Social media analytics for societies and businesses: Bibliometric analysis

[1]Ms. Madhavi Shamkuwar, [2]Mr. Jayesh Katkar, [3]Mr. Pandurang Patil, [4]Dr. Rahul More

[1,2,3]Assistant Professor, Zeal Institute of Business Administration, Computer Application & Research, Pune

[4]Assistant Professor, Sinhgad Institute of Management & Computer Application, Pune

Abstract:

By 2022, there are expected to be 4.62 billion active social media users, or 58% of the world's population, and this number is expected to increase quickly. Through social media, the consumer in this new era can communicate directly with other people, businesses, and the government. Social media is without a doubt the most abundant source of human-generated text input. Opinions, feedbacks, perspectives, thoughts, and critiques offered by internet users represent attitudes and sentiments toward particular topics, products, companies, or services in many ways. The gigantic amount of hetergogenuos data thus generated through various social media platforms provides a rich and a collaborative way for consumers to stay connected across both public and private forums.

To acquire a broad perspective on social media big data analytics, this study provides an overview of recent studies in social media, data science, and machine learning.

We further investigated the various applications and uses of social media analytics tools inside local government in our research and discovered that social media analytics may be incredibly useful for the government in both exceptional events and everyday operations. The main uses of social media analytics are highlighted, along with prospective research questions and problems that merit investigation, such as enhancing information flow and using analytics for everyday tasks.

To fulfil the need and offer clarity, we develop a taxonomy on social media analytics. This research effort includes discusses supporting data types, tools, and approaches. As a result, selecting the social data analytics that best meet their demands will be simpler for researchers.

Keywords: social media analytics, artificial intelligence, bibliometric analysis, social media methods, social media techniques, social media platforms

I. Introduction

This creates the opportunity to make the “Big Social Data” handy by implementing machine learning approaches and social data analytics.

A comprehensive list of relevant statistical/machine learning methods to implement each of these big data analytics is discussed in this work. Social networks' popularity and use offer a wealth of data that may be used to address a variety of research topics from other fields. However, the study is complicated by the nature of social media data. This study's objective is to present a comprehensive review of the studies that have examined social media data since 2012 to 2022. The conclusion that there are not established nor widely used unambiguous definitions came from a thorough literature study based on 57 research articles. Marketing, hospitality and tourism, disaster management, and disruptive technologies are the key research areas. The vast majority of the social media analysis's data comes from Twitter. The methods now in use are sentiment and content analysis.

II. Systematic literature review

A comprehensive and methodical technique for examining the literature on a specific topic or research topic is called a systematic literature review. It involves a methodical and organized process to find, extract, evaluate and synthesize relevant research to provide an objective and comprehensive overview of the current state of the art. The purpose of the literature review is to provide transparency in the selection and evaluation of studies while minimizing bias.

2.1 PRISMA protocol

Because of the solid foundation of a strong and robust literature review, content analysis must be performed as part of a meta-analysis. The method of quantitative analysis was chosen according to the proposal. PRISMA stands for Preferred Reporting Units for Systematic Reviews and Meta-Analysis. It is a general set of guidelines useful for conducting systematic reviews and meta-analyses. The aforementioned PRISMA protocol consists of a control section and a study flow chart that explains the purpose of systematic reviews and meta-analyses so that the entire process is transparently visualized for stakeholders. The checklist consists of important parts - title, abstract, methods, results, discussion and funding. These components are very useful for search strategy, filtering strategy, data collection, data extraction and analysis, and interpretation of results.

The PRISMA flowchart is a four-step pictorial representation of the process - scientific articles Identification, Screening, Eligibility and Inclusion.

In order to create a solid foundation for a strong and robust literature review, content analysis analysis is performed as part of the meta-analysis. Quantitative analysis methods were selected as proposed.

2.2 VISUAL LITERATURE REVIEW

Research subtopics are highlighted by performing a literature search using various research tools, thereby creating a complete picture of the review at a glance. The tools used for this are Carrot 2, Open Knowledge Maps, and Citation Gecko. Carrot 2 is an open source engine [2] that creates clusters of documents and renders them graphically. It supports five languages: English, French, German, Italian, and Spanish, and supports documents from eight countries: Austria, France, German, United Kingdom, Liechtenstein, Italy, Spain, and Switzerland. Below is a visual representation of Figure 1 created when the search term was "social media analytics."

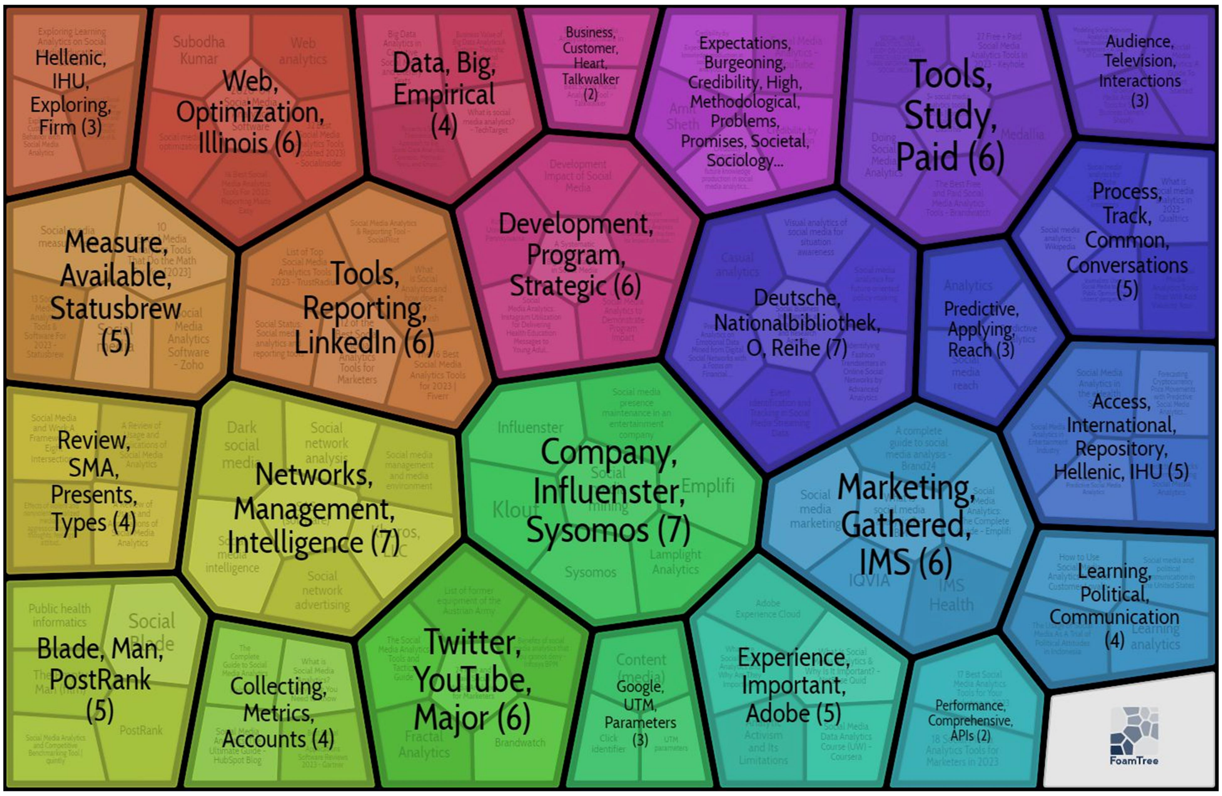

Figure 1: Treemap generated with keyword ‘ social media analytics’ from Carrot2

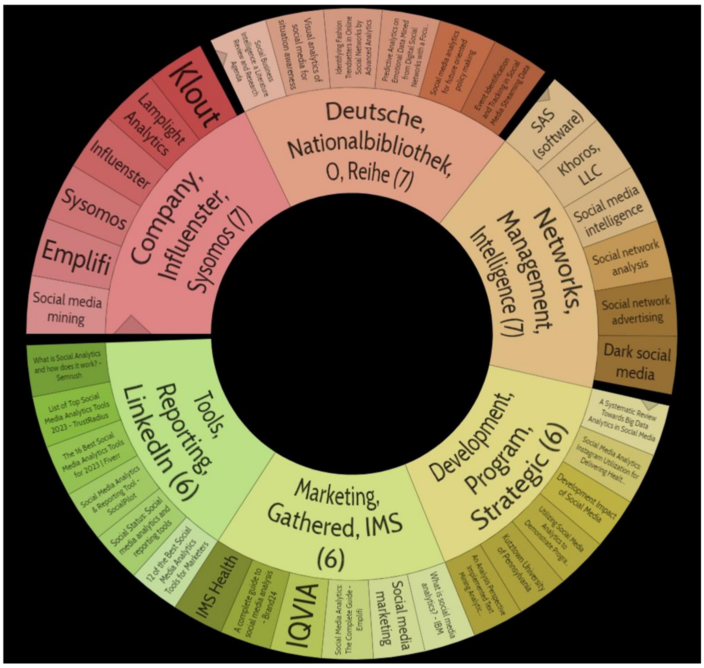

Figure 2: Sunburst chart generated with keyword 'Social media analytics' from Carrot2

Citation Gecko [2] is a graphical user interface designed solely for finding relationships between different articles on related keywords. Below is the visual generated when the search term is "smart agriculture" and all papers are considered "seed papers". As a result, his two visualizations, List of Seed Papers and Clusters of Seed Papers and Related Papers, were created. Visualizations are listed below.

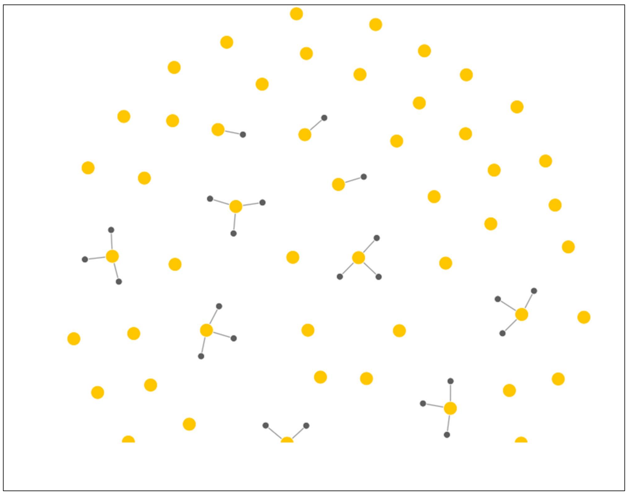

Figure 3: List of seed papers generated with keyword 'Social media analytics' from Citation Gecko

As shown in the figure, there is a seed paper in the center of the cluster, and the paper citing the seed paper is shown in yellow on the other side. Note that in most cases the seed paper is surrounded by many other papers, but in some cases the seed paper is solitary.

I. BIBLIOMETRIC ANALYSIS

The impact and visibility of research articles in a particular subject or field is assessed using bibliographic analysis, a type of quantitative data analysis. Bibliographic analysis [3], a type of quantitative data analysis, is used to assess the impact and visibility of research articles within a particular subject area or field. [4-5]

Database

To begin a bibliographic search, the first step is to determine the appropriate database from which to retrieve relevant documents. To perform a bibliographic analysis, use IEEE Xplore (Institute of Electrical and Electronics Engineers) – a research database containing a diverse collection of research papers in electrical engineering, electronic engineering, computer science and related fields [6]. It consists of 4 million documents, including research papers, conference proceedings, standards, and books [7]. and other publishers. IEEE Xplore accepts data in a variety of formats, including .ris, .bib, and .csv formats. IEEE Xplore performs simple and advanced searches and applies basic query or filter methods. Additionally, the platform provides data in multiple dimensions such as article type, publication year, author, affiliation, publication title, publisher, conference location, default status, default type, publication topic, etc. To do. This allows users to search for terms such as title, title/abstract, journal name, author name, and affiliation, facilitating accurate and targeted searches within the database. Data was extracted from approximately 2000 research papers and downloaded in CSV format.

Search strategy

The search query ‘social media analytics’ from time frame 2001 to 2023 year were being searched for the ‘full text and the meta data’, this was done to gain more dimensions about the query terms. A metric known as centrality measures how much a network of keywords interacts with other keyword networks. It gauges the degree to which various networks are interconnected. Contrarily, density measures a network's internal cohesiveness or strength by examining how closely and intricately interwoven its words are. We may divide the themes into four different groups based on their varied levels of centrality and density by taking both centrality and density into account. [8].

Advance search:

("Document Title": social media analytics) OR ("Document Title": analytics”) AND ("Document Title": big data) AND ("Author Keywords": climate) OR ("Author Keywords": environment, global warming) OR ("Author Keywords": temperature, hot, heat, season, wind, air, rain,

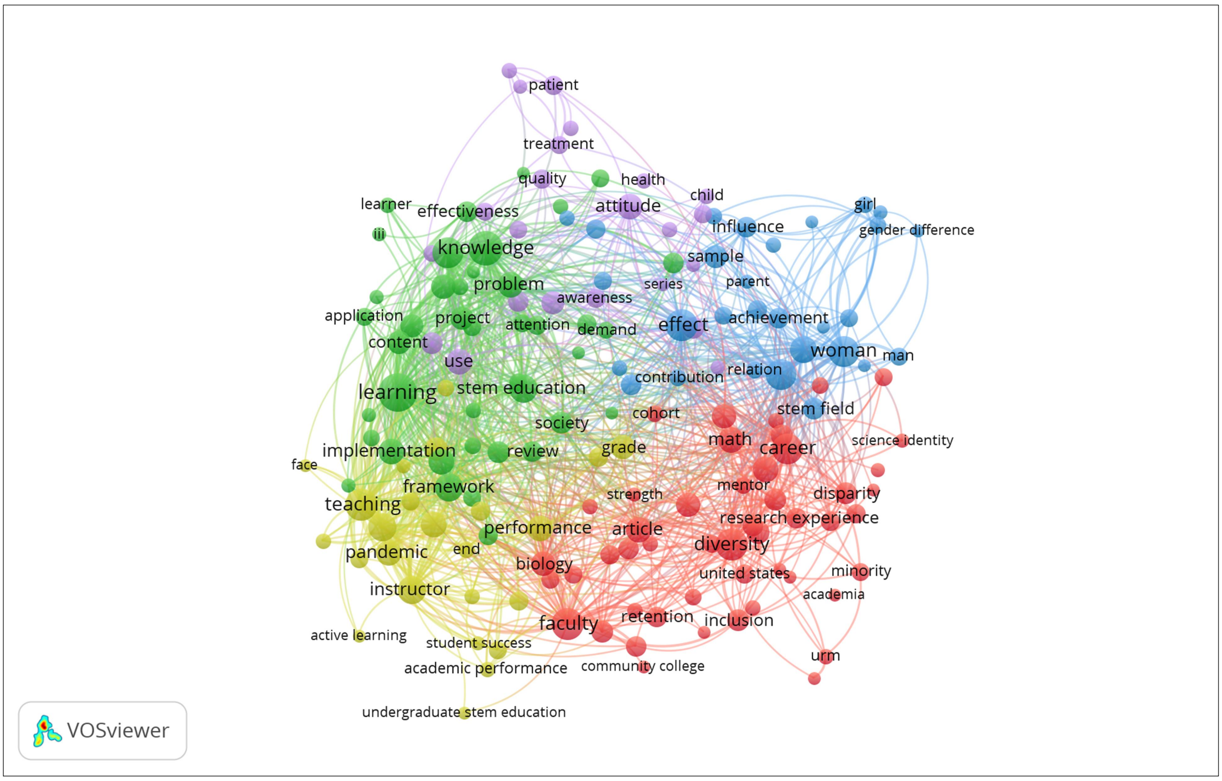

Figure 4: Binary counting 409/845

Data were exported to Microsoft Excel from the received literature. The exported data included information on annual growth in publications, document types, languages, countries, authors, institutions, journals, citations and funding bodies. Purchased materials were also exported to VOSviewer to create network visualization maps. VosViewer software has the ability to build networks between concepts, terms and keywords. VosViewer uses the VOS (Visualization of Similarity) algorithm to generate maps that visually represent these networks. The distance between pairs of objects on a map accurately represents their similarity with high mathematical precision [9]. Use the feature that allows you to create co-occurrence networks using keywords. In these networks, keywords used by authors are represented as nodes. The size of each node reflects the frequency of occurrence of that keyword. Moreover, the connectivity between nodes is proportional to the degree of co-occurrence between related words [10].

Data were exported to Microsoft Excel from the received literature. The exported data included information on annual growth in publications, document types, languages, countries, authors, institutions, journals, citations and funding bodies. Purchased materials were also exported to VOSviewer to create network visualization maps. VosViewer software has the ability to build networks between concepts, terms and keywords. VosViewer uses the VOS (Visualization of Similarity) algorithm to generate maps that visually represent these networks. The distance between pairs of objects on a map accurately represents their similarity with high mathematical precision [9]. Use the feature that allows you to create co-occurrence networks using keywords. In these networks, keywords used by authors are represented as nodes. The size of each node reflects the frequency of occurrence of that keyword. Moreover, the connectivity between nodes is proportional to the degree of co-occurrence between related words 10].

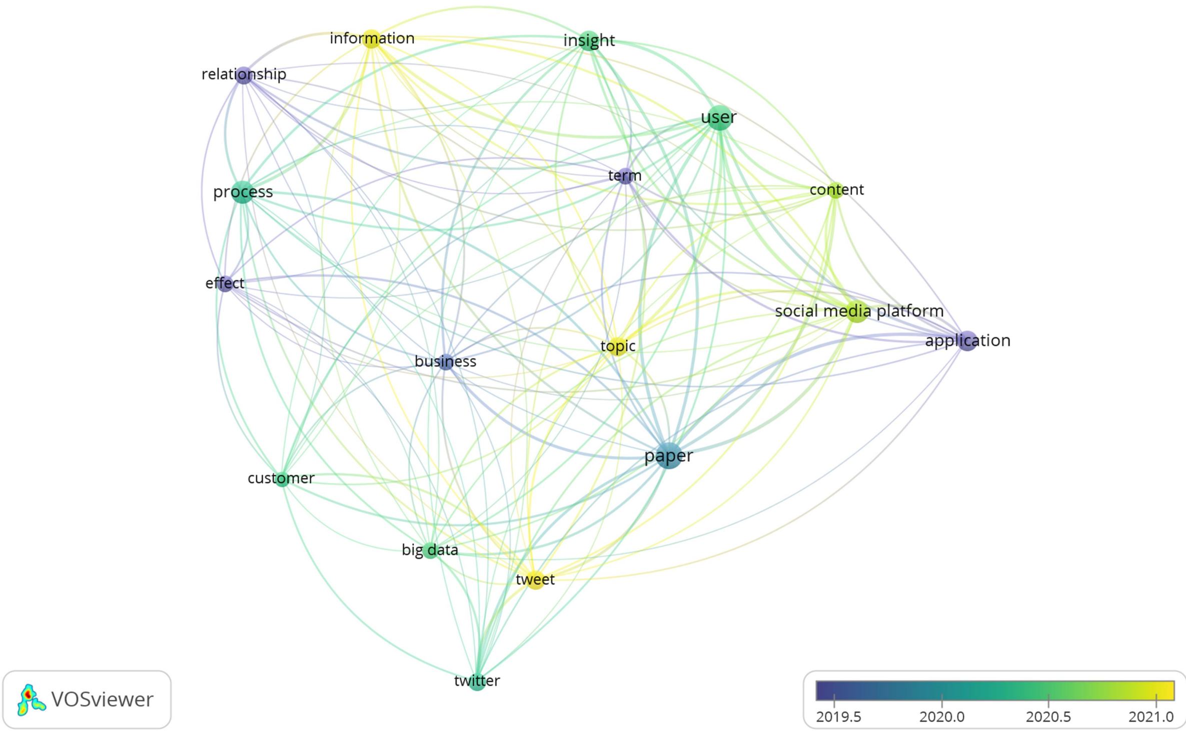

Figure 5: Visualizations from 'Ris' data created from VOSviewer(Citations)

As shown in figure, the three clusters thus formed gives list of items such as:

Cluster no	Cluster color	Name of items arranged in ascending order
1	Red	Academia, Career, disparity, diversity, faculty, inclusion, retention,
2	Green	Application, Learner, framework, knowledge, project, stem education
3	Blue	achievement, contribution, Effect, influence
4	Yellowish Green	instructor, pandemic, performance, Teaching

Table 1: Types of cluster, cluster color and the related significant words

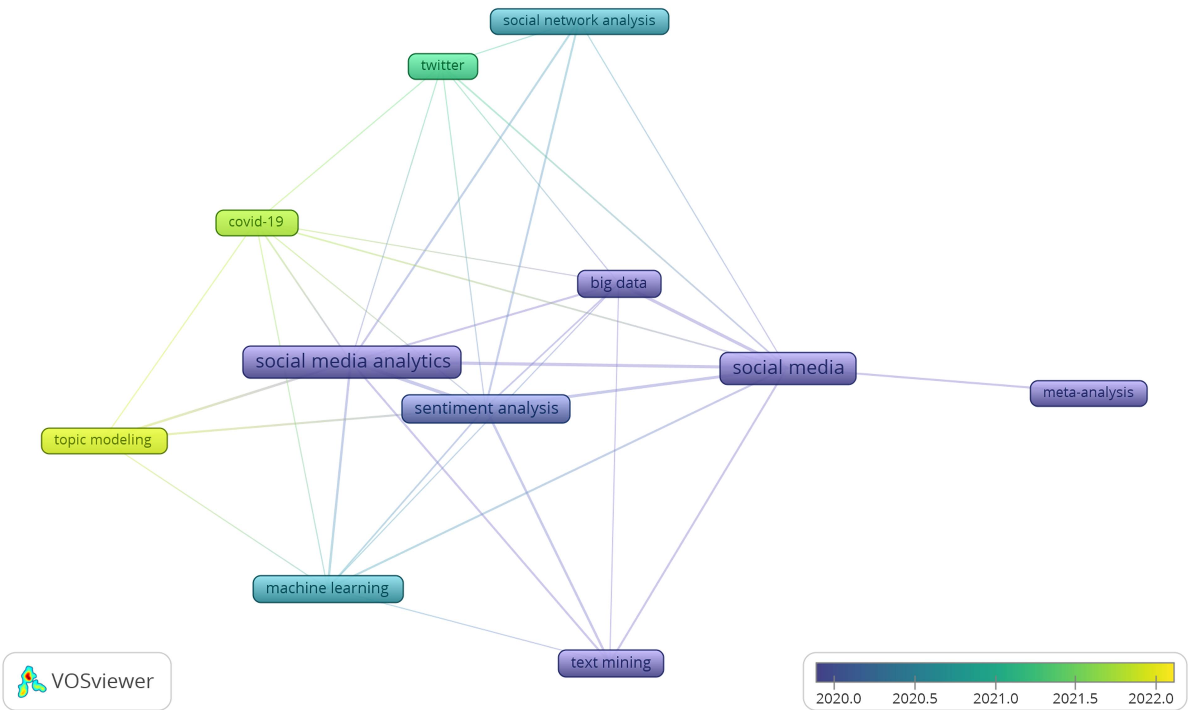

Figure 6: Visualizations from 'Ris' data created from VOSviewer (citations and co occurrences)

As part of the visual representation of the literature search, we decided to determine article citations and patent citation numbers. As a result, the following visualization was created using Tableau Desktop specifically. We found that the number of paper citations gradually increased, and is significant in 2024 and then the downfall can be noted. However, the number of article citations shows a sharp increase in "spikes" in 2014, 2017 and 2019. The figure shows the number of paper citations and patent citations from 2001 to 2020.

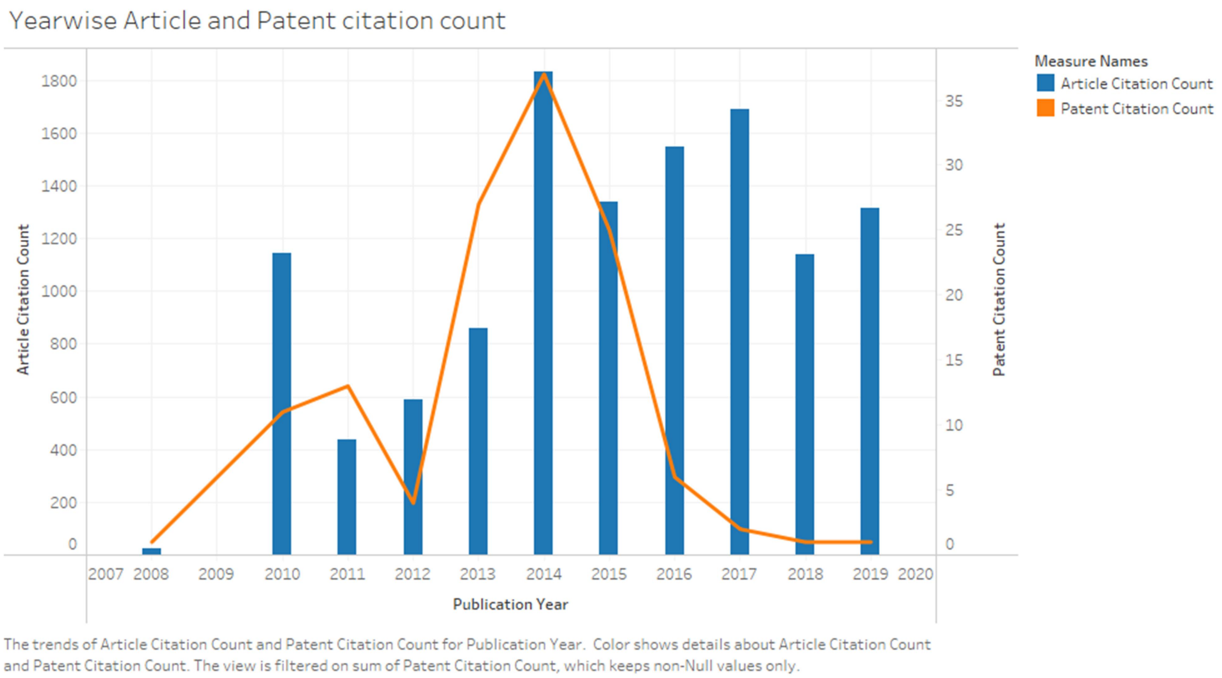

Figure 7: Yearwise Article and Citation count

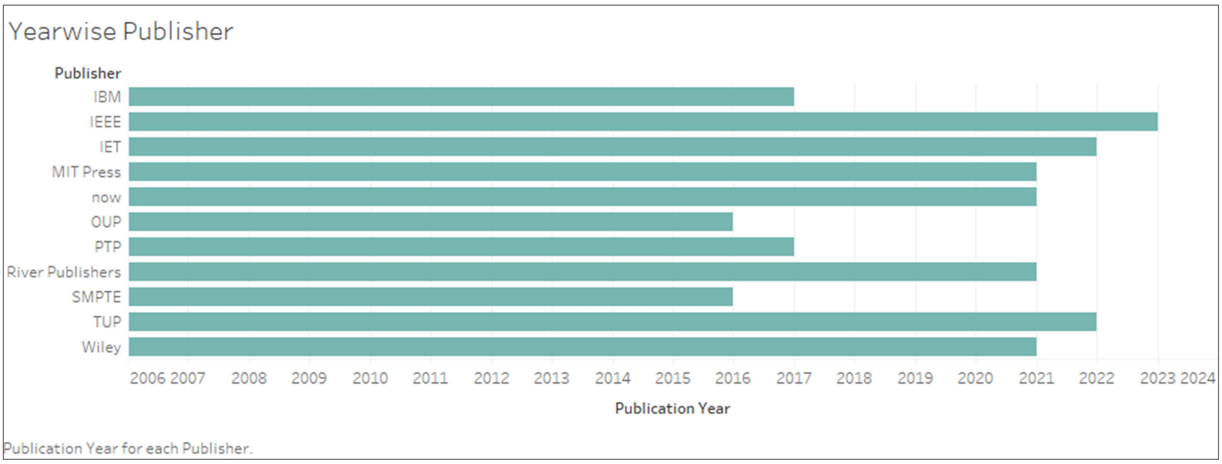

Figure 8: Yearwise Article and Citation count

Applications of Social media analytics for societies and businesses

Natural Disaster

In this study, we introduced ACT, an automated method for real-time social media analysis in emergency situations.

On the incoming streams of data from Twitter and Instagram, ACT uses a variety of analytics. ACT can identify tweets and photographs associated with an event by combining categories and location-based filters, giving users the option to "drill down" to get understanding rather than just viewing a single tweet or image.

In order to speed up the necessary response and recovery measures during natural disasters, ACT also offers users an interactive visualisation platform where they may investigate the events. It also minimises the human labour required to extract pertinent information from social media data. The Australian Red Cross tested ACT throughout the 2013–2014 bushfire season. [11] The Twitter analytics module used with Azure cloud services was the topic of this investigation. We demonstrated the framework's initial implementation using static data. The whole architecture for real-time streaming of Twitter data will be developed and deployed as we intend to establish a Twitter analytics as a service platform. Every recognised service in our architecture will have a microservice created in order to connect to the Azure event hub and stream tweets.

Microservices have some restrictions as well, such as the potential for extra overhead from interprocess communication. We attempted to discover the micro-services that were loosely linked with the least amount of interconnection messages possible using our framework. However, further effort is required to concentrate on the efficiency of the system and the communication between the microservices. [12]

According to the authors, this study indicates potential fairness problems in disaster informatics tasks, which has two implications. This work can provide a set of mitigation strategies to address the fairness issues and can help disaster informatics researchers become aware of potential biases in data sources, analytical models, and outcomes that may impact the well-being of the poor, the elderly, and minorities.

Additionally, this paper gives machine learning researchers the chance to pinpoint actual flaws with fairness in the real world. This paper's practical contribution mostly consists of (1) scenarios and explanations of how biases are introduced in typical disaster analysis methods, and (2) a methodology for assessing and mitigating these biases.[13]

We have gathered tweets from tweepy on the Nepal earthquake and identified keywords for algorithmic analysis. Additional research was conducted using these terms, and the output was an overview of disaster occurrence and its effects. A catastrophe management module was created as a result of automating the identification of keywords for the purpose of further investigation. Future work will involve developing a generic module that can analyse any hashtag.[14]

The study's overall findings demonstrate that sentiment data can accurately reflect people's feelings amid a natural disaster.

Most people have a more upbeat attitude on life. Twitter data has limitless potential, but acquiring and understanding the data has proven to be challenging. By assessing the issues raised by those who are affected and the sentiments of users in reaction to existing response efforts on a real-time basis, this data can significantly contribute in the improvement and optimisation of relief operations here and in other countries. The sentiments of people throughout the tragedy were identified and found to be largely favourable using twitter data from the cities hit by Hurricane Irma.[15]

By combining several data mining approaches with geotagging, a coherent set of interconnected components for extracting situation awareness is built. An API is offered by Twitter and News that can be used to do searches using keywords and hashtags, although it can only access a certain amount of entities. It is necessary to extract these keywords from the public Twitter stream. Filtering methods are employed to lower the noise. NER and the Google Maps Geocoding API were used to determine the postings' locations. In order to determine the sort of disaster and its location with a trueness value of 0, 1, 2, or 3, each Twitter data point was compared with news data from internet sources. We can tell the technique is effective since we can compare Twitter data to news.

In order to analyse earthquakes and characterise the situation from social media messages, this article suggests an exploratory visualisation that is interactive. Users can explore and identify patterns and associated events with the help of the general view and brushing/linking features. The VAST Challenge 2019: Mini-Challenge 3 dataset application of this technique further validates the design decisions of numerous linked views assist users to express the messages from the perspectives of community, location, and number of related posts.[16]

The purpose of this study is to undertake a thorough analysis of the literature in order to determine how artificial intelligence, in particular machine learning, can be used to analyse large amounts of social media data for disaster management. In an initial study using the systematic literature review, we discovered 17 relevant papers that address the goal we were trying to accomplish. We were able to determine from this review that the majority of studies concentrated on text classification and picture classification. Surprisingly, we were unable to locate any research papers on speech recognition or video categorization that were relevant to disaster response. Therefore, more research is required to determine how well AI can be used to analyse large amounts of social media data for disaster response.[17]

The application demonstrates the enormous potential of social media and smartwatches in the healthcare industry. The difficult challenge of predicting mental health illnesses shouldn't solely rely on psychometric exams. Decisions about your health should be made after taking into account a number of considerations.

Results from social media and smartwatches can be combined with information from other sources, such as psychometric exams, to produce accurate results. Mental health professionals can identify the patient's issues and take the required action with the help of this programme.[18] By utilising sentiment-based features from blog posts written by communities' members, this study looks into the issue of identifying hypercommunities across three groups of online communities that contain and exclude members with mental illnesses.

We derive latent themes from the corpus, which was constructed using mood tags, general terms, and affective words in the blog posts written by members of the communities, using the Bayesian nonparametric approach HDP. To identify meta-communities among the communities, we use a nonparametric clustering approach. Using latent themes to visualise the detected meta-communities reveals a division between the groupings.

This shows the evidence of differentiating emotional expression in online forums focused on mental health, offering a potential direction for assistance and intervention.[19]

By examining numerous threads, we were able to characterise and deduce the domain objectives and tasks that OHC administrators try to accomplish in this design research. Our visual analytics solution collected secret thread dimensions throughout the iterative design stages, allowing the study's administrators to investigate, contrast, and locate related threads utilising a variety of views and interactions. Our study highlights the significance of documenting the analytical unit that impacted the layout of our main view and work flow. By offering a combinatorial prototype where users could freely come up with and validate combinations of metrics and visual encodings, we also shared our lessons on design research technique.[20]

This research focuses on the latter part and describes how a socially influencing tool for quitting smoking was designed using the PSD model. The suggested system develops unique computer-mediated user-interactive features that tap into the power of social relationships using the social influence principles stated in the PSD model. By utilising social diffusion mechanisms seen in online social communities, these features hope to encourage significant network affiliations between users and intervention information, which will lead to an increase in health-related behaviours. [21]

Furthermore, according to 61% of poll participants, how the text is written matters. This study also acknowledges that there may be additional elements, relating to the characteristics of the material, which are leading to good post performance because 33% of respondents believe that the text's emotions are significant. We therefore intend to assess content features (both for textual material and photos), applied to a Facebook dataset that contains 153 public health organisations, as part of our ongoing research.[22]

Conclusion

One of the main conclusions from our study is that health care organisations frequently use social media because of its user base's popularity and to promote their messages more quickly. Practical ramifications are included in half of the research. On the basis of the literature review, concise definitions are offered, and potential directions for top-notch future study are indicated. The future lies in the implementation of the Social media analytics.

References:

1. **https://carrotsearch.com/**
2. **https://www.citationgecko.com/**
3. Sweileh WM, Wickramage K, Pottie K, Hui C, Roberts B, Sawalha AF, Zyoud SH. Bibliometric analysis of global migration health research in peer-reviewed literature (2000-2016). BMC Public Health. 2018;18(1).
4. Aria and Cuccurullo, 2017 M. Aria, C. Cuccurullo, bibliometrix: an R-tool for comprehensive science mapping analysis, J. Informetr., 11 (2017), pp. 959-975
5. Van Eck and Waltman, 2019 N.J. Van Eck, L. Waltman, VOSviewer Manual. Version 1 .6.13, Universiteit Leiden (2019), https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.13.pdf
   A. Durniak, "Welcome to IEEE Xplore," in IEEE Power Engineering Review, vol. 20, no. 11, pp. 12-, Nov. 2000, doi: 10.1109/39.883281.
6. S. Sinha, B. Lacquet and B. T. J. Maharaj, "IEEE Xplore digital library indexes the Transactions of the SAIEE, 1909 to date," in SAIEE Africa Research Journal, vol. 112, no. 4, pp. 158-159, Dec. 2021.
7. Cobo et al., 2015 M.J. Cobo, M.A. Martínez, M. Gutiérrez-Salcedo, H. Fujita, E. Herrera-Viedma, 25 years at Knowledge-Based Systems: a bibliometric analysis Knowl.-Based Syst., 25th anniversary of Knowledge-Based Systems, 80 (2015), pp. 3-13
8. Van Eck and Waltman, 2014 N.J. Van Eck, L. Waltman, Visualizing bibliometric networks, Y. Ding, R. Rousseau, D. Wolfram (Eds.), Measuring Scholarly Impact, Springer Inter. Pub., Cham (2014), pp. 285-320

9. Tang et al., 2002 S. Tang, Q. Zhu, X. Zhou, S. Liu, M. Wu, Aconception of digital agriculture, International Geoscience and Remote Sensing Symposium (IGARSS). Toronto, Ont. (2002), pp. 3026-3028
10. Yong et al., 2002 L. Yong, L. Xiushan, Z. Degui, L. Fu, The main content, technical support and enforcement strategy of digital agriculture Geo Spatial Inf. Sci., 5 (2002), pp. 68-73
11. [W. Sherchan, S. Pervin, C. J. Butler, J. C. Lai, L. Ghahremanlou and B. Han, "Harnessing Twitter and Instagram for disaster management," in IBM Journal of Research and Development, vol. 61, no. 6, pp. 8:1-8:12, 1 Nov.-Dec. 2017, doi: 10.1147/JRD.2017.2729238.]
12. A. Khaleq and I. Ra, "Cloud-Based Disaster Management as a Service: A Microservice Approach for Hurricane Twitter Data Analysis," 2018 IEEE Global Humanitarian Technology Conference (GHTC), San Jose, CA, USA, 2018, pp. 1-8, doi: 10.1109/GHTC.2018.8601887.
13. Y. Yang, C. Zhang, C. Fan, A. Mostafavi and X. Hu, "Towards Fairness-Aware Disaster Informatics: an Interdisciplinary Perspective," in IEEE Access, vol. 8, pp. 201040-201054, 2020, doi: 10.1109/ACCESS.2020.3035714.
14. B. Shah, V. Agarwal, U. Dubey and S. Correia, "Twitter Analysis for Disaster Management," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4, doi: 10.1109/ICCUBEA.2018.8697382.
15. I. Vayansky, S. A. P. Kumar and Z. Li, "An Evaluation of Geotagged Twitter Data during Hurricane Irma Using Sentiment Analysis and Topic Modeling for Disaster Resilience," 2019 IEEE International Symposium on Technology and Society (ISTAS), Medford, MA, USA, 2019, pp. 1-6, doi: 10.1109/ISTAS48451.2019.8937859.
16. H. N. Nguyen and T. Dang, "EQSA: Earthquake Situational Analytics from Social Media," 2019 IEEE Conference on Visual Analytics Science and Technology (VAST), Vancouver, BC, Canada, 2019, pp. 142-143, doi: 10.1109/VAST47406.2019.8986947.
17. 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," 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Sendai, Japan, 2018, pp. 1-4, doi: 10.1109/ICT-DM.2018.8636388.

18. N. Amate, S. Patil, P. Jojan and S. Morankar, "Use of Social Media and Smartwatch Data Analytics for Mental Health Diagnosis," 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2021, pp. 1-6, doi: 10.1109/ICITIIT51526.2021.9399591.
19. B. Dao, T. Nguyen, S. Venkatesh and D. Phung, "Nonparametric discovery of online mental health-related communities," 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 2015, pp. 1-10, doi: 10.1109/DSAA.2015.7344841.
20. B. C. Kwon, S. -H. Kim, S. Lee, J. Choo, J. Huh and J. S. Yi, "VisOHC: Designing Visual Analytics for Online Health Communities," in IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, pp. 71-80, 31 Jan. 2016, doi: 10.1109/TVCG.2015.2467555.
21. S. Myneni and S. Iyengar, "Socially Influencing Technologies for Health Promotion: Translating Social Media Analytics into Consumer-facing Health Solutions," 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 2016, pp. 3084-3093, doi: 10.1109/HICSS.2016.388.
22. N. Straton, R. Vatrapu and R. R. Mukkamala, "Facebook and public health: A study to understand facebook post performance with organizations' strategy," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 3123-3132, doi: 10.1109/BigData.2017.8258288.