Big Data in Media and Entertainment

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

We are living in the era of big data, which is huge amounts of data in digital form. From healthcare to finance, big data is being used to transform how industries function, enabling business enterprises to create new revenue streams, enhance customer experiences, and increase operational efficiency. Big data plays a crucial role in the media and entertainment industry by enabling companies to understand audience behavior, personalize content, and optimize marketing efforts. It significantly impacts the media and entertainment industry by enabling data-driven decision-making, improving user experiences, and streamlining operations. It is now the real hero for the media and entertainment industry. This paper presents an overview of the state of the art of big data in the media and entertainment industry.

KEYWORDS: big data, big data analytics, media and entertainment, M&E, M&E industry

INTRODUCTION

Today, the term "media" encompasses not only television, radio and print, but also phone calls, text messaging, social platforms, and video chatting — any channel through which information and entertainment is disseminated. The media and entertainment industry is all about art and employing big data in it. Publishers, broadcasters, news organizations, cable companies, and gaming companies in the media and entertainment industry are facing new business models for the way they create, market, and distribute their content. This is happening because today's consumers search and access content anywhere, at any time, and on any device. For the media and entertainment industry, their customers are the real kings and big data is helping them to treat their customers like a one. With millions of digital consumers, media and entertainment companies are in a unique position to leverage their big data assets for more profitable customer engagement [1].

The media and entertainment industries have frequently been at the forefront of adopting new technologies. With this rapidly growing big data market, organizations are leveraging big data to gain insights that help them make better decisions, improve operations and ultimately drive optimal growth. The exploitation of data in the media industry has always played an important role, especially nowadays, when people interact with various sources of information and spend more time online, producing data through their devices (smartphones, tablets, laptops, etc). Media platforms use big data to track content performance across various platforms, such as social media, streaming services, and websites. This can help companies identify trends and optimize their content strategy. For example, media companies like Disney track the performance of their movies and TV shows across various platforms to understand audience engagement and optimize their content strategy.

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 [2]. 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 1 [3]. The cloud word for big data is shown in Figure 2 [4].

Big Data is essentially classified into three types [5]:

➤ *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 3 [6].

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 [7]. 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.

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 [8].

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 4 [9], 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 [9]:

➤ *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) [10].

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. Structured and unstructured data are generated in various types [11-14].

BIG DATA IN MEDIA AND ENTERTAINMENT

The media and entertainment industry is evolving constantly, and data analytics is emerging as a game-changer in shaping the landscape. It is evolving at an unprecedented rate, driven by the twin needs to reduce operating costs and simultaneously generate more revenue from increasingly competitive and uncertain markets. This transformational integration of big data and analytics is not just reshaping content creation and distribution but is also redefining the landscape of media and entertainment software development. The transformative impact of big data and analytics in the media and entertainment industry is enabling the industry to strategically use data analytics to reshape the industry in several key areas, drive innovation, and enhance user experiences. Figure 5 shows a representation of the media and entertainment industry [15].

Over the past few years, big data has grown crucial in the media and entertainment industries. Entertainment businesses have obtained through insights into their clients, systems, and processes by embracing big data analytics. Big data in media and entertainment is not only assisting businesses to gain hidden insights into consumer behavior but is also helping the delivery of personalized content. Today, we can have access to our favorite shows and movies anytime, anywhere with the advancements of services by using big data analytics. We now have access to everything at our fingertips and big data has been the backbone of this amazing transformation. Because of this, technological development has become essential in bringing outside entertainment to consumers' homes.

EXAMPLE OF BIG DATA IN MEDIA AND ENTERTAINMENT

Big data in the media industry is a big deal. The most competitive media companies are adopting big data solutions to help manage their data, generate new insights, and improve their services. For example, the most pre-eminent players in the media and entertainment industry, such as Netflix, Amazon, Hulu, and Disney, have already been leveraging big data as part of their operations to enhance the customer experience. The following are examples of such companies [16]:

➤ *Spotify* is a music streaming service using listening data to curate personalized experiences for its users. Spotify has more 80 million users, which translates to a ton of listening data. So much data, in fact, that Spotify has a blog that displays the different ways data is understood on the platform.

➤ *Instagram* is a social platform that connects users through photos and videos. Acquired by Facebook in 2012, the widely used site has upwards of one billion users. That means it also has mountains of data. Instagram offers many kinds of profiles, including business profiles that offer users data insights about things like the reach of individual posts, past post comparisons, and the origins of audience impressions.

➤ *Ampersand* is a digital advertising solution specializing in targeted TV ads. The company manages and analyzes large volumes of data to ensure effective targeting and campaign optimization. Ampersand works to equip advertisers with insights and solutions to help them make data-driven decisions that maximize the impact of their advertising campaigns.

➤ *NBCUniversal* is a titan of media and entertainment that captivates audiences with shows and movies distributed through an array of network television, film studio, and streaming brands. Data plays an important role in NBCUniversal's business strategy. The company leverages analytics tools to access user insights that inform decision making — an approach that has led to increased viewership.

➤ *Hulu* is a streaming service that provides access to a library of TV shows and movies, original content, and current seasons of television series airing on stations like FOX, ABC, and NBC. Accessible across devices, Hulu regularly adds content and recently expanded to offer a live TV service.

➤ *Netflix*'s leadership in the marketplace can be directly attributed to successfully harnessing the power of big data. Big data and a robust analytics program have been a competitive advantage for Netflix over the past decade. With over 200 million subscribers, Netflix leverages its wealth of audience data to deliver personalized content recommendations and platform experiences. As

the world's most popular streaming platform, Netflix analyzes its financial data to forecast demand for content, estimate its financial impact, and invest dollars to bring audiences the content they want.

APPLICATIONS OF BIG DATA MEDIA AND ENTERTAINMENT

The integration of big data and analytics is foundational to the new era in the entertainment industry. Various areas where big data can be used in media and entertainment include consumer care, advertising, content monetization, and data journalism. Figure 6 shows some use cases of big data analytics in M&E industry [17]. Common applications of big data in the M&E industry include the following [18-20]:

➤ *Customer Care*: For a media and entertainment company, nobody is more important for them than its users and keeping its users happy is their toughest task. Companies need to ensure that they fulfill each of their customer's wishes. To attain the same, they must be aware of what their customers need. Big data analytics are used to recommend users shows or movies according to their preferences. Companies can even gain in-depth details about the other important things such as viewing history, ratings, reviews, data from social media, etc. Big data analytics is helping companies to connect with their customers in a much better way than they ever did.

➤ *Advertising:* Running a media and entertainment company without advertising is like winking at a person in the dark; you know what you are doing, but nobody else does. Advertising to media and entertainment companies is what food is to soul. A media company's success heavily depends on their advertising strategies. Big data helps media companies target advertisements more effectively by understanding consumer demographics, interests, and online behavior. Big data analysis helps companies to develop more personalized ads and provides insights about the best time to stream those ads to seek the attention of the maximum number of customers. As big data has made it possible for the media houses to understand their customer's exact preferences; it is quite easy for them to engross the customers.

➤ *Content Monetization:* Like all businesses, media companies aim to maximize revenue, minimize costs, and improve decision-making and business processes. As consumer interests shift from analog to digital media, there are substantial opportunities to monetize content and to identify

new products and services. Entertainment and media companies can use big data to understand what content, products and features consumers want. Product updates have become more cost effective and time effective, thanks to the analysis of customer data. You will never know what features your consumers want or need you to release if you do not dive into the data; it could give you a competitive edge, increase revenue, and brand loyalty.

➤ *Data Analysis:* Media and entertainment companies generate and collect data from a variety of sources. They need to analyze data not only at the customer and product levels, but also at network and infrastructure levels. Audience data breaks down into three broad categories: personal data, demographic data, and behavioral data. Key technologies in the coming years will be descriptive analytics, more sophisticated customer relationship management solutions, and lastly data visualization solutions that are accessible to a wide range of users in the enterprise. It is only by "humanizing" these tools that big data will be able to deliver the benefits that data-driven businesses increasingly demand.

➤ *Data Journalism:* Journalism incorporated big data into its practices in a way that influenced the internal logic of the profession. In the last 30 years, digital technologies with the introduction of various tools have made journalistic work easier. However, they have also made journalist work more difficult, because they have overwhelmed journalists with more information than can be handled by their investigative toolboxes. Data journalism emerges as a result of these changes, and it is related to data-driven journalism. Specifically, the introduction of information and communications technology (ICT) and the availability big data have turned data journalism into its current form. Data journalism promotes open journalism and open data. The term open data is related to transparency, accountability, accessibility, and free, public, and recyclable use.

➤ *Audience Participation:* Although audience participation has always been part of the journalism practice, the diffusion of Web 2.0 tools along with the socio-economic circumstances have led to the proliferation of user-generated content and increased users' involvement in the news production process. Audience participation in news production can be enabled by data journalism projects as well. Thanks to the participation of victims and

witnesses, a number of media organizations in Latin America have revealed situations involving huge breaches of human rights not identified in official records. Figure 7 shows an example of audience participation [21].

➢ *Data Mining:* While it is clearly evident that there are multitudes of potential applications for big data in the media industry, the fact remains that these datasets are so large and complex that in practice they are particularly unwieldy. Data mining is defined as a logical procedure used in order to search through very big amounts of data, with the purpose of discovering new, non-trivial information, which can subsequently be used to arrive at previously unknown conclusions. By this definition, it is immediately obvious that data mining and big data go hand-in-hand when it comes to journalistic practices. As stated before, the nature of big data renders them inaccessible to being processed by humans, or even by simple software, because of various factors that make them hard to understand and compute.

BENEFITS

Integrating big data and analytics into the media and entertainment industry is laying the foundation of a new era. As these technologies continue to evolve, they are set to unlock more possibilities for creating engaging and innovative entertainment experiences. Big data in the media industry can yield three different types of insights: diagnostic, predictive, and prescriptive. By using big data analytics, entertainment companies have been able to gain detailed insights regarding not only their customers but also their systems and processes. Other benefits of big data in the M&E industry include the following [17,22]:

➢ *Personalized Content:* By analyzing viewer's viewing habits, preferences, and interactions, these technologies will enable entertainment platforms to design content that resonates with individual tastes.

➢ *Informed Decision-making:* Data analytics presents deep insights into viewer preferences, demographics, and behaviors, thereby equipping content creators and distributors with valuable insights for informed decision-making. This data is instrumental in shaping content strategies, scheduling releases, and guiding the creation of new content as per audience demands.

➢ *Predictive Analysis:* This helps anticipate future trends and viewer preferences, enabling the sector to stay ahead of the curve. By predicting potential successes, media and entertainment studios can

make informed decisions about different projects, thus reducing the financial risks associated with content production.

➢ *Prescriptive Insights:* These are recommendations to make a business decision or take action in a certain way. Prescriptive insights are also the domain of artificial intelligence, requiring the AI application to translate its forecasts into actionable recommendations that support strategic business objectives. Media companies who leverage AI to analyze their big data can generate more accurate and higher-quality insights than those who stick to manual methods

➢ *Data Management:* Data management, including aggregation and normalization, is a time-consuming task for media companies who still depend on manual processes. Enterprise AI solutions with multi-platform integration features can streamline and automate the process of aggregating data from multiple sources into a centralized repository and normalizing the data to prepare it for analytics.

➢ *Targeted Marketing:* With detailed audience insights, organizations can create highly targeted marketing campaigns to resonate with specific viewer segments. This level of targeting will help enhance viewer engagement while ensuring a higher return on investment for marketing campaigns.

➢ *Advertising:* The primary factor determining a company's market worth and profitability is still advertising. Advertisements seem to be a natural component of any entertainment industry. This assists the businesses in acting as retargeting agents so that advertising will appear if people are watching a show or a movie with a connection to the items. Big data applications assist in analyzing user behavior and what they are likely to buy through targeted advertisements.

➢ *Content Optimization:* This is the ongoing practice of maximizing revenue from content distribution and licensing agreements. Enterprise AI solutions can help media companies forecast audience demand and assess the revenue potential of various types of content. As a result, media companies can make better strategic decisions about licensing or producing content and distribution teams can leverage insights to maximize licensing terms.

➢ *Revenue:* Another crucial aspect where big data has proven to be invaluable is revenue. It analyzes what customers want, what is in vogue in the market, the target audience's viewing history, etc.

to offer recommendations about generating better sales, adapting marketing strategies, fine-tune the when and where of content delivery, etc. All these measures seek to ensure high customer engagement rates and thus, better revenue for the company. Media organizations realized that by studying content consumption data, they can extract useful information which may help in designing successful publishing strategies and lead to new revenue opportunities.

CHALLENGES

In spite of its benefits, integrating big data and analytics in entertainment software development comes with challenges, including data privacy concerns, managing vast data volumes, and ensuring data accuracy. Most media firms are already doing some big data analytics, but the technical challenges of efficiently pooling data from multiple sources and extracting insights may be preventing them from using their data to its full potential. Attracting and keeping customers engaged are the biggest challenges faced by media and entertainment companies across the world. Big data as a phenomena is still, by their very nature, hard to access and work. Other challenges of big data in the M&E industry include the following [22]:

➤ *Siloed Data:* Big data in the media industry may be collected from subscribers, generated internally and stored in a database, provided by distribution partners, or sourced from a third-party organization. These multiple sources of data live in separate systems (data silos), and this segmentation means they cannot easily be integrated for analytics applications.

➤ *Non-standardized Formats:* Big data in the entertainment industry comes from multiple sources and may be in a variety of different formats. A media firm that licenses content to ten different distributors might receive ten different profit-sharing reports, each with their own fields and format. The lack of a standardized format means that this data must be normalized before it can be analyzed effectively.

➤ *Manual Data Processes*: While some media companies are using complex algorithms to process big data, others are still dependent on manual processes for data aggregation, normalization, analytics, and reporting. But as big data continues to grow, manual processes become more time-consuming, insights are delayed, and the overall impact and value of big data diminishes.

➤ *Ethical Considerations:* It is crucial for media companies to use big data responsibly and ethically, ensuring that data is used to benefit users and not exploit them.

➤ *Infrastructure:* While start-ups and SMEs can operate efficiently with open source and cloud infrastructure, for larger, older players, updating legacy IT infrastructure is a challenge. Legacy products and standards still need to be supported in the transition to big data ways of thinking and working. Failure to transform the culture and skillset of staff could impact companies who are profitable today but cannot adapt to data-driven business models.

➤ *Consumer Awareness:* There is increased consumer awareness and concern about how personal data is being used. There is regulatory uncertainty for European businesses that handle personal data, which potentially puts them at a disadvantage compared to, say, US companies who operate within a much more relaxed legal landscape.

CONCLUSION

Media and entertainment industry is advancing at an unprecedented, governed by dual requirements to minimize costs while at the same time generating more revenues from a highly competitive and uncertain market. The media and entertainment sector is in many respects an early adopter of big data technologies, but much more evolution has to happen for the full potential to be realized. The sector is experiencing a significant transformation with the integration of big data and analytics. Big data and analytics are key drivers for industry growth, extending beyond content personalization to strategic decision-making and significantly influencing the entertainment sector's overall success.

With the influence of big data and analytics in media and entertainment on the rise, it is becoming evident that these technologies are pivotal to the industry's future trajectory. The future of entertainment is here, and it is decidedly data-driven. More information about big data in the media and entertainment industry can be found in the books in [23-25] and the following related journal: *Journal of Big Data*

REFERENCES

[1] "Big data in media and entertainment," https://www.qubole.com/big-data-in-media-and-entertainment

[2] 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.

[3] A. Slamecka, "Big data explosion," April 2022, https://blogs.cisco.com/financialservices/big-data-explosion

[4] "Passion points: Analytics in the sports, media & entertainment industries," February 2017, https://gradblog.schulich.yorku.ca/event/passion-points-analytics-in-the-sports-media-entertainment-industries/

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

[6] 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

[7] 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.

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

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

[10] 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.

[11] J. Moorthy et al., "Big data: Prospects and challenges," *The Journal for Decision Makers*, vol. 40, no. 1, 2015, pp. 74–96. https://www.grandviewresearch.com/industry-analysis/industrial-wireless-sensor-networks-iwsn-market

[12] A. K. Tiwari, H. Chaudhary, and S. Yadav, "A review on big data and its security," *Proceedings of IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems*, 2015.

[13] M. B. Hoy, "Big data: An introduction for librarians," *Medical Reference Services Quarterly*, vol. 33, no 3. 2014, pp. 320-326.

[14] M. Viceconti, P. Hunter, and R. Hose, "Big data, big knowledge: Big data for personalized healthcare," *IEEE Journal of Medical and Health Informatics*, vol. 19, no. 4, July 2015, pp. 1209-1215.

[15] R. Williamson, "Media and entertainment: How this industry is impacted by big data," January 2021, https://www.datasciencecentral.com/media-and-entertainment-how-this-industry-is-impacted-by-big-data/

[16] A. Schroer, "Big data in media & entertainment: 15 examples to know," August 2024, https://builtin.com/articles/big-data-media

[17] "Data analytics in media and entertainment (M&E) industry," https://www.sganalytics.com/blog/data-analytics-in-media-and-entertainment-industry/

[18] "Big data in media and entertainment – The new hero in industry," https://data-flair.training/blogs/big-data-in-media-and-entertainment/

[19] "5 Ways big data plays a major role in the media and entertainment industry," July 2018, https://www.maropost.com/blog/5-ways-big-data-plays-a-major-role-in-the-media-and-entertainment-industry/

[20] A. Veglis et al., "Applications of big data in media organizations," *Social Sciences,* vol. 11, no. 9. 2022.

[21] "Data-driven decisions: How big data and analytics are shaping the future of media & entertainment," April 2024, https://www.sprinterra.com/data-driven-decisions-how-big-data-and-analytics-are-shaping-the-future-of-media-entertainment/

[22] "5 AI-Driven use cases for big data in media & entertainment," October 2022, https://www.symphonyai.com/resources/blog/media/5-ai-driven-use-cases-for-big-data-in-media-entertainment/

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

[24] P. C. K. Hung, *Big Data Applications and Use Cases (International Series on Computer, Entertainment and Media Technology)*. Springer, 2016.

[25] T. Hennig-Thurau and M. B. Houston, *Entertainment Science: Data Analytics and Practical Theory for Movies, Games, Books, and Music*. Springer, 2018.

Figure 1 Different components of big data [3].
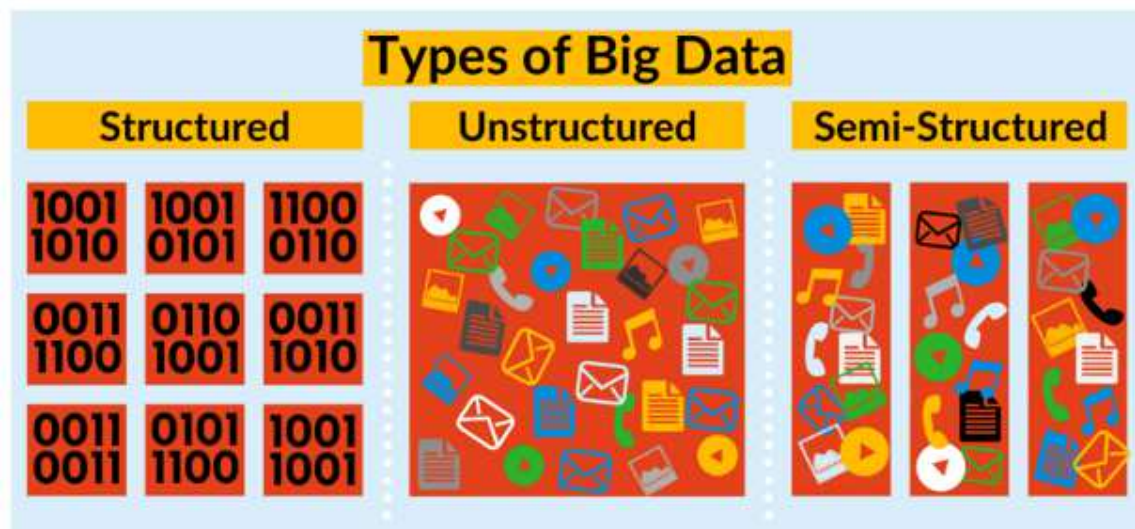

Figure 2 The cloud word for big data [4].
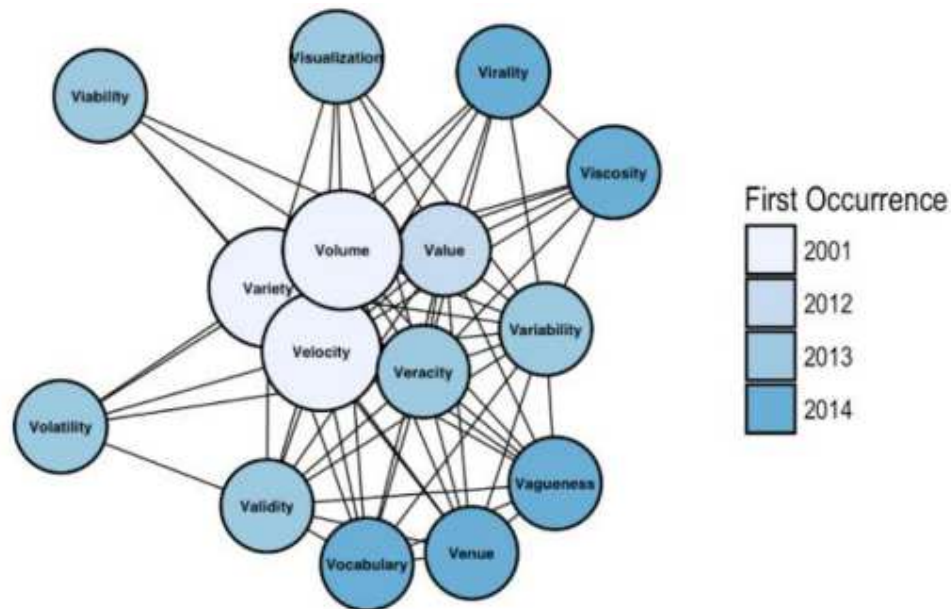

Figure 3 Types of big data [6].

Figure 4 The 42 V's of big data [9].

Figure 5 A representation of the media and entertainment industry [15].
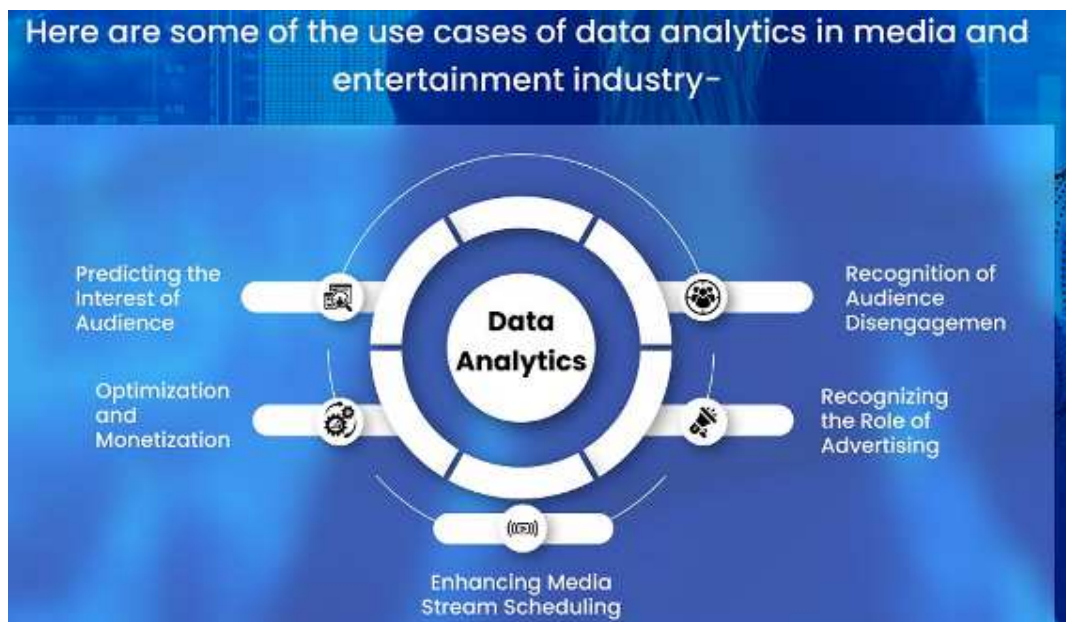

Figure 6 Some use cases of big data analytics in M&E industry [17].

Figure 7 An example of audience participation [21].