Review Article

A research framework for passive surveillance for food safety from social media: Identification and evaluation of customer reviews for regulatory use and case study of 30 restaurants

Akash Gajanan Prabhune[1],*, Neeraj Kumar Sethiya[2], Heemanshu Arora[3]

[1]Population Health Informatics, DIT University, Dehradun, Uttarakhand, India
[2]Dept. of Pharmacology, DIT University, Dehradun, Uttarakhand, India
[3]Foundation of Healthcare Technologies Society, New Delhi, India

ARTICLE INFO

ABSTRACT

The primary objective of this paper is to develop a framework for continuous monitoring of the safety of food business operators without overburdening established regulatory systems through social media for food safety. A phase-wise methodology was adopted, wherein Phase 1 was dedicated to identifying available literature on Adverse Drugs Reactions (ADR) reporting using Social Media data. Phase 2 used the data from google maps review of the restaurants to replicate a similar methodology for Food Safety Surveillance. We identified 5 themes for a complete Surveillance framework, theme 1 involves data collection from social media, theme 2 involves pre-processing of data for analysis, theme 3 involves data annotations, theme 4 involves Identifying the relationship between regulatory violation and event, and theme 5 involves evaluation of the model. We were able to demonstrate the ADR reporting methodology could be adopted till theme 3, whereas theme 4 requires the development of an algorithm to assess the causality of an event with the Food Safety Code. According to our research, it is possible to develop a passive surveillance system for food safety that adheres to the principle of ADR reporting; however, the main obstacle is the absence of a causality assessment algorithm that can link an event to the food safety code and help regulators take immediate action.

For reprints contact: reprint@ipinnovative.com

1. Introduction

Food Safety is an important aspect of health and well-being and the very nature of the food industry in India is a challenge for a regulator to monitor the quality of food supplied through the street cart vendor, a low-key restaurant near an office complex, hostel mess of a university or a restaurant in a 5-star establishment. The limited manpower and resources of the Food Safety Standards Authority of India (FSSAI) also complement the problem that the regulator is under-resourced to maintain vigilance at par with its global counterparts.[1]

Developing nations like India have a highly fragmented and unorganized food industry and limited resources at their disposal, thus the need for a passive surveillance system has been necessitated.[2] The Global Foodborne Infections Network (GFN),[3] is one such system which aims to build the capacity to detect, control and prevent foodborne and other enteric infections from farm to table. The proposed systems are aimed at promoting integrated, laboratory-based surveillance through intersectoral collaboration among human health, veterinary and food-related disciplines. Another well-known active surveillance system "Food Net" coordinated by the Centre for Disease Control (CDC) in

* Corresponding author.
E-mail address: prabhunesky@gmail.com (A. G. Prabhune).

North America is focused on nine organisms and related illnesses, such as hemolytic uremic syndrome (HUS), associated with Escherichia coli.[4]

FSSAI the apex body under the Ministry of Health and Family Welfare, Government of India for regulation of the food industry has incorporated Active Surveillance as a key mandate under its purview.[5,6] The Manual of Food Safety Officers published by FSSAI divides the task of regulating and licensing the industry into three tiers: a) Central Licensing Authority, b) State Licensing Authority, and c) Registration Authority. The three tiers are coordinated through State and National Commissioner, the food industry is broadly categorized into four categories a) Manufacturers/Millers, b) Hotel 3 stars and above, c) All food service providers including restaurants, boarding houses, clubs, canteens, caterers, banquet halls with food catering arrangements, d) Any other food business operator (Street vendors) The FSSAI has annual active surveillance plan in place targeting state-specific food products with historical evidence of adulteration (Milk, Ghee, Diary Products), contamination (Seafood, Meat) across specific time of a year. Vegetables are to be tested in monsoon season for e-Colli, and milk is also to be tested for e-Colli in monsoon.

Though there are provisions for passive surveillance under the Food Safety and Security (FSS) act 2006, few details are presented in Manuals, and little literature is published or grey literature available, indicating know how's of undertaking a passive surveillance system.[1,7] This presents a unique opportunity to use two rapidly growing fields of Digital Epidemiology and Consumer Health Informatics to develop a first-line passive surveillance network using principles of population health informatics. Much work has been done to monitor the adverse events related to various drugs. Web-Recognizing Adverse Drug Reaction (WEB RADR),[8] by Innovative Medicines Initiative, Vigi4Med Project,[9] and EudraVigilance,[10] are some of the well-established adverse drug event reporting programs utilizing data from social media, and online forums. In this paper, we propose to develop a framework for using data available on social media to generate credible evidence for preventive action by regulatory agencies in India.

2. Objectives

The primary objective of this paper is to develop a framework for continuous monitoring of the safety of food business operators without overburdening established regulatory systems through social media for food safety. The secondary objectives include the development of a clear pathway of valid signal detection in food safety reporting from social media; defining technological needs for signal detection; specifying the data protection and confidentiality aspects to use public social media data

3. Methodology

This paper was written using a two-phase methodology. Phase 1 included a review of the literature available on the use of data available on social media for Adverse Drug Reaction (ADR) monitoring to derive a comprehensive framework for passive surveillance of food business operators. Phase 2 included Use data from social media to evaluate the conceptual feasibility of developing a strong signal using a developed framework.

Under Phase 1, we developed a search strategy based on the Participants, Interventions, Comparison and Outcome (PICO) format. The key search terms include terms such as: "Pharmacovigilance, Adverse Drug Reaction Report, ADR Reporting, ADR Monitoring, Drug Monitoring, Adverse Events, Social Media, Twitter, Facebook, Online Forums, Secondary data, Crowd Sourced data, Framework, Evaluation of systems, Assessment". Suitable bullion operators "or, and" were used. PubMed and Google Scholar search resulted in 155 articles. According to the following criteria, we screened the articles for eligibility.

Studies using secondary data related to ADR monitoring. Studies published in English were included in the review. While studies using primary data collection using the internet for ADR monitoring and studies published in a language other than English were excluded from the review. The process is presented in Figure 1. This led to 22 papers being shortlisted for full-text reading. We screened through the 22 articles and stratified the articles into categories of a) Theme 1: Data Extraction, b) Theme 2: Pre-Processing, c) Theme 3: Data Annotation, d) Theme 4: Identifying the relationship between regulatory violation and event, e) Theme 5: Evaluation.

We presented various frameworks or procedures available across the literature and drew parallels between ADR reporting a Food Safety Surveillance. A combined framework was thus developed from phase 1. Under Phase 2, we used data from a Google Maps Review of 10 Food Business Operators (FBOs) in Byatrayanpura Wards of Bhruhat Bengaluru Mahanagar Palike, from 2016 to 2021 to obtain the results from the framework in line with regulatory requirements. We included FBOs with a minimum of 100 reviews, operating in dine-in mode. FBOs serving alcohol were excluded, FBOs operating for less than 1 year and only in the delivery mode were excluded.

Data Extraction and processing were carried out as per outlined framework in phase 1, descriptive data analysis was undertaken along with evaluation analysis using F measure, Precision, Recall and Accuracy.

4. Results

I have presented the results of Phase 1 and Phase 2, according to the themes, highlighted in the methodology section.
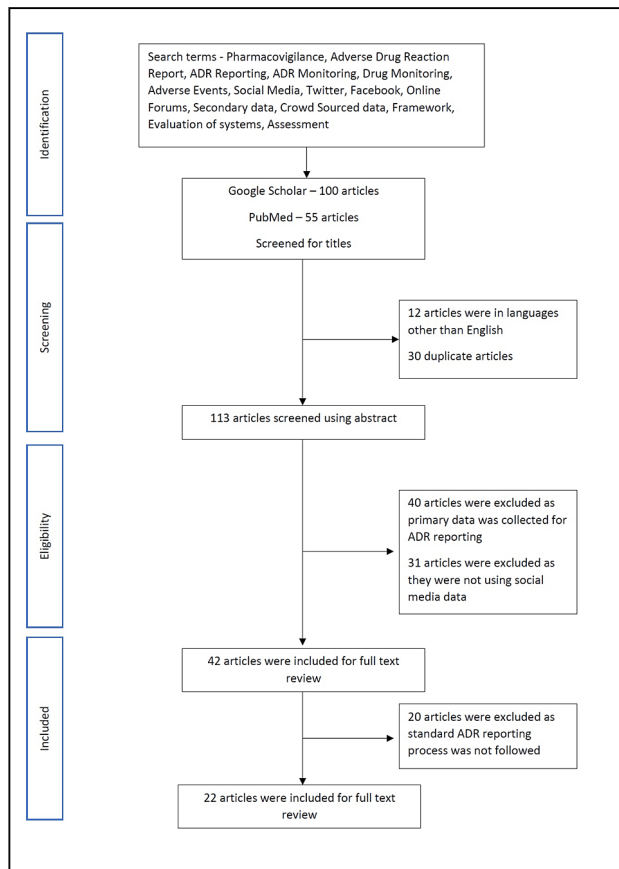
Fig. 1: Study flow chart (PRISMA flowchart)

Theme 1: Data Extraction - Under data extraction, we tried to look for specific answers like "What are the possible sources of data? what are the ethical implications of the data available on social media? What are the processes associated with the extraction of quality data from social media?" Social media is defined as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 that allow the creation and exchange of user-generated content based on mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content".[11] A report from the WEB RADAR study by Cater et al[12] found that Twitter and Facebook data was not useful for signal detection in Pharmacovigilance. Another paper by Sarker et al.[13] presented a key reason for lower F values from ADR reporting from data derived from Twitter was vague or too descriptive data. Studies by Comfort et al, and Pierce et al, used Twitter, Facebook, and Tumblr data with limited success.[14,15] We tried to search for sources of FBO data, while Twitter, Facebook, Tumblr, and Instagram had videos, and pictures related to food, most were promotional content.

Another source was reviews left by dinners over Google Maps, and Zomato. Amazon, Flipkart and other electronic marketplaces also had product reviews related to food, however, they were mostly for groceries, and ready-to-eat meals, related to Type A FBOs consisting of manufacturers and millers. Thus, we focused on the data available over Google Maps specifically.

4.1. Typical review on Google Maps for an FBO is as follows

"Food was very good, but considering the current covid pandemic, I'd say the Restaurant is not as prepared. The spoons are kept in water for folks to pick from... Felt very unhygienic because it's unclear if that water is replaced. Five stars for food, but only 2 for seating arrangements. And considering close together the seats I'm going to give them 3 stars for dine-in. I recommend taking it out. Overall, 4 stars" - A Level 7 Local Guide on Google Maps.

The Local Guide program by Google Maps is dedicated to a global community of explorers who write reviews, share photos, answer questions, add or edit places, and check facts on Google Maps. Millions of people rely on contributions "to decide where to go and what to do".[16] A Google guide has various levels from Level 1 to Level 10, as the level increases the credibility of information also increases proportionally.

The question of ethics was also explored from the literature, many ADR reporting using social media had put forth ethical benchmarks while using data from social media. A paper by Bousquet et al.[8] mentions that privacy becomes a key issue will using data over social media as the ownership of data will remain with the original contributors. Another paper by Azam et al[17] mentioned that consent for use of data over social media is not guaranteed, as many people are not aware of the real terms of the use of data they put over social media.

The Google Maps end-user policy highlights the reviews added by its users will be available for all, however, users can control access to personal information. In view of the challenges and policies, the Ethical framework adopted for this study is; a) The FBOs name will be visible as the FBO has voluntarily registered on Google maps for business visibility and the review data is the ownership of individual local guides or reviewers. b) The privacy of the local guide or reviewer will not be disclosed, and all data will be anonymized, the consent from each local guide or reviewer will not be feasible and will consider the user has read through the End User License Agreement provided by Google Maps. The data from Google Maps was extracted using Anaconda Script in Python and exported to the excel sheet.

Theme 2 Pre-Processing: The Data Pre-processing step prepares the raw data for analysis. The data from social media is usually in form of free text. Data pre-processing

consists of two steps: text cleaning and sentence boundary detection.[18] Various methods are marked in the literature for preprocessing of data, including Sentence Splitting, Parsing, Stemming, and Lemmatization.[19] We used the Method prescribed by Liu et al[18] of two-step pre-processing

Text cleaning for punctuation removal, personal identifier removal, and URL removal. A Google Map Review about Multicuisine restaurant was written as

"Stepped in as the reviews were good but stepped out with a bad taste after experiencing cockroaches crawl out of our table and over our plates.

The reason given was pest control was recently done.

We changed tables but the next family that walked in was seated at the same until we told them about the cockroaches."

After the first step of text cleaning the review was processed as

"Stepped in as the reviews were good but stepped out with a bad taste after experiencing cockroaches crawl out of our table and over our plates the reason given was pest control was recently done we changed tables but the next family that walked in was seated at the same until we told them about the cockroaches."

The second text was on space boundary detection, herein the review was split based on bullion operators stepped in as the reviews were good———1

Stepped out with a bad taste after experiencing cockroaches crawl out of our table and over our plates the reason given was pest control was recently done we changed tables ———2

The next family that walked in was seated at the same until we told them about the cockroaches——3

Theme 3 Data Annotation: Under theme three the literature presented options of Dictionary or Lexicon based, Rule-based, Machine Learning based techniques.[13,19,20] We used the New Mexico Restaurant Association (NMRA) Dictionary[21] on food safety as it was the most comprehensive open-access dictionary available on food safety data on social media. Another tool at our disposal was the FDA Guidelines for Confirmation of Foodborne Disease Outbreaks.[22] We subjected 10 reviews randomly selected from the 10 restaurants sample to the NMRA Dictionary and FDA Guidelines for Confirmation of Foodborne Disease Outbreaks. Using NMRA Dictionary we used the shortest dependency pathway to extract Food Violations, the shortest dependency pathway is presented in Figure 2.

Based on the data available from 10 restaurants and 10505 google maps reviews the data was extracted, cleaned for punctions and grammatical errors and split, the common keywords that were extracted from the reviews were presented in table 1 along with frequency. From the 10505 reviews, 263 reviews were having more than three words comprehension, and 125 reviews were evaluated for
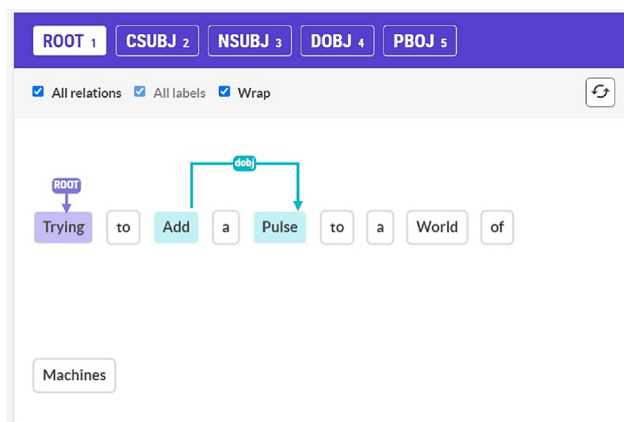

Fig. 2: Shortest dependency path algorithm

keywords after excluding reviews without a proper semantic framework. The keywords are summarized in Table 1 in terms of their frequency across the 125 reviews.

The thematic area-wise frequency of keywords is presented in Table 2.

When looking across 10 restaurants the Health Issues (N=7) of Stomachache, Food Poisoning were attributed to biryani, rice, and fried rice. Almost all restaurants had a review mentioning pests like cockroaches, mosquitoes, and mosquitoes in the dining area and 1 restaurant had a review mentioning food items having worms. All the restaurants had negative reviews with respect to service, staff, and food taste.

Theme 4 Identifying the relationship between regulatory violation and event: For the relationship between keywords and food safety violations, the ADR reporting papers suggest a Rule-based assessment - Kramer algorithm, World Health Organisation (WHO) algorithm for severity assessment and Statistical assessment to link an adverse event with a drug based on the time of drug intake and symptoms and determine the causality of an event. For the proposed Food Safety, the FDA Guidelines for Confirmation of Foodborne Disease Outbreaks was a standard algorithm, however, it included confirmation of organisms for laboratory and medical matching of symptoms, which were not feasible from the data available from the Google Maps reviews. Statistical Method was an option; however, it requires stakeholder discussion to define the hypothesis and define the signal from a regulatory perspective.

Theme 5 Evaluation: To evaluate the performance of the system, we propose to use statistical measures of F Square, Accuracy, Precision, and Recall. As the data available in the feasibility study was limited to 125 reviews, the evaluation of the system was not undertaken. The structured proposed framework is presented in Figure 2.

Table 1: Frequency of keywords across the sample restaurants

Themes	Keywords	Frequency
Food Taste	Oily	4
Food Taste	Spicy	1
Cost	Pricey	7
Cost	Expensive	5
Hygiene	Not Hygienic	7
Hygiene	Not Clean	2
Pest	Cockroaches	3
Food Taste	Bad taste	5
Food Taste	Salty	3
Uncategorised	Not Good	5
Cost	Price Hiked	1
Service related	Late Service	1
Uncategorised	Worst Restaurant	1
Uncategorised	Bad Food	8
Food Quality	Not Cooked Properly	3
Service related	Wrong order	1
Service related	Worst Service	14
Service related	Pathetic Service	6
Service related	Worst Costumer Service	4
Health Issues	Fell ill	1
Uncategorised	Sub Standard	1
Service related	Slow Service	4
External items in food	Stapler Pin	1
Health Issues	Suffered from very bad health	1
Food Taste	Horrible Food	2
Service related	Order Delay	2
Food Quality	Smelly	3
Pest	Mosquitoes	4
Food Taste	Tasteless	1
Service related	Untidy Waiters	1
Health Issues	Food Poisoning	3
Uncategorised	Noisy	1
Food Quality	Food Sucks	1
Health Issues	Stomach Pain	2
Uncategorised	Not recommended	12
Uncategorised	Not Fresh	1
Pest	Worms	1
External items in food	Hair in the food	1
Uncategorised	Soda	1
Total		125

Table 2: Thematic area wise frequency of keywords

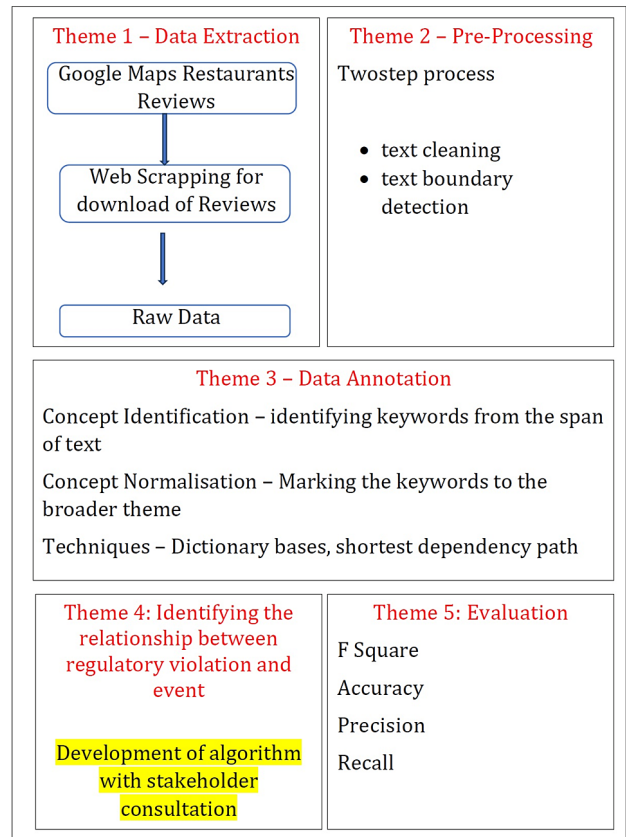| Themes	Frequency	%
Food Taste	16	12.8
Cost	13	10.4
Hygiene	9	7.2
Pest	8	6.4
Service Related	33	26.4
Health Issues	7	5.6
External items in food	2	1.6
Food Quality	7	5.6
Uncategorised	30	24
Total	125	100

Fig. 3: Proposed framework for the food surveillance

5. Discussion

Various research papers have used Restaurant reviews on Google Maps, and Yelp, to analyse using sentiment analysis techniques. A paper by Krishna et al., Hossain et al., and Adi et al., all presented machine learning techniques using the Bayesian approach and presented the sentiment analysis that can be used to identify the sentiments of the population visiting the restaurants.[23–25] The analysis was used to classify the restaurant based on the services offered and the general cause of negative or positive feedback. A paper published by Harris et al.[26] used a different approach wherein machine learning and human analysis to classify the tweets relevant to food poisoning and automate reply to the individuals to report the food poisoning to local health authority for increased reporting of food poisoning outbreak. Indicating the need for an algorithm to validate the causality between event and outcome.

This study is the first attempt to our knowledge to develop a framework for a passive surveillance network using social media data and providing high-quality signals with regulatory value. The efforts of our study have shown,

the data from Google Reviews can be used for general sentiment analysis around the restaurant business. The methodology prescribed for ADR reporting proved to be closely linked to our approach to Food Safety, however, we were not able to identify the algorithm which would assess causality between regulatory violation and an event. This necessitates initiating separate studies to consult with various stakeholders (specifically regulatory bodies) to understand their needs, understand the current method, develop a consensus method to build an algorithm and test the validity of such algorithm.

6. Conclusions

The review of various articles indicates the work on using social media to derive Regulatory signals specifically in pharmacovigilance has been successfully carried out and success, the key barrier to replicating the same model in Food Safety and Regulation is the non-availability of causality assessment algorithms which would link and event with a food safety violation. The current approach of using review data to generate sentiment around FBOs is statistically robust. The sentiment analysis techniques do indicate the sentiment behind the review left by the customer. Further segmentation and analysis had potential to positively impact the quality of food and customer experience. The study indicates the need to use informatics tools to develop and pilot tech enabled model in accordance to current Food Safety Code.

7. Source of Funding

None.

8. Conflict of Interest

None.

References

1. Gardner S. Consumers and food safety: A food industry perspective; 1993. Available from: https://www.fao.org/3/v2890t/v2890t05.htm.
2. Sharma LL, Teret SP, Brownell KD. The Food Industry and Self-Regulation: Standards to Promote Success and to Avoid Public Health Failures. *Am J Public Health*. 2010;100(2):240–6.
3. Desai N, Joshi H, Chiller TM. Global foodborne infections network (GFN): An opportunity for capacity-building in enteric diseases in India. In: 141st APHA Annual Meeting and Exposition; 2013.
4. Mccabe-Sellers BJ, Beattie SE. Food safety: Emerging trends in foodborne illness surveillance and prevention. *J Am Diet Assoc*. 2004;104(11):1708–17.
5. FSSAI Chapter 6 Inspection of Food Establishment. Available from: https://www.fssai.gov.in/upload/uploadfiles/files/Chapter6.pdf.
6. FSSAI Chapter 8 Annual Surveillance Plan. Available from: https://www.fssai.gov.in/upload/uploadfiles/files/Chapter8.pdf.
7. Freeman J, Ryan N, Glenesk J. Strategic Surveillance for Food Safety: Designing a surveillance approach and considerations for implementation; 2019. Available from: https://www.rand.org/pubs/research_reports/RR2519.html.
8. Bousquet C. The Adverse Drug Reactions from Patient Reports in Social Media Project: Five Major Challenges to Overcome to Operationalize Analysis and Efficiently Support Pharmacovigilance Process. *JMIR Res Protoc*. 2017;6(9):e179.
9. Audeh B, Bellet F, Beyens MN, Louët ALL, Bousquet C. Use of Social Media for Pharmacovigilance Activities: Key Findings and Recommendations from the Vigi4Med Project. *Drug Saf*. 2020;43(9):835–51.
10. Postigo R, Brosch S, Slattery J, Haren AV, Dogné JM, Kurz X, et al. EudraVigilance Medicines Safety Database: Publicly Accessible Data for Research and Public Health Protection. *Drug Saf*. 2018;41(7):665–75.
11. Council conclusions on shaping Europe's digital future; 2020. Available from: https://www.consilium.europa.eu/media/44389/st08711-en20.pdf.
12. Caster O, Dietrich J, Kürzinger ML, Lerch M, Maskell S, Norén GN. Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project. *Drug Saf*. 2018;41(12):1355–69.
13. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. *J Biomed Inform*. 2015;54:202–12.
14. Pierce CE, Bouri K, Pamer C, Proestel S, Rodriguez HW, Le HV, et al. Evaluation of Facebook and Twitter Monitoring to Detect Safety Signals for Medical Products: An Analysis of Recent FDA Safety Alerts. *Drug Saf*. 2017;40(4):317–31.
15. Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting Through the Safety Data Haystack: Using Machine Learning to Identify Individual Case Safety Reports in Social-Digital Media. *Drug Saf*. 2018;41(6):579–90.
16. Overview - Local Guides Help. Available from: https://support.google.com/local-guides/answer/6225846?hl=en.
17. Azam R. Accessing social media information for pharmacovigilance: what are the ethical implications? *Ther Adv Drug Saf*. 2018;9(8):385–7.
18. Liu J, Zhao S, Wang G. SSEL-ADE: A semi-supervised ensemble learning framework for extracting adverse drug events from social media. *Artif Intell Med*. 2018;84:34–49.
19. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. *BMC Med Inform Decis Mak*. 2018;18(1):38.
20. Rees S, Mian S, Grabowski N. Using social media in safety signal management: is it reliable? *Ther Adv Drug Saf*. 2018;9(10):591–9.
21. Food Safety Vocabulary. NMRA; 2020. Available from: https://www.nmrestaurants.org/food-safety-vocabulary/.
22. B Appendix. Guidelines for Confirmation of Foodborne-Disease Outbreaks. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/ss4901a3.htm.
23. Krishna A, Akhilesh V, Aich A, Hegde C. Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques. In: Sridhar V, Padma MC, Rao KAR, editors. Emerging Research in Electronics, Computer Science and Technology. Springer; 2019. p. 687–96.
24. Hossain N, Bhuiyan MR, Tumpa ZN, Hossain SA. Sentiment Analysis of Restaurant Reviews using Combined CNN-LSTM. In: 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT); 2020. p. 1–5.
25. Laksono RA, Sungkono KR, Sarno R, Wahyuni CS. Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes. In: 12th International Conference on Information & Communication Technology and System (ICTS); 2019. p. 49–54.
26. Harris JK, Hawkins JB, Nguyen L, Nsoesie EO, Tuli G, Mansour R, et al. Using Twitter to Identify and Respond to Food Poisoning: The Food Safety STL Project. *J Public Health Manag Pract*. 2017;23(6):577–80.

Author biography

Akash Gajanan Prabhune, PhD Fellow

Neeraj Kumar Sethiya, Associate Professor

Heemanshu Arora, Public Health Researcher