International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Big Data Methods and Analytics for Text, Audio, Video, and Social Media Data Concepts: Methods and Analytics for Text, Audio, Video, and Social Media Data

Shikha Tiwari[1], Dr. Preeti Bala[2], Somi Goel[3]

[1,2]Assistant Professor IMS Ghaziabad
[3]Student, BCA 3rd Year Ghaziabad

ABSTRACT

This paper provides a unified description of big data by integrating ideas from academics and practitioners. The primary focus of the paper is the analytical methods applied to big data. The focus this study places on analytics for unstructured data—which accounts for 95% of big data—makes it particularly noteworthy. The massive volumes of data produced by various machine-driven automated systems, including statistical, text, audio, video, sensor, and biometric data, are collectively referred to as "big data." In this essay, we consider the dimensions of various types of large data while addressing their challenges, applications, and problems. In this section, we'll cover social media data analytics along with content-based, text, audio, and video analytics.

Keywords: Big data analytics, Big data definition Unstructured data analytics, Predictive analytics

1. INTRODUCTION

Big data is the term used to describe the vast amount of structured and unstructured data that is currently available and will likely continue to grow over time. In summary, the sheer volume and complexity of this data prevents any conventional data management system from being able to store or analyze it efficiently. Big data is important because it offers numerous benefits for improved operational efficiency, which can lead to more precise outcomes and even the possibility of streamlining business procedures. Big data is extremely important for businesses and society as a whole. The data came from several sources, such as social media posts and shares, video, audio, and movie files, sensors that were used to record climate data, and more. Movies, videos, and audio files can all be uploaded or posted on social networking sites with this variety. We call this collection of data "BIG Data."

These days, big data is being used in a variety of ways to grow businesses and use outside intelligence in decision-making.

In most enterprise scenarios, there is data that needs more processing power than is currently available, moves too quickly, or is too big. Using big data, businesses may be able to improve operations and make decisions more quickly and intelligently.

Large data sets are typically ones that are larger than what can be handled by frequently used software tools in terms of capturing, organizing, managing, and processing in a reasonable amount of time. Big data is a collection of methods and tools that need new ways to be integrated to extract significant hidden values from enormous, diverse, complex, and scaled datasets. Wal-Mart is in charge of over one million customer interactions each hour. Facebook processes 40 billion images from its user population. To process massive amounts of data effectively, big data requires certain technologies. It makes use of various technologies, including time series analytics, data fusion and integration, genetic algorithms, machine learning, signal processing, simulation, natural language processing, and visualization. The structure of this paper is as follows. We define big data at the outset of the paper. We draw attention to the fact that big data has multiple dimensions, of which size is just one. When characterizing big data, other attributes like the rate of data generation are just as crucial. We then go into more detail about the different kinds of big data, including social media, text, audio, and video. We approach the big data conversation through the lens of analytics. As a result, when we talk about data in video format, we mostly talk about the techniques and resources for doing so.

1.1. Characteristics of Big Data:

Volume: The term "big data" itself refers to a very large quantity. The size of the data is a critical factor in extracting value from it. Furthermore, the volume of data affects whether a given set of data is truly Big Data or not. Therefore, one factor that must be taken into account when working with big data is "volume." [10, 12, 13, 15, 16].

Velocity: Data generation speed is referred to by the term "velocity." The real potential in the data is determined by how quickly it is generated and processed to meet demands. Big Data velocity is the rate at which information moves through various sources, including sensors, mobile devices, networks, social media sites, application logs, business processes, and so on. The flow of data is massive and continuous.

Variety: This includes both structured and unstructured data, as well as a range of sources. Spreadsheets and databases were the only data sources that the majority of applications took into consideration in the past. These days, analysis apps also take into account data from emails, pictures, videos, monitoring devices, PDFs, audio files, and other sources. The diversity of unstructured data presents challenges for data mining, storage, and analysis.

Variability: Data flows can exhibit significant irregularities with recurring peaks, in addition to the growing speeds and types of data. Peak data loads that are daily, seasonal, or event-triggered can be difficult to handle. especially when dealing with unstructured data.

Complexity: There are several sources of data in today's world. Furthermore, linking, matching, cleaning, and transforming data across systems remains a challenge. But, to prevent your data from rapidly getting out of control, you must connect and correlate relationships, hierarchies, and multiple data linkages.

Value: It covers how we can improve business and lifestyle through the use of this big data. We are aware that various business and social application kinds produce various kinds of data. One of the main problems is still finding values from big data in their application areas.

2. BIG DATA ANALYTICS

The process of analyzing vast and diverse data sets, or big data, to find information that can assist businesses in making wise decisions is known as big data analytics. This information can include hidden patterns, unidentified correlations, market trends, and consumer preferences.

Broadly speaking, data analytics tools and methods offer a way to examine data sets and derive conclusions from them that support businesses in making wise business decisions. Basic inquiries concerning the performance and operations of businesses are addressed by business intelligence (BI)

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

queries. Advanced analytics, or big data analytics, is a type that uses high-performance analytics systems to power complex applications with features like statistical algorithms, what-if analysis, and predictive models. Big data analytics applications enable big data analysts, data scientists, predictive modelers, statisticians, and other analytics professionals to analyze growing volumes of structured transaction data, plus other forms of data that are often left untapped by conventional BI and analytics programs. A variety of semi-structured and unstructured data types are included in this, such as clickstream data from the internet, web server logs, social media posts, text from emails and survey responses from customers, mobile phone records, and machine data obtained from sensors linked to the internet of things (IOT). By using appropriate analytics organizations can increase sales, increase customer service, and can improve operations. Predictive Analytics allow organizations to make better and faster decisions.

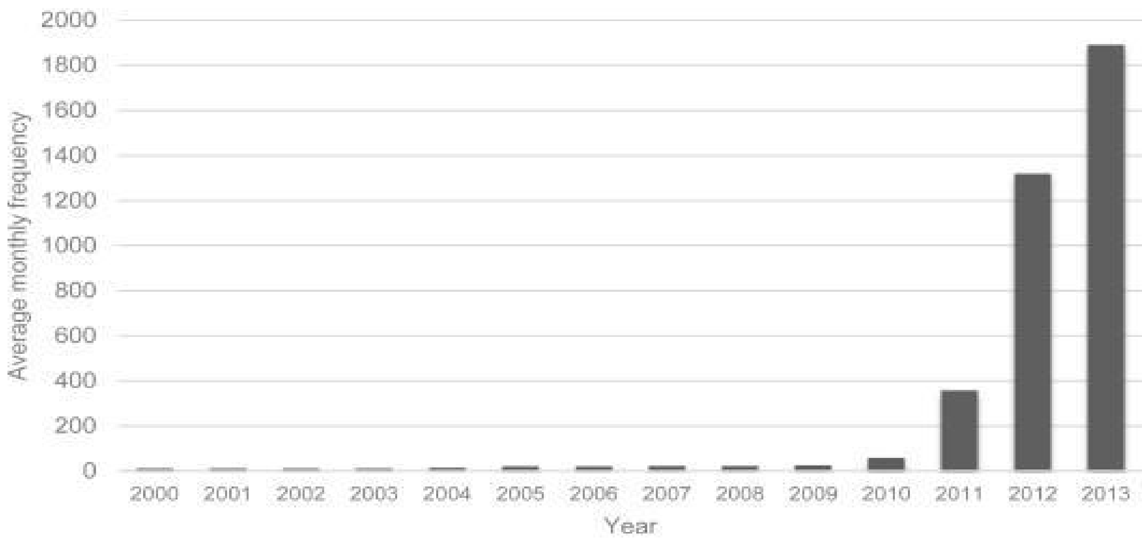


Fig. 1. Frequency distribution of documents containing the term "big data" in ProQuest Research Library.

2.1 Big Data Analytics usage in India

Big Data is being used in India in many innovative ways, ranging from predicting train ticket confirmations to looking for leaks in the water supply to even helping find the ideal bride and groom. Here are a few applications of big data analytics that have been used in India recently.

1. Win elections (exit poll).
2. Finding a perfect match.
3. Detecting water leakages.
4. Gaining insights into shopping behavior.
5. Ensuring proper water supply.
6. Improve India's financial inclusion ratio.
7. Improve product development.
8. Predict ticket confirmations for trains.

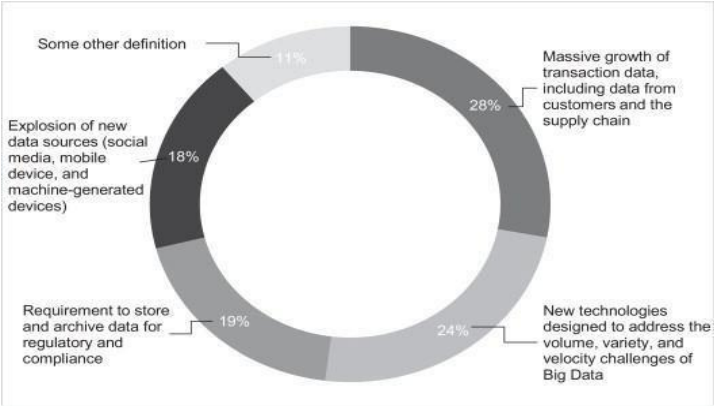


Fig. 2. Definitions of big data based on an online survey of 154 global executives in April 2012.

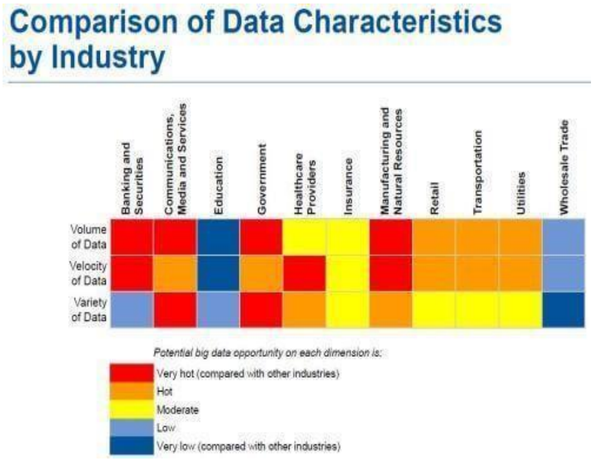


Fig.3 Comparison of data Characteristics

3. Predictive Analytics

Predictive analytics is a subfield of advanced analytics that involves forecasting future events that are not yet known. Numerous methods are used in predictive analytics.

from artificial intelligence, machine learning, modeling, statistics, data mining, and machine learning to examine current data and forecast future events.

It combines management, information technology, and business process modeling with a variety of data mining, predictive modeling, and analytical techniques to forecast future events. Patterns in transactional and historical data can be used to pinpoint future opportunities and risks. With a specific set of conditions, predictive analytics models evaluate risk by capturing relationships between numerous factors and allocating a score, or weightage. Businesses can effectively interpret big data for their benefit by applying predictive analytics.

Business users can generate predictive intelligence by identifying patterns and relationships in both structured and unstructured data through the use of data mining, text analytics, and statistics. Structured data, such as age, gender, marital status, income, and sales figures, are easily analyzed. Unstructured data are textual records found in social media posts, call center notes, and other open text that must have the sentiment taken out of them before being utilized in the model-building process. With the help of predictive analytics, businesses can become more proactive and forward-thinking, projecting results and actions based on data rather than conjecture or gut feeling. Prescriptive analytics goes a step further by offering options for decisions that will help users take advantage of the predictions and their implications as well as actions to take.

Fig.4. Predictive analytics process

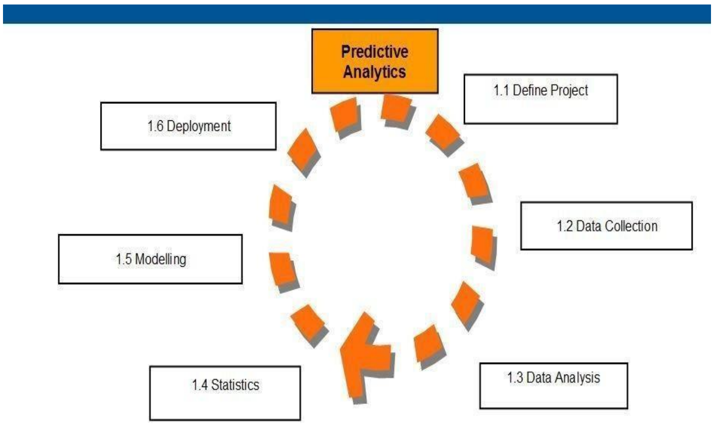
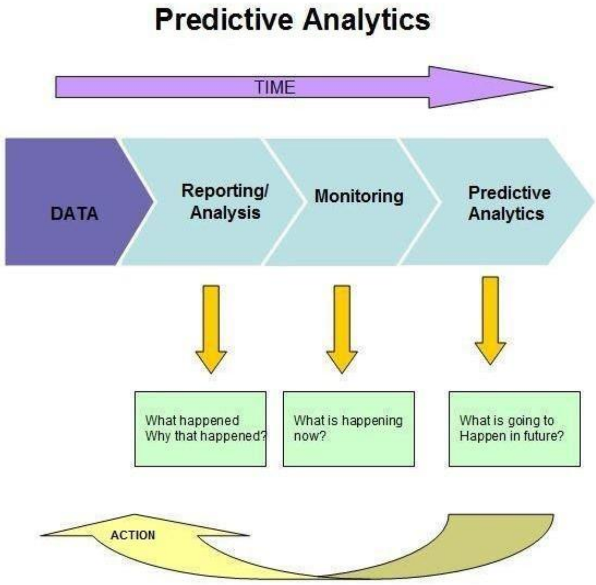


Fig.5. Predictive analytics value chain



Marketing campaigns, sales, customer service, and other CRM (customer relationship management) domains can all benefit from predictive analytics. It is their goal to efficiently analyze in-demand products and forecast the purchasing patterns of their clientele.

Clinical Decision Support: Diabetes, asthma, lifetime illnesses, and other conditions can be predicted for a patient using predictive analytics.

Collection analytics: By determining the best collection agencies, contact tactics, etc. for each customer, predictive analytics assists financial institutions in allocating resources for collection.

Cross-selling: Predictive analytics can assist a business that sells a variety of products by analyzing the spending and behavior of its customers. This has the potential to stimulate cross-selling or the sale of additional goods to existing clients.

Retention of Customers: Given the quantity of with the rise of competing services, companies must constantly prioritize keeping customers happy, rewarding devoted patrons, and reducing the number of lost customers. By regularly analyzing consumer usage, spending, and behavior patterns, predictive analytics, when used appropriately, can result in an active retention strategy.

Direct marketing: Keeping up with rival goods and customer trends can be difficult when promoting goods and services to consumers. Predictive analytics is useful not only for prospect identification but also for determining the best product versions, advertising, communication channels, and timing to target a particular customer.

Fraud detection: Fraud can take many different forms and affects a lot of businesses. Some examples include false insurance claims, fraudulent transactions, erroneous credit applications, and identity theft. These issues afflict businesses of all sizes across numerous industries. Credit card issuers, insurance providers, retailers, manufacturers, suppliers to other businesses, and even service providers are a few instances of potential victims. Predictive analysis can help to identify high-risk fraud candidates in business or the public sector.Predictive analytics can be used to solve problems related to portfolio, product, or economy-level forecasting by utilizing time series techniques. These issues can also be resolved through machine learning techniques, which convert the initial time series into a feature vector space and use an algorithm to identify patterns with predictive ability.

Risk management: The goal of using risk management strategies is to anticipate and profit from potential future events. Organizations and commercial enterprises can identify future risks, such as natural disasters and their effects, with the use of predictive analysis. Risk management enables them to accept the right choice at the right moment.

Underwriting: Because they provide a variety of services, many companies have to account for risk exposure and calculate the necessary costs to cover the risk. For instance, auto insurance companies must precisely calculate the premium amount required to insure each vehicle and driver. Predictive analytics can be used by a health insurance provider to estimate an enrollee's future costs by analyzing a few years' worth of medical claims data as well as lab, pharmacy, and other records, if they are available. By projecting the likelihood of disease, default, bankruptcy, etc., predictive analytics can assist in underwriting these quantities. By employing application level data to forecast a customer's future risk behavior, predictive analytics can expedite the customer acquisition process.

Underwriting: Because of their diverse service offerings, many companies must assess the risk they pose and calculate the necessary expenditure to mitigate it. For each vehicle and driver, for instance, auto insurance companies must precisely calculate the premium amount to be charged. Predictive analytics can help a health insurance provider estimate how costly an enrollee will likely be in the future by analyzing a few years' worth of historical medical claims data along with lab, pharmacy, and other records if they are available. Through the prediction of illness, default, bankruptcy, and other events, predictive analytics can assist in underwriting these quantities. Using application-level data, predictive analytics can forecast a customer's future risk behavior, which can expedite the customer acquisition process.

4. SOCIAL MEDIA ANALYTICS

With the rise of Web 2.0 technologies and the popularity of the Internet, big data analytics has become a significant area of research. Furthermore, social media applications' widespread use and adoption have presented researchers and practitioners with several opportunities as well as difficulties. Users of social media platforms generate enormous amounts of data as a result of integrating their daily activities and background information. Recently, a great deal of research has been done on "big data," which refers to this massive amount of generated data. To get a broad overview of the social media big data analytics research topic, a review of recent works is presented. We categorize the literature according to significant features. This study also compares possible big data analytics techniques and their quality attributes. Moreover, we provide a discussion on the applications of social media big data analytics by highlighting the state-of-the-art techniques, methods, and the quality attributes of various studies. Open research challenges in big data analytics are described as well. These Days Social media is the greatest channel for quickly and easily leveraging social media for product and business marketing, as well as for understanding the choices and sentiments of customers in real time. For search, user recommendations, and merchandising, eBay.com leverages a 40PB Hadoop cluster in addition to two data warehouses with 7.5 petabytes and 40PB in capacity. within the 90PB data warehouse of eBay. Every day, millions of back-end operations and inquiries from more than 500,000 third-party sellers are handled by Amazon.com. Linux is the fundamental technology that powers Amazon, and as of 2005, the company boasted the three largest Linux databases in the world, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB. 50 billion photos are managed by Facebook from its user base. Google was processing about 100 billion searches a month as of August 2012.

Even though social network research dates back to the early 1920s, social media analytics is a relatively new field that arose in the early 2000s with the introduction of Web 2.0. The data-centric nature of contemporary social media analytics is one of its main features. Numerous academic fields, including psychology, sociology, anthropology, computer science, mathematics, physics, and economics, have contributed to the field of social media analytics research. The main use of social media analytics in recent years has been in marketing. This is explained by the fact that social media is being widely and increasingly embraced by consumers around the globe. According to Forrester Research, Inc., social media will grow at the second-fastest rate among marketing channels in the US between 2011 and 2016.

4.1. Application areas

1. Behavior Analytics
2. Location-based interaction Analytics
3. Recommender systems development
4. Link prediction
5. Customer interaction and Analytics & marketing
6. Media use
7. Security
8. Social studies

4.2. Challenges of Social Media Analytics

1. Massive amounts of data require lots of storage space and processing power.
2. Shifting social media platforms.
3. Worldwide online accessibility provides more data in many languages.
4. Evolution of online language.

5. CONTENT-BASED ANALYTICS

Content-based analytics refers to any type of data stored on the back-end website of social media. Facebook users, for instance, store their videos, photos, and data on Facebook storage. Social networking sites like Facebook, Twitter, and WhatsApp need to increase their storage capacity daily due to the rapidly growing number of users. The problem is that they are unsure of how much more storage capacity they need to add. To generate recommendations, content-based predictive analytics recommender systems primarily compare features (tagged keywords) between related items and the user's profile. Items with features that correspond with the original item's features will be suggested to the user when they buy an item with tagged features. There's a greater chance the user will find the recommendation appealing if more features match. Precision refers to this probability degree. [4,6,13] However, user-based tagging highlights additional issues with collaborative filtering and content-based filtering systems, such as:

1. Credibility: Users with a limited rating history may distort the data, and not all customers, particularly those who make purchases online, are trustworthy. Furthermore, some vendors might promote others to give favorable ratings to their products while devaluing those of their rivals.
2. Scarcity: Not every item will be rated, or there won't be enough ratings to generate meaningful information.
3. Inconsistency: Although the meaning of an item may be the same, different users may not tag it with the same keywords. Furthermore, some characteristics may be arbitrary. For instance, a movie may be deemed too long by one viewer and too short by another.

Precision with constant feedback

Whenever feasible, get feedback from users to help the system make more accurate recommendations. There are numerous methods and channels available for gathering client feedback. After a purchase, some businesses request that customers rate a product or service. Some systems allow users to ―like or ―dislike a product by providing links similar to social media.

Measurement for the effectiveness of system recommendations

The degree to which a system satisfies the following two requirements determines how successful its recommendations are: recall (see it as a set of potential matches, often bigger than perfect matches) and precision (think of it as a set of perfect matches, generally a tiny set). Problems with effectiveness measurement:

1. Precision quantifies the accuracy of the system's suggestion. Because precision may be arbitrary and hard to measure, it is challenging to quantify.
2. While some suggestions could be related to the client's interests, the buyer could still decide not to purchase. The best proof that a suggestion is accurate comes from observable facts: the buyer purchases the product. As an alternative, the user may be specifically asked to score the system's recommendations.
3. Recall quantifies the range of plausible, insightful suggestions your system generates. Recall might be thought of as a list of potential ideas, not all of which are ideal suggestions. Recall and accuracy typically have an inverse relationship. In other words, accuracy decreases when recall increases and vice versa.

A system with both high recall and precision would be ideal. But in practical terms, finding a fine balance between the two is the wisest course of action. The emphasis you place on recall or precision largely relies on the issue you're trying to address.

6. TEXT ANALYTICS

Text analytics, also known as text mining, involves utilizing methods to derive insights from written data. Organizations store various forms of textual data such as social media posts, emails, blogs, and customer feedback. By employing statistical analysis, computational linguistics, and machine learning, text analytics helps in transforming large amounts of text into valuable summaries that aid in making informed decisions. For example, text analytics can be utilized to forecast stock market trends by analyzing financial

news. Below, we provide a concise overview of text analytics techniques.
Information extraction (IE) methods are utilized to extract organized data from unstructured text. For instance, IE algorithms can retrieve organized details like drug names, dosages, and frequencies from medical prescriptions. Two key components of IE include Entity Recognition (ER) and Relation Extraction (RE). ER identifies names in the text and categorizes them into specific groups such as person, date, location, and organization. On the other hand, RE identifies and extracts semantic connections between entities (e.g., individuals, organizations, medications, genes, etc.) in the text. For instance, when presented with the sentence "Steve Jobs co-founded Apple Inc. in 1976", an RE system can extract relationships like Founder Of [Steve Jobs, Apple Inc.] or Founded In [Apple Inc., 1976].
Text summarization techniques automatically generate a concise summary of one or more documents, conveying the main information from the original text(s). These techniques are used in various applications such as scientific papers, news articles, advertisements, emails, and blogs. There are two main approaches to summarization: extractive and abstractive. Extractive summarization involves creating a summary by selecting and combining original text units, typically sentences, to form a subset of the original document. This approach focuses on identifying important text units based on their location and frequency in the text. Extractive techniques do not require a deep understanding of the text. On the other hand, abstractive summarization involves extracting semantic information from the text to create summaries that may include text units not present in the original text. This approach utilizes advanced Natural Language Processing (NLP) techniques to analyze and generate summaries. Abstractive systems generally produce more coherent summaries compared to extractive systems, although the latter are more straightforward to implement, especially for large datasets.

Steps for Text Analytics system:

a) Text: In initial stage data is unstructured.
b) Text processing: All information will transfer in Semantic Syntactic text.
c) Text transformation: In it important text will extract for future use.
d) Feature selection: In it data is counted and display in Statistics format.
e) Data mining: All data is classified and clustered.

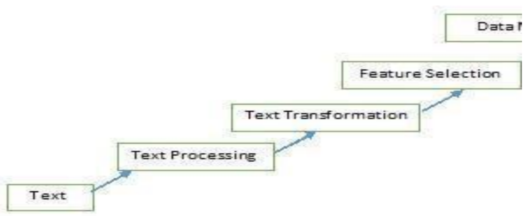


Fig.; Steps for Text Analytics system

Text Analytics applications area:-

1. The security application is designed to monitor and analyze various online platforms such as internet blogs, news websites, and social media sites for national security. Its primary function is to detect and identify any unethical activities or content on the internet.
2. Through the analysis of text data, the marketing application enables us to determine the preferences of customers and identify the type of products they are most likely to favor.
3. In survey research, companies often pose open-ended questions to customers, asking for their opinions, pros and cons about certain products, or suggestions. Text analytics is necessary for analyzing this

type of data and extracting valuable insights.

4. Utilizing big data analytics, the automatic process for emails and messages allows for the efficient filtering of large volumes of incoming emails based on specific terms or keywords. This technology is also beneficial when it comes to automatically redirecting messages or emails to the appropriate department or section.

Distinct Aspects of Text in Social Media:

1. The real-time nature of social media services is a crucial aspect. With the constant evolution of content and communication styles, the text also changes. As the textual data becomes time-sensitive, people's thoughts also shift accordingly.
2. Efficient processing of short texts is vital for text analytics. Due to the brevity of messages, individuals can engage more effectively on social networking platforms. Short messages, comprising only a few phrases or sentences, are commonly utilized in social media.
3. The quality of content differs significantly between text in social media and traditional media. Various individuals share diverse content based on their knowledge, ideas, and thoughts. When crafting messages, numerous new abbreviations and acronyms are employed, such as "How r u?" or "Gr8," which, although not actual words, are widely used in social media.

Applying Text Analytics to Social Media:

A multitude of questions and answers flood social networking websites as people seek to address the inquiries posed by their peers. The act of tagging data has experienced a significant surge, exemplified by instances where users search for recent events such as the "Bihar Election."

In response, the system retrieves results that have been labeled with tags like "Bihar" or "Election." The textual data found within social media platforms not only offers a wealth of information but also presents users with a diverse and distinctive range of content in the form of comments, posts, and tags.

7. AUDIO ANALYTICS:

Audio analytics is a method used to analyze and extract information from unstructured audio data. When applied to human spoken language, it is commonly referred to as speech analytics. The terms audio analytics and speech analytics are often used interchangeably since they are primarily used for analyzing spoken audio. Currently, the main areas where audio analytics is applied are customer call centers and healthcare.

In call centers, audio analytics is utilized to efficiently analyze large volumes of recorded calls, which can amount to thousands or even millions of hours. These techniques are beneficial in various ways, such as improving customer experience, evaluating agent performance, increasing sales turnover rates, monitoring compliance with policies (such as privacy and security), gaining insights into customer behavior, and identifying issues related to products or services. Audio analytics systems can also be designed to analyze live calls, allowing for real-time analysis and decision-making.

1. Event Detection involves monitoring various data sources to identify specific events that occur within them, such as images, videos, audios, and text documents.
2. With the rise of social networking websites, Collaborative Question Answering services have become more prevalent. These services offer expert recommendations for cross-selling and up-selling based on customer interactions, providing real-time feedback to agents. Additionally, Interactive Voice Response (IVR) platforms are utilized by automated call centers to recognize and assist frustrated callers.

Audio analytics plays a crucial role in healthcare by aiding in the diagnosis and treatment of medical conditions that impact a patient's communication patterns. Additionally, it can be utilized to analyze an infant's cries, which provide valuable information about their health and emotional well-being. The abundance of data collected through speech-driven clinical documentation systems further drives the integration of audio analytics in the healthcare industry. There are two primary technological approaches in speech analytics: the transcript-based approach, also known as large-vocabulary continuous speech recognition (LVCSR), and the phonetic-based approach. Let's delve into each of these approaches.

LVCSR systems follow a two-phase process: indexing and searching. During the indexing phase, the system aims to transcribe the speech content of the audio. This is achieved through automatic speech recognition (ASR) algorithms that match sounds to words. The words are then identified based on a predefined dictionary. If the system fails to find an exact match in the dictionary, it provides the closest alternative. The output of this phase is a searchable index file that contains information about the sequence of words spoken in the speech. In the searching phase, standard text-based methods are employed to locate the desired search term within the index file.

On the other hand, phonetic-based systems work with sounds or phonemes. Phonemes are distinct units of sound in a specific language that differentiate one word from another. Phonetic-based systems also consist of two phases: phonetic indexing and searching. In the indexing phase, the system translates the input speech into a sequence of phonemes, as opposed to LVCSR systems that convert speech into a sequence of words. In the searching phase, the system scans the output of the first phase for the phonetic representation of the search terms.

Application Area of Audio Analytics:

The audio file format is utilized for transferring data from one location to another. Audio analytics is employed to verify if the audio data provided is in the correct format or a similar format to what was sent by the sender. There are various applications of audio analytics:

a. Surveillance Application: This application involves the systematic selection of audio classes to detect crimes within society. Utilizing an audio analytics framework in surveillance applications is crucial for identifying suspicious activities. Moreover, this application can be used to promptly relay important information to surveillance teams during crisis situations.
b. Threat Detection: Audio analytics can help in identifying threats that occur between the sender and receiver.
c. Tele-monitoring System: Modern technology includes cameras with audio recording capabilities. Audio analytics can effectively detect sounds such as screams, breaking glass, gunshots, explosions, and calls for help. Integrating audio analytics with video analytics in a single monitoring system enhances threat detection efficiency.
d. The mobile networking system facilitates communication and data transfer between different locations. In cases where network issues arise and audio quality is compromised, Audio Analytics can be utilized to identify and address information that was not transmitted correctly due to such problems.

8. VIDEO ANALYTICS

Video analytics, also referred to as video content analysis (VCA), encompasses a range of techniques for monitoring, analyzing, and extracting valuable information from video streams. While video analytics is still in its early stages compared to other forms of data mining, several methods have already been developed to process both real-time and pre-recorded videos. The proliferation of closed-circuit television

(CCTV) cameras and the surging popularity of video-sharing platforms are the primary drivers behind the advancement of computerized video analysis. However, a significant hurdle lies in the vast amount of video data. To put this into perspective, a single second of high-definition video is equivalent to more than 2000 pages of text in terms of size. Now, consider the fact that 100 hours of video are uploaded to YouTube every minute (YouTube Statistics, n.d.).

Big data technologies have transformed this challenge into an opportunity by eliminating the need for costly and risky manual processing. Leveraging big data technologies enables automatic sifting through and extraction of intelligence from extensive video footage. Consequently, big data technology stands as the third contributing factor to the advancement of video analytics.

Version 1:

In recent years, video analytics have primarily been utilized in automated security and surveillance systems. Unlike labor-intensive surveillance systems, which are costly and less efficient, video analytics can effectively carry out surveillance tasks such as detecting breaches in restricted areas, identifying abandoned or removed objects, recognizing suspicious activities, and detecting camera tampering. When a threat is detected, the surveillance system can promptly alert security personnel or initiate an automatic response.

The data captured by CCTV cameras in retail stores can be utilized for business intelligence purposes, particularly in marketing and operations management. By employing intelligent algorithms, retailers can gather demographic data about customers, including age, gender, and ethnicity. Additionally, retailers can track customer traffic, analyze movement patterns, measure dwell time in different areas, and monitor queues in real time. By correlating this data with customer demographics, valuable insights can be derived to inform decisions related to product placement, pricing, assortment optimization, promotional strategies, cross-selling, store layout, and staffing. An additional possible use of video analytics in the retail sector involves analyzing the purchasing habits of groups. When families shop together, often only one person interacts with the cashier, resulting in traditional systems overlooking data on the shopping behaviors of other family members. Video analytics can assist retailers in seizing this overlooked opportunity by offering insights into the group's size, demographics, and the individual members' purchasing habits.

Application of video analytics

a) CCTV cameras are extremely useful in accident cases as they allow us to determine exactly what occurred at the time of the accident. Additionally, they serve as a valuable tool for security purposes and monitoring parking areas.

b) In various settings such as schools and traffic police departments, it is essential to index multimedia content for easy search and retrieval. Video indexing can be accomplished by utilizing different levels of information present in the video, including metadata, soundtracks, transcripts, and visual content. Relational database management systems (RDBMS) are commonly employed for video search and retrieval using a metadata-based approach. Furthermore, audio analytics and text analytics techniques can be applied to index videos based on their associated soundtracks and transcripts, respectively.

When it comes to the system architecture, there are two approaches to video analytics: server-based and edge-based.

In a server-based architecture, the video captured by each camera is transmitted back to a centralized and dedicated server for video analytics. To accommodate bandwidth limitations, the video is often compressed by reducing frame rates and image resolution. However, this compression may result in a loss of information, potentially affecting the accuracy of the analysis. Nevertheless, the server-based approach offers economies of scale and facilitates easier maintenance.

On the other hand, an edge-based architecture involves performing analytics at the "edge" of the system. This means that the video analytics are conducted locally and directly on the raw data captured by the camera. Consequently, the entire video stream is available for analysis, enabling a more effective content analysis. However, edge-based systems tend to be more expensive to maintain and have lower processing power compared to server-based systems.

c) Video analytics algorithms are utilized to analyze videos, a task that poses challenges and is extremely time-consuming for human operators, especially when dealing with a large amount of data. Video analytics enables us to search for specific videos when needed.

d) Video analytics for Business Intelligence involves extracting statistical and operational data. Instead of having operators manually review all videos to count people or cars in a certain area, or determine the most common traffic routes, video analytics can automate these processes.

e) Target and Scene Analytics are essential components of Video Analytics for Business Intelligence. Target Analytics provides detailed information about target movements, patterns, appearances, and other characteristics that can aid in target identification.

f) Direction Analytics involves assigning specific values (ranging from low to high) to areas within a camera's field of view in order to distinguish behaviors.

g) By automating processes, video analytics eliminates the need for human monitoring for extended periods of time. This automation allows human judgment to be inserted at the most critical moments in the surveillance process.

The importance of statistical significance, the obstacles related to computational efficiency, and the distinct features of big data mentioned earlier emphasize the necessity of creating novel statistical methods for extracting valuable insights from predictive models.

Concluding future aspects

The purpose of this study is to review, analyze, and discuss extensive data. We have provided various definitions of big data, emphasizing that size is just one aspect of it. Other aspects, such as velocity and variety, are equally significant. Our main focus has been on analytics to extract valid and valuable insights from big data. We emphasize that predictive analytics, which mainly deals with structured data, tends to overshadow other forms of analytics used for unstructured data, which accounts for 95% of big data. We have examined analytics techniques for text, audio, video, and social media data, as well as predictive analytics. The study argues for developing new statistical techniques for big data to address the unique characteristics that differentiate it from smaller data sets. Most statistical methods currently in use have been designed for smaller data sets consisting of samples.

Advancements in storage and computing technologies have made it possible to capture the informational value of big data in a cost-effective and timely manner. As a result, there has been a significant increase in the real-world application of analytics that were previously not economically viable for large-scale use before the era of big data. For instance, sentiment analysis (opinion mining) has been around since the early 2000s. However, big data technologies have allowed businesses to implement sentiment analysis to extract valuable insights from millions of opinions shared on social media. The analysis of unstructured text driven by the vast amount of social media data is creating business value by utilizing traditional sentiment analysis techniques (pre-big data) that may not be best suited to leverage big data.

REFERENCES

1. Web content is available on the link:- https://www.guru99.com/what-is-big-data html#4
2. Web content is available on the link:- https://searchbusinessanalytics.techtarget. com/definition/big-data-analytics?amp=1
3. Web content is available on the link:- https://www.google.com/amp/s/www.predictiveanalyticstoday.com Web content available on the link https://www.google.com/amp/s/www.dqin dindia.com/8-innovative-examples-of-big-data-usage-in-India/amp/
4. Web content is available on the link:—https://www.statsoft.com/textbook/text- mining
5. Web content available on the link:-https://www.sciencedirect.com/science/article/pii/so7475632183041X
6. **Big data analytics meets social media: A systematic review of techniques, open issues, and future directions** Author Elkin L.S., Topal K., Bebek G.
7. **A new big data approach for topic classification and sentiment analysis of Twitter data** Author Chiplunkar A.
8. Benchmarking big data architectures for social networks data processing using public cloud platforms Author Persico V., Pescapé A., Picariello A., Sperlí G.
9. **Data Mining: Concepts and Techniques** Author: Jiawei Han, Micheline Kamber, Jian Pei
10. **Big data analytics meets social media: A systematic review of techniques, open issues, and future directions** Author Elkin L.S., Topal K., Bebek G.