity is hugely important for transportation. As connected networks and technologies like 5G, IoT, and AI evolve, smart cities will get even smarter. In the city of the future, buses, trains, cars, and even roadways will be interconnected. This will enable digital systems to "talk" with one another, and help cities to move people more efficiently than ever.
- *Better Safety:* Big data and the IoT can help in mitigating faults of signals. Passengers and cargo are advantageous to big data analytics in ensuring they stay safe while traveling. You can reduce accidents, prevent theft, and enhance overall security.
- *Cost Savings*: A well-optimized logistics system reduces operational costs by minimizing unnecessary inventory holding, streamlining transportation and distribution, and optimizing routes. This leads to reduced warehousing expenses, lower transportation costs, and efficient use of resources, ultimately increasing profitability.
- *Enhanced Knowledge*: Big data and IoT are capable of making customers or users aware of the most efficient form of transport at any given time. There are several train operating companies that have already started using big data to process availability of seat data in real-time and also to inform the passengers waiting on platforms about the carriages having the most number of available seats. This benefit of big data not only enhances the experience of customers but also adds up to their knowledge.
- *Improved Customer Service*: One of the most significant benefits of big data is the fact that it has enhanced the experience of customers by improving their knowledge. Real-time data insights enable better tracking of shipments and more accurate delivery estimates, enhancing customer satisfaction. An effective logistics system ensures timely and accurate deliveries, leading to improved customer service. On-time deliveries, accurate order fulfillment, and real-time tracking capabilities enhance customer satisfaction, trust, and loyalty. Meeting or exceeding customer expectations is crucial for maintaining a competitive edge.
- *Competitive Advantage:* Big data empowers organizations to stay competitive, optimize their operations, and drive strategic decisions with confidence. Effective logistics systems enable businesses to respond quickly to market changes and customer demands. This agility provides a competitive advantage, allowing companies to adapt to evolving conditions, optimize inventory levels, and reduce lead times, ensuring they remain at the forefront of their industry.
- *Market Expansion:* Efficient logistics enable businesses to reach new markets and customers. With streamlined supply chain processes, companies can enter new regions or expand their customer base, unlocking growth opportunities and increasing market share.
- *Efficient Operation:* Big data can be used to eliminate errors and minimize unnecessary spending. It can be leveraged to identify the problems related to delays and downtimes for transport maintenance. Strukton rail in the Netherlands is a very good example of how the benefits of big data can make operations efficient.

Some of these benefits are displayed in Figure 10 [18].

CHALLENGES

There are also significant challenges associated with transportation analytics. Extracting and processing transportation big data is complex. Capturing and mining video data requires a robust, low-latency data governance and management platform. There is a far more important question of personal data privacy. Various factors such as privacy, regulations, and confidentiality must be taken care of to use data 100% effectively to provide fruitful analysis. Other challenges include the following [14]:

- *Data Privacy and Security:* One of the main issues with data privacy and security is the potential for sensitive data to be accessed by unauthorized individuals or entities. As a matter of fact, data from a GPS device may contain information about the location and movements of individuals. And they could serve malicious purposes if it falls into the wrong hands. It is critical to control access to your data by permitting authorized individuals only. Others will be blocked right after entering the data system. Fortunately, best practices are emerging to ensure privacy protection.

- *Data Quality:* Data quality refers to the completeness, homogeneity, and trustworthiness of the data. With many forms of big data, quality and accuracy are less controllable, as large data volumes often make up for the lack of quality or accuracy. There exists a stereotype that analyzed data must be accurate, complete, and reliable. In the transportation industry, these requirements are considered big challenges due to the large volume of data generated from multiple sources. The main issue is the potential for errors or inconsistencies in the data. Governments or transportation service providers should take into account robust data quality control processes. They must check data validation for errors and inconsistencies and clean data to remove or correct inaccurate information.
- *Data Integration:* Integrating multiple data sources into a single, unified dataset has never been an easy task, particularly in the transportation industry. Data stored in different formats or structures prevents us from combining and analyzing them effectively. Data transformation techniques enable data scientists to convert data from different formats into a common one for practical analysis. Besides, they ought to implement data alignment to process multiple-source data together.
- *Data Cost:* Engineering and analyzing data cost organizations a lot of money. The larger volume of data you want to collect and store, the higher expense you have to spend. This also happens to processing and analyzing that data. It is necessary to provide clear and cost-effective data management strategies. This may involve implementing data archiving techniques to keep older data on less expensive storage media.
- *Data Storage:* Data storage and processing should take place in a secure data repository protected by multi-layered network security architecture, and supported by system audits and controls.
- *Transparency:* At a minimum, transportation data providers should be able to explain the modeling behind a transportation data algorithm, including the data sources, how the data is handled, and the algorithm's capabilities. Transparency is essential for assessing today's complicated data sets.
- *Sustainability:* Sustainability is a pressing concern for modern transportation systems. Big Data analytics can assess the environmental impact of various transportation modes, while AI can suggest optimal routes for electric and hybrid vehicles, minimizing energy consumption. Together, these technologies can help cities achieve their sustainability goals, reducing carbon emissions and promoting greener alternatives.

CONCLUSION

Big data has remarkably impacted the transportation industry. Transportation data analytics are being used by an increasing number of cities, transit organizations, transportation departments, and other entities to solve problems, prioritize investments, and gain stakeholder support. It can give complete trip information from start to finish, including origins and destinations, routes, trip distances, and journey time. Transportation experts use analytics tools such as R, Python, and MATLAB to help analyze complex data effectively. Transportation professionals must make informed decisions based on current, reliable data, not on educated assumptions or the opinions of a few loud stakeholders. They may instantly obtain reliable data for every route in the country, every day of the year, using data analytics [19]. More information about big data in transportation can be found in the books in [20-26] and the following related journals:

- Journal of Big Data
- Big Data and Cognitive Computing
- IET Intelligent Transport Systems
- Journal of Traffic and Transportation Engineering

REFERENCES

[1] "What are big data analytics in transportation?" https://www.streetlightdata.com/what-transportation-big-data-analytics/

[2] Z. Luo et al., "Insights into transportation CO_2 emissions with big data and artificial intelligence," *Patterns*, vol. 6, no. 4, April 2025.

[3] M. N. O. Sadiku, M. Tembely, and S.M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research,* vol. 3, no. 2, 2016, pp. 106-108.

[4] "Big data: What it is and why it matters?" August 2024, https://www.inventateq.com/top-stories/big-data-what-it-is-and-why-it-matters/

[5] L. Rembert, "How accounting teams can leverage big data," https://tdwi.org/articles/2020/03/03/adv-all-how-accounting-teams-can-leverage-big-data.aspx

[6] "The complete overview of big data," https://intellipaat.com/blog/tutorial/hadoop-tutorial/big-data-overview/

[7] R. Allen, "Types of big data | Understanding & Interacting with key types (2024)," https://investguiding-

com.custommapposter.com/article/types-of-big-data-understanding-amp-interacting-with-key-types

[8] P. Baumann et al., "Big data analytics for earth sciences: The earthserver approach," *International Journal of Digital Earth*, vol. 19, no. 1, 2016, pp.3-29.

[9] X. Wu et al., "Knowledge engineering with big data," *IEEE Intelligent Systems*, September/October 2015, pp.46-55.

[10] "Comprehensive guide to big data analysis," May 2024, https://www.sprinkledata.com/blogs/comprehensive-guide-to-big-data-analysis

[11] "The 42 V's of big data and data science," https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

[12] P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in N. Bouchemal (ed.), *Intelligent Systems for Healthcare Management and Delivery*. IGI Global, chapter 4, 2019, pp. 72-100.

[13] N. Joshi, "This is why big data in transportation is a big deal," September 2017 https://www.allerin.com/blog/this-is-why-big-data-in-transportation-is-a-big-deal

[14] T. Nguyen, "The potentials of big data analytics in transportation," https://www.neurond.com/blog/big-data-analytics-in-transportation

[15] V. Terekhov, "How to use big data analytics in transportation planning," May 2024, https://attractgroup.com/blog/big-data-analytics-in-transportation-planning/

[16] J. R. Montoya-Torres et al., "Big data analytics and intelligent transportation systems," *IFAC-Papers OnLine,* vol. 54, no. 2, 2021, pp. 216-220.

[17] N. Singh, "The role of big data analytics and AI in smart transportation," September 2024, https://www.aeologic.com/blog/the-role-of-big-data-analytics-and-ai-in-smart-transportation/

[18] "Analyzing benefits of big data: The real application of big data in transportation," September 2024, https://www.quantzig.com/blog/benefits-of-big-data-transportation/

[19] "What are big data analytics in transportation?" https://www.streetlightdata.com/what-transportation-big-data-analytics/

[20] M. N. O. Sadiku, U. C. Chukwu, and P. O. Adebo, *Big Data and Its Applications.* Moldova, Europe: Lambert Academic Publishing, 2024.

[21] A. E. Hassanien et al. (eds.), *Big Data in Complex Systems: Challenges and Opportunities*. Springer, 2015.

[22] E. Simoudis, *The Big Data Opportunity in Our Driverless Future.* Corporate Innovators, 2017.

[23] E. L. Sanfilippo, *GPS Big Data and Mobility Analysis: A Practical Guide With 18 Real Case Studies To Effectively Understand and Use Big Data in Urban Planning, Transportation, And Traffic Models (Transport Big Data).* Independently Published, 2024.

[24] S. Moridpour, A. T. Pour, and T. Saghapour, *Big Data Analytics in Traffic and Transportation Engineering: Emerging Research and Opportunities (Advances in Civil and Industrial Engineering).* Engineering Science Reference, 2019.

[25] G. Zhao and G. Zhu, *Big Data Transportation Systems.* World Scientific Publishing Company, 2021.

[26] C. Yang and S. V. Ukkusuri (eds.), *Transportation Analytics in the Era of Big Data.* Springer, 2018.

Figure 1 A typical urban traffic [1].

Figure 2 Different components of big data [4].

Figure 3 The cloud word for big data [5].

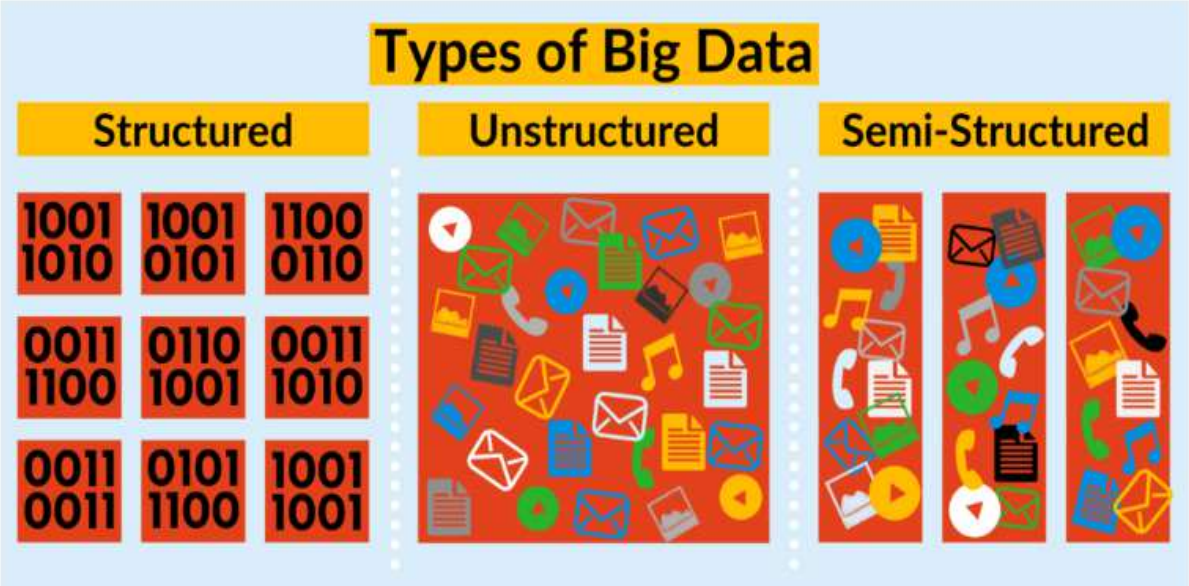

Figure 4 Types of big data [7].

Figure 5 Big data analytics [10].

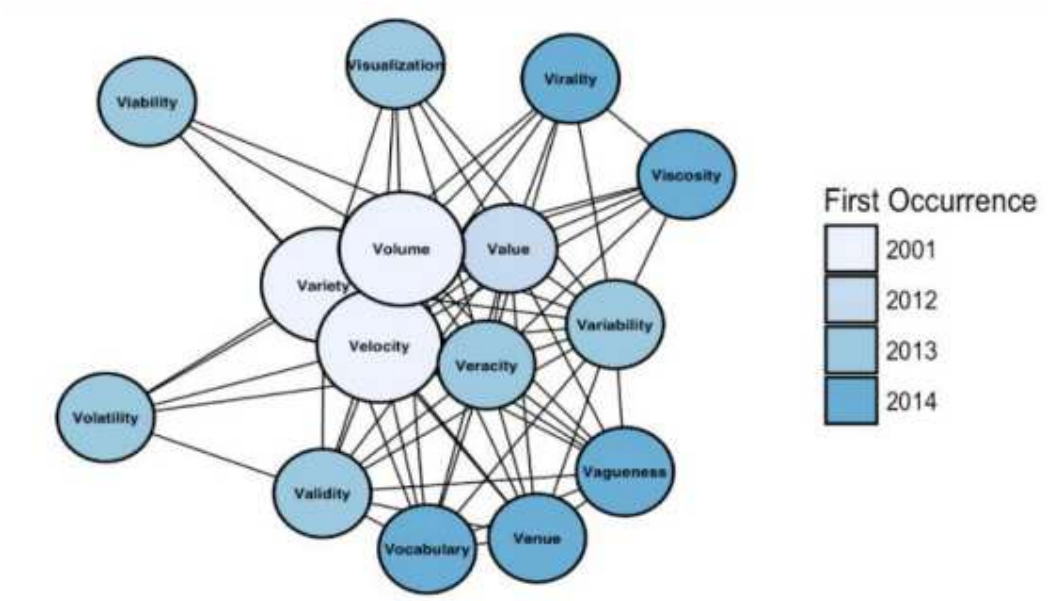

Figure 6 The 42 V's of big data [11].

Figure 7 Big data in transportation [13].

Figure 8 A representation of big data analytics in transportation [14].

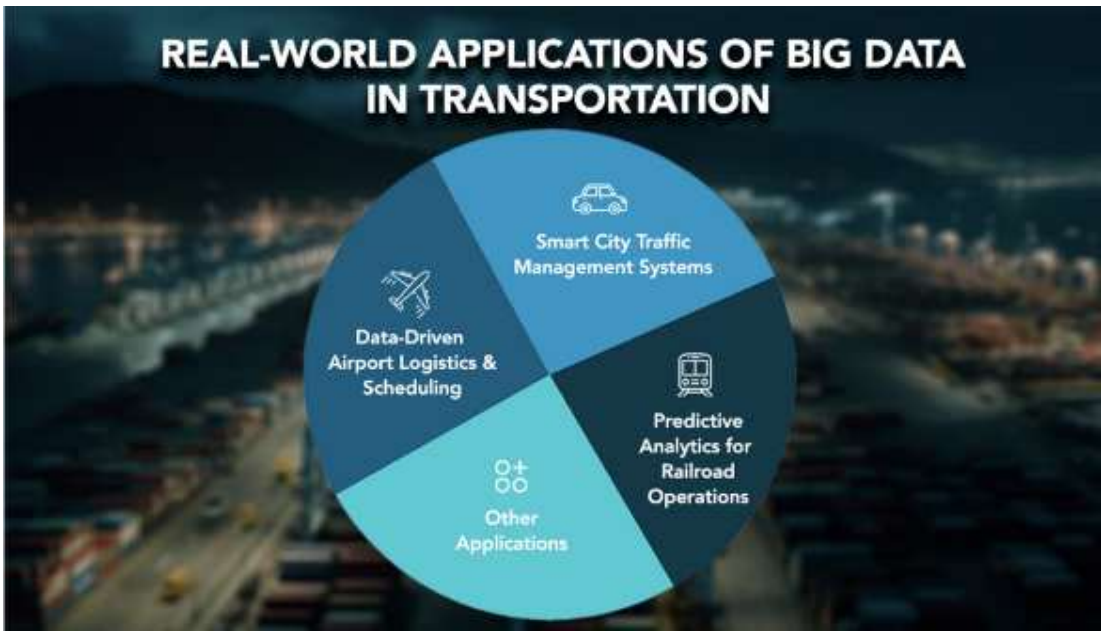

Figure 9 Some real-world applications of big data in transportation [15].

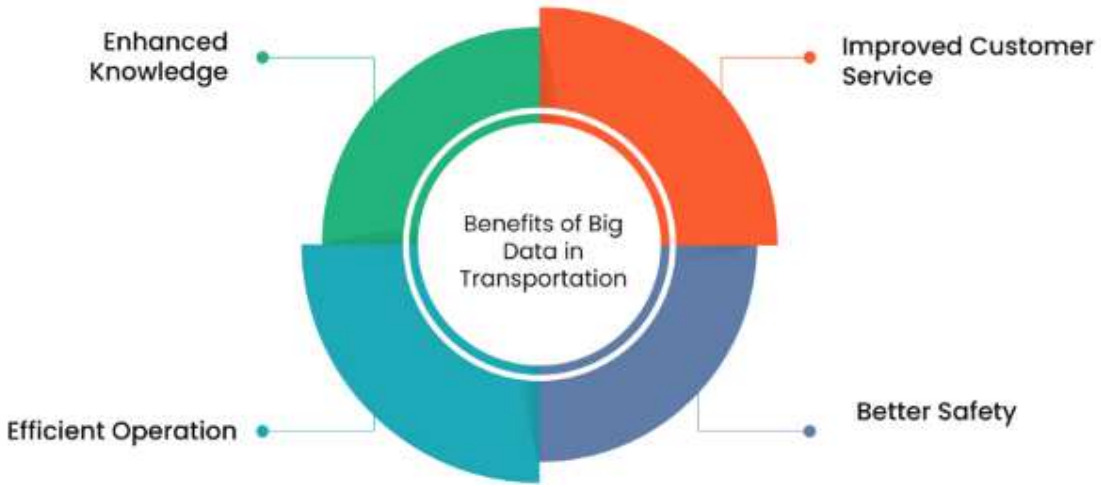

Figure 10 Some of the benefits of big data in transportation [18].