International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 04 | April -2020 ISSN: 2582-3930

Analysis of Social Media Content Using Text Mining On Big Data: Literature Review (Survey)

Ankita A. Kathane[1], Vivek B. Kute[2]

[1]Student, Department of Computer Engineering, St. Vincent Pallotti College of Engineering Nagpur,India
[2]Associate Professor, Department of Computer Engineering, St. Vincent Pallotti College of Engineering Nagpur,India

---

Abstract -Social media platforms are a boon for companies to reach out to their audiences. Thus, it is important to identify and rectify the users and current trends in society. Each passing day brings in a new concept in lifestyle and technology and thus it is important to get to the very basics of the utilization of technology to yield the maximum profit. Striking the right tune when it comes to social media content is crucial for branding and growing the product. Moreover, it helps in monetizing the aspects of brands by fetching the companies, an appropriate target audience's data. The purpose of this research paper is to determine the progress done in the field of text mining and text analytics on social media content as well as the big data. The outcomes of this survey research paper aim to find out more opportunities in classifying the unstructured data to get a proper record of the target audience. It also aims to elucidate how, by the use of text mining on big data of the various social media sites, one can get to better production and organization of his or her products and make the optimum decisions.

***Key Words*:**Text mining, big data, social media, big data analytics, text analysis, social media data, data mining, social media analysis.

1.INTRODUCTION

Social media consists of an innumerable amount of data that needs to be sorted to get a fruitful outcome from it. this data is in the form of either structured or instructed data. Most of the data that is present on social media is instructed. Processing such data to yield maximum outcome from it is next to impossible if it is not sorted or strutted. The sites that are on full running on social media get hundreds and thousands of visitors each day and they search on various topics. This means that they enter a query in the natural language or the layman language that is not understood by the machines. If we take a closer look at this process, there are some specific keywords that people tend to search for, however, they enter it in the natural language that humans speak. They enter in complete sentences as it comes to their mind hit the search button. This sentence is a query, which may contain one two or three main keywords and some subsidiary words. The other parts of the sentence that has the adjective, pronouns and other parts are of no use to the machine. Our agenda while doing this research is to find out or develop if needed, algorithms that do this job easily and efficiently. The outcomes of this research will give the site owners the maximum keywords hits for a specific period, which will help them in creating more relatable content for the audience that visits their page.

2. BACKGROUND

Data mining on big data has become a huge trend these days with the incorporation of other technologies such as text mining and artificial intelligence. The data that is found on social media sites is mostly unstructured or semi-structured. To reap maximum benefit from it, it needs to be analyzed. However, before analyzing it, it needs to be fetched. Here comes the role of text mining. While the data is in the form of human language, the machine needs to be trained about what is data is necessary to process the information and what part of the data can be omitted or in other words is unnecessary. Text mining, text analysis, natural language processing are the parts, which deal with the above issue. On the other hand, sentiment analysis is used to determine the tone of the user. This comes in the picture when the data that is fetched needs to be analyzed to yield a fruitful outcome.

2.1 Text Mining

Text mining is a process in which the unstructured data is processed to derive high-quality data or information from the text. This means that the information that is being extracted needs to be of use to the user. However, it is often confused with data mining, which is a similar process but not the same. The difference between text mining and data mining is that text mining focuses on unstructured data and data mining focuses on structured data. Text mining approaches can further use machine learning and artificial intelligence to advance the application of text mining

2.2 Text Analytics

Although text mining and text analytics aim to solve the same problem i.e. fetching the unstructured data to get more quality outputs, there is some difference between them. Text mining is an approach that identifies the information within the text and then gives out quality results. As for the text analytics is concerned, it gives out information that is easy to read visually- graphs, tables, and reports.

2.3 Natural Language Processing (NLP)

Natural language processing aims at mapping the human entered language i.e. the natural language into machine language i.e. the language that is processed by the system. Various algorithms are used to do this process in natural language processing (NLP). Furthermore, natural language processing can be used for predicting the behavior of the audience that is visiting a specific page. It works by fetching the input data and retrieving necessary information in the form of words that have a meaning. These words can further be used or manipulated as per requirements.

2.4 Big Data

The term "big" is specifically used in the term big data to make us realize that the amount of data that it consists of is huge. Big data consists of a huge amount of data that is both structured and unstructured. We used the term big data in our research because the data that the web contains involves both, structured as well as unstructured data. As a matter of fact, big data is said to be so enormous that it is next to impossible to process it using traditional techniques.

2.5 Sentiment Analysis

The next big thing that comes out when social media and text mining that is related to human behaviors is sentiment analysis. Sentiment analysis is a process of computing, identifying and categorizing opinions expressed in a piece of text. This is done especially to determine whether the writer's attitude towards the topic, product, etc. is positive, negative, or neutral. This is very essential when it comes to analyzing the text extracted from the text mining process. Using sentiment analysis, the brands can focus more on the quality and demand of the product they are selling. This also helps in deriving the feedback of the consumer through their reviews, which further leads to enhancement in the quality of products and services.

2.6 Categories Of Social Media [1]

The term "social media" comprises various categories that are present on the world wide web and are used by people for various purposes such as connecting, sharing personal experiences, sharing information and knowledge, buying, selling and renting items, etc. At the core, each site is a collection of web pages that has information in either structured, semi-structured or unstructured data. These are dynamic web pages and they can take some information from the user as well as at the backend, they also map information about the visitor.

Most common social media platforms:

Social networking sites – Facebook, Twitter, Instagram, LinkedIn, Tinder, etc.

Blogging – Wordpress, blogger, community blogs, etc.

Bookmarking sites – Pinterest, Tumblr, digs, Flipboard, etc.

However, if one observes closely, all the sites are hybrid and while they sever the main purpose of connectivity and information sharing, they broadly have similar characteristics. While studying these aspects, we try to hit on similar characteristics to yield results that can be applied to all of them.

3. EXISTING WORK IN TEXT MINING AND TEXT ANALYTICS

There is a wide area of research that needs to be considered when it comes to text mining. Different languages[2], sentence structure, word meanings are some of the aspects that need to be explored. Moreover, there are various fields of applications[3] other than social media that need text mining techniques and can be used to get more data in terms of usage. Health care is one of the leading industries in it wherein AI uses sentiment analysis and text analysis to determine or diagnose a disease. It is also important to extract logical patterns[4] from the unstructured data that has been mined and make a proper prediction of it. It is also seen that the classification [5] of the mined data is essential to make comparisons that lead to a required output. Various algorithms are used to make the extraction and analysis of the text useful.

4. CONCLUSIONS

Observing the huge amount of data that is present on the world wide web in an unstructured format, it is necessary to get it a format that yields us some benefit. Text mining algorithms that are available are of great use however, with each passing day as the technology advances, the renovation of the algorithms needs to take place. More speed, accuracy, and lesser costs are needed for the execution of this process. Having said that, we aim to develop an algorithm that makes the use of visitor's data and the keywords that they search to find out what kind of information they want. Furthermore, we intend to develop a system that fetches the high-quality information for the content creators to analyze and utilize these keywords, their frequency, and their urgency to procedure content related to the consumer's demands. Moreover, we also tend to fetch the timings, areas, and interests of the user depending on their activity.

REFERENCES

[1] Current Trends in Text Mining for Social MediaTajinder Singh*, MadhuKumari, TriveniLal Pal and AhsanChauhan*Department of Computer Science & Engineering, National Institute of Technolgy, Hamirpur, H.P, 177005, India.*

[2] A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives
Said A. Salloum1, 2 *, Mostafa Al-Emran3, Azza Abdel Monem4, Khaled Shaalan1
1Faculty of Engineering & IT, The British University in Dubai, UAE.
2University of Fujairah, UAE.
3Faculty of Computer Systems and Software Engineering, University Malaysia Pahang, Malaysia.
4Faculty of Computer and Information Sciences, Ain Shams University, Egypt.

[3]Applications of Social Media Text Analysis
AtefehFarzindar Diana Inkpen
NLP Technologies Inc.
Université de Montréal
farzindar@nlptechnologies.ca
University of Ottawa
diana.inkpen@uottawa.ca

[4] A Survey on Text Mining in Social Networks
R I Z W A N A I R F A N 1, C H R I S T I N E K. K I N G1, D A N I E L G R A G
E S1, S A M E W E N1, S A M E E U. K H A N1, S A J J A D A. M A D A N I2, J
O A N N A K O L O D Z I E J3, L I Z H E W A N G4 , D A N C H E N5, A M M A
R R A Y E S6, N I K O L A O S T Z I R I T A S4, C H E N G - Z H O N G X U4, A
L B E R T Y. Z O M A Y A7, A H M E D S A E E D A L Z A H R A N I8, and H O
N G X I A N G L I9
1North Dakota State University, Fargo, North Dakota, USA;

email: rizwana.irfan@ndsu.edu, christine.k.king@ndsu.edu, daniel.grages@ndsu.edu, sam.ewen.2@ndsu.edu, samee.khan@ndsu.edu;
2COMSATS Institute of Information Technology, Islamabad, Pakistan;
email:madani@ciit.net.pk;
3Cracow University of Technology, Cracow, Poland;
email:jkolodziej@uck.pk.edu.pl;
4Chinese Academy of Sciences, China;
email:lzwang@ceode.ac.cn;
email:cz.xu@siat.ac.cn;
email:nikolaos@siat.ac.cn;
5China University of Geosciences, Wuhan, China;
email:chendan@pmail.ntu.edu.sg;
6CISCO Systems, San Jose, CA, USA;
email:rayes@cisco.com;
7University of Sydney, NSW, Australia;
email:albert.zomaya@sydney.edu.au;
8King Abdulaziz University, Saudi Arabia;
email:asalzahrani@kau.edu.sa;
9University of Louisville, KY, USA;
email:h.li@louisville.edu;

[5] A SURVEY ON TEXT ANALYTICS AND CLASSIFICATION TECHNIQUES FOR TEXT DOCUMENTS
NiharRanjan, Abhishek Gupta, IshwariDhumale, PayalGogawale, and *RugvedGramopadhye
Department of Computer Engineering, Sinhgad Institute of Technology and Science, SavitriBaiPhule Pune University, Pune, India

[6] Extracting Information from Social Network using NLP
CharuVirmani
Research Scholar, YMCA UST, India.
Dr. AnuradhaPillai
Ymcaust, Faridabad, India.
Dr. Dimple Juneja
NIT, Kurukshetra, India.

[7] Text Mining with Information Extraction
Raymond J. Mooney and Un Yong Nahm
Department of Computer Sciences,
The University of Texas, Austin, TX 78712-1188

[8] Review of social media analytics process and Big Data pipeline

[9] Text Mining: Techniques, Applications, and Issues
RamzanTalib_, Muhammad KashifHanify, ShaeelaAyeshaz, and Fakeeha Fatima
Department of Computer Science,
Government College University, Faisalabad, Pakistan

[10] Text Mining for Big Data Analysis in Financial Sector: A Literature Review
MirjanaPeji´c Bach 1,*, ŽivkoKrsti´c 2, SanjaSeljan 3 and LejlaTurulja 4
1 Faculty of Economics & Business, University of Zagreb, 10000 Zagreb, Croatia
2 Atomic Intelligence, 10000 Zagreb, Croatia; zivko.krstic@live.com
3 Faculty of Humanities and Social Sciences, Information and Communication Sciences, University of Zagreb,
10000 Zagreb, Croatia; sanja.seljan@ffzg.hr
4 School of Economics and Business, University of Sarajevo, 71000 Sarajevo, Bosna i Hercegovina;
lejla.turulja@efsa.unsa.ba
* Correspondence: mpejic@efzg.hr

[11] USING TEXT MINING TO CLASSIFY RESEARCH PAPERS
Assoc. Prof. Ph.D. Snezhana Sulova1
Chief Assist. Prof. Ph.D. Latinka Todoranova2
Assist. Prof. Ph.D. Bonimir Penchev3
Assist. Prof. Radka Nacheva4
1 University of Economics - Varna, Bulgaria
2 University of Economics - Varna, Bulgaria
3 University of Economics - Varna, Bulgaria
4 University of Economics - Varna, Bulgaria

[12] R. Rajendra and V. Saransh, “A Novel Modified Apriori Approach
for Web Document Clustering,” International Journal of Computer
Applications, pp. 159–171, 2013.

[13] K. Sumathy and M. Chidambaram, “Text mining: Concepts, applications,
tools and issues-an overview,” International Journal of Computer
Applications, vol. 80, no. 4, 2013.

[14] P. J. Joby and J. Korra, “Accessing accurate documents by mining
auxiliary document information,” in Advances in Computing and
Communication Engineering (ICACCE), 2015 Second International
Conference on. IEEE, 2015, pp. 634–638.

[15] Z. Wen, T. Yoshida, and X. Tang, “A study with multi-word feature
with text classification,” in Proceedings of the 51st Annual Meeting of
the ISSS-2007, Tokyo, Japan, vol. 51, 2007, p. 45.

[16] V. Gupta and G. S. Lehal, “A survey of text mining techniques and
applications,” Journal of emerging technologies in web intelligence,
vol. 1, no. 1, pp. 60–76, 2009.

[17] R. Agrawal and M. Batra, “A detailed study on text mining techniques,”
International Journal of Soft Computing and Engineering (IJSCE) ISSN,
pp. 2231–2307, 2013.

[18] D. S. Dang and P. H. Ahmad, “A review of text mining techniques
associated with various application areas,” International Journal of Science and Research (IJSR), vol. 4, no. 2, pp. 2461–2466, 2015.

[19] R. Steinberger, “A survey of methods to ease the development of
highly multilingual text mining applications,” Language Resources and evaluation, vol. 46, no. 2, pp. 155–176, 2012.

[20] A. M. Cohen and W. R. Hersh, “A survey of current work in biomedicaltext mining,” Briefings in bioinformatics, vol. 6, no. 1, pp. 57–71, 2005.