

# Extract Genuine Healthcare Posts on Social Media



K. Alakananda, K Rajendra Prasad, C Raghavendra

**Abstract:** Social media plays an important role in spreading the news. People who search for health online and those who rely on social media for small health issues and diet management are increasing periodically. Nowadays, food habits are greatly influenced by social media. Health-related data such as home remedies, diet management, and beauty tips are mainly focused in this paper. Such data available in social media may be genuine or might be not, just because of business strategy to promote products suggested. There is a huge scope for misleading vital content in this scenario. This paper gives an overview on how to process data using machine learning techniques and/or deep learning techniques available on social media by applying social media analytics and revile trustworthy information at one place. As well as describes how to create a platform for genuine information about health care.

**Keywords:** Social media analytics, Health Care, Machine Learning, Deep Learning.

## I. INTRODUCTION

Many posts related to health on social media have no scientific approval which may immediately affect people in critical condition, the people who are very busy with their hectic work, and some emergency conditions like unavailability of doctors. In all cases the people don't have the right option and they completely fed on social media for the sake of cure or instant short term relief. But, very few cases it is successful and the majority of them are not happy with this sort of procedure. There is some genuine information, which has to be extracted and focused on.

According to dot Health nationwide survey [1], 1,509 online respondents responded across America on health-related searches. As per the survey 53% of people search online before consulting a doctor, among those, 63% are women and 52% are men as shown in below fig 1.

A new study describes more than 92% of web users across the world are doing research on product or service before going to buy and more than 84% of users, top it.

A survey conducted by the Pew Internet and American Life project, people who search for a specific disease is 63%, about medical treatment is 47%. Other than these, they also prefer to search for a diet, nutrition, and vitamins is 44% and content related to exercise, fitness is 36%. Other searches are alternative treatments, depression, and anxiety or stress management.

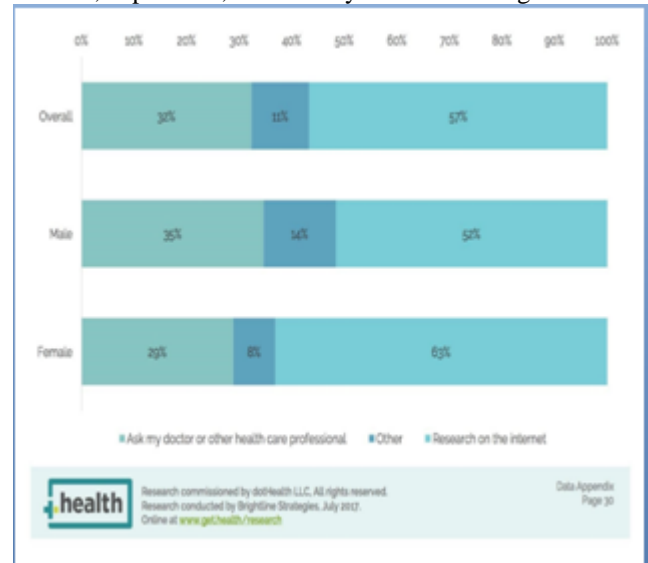


Fig 1: Researching specific conditions or questions

According to the above statistics, most people browse the internet for health-related issues. But there is a lot of fake news on social media because there are some people who want to promote it as a part of business strategy. Among these posts, home remedies, diet management, and fitness are worthy, which can be easily accessed, which cures and sometimes prevent diseases in the long run. It is more effective in problems related to skin, bodyweight management, and hair loss and hair therapy. To check efficiency of these health care issues, analytics is required to analyze public comments, feedbacks which will give good health measures to the people in the society [2,3].

This can be achieved in two models. The first model is to extract data from social media through data analytics and validate it. The second model is to create a new separate platform for genuine information. This paper gives an overview in two phases, firstly, how data is extracted through social media analytics and provide genuine information in a website in SOCIAL MEDIA DATA EXTRACTION AND ANALYSIS is discussed in model I and secondly, how to create a new platform that is website or application with genuine information is discussed in model II that is CREATING A NEW SOCIAL MEDIA NETWORK AND ANALYSING.



Revised Manuscript Received on October 30, 2019.

\* Correspondence Author

**K Alakananda**, Dept. of CSE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: kasampravin@gmail.com

**Dr. K Rajendra Prasad\***, Professor and Head, Dept. of CSE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: krprgm@gmail.com

**C Raghavendra\***, Asst. Professor, Dept. of CSE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: crg.svch@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

### II. SOCIAL MEDIA DATA EXTRACTION AND ANALYSIS

To extract data and processing is known as social media analytics. Tools are built using social media analytics to achieve a specific task. There are some built-in-tools to get data like keyHole, vaizle, etc. Languages supporting these social analytics are Python, R, etc.

#### Data Extraction

Extract health-related data from social media using official API or scraping or crawling. API is specific to a particular social media network. Scraping is getting data from the HTML page of web page. Crawling is getting data from URLs and links associated with subpages. Among these API is much effective [4]. This process called Data extraction. Here we extract posts, comments, likes, etc, according to specific social media network. We extract data from social media based on health-related keywords.

#### Data Cleaning

In general, while extracting data from social media is contains noise. To remove noise from data we use some noise removal techniques. It includes pre-processing, text normalization, duplication removal.

#### Data Analysis

Once getting noisy fewer data from the web that data is analyzed using machine learning and deep learning techniques [5]. In machine learning this process comes under unsupervised analysis. Collect the right data from social media it is required to use natural language processing (NLP) techniques such as text analytics, Graph mining techniques. The reason for using NLP techniques is across the world people use different languages and symbols. To analyze these texts it is necessary to use NLP techniques. Using the above techniques every post is analyzed and grouped as per the health issue it cures [6,7].

The metric to measure the home remedy or cure related to a health issue is based on public review through comments, likes, and shares. These are processed using above techniques. If the review by the public for the health tip is considered as more than 50% sorted out positively.

Based on critics (doctors or scientists) review, a cure may be considered. If review is less than 50% positive then it is verified by critics. Sometimes based on poor reviews due to lack of awareness, in such cases the ayurvedic doctor's opinion and comments will help for coming out from these issues as home remedies [8].

#### Processed Data

The trained dataset is obtaining after applying NLP techniques, this valid data sets separated from fake news is made available to the public through a separate website or an application [9].

In this place, health remedy has to be arranged as per health issue it cures. Again, it has to be arranged as per different stages of a specific disease like starting stage, normal stage, advanced stage. For this stages of diseases cures are mapped as per the reviews or comments given by the public that is, at which stage they got cured by this remedy. It reduces risk of browsing misleading content [10].

This kind of classification is achieved through machine learning or deep learning. Here user can browse data disease wise and stage-wise. The stage and disease are determined by certain symptoms. These symptoms mapped with information available in database about the disease. So user can get the perfect information. A large amount of data is gathered in database for obtaining more accuracy. For the process of data elicitation web scraping can be used and stored it into database. This kind of approach makes accurate remedy available to browsers.

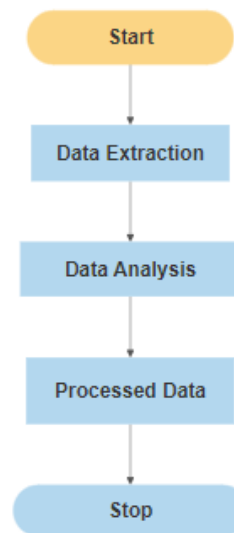


Fig 2: Flow chart representing model-I

### III. CREATING A NEW SOCIAL MEDIA NETWORK AND ANALYSING

Model II is creating a new social media platform especially for health care and it may be an application or website. The advantage of this platform from others is due to the features given below.

This platform allows the browser to post related posts for health issues. This website provides genuine information which helps to cure health issues. This will helps the people to cure health issue which should be strongly recommended by patients. This is the main difference from other bogus health care websites. By allowing users to provide feedback, rating, likes, comments, etc. which will enhance efficiency and update accurate information on the web.

If the rating given by users for a particular home remedy or cure is less than 50% or negative comments are more than positive comments immediately the home remedy is removed.

Here people can browse the remedy disease wise and the stage-wise. Here, the users need to enter the symptoms and can get conclude the stage of disease. The suitable resultant information will be verified in the database and appeared on user screen. This is not suitable for normal and simple health issues.

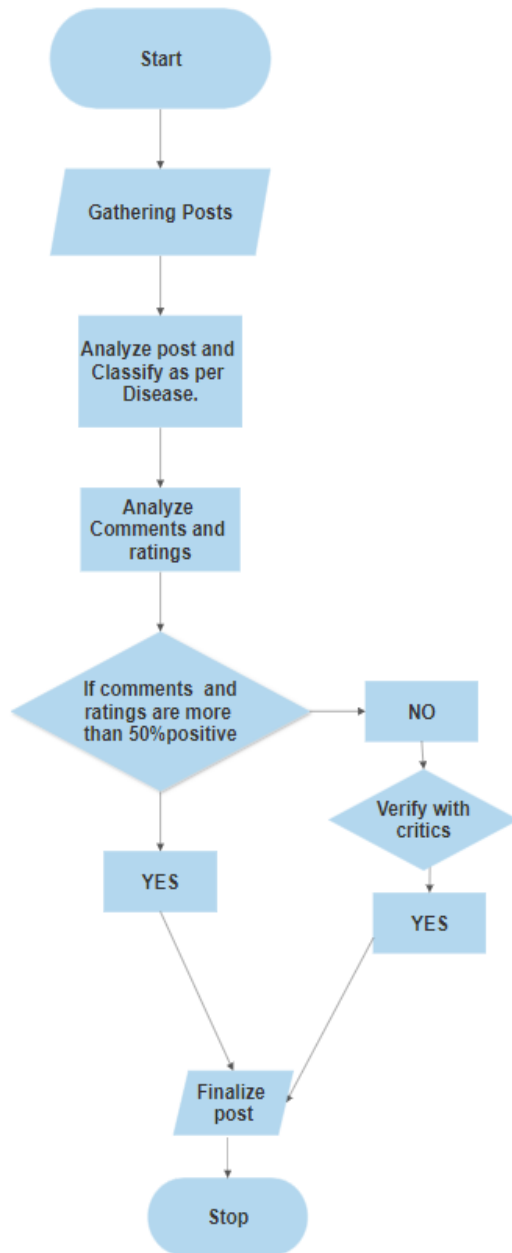


Fig 3: Workflow diagram for model-II

The difference between model-I and model-II is, in model-I, we are gathering posts from various social media platforms and process it for genuine information and making it available as per health issue.

Whereas in model-II, posts are gathered from web users across the world (newly created) and process it for genuine information which is available as per disease and stage of disease.

## IV. CONCLUSION:

This work gives an overview to make genuine health care information available for people either by extracting data using social media analytics or by creating a new social media network. This requires a database to store the huge amount of data, machine learning, deep learning, natural language processing, social media analytics, and other tools for achieving accurate or genuine information from web.

## REFERENCE:

1. www.get.health/research
2. K. Rajendra Prasad, C. Raghavendra, K Sai Saranya, "A Review On Classification Of Breast Cancer Detection Using Combination Of The Feature Extraction Models", International Journal of Pure and Applied Mathematics, Volume 116, No. 21, September 2017, pp 203-208.
3. C. Raghavendra, K. Rajendra Prasad, "Effective Mammogram Classification using Various Texture Features", Journal of Advanced Research in Dynamical and Control Systems, Vol. 9. Sp Issue, December 2017.
4. <https://www.healthline.com/health-news/how-fake-health-news-may-be-influencing-you-to-make-dangerous-decisions#How-fake-health-news-hurts-your-health>
5. <http://www.nbcnews.com/id/3077086/t/more-people-search-health-online/#.XVwdWeMzbiU>
6. <https://www.pm360online.com/how-consumers-find-and-use-online-health-related-content-in-2017/>
7. <https://www.healthline.com/health/real-or-fake-medical-information-online#1>
8. <https://www.statepress.com/article/2018/02/spmagazine-home-remedies-vs-prescription-meds>
9. <https://www.sciencedirect.com/science/article/pii/S1877050915000964>
10. "Python Social Media Analytics", Siddhartha Chatterjee, Michal Krystianczuk, Packt, Birmingham-Mumbai, 2017.

## AUTHORS PROFILE



**Ms. K Alakananda** is pursuing her B.Tech degree in Institute of Aeronautical Engineering, Hyderabad. Her interested areas are Big data and Data mining.



**Dr. K Rajendra Prasad** is professor and head in the department of CSE. He received B.Tech degree in CSE from JNT University, Hyderabad; M.Tech degree in CSE from VTU, Belgaum; Ph.D. in CSE from JNT University, Anantapur. He has 14 years of indispensable experience in teaching. Previously he served as a member in Board of Studies where he was actively involved in TEQIP work. He has published 15 papers in various international conferences and 8 papers in reputed journals. He is a member of CSI, IEEE professional societies. His major research interests are Data Mining, Pattern Recognition and Speech Processing.



**Mr. C. Raghavendra** Graduated in B.Tech (CSE) from Jawaharlal Nehru Technological University, Hyderabad in 2008. He received Master's Degree in M.Tech (CSE) from Bharath INSTITUTE OF Higher Education and Research (BIHER), Chennai, in 2011. He pursuing a Ph.D. in Computer Science & Engineering from BIHER, Chennai. Presently, he is working as Assistant Professor of CSE Dept., Institute of Aeronautical Engineering, Hyderabad. He has published more than 20 Publications in various International Journals and conferences. His research interests are Image processing & Security, Big data and Databases. He is a life member of CSI, IAENG, CSTA professional societies.