Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776568

RESEARCH AND DESIGN BIG DATA TECHNIQUES FOR SOCIAL NETWORKS: A VARIETY

Raghupathi K and Dr. Kailash Patidar
Department of Computer Science Engineering , Dr. A. P. J. Abdul Kalam University, Indore (M.P.), India 452016
Corresponding Author : - Raghupathi K

ABSTRACT
Social media sites like Facebook, Twitter, and others have contributed to a dramatic increase in the volume of data travelling over the internet. Information can be gathered from a variety of sources, including satellite-based weather reports, social media posts, and digital photographs and videos. Data of this magnitude lends itself to scientific collection and analysis. Still, one could call this a "big data" set. What we mean when we talk about "big data" is a swath of information that is large, complex, and contains both structured and unstructured forms of information. A variety of sources provide data, including weather sensors, social media, digital photographs and videos, and more. Big data refers to these massive amounts of information. Big data naturally includes information gathered from social networks. This means that the methods used in big data analysis can also be applied to the study of data gleaned from social networks. The goal of this paper is to introduce readers to a variety of techniques used in the study of social network data using big data analysis. We also outline some solvable issues for further study in this area.
KEYWORDS: Social networks, Big data, Data mining, Hadoop, MapReduce

1. INTRODUCTION

A social network consists of a finite number of distinct individuals. These entities could be individuals, businesses, or even entire neighborhoods. They're related to each other, thus there's some bond between them. The rise in popularity of social networks can be attributed to the fact that they permit researchers to investigate not just social actors but also their social connections. Consequently, the study of social networks may be traced back to many parts of basic social science. Systematically analyzing network information can reveal useful social and organizational trends. Changes that should be made to the social structure and to individuals' roles within it are also discussed. Connected networks serve as the body of society's most important organs. Social network analysis has become more important in the field of social and behavioral sciences as researchers come to appreciate the fundamental role that networks play in contemporary society. The uniqueness and importance of a social network cannot be overstated, and this is why relational data is so important. Social network analysis is concerned with discovering patterns in interconnections. It has seen extensive usage in the business world for purposes like as spotting fraud and tracking down key players in a network. Analyzing the connections between people is useful in many fields, including epidemiology, computer networking, and the study of emergent behavior in both physical and biological systems. One of the main responsibilities of a network analyst is to create a model of these interconnections that accurately represents the make-up of a specific group. This framework allows us to examine how the group as a whole functions and how its members are affected. Because of this, the social network perspective provides a fresh point of view, one that allows for the inclusion of a wide range of possible structure types, including but not limited to behavioral, social, and political.

The scientific examination of data acquired from social networks may disclose several significant behaviors of the components involved and social trends, providing insight into suitable modifications to the social structure and the roles of persons within it. There is a wealth of data suggesting that social network analysis is a useful tool for understanding how communities act and how they affect one another's well-being. The study of social networks provides a theoretical and systematic framework for thinking about human relationships. According to Freeman's reasoning, social network analysis techniques allow for forensic claims regarding social characteristics and processes. As a result, social network analysis's core ideas and procedures may be readily altered to serve a wide range of purposes. It provides a formal, conceptual framework for societal thought. It assumes that links between interacting pieces are essential. In order to advance methodologies for social network analysis, it is essential to recognize that the unit of study in network analysis is not the individual but rather an entity consisting of a collection many individuals, each of whom is related to a few, some, or many others. We focus on analytical problems that aren't the norm. To make sense of the data, which is considerably different from the data often utilized by social and behavioral scientists, social network analysis is applied. To social network analysts, the importance of people's and communities' interconnection cannot be overstated. Relationships between the units can be measured with structural variables. Theories that employ network notions can be recognized by their unique assertions concerning the relationships between persons and groups. The essential tenet of these ideas is that individual units are not acting independently of one another but are instead impacted by one another. When such structural factors are included, data analysis and model building are subject to different possibilities and limits. The interconnections between social entities or actors provide a framework or structure in social network analysis, which is used to examine and analyze social systems apart from the individual behaviors, attitudes, and beliefs that make up such systems. It is possible to uncover hidden patterns by examining the ties that bind individuals together. Conventional methods that focus on researching individual social actors' features have been shown to be much less successful than social network analysis. Social network analysis places a higher value on the connections between nodes in a network than on any one node's features. Many phenomena in the real world, including the "small world phenomenon," have been seen and understood by employing social network research techniques. Over the past few decades, information has expanded enormously due to the rapid development of IT. The web, the utmost repository of data, has become a scalability problem due to the proliferation of user-generated content such as blogs, tweets, social-network interactions, and photographs; servers, which generate activity logs constantly; scientists, which collect measurement data about the world in which we live. As a result, experts in the social media industry have turned to big data analysis techniques.

BIG DATA ANALYSIS

The term "big data" was coined to describe the proliferation of unstructured and structured data sets that exceed the capacity of conventional database and software tools. Companies can learn unprecedented amounts about their operations with the help of big data, which can then be used to make better decisions and boost productivity. The value of big data lies in our ability to make good use of it; data of this nature can be collected from virtually any source, and it can be analyzed in any number of ways to facilitate cost savings, time savings, the creation of new products, the improvement of existing products, and the making of sound decisions. The concept of "big data" first appeared infrequently in the early 1990s, but its significance and prevalence quickly grew. Many businesses today consider big data to be fundamental to their data strategy. The benefits and drawbacks of big data initiatives can be better comprehended by understanding the unique characteristics and properties of big data. First, we'll take a look at some of the salient features of big data sets, and then we'll move on to describing analysis methods. The properties used to describe these data sets have been in place since their inception and begin with the letter V in the English alphabet. Three Vs was the initial count, and now that figure is steadily climbing. Below, we will discuss 7 of these Vs that define big data.

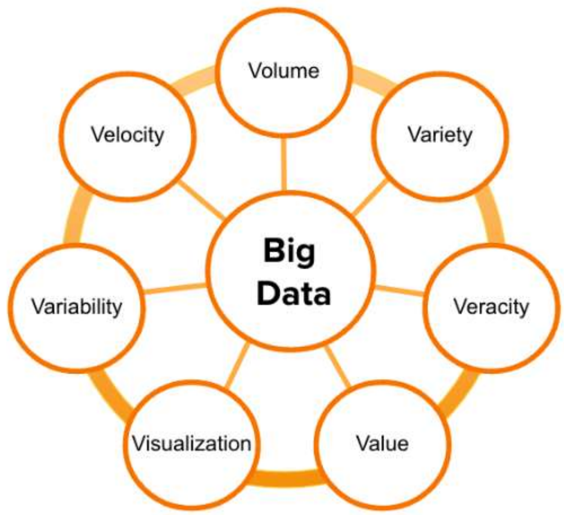

Figure 1. 7'vs of Big Data

1. Volume

Since more than 90% of all data in existence today was produced within the past two years, it is not surprising that the sheer volume of big data is its defining characteristic. The sheer magnitude of modern knowledge is often overwhelming. Take a peek at these few examples: Every minute, quite so much as 300 hours of new video are uploaded to YouTube. There were around 1.1 trillion images taken in 2016, and projections for 2017 put that number at 9 percent growth. It is estimated that by 2017, the overall number of images stored will have increased to $4.7 trillion from 2016's $3.9 trillion. This is due in large part to the fact that people tend to store the same image in multiple places, including on their personal devices, in cloud storage, and on various social media platforms. In 2016, experts estimated that monthly mobile traffic globally will amount to 6.2 Exabytes. That's the equivalent of 6.2 billion gigabytes, for those keeping score at home.

2. Velocity

The rate of data creation, production, or updating is what we mean when we talk about data velocity. The fact that Facebook's data warehouse can hold over 300 petabytes of information may sound impressive, but it's important to remember that new information is being created at an exponential rate. According to Facebook, 600 terabytes of new data is uploaded every day. More than 3.5 billion searches occur every day, with Google handling an estimated 40,000 queries at any given moment.

3. Variety

When dealing with big data, we need to be able to process not only structured data, but also semi-structured and largely unstructured data. The preceding examples may lead you to believe that the vast majority of big data consists of unstructured information; however, in addition to audio, image, and video files, social media updates, and other text formats, there are also log files, click data, machine and sensor data, etc.

4. Variability

Different things can be meant by "variability" when discussing big data. Consider the fact that there are numerous discrepancies in the information. Anomaly and outlier detection techniques are required to locate these before useful analytics can be performed. Due to the many different data dimensions that emerge from using a wide variety of data sources, big data can also be quite unpredictable. The rate at which large datasets are loaded into your database can also be considered variable.

5. Veracity

Unfortunately, that's just how big data works. The data's veracity (its ability to inspire trust) decreases as the aforementioned characteristics increase. Similar to validity and volatility, but not the same as (see below). Data veracity is more concerned with the legitimacy of the data itself, its context, and the significance of any inferences drawn from the data. Think of database containing information about restaurant sales and prices over the course of five years, for instance. Perhaps you are wondering: Who made up the foundation on which everything is built? How did they go about gathering the information? Only some kinds of eateries or restaurants were considered? The information was summarized by the data creators? Is there any indication that the data has been altered by any third party? If we want to know if this information is reliable, we need to know the answers to these questions. Having this information allows us to assess the potential dangers of basing our analyses and business decisions on this data set without first verifying its accuracy.

6. Validity

Data validity is the degree to which it is reliable and appropriate for its intended purpose, much like data veracity. It is estimated that up to 60% of a data scientist's time is spent on data cleaning before any analysis can be performed (source: Forbes). Adopting good data governance practices to ensure consistent data quality, common definitions, and metadata is essential to maximizing the value of big data analytics.

7. Value

Value is the final consideration, but it is also one of the most crucial ones. If you can't use the information to improve your business, none of the other benefits of big data matter. Big data has substantial value because it can help you learn more about your customers, refine your advertising strategies to reach them, streamline internal operations, and boost the efficiency of your machines and business. Before launching a big data strategy, you should be aware of its potential and the challenges it presents.

BIG DATA APPLICATIONS

The use of big data tools and processing has effectively revolutionized numerous fields. Social media analysis, the advertising industry, and the management of large computer systems are just few of the areas that have profited from this. There are many other fields where big data has a good chance of succeeding, however:

- **Healthcare** - Big data has the potential to significantly impact the Healthcare sector in a variety of ways. Although the potential for advancement in this field is boundless, it has been explored in some specific areas, such as error checking and anomaly detection in medical research datasets.
- **Governance** - By expanding the idea of making comprehensive user models, governments are striving to become more people-oriented. A direct result of this will be tailored citizen-government interactions. Municipal officials hope to prevent problems before they happen by using analytics, models, and simulations.
- **Smart Cities** - The next stage is to combine these systems after they have automated many crucial city-wide activities and numerous sensors have been put wherever practicable. Complex models are being built that incorporate real-time data from the several systems and sensors with simulations and humans-in-the-loop in order to ease city-wide operations and reduce knock-on effects caused by faults in one or more systems.
- **Industry 4.0** - Systems powered by analytics and simulation will be the next big thing in industrial activity because of their ability to adapt swiftly to issues and variances in the production environment. Intelligent Manufacturing Systems (IMS) are being created with the use of virtual, fractal, bionic, and holonic production methods with the intention of improving production management. Decision-making models inside the system consider not only the factory floor, but the whole supply chain as a whole.
- **The Internet of Things (IoT)** - Integration of Internet functionality into a wide variety of products and sensors will drive the IoT revolution by enabling them to record and transmit data for analysis. Arguably, the Internet of Things will serve as the technological engine that propels the next generation of smart cities and Industry 4.0 applications.

To keep up with the needs of these industries, future big data applications will need to include machine learning to a greater and greater extent, to help generate logical conclusions from the quantity of accessible data. Many big data applications have historically relied on data collection, storage, and processing in batches. However, real-time processing of data is essential for applications in the Internet of Things (IoT), the healthcare industry, and smart cities, which need immediate insights from data streams. In addition, there will be a need for queries to existing data sources to return results in under one second.

2. LITERATURE REVIEW

Ebrahim Mahdipour et al (2021), Social networking sites (SNSs) facilitate global communication by allowing users to share and explore the work of others, as well as provide and receive feedback in the form of ratings and comments. The 5 V's of big data velocity, volume, value, breadth, and veracity apply equally to online communities. Therefore, large data analytic frameworks and tools (SNA) are commonly used by SNA. The meteoric rise in popularity of social networks has sparked a surge of interest in the study of social data, namely its description and discovery of communication patterns among users and insight into their activities. In this study, we present a comprehensive analysis of big data analytic methodologies in social networks, based on a search of the literature from 2013 through August 2020, which yields 74 papers, and explain how big data analytics may be used to social media. Major journals/conferences, yearly distributions, and a variety of publishers are discussed in this paper's results. There are also two primary schools of thought when it comes to analyzing big data: content-oriented and network-oriented. We also discuss the fundamental ideas, evaluation standards, tools, methods, outcomes, and challenges associated with assessment. We conclude with a discussion of open questions and potential future research directions.

Dr.S.Matilda et al (2018), Two cutting-edge factors, big data and social media, have catapulted IT to its current height. Data mining workloads are a part of big data, and the results of these workloads are of importance to business executives and analysts across all industries. These burdens might be taken from many different places. The flood of data produced by social media is being held up as a source of insights into human nature. Today, it's more important than ever to mine actionable intelligence from this deluge of data, which comes at you at varying rates of volume, variety, and velocity and must be evaluated in the context of your company's requirements. As a result, it is crucial to create innovative systems for processing and interpreting massive data, as well as the necessary tools and procedures. The convergence of big data and social media offers a new window of opportunity to push management theory and practice forward.

Sheela Gole et al (2015), population greatest in the world The impact of social media is such that the average internet user now spends nearly 2.5 hours per day liking, chatting, poking, and tweeting on social media, which has become a vast source of unstructured data because of the number of times the Facebook 'Like' button is pressed every day across the web, revealing what people care about. Traditional databases and architecture have a hard time modifying, grilling, and then structuring enormous data, yet this data may lead to numerous consumer insights that can aid in the development of win-win circumstances. Large data sets must be mined for insights in order to demonstrate causality, interdependence, and enable prediction of events and behavior. The 5 Vs of Big Data are volume, velocity, variety, trustworthiness, and value. Using data mining methods, tools, and the Hadoop architecture, this article discusses the 5Vs, features, problems, and future of Big Data in the social media arena.

Md. Saifur Rahman et al (2022), With the advent of web 2.0 and social media, the internet has become a global platform for communication. In today's world, people can communicate with each other, with businesses, and with governments in unprecedented ways. People are often comfortable voicing their thoughts, feelings, and ideas aloud. Because of this, we can put machine learning methods and social data analytics to use to make "Big Social Data" more useful. This research surveys contemporary literature in the fields of social media, data science, and machine learning to provide a holistic picture of big data analytics in the social media sphere. In this article, we'll discuss how social media data may contribute to better data-driven decision-making.

S. Arista et al (2016), Science, technology, engineering, medicine, healthcare, economics, and even society itself are all being altered by the advent of Big Data. The potential of Big Data is widely recognized and praised. Today, Big Data is seen as a potential gold mine by many businesses. Researchers in Information Sciences, as well as policy and decision makers in governments and businesses, have all taken an interest in Big Data. However, with the advent of Big Data comes a new scientific paradigm: data-intensive scientific discovery (DISD). Key characteristics of Big Data, Data Analytics applications, and enabling technology are discussed, along with a summary of the most important background material. Improving corporate intelligence and following current market trends are two of the main impetuses behind this publication, which aims to promote the sharing of cutting-edge research in the burgeoning disciplines of Big Data and Analytics.

Ajith Abraham et al (2018), The proliferation of social media over the past several years has led to a daily uptick in the volume of information being generated. In addition, social network mining is complicated by the wide range of difficulties that might arise and the inherent interconnectedness of people's social circles. Such complicated and large data sets are beyond the capabilities of traditional algorithm software, calling for the creation of new analytic methodologies and tools. Topics related to social networks in big data analysis are covered in this comprehensive guide. Discussions of social networking theory, methods, and problems are included. This book covers a wide range of topics, including but not limited to neural networks, deep learning, AI, visualization, online education and healthcare, and computer security and intrusion detection in the workplace.

Bikash Chandra Pattanaik et al (2022), In the last decade, several sectors have flourished thanks to the development of new technology. These breakthroughs are improving the quality of life for people everywhere and increasing profits for companies like Netflix, Alibaba, Flipkart, etc. The majority of people in today's society either makes use of or are surrounded by smart devices. However, analytics-based industries and enterprises are improving the quality of life for everyone. Numerous businesses, for instance, get information on customers' theater, film, and music tastes. These recommendations are being implemented, and the related users are enjoying enhanced service as a result. They are the work of data scientists. Data science, on the other hand, is a cross-disciplinary field that uses scientific methods, processes, algorithms, and systems to mine both labelled and unlabeled data for useful information and insights. Data scientists also use Big Data Analytics, an analysis tool developed specifically for use in the field of data science. Hadoop and other tools are used among many others to analyze the massive data sets and draw conclusions or make predictions.

Zhiyong Zhang et al (2017), Multimedia social networks (MSNs) have proliferated and evolved over the past decade, expanding at an unprecedented rate to permeate every aspect of our daily lives, from pleasure to business. The flexibility of mobile Internet and mobile terminals allows users to access MSNs under any identity, regardless of location, role, or group membership. As a result, user behavior in relation to MSNs is getting increasingly rich and nuanced. The primary contribution of this study is an expansion and improvement of the SocialSitu scenario analytics framework, as well as the introduction of a unique algorithm for users' intention serialization analysis that builds on the principles of the classic Generalized Sequential Pattern (GSP).

Kumar Rahul et al (2020), When it comes to big data analytics (BDA), machine learning algorithms (MLAs) are a useful method or instrument. When it comes to analyzing massive amounts of data produced by an application, this instrument is ideal. The goal of machine learning algorithms is to discover useful information and data for business purposes. Big data analytics (BDA) includes this as one of its many offerings. Risk management, root-cause analysis, spotting fraudulent activity in procurement records, and a slew of other tasks may all benefit from the application of big data analytics (BDA). This article discusses the research conducted in this area to evaluate the significance of machine learning tools and techniques and determine the fields most suited for employing them. These fields include, but are not limited to, marketing, human resources, healthcare, insurance, banking, the automotive industry, and many more. Various difficulties associated with machine learning tools and technologies, as well as the present state of their acceptance in industries, are discussed in this study.

Ankit Didwania et al (2016), For communication between nodes in a mobile social network to take place, the network as a whole must be able to tolerate delays. The system relies on a store-carry-forward concept to function. We benefit greatly from the community aspect of the mobile social network since human beings are sociable creatures that thrive in close quarters. Human-carried gadgets may effectively communicate with one another thanks to the community structure provided by the internet. Our research has uncovered the best community identification strategies for use in a mobile social network. We have also performed a comprehensive evaluation of the available distributed community detection techniques for Social networks, taking into account such critical criteria as complexity and community type. With this kind of study, we can better understand the advantages and disadvantages of various algorithms now in use.

M. Usman Nisar et al (2014), Graphs are extremely valuable due of their adaptability and expressiveness. They are a useful metaphor for internet communities, search engines, and even genome sequencing. Graph pattern matching is a highly significant area with several practical applications. The goal, conceptually, is to identify sub graphs of a given graph that exhibit a specific pattern. Sub graph isomorphism and Regular Expression matching are only two of the many approaches that have been developed to this problem. With Big Data, scientists are constantly confronted with large graphs, which have greatly increased the difficulty of this field. Three different distributed algorithms for approximate graph pattern matching are analyzed for their speed improvements and communication patterns. We also investigate how various graph partitioning strategies affect execution time and network I/O. Our exhaustive analysis demonstrates that the algorithms have highly scalable behavior, and that min-cut partitioning can increase speed in some cases while also significantly lowering network traffic.

Ajmera Rajesh et al (2018), There has been a significant increase in recent years in the use of the Internet for long-distance social networking communication, social insurance, online business, bank transactions, and many other forms of administration. These web programs need a comfortable level of safety and protection to be usable. However, our computers are vulnerable to a wide variety of attacks. The availability of tools and traps for breaching and attacking networks is growing. Abnormal activities in social organizations refer to illegal activities that display different behaviors than those seen in other groups with a similar structure. This research explores the innovative ordering of anomalies as a function of a variety of characteristics. This document includes an overview of several methods for detecting and identifying anomalies, as well as the underlying assumptions and reasons for the prevalence of such discrepancies.

3. RESEARCH METHODOLOGY

The big data, as modern datasets are also known, deviates from their historical forebears in three key respects: volume, velocity, and variety. Today, information is produced rapidly and in vast quantities from a dizzying array of different sources. If this information is used wisely, we can finally enter the "information age." After performing smart processing and analytics on the data at hand, useful insights can be gleaned. This section discusses the methods (especially those associated with machine learning) used to collect, store, process, and analyse such massive quantities of data. We also make an effort to relate the concepts discussed and the examples used to illustrate them to humanitarian progress. The purpose of this section is to give readers some context for the techniques discussed in the context of humanitarian development by providing a brief history and a list of related works.

BIG DATA TECHNIQUES

There are two broad categories of data processing methods: batch processing and stream processing. Two distinct processing methods, shown as Batch processing and Stream processing in Figure 2, are described below. These two works focus on contrasting aspects of the same topic. Whereas batch processing utilizes previously collected data, real-time processing handles data that is either time-sensitive or can only be processed in the moment. One can classify processing frameworks into three types based on the various ways in which big data is processed. Hadoop, a framework designed for batch processing; Apache Strom, a framework designed for streaming data only; and hybrid frameworks (Apache Spark).

1) Apache Hadoop

When processing data in batches, Apache Hadoop is what's used. The first framework of its kind with such strength was open source. This facilitates the processing of enormous data sets. Hadoop uses the Map-Reduce framework to process this massive amount of data. Map-Reduce is based on the strategy of breaking down a problem into manageable chunks, or sub problems, before tackling each one individually. There are two main types of nodes that make up a Hadoop cluster: the Master and the worker nodes. The master node assigns tasks to the workers. When their tasks are complete, worker nodes communicate their findings to the controlling node, or master. The data transmitted to the master node from the worker nodes is compiled, and from there a solution to the problem is formed. The Apache Hadoop framework may be used for many different purposes. Hadoop is being used by some of the biggest names in business, like Amazon, Microsoft, and Google. Hadoop has many benefits, such as its suitability for large datasets, its ability to perform distributed data processing, its ability to easily handle partial failure, and its simple programming model. On the other hand, Hadoop also has many drawbacks, such as the fact that its processing becomes slow due to the amalgamate functioning of the master node, the existence of only one master node, the lack of transparency in the configuration of the nodes, and its restrictive nature.

2) Apache Storm

Big data has the basic 3 V's (volume, variety, and velocity), making stream processing a time-consuming task. While Hadoop is capable of handling these 3Vs, Apache Strom is needed when processing data in real-time, where velocity is of paramount importance. Hadoop's shortcomings include the fact that it is difficult to incorporate real-time data into its ecosystem, that parallel processing cannot be performed in real time, and that it is difficult to retrieve the results of processing in a timely manner due to the nature of Hadoop's architecture. Apache Storm is up to the task of handling any of the aforementioned difficulties. Data ingestion can be handled by Kafka, while the second challenge can be met by Strom, which retrieves the data from Kafka and performs parallel processing of data, and finally generates the correlation of large datasets. The primary goal is to provide the fastest and most reliable stream processing with minimal latency. Data sets of any size are no problem for it. Scheduling of Directed Acyclic Graphs is another area where this tool shines (DAGs). In terms of storm technology, there are three main components: streams, spouts, and bolts. The stream represents the massive amount of information that is continuously being generated and sent to the system from a wide variety of sources. The data is retrieved from the stream using technologies like Kafka and then made available to Bolt via Apache spout. The streams are used by the processing logic of Bolt, which performs actions on them and generates a new stream. Finally, the systems can take the Bolt output into account. Strom does not promise that messages will be delivered in any particular order. Each message is guaranteed to be processed at least once thanks to this feature. When things go wrong, however, you risk receiving duplicate messages.

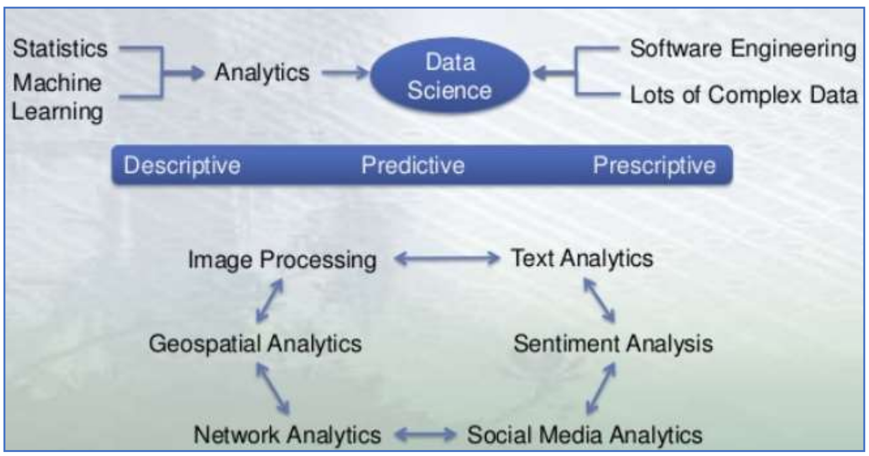

Figure 2. Big Data Analytics

There are currently three main types of Big Data tools: those that process data in batches, those that process data in real time, and those that allow for interactive analysis. Apache Hadoop is a well-known and capable batch-processing-based Big Data technology. Its architecture and platforms may be used to build customized Big Data applications.

Table 1: Batch Processing-Based Big Data Tools

Big Data Techniques	Usage	Benefit
Apache Hadoop	Platform, infrastructure	High scalability, dependability, and comprehensiveness

Dryad	Platform, infrastructure	Good programmability and a high-performance distributed execution engine
Apache Mahout	Business ML Algorithm	Excellent Maturity
Jasper soft BI Suite	Business Intelligences Software (BIS)	Scalable, cost-effective self-service BI
Business Analytic	Platform for Business Analytic	Knowledge discovery must be robust, scalable, and flexible.
Sky tree Server	ML advanced analytic	Process huge datasets accurately and quickly.
Tableau	Data Visualization and the Business analytics	Dashboards that are faster, smarter, more fit, and easier to use
Karma sphere Studio and Analyst	Big Data workspace	Standards and collaboration
Talend Open Studio	Data Management, application integrations	graphics environment based on Eclipse that is simple to use

Analytics for massive amounts of data are still in their infancy. Finding patterns is an art, but it's still important to have a system in place to evaluate your results and report accurately. With pattern-matching techniques, the scope of possible studies is practically limitless. Graphs are crucial to the functioning of modern social networks. This is due to the fact that graphs are a natural way to represent any kind of relationship, no matter how intricate. Examples of social network and media graphs are all around us, from our Face book friends and Twitter followers to our LinkedIn connections' recommendations and Amazon customers' product preferences. The context of a graph's application can have a significant impact on the characteristics it exhibits. While the Map Reduce (MR) paradigm has seen widespread adoption, it is not well suited to the iterative nature of many graph analytics tasks. Even with these advances, scalable graph analytics presents new challenges and avenues for investigation. More research is needed, especially on the topic of massively dispersed graphs, which are common in large data sets. Data and analytics are readily available; however, existing organizational systems rarely incorporate them. To fully reap the benefits of big data, we need to design efficient interfaces for integrating with existing computational models and frameworks.

Table 2: Tools for big data based on stream processing

Name	Usage	Benefits
Storm	System for real-time calculation	Scalable, fault-tolerant, and simple to set up and run
S4	Processing streams of data that never end	Platform that has been proven, distributed, scalable, fault-tolerant, and pluggable.

SQL Stream Server	The application of sensors, M2M, and telematics	Big Data platform with SQL-based real-time streaming
Splunk	Get machine data and use it.	Rapid and simple operation, dynamic situations, scaling from laptop to datacentre
Apache Kafka	System for sending messages to many people at once	unchangeable activity data stream with high throughput
SAP Hana	Platform for doing business in instantaneously	real-time analytics and rapid in-memory computing

Research into the siloed work practices that stunt the expansion of big data is an urgent priority. This would result in a consolidated view of big data analytics across all business functions. The administration of storage spaces and the mining of relevant data are two other areas where study is still in its infancy. The storage and processing needs of big data analytics make cloud computing an ideal host. Social media and business networking sites like Facebook, LinkedIn, Twitter, Amazon, eBay, and Google+ are all examples of data-heavy applications that have contributed to a rise in demand for cloud-based data storage and processing. The majority of data, according to Schouten's (2012) projections, will be stored in the cloud by 2016. Consequently, we can now consider cloud-based analytics a serious area of study. There is also a significant lack of study in the areas of security and privacy, web intelligence, Intelligent Management, and data modeling for MIS. While there has been significant progress in recent research activity across all aspects of big data, many challenges still need to be addressed.

Table 3: Performance: Time comparison for task completion

Task	**Execution Time (seconds), normalized values**	
	Apache Hadoop	**Apache Storm**
Word Count	0.36	0.42
Grep	0.04	0.73
Tera Sort	0.62	0.95
connected components	0.19	0.82
Page rank	0.16	1
k-means	0.13	0.25
Average	**0.216**	**0.883**

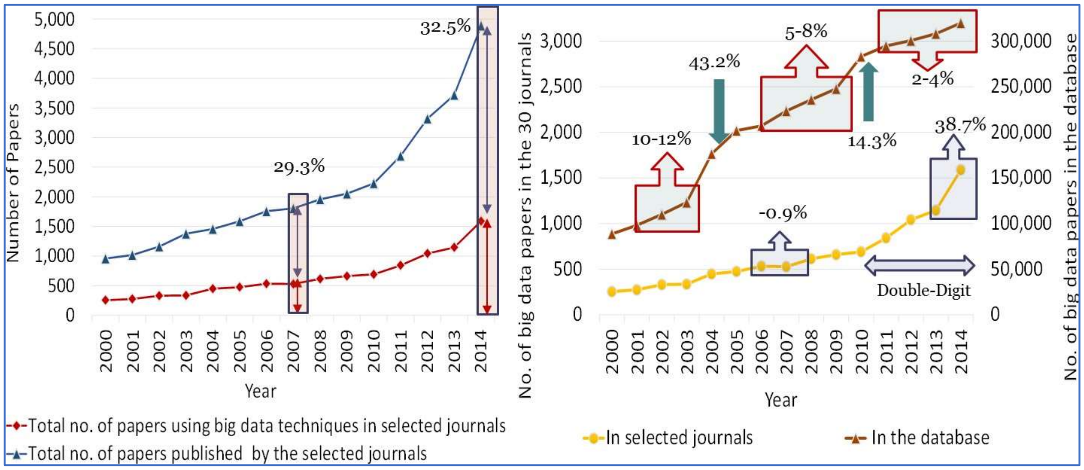

Figure 3 Comparisons between the Selected Journals and Database

4. CONCLUSION

Big data analytics and social media marketing are just getting started. The developing world is still very far behind, and it is not clear who is responsible for moving things along. While social media can be an effective supplement, slow progress is often indicative of a lack of technological infrastructure and data management and analysis experts. Companies can still make use of the data they collect if they adopt the proper mentality, develop the appropriate strategy, and implement the appropriate technology. Learning how to accurately quantify the social value of business activities is essential for businesses looking to expand their operations and better market their offerings. In addition, each article's central ideas, benefits, drawbacks, evaluation approaches, instruments, and evaluation parameters were dissected and discussed. While accuracy (20%), time (16%), and scalability (12%) were all taken into account during the evaluation process, privacy, reliability, and security measures were somewhat overlooked. If we look at the studies that were chosen and examine the tools that were used, We discover that while Python became the most widely used programming language, Hadoop was utilized significantly more frequently. According to the findings of this SLR, existing social big data analytic systems lack the capacity to ensure privacy and scalability and have experienced various outstanding difficulties, including such delay, real-time processing, and excessive feature selection run-time. There are still many open questions in the field of opinion and sentiment analysis, including how to deal with issues like domain dependency, a lack of available resource languages, the identification of sarcasm and slang, the identification of subjectivity, and the integration of data from multiple sources.

5. REFERENCES

[1]. Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662- 679.

[2]. Bertolucci,J., (2013). Big data Analytics: Descriptive vs. Predictive vs. Prescriptive, Big data Analytics, Information Week.

[3]. Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. McKinsey Quarterly, 4, 24-35

[4]. De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. In AIP Conference Proceedings (Vol. 1644, pp. 97-104).

[5]. Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data'revolution in healthcare. McKinsey Quarterly.

[6]. Beyer, M. A., & Laney, D. (2012). The importance of 'Big Data': A definition (Gartner Report G00235055). Retrieved from https://www.gartner.com/doc/2 057415?ref=clientFriendlyURL

[7]. Bradbury, D. (2013, June). Effective social media analytics. The Guardian. Retrieved from http://www.theguardian.com/technology/2013/jun/10/ effective-social-media-analytics

[8]. Einav, L. & Levin, J.D. (2013). The data revolution and economic analysis (NBER Working Paper no. 19035). Retrieved from http://www.nber.org/papers/ w19035

[9]. e-skills UK (2013). Big data analytics. An assessment of demand for labour and skills, 2012–2017 (E-skills UK report on behalf of SAS UK). Retrieved from https://www.e-skills.com/Documents/Research/General/BigDataAnalytics_ Report_Jan2013.pdf

[10]. Hoit, D. M. (2013). Big data, big expectations (Centre for Digital Education Report Q2 2013). Retrieved from http://www.centerdigitaled.com/paper/2013-Q2- Special-Report-Big-Data-Big-Expectations.html

[11]. Kalil, T. (2012). Big data is a big deal. Retrieved from http://www.whitehouse.gov/ blog/2012/03/29/big-data-big-deal

[12]. KDnuggets (2014). Online education in analytics, big data, data mining, and data science. Retrieved from http://www.kdnuggets.com/education/online.html

[13]. Lohr, S. (2012, February 11). The age of big data. The New York Times. Retrieved from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impactin-the world.html?pagewanted=all&_r=0.

[14]. Mayer-Schönberger, V. & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt.

[15]. Nerney, C. (2013). Universities Expanding Big Data Analytics Courses with IBM Aid. Retrieved from http://data-informed.com/universities-expanding-big-dataanalytics-courses-with-ibm-aid.

[16]. Sheela Gole, Bharat Tidke, "A survey of big data in social media using data mining techniques" (2015).

[17]. S. Arista,A comprehensive study on big-data analytics-tools, techniques, technologies and applications (2016)

[18]. Bikash Chandra Pattanaik, Snigdha Mohapatra, Sibani Samal, GYANENDRA KUMAR PALLAI, "Big Data Analytics and Data Science Emerging Trends and Challenges" (2022)
[19]. Zhiyong Zhang, "A Situational Analytic Method for User Behavior Pattern in Multimedia Social Networks" (2017)
[20]. Kumar Rahul, Rohitash Kumar Banyal, Puneet Goswami, and Vijay Kumar, "Machine Learning Algorithms for Big Data Analytics" (2020).