BIG DATA & SOCIAL MEDIA

[1] Mr.Pradip Ingle ,[2]Anita Patare,[3] Harshada Bankar,[4] Abhijit Sarnaik ,[5]Anil Dhasal

[1]Assistant Professor of Department of Information Technology, Anuradha Engineering College, Chikhli

[1]pradipingle2009@gmail.com ,[2]apatare777@gmail.com ,[3]harshadabankar89@gmail.com,[4]. Abhijitsarnaik6@gmail.com , [5] anildhasal156@gmail.com

ABSTRACT: Big Data and Social Media Social media platforms generate vast amounts of data characterized by high volume, variety, and velocity. This data serves as a rich source for big data analytics, enabling marketers and businesses to gain valuable insights. Key applications include personalization, predictive analytics, enhanced targeting, real-time marketing, and campaign optimization. Tools such as text analytics, sentiment analysis, and data mining are commonly used to process and interpret social media data. However, challenges such as privacy concerns and data quality must be addressed to ensure compliance and accuracy. Future trends point towards increased data collection from diverse sources, real-time analytics, and the integration of AI and machine learning to drive innovation in social media marketing. The interplay between big data and social media is crucial for businesses aiming to enhance their marketing strategies and improve user engagement.

Keywords: Big Data, Social Media, Data Analytics
Personalization, Predictive Analytics, Targeting Real-time Marketing, Data Mining.

INTRODUCTION:

Social media data comes from comments, posts, images, and videos. A report says that by 2023, 2.72 billion people will regularly watch videos on their smartphones due to better 4G and 5G networks. While social media trends change quickly, they still consistently produce a lot of digital data. Big data helps target users based on their interests and personalizes ads. People are grouped by age, gender, interests, and location. Those who live nearby and share similar online habits are likely to like the same products. However, there are issues with how big data is used in marketing. For instance, in 2018, Netflix was accused of racism for showing different movie posters to users based on their gender and ethnicity. Additionally, big data helps media measure their audience. It can even affect what they choose to publish. For example, the Huffington Post uses a system to track visits, comments, and other user activities, and it creates reports based on this data .The Huffington Post's system checks how well headlines attract readers and finds the best ways to share content with different users. It discovered that parents often read articles on their phones late at night after putting their kids to bed, so it posts parenting content at those times. Big data have become a valuable resource. The utilization of big data is everywhere, starting from social networks, academia, healthcare, aerospace, transport planning, oil and gas development to telecoms, e-commerce, finance and insurance, military and surveillance, and a variety of other fields [1].However, this massive amount of data will become an asset only if we know how to make data talk. Data analytics is that tool that makes data convey stories in a clear and accessible manner.

A well-known social listening platform—Brand watch, says that over 3.4 billion people worldwide were using social media sites in May 2019[2].The reason for this is that a social media based platform allows for faster information sharing, supports text, image, audio, and video sharing, and allows the individual user to communicate with a huge number of other users at the same time. Surprisingly, social media's popularity has recently risen to the point that many use it as their principal communication route to report to the public or emergency personnel [5]. Social media has become the most effective means of networking and hobby of any human being. With the proliferation of smart devices and services extended by social networks, it is gaining importance and attracts people in large numbers. The interaction between human beings via social media is a potential source of unstructured, finer-grained and larger-scale digital data. The term "Big data" is pervasive and includes huge quantities of data, social media analytics, next generation data management capabilities, real time data and much more. It is also the shorthand for advancing trends in technology that open the door to a new approach to understand the latest trends and making intelligent decisions (Schroeck, Shockley, Smart, Morales & Tufano, 2014 [3].

Big Data

Business architecture and marketing strategy. The Big data is the term which refers to a collection of data so large and complex that cannot be processed using existing database management tools or traditional data processing applications. The procedures and processes for capture, curation, storage, search, sharing, transfer, analysis, and visualization are different for big data (Fonseca & Boutaba, 2015). The term was coined to describe the exploding growth of data observed in the order of zetabytes, that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions (Stawski, 2015).This data comes from everywhere: sensors used to gather information, posts and responses in social media sites, multimedia contents, online purchase transaction, and cell phone signals to name a few. The click of a mouse, a keystroke or a single touch on the mobile screen contributes to the big data. Therefore in the digital space, consumers, suppliers and organizations are creating and consuming vast amount of information. Gartner predicts that enterprise data in all forms will grow 650 percent over the next five years. IDC enterprise data growth stat envisages that the world's volume of data doubles every 18 months. This flood of big data is a golden goose for business leaders over which they can build their real challenge lies in harnessing this data for various applications using advanced tools and techniques [3].

Technological advances in recent years have led to a significant amount of data which is now generated in everyday life, such as shopping, travelling, banking, manufacturing and trading, public utilities, state and governance, sports, entertainment, science, education and health. Commercial organisations, research bodies and governments have started to realise the importance of using this data for their growth. As a result, the study of big data has gained prominence among scholars in different areas of research (Einav & Levin, 2013; Mayer-Schönberger & Cukier, 2013) as well as generating interest from the non-academic world (BBC, 2013; Lohr, 2012) [4]. The concept of big data encompasses the collection of data, the combination of the data collected from various sources, processing it and using the results so obtained. Specifically, big data is a term used for large databases requiring complex processing and visualisation which cannot be efficiently handled by traditional data processing software (Wikipedia, 2014a). According to the McKinsey

Global Institute, “Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze” (Manyika et al., 2011). A well known model (known as 3V’s model) of big data attributed to Gartner Inc. defines it as “Big data is high volume, high velocity, and/ or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization” (Beyer & Laney, 2012). The term ‘volume’ here indicates the complexity of datasets and not necessarily their size. ‘Variety’ refers to the different type of structured or unstructured data such as text and numeric, video and audio and log files. ‘Velocity’ refers to the speed with which data can be made available for analysis. Sometimes other V’s such as ‘Veracity’ (aiming at data integrity and the ability of the organisation to confidently use the data) or ‘Value’ (does new data enable an organisation to get more value?) are highlighted as well (Swoyer, 2012; Villanova University, 2014) [4].

The society continues to become more and more addicted with Social media which generates more information in a short duration than any other contemporary approaches. Data lies scattered across the Internet, while businesses, suppliers and customers are creating and consuming large amount of information. Understanding what it means for business and making intelligent decisions will give a competitive edge over one’s peers in the global market. The quote on CNBC by Wilbad (2012), really exemplifies this “Data is the new Oil.” which has the potential to turn business in any desired direction. Traditional approaches to large data are bound to break down soon under large exploding growth of data collected from social networks. This requires new technologies and techniques such as those used in big data. The ability to analyze and mine data from social networks and utilize the knowledge inside them will help in developing new business models, redesign business processes, accelerate one business or control the direction of the business trend. The data analyzed can paint a picture of the needs and behavioral patterns of groups and individuals, of any product, on regional basis. Researchers and policymakers are beginning to realize the potential for converting these torrents of data into useful information that can be used to identify needs, develop new products, provide services, and predict and prevent crisis for the benefit of the organization. Better predictions can be made to target more-effective interventions, in areas that have been dominated by gut and intuition rather than by data. For this reason it has the potential to revolutionize management [3].

Social media, a range of websites and applications, has evolved from a social networking tool to a crucial tool for news dissemination, enterainment, and commerce. It serves as a vital tool for personal and business interactions, enabling communication and customer engagement .Business to-consumer websites now feature social components for easy engagement, with tools for tracking brand perception across platforms like Facebook, TikTok, Instagram , and LinkedIn with mobile application their reach.

Social Media :

The society continues to become more and more addicted with Social media which generates more information in a short duration than any other contemporary approaches. Data lies scattered across the Internet, while businesses, suppliers and customers are creating and consuming large amount of information. Understanding what it means for business and making intelligent decisions will give a competitive edge over one’s peers in the global market. The quote on CNBC by Wilbad (2012), really exemplifies this “Data is the new Oil.” which has the potential to turn business in

any desired direction. Traditional approaches to large data are bound to break down soon under large exploding growth of data collected from social networks. This requires new technologies and techniques such as those used in big data. The ability to analyze and mine data from social networks and utilize the knowledge inside them will help in developing new business models, redesign business processes, accelerate ones business or control the direction of the business trend. The data analyzed can paint a picture of the needs and behavioral patterns of groups and individuals, of any product, on regional basis. Researchers and policymakers are beginning to realize the potential for converting these torrents of data into useful information that can be used to identify needs, develop new products, provide services, and predict and prevent crisis for the benefit of the organization. Better predictions can be made to target more-effective interventions, in areas that have been dominated by gut and intuition rather than by data. For this reason it has the potential to revolutionize management.[6]

RESEARCH IN SOCAIL MEDIA BIG DATA FOR PUBLIC HEALTH INTERVENTIONS :

The reports described in the previous section have illustrated "infodemiology" techniques that leverage big data from social media for timely insights that could inform the development of critical public health responses. Other researchers have further described the amalgamation of multimodal data from social media complemented by other sources such as traditional news media and online behaviour/market research agency platforms to inform the development of evidence based public health interventions. These capabilities have been demonstrated for the evaluation of the public's compliance to public health measures as well as the evaluation of national responses to help control the pandemic in China, using data from social media such as Weibo and Tik Tok, the People's Daily major Chinese newspaper, and online market research agency platforms such as Mob-Tech research institute (Hua and Shaw, 2020)

Researchers have further demonstrated the potential utility of new techniques such as Online Ecological Recognition (OER) that combine big data with other emerging technology domains like artificial intelligence (AI) to develop predictive models (Li S. et al., 2020). This facilitated additional applications beyond the surveillance of information, to evaluating the mental health impact of the pandemic itself. The study demonstrated that negative emotions and sensitivity to social risks increased while positive emotions and life satisfaction plummeted (Li S. et al., 2020). The results from this particular study corroborated with the Behavioural Immune System (BIS) theory that people tend towards negative emotions when threatened by disease, whereby the spike in negative emotion was heightened during COVID-19 due to the infodemic. However, another study of 17,865 Weibo users in China highlighted a silver lining regarding the impact of social media during this pandemic, whereby initial negative emotions (after COVID 19wasreportedwidely) were subsequently balanced by positive emotions as users leveraged social media platforms for peer-support, with trending topics such as "faith" and "blessing". (Li S. et al., 2020). These terms reflect greater group cohesiveness given the threat to greater public, and these findings were further replicated in Lombardy based on data from Italy (Su et al., 2020). Notably, the increased group cohesion occurred in tandem with more monetary and supply donations to regions of need and key organisations including the Hubei Red cross (Li S. et al., 2020). It is thus evident that social media can be leveraged for positive impact, by helping to connect individuals during a crisis and improve individual alignment for common good. This has additional implications for other aspects of medication, including the use of these platforms for health

promotion and raising awareness about critical health-related problems (Horrell et al., 2019). Through further investigation and refinement of these methods, public health organisations will be able to optimise response strategies in real-time by extrapolating trends in transmission, communication content, information f low, and population sentiment [6].

WHAT LIES AHEAD :

Our article has highlighted the potential impact of social media big data to be a double-edged sword. Presently, the negative impact has gained much visibility and criticism, due to limited mechanisms for differentiating reliable information from misinformation, and mitigating the risks of the latter. Fortunately, increasing coordination between social media platform providers, non governmental organisations, and governments have given rise to promising collaborations such as the "Share verified" initiative led by the United Nations (UN) to build a freely-accessible resource of reliable health content and front-end flags to redirect individuals to reliable sources in order to address misinformation. Ultimately, long-term solutions may require new legislation to govern the creation and dissemination of misinformation online. Regulations have been effectively applied for other public health challenges, such as tobacco advertising regulations to reduce population exposure to marketing and cues to smoke (Henriksen, 2012).[11]

How ever, in the case of online misinformation, then for cement of such legislation will be significantly more complexed, given the scale of individuals as potential sources as opposed to corporations that are stakeholders within the tobacco industry. This will likely require methods such as COSMO for big data surveillance, with incorporation of AI analytics of the social media big data to scale up enforcement. This also begets consideration of developing alternatives and complements to social media as sources of reliable health information hosting and exchange, with several recently launched in response to COVID-19 misinformation. Online health communities (OHCs) have drawn increasing interest in the domain of virtual social networks due to their potential to amplify positive impact such as peer-support and quality data as a source of health evidence (Smith et al., 2017; Audrain-Pontevia et al., 2019), as well as mitigate against negative impact through policies against the promotion of inaccurate information, as well as configurations that involve medical practitioners in moderation and content generation (Eysenbach, 2000; AskDr, 2020)[6].

However, even with these measures in place, studies have highlighted the potential for lapses to occur that can be difficult to detect (Huh et al., 2016). Therefore, there is a growing need for OHCs that leverage the strengths of social media platforms with additional embodiments that mitigate against its weaknesses. These may be configurations that empower verified medical experts with digital tools to moderate the content and flow of information. These applications of OHCs for patients with chronic pain and mental health disorders that are likely to progress and increase in prevalence during COVID-19 have been described in earlier reviews led by relevant specialist (Chew et al., 2020; Li L. W. et al., 2020). These digital platforms represent potential areas for future research and cross-disciplinary collaborations between technology partners, clinicians and regulators to enhance public health responses [6].

Limitations:

Many disciplines and sectors have advanced as a result of the widespread use of social media data and big data analytics. There are numerous hurdles and limitations to working in this field. With the increasing abundance of social media data, files are now being distributed over multiple

physical sites. Public access is becoming difficult and technical skill is needed to access these data. The maintenance of large social datasets is challenging and expensive. Consumes continuously sharing status updates, photos, videos, etc., are not always useful for analysis. Data cleaning and filtering are required to extract necessary data from this complex dataset that is costly and time-consuming.

Cyber-attacks have a severe impact on social data during sensitive events such as elections, which could result in a faulty conclusion. There is a great chance of getting unreliable and incomplete data. Like noise, misspelling, foreign data in English, etc. Integrating and combining social data from many platforms is a difficult task.

Big Data in Social Media :

Big data analytics on social media provide substantial benefits to social media, businesses, and individuals in terms of making wise decision [7]) . Several issues, however, remain unresolved. To tackle some big data difficulties, the big data research community, industry, and government must work together. such challenges these are:

A. User-generated content (UGC)

This is a type of data that is frequently in the form of text, images, and videos. This information is contributed by regular people, and it is not necessarily in a consistent manner. As a result, because the data provided by social media sites is inherently inaccurate and unstructured, user generated content performance ranges from extremely high to extremely bad. All of this data might include people's own beliefs, habits, and ideas, making it more difficult to extract useful information from particular data.[8] It is a rich topic for corporations and researchers to examine because of the availability of consumer information that may give useful and data on approximately.

B. Dataset Availability

Machine learning methods are heavily reliant on information available. As a result, before machine learning can take place, the time all the time must be available. A model also often learns from the training set before applying what it has learned to tasks like recognition and diagnosis. However, in the case of data flowing, where new data is constantly arriving, this scenario is impossible to realise . [9] [8]

C. Velocity

Learning cannot commence when the complete dataset is unavailable since most learning algorithms rely on data availability. However, in the event of data streaming, data availability will be impossible to provide owing to the constant arrival of new data. [10] Big data velocity refers to the speed and pace at which data is processed. Rapid technological advancements, such as cell devices and real-time sensor network, enable quick connection between our surroundings and other smart equipment. As a result, during the development phase, huge data velocity must be handled.

D. Natural language processing (NLP)

NLP is a type of intelligence work that identifies the meaning of human-written content. Many tasks, including as link extraction, speech recognition, automated summarization, topic separation,

sentiment classification, and entity recognition, may be accomplished using structural knowledge building. This processing includes subject separation, connection extraction, entity recognition (identification, automatic text classification, and separating). [7] It is used in text categorization, machine learning, and computerised question answering. In the world of massive social data, It is recognised as a difficult problem. Language acquisition is rarely exact or straightforward.

E. Data Locality

The fundamental issue associated with diverse data types is data localization. Furthermore, all computational methods need the storage of datasets in memory. This assumption, however, does not apply to large data since the entire data set cannot be stored in memory. The data is often split among several files located in distinct physical places. Big data is not the same as traditional machine learning, which just sends data to a computer facility. [7] As a result, in need for a big data platform to be efficient, the problems of data localization, processing latency, and traffic patterns must be addressed. Language acquisition is rarely precise or direct.

VIII. COMMON SOCIAL MEDIA PLATFORM

Social networks offer free and fast communication with friends usually in multiple forms such as tweets, pictures, videos and texts. Another attractive feature is creation of groups for multicast communication with friends, colleagues or family members. The most popular social media platforms are given below.

A. Instagram:

Big data has revolutionized how social media platforms, including Instagram, collect, analyze, and utilize data to enhance user experience, improve marketing strategies, and influence trends. Instagram, with over 2 billion monthly active users, generates vast amounts of data daily, making it a prime example of big data applications in social media.

B. Facebook:

A social networking website that makes it easy to connect and share with family and friends. For availing the services users have to create personal profiles. Interactions are in the form of ‘posts’ which may be text, images or videos, and users can indicate their preferences for the content, articles, products and services through a ‘Like’ function.

A. Google+:

It is a platform focused on bringing all the services of Google together for users to enjoy social networking. Features include the ability to post photos and status updates to interest based communities, group different types of relationships into Circles, multi-person instant messaging, text and video chat called Hangouts, events, location tagging, and the ability to edit and upload photos to private cloud-based albums (Gundotra, 2011; Lytle, 2013).

B. LinkedIn:

LinkedIn is a social networking forum designed specifically for professionals. The goal is to allow registered members to establish a network of people they know and trust professionally.

C. YouTube:

YouTube is a social media platform and website for sharing video online. Allows users to upload, view and comment on videos. It was founded in 2005 by Chad Hurley, Steve Chen , and Jawed

Karim. Google purchased YouTube in 2006 for $1.65 billion.

Conclusion :

In conclusion, the integration of big data and social media has transformed how individuals and organizations understand, interact, and influence the world around them. By leveraging vast amounts of data generated through social media platforms, businesses, governments, and researchers are able to gain real-time insights into user behavior, preferences, and emerging trends. This provides opportunities for more personalized services, targeted advertising, and improved decision-making**.**

However, this transformation is not without challenges. Ethical concerns surrounding data privacy, the accuracy of data analysis, and the potential for misuse of information remain pressing issues. Additionally, the sheer volume and complexity of data require advanced analytics and artificial intelligence tools to extract meaningful insights. As the use of big data continues to expand within social media, it is crucial for stakeholders to balance innovation with responsible practices to ensure that data-driven solutions benefit society as a whole. Overall, the convergence of big data and social media holds immense potential, but it also demands careful consideration of privacy, security, and ethical standards to ensure sustainable and equitable development in the digital age.

REFERENCES

[1] V. Dhawan and N. Zanini, Big data and social media analytics, Res. Matters A Cambridge Assess. Publ., no. 18, pp. 36–41, 2014.

[2] K. Smith, 126 amazing social media statistics and facts, https://www.brandwatch.com/blog/amazing-social media-statistics-and-facts/, 2019.

[3] Author : Dr .S .Matilda, Professor & Vice Principal, IFET College of Engineering, Tamil Nadu, India-BIG DATA IN SOCIAL MEDIA ENVIRONMENT

[4] Vikas Dhawan and Nadir Zanini Research Division Big data and social media analytics .

[5] V. Nunavath and M. Goodwin, The role of artificial intelligence in social media big data analytics for disaster management–initial results of a systematic literature review, in Proc. 2018 5th Int. Conf. Inf. Commun. Technol. Disaster Manag. (ICT-DM), Sendai, Japan, 2018, pp. 1–4.

[6] Alton M. K. Chew1,2 and Dinesh Visva Gunasekeran1* extra-(1Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore, Singapore, 2UCL Medical School, University College London (UCL), London, United Kingdom) Big Data and Social Media : The Good, The Bad and The Ugly (un)truths

[7] Rahman, Md. Saifur & Reza, Hassan. (2021). Toward Big Analytics in Social Media: Study of Types, Techniques, and Taxonomy. Data

[8] Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2019). Social media big data analytics: A survey. Computers in Human Behavior, 101, 417 428. https://doi.org/10.1016/j.chb.2018.08.039

[9] Sahatiya, P. (2018). Big Data Analytics on Social Media Data: A Literature Review. International Research Journal of Engineering and Technology (IRJET), 2018 (05)(02).

[10] R. (2022b, februari 27). Top 9 Big Data Challenges (And How You Can Solve Them Easily). SyncApps, the #1 Integration Cloud. Geraadpleegd op 16 maart 2022, van https://www.cazoomi.com/blog/top-9-big-data-challenges and-how-you-can-solve-them easily

[11] Henriksen, L. (2012).Comprehensive Tobacco Marketing Restriction : Promotion , Packaging ,Price and Tob.Contol.21(2),147-153.Doi :10.1136/050416