International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2022): 7.741

Review on Practical Applications of Social Media Data Mining

Sibi Mathew

TKM College of Engineering Kollam, Kerala, India, 691005
Email: sibimathew2019 [at]gmail.com

Abstract: ***Social media accommodates about 5 billion users' worldwide sharing events, comments, and sentiments, multimedia, documents, etc. It handles large volumes of data on a daily bases. There has been a wide variety of research evolving around this huge chunk of data. Even though it put forward the question of privacy, still ongoing research proves its applications. This research was conducted to examine the tremendous social and commercial value of social media data. This review examines previous literature and existing applications from a practical perspective. It provides an overview of commonly used pipelines when building social media-based applications and describes available analytical techniques such as topic analysis, time series analysis, sentiment analysis, and network analysis. We then present the impact of such applications in three different areas including disaster management, healthcare, business, etc. Finally, we enumerate existing challenges and suggest possible future research on the area.***

Keywords: Social media, prediction analysis, applications

1. Introduction

The rise of social media has made sharing and accessing information faster and easier. Social media can serve as a tool for raising awareness, especially during emergencies. When an alarming cluster of wildfires broke out in the Amazon rainforest, he first posted about the wildfires on his Twitter account on Aug.6, 2019, two weeks before Cable News reported the incident. Much research has focused on extracting anomalous events such as disease outbreaks, natural disasters, and infrastructure outages from social media data. Additionally, social media helps connect individuals and break down communication barriers. When Japan was hit by the earthquake and tsunami in March 2011, millions of people around the world turned to social media to find family and friends. Additionally, social media provides an opportunity for everyone to tell their story and share their opinions. Businesses can apply text mining and sentiment analysis to social media data to understand public opinions and comments about their products and services to better understand the market and prepare for the future. It also enables different studies on human behavior as well.

Social media can be categorized by type. of the content, it generates. For example Wikipedia allows collaborative editing. Blog etc. WordPress regularly publishes articles about events. subjects, similar to Twitter, some other social media such as Instagram, and Facebook help in connecting people as well as sharing multi-media, etc. There are some other social media websites that even rate people and places. Websites like Stack Overflow, W3schools, and Brainy serve as a platform to ex-change queries and answers, etc. Even there are websites that connect professionals such as Linked In. With the widening verities of social media, it has become very difficult to measure its scope or depth in our day-to-day life.

According to Kepios, there are more than 4.74 billion social media users in October 2022 which account for about 59.3 percent of the global population. Due to the sheer number of users and the sheer amount of content created by them, social media is considered a valuable data source in both science and industry. Numerous applications have been developed using social media data to provide value and insight in a variety of areas. This chunk of data put forward some challenges in terms of its volume, speed of processing, etc. Unlike traditional data sources, which typically have centralized agents to generate content, social media allows users to generate content themselves. Social media has also begun to influence the cultural, psychological, and behavioral patterns of both adults and teenagers. [2] Due to poor quality control, Information circulating on social media cannot be given the same reliability and completeness as official news sources. [1]

2. Common Application

A. Data collection

Most social media platforms such as Twitter and Facebook offers developers a robust official API Collect data. Some APIs, such as the Twitter Streaming API, enables real-time data streaming with concurrency. A large amount of data is collected and shared in live stream data services which can be used for event recognition.

Uses web crawling to collect data when public APIs. A web crawl also called web scraping or spider is a computer program used to scrape information from websites. However, web crawling can have negative connotations if it scrapes personal or sensitive information, ignores the website's privacy policy or terms of service, or scrapes without the owner's permission. It's important to review policies and regulations before crawling. When implementing data protection regulations, privacy is becoming an issue as social media platforms collect vast amounts of data. [4]

B. Data storage

It requires appropriate technology for storing and processing large amounts of data. Relational databases (such as Microsoft SQL Server, MySQL) and non-relational databases (Cassandra, and Redis) are commonly used for storing social media. Media data is currently under study. Despite the spread there is another NoSQL-based database research. This research is not focused on research data storage solutions, but extensive research in this area can be found in Stieglitz's studies [7] Commonly available data storage solutions. Moara provided a thorough review of data warehouse design for social media data. Cao [9] described data warehousing and online analytical processing technology in the field of business intelligence.

C. Data analysis

Large-scale application with social media Data is used in detecting events using topic analytics technology. Extracted events from nature from disasters to disease outbreaks, etc.

Some applications are designed for study time Series information on SNS. Time Serial data provide an excellent resource for research Analysis of past behavior, current trends, Identify, and predict anomalies or bursts Analysis of future and seasonal variations. Social media offers a rich source of user-generated content. Sentiment analysis is particularly popular in business and marketing. People share their thoughts and feelings about the product, Social media events and news. Some applications look at networks of social media virtual communities to identify influencers or investigate information diffusion pat-terns. Such applications analyze relationships between users. Such relationships include 'Follow', 'Like', and 'Report'. Diagnostic analysis is an extended type of analysis characterized by techniques such as data discovery, drill-down, data mining, and data correlation. Predictive analytics transforms data into profitable and actionable information. Prescriptive analytics Prescriptive analytics can similarly propose decision opportunities on exploiting a future opportunity or lessening future risks. The paper will be looking into some of the many analysis areas.

D. Data visualization

Data visualization is to present information in a graph, chart, or other visual formats, so as to make it in an easily understandable form. Data visualization help understand patterns and similarities in data very easily. As social media data are always on a large scale and it is challenging to interpret increasingly large batches of data, data visualization has become an indispensable tool. Applications are providing innovative visualizations to better present the information from social media. There are a large variety of visualization tools that enable us to understand and analyze data and also obtain the essence of data. Visualization is not only a tool to help researchers better understand information during data analysis but also serves as a demonstration to communicate the analysis results to the audience.

E. Computational intelligence

Social media platforms have grown in popularity as interactive and behavior-rich resources, which present the potential for big data analytics. Additionally, in order to grasp opinions on these platforms, artificial intelligence must be actively used. Computational Intelligence is an adaptive processes that "allow or assist intelligent behaviors in complicated and changing situations" is how this new scientific field of intelligence is defined. Artificial neural networks (ANN), fuzzy systems, swarm intelligence, evolutionary computation, and deep learning are the classification's foundations, used in "Computational Intelligence". [3]

3. Analysis Techniques

A. Topic analysis

Social media has become an important news source. Evan Williams, one of Twitter's founders, defined the service as "What we have to do is bring people the best the latest and most relevant information possible. we Think of Twitter as not a social network but as an information network it tells people that they are interested in what's happening in the world". This strategy changed Twitter around 2010.

"Twitter defines low-level information News Flash Portal". Even though it is not considered an alternative to authoritative news sources it can sometimes provide real-time information. Inform online communities ahead of newspapers of new problems. Reporting an incident always requires a certain amount of time, but ordinary people can easily post what's going on on Twitter. Therefore, researchers are interested in Development of meaningful topic extraction method from millions of data points. In the previous work, there are three major solutions suggested.

1) Use of Hashtags

Hashtags are defined as words or acronyms Terms that begin with the "#" symbol. it is widely used in Social media such as Twitter and Instagram. The hashtag is considered important metadata for classifying posts To spread an idea or topic. The use of hashtags is an integral part of social media. [15]

These act as sources for identifying social media themes.

2) Queries to search

Search queries are also important For example, search on Google using your keywords. "NASA" provides a list of articles about NASA, so you can think of it as a "NASA" article group topic. The platform collects social media data in order to interrogate different topics. For example, suppose we are considering the phenomenon of landslides from Twitter data. Twitter Tweets are used to feed the system captured by the query "landslide". However, it can introduce noise in search keywords' multiple meanings. This introduced a requirement for a classification model. The winner of the election is found from the classification model.

3) Topic modeling

A topic model is defined as a kind of model for discovering abstract themes in a collection of documents. Theme modeling is commonly used as a text mining tool for discovering the semantic structure hidden in the text body. LDA is a popular topic modeling technique. Expect it to improve accuracy; researchers are expanding LDA by including additional information such as Hashtags, user profiles, geolocation, queries, etc. Minutes proposed TLDA to bridge hashtags and topics. TLDA extends LDA with observed hashtags and associates hashtags with common thematic combinations [12], [13], [14] to discover potential themes and integrate theme-specific nor-mal distribution of location (latitude) and longitude). It also adds a sample base parameter estimation. Works used in the past Spatial information for improved topic extraction used the relevance and similarity of resources, they Estimated directly by the distance between resources Sites. Others took user and social features and consider networks to improve the topic accuracy model.

B. Descriptive analytics

Descriptive analytics is the first step in the data processing process that gives all the historical/past data required to provide useful information. Additionally, it can be applied to facilitate additional data analysis. Historical data are provided a thorough understanding of the successful and unsuccessful components of previous data. Based on this knowledge, descriptive analytics offers methods for data analysis. Descriptive analysis, often known as ”post-mortem analysis, ” is frequently used in businesses to report on events, such as the supervision of sales, and departmental, and financial activities. The descriptive models are able to measure, recognize, and classify different connections and relationships in data. Further models that can simulate a sizable number of customized agents and generate predictions can be improved using descriptive modeling techniques. [3]

C. Diagnostic analytics

Diagnostic analytics is an extended type of analysis characterized by techniques such as data discovery, drill-down, data mining, and data correlation. Diagnostic analytics examines data and answers questions such as ”How did that happen?”. This analysis delves into the data to understand the detailed behavior and causes of the event. Diagnostic analytics give you the opportunity to quickly understand your data and quickly answer key employee questions. Cornerstone View is an example of diagnostic analytics that provides executives/organizations with the fastest and easiest way to gain valuable knowledge related to complex issues. Use data visualization as an interactive tool, allowing administrators to consolidate data and easily search and filter people. For example, a manager may reveal important knowledge about an employee when applying for an important position or succession.

D. Predictive analytics

Predictive analytics turns data into useful and actionable in-formation. This analysis uses data to make decisions about the future outcome of an event or the likelihood that a particular situation will occur. Predictive analytics includes a range of statistical techniques, from modeling, machine learning, and game theory to analyzing current and past facts, to predict future events. Historical data found in predictive models are used by companies to identify risks and opportunities. The model identifies relationships between various factors to enable risk assessment based on a specific set of conditions that guide the decision-making process. Visionary the analysis consists of three main phases: modeling (prediction), decision tests and optimization, and transaction profiling. Optimizing customer relationship management systems is an example of predictive analytics. It enables organizations to analyze all customer data and predict customer behavior. Overall, predictive analytics can lead to big gains and some strong customer relationships.

E. Prescriptive analytics

Similarly, prescriptive analytics can suggest decision-making opportunities to take advantage of future opportunities or mitigate future risks. This analysis shows the impact of each choice. In practice, prescriptive analytics can continuously and spontaneously process new data to improve accuracy and provide better decision-making options. A prescriptive approach is a focused approach that examines possible options, the relationships between options, and the implications and consequences of those options, ultimately suggesting the ideal option in real-time. The efficiency of predictive analytics depends on the sufficiency of the decision model to capture the impact of the analyzed decisions. Additionally, optimization, game theory, simulation, and decision-making techniques are specific approaches used in prescriptive analytics.

F. Time series analysis

A time series is an indexed series of data points that are chronologically ordered. Time series carry deep importance in examining and comparing Past Behavior past, and current trends, identification of anomalies, or bursting, predicting future trends, analyzing seasonal data variations, etc. Time series analysis has become an important analytical method in social media research. There are two features of social media that make it even more likely Interesting and challenging. First, social media Contains transient types of events Correlate with each other, but you can also capture various aspects of the event. Second, there is Random, unpredictable, and obnoxious noise on social networks media data. Three steps of time series analysis are shown as follows.

1) Data selection and preprocessing:

Data are filtered based on research questions. Predicted the purchases of digital cameras and personal computers, so they collected tweets written by users who explicitly admitted they purchased the products. Comito [18] wanted to capture relevant events in a geographic area, so they collected data from one location.

2) Data extraction and transformation:

Time Series analysis requires data points arranged as follows: Chronological. Each data point can be a frequency of a particular event, text, or place. Social media data contains a lot of information, including hashtags. Geographic information, text, and user profiles. You can extract interesting features from your data, such as: Hashtags, User Connections, Repost Frequency action. Studies on mail networks show that they extracted the frequencies of communication and use them on other functions across accounts analysis. Zhao et al. and Dahouei [17] hotly identified Topics and frequency of use from social media of hashtags as a function and use LDA to extract and categorize main topics and categorize topics used for time series analysis.

3) Time series analysis. Research purposes, different tech-niques which are applied to the extracted features.

- **Exploratory analysis**

It is a way that generally uses visible techniques to summarize the primary functions of the dataset. Some researchers lay out and construct gear to assist others to discover time-collection data. TwitterMonitor corporation's records are primarily based totally on key phrases and visualize the trends. Some others supplied a visible analytic device to research stance phenomena on time collecting data from social media.

- **Smoothing techniques**

These are commonly used to clean abnormal roughness to look a clearer sign in time collection. Dahouei [17] analyzed the distribution of hashtags from warm subjects and tried to discover styles be-low the distribution. In different words, they desired to clean the distribution with the aid of figuring out the essential points. Some studies had been inquisitive about the neighborhood country smoothness and correlation throughout a couple of streams to discover occasion bursts. The Perceptual Import Point (PIP) approach to extract the best beneficial information and clean time collection trends were also in use.

- **Clustering**

Clustering with time collection is to partition facts into corporations primarily based totally on similarity or distance, in order that time collection styles within side the equal cluster are similar. Some researchers [2] studied temporal styles associated with online content material and the way the content material‘s popularity grows and fades over time. They tackled the problem with a time collection clustering technique, the K-Spectral Centroid (K-SC) clustering algorithm. They located most press organization information showcase a completely speedy upward push followed through extraordinarily sluggish decay, whereas, bloggers play a very crucial position in figuring out the sturdiness of information on the web. The outcomes may be used for higher placing content material to maximize click-via charges and additionally for locating influential blogs and tweets.

- **Peak detection**

Peak detection is to discover surprising surges, which are inherently interesting, as they recommend the occurrence of an influential occasion in social media. In a commercial enterprise context, activities that generate sufficient pastimes to result in flurries of social media sports are significant to commercial enterprise analysts. For social goods, surges may also be herbal disasters, sickness outbreaks, or infrastructure breakdowns and height detection may be used to comprehend emergency detection. There are a variety of height detection techniques, together with Palshikar‘s height detection algorithm, Wavelet remodel algorithm and Lehmann‘s height detection algorithm. These algorithms were evaluated for so-cial media occasion detection. Comito [18] extracted space-time functions from social media and implemented height detection to discover deviations from consumer's everyday behavior, within side the wish to find significant activities.

- **Prediction**

Prediction is to forecast destiny primarily based totally on historic observations. Zhao et al. [16] anticipated the reposting counts of micro-weblog messages on Sina Weibo using curve becoming optimized through the empirical model. Ni [17] forecast the subway passenger glide with social media records. Prediction with social media records is likewise popular with inside the commercial enterprise context. A huge quantity of traditional time-collection fashions is expecting the destiny income of a product every so often to offer unpromising results, due to the fact they have majorly trusted the historic income and ignored the dynamic effect of new events. Customers‘ current opinions and feedback approximately organizations and products could in large part have an effect on shopping behaviors [20]. Social media is a superb supply of such information. Therefore, it was proposed to mix social media prediction and current analytical forecasting fashions. Some studies such as those of Han [21] emphasized extracting sentiment from social media for income prediction.

G. Social network analysis (SNA)

Social network analysis is based on identifying social relationships between different social media users using the network theory of nodes and connections. Social Network Analysis involves Cognition and sociology. This analysis also has significant support in the fields of medicine, anthropology, biology, and information science. SNA has been the subject of speculation and research. SNA experienced a renaissance due to the ubiquity and quantity of content. Social networks are embedded at many levels in many data sources. Social networks can emerge from information from sources such as text, databases, sensor networks, communication systems, and social media. [23], [24], [25]

H. Sentiment analysis

Social media is one of the richest reasserts of publicly generated text. Sentiment evaluation has ended up an essential place of labor to recognize the emotions and feelings of positive organizations due to the fact such a lot of humans share their mind and emotions toward merchandise, events, and information on social media. Business proprietors would want to examine how their merchandise or carrier makes humans feel; clients seek advice from online opinions while making buy decisions, politicians gauge public reactions to their policies; quick-time period inventory expenses are influenced via way of means of current public opinion on companies; and the revenue of movies, concerts, and performances are pushed via way of means of target market opinions. Some studies have used real-time records from Twitter to construct Business Intelligence (BI) structures to forecast income and revenues. Others investigated the impact of social media on quick time period firm inventory marketplace performances via way of means of reading sentiment in wealthy texts.

Sentiment evaluation objectives to decide sentiment polarity. Most present evaluation strategies are consciousness of differentiating between fine and bad feelings, or a few would possibly have impartial as an extra category. He [27] tried to gain finer-grained sentiment evaluation, which could yield greater precise and greater actionable effects with targeted bad emotion subcategories, along with anger, disappointment, and anxiety, and fine emotion subcategories, along with happiness and excitement. There are two major approaches discussed in the previous literature. [26], [28], [29]

- **Lexicon-based approaches**

Lexicon-primarily based totally methods are famous with-inside the early stages. They become aware of the dominant polarity of textual content with the aid of using trying to find emotional signs primarily based totally on lexicons. The overall performance closely relies upon the fine of sentiment lexicons. Furthermore, accuracy is restricted because of semantic ambiguity. The widely-used English sentiment lexicons encompass Linguistic Inquiry and Word Count (LIWC), General Inquirer Lexicon, and SentiWordNet1. To examine different languages, sentiment lexicons within side the corresponding languages should be built, or translating centered texts into English or translating English lexicons into corresponding languages is needed. Some studies used Chinese Emotion Word Ontology to behavior a Simple Sentiment Word Count Method (SSWCM) on Weibo, a Chinese social media platform.

- **Learning-based approaches**

Learning-based approaches, including supervised and un-supervised learning approaches, have become very successful in the field. The approaches derive sentiment polarity from the relationship between the features of the targeted text. Earlier works focus on bag-of-words and linguistic features. Later works start exploring metadata, including reposts, hashtags, likes, and counts. The main blockers to building machine learning models are learning speed and effectiveness. It is difficult to train a model with too many features on a huge training dataset. Therefore, feature selection is the key to building effective sentiment classification models. Studies proposed to use of Chi-Square feature selection to select significant features and remove irrelevant, redundant, and noisy data to improve performance. They also discussed how to handle negation and emoticons to further improve the accuracy of machine learning-based approaches.

Perikos and Hatzilygeroudis [73] brought the ensemble classifier schema that mixes lexicon-primarily based to-tally and learning-primarily based totally methods to research big social statistics and version feelings in social media. The lexicon-primarily based on the total version does now no longer require schooling and it could be used immediately and successfully for big quantities of texts. The learning-primarily based total version can similarly increase accuracy. The purpose of the ensemble classifier schema is to effectively leverage the benefits of the bottom classifiers, aiming to boom ordinary performance and accuracy.

- **Sentence-level sentiment analysis** Document-level sentiment analysis can be too coarse-grained for certain purposes. Much of the early work on sentence-level analysis focused on subjective sentence identification. However, it requires complex tasks, such as handling conditional statements, handling ironic statements, and so on. In such cases, sentence-level sentiment analysis is preferable. [29]

- **Document-level sentiment analysis** The biggest challenge in document-level sentiment analysis is cross-domain sentiment analysis and cross-language sentiment analysis. Specific domain-oriented sentiment analysis has been shown to achieve remarkable accuracy with high sensitivity domain. The feature vectors used in these tasks contain sequences of words restricted to specific domains. Sentiment classifiers are applied because annotating the data for each new domain is costly. Alignment of spectral features, structure correspondence learning, and emotion-sensitive thesauri are three classic techniques. They differ in terms of feature vector expansion, word relevance measures, and finally the classifiers used for classification.

I. Word-level sentiment analysis

The document-level analysis focuses on distinguishing the entire document as subjective or objective, positive or negative since the document contains both subjective and objective sentences, sentence-level analysis is more effective than document-level analysis. Words are the basic units of language, but word polarity is closely related to the subjectivity of the corresponding sentences and documents, there is a great possibility that sentences containing adjectives are subjective sentences. Furthermore, expression word choice reflects not only demographics such as gender and age, but also motivations, personality, social status, and other psychological or social characteristics. Words are therefore the basis for text sentiment analysis. Currently frequent Methods used include natural language processing-based approaches and ma-chine learning-based approaches approach. For sentiment analysis of micro blogging texts, most researchers suggest applying concept-based methods. The emotional concept is a connecting link between the emotional direction of the text and the individual words. A collection of specific types of perspective information that indicates the sentiment and subjectivity of text. There are several studies on the emotional orientation of words based on the use of emotional vocabulary.

J. Network analysis

Social media creates a digital network wherein customers are grouped consistently with their preferences. Information is transmitted inside and among groups. Previous programs are dedicated to reading the relationships on this digital network and to informing how facts flow and who contributes to the dissemination of facts. There are 5 not unusual places sorts of consumer interactions on social media being tagged on a post, commenting on a post, sharing a post, liking a post, and following others. Generally, in social community evaluation, we gift taking part customers as vertices or nodes, and we outline edges or hyperlinks as connections among customers. Some researchers do not forget direction, whilst others do not. Most of them expect that any hyperlink among vertices shows their similarity. The 3 important functions of social community evaluation are defined below.

- **Identify influencers.**

Generally, influencers are people or corporations which have installed credibility in a particular industry. Social media influencers can attain a big target market and persuade others with their authenticity and affect. Influencers are uncommon in social networks, however, their effect can speedy unfold to maximum nodes within the community. Identifying influencers permits us to boost records propagation, conduct hit e-trade advertisements, etc. Page Rank is a famous set of rules used to calculate users‘ effects. Han [31], [32] considered consumer affect to enhance landslide detection with Twitter statistics. They brought the ideas of applicable and inappropriate digital groups primarily based totally on whether the published messages are applicable to landslides or not long after which implemented the Page Rank set of rules to pick out influential nodes in every community. Page Rank was used to calculate users‘ authority to discover rising subjects on Twitter. From graph perspectives, diverse centrality measures had been provided to pick out critical nodes, together with Degree Centrality, Betweenness Centrality, and Closeness Centrality. Additionally, there are new algorithms proposed to similarly enhance the measurements with the aid of using topological connections. The IC version is an extensively used records diffusion version that received influential nodes with the aid of using an extended IC version to the second-order IC version. The second-order IC version considers the effects of direct pals and additionally takes the preceding effect into consideration. It was proposed to apply affiliation rule mastering to discover the relationships among users. Association rule mastering is a machine mastering method that pursues to discover how one object influences other with the aid of using studying how often positive gadgets seem together. Some studies used statistics from Location-Based Social Networks (LBSNs) and proposed a singular community version and a power propagation version, which have a look at effect propagation in each online social networks and the bodily world.

- **Predict links.**

Link prediction in social networks is described because of the project of predicting whether or not a link among particular nodes will seem within side the future. Link prediction can offer insights concerning how the community will develop or fall apart and what number of connected additives will exist within the community. A graph may be expressed as an adjacency matrix. Link prediction may be regarded as a binary category problem, classifying the factors within the adjacency matrix as zero (linked) or one (linked). To construct a category model, preceding paintings have explored distinctive features, which include graph structural properties and attributes of the nodes. However, because of the huge scale of social networks, the category method is time-consuming. Some studies proposed a green answer for predicting key community degree values with time collection forecasting.

- **Visualize networks.**

Network evaluation is essential and powerful; however, it's also complicated. Analysis of big networks calls for specialized area knowledge. Some researchers work on growing gear to assist non-programmers apprehends and examine complicated networks. They desire to allow greater full-size community evaluation studies and assist researchers to examine the networks with much less effort. Node XL is this kind of device that permits customers to discover networks and find out insights with an easy-to-use interface.

K. Text mining

Text mining has emerged as one of the most popular techniques in social media to extract information from various types of unstructured content such as text, images, and multimedia. A detailed overview of text mining in social media is presented by [38]. In addition, [39] used text mining techniques to compare data on dietary habits in disability communities based on fitness tracking techniques.3 Types of Eating Disorders Associated with Dates collected from comments posted on social media for 3 years. The data consists of 6 subreddits. Results confirm that the pre-existing conditions subreddit recovers less than the eating disorders subreddit, which has the highest frequency of fitness tracker terms. Some research investigated whether computational linguistic analysis methods could be used to assess suicide risk and emotional distress in social media in China. A framework for social media content has proposed some research. The authors used an empirical study to test the framework during the 2014 World Cup.

4. Application Impact

Social media-primarily based totally packages have an extensive variety of impacts. Our survey specializes in their effects within the fields of healthcare, catastrophe management, and business.

A. Healthcare

Traditionally, public fitness surveillance is supported with the aid of using the reviews of particular illnesses from fitness care vendors to public fitness officials. The technique is classed as indicator-primarily based totally surveillance. Because fitness care vendors are advised to file instances that contain laboratory confirmations or meet sure definitions, the reviews are normally credible. However, the restrictions of the technique are insurance and timeliness [32]. The reviews come from the locations in which there are fitness infrastructures and fitness vendors are inclined to file the instances. Also, the file can arise after unwell people search for scientific attention. The fast increase in social media utilization and the innovation of massive information evaluation offer the possibility to discover fitness-associated occasions via unstructured and non-standardized or subjective reports, blogs, stories, tweets, and different facts. The method is assessed as event-primarily based totally surveillance. It is reasonably priced and flexible, consequently fending off the restrictions of the conventional method. The method may be used anywhere, and may even discover a plague earlier than the affected person seeks scientific treatment. However, every other set of issues exists. Since facts aren't always continually monitored with the aid of using professionals, the trustworthiness of the outcomes has been questioned. The acceptability amongst public fitness practitioners and policymakers is every other issue. Social media is extensively utilized in public fitness practice. A massive variety of good-sized and heterogeneous social media-primarily based totally prototypes were advanced to reap ailment surveillance and outbreak management. Some programs were followed with the aid of using the fitness government data. We recognized 13 examples as representatives. Twitter is the source of a first-rate fact. Most structures are funded with the aid of using governments, universities, and hospitals, inclusive of the Global Public Health Intelligence Network (GPHIN). Building multilingual programs is vital to cowl worldwide occasions. GPHIN, constructed in 1997, can help nine languages and MedISys, constructed in 2008, can help 25 languages. With the additional, improvement of records evaluation techniques, outbreak detection has emerged as extra correct and efficient. The recognition amongst public fitness practitioners and policymakers has dramatically increased. The World Health Organization (WHO) makes use of the Canadian-primarily based totally GPHIN in its international alert and reaction activities. The European Centre for Disease Prevention and Control makes use of MedISys, a web-primarily based totally device for epidemic intelligence activities. The US Centers for Disease Control and Prevention (CDC) has a unique program, the Global Disease Detection (GDD) Operations Center, committed to detecting and tracking international public fitness occasions thru Event-Based Surveillance (EBS). The Department of Health in Chicago and New York use Twitter and Yelp to identify foodborne illnesses. Some packages are cognizant of one sort of disease, together with Foodborne New York and Foodborne Chicago. Others generally tend to cowl a couple of diseases, like GPHIN and Health map. However, nearly all packages have a predetermined set of sickness sorts and gather social media statistics primarily based totally on those sickness sorts. In such a way, the packages can't perceive outbreaks of recent sicknesses sorts. For instance, in December 2019, the brand new coronavirus, COVID-19, became first reported. A utility that collects social media statistics with a predetermined sickness call will now no longer be capable of stumbling on this outbreak due to the fact that it is a brand-new sickness type. However, COVID-19 signs encompass fever and problems breathing, which can be much like many different breathing diseases. Thus, the insurance of the packages may be advanced with the aid of using amassing statistics primarily based totally on signs. With the assistance of wealthy statistics from social media and superior textual content mining technology, social media-primarily based totally biosurveillance structures may be used as early caution structures for upcoming public fitness emergencies. The structures file the effect of interventions, track development closer to precise goals, display and make clear the epidemiology of fitness problems, set priorities, and offer facts for public fitness rules and strategies.

B. Disaster Management

Dedicated bodily sensors are historically required to understand natural disasters, which include floods, earthquakes, hurricanes, and wildfires. According to the National Severe Storms Laboratory, the WSR-88D radars graphically show the detected precipitation on a map, which enables trouble to flash flood declarations, watches, or warnings. However, floods can occur for the duration of heavy rains, however, they also can occur while waves come on shore, while the snow melts too fast, or while dams or levees break. When detecting natural disasters, we want to understand what's going to occur, earthquakes, floods, or landslides, we need to understand how long it will occur and who could be affected. As social media is turning into a real-time facts dissemination platform, researchers have studied using social media in natural catastrophe detection. Some studies [31], [32] supplied case research of tweets for the duration of Australian floods and Haitian Earthquakes. Other research proposed techniques to elevate situational cognizance for the duration of Natural disasters. Some research devised a Landslide Information System (LITMUS) with a couple of facts sources, such as Twitter, Facebook, and Instagram.

They display that the pipeline of LITMUS may be without difficulty generalized for different occasion types, which includes infrastructure breakdowns.

Social media offers new ways to explore distributed communication in times of disaster [35], [36]. As citizens of the ubiquitous web, social media users collect and share their observations and views online. Social media improves situational awareness, especially during disaster management. The 2004 tsunami in the Indian Ocean and 2005 Hurricane Catalina created social media online communities for disaster preparedness. Some communities reported on the latest infrastructure status, while others assisted in contacting missing family members. Smartphones have many built-in sensors such as gyroscopes, magnetometers, a global positioning system (GPS), proximity sensors, ambient light sensors, a micro-phone, touch screen sensor, fingerprint sensors, pedometers, barometers, heart rate sensors, thermometers, humidity sensors that produce some useful data. Citizens will have a greater impact on disaster response if future social media can enable sharing of such sensor data.

C. Business

Mutual communication between consumers and providers is established through social media. Businesses use free social media platforms to provide businesses with a cost-effective way to advertise in their niche. Viral marketing is becoming cheaper, more efficient, and more widespread. Social media, on the other hand, is a platform for consumers to share their comments. Such user-generated content provides businesses with insight into their products and services and helps drive the development of new products. Existing applications examine social media data to assess the effectiveness of digital marketing and brand equity. They want to effectively create ads that effectively communicate their products and services to potential consumers. Desai and Han [37] used data from Instagram to investigate the impact of text comments on college brand equity. This is an important finding, as it should be encouraged To make the university into a diverse group of people in society Media Considering Universities Are Recruiting Students from all over the world.

Some research focuses on helping businesses better monitor and understands customer feedback and comments on social media. Some studies proposed a framework for a social broadcast-based business intelligence system that uses real-time information extracted from social media using text mining techniques. Business intelligence (BI) platforms are designed to acquire, interpret, collect, analyze, and explore information to support business functions. Traditionally, BI analyzes internal data sources such as operational data. The explosive growth of user-generated content It provided a new perspective on BI. By knowing public attitudes toward their products and services, companies can predict sales, revenue, and even corporate value. Most traditional time series models for forecasting sales are based on historical and seasonal sales and fail to capture the dynamic impact of recent events. Combining social media predictive models with traditional predictive models is valuable when consumer behavior is a key factor. Researchers presented a predictive model for monthly car sales using social media data. Some studies introduced a social data mining system to predict the box office performance of movies. Others examined the impact of social media on the short-term stock market performance of firms. They applied advanced sentiment analysis and used stock returns and risks as indicators of a company's short-term performance. Their results suggest that companies should carefully consider the importance of different media when implementing social media marketing strategies.

However, two-way communication between customers and providers in social media poses many challenges for businesses. Traditionally, advertising only allowed businesses to talk to their customers, whereas social media platforms al-lowed customers to talk directly to each other. Businesses can-not control the content, timing, or frequency of social media-based conversations between consumers. Therefore, it is very important for businesses to quickly monitor and understand customer comments. Additionally, businesses must learn to steer consumer discourse in a direction that aligns with the organization's mission and performance goals.

5. Challenge & Opportunity

A. Data privacy

"Knowledge is power, and in the Internet age, knowledge comes from data. Our personal data is the power and wealth it creates in today's data-driven economy. The time has come, and that includes going private, " said Attorney General Becerra [40]. Since 2018, many countries around the world are reviewing and discussing draft privacy laws. The General Data Protection Regulation (GDPR), which came into force on May 25, 2018, regulates data protection and data subject protection. GDPR will give businesses more control over the personal data they collect. In the United States, the California Consumer Privacy Act (CCPA) went into effect on January 1, 2020. This is a law that gives consumers new rights regarding access, deletion, and disclosure of personal data collected by businesses. South Korea is updating its regulations in hopes of achieving adequacy. Brazil's data protection law will come into force on August 15, 2020. The enforcement of data protection regulations; GDPR and CCPA, as well as other new data protection laws, have increased public awareness of data protection, especially in organizations that collect and process personal data. Some researchers emphasize the necessity and complexity of investigating the public interest while protecting user privacy. Privacy laws may vary from region to region, but the goal is the same: to protect your privacy. Using the GDPR as an example, this section focuses on its impact on social media-based applications and developing GDPR-compliant applications. Data protection concerns personal data or the use of personal data. Article 4 (1) of the GDPR defines "personal data" as any information relating to an identified or identifiable natural person (data subject) [96]. Information shared on social media platforms can be classified as personal data because it relates to an identified or identifiable natural person. Personal data from social media is owned by the data subject, not by the platform. Under the GDPR, there are important differences between data collection from data subjects and data collection from third parties. This is because the scope of collection from third parties is usually much greater than the scope of collection from the data subject himself. or The applications featured in the study collect and process data from third-party social media platforms. The GDPR concerns when a data subject must be notified about data processing by a third party. The principle of transparency is put into practice. Transparency addresses the right of data subjects to know and understand how their data is being used. While it seems relatively easy to notify the data subject when we collect data directly from them, this can be difficult when we collect data from third parties such as: B. Collect data via the Twitter Stream API. The GDPR incorporates several exceptions that codify and specify the conduct of investigations. Exceptions are social media platforms, through appropriate terms and conditions, where data is shared with third parties (e. g. via APIs), or where large amounts of data collected are disproportionate to inform all parties. It assumes that you have already notified the user that effort is required. Once we have lawfully received your personal data, you should check whether we have a legal basis for processing the data. Article 6 of the GDPR lists six legal bases, three of which are relevant to scenarios: (1) data subject consent, (2) carried out in the public interest and (3) Legitimate Interests of Controller. Consent is not required where the processing is in the public interest. Some regions, such as Sweden, allow considerable latitude for research conducted by public institutions. However, this freedom is not available to commercial organizations that rely on it. Consent or Legitimate Interests. Appropriate safeguards must be applied when processing data Especially for sensitive data such as ethnicity and sexual orientation. GDPR also emphasizes the concept of profiling. Profiling is defined as automated analysis to identify correlations and apply those correlations to individuals to identify current or future behavioral characteristics. Given the definitions above, most of the research presented in the previous section can be classified as profiling. Such investigations may require the conduct of a Data Protection Impact Assessment (DPIA). This is usually carried out by a data protection officer.

Data retention is another GDPR concern. If you want to store the collected data after processing, you must ensure that the data subject has access to the stored data and that the data subject can withdraw it (GDPR Article 15)., the right to rectification (Article 16 GDPR), the right to be forgotten (Article 17 GDPR) and the right to restriction of processing (Article 18 GDPR). Consider restrictions on the collection and processing of personal data due to legal obligations arising from the GDPR. Build GDPR-compliant applications and respect data subject rights. We recognize that data subjects must be treated with dignity and respect, regardless of the purpose of the research. Meanwhile, the possibilities of the future are revealed. Notification systems should be designed and developed to enable mass notification of data subjects for data collection and data processing by third parties. Standard and automated tools are expected to meet data subject requests, such as data withdrawal and correction. The development of such tools that allow communication between social media users, who are data subjects, and researchers who process the data, will ease the burden of privacy compliance in future research.

In addition, social media platforms should also implement privacy protection mechanisms to protect user privacy. Care-less posting of content online can lead to serious exposure to confidential information. Studies proposed a privacy frame-work to manage data disclosure behavior to hide sensitive physical profiles. They further improved the framework to ensure high performance while protecting privacy and ensuring fairness among users. Tsai [41] states that an attacker can use a user's profile and social relationships to predict sensitive information from social networks. To prevent this, they pro-posed a data-scrubbing technique that collectively manipulates user-profiles and social relationships to make predictions less predictive of an adversary's sensitive information.

B. Rumor & fake news

With billions of users worldwide, social media is becoming one of the most important sources of information. However, because social media is less controlled and restricted than official news sources, false claims are more common and popular. Spread rumors very easily and quickly. This is a big deal as it can cause economic or political turmoil. Researchers explained that there are two forms of false information. deceive the audience). Rumors on SNS are information that has not been confirmed to be true or false at the time of posting. Rumors can spread misinformation and disinformation. Fake news consists of information that is intentionally and blatantly false and has the motive to mislead readers. Many researchers have focused on identifying rumors and fake news on social media. There are five general approaches:

1) Classification-based approach.

Rumor detection can be viewed as a binary classification problem. Use supervised learning approaches such as Support Vector Machine and Logistics Regression to classify trending topics as fake or not. Researchers use feature engineering to improve the performance of rumor detection classification models. Some studies showed user content-based, user-based, behavior-based, and distribution-based features and uses extracted multimedia-based and location-based features. Researchers proposed a model based on temporal, structural, and linguistic features of rumor propagation.

2) Pattern-based approach.

This approach summarizes patterns associated with rumors and uses matching techniques to do so highlight controversial social media posts. Some other techniques identified rumors based on query phrase patterns.

3) Simulation-based approach.

A microscale generative model of context-driven viral activity has been proposed by some research. In this work, we use the probabilistic discrete event-based simulator PhySense to simulate viral activity cascades in synthesized network topologies and strategies to control such activity cascades via appropriate metrics to investigate the facts.

4) Time series-based approach.

The technique was proposed using temporal characteristics of tweets to quickly detect rumors. Without many features, it can significantly reduce training and prediction times.

5) Crowdsourcing-based approach.

Some other research proposed a rumor control framework, called cloud blocking that allows users to implement control schemes in a collaborative and distributed manner. They designed two incentive mechanisms to motivate more users to actively participate in rumor-blocking activities. Our previous rumor detection efforts provide a solid foundation for the further evolution of social media-based event detection applications. We examined applications developed to enable natural disaster detection and biomonitoring. Future work may integrate the rumor detection procedure into an event detection application and evaluate the results systematically.

C. Image & Video

With the rapid development of internet speed, images and videos are becoming an important part of social media such as Instagram (images) and YouTube (videos). Image analysis is also mature. You can use machine learning and neural networks to automatically annotate images and segment-specific objects in images. Images from social media have been used as training datasets for computer vision projects. Text and geo-tagging make it easier to understand the content of social media images. Hata [42] proposed to classify high-resolution urban remote sensing images using convolutional neural networks and fully convolutional networks, and using labeled social media images as a training dataset, followed by manual labeling. This approach reduces the cost of producing conventional datasets. One can use your existing social media-based applications and develop ways to add image analysis to further improve their performance. For example, emergency management systems can be enhanced by incorporating image classifiers in addition to text analysis.

D. Multilingual Support

Due to economic globalization, most social media channels including Facebook and Twitter are open to all languages and people around the world. Although English has been adopted by many as the default international standard language, people from all countries generally exchange information in their own language. On Twitter, a 2013 study found that only 34% of tweets were in English. Therefore, applications that use social media as a data source may need to support languages other than English. Discussions about events that occur in a country are more likely to be posted in the local language in addition to English, so if a disaster detection application can analyze data in different languages, event convergence will improve. Designers of social media applications with global reach are faced with a dilemma between supporting English only (with no local information) and tediously localizing the application in all languages of interest. Difficult trade-offs are often faced here. Trying to integrate social media information in many languages is even more difficult for programmers and translators. Native language skills are required to extend the application to support additional languages. A machine learning model needs a suitable language training dataset to train a model for a new language. Studies suggested using online machine translation tools such as Google and Bing Translator as a cheaper and faster alternative. They made a systematic comparison to investigate to see design alternatives that use an increased number of manually developed filter stages. Test results show that automated translation can achieve comparable or better results. The study also examined the use of machine translation. To achieve multilingual sentiment classification, [43] used Google Machine Translation to create a comparable Malay dataset from the English dataset, and then used the English and Malay datasets. into a single dataset to create a deep-learning model with multilingual word embed-dings. To develop a cross-language mood classification so that resources in resource-rich languages (such as English) can be used to classify the mood polarity of texts in resource-poor languages (such as Japanese). was intended. The main issue is the lexical gap between languages, and we propose a model for learning expressions between languages, BiDRL, which learns words and sentences of both languages simultaneously. Researchers developed MuSES, a multilingual emotion identification system with a novel zero-effort labeling approach that leverages knowledge bases such as Wikipedia to label word-level emotions for non-English words. Even more interesting is in a multi-ethnic and multicultural country like Malaysia with a population made up of Malays, Chinese, Indians, and other immigrants. People may express their opinions in a mixture of multiple languages. There is an urgent need to support multiple languages in social media-based applications, especially healthcare and disaster management applications. While previous studies have explored the potential benefits of introducing machine translation and creating multilingual word embeddings, further improvements in multilingual sup-port for the timely processing of large amounts of data include more research that needs to be done.

6. Conclusion

Social media offers a wide variety of information that gets exchanged from person to person. This research comprehensively reviewed some of the previous literature on extract-ing value and insight from social media data and practical implementation to evoke useful results. We systematically compared existing applications and categorize them based on analytical techniques and scope of impact. From a practical point of view, we discussed the application of social media data and the impact it has made in the area of research. We discussed some fields of the impact that uses social media data such as healthcare, disaster management, and economics, and hope to foster broader research of such applications in big data analytics as well as social media data mining. Finally, We discussed some of the challenges and perspectives on future research directions in the areas of social media-based applications related to data protection, multilingual support, etc.

References

[1] Kawaljeet Kaur Kapoor, Kuttimani Tamilmani, Nripendra P. Rana, Pushp Patil, Yogesh K. Dwivedi & Sridhar Nerur ”Advances in Social Media Research: Past, Present and Future”pages531-558 (2018)

[2] Brooke Auxier and Monica Anderson “Social Media Use in 2021”Pew Research Center, April 2021, “Social Media Use in 2021”

[3] Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2018). Big Social Media Data Analytics: A Survey. Computers in Human Behavior. doi: 10.1016/j. chb.2018.08.039 10.1016/j. chb.2018.08.039

[4] Stefan Stieglitz Milad Mirbabaie Bjo¨rn Ross Christoph Neuberger Beier & Wagner, ”Social media analytics - Challenges in topic discovery, data collection, and data preparation” International Journal of Information Management (2018)

[5] Data never sleeps 5.0, https://www. domo. com/learn/datanever-sleeps-5, 2020.

[6] E. Ortiz-Ospina, The rise of social media, https://ourworldindata.org/rise-of-social-media, 2019.

[7] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, Social media analytics—Challenges in topic discovery, data collection, and data preparation, International Journal of Information Management, vol.39, pp.156- 168, 2018.

[8] I. Moalla, A. Nabli, L. Bouzguenda, and M. Hammami, Data warehouse design approaches from social media: Review and comparison, Social Network Analysis and Mining, vol.7, no.1, p.5, 2017.

[9] J. W. Cao, K. Basoglu, H. Sheng, and P. B. Lowry, A systematic review of social networking research in information systems, Communications of the Association for Information Systems, vol.36, no.1, 2015

[10] Norjihan Abdul Ghania, , Suraya Hamida, Ibrahim Abaker Tar-gio Hashemb, Ejaz Ahmedc, ”Social media big data analytics: A survey”, 0747-5632/ © 2018 Elsevier Ltd

[11] Qixuan Hou, Meng Ha, and Zhipeng Cai, ”Survey on Data Analysis in Social Media: A Practical Application Aspect” BIG DATA MINING AND ANALYTICS ISSN 2096-0654 02/06 pp259-279 Volume 3,

[12] Number 4, December 2020 DOI: 10.26599/BDMA.2020.9020006

[13] V. A. Rohani, S. Shayaa, and G. Babanejaddehaki, Topic modeling for social media content: A practical approach, presented at 2016 3rd Int. Conf. Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 2016, pp.397-402.

[14] A. Argyrou, S. Giannoulakis, and N. Tsapatsoulis, Topic modelling on Instagram hashtags: An alternative way to automatic image annotation? presented at 2018 13th Int. Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), Zaragoza, Spain, 2018, pp.61- 67.

[15] A. Ejaz, S. K. Fatima, Q. N. Rajput, and S. A. Khoja, Analyzing news from electronic media and topics discussed on social media using ontology, presented at 2018 5th Int. Conf. Social Networks Analysis, Management and Security (SNAMS), Valencia, Spain, 2018, pp.349-354.

[16] Z. Q. Ma, W. W. Dou, X. Y. Wang, and S. Akella, Taglatent dirichlet allocation: Understanding hashtags and their relationships, presented at 2013 IEEE/WIC/ACM Int. Joint Conf. Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 2013, pp.260- 267.

[17] Kim, J., & Hastak, M. (2018). Social network analysis. International Journal of Information Management: The Journal for Information Pro-fessionals, 38 (1), 86-96.

[18] M. Ni, Q. He, and J. Gao, Forecasting the subway passenger flow under event occurrences with social media, IEEE Transactions on Intelligent Transportation Systems, vol.18, no.6, pp.1623-1632, 2017.

[19] E. A. Dahouei, A cloud-based dashboard for time series analysis on hot topics from social media, presented at 2017 Int. Conf. Energy, Com-munication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp.1-6.

[20] C. Comito, D. Falcone, and D. Talia, A peak detection method to uncover events from social media, presented at 2017 IEEE Int. Conf. Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 2017, pp.459-467.

[21] L. Y. Liu, J. L. Priestley, Y. Y. Zhou, H. E. Ray, and M. Han, A2Text-Net: A novel deep neural network for sarcasm detection, presented at 2019 IEEE 1st Int. Conf. Cognitive Machine Intelligence, Los Angeles, CA, USA, 2019, pp.118-126.

[22] S. Desai and M. Han, Social media content analytics beyond the text: A case study of university branding in Instagram, in Proc.2019 ACM Southeast Conf., Kennesaw, GA, USA, 2019, pp.94-101.

[23] J. S. He, M. Han, S. L. Ji, T. Y. Du, and Z. Li, Spreading social influence with both positive and negative opinions in online networks, Big Data Mining and Analytics, vol.2, no.2, pp.100-117, 2019.

[24] Serrat, O. (2017). Social network analysis Knowledge solutions. Springer39-43. Shanthi, C., & Pappa, N. (2017). An artificial intelli-gence based improved classification of two-phase flow patterns with feature extracted from acquired images. ISA Transactions, 68, 425-432.

[25] Cybenko, G. (2017). Parallel computing for machine learning in social network analysis. Paper presented at the parallel and distributed pro-cessing symposium

workshops (IPDPSW), 2017 IEEE international.

[26] Evans, P. (2017). Generation and analysis of a social network. Hamlet. Fan, Z., Bi, D., He, L., Shiping, M., Gao, S., & Li, C. (2017). Low-level structure feature extraction for image processing via stacked sparse denoising autoencoder.

[27] Poria, S., Cambria, E., Hussain, A., & Huang, G. -B. (2015). Towards an intelligent framework for multimodal affective data analysis. Neural Networks, 63, 104-116. Povoda, L., Burget, R., Dutta, M. K., & Sengar,

[28] N. (2017). Genetic optimization of big data sentiment analysis. Paper presented at the signal processing and integrated networks (SPIN), 2017 4th international conference on.

[29] S. Desai and M. Han, Social media content analytics beyond the text: A case study of university branding in Instagram, in Proc.2019 ACM Southeast Conf., Kennesaw, GA, USA, 2019, pp.94-101.

[30] J. S. He, M. Han, S. L. Ji, T. Y. Du, and Z. Li, Spreading social influence with both positive and negative opinions in online networks, Big Data Mining and Analytics, vol.2, no.2, pp.100-117, 2019.

[31] Lin Yue, Weitong Chen, Xue Li, Wanli Zuo & Minghao Yin ”A survey of sentiment analysis in social media” 04 July 2018

[32] I. Perikos and I. Hatzilygeroudis, A framework for analyzing big social data and modeling emotions in social media, presented at 2018 IEEE 4th Int. Conf. Big Data Computing Service and Applications (BigDataService), Bamberg, Germany, 2018, pp.80-84.

[33] M. Han, M. Y. Yan, Z. P. Cai, Y. S. Li, X. Q. Cai, and J. G. Yu, Influence maximization by probing partial communities in dynamic online social networks, Transactions on Emerging Telecommunications Technologies, vol.28, no.4, p. e3054, 2017.

[34] M. Han, Z. J. Duan, C. Y. Ai, F. W. Lybarger, Y. S. Li, and A. G. Bourgeois, Time constraint influence maximization the algorithm in the age of big data, International Journal of Computational Science and Engineering, vol.15, nos.3 & 4, pp.165-175, 2017.

[35] Event-based surveillance, https://www. cdc. gov/globalhealth/ healthpro-tection/gddopscenter/how. html, 2020.

[36] Severe weather 101-Floods, https://www. nssl. noaa. gov/ education/svrwx101/floods/detection/, 2020.

[37] A. Musaev, Z. Jiang, S. Jones, P. Sheinidashtegol, and M. Dzhumaliev, Detection of damage and failure events of road infrastructure using social media, in International Conference on Web Services, H. Jin, Q. Wang, and L. J. Zhang, eds. Seattle, WA, USA: Springer, 2018, pp.134- 148.

[38] D. Murthy and A. J. Gross, Social media processes in disasters: Implications of emergent technology use, Social Science Research, vol.63, pp.356-370, 2017.

[39] S. Desai and M. Han, Social media content analytics beyond the text: A case study of university branding in Instagram, in Proc.2019 ACM Southeast Conf., Kennesaw, GA, USA, 2019, pp.94-101.

[40] Reddick, C. G., Chatfield, A. T., & Ojo, A. (2017). A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government facebook use. Government Information Quarterly, 34 (1), 110-125.

[41] McCaig, D., Bhatia, S., Elliott, M. T., Walasek, L., & Meyer, C. (2018). Text-mining as a methodology to assess eating disorder-relevant factors: Comparing mentions of fitness tracking technology across online communities. International Journal of Eating Disorders.

[42] Attorney general Becerra publicly releases proposed regulations under the California consumer privacy act, https://oag. ca. gov/news/press-releases/attorney-generalbecerra-publicly-releases-proposed-regulations-under California, 2019.

[43] Z. P. Cai, Z. B. He, X. Guan, and Y. S. Li, Collective data-sanitization for preventing sensitive information inference attacks in social networks, IEEE Transactions on Dependable and Secure Computing, vol.15, no.4, pp.577-590, 2018.

[44] Y. Q. Qin, M. M. Chi, X. Liu, Y. F. Zhang, Y. J. Zeng, and Z. M. Zhao, Classification of high-resolution urban remote sensing images using deep networks by integration of social media photos, presented at IGARSS 2018-2018 IEEE Int. Geoscience and Remote Sensing Symp., Valencia, Spain, 2018, pp.7243-7246.

[45] M. J. Fuadvy and R. Ibrahim, Multilingual sentiment analysis on social media disaster data, presented at 2019 Int. Conf. Electrical, Electronics, and Information Engineering (ICEEIE), Denpasar, Indonesia, 2019, pp.269-272