Supplementary Material

**Social media analytics in nutrition research: A rapid review of current usage in investigation of dietary behaviours**

Part 1: Database searches across Scopus, Medline, CINAHL, ACM Digital Library and Engineering Village

Part 2: Original data charting tables

Table 2A: Eligible studies with social media analytics methodology in nutrition related research

Table 2B: Study author(s) disciplines or fields of research

Table 2C: Ethical approval, mentions and considerations

Part 1: Part 1: Database searches across Scopus, Medline, CINAHL, ACM Digital Library and Engineering Village

Database: Scopus Search title: Social Media Analytics in Nutrition Research with Scopus

Date searched: March 9, 2020 Results: n=1,577

Export method: Scopus supports .RIS export for up to 2,000 records.

|  |  |  |  |
| --- | --- | --- | --- |
| Social media related keywords: | Data analytic and SMA keywords: | Nutrition (including targeted food and health) keywords: | |
| Social media  Social media platform  Social listening  Social network\*  Website  Online  Internet  Social communication  Mobile communication  Social networking site  Facebook  Twitter  Instagram  YouTube | Data mining  Data analy\*  Big data  Machine learning  Infoveillance  Infodemiology  Dashboarding  Real time monitoring  Informatics  Cloud computing  Artificial intelligence  Business analytics  Metrics | Nutri\* (nutrition, nutrients)  Diet\* (dietetics; dietitian)  Health  Public health  Diabetes  Heart Disease  Cardiovascular Disease  Cancer  Digest\*  Gastro\*  Gut  Overweight  Obesity | Eating  Eating out  Meal\*  Breakfast  Dinner  Dining  Lunch  Snack\*  Beverage  Drink\*  Alcohol  Fast food  Takeaway  Junk food  Food purchase  Food shopping  Supermarket  Restaurant |
| **Social media limit to articles 2014-2020 in English = 216,690**  **TRIAL SEARCH**  TITLE-ABS-KEY ( "Social media"  OR  "Social media platform"  OR  "Social listening"  OR  "Social network\*"  OR  website  OR  online  OR  internet  OR  "Social communication"  OR  "Mobile communication"  OR  "Social networking site"  OR  facebook  OR  twitter  OR  instagram  OR  youtube )  AND  ( LIMIT-TO ( PUBYEAR ,  2019 )  OR  LIMIT-TO ( PUBYEAR ,  2018 )  OR  LIMIT-TO ( PUBYEAR ,  2017 )  OR  LIMIT-TO ( PUBYEAR ,  2016 )  OR  LIMIT-TO ( PUBYEAR ,  2015 )  OR  LIMIT-TO ( PUBYEAR ,  2014 ) )  AND  ( LIMIT-TO ( DOCTYPE ,  "ar" ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )  **AND**  **Data analytic limit to articles 2014-2020 in English**  TITLE-ABS-KEY ( "Data mining"  OR  "Data analy\*"  OR  "Big data"  OR  "Machine learning"  OR  infoveillance  OR  infodemiology  OR  dashboarding  OR  "Real time monitoring"  OR  informatics  OR  "Cloud computing"  OR  "Artificial intelligence"  OR  "Business analytics"  OR  metrics )  AND  ( LIMIT-TO ( PUBYEAR ,  2019 )  OR  LIMIT-TO ( PUBYEAR ,  2018 )  OR  LIMIT-TO ( PUBYEAR ,  2017 )  OR  LIMIT-TO ( PUBYEAR ,  2016 )  OR  LIMIT-TO ( PUBYEAR ,  2015 )  OR  LIMIT-TO ( PUBYEAR ,  2014 ) )  AND  ( LIMIT-TO ( DOCTYPE ,  "ar" ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )  **AND**  **Nutrition limit to articles 2014-2020 in English**  TITLE-ABS-KEY ( nutri\*  OR  diet\*  OR  health  OR  "Public health"  OR  diabetes  OR  "Heart Disease"  OR  "Cardiovascular Disease"  OR  cancer  OR  digest\*  OR  gastro\*  OR  gut  OR  overweight  OR  obesity )  AND  ( LIMIT-TO ( PUBYEAR ,  2019 )  OR  LIMIT-TO ( PUBYEAR ,  2018 )  OR  LIMIT-TO ( PUBYEAR ,  2017 )  OR  LIMIT-TO ( PUBYEAR ,  2016 )  OR  LIMIT-TO ( PUBYEAR ,  2015 )  OR  LIMIT-TO ( PUBYEAR ,  2014 ) )  AND  ( LIMIT-TO ( DOCTYPE ,  "ar" ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )  **Nutrition limit to articles 2014-2020 in English**  **OR**  TITLE-ABS-KEY ( eating  OR  "Eating out"  OR  meal\*  OR  breakfast  OR  dinner  OR  dining  OR  lunch  OR  snack\*  OR  beverage  OR  drink\*  OR  alcohol  OR  "Fast food"  OR  takeaway  OR  "Junk food"  OR  "Food purchase"  OR  "Food shopping"  OR  supermarket  OR  restaurant )  AND  ( LIMIT-TO ( PUBYEAR ,  2019 )  OR  LIMIT-TO ( PUBYEAR ,  2018 )  OR  LIMIT-TO ( PUBYEAR ,  2017 )  OR  LIMIT-TO ( PUBYEAR ,  2016 )  OR  LIMIT-TO ( PUBYEAR ,  2015 )  OR  LIMIT-TO ( PUBYEAR ,  2014 ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )  **Nutrition limit to articles 2014-2020 in English**  **FINAL n=4,061**  ( TITLE-ABS-KEY ( eating  OR  "Eating out"  OR  meal\*  OR  breakfast  OR  dinner  OR  dining  OR  lunch  OR  snack\*  OR  beverage  OR  drink\*  OR  alcohol  OR  "Fast food"  OR  takeaway  OR  "Junk food"  OR  "Food purchase"  OR  "Food shopping"  OR  supermarket  OR  restaurant ) )  OR  ( TITLE-ABS-KEY ( nutri\*  OR  diet\*  OR  health  OR  "Public health"  OR  diabetes  OR  "Heart Disease"  OR  "Cardiovascular Disease"  OR  cancer  OR  digest\*  OR  gastro\*  OR  gut  OR  overweight  OR  obesity ) )  AND  ( TITLE-ABS-KEY ( "Data mining"  OR  "Data analy\*"  OR  "Big data"  OR  "Machine learning"  OR  infoveillance  OR  infodemiology  OR  dashboarding  OR  "Real time monitoring"  OR  informatics  OR  "Cloud computing"  OR  "Artificial intelligence"  OR  "Business analytics"  OR  metrics ) )  AND  ( TITLE-ABS-KEY ( "Social media"  OR  "Social media platform"  OR  "Social listening"  OR  "Social network\*"  OR  website  OR  online  OR  internet  OR  "Social communication"  OR  "Mobile communication"  OR  "Social networking site"  OR  facebook  OR  twitter  OR  instagram  OR  youtube ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )  AND  ( LIMIT-TO ( DOCTYPE ,  "ar" ) )  AND  ( LIMIT-TO ( PUBYEAR ,  2019 )  OR  LIMIT-TO ( PUBYEAR ,  2018 )  OR  LIMIT-TO ( PUBYEAR ,  2017 )  OR  LIMIT-TO ( PUBYEAR ,  2016 )  OR  LIMIT-TO ( PUBYEAR ,  2015 )  OR  LIMIT-TO ( PUBYEAR ,  2014 ) )  **REFINE**  Top 2,000 articles by date ranged searched for relevance by abstract and title n=48  Next step rationalized social media keywords to the following:  Social media  Social media platform  Social listening  Social network\*  Social networking site  Facebook  Twitter  Instagram  YouTube  Add: Blog  TITLE-ABS-KEY ( "Social media"  OR  "Social media platform"  OR  "Social listening"  OR  "Social network\*"  OR  "Social networking site"  OR  facebook  OR  twitter  OR  instagram  OR  youtube  OR  blog ) n = 196,592  Replace.  Search as above.  **Final search**  ( TITLE-ABS-KEY ( nutri\*  OR  diet\*  OR  health  OR  "Public health"  OR  diabetes  OR  "Heart Disease"  OR  "Cardiovascular Disease"  OR  cancer  OR  digest\*  OR  gastro\*  OR  gut  OR  overweight  OR  obesity ) )  OR  ( TITLE-ABS-KEY ( eating  OR  "Eating out"  OR  meal\*  OR  breakfast  OR  dinner  OR  dining  OR  lunch  OR  snack\*  OR  beverage  OR  drink\*  OR  alcohol  OR  "Fast food"  OR  takeaway  OR  "Junk food"  OR  "Food purchase"  OR  "Food shopping"  OR  supermarket  OR  restaurant ) )  AND  ( TITLE-ABS-KEY ( "Data mining"  OR  "Data analy\*"  OR  "Big data"  OR  "Machine learning"  OR  infoveillance  OR  infodemiology  OR  dashboarding  OR  "Real time monitoring"  OR  informatics  OR  "Cloud computing"  OR  "Artificial intelligence"  OR  "Business analytics"  OR  metrics ) )  AND  ( TITLE-ABS-KEY ( "Social media"  OR  "Social media platform"  OR  "Social listening"  OR  "Social network\*"  OR  "Social networking site"  OR  facebook  OR  twitter  OR  instagram  OR  youtube  OR  blog ) )  AND  ( LIMIT-TO ( LANGUAGE ,  "English" ) )  AND  ( LIMIT-TO ( DOCTYPE ,  "ar" ) )  AND  ( LIMIT-TO ( PUBYEAR ,  2020 )  OR  LIMIT-TO ( PUBYEAR ,  2019 )  OR  LIMIT-TO ( PUBYEAR ,  2018 )  OR  LIMIT-TO ( PUBYEAR ,  2017 )  OR  LIMIT-TO ( PUBYEAR ,  2016 )  OR  LIMIT-TO ( PUBYEAR ,  2015 )  OR  LIMIT-TO ( PUBYEAR ,  2014 ) ) | | | |

Database: Medline via Ovid

Search title: Social Media Analytics in Nutrition Related Research with Medline

Date searched: March 9, 2020 Results: n = 85

Export method: Medline supports .RIS export.

|  |  |  |  |
| --- | --- | --- | --- |
| Social media related keywords: | Data analytic and SMA keywords: | Nutrition (including targeted food and health) keywords: | |
| Social media  Social media platform  Social listening  Social network\*  Social networking site  Facebook  Twitter  Instagram  YouTube  Blog | Data mining  Data analy\*  Big data  Machine learning  Infoveillance  Infodemiology  Dashboarding  Real time monitoring  Informatics  Cloud computing  Artificial intelligence  Business analytics  Metrics | $Nutri\* (nutrition, nutrients)  $Diet\* (dietetics; dietitian)  Health  Public health  Diabetes  Heart Disease  Cardiovascular Disease  Cancer  Digest\*  Gastro\*  Gut  Overweight  Obesity | Eating  Eating out  Meal\*  Breakfast  Dinner  Dining  Lunch  Snack\*  Beverage  Drink\*  Alcohol  Fast food  Takeaway  Junk food  Food purchase  Food shopping  Supermarket  Restaurant |
| Limits: English language; Humans; 2014-current  ("Social media" or "Social media platform" or "Social listening" or "Social network\*" or "Social networking site" or Facebook or Twitter or Instagram or YouTube or Blog).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]  n=12191  AND  (Data mining or "Data analy\*" or "Big data" or "Machine learning" or Infoveillance or Infodemiology or Dashboarding or "Real time monitoring" or Informatics or "Cloud computing" or "Artificial intelligence" or "Business analytics" or Metrics).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]  n= 52793  AND  ($Nutri\* or $Diet\* or Health or "Public health" or Diabetes or "Heart Disease" or "Cardiovascular Disease" or Cancer or Digest\* or Gastro\* or Gut or Overweight or Obesity).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]  n=1170178  OR  (Eating or "Eating out" or Meal\* or Breakfast or Dinner or Dining or Lunch or Snack\* or Beverage or Drink\* or Alcohol or "Fast food" or Takeaway or "Junk food" or "Food purchase" or "Food shopping" or Supermarket or Restaurant).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]  N=81713 | | | |

Database: Social Media Analytics in Nutrition Related Research with CINAHL

Date searched: March 9, 2020

Results: (Peer reviewed; 2014-2019; English language; Human) n= 1,259

Export method: Direct export x 2 batches to Endnote Web

|  |  |  |  |
| --- | --- | --- | --- |
| Social media related keywords: | Data analytic and SMA keywords: | Nutrition (including targeted food and health) keywords: | |
| Social media  Social media platform  Social listening  Social network\*  Social networking site  Facebook  Twitter  Instagram  YouTube  Blog | Data mining  Data analy\*  Big data  Machine learning  Infoveillance  Infodemiology  Dashboarding  Real time monitoring  Informatics  Cloud computing  Artificial intelligence  Business analytics  Metrics | $Nutri\* (nutrition, nutrients)  $Diet\* (dietetics; dietitian)  Health  Public health  Diabetes  Heart Disease  Cardiovascular Disease  Cancer  Digest\*  Gastro\*  Gut  Overweight  Obesity | Eating  Eating out  Meal\*  Breakfast  Dinner  Dining  Lunch  Snack\*  Beverage  Drink\*  Alcohol  Fast food  Takeaway  Junk food  Food purchase  Food shopping  Supermarket  Restaurant |
| Limits: Journals; English language; Humans; 2014-current  "Social media" OR "Social media platform" OR "Social listening" OR "Social network\*" OR "Social networking site" OR Facebook OR Twitter OR Instagram OR YouTube OR Blog  n= = 8,639  AND "Data mining" OR "Data analy\*" OR "Big data" OR "Machine learning" OR Infoveillance OR Infodemiology OR Dashboarding OR "Real time monitoring" OR Informatics OR "Cloud computing" OR "Artificial intelligence" OR "Business analytics" OR Metrics  N= =130,280  AND  $Nutri\* OR $Diet\* OR Health OR "Public health" OR Diabetes OR "Heart Disease" OR "Cardiovascular Disease" OR Cancer OR Digest\* OR Gastro\* OR Gut OR Overweight OR Obesity  N = 350,324  OR Eating OR "Eating out" OR Meal\* OR Breakfast OR Dinner OR Dining OR Lunch OR Snack\* OR Beverage OR Drink\* OR Alcohol OR "Fast food" OR Takeaway OR "Junk food" OR "Food purchase" OR "Food shopping" OR Supermarket OR Restaurant  N= = 39,775 | | | |

Database: Social Media Analytics in Nutrition Related Research with ACM Digital Library Full-Text Collection

Date searched: MARCH 9, 2020

Results: n = 1,608

Export method: Direct export to Endnote Web

|  |  |  |  |
| --- | --- | --- | --- |
| Social media related keywords: | Data analytic and SMA keywords: | Nutrition (including targeted food and health) keywords: | |
| Social media  Social media platform  Social listening  Social network  Social networking site  Facebook  Twitter  Instagram  YouTube  Blog | Data mining  Data analytics  Big data  Machine learning  Infoveillance  Infodemiology  Dashboarding  Real time monitoring  Informatics  Cloud computing  Artificial intelligence  Business analytics  Metrics | Nutrition  Diet  Health  Public health  Diabetes  Heart Disease  Cardiovascular disease  Cancer  Digestion  Gastro  Gut  Overweight  Obesity | Eating  Eating out  Meal  Breakfast  Dinner  Dining  Lunch  Snack  Beverage  Drink  Alcohol  Fast food  Takeaway  Junk food  Food purchase  Food shopping  Supermarket  Restaurant |
| Limits: 2014-current; Journals  "query": { (("Social media" OR "Social media platform" OR "Social listening" OR "Social network" OR "Social networking site" OR Facebook OR Twitter OR Instagram OR YouTube OR Blog) AND ("Data mining" OR "Data analytics" OR "Big data" OR "Machine learning" OR Infoveillance OR Infodemiology OR Dashboarding OR "Real time monitoring" OR Informatics OR "Cloud computing" OR "Artificial intelligence" OR "Business analytics" OR Metrics) AND (Nutrition OR Diet OR Health OR "Public health" OR Diabetes OR "Heart Disease" OR "Cardiovascular disease" OR Cancer OR Digestion OR Gastro OR Gut OR Overweight OR Obesity OR Eating OR "Eating out" OR Meal OR Breakfast OR Dinner OR Dining OR Lunch OR Snack OR Beverage OR Drink OR Alcohol OR "Fast food" OR Takeaway OR "Junk food" OR "Food purchase" OR "Food shopping" OR Supermarket OR Restaurant)) }  "filter": {"publicationYear":{ "gte":2014 }},  {owners.owner=GUIDE} | | | |

Database: Social Media Analytics in Nutrition Related Research with Engineering Village

Date searched: October 9 2019\* Subscription to this database ended in 2019 so the search could not be update in March 2020

Results: Journal articles and in press; English language; 2014-current n= 691

Export method: Can only export 500 .RIS files at one time. Two batches of export completed and combined in Endnote folder. Need to select Ei Compendex filter.

|  |  |  |  |
| --- | --- | --- | --- |
| Social media related keywords: | Data analytic and SMA keywords: | Nutrition (including targeted food and health) keywords: | |
| Social media  Social media platform  Social listening  Social network\*  Social networking site  Facebook  Twitter  Instagram  YouTube  Blog | Data mining  Data analy\*  Big data  Machine learning  Infoveillance  Infodemiology  Dashboarding  Real time monitoring  Informatics  Cloud computing  Artificial intelligence  Business analytics  Metrics | $Nutri\* (nutrition, nutrients)  $Diet\* (dietetics; dietitian)  Health  Public health  Diabetes  Heart Disease  Cardiovascular Disease  Cancer  Digest\*  Gastro\*  Gut  Overweight  Obesity | Eating  Eating out  Meal\*  Breakfast  Dinner  Dining  Lunch  Snack\*  Beverage  Drink\*  Alcohol  Fast food  Takeaway  Junk food  Food purchase  Food shopping  Supermarket  Restaurant |
| Limits: Journal articles and in press; English language; 2014-current  (((("Social media" OR "Social media platform" OR "Social listening" OR "Social network\*" OR "Social networking site" OR Facebook OR Twitter OR Instagram OR YouTube OR Blog))) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA)))  n=2594  AND  ((("Data mining" OR "Data analy\*" OR "Big data" OR "Machine learning" OR Infoveillance OR Infodemiology OR Dashboarding OR "Real time monitoring" OR Informatics OR "Cloud computing" OR "Artificial intelligence" OR "Business analytics" OR Metrics)) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA)))  n=237815  AND  (((Nutri\* OR Diet\* OR Health OR "Public health" OR Diabetes OR "Heart Disease" OR "Cardiovascular Disease" OR Cancer OR Digest\* OR Gastro\* OR Gut OR Overweight OR Obesity)) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA)))  N=350313  Final n=691 redo see saved  ( (((((((Eating OR {Eating out} OR Meal\* OR Breakfast OR Dinner OR Dining OR Lunch OR Snack\* OR Beverage OR Drink\* OR Alcohol OR {Fast food} OR Takeaway OR {Junk food} OR {Food purchase} OR {Food shopping} OR Supermarket OR Restaurant)) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA))) AND ((2019 OR 2018 OR 2017 OR 2016 OR 2015 OR 2014) WN YR))) AND (1884-2019 WN YR)) OR (((((Nutri\* OR Diet\* OR Health OR {Public health} OR Diabetes OR {Heart Disease} OR {Cardiovascular Disease} OR Cancer OR Digest\* OR Gastro\* OR Gut OR Overweight OR Obesity)) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA)))) AND (2014-2019 WN YR)))) AND ( (((({Data mining} OR {Data analy\*} OR {Big data} OR {Machine learning} OR Infoveillance OR Infodemiology OR Dashboarding OR {Real time monitoring} OR Informatics OR {Cloud computing} OR {Artificial intelligence} OR {Business analytics} OR Metrics)) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA)))) AND (2014-2019 WN YR)) AND ( ((((({Social media} OR {Social media platform} OR {Social listening} OR {Social network\*} OR {Social networking site} OR Facebook OR Twitter OR Instagram OR YouTube OR Blog))) AND ((({ja} OR {ip}) WN DT) AND ({english} WN LA)))) AND (2014-2019 WN YR)) | | | |

**Part 2:** **Data charts**

**Table 2A: Eligible studies with social media analytics methodology in nutrition related research (alphabetical by first author)\***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | **References**  **Author, Year, Country** | **SoMe**  **Platform(s)** | **Dataset size**  **SMA summary** | **Nutrition relevance(s) and taxonomy** | **SMA Steps**  **Data**  **Discovery** | **SMA Steps**  **Data Collection** | **SMA Steps**  **Data Preparation** | **SMA Steps**  **Data Analysis** |
|  | Abbar et al. 2015, Qatar on USA data (1) | Twitter | 892K tweets | Collected and analysed tweets from a sample of users and aimed to correlate food related messages with state-wide health data (obesity rates).  Taxonomy: content analysis | 🗸 | 🗸  Twitter API | 🗸 | 🗸  LIWC  Genderize API |
|  | Alhabash et al. 2018, USA (2) | Twitter | 47.5M tweets | Collected a large sample of tweets and performed content analysis to examine how the prevalence of tweeting about alcohol, along with tweet features, can predict tweet virality.  Taxonomy: content analysis, engagement | 🗸 | 🗸  Twitter API and geotagged tweets | 🗸 | 🗸  LIWC |
|  | Alajajian et al. 2017, USA(3) | Twitter | ~50M tweets | Developed a machine learning device called “Lexicocalorimeter” with the attempt to measure energy/caloric content of foods mapped across the USA population.  Taxonomy: Content analysis | 🗸 | 🗸 | 🗸 | 🗸 |
|  | Blackstone, et al. 2018, USA (4) | Facebook | 617 individual posts of user data (original posts and comments) from 2 public Facebook fitness groups | Analysed posts from extreme fitness and nutrition groups for harmful health messages  Taxonomy: content analysis | 🗸  Searched platform to obtain popular groups by member number | x  Manual compiled into Excel | ?  No mention of cleaning data | x  Manual codebook |
|  | Carrotte et al. 2017, Australia (5) | Instagram  Facebook  Twitter  Tumblr | 360  12  4  39 | Collected posts with #fitspo to characterize food, body image and dieting messages by gender.  Taxonomy: content analysis. | 🗸  Tagboard.com  Platform search | x | x | x |
|  | Cavazos-Rehg et al. 2015, USA (6) | Twitter | 11,966,381 tweets  Subset 5,000 | Assessed the content of tweets with alcohol or drinking-related keywords to explore binge drinking behaviours.  Taxonomy: content analysis | 🗸  Purchased data from Simply Measured a private analytics company using Gnip to access Twitter firehose. After preparing keywords with [Topsy](http://topsy.thisisthebrigade.com/) | 🗸  Purchased data from Simply Measured.  Assessed Klout scores | 🗸 | 🗸/x  Manual coding of subset by crowdsourcing using CrowdFlower |
|  | Chen et al. 2014, USA (7) | Twitter | 81,543 tweets  Subset ~350 | Collected tweets with mentions of grocery stores and fast food outlets to map relationship between food choices and exposure to local food environment.  Taxonomy: content analysis | 🗸  Open Source Python library and key word search | 🗸  Open Source Python library | 🗸/x  ESRI ArcMap 2.0  Manual coding | 🗸/x  Manual coding  ANOVA |
|  | ElTayeby et al., 2018, USA (8) | Facebook group | 4266 posts (text, images, video) within in one Facebook group “I’m Shmacked” | Investigated the feasibility of mining Facebook posts and machine learning to identify between alcohol drinking and non-alcohol posts with a view to exploring drinking behaviours in US College students  Taxonomy: content analysis | 🗸  Facebook Group member number and media reports | 🗸  [Facebook Platform Python SDK](https://github.com/mobolic/facebook-sdk) and API | 🗸  Human annotation to train machine learning  Cleaned links, hashtags, emoticons | 🗸  Text classification using SVM or LDA  Image and Video classification using AlexNet |
|  | Fried et al. 2014, USA(9) | Twitter | 3M | Analysed food related tweets by hashtags (such as #dinner and #breakfast) and developed machine learning system to predict population characteristics such as overweight.  Taxonomy: content analysis | 🗸 | 🗸 | 🗸 | 🗸  Standford CoreNLP  LDA |
|  | Huang et al. 2017, USA (10) | Twitter | 50M | Applied Natural Language Processing and machine learning methods to examine alcohol (and tobacco) tweets for temporal patterns by age group (including underage).  Taxonomy: content analysis | 🗸 | 🗸  Twitter API | 🗸  Python | x/🗸  Python  Amazon Mechanical Turk |
|  | Kershaw et al. 2014, UK(11) | Twitter | 31.6M | Designed exploratory machine learning model with the aim to assess alcohol related tweets for consumption patterns and compared to an existing health dataset.  Taxonomy: content analysis | 🗸 | 🗸  Twitter API | 🗸 | 🗸 |
|  | Karami et al. 2018, USA (12) | Twitter | 4.5million tweets | Analysed using machine learning text in tweets to characterize health opinions in diabetes, diet, exercise and obesity (DDEO)  Taxonomy: content analysis | 🗸  Used keys word and hashtags on twitter | 🗸  Used Twitter API for real-time data collection of 1% of tweets in time period | 🗸  Standard list for removing stop words. Log-likelihood estimation to find optimum number of topics | 🗸  Topic modelling using Mallet implementation of LDA and LIWC to find health related topics and subtopics |
|  | Mejova et al, UK and Qatar using USA data (13) | Instagram  Foursquare | 20K  194K | Designed machine learning model to analyse posts for food consumption patterns with the aim to correlate with obesity rates and location of fast food restaurants.  Taxonomy: content analysis | 🗸 | 🗸 | 🗸 | 🗸/x  Datascience toolkit API  Instagram Endpoints API  Crowdflower |
|  | Nguyen et al. 2016, USA (14) | Twitter | 80 million geotagged tweets | Validated machine learning for constructing indicators of happiness, food and physical activity via location (neighbourhoods) compared to census data  Taxonomy: content analysis | 🗸  Twitter API and geolocator | 🗸  Extracted 1% tweets using Twitter API and geolocator and used Python to assign zipcode | 🗸  Divided text into tokens using Stanford Tokenizer | 🗸  Sentiment analysis using machine learning (MALLET) and extraction of food and physical activity mentions and frequencies matched to constructed food list |
|  | Nguyen et al. 2017, USA (15) | Twitter | 422,094 tweets | Collected food and physical activity tweets sent from one USA state (Utah) and compared sentiment to health data on overweight and obesity to demonstrate validity of Twitter-derived neighborhood characteristics  Taxonomy: content analysis | 🗸  Twitter API and geolocator | 🗸  Extracted 1% tweets using Twitter API and geolocator and used Python to assign zipcode | 🗸  Divided text into tokens using Stanford Tokenizer | 🗸  Sentiment analysis using machine learning (MALLET) and extraction of food and physical activity mentions and frequencies matched to constructed food list |
|  | Nguyen et al. 2017, USA (16) | Twitter  Yelp | 4,041,521 tweets  505,554 reviews | Collected food data from Twitter and Yelp to characterize food environments of different locations to create a state level food database in the USA and compare to Government data.  Taxonomy: content analysis | 🗸 | 🗸  Twitter API  Yelp search API  OpenStreetMap | 🗸 | 🗸  MALLET  Python |
|  | Ofli et al. 2017, USA and Qatar(17) | Instagram | 1.9M images | Collected Instagram food images and hashtags and used image recognition technology to study differences in how a human describes the food compared to machine learning.  Taxonomy: content analysis | 🗸 | 🗸  Python using Shapely | 🗸 | x/🗸  Manual classification using Amazon Mechanical Turk |
|  | Pang et al. 2015, USA(18) | Instagram | 195,000 images | Collected Instagram images with alcohol hashtags and used facial recognition technology to determine patterns of drinking in underage consumers.  Taxonomy: content analysis | 🗸/ x | 🗸 | 🗸 | 🗸  Face processing toolkit Face++ |
|  | Phan et al. 2019, Switzerland(19) | Instagram | 1.6M | Analysed Instagram alcohol related posts and compared to an existing dataset in order to develop machine learning systems that monitor alcohol consumption.  Taxonomy: content analysis | 🗸 | 🗸 | 🗸 | 🗸 |
|  | Primack et al. 2015, USA(20) | YouTube | 200 videos retrieved and 70 videos coded related to alcohol intoxication | Collected and manually coded YouTube videos for sentiment related to alcohol intoxication to characterize the popular content  Taxonomy: content analysis | 🗸  Keyword search on platform to obtain popular videos by views | x  Manual  collection of views, likes, dislikes | x  Manual  sorting | x  Manual  coding including of content sentiment using codebook |
|  | Rahman et al. 2016, Bagladesh(21) | Foursquare and Twitter | 731 Twitter users | Collected tweets on Foursquare restaurant checkins and designed predictive machine learning model with the aim to assess if tweet words can predict eating out preferences.  Taxonomy: content analysis | 🗸  Greptweet.com | 🗸 | 🗸 | 🗸  LIWC  WEKA |
|  | Rich et al. 2016, UK(22) | Instagram | 800,000 images | Analysed Instagram food images with the aim to improve machine learning food image recognition.  Taxonomy: content analysis | 🗸  Instagram API | 🗸  Instagram API | 🗸 | 🗸 |
|  | Shah et al. 2020, USA (23) | Twitter | ~10M tweets  50,000 food subset | Designed a high performance machine learning model to classify food tweets across Canada and demonstrated correlation between the location of high caloric food tweets and Government consumption data.  Taxonomy: Content analysis | 🗸  Twitter API | 🗸  Customised machine learning model | 🗸  Customised machine learning model | 🗸  Customised machine learning model |
|  | Sharma et al. 2015, USA (24) | Instagram | 1.8M | Collected posts with food hashtags and designed a machine learning model to link posts to USDA National Nutrient Database and estimate calorie information.  Taxonomy: content analysis | 🗸 | 🗸  Instagram API | 🗸 | x/🗸 |
|  | Silva et al. 2017, Brazil and UK (25) | Foursquare | Varied ~5M | Collected data from location based network (Foursquare) checkins on eating and drinking habits and proposed a methodology to identify boundaries and similarities across populations.  Taxonomy: content analysis | 🗸 | 🗸  Foursquare/Twitter API | 🗸 | 🗸 |
|  | Sullivan et al. 2016, USA (26) | Amazon.com user-generated reviews | 40,0000 reviews of 2,708 products | Collected Amazon.com reviews of nutritional supplements and used unsupervised natural language processing techniques to capture adverse reactions and design system to score products on adverse reaction potential  Taxonomy: content analysis and surveillance | 🗸  Web crawler to obtain reviews under set categories e.g. “Herbal supplement” | 🗸  Web crawler to obtain reviews under set categories e.g. “Herbal supplement” | 🗸  Topic modelling using unsupervised Latent Dirichlet Modelling (LDA) and subfield of Natural Language Processing | 🗸  Topic modelling using unsupervised Latent Dirichlet Allocation (LDA) a subfield of Natural Language Processing to develop scoring system of risk for each product |
|  | Sun, et al. 2018, China and UK using USA data (27) | Twitter | 41 million tweets (compared to USA CDC data) from 110 major cities over 12 months | Collected and analysed tweets to show relation with Government obesity data and Twitter derived dietary habits of users.  Taxonomy: content analysis (view to surveillance) | 🗸 | 🗸 | 🗸 | 🗸 |
|  | Turner-McGrievy et al. 2015, USA (28) | Twitter | ? | Collected tweets over 1 year with #fitness, #diet, #health and “weight” to explore temporal trends (influence of holiday period on frequency).  Taxonomy: content analysis, surveillance | 🗸  Hashtagify.me | 🗸  peoplebrowsr | ? | ? |
|  | Vidal et al. 2015, Uruguay & NZ (29) | Twitter | 48,746 original tweets and then subset 16,000 per 4 topics and manual coding 12,000 | Explored main topics in tweets on breakfast, lunch, snack and dinner to evaluate Twitter’s potential as a research tool.  Taxonomy: content analysis | 🗸  Tweets containing key words breakfast, lunch, dinner and snack from researchers Twitter account | 🗸  TwitterR Package | 🗸  TwitterR Package removed duplicates and retweets and cleaned data | x  Manual coding of subset of tweets |
|  | Vydiswaran et al., 2019, USA (30) | Twitter | 21.19M Tweets  Multiple subsets | Demonstrated that Twitter can be used to characterize neighborhood-level food-related behaviours and attitudes / sentiment in food-related tweets.  Taxonomy: content analysis | 🗸  Twitter API | 🗸 | 🗸/x | 🗸/x  Manual coding of subsets |
|  | Widener et al. 2014, USA (31) | Twitter | 148,533 tweets | Collected geolocated Twitter data in the USA and applied advanced data-mining framework to explore if prevalence of healthy and unhealthy food mentions could be mapped in line with USDA data.  Taxonomy: content analysis | 🗸  Twitter API | 🗸 | 🗸 | 🗸  Alchemy API for sentiment analysis |
|  | Wombacher et al. 2017, USA (32) | Twitter | 851 tweets | Analysed Twitter data using hashtag #NeckNominate to identify normative forces at play in a popular, dare based alcoholic drinking game  Taxonomy: content analysis | 🗸  Tweets detected using Topsy social media analytics website | 🗸  Tweets detected using Topsy social media analytics website | ?  undescribed | x  Manual coding of text, themes, sentiment, links, retweets |
|  | Yan et al. 2015, China (33) | Koubei (online community) | 10,136 restaurant reviews in Harbin China location on Koubei | Analysed online restaurant reviews to assess revisit intent and obtain evaluation indicators of which healthy was important in food quality.  Taxonomy: content analysis | ? | ? | ? | 🗸  Used text mining technology |
|  | Zhou et al. 2018, China (34) | Weibo (Chinese) | 3,975,800 microblogs | Used popular Chinese social media site to detect different preferences in types of meals using sentiment analysis and trends over time.  Taxonomy: content analysis and surveillance | 🗸 | 🗸 | 🗸 | 🗸 |

\*Social media analytics (SMA) key:

🗸 SMA was described for the relevant step

X Manual extraction, coding or analysis by humans was described for the relevant step

🗸/x A combination of SMA and manual processes were described for the relevant step

? The details could not be clearly identified for the relevant step

**Table 2B Data chart: Study author(s) disciplines or fields of research (alphabetical by first author) where identifiable\*\***

|  |  |  |  |
| --- | --- | --- | --- |
| **Study** | **SoMe**  **Platform(s)** | **Health, Medicine, other related** | **Business, Informatics, Computer Science, other related** |
| Abbar et al. (1) | Twitter | x | Computing |
| Alhabash et al. (2) | Twitter | Health | Computer science, Advertising, Data mining |
| Alajajian et al. | Twitter | Culinary Arts and Food Science | Mathematical and Computer Science |
| Blackstone et al. (4) | Facebook | Public Health, Health Science | x |
| Carrotte et al. (5) | Instagram  Facebook  Twitter, Tumblr | Health, Medicine | x |
| Cavazos-Rehg et al. (6) | Twitter | Psychiatry, Medicine | *Assisted by external social media analytics companies* |
| Chen et al. (7) | Twitter | x | Geography |
| ElTayeby et al. (8) | Facebook group | Health Science | x |
| Fried et al. (9) | Twitter | ? | ? |
| Huang et al. (10) | Twitter | Public Health | Statistics, Engineering |
| Karami et al. (12) | Twitter | Public Health | Information Science |
| Kershaw et al. (11) | Twitter | x | Computing, Management Science |
| Mejova et al. (13) | Instagram  Foursquare | ? | Computing |
| Nguyen et al. (14) | Twitter | Health, Global Health | Computing, Geography, Sociology |
| Nguyen et al. (15) | Twitter | Epidemiology, Biostatistics, Public Health | Computing, Geography, Sociology |
| Nguyen et al. (16) | Twitter  Yelp | Epidemiology, Health | Computing, Geography, Sustainability |
| Ofli et al. (17) | Instagram | x | Computing |
| Pang et al. (18) | Instagram | ? | ? |
| Phan et al. (19) | Instagram | x | Artificial Intelligence |
| Primack et al. (20) | YouTube | Medicine, Public Health | x |
| Rahman et al. (21) | Twitter  Foursquare | x | Engineering and Technology |
| Rich et al. (22) | Instagram | x | Engineering, Computer Science |
| Shah et al. | Twitter | Medicine | Mathematics and computer science |
| Sharma et al. (24) | Instagram | x | Technology |
| Silva et al. (25) | Foursquare | x | Computer Science, Informatics, Geography |
| Sullivan et al. (26) | Amazon | Biomedical Informatics | Biomedical Informatics |
| Sun et al. (27) | Twitter | x | Computer science, Network computing and security |
| Turner-McGrievy et al.(28) | Twitter | Health Promotion, Exercise Science, RD\* | x |
| Vidal et al. (29) | Twitter | Psychology, Food | x |
| Vydiswaran et al. | Twitter | Public Health, Nutritional Science\* | Information, Urban planning |
| Widener et al. (31) | Twitter | x | Geography, Geospatial analysis and computation |
| Wombacher et al. (32) | Twitter | Health | Communication |
| Yan et al. (33) | Koubei | x | Management Science and Engineering |
| Zhou et al. (34) | Weibo | x | Information management |

\*Inclusion of a nutritionist or dietitian in the research team. RD is the USA credential for Registered Dietitian

\*\*Studies where author affiliation not clearly identifiable: Fried et al., Mejova et al., Pang et al.

**Table 2C Data chart: Ethical approval, mentions and considerations (alphabetical by first author)**

|  |  |  |
| --- | --- | --- |
| **Study** | **SoMe**  **Platform(s)** | **Ethics approval or mention** |
| Abbar et al. (1) | Twitter | x |
| Alhabash et al. (2) | Twitter | x |
| Alajaijan et al. | Twitter | x |
| Blackstone et al. (4) | Facebook | Approved by the IRB at the researchers’ universities. |
| Carrotte et al. (5) | Twitter  Facebook  Tumblr  Instagram | x |
| Cavazos-Rehg et al. (6) | Twitter | The researchers stated: “The Twitter data in this current study are public. Washington University’s Institutional Review Board reviewed our study protocol, and our research received an institutional review board exemption.” |
| Chen et al. (7) | Twitter | x |
| ElTayeby et al. (8) | Facebook group | The researchers stated that: “We deidentified posts…”. |
| Fried et al. (9) | Twitter | x |
| Huang et al. (10) | Twitter | x |
| Karami et al. (12) | Twitter | The researchers stated: “This paper used the real-time method to randomly collect 10% of publicly available English tweets…” |
| Kershaw et al. (11) | Twitter | x |
| Mejova et al. (13) | Instagram  Foursquare | x |
| Nguyen et al. (14) | Twitter | x |
| Nguyen et al. (15) | Twitter | The University of Utah Institutional Review Board |
| Nguyen et al. (16) |  | The University of Utah Institutional Review Board |
| Ofli et al. (17) | Instagram | x |
| Pang et al. (18) | Instagram | x |
| Phan et al. (19) | Instagram | x |
| Primack et al. (20) | YouTube | x |
| Rahman et al. (21) | Twitter  Foursquare | x |
| Rich et al. (22) | Instagram | x |
| Shah et al. (23) | Twitter | x |
| Sharma et al. (24) | Instagram | x |
| Silva et al. (25) | Foursquare | x |
| Sullivan et al. (26) | Amazon.com reviews | x |
| Sun, et al. (27) | Twitter | x |
| Turner-McGrievy et al. (28) | Twitter | The researchers stated that: “All procedures, including the informed consent process, were conducted in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000”. |
| Vidal et al. (29) | Twitter | *x* |
| Vyidiswaran et al. | Twitter | University of Michigan Institutional Review Board |
| Widener et al. (31) | Twitter | x |
| Wombacher et al. (32) | Twitter | The researchers stated: “As messages on Twitter are publicly available, it was the most direct method of gathering data…”. |
| Yan et al. (33) | Koubei | x |
| Zhou et al. (34) | Weibo (Chinese) | x |

**References:**

1. Abbar S, Mejova Y Weber I (2015) You Tweet What You Eat: Studying Food Consumption Through Twitter, vol. 2015-, pp. 3197-3206.

2. Alhabash S, VanDam C, Tan PN *et al.* (2018) 140 Characters of Intoxication: Exploring the Prevalence of Alcohol-Related Tweets and Predicting Their Virality. *SAGE Open* 8.

3. Alajajian SE, Williams JR, Reagan AJ *et al.* (2017) The Lexicocalorimeter: Gauging public health through caloric input and output on social media.(Research Article)(Report). *PLoS ONE* 12, e0168893.

4. Blackstone SR & Herrmann LK (2018) Extreme body messages: themes from Facebook posts in extreme fitness and nutrition online support groups. *Began with 2015* 4, 33.

5. Carrotte ER, Prichard I Lim MSC (2017) "Fitspiration" on Social Media: A Content Analysis of Gendered Images. *Journal of medical Internet research* 19, e95.

6. Cavazos-Rehg PA, Krauss MJ, Sowles SJ *et al.* (2015) "Hey Everyone, I'm Drunk." An Evaluation of Drinking-Related Twitter Chatter. *Journal of studies on alcohol and drugs* 76, 635.

7. Chen X & Yang X (2014) Does food environment influence food choices? A geographical analysis through “tweets”. *Applied Geography* 51, 82-89.

8. ElTayeby O, Eaglin T, Abdullah M *et al.* (2018) A feasibility study on identifying drinking-related contents in Facebook through mining heterogeneous data. *Health Informatics Journal*, 1460458218798084.

9. Fried D, Surdeanu M, Kobourov S *et al.* (2014) Analyzing the language of food on social media, pp. 778-783.

10. Huang T, Elghafari A, Relia K *et al.* (2017) High-resolution temporal representations of alcohol and tobacco behaviors from social media data. *Proceedings of the ACM on Human-Computer Interaction* 1.

11. Kershaw D, Rowe M Stacey P (2014) Towards tracking and analysing regional alcohol consumption patterns in the UK through the use of social media, pp. 220-228.

12. Karami A, Dahl AA, Turner-McGrievy G *et al.* (2018) Characterizing diabetes, diet, exercise, and obesity comments on Twitter. *International Journal of Information Management* 38, 1-6.

13. Mejova Y, Haddadi H, Noulas A *et al.* (2015) #FoodPorn: Obesity Patterns in Culinary Interactions. In *Proceedings of the 5th International Conference on Digital Health 2015*. Florence, Italy: ACM.

14. Nguyen QC, Li D, Meng HW *et al.* (2016) Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity. *JMIR Public Health and Surveillance* 2, e158.

15. Nguyen QC, Brunisholz KD, Yu W *et al.* (2017) Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Scientific Reports* 7.

16. Nguyen QC, Meng H, Li D *et al.* (2017) Social media indicators of the food environment and state health outcomes. *Public Health* 148, 120-128.

17. Ofli F, Aytar Y, Weber I *et al.* (2017) Is Saki #delicious?: The Food Perception Gap on Instagram and Its Relation to Health. In *Proceedings of the 26th International Conference on World Wide Web*. Perth, Australia: International World Wide Web Conferences Steering Committee.

18. Pang R, Baretto A, Kautz H *et al.* (2015) Monitoring adolescent alcohol use via multimodal analysis in social multimedia. In *Proceedings of the 2015 IEEE International Conference on Big Data (Big Data)*: IEEE Computer Society.

19. Phan T-T, Muralidhar S Gatica-Perez D (2019) Drinks & Crowds: Characterizing Alcohol Consumption through Crowdsensing and Social Media. *Proceedings of ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 1-30.

20. Primack BA, Colditz JB, Pang KC *et al.* (2015) Portrayal of Alcohol Intoxication on YouTube. *Alcoholism: Clinical & Experimental Research* 39, 496-503.

21. Rahman MM, Majumder MTH, Mukta MSH *et al.* (2016) Can we predict eat-out preference of a person from tweets? In *Proceedings of the 8th ACM Conference on Web Science*. Hannover, Germany: ACM.

22. Rich J, Haddadi H Hospedales TM (2016) Towards Bottom-Up Analysis of Social Food. In *Proceedings of the 6th International Conference on Digital Health Conference*. Montreal, Quebec, Canada: ACM.

23. Shah N, Srivastava G, Savage DW *et al.* (2020) Assessing Canadians Health Activity and Nutritional Habits Through Social Media. *Frontiers in Public Health* 7.

24. Sharma S & De Choudhury M (2015) Measuring and Characterizing Nutritional Information of Food and Ingestion Content in Instagram, pp. 115-116.

25. Silva TH, Melo POSVd, Almeida JM *et al.* (2017) A large-scale study of cultural differences using urban data about eating and drinking preferences. *Inf Syst* 72, 95-116.

26. Sullivan R, Sarker A, O'Connor K *et al.* (2016) Finding Potentially Unsafe Nutritional Supplements from User Reviews with Topic Modeling. *Pacific Symposium on Biocomputing* 21, 528-539.

27. Sun Q, Wang N, Li S *et al.* (2018) Local spatial obesity analysis and estimation using online social network sensors. *Journal of Biomedical Informatics* 83, 54-62.

28. Turner-McGrievy GM & Beets MW (2015) Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. *Translational Behavioral Medicine* 5, 160-166.

29. Vidal L, Ares G, Machín L *et al.* (2015) Using Twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations". *Food Quality and Preference* 45, 58-69.

30. Vydiswaran VGV, Romero DM, Zhao X *et al.* (2020) Uncovering the relationship between food-related discussion on Twitter and neighborhood characteristics. *Journal of the American Medical Informatics Association : JAMIA* 27, 254-264.

31. Widener MJ & Li W (2014) Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography* 54, 189-197.

32. Wombacher K, Reno JE Veil SR (2017) NekNominate: Social Norms, Social Media, and Binge Drinking. *Health Communication* 32, 596-602.

33. Yan X, Wang J Chau M (2015) Customer revisit intention to restaurants: Evidence from online reviews. *Information Systems Frontiers* 17, 645-657.

34. Zhou Q & Zhang C (2018) Detecting users' dietary preferences and their evolutions via Chinese social media. *Journal of Database Management* 29, 89-110.