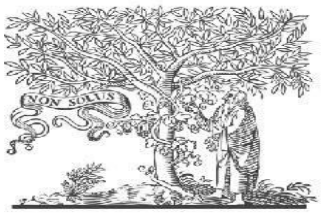
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

Link : https://ijiemr.org/downloads/Volume-12/Issue-05

10.48047/IJIEMR/V12/ISSUE05/27

Title **A Systematic Review Towards Big Data Analytics in Social Media**

Volume 12, Issue 05, Pages: 267-278

Paper Authors

MJ. Ravichandra Reddy, Abdul Raqeeb, . Abdul Samad, Md. Noman Khan, Viqar Khan

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A Systematic Review Towards Big Data Analytics in Social Media .

1. **J. Ravichandra Reddy**, Assistant Professor, MTech(Phd), Department of CSE, (Vijay Rural Engineering College(VREC)), ravichand48@gmail.com

2. **Abdul Raqeeb,** BTech, Department of CSE, (Vijay Rural Engineering College(VREC)), raqeebk836@gmail.com

3. **Abdul Samad**, B.Tech, Department of CSE, (Vijay Rural Engineering College(VREC)), abdsam5055@gmail.com

4. **Md. Noman Khan**, B.Tech, Department of CSE, (Vijay Rural Engineering College(VREC)), nomantcj@gmail.com

5. **Viqar Khan,** B.Tech, Department of CSE, (Vijay Rural Engineering College(VREC)), viqarkhan.987@gmail.com

ABSTRACT: The current breakthrough in internet 2.0 technology opens up the possibility of connecting individuals all over the globe via society 2.0 and web 2.0 technologies. This new age enables consumers to communicate directly with other people, businesses, and the government. People are willing to share their thoughts, points of view, and ideas on any issue in a variety of media. This opens up the possibility of making "Big Social Data" useful by integrating machine learning methodologies and social data analytics. To acquire a broad perspective on social media big data analytics, this research provides an overview of current studies in social media, data science, and machine learning. We show why social media data are important components of a better data-driven decision-making process. By integrating 5 V's and 10 Bigs, we propose and create the "Sunflower Model of Big Data" to describe big data and bring it up to speed with technology. We identify the top 10 social data analytics for use on social media sites. This paper discusses a thorough list of applicable statistical/machine learning approaches for implementing each of these big data analytics. To date, "Text Analytics" is the most often utilised analytics in social data analysis. To address the demand and give a clear understanding, we develop a taxonomy on social media analytics. This research paper includes discusses tools, approaches, and supporting data types. As a consequence, researchers will have an easier time determining which social data analytics would best meet their requirements.

Keywords *– big data, social media, big data analytics, social media analytics.*

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

1. INTRODUCTION

Big data has evolved into a useful resource. Big data is used everywhere, from social networks to academics, healthcare, aerospace, transportation planning, oil and gas development, to telecommunications, e-commerce, banking and insurance, military and surveillance, and a variety of other industries. However, this vast volume of data will only become a value if we can figure out how to make data communicate. Data analytics is the instrument that allows data to tell stories in a clear and understandable way. According to Brandwatch, a well-known social listening platform, over 3.4 billion individuals globally used social media sites in May 2019. This social media-based platform generates a tremendous volume of structured, semi-structured, and unstructured data in a short period of time. This is because a social media-based platform enables speedier information sharing, offers text, picture, audio, and video sharing, and allows an individual user to connect with a large number of other users at the same time. Surprisingly, social media has lately grown in popularity to the point that many people utilise it as their primary communication channel to report to the public or emergency officials. Because of its ease of use and worldwide reach, the CMO (Chief Marketing Officer) of many large commercial companies began to reply to questions presented on social media.

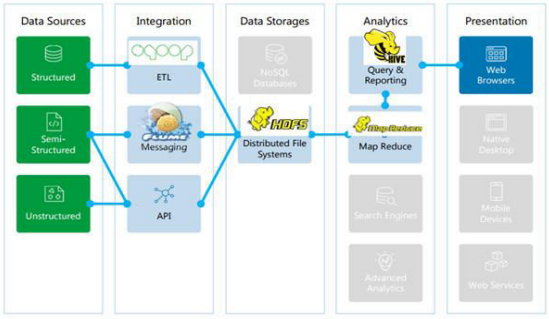


Fig.1: Example figure

According to the statistics, 40.8 percent responded through Twitter, 26.2 percent via Facebook, and 16.5 percent via LinkedIn[8]. As a consequence, massive volumes of data are becoming the norm for reflecting civilisation across the planet. In recent years, some firms have made considerable investments in social media-driven decision-making processes, making this medium a popular option for consumer data research and company improvement[9]. It allows firms to participate in instant customer exposure, maybe in the most premium manner conceivable, with more efficacy than traditional marketing services and technology. In several areas, the capacity to assess, correlate, and learn from enormous volumes of data is becoming more important for making predicted decisions.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

2. LITERATURE REVIEW

Beyond the hype: Big data concepts, methods, and analytic:

When the term "big data," the first and, at times, only dimension that comes to mind is size. This work seeks to provide a fuller definition of big data that includes its other distinguishing and distinguishing properties. Industry's quick expansion and acceptance of big data has pushed the debate to popular channels, requiring the scholarly press to catch up. Academic publications in a variety of areas have yet to address the issue, despite the fact that they might benefit from a meaningful discussion of big data. This study provides a unified definition of big data by combining concepts from practitioners and academia. The major emphasis of the study is on big data analysis approaches. This study is notable for its emphasis on analytics linked to unstructured data, which accounts for 95% of big data. The need of developing suitable and efficient analytical tools to harness vast amounts of heterogeneous data in unstructured text, audio, and video forms is highlighted in this research. This report also emphasises the need of developing new predictive analytics tools for structured large data. In practise, statistical techniques were developed to infer from sample data. The heterogeneity, noise, and vast amount of structured big data necessitate the development of computationally efficient algorithms capable of avoiding big data hazards like false correlation.

Social media big data analytics: A survey:

Because of the ubiquity of the Internet and the development of Web 2.0 technologies, big data analytics has lately arisen as a significant study subject. Furthermore, the growth and use of social media apps has given scholars and practitioners with several possibilities and problems. The vast volume of data created by social media platform users is the consequence of the integration of their personal information and everyday activities. This massive amount of created data, dubbed "big data," has lately received a lot of attention. To get a wide perspective on the social media big data analytics research issue, an overview of current works is offered. We categorise the literature based on crucial factors. This research also compares several big data analytics methodologies and their quality characteristics. Furthermore, we examine the applications of social media big data analytics by emphasising cutting-edge approaches, methods, and quality aspects of diverse research. Open research problems in large data analytics are also discussed.

A survey on big data analytics using social media data:

Analytics is critical in all industries for making choices based on specific information. The practise of gathering information from numerous social media platforms, websites, and blogs is known as social media analytics. This analytics is performed in order to get sound business judgements. In today's environment, using social media has become the current trend. Social data analytics is more than simply gathering likes and comments posted by people; it has also become a platform for numerous businesses to promote themselves. Social data was frequently utilised in applications such as marketing and elections to generate prediction choices. Some of the tactics used include developing hypotheses, delving deeply into data, charting occurrences, and so on. This analytics may also be used in applications such as business, legislative changes, education, and demonetization. The issues include measurements generated by social media reaching the correct individuals and unstructured data being difficult to interpret. This study addresses the model, topic, performance assessment, benefits, and drawbacks as part of a literature review.

The role of artificial intelligence in social media big data analytics for disaster management–initial results of a systematic literature review:

When a catastrophe happens, people who are both directly and indirectly impacted by the event often upload massive amounts of data (e.g., photographs, text, audio, video) on a variety of social media sites. This is due to the fact that social media has lately become a main communication medium for individuals to report to the public or emergency personnel (ERs). ERs from different emergency response organisations (EROs) normally consider gaining knowledge of the situation in order to react to a catastrophe that has happened. However, when a crisis occurs, social media platforms are swamped with different types of data, which overwhelms ERs with large data within minutes. Furthermore, the bulk of the data in this supplied data may be repetitive and meaningless. As a result, it becomes difficult for ERs to make sense of and make choices based on the available huge data. Despite recent technological developments, collecting and interpreting disaster-related social media large data remains a difficult undertaking. As a result, the purpose of this study is to give an early analysis of a systematic literature review on the use of artificial intelligence to analyze/process social media large data for effective disaster management. 68 papers were discovered using a

systematic review method. Following that, we examined all of the articles that had been identified. We infer from our study that the majority of the evaluated publications focus on text and picture classification, with convolutional neural networks being the most often used classification method.

Understanding customer experience diffusion on social networking services by big data analytics:

Customers are increasingly engaged with social networking services (for example, Facebook and Twitter). Most companies, in particular, are attempting to benefit from such social networking services, as it has rapidly evolved into an information carrier for clients, conveying the most recent information about goods and services. As a result, this research investigates how information supplied by businesses is transmitted and what the critical aspects in understanding information dissemination are. More crucially, this research categorises the sorts of tweets sent by a corporation and then investigates the influence of these tweets on dissemination. This research classified three categories utilising con- tent analysis: I information provision (IF), I I advertising (AD), and I I I both (IFAD), each with eight particular ideas. These findings indicate that the disparities for all three categories of information content are considerable. It demonstrates that by offering the IFAD type rather than the AD type, businesses may distribute information more rapidly.

3. METHODOLOGY

According to data, 40.8 percent responded through Twitter, 26.2 percent via Facebook, and 16.5 percent via LinkedIn. As a consequence, massive volumes of data are becoming the norm for reflecting civilisation across the planet. In recent years, some firms have made considerable investments in social media-driven decision processes, making this platform a popular alternative for consumer data research and company development. It allows firms to participate in instant customer exposure, maybe in the most premium manner conceivable, with more efficacy than traditional marketing services and technology. In several areas, the capacity to assess, correlate, and learn from enormous volumes of data is becoming more important for making predicted decisions.

Disadvantages:

1. Effective methods and analytical tools are required for reviewing the ever-increasing data offered by different social media apps.
2. The quantity of research conducted on social media has expanded considerably.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

To acquire a broad perspective on social media big data analytics, this research provides an overview of current studies in social media, data science, and machine learning. We show why social media data are important components of a better data-driven decision-making process. By integrating 5 V's and 10 Bigs, we propose and create the "Sunflower Model of Big Data" to describe big data and bring it up to speed with technology. We identify the top 10 social data analytics for use on social media sites. This paper discusses a thorough list of applicable statistical/machine learning approaches for implementing each of these big data analytics. To date, "Text Analytics" is the most often utilised analytics in social data analysis. To address the demand and give a clear understanding, we develop a taxonomy on social media analytics. This research paper includes discusses tools, approaches, and supporting data types.

Advantages:

1. It will be easy for researchers to choose which social data analytics will best meet their goals.
2. We show why social media data are important components of a better data-driven decision-making process.

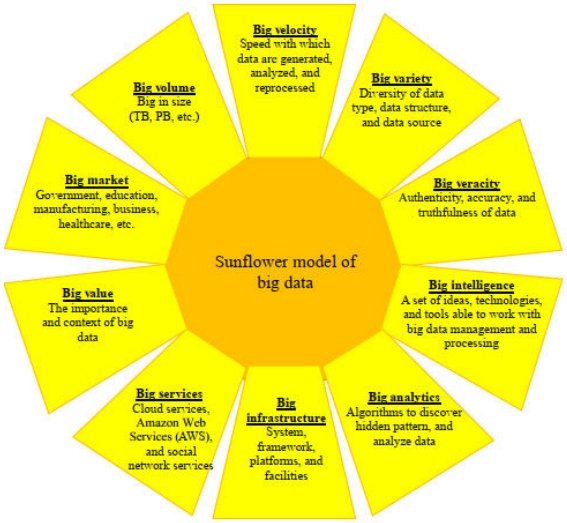


Fig.2: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Data exploration: we will put data into the system using this module. Processing: we will read data for processing using this module.
- Data splitting into train and test: Using this module, data will be split into train and test.
- Create models such as LR, RF, Adaboost, SGD, KNN, DT, NB, SVM, MLP, Gradient boosting, voting classifiers, LSTM, RNN, and CNN. Calculated algorithm accuracy.

- User signup and login: Using this module will result in registration and login. User input: Using this module will result in predicted input.
- Prediction: final predicted shown

4. IMPLEMENTATION

ALGORITHMS:

LR: Logistic regression is a Machine Learning classification approach that predicts the likelihood of certain classes based on various dependent variables. In summary, the logistic regression model computes the logistic of the outcome by adding the input characteristics (in most situations, there is a bias component).

RF: A Random Forest Method is a supervised machine learning algorithm that is widely used in Machine Learning for classification and regression issues. We know that a forest is made up of many trees, and the more trees there are, the more vigorous the forest is.

Adaboost: Any machine learning algorithm may benefit from the usage of AdaBoost. It works well with slow learners. On a classification task, these are models that reach accuracy slightly above random chance. Decision trees with one level are the best suited and hence most often used algorithm with AdaBoost.

SGD: Stochastic Gradient Descent (SGD) is a simple yet very efficient method for fitting linear classifiers and regressors to convex loss functions such as (linear) Support Vector Machines and Logistic Regression.

KNN: The k-nearest neighbours method, often known as KNN or k-NN, is a non-parametric, supervised learning classifier that employs proximity to classify or predict the grouping of a single data point.

DT: A decision tree is a graph that illustrates every potential outcome for a given input using a branching mechanism. Decision trees may be hand-drawn or generated using a graphics application or specialist software. When a group has to make a decision, decision trees may help concentrate the debate.

NB: A naïve Bayes classifier is an algorithm that classifies things using Bayes' theorem. Naïve Bayes classifiers presume high, or naive, independence between data point properties. Spam filters, text analysis, and medical diagnosis are all common applications for naive Bayes classifiers.

SVM: A support vector machine (SVM) is a deep learning technique that uses supervised learning to classify or predict data groupings. Supervised learning systems in AI and machine

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

learning give both input and intended output data, which are labelled for categorization.

MLP: A feedforward artificial neural network that creates a set of outputs from a set of inputs is known as a multilayer perceptron (MLP). An MLP is defined by numerous layers of input nodes that are linked as a directed graph between the input and output layers.

Gradient boosting: Gradient boosting is a sort of boosting in machine learning. It is based on the assumption that the best next model, when merged with past models, minimises the total prediction error. The main concept is to define the desired outcomes for this next model in order to reduce error.

voting classifier: A voting classifier is a machine learning estimator that trains many base models or estimators and predicts by aggregating their results. Aggregating criteria may be coupled voting decisions for each estimator output.

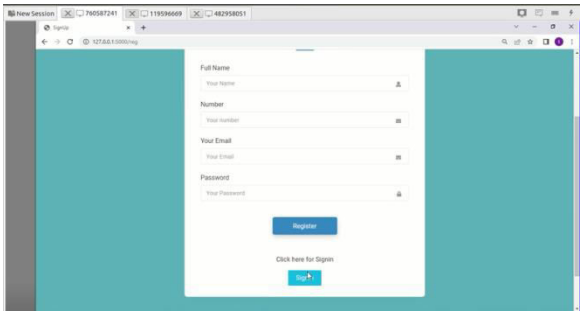
LSTM: LSTM is an abbreviation for Long-Short Term Memory. In terms of memory, LSTM is a sort of recurrent neural network that outperforms standard recurrent neural networks. LSTMs perform far better when it comes to learning specific patterns.

RNN: Recurrent neural networks (RNNs) are the cutting-edge method for sequential data, and they are employed in Apple's Siri and Google's voice search. It is the first algorithm to recall its input thanks to its internal memory, making it ideal for machine learning issues involving sequential data.

CNN: A CNN is a kind of network design for deep learning algorithms that is primarily utilised for image recognition and pixel data processing applications. There are different forms of neural networks in deep learning, but CNNs are the network design of choice for identifying and recognising things.

5. EXPERIMENTAL RESULTS

Fig.3: Home screen



International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig.4: User signup

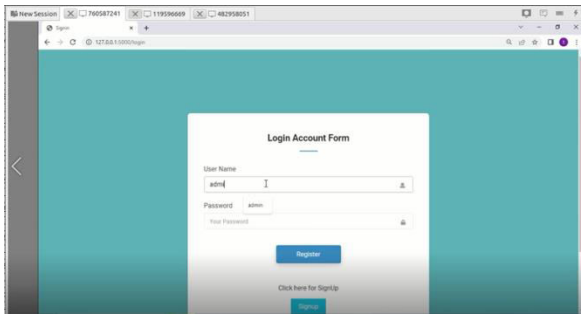


Fig.5: User signin

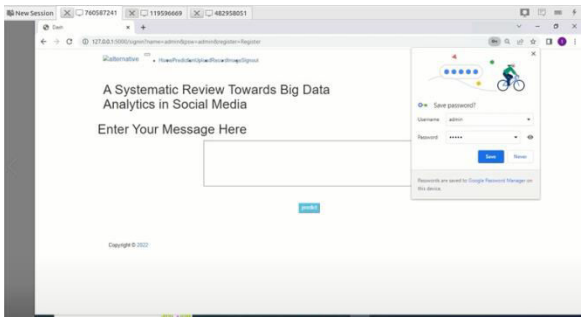


Fig.6: Main screen

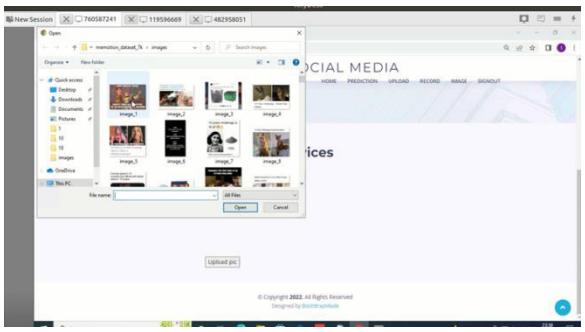


Fig.7: User input

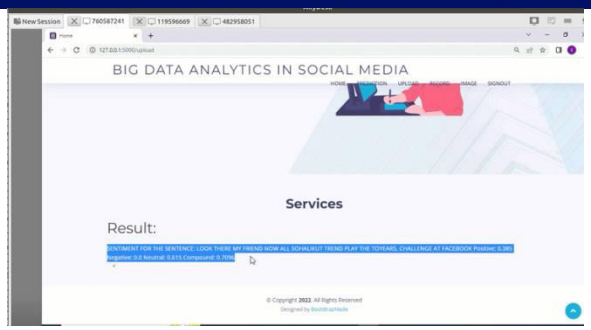


Fig.8: Prediction result

6. CONCLUSION

Big data has emerged as a crucial data analytics tool for understanding human behaviour through analysing data from social media, thanks to developments in computer technologies. Social data, data science, and social data analytics may assist many sorts of organisations, from business to government. The purpose of this study is to address a research vacuum by identifying the 10 most generally accepted and utilised big data analytics for evaluating social data and making judgements. Given the overlap in social media analytics methodologies, we created a taxonomy of big data analytics in the social media sector. We establish three major categories and then allocate these 10 analytics to each one based on the objective, type of use, and working location. Machine learning methods help in data analysis in social media. Each of these social data analytics is coupled with a large number of machine learning or statistical approaches. We provide social data analytics as well as the

techniques for doing so. Until recently, "Text Analytics" was the most extensively used analytics for social data analysis. Furthermore, academics are improving their capacity to extract relevant information from social data images, audio, and video. Because it enables individuals to share their thoughts on the most recent event, product, tools, skills, and other issues, social media platforms regularly supply data to examine. We must capitalise on the advantages of this vast volume of data.

7. FUTURE SCOPE

This study takes a wide, general approach to big data analytics in social media. To achieve the same goal, a specialised topic of interest, such as business analytics in social media, geospatial/location based analytics, social media data analysis for political science research, and so on, might be investigated. We will continue our analysis by concentrating on a few social media networks, including Facebook, Twitter, and Snapchat. Any of the 10 big data analytics listed above may be further investigated. We do not have the time or resources to thoroughly study each big data analytics' list of machine learning methods. We want to continue working on this project in order to develop a shortlist of suitable algorithms for each of the 10 big data analytics categories. We also want to identify and agree on a few common attributes/characteristics of big data analytics so that we may fine-tune one analytics and undertake comparative performance study.

REFERENCES

[1] V. Dhawan and N. Zanini, Big data and social media analytics, Res. Matters A Cambridge Assess. Publ., no. 18, pp. 36–41, 2014. [2] K. Smith, 126 amazing social media statistics and facts, https://www.brandwatch.com/blog/amazing-socialmedia-statistics-and-facts/, 2019.

[3] A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods, and analytics, Int. J. Inf. Manage., vol. 35, no. 2, pp. 137–144, 2015.

[4] N. A. Ghani, S. Hamid, I. A. Targio Hashem, and E. Ahmed, Social media big data analytics: A survey, Comput. Human Behav., vol. 101, pp. 417–428, 2019.

[5] P. V. Paul, K. Monica, and M. Trishanka, A survey on big data analytics using social media data, in Proc. 2017 Innov. Power Adv. Comput. Technol. (i-PACT), Vellore, India, 2017, pp. 1–4.

[6] F. Shaikh, F. Rangrez, A. Khan, and U. Shaikh, Social media analytics based on big data, in Proc. 2017 Int. Conf. Intell. Comput.

Control. (I2C2), Coimbatore, India, 2017, pp. 1–6.

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

[8] F. Piccialli and J. E. Jung, Understanding customer experience diffusion on social networking services by big data analytics, Mob. Networks Appl., vol. 22, no. 4, pp. 605–612, 2017.

[9] P. Ducange, R. Pecori, and P. Mezzina, A glimpse on big data analytics in the framework of marketing strategies, Soft Comput., vol. 22, no. 1, pp. 325–342, 2018.

[10] P. Grover and A. K. Kar, Big data analytics: A review on theoretical contributions and tools used in literature, Glob. J. Flex. Syst. Manag., vol. 18, no. 3, pp. 203–229, 2017.