

RESEARCH

Open Access



# Modeling the public attitude towards organic foods: a big data and text mining approach

Anupam Singh\*  and Aldona Glińska-Neweś

\*Correspondence:  
anupam.nitb@gmail.com  
Nicolaus Copernicus  
University, Torun, Poland

## Abstract

This study aims to identify the topics that users post on Twitter about organic foods and to analyze the emotion-based sentiment of those tweets. The study addresses a call for an application of big data and text mining in different fields of research, as well as proposes more objective research methods in studies on food consumption. There is a growing interest in understanding consumer choices for foods which are caused by the predominant contribution of the food industry to climate change. So far, customer attitudes towards organic food have been studied mostly with self-reported methods, such as questionnaires and interviews, which have many limitations. Therefore, in the present study, we used big data and text mining techniques as more objective methods to analyze the public attitude about organic foods. A total of 43,724 Twitter posts were extracted with streaming Application Programming Interface (API). Latent Dirichlet Allocation (LDA) algorithm was applied for topic modeling. A test of topic significance was performed to evaluate the quality of the topics. Public sentiment was analyzed based on the NRC emotion lexicon by utilizing *Syuzhet* package. Topic modeling results showed that people discuss on variety of themes related to organic foods such as plant-based diet, saving the planet, organic farming and standardization, authenticity, and food delivery, etc. Sentiment analysis results suggest that people view organic foods positively, though there are also people who are skeptical about the claims that organic foods are natural and free from chemicals and pesticides. The study contributes to the field of consumer behavior by implementing research methods grounded in text mining and big data. The study contributes also to the advancement of research in the field of sustainable food consumption by providing a fresh perspective on public attitude toward organic foods, filling the gaps in existing literature and research.

**Keywords:** Organic foods, Text mining, Big data, Sentiment analysis, Latent Dirichlet Allocation (LDA), Topic modeling, Machine learning

## Introduction

The most common approach to study customer behavior, including food-related beliefs and emotions, uses the self-reported method, e.g., participants declare themselves their feelings and emotions in interviews or verbal questionnaires [1, 2]. It stands for

significant limitation of such studies. While self-report methods have many advantages, they might be affected by biases such as social desirability bias, response bias, and common method bias [3, 4]. Moreover, some problems are very hard to capture; for example, emotions are, by nature, subjectively experienced in a specific context, so their objective assessment is very difficult [5, 6]. Wang et al. [7] and Feizollah et al. [8] argue that state-of-art computational techniques based on machine learning methods can potentially eliminate such methodological issues.

There is growing interest in understanding consumer choices for foods [9, 10]. It is motivated, first of all, by climate change a serious global issue that poses an urgent and perhaps one of the greatest challenges facing humankind [11, 12]; the food industry contributes predominantly to these effects. Agriculture accounts for over a quarter of global greenhouse gas emissions; half of the total habitable land is used for agriculture; it uses 70% of global freshwater withdrawals and causes 78% of waterways' pollution; 94% of mammal biomass (excluding humans) is livestock [13, 14]. Unsustainable food consumption and its tangible consequences on human health and the environment, direct the public attention to organic foods which are viewed as ecological and healthy [9, 10]. European Union (EU) describes organic production as "an overall system of farm management and food production that combines best environmental practices, a high level of biodiversity, the preservation of natural resources and the application of high animal welfare standards" [15].

The objective of this study is to identify the topics that users post on Twitter about organic foods and to analyze the emotion-based sentiment of those tweets. People find social media platforms convenient to share their experiences, attitude, and opinion. The rise in the number of active users and users' contents on Twitter delivers the scope for scientific exploration of public attitudes. In the present study, we extract the latent topics on organic foods from Twitter data and analyze the public sentiments about such foods. The study is the first of its kind, in this line of research, to use text mining based on unsupervised machine learning technique, which eliminates the limitations of sample size and subjectivity of conventional methods.

So far, the studies on food consumption have utilized rather subjective traditional research methods, such as questionnaire surveys and interviews [16–19]. Although the body of research on organic food consumption is quite large, the question of why people buy organic food is still not fully understood [16, 20, 21]. Moreover, research shows that while consumers report favorable attitudes toward organic food, they often do not display subsequent buying actions; in other words, there is the intention-behavior gap in the consumer decision-making process regarding organic food purchases [22–24]. Adoption of more objective research methods in the studies on food consumption can help to explain this gap.

To the best of our knowledge, there has been no study analysing organic food consumption behaviour using big data and machine learning techniques. Danner and Menapace [10] explored beliefs regarding organic foods but based on content analysis of online comments posted in news websites and forums. However, our study uses machine learning-based text mining techniques to identify the most discussed themes on organic foods, and analyses also the underlying emotions and sentiments about the same. The study contributes to the field of consumer behavior by implementing research methods

grounded in text mining and big data. The study contributes also to the advancement of research in the field of sustainable food consumption by providing a fresh perspective on public attitude toward organic foods using the aforementioned objective methods of data collection and analysis, filling the gaps in existing literature and research.

In the following, the relevant literature on sustainable food consumption, and consumer beliefs and emotions about organic foods are reviewed. Subsequently, the methodology used in the study is explained, followed by results and discussion. Finally, the conclusion of the study is drawn with the limitations and the suggestions for future research.

### Literature review

Organic food production is based on environmentally sustainable techniques excluding chemical or synthetic additives; its ecological nature strongly supports ecosystems and bio-diversity [25–27]. Organic food production and consumption are also considered as animal-friendly and related to fairness and human health [28, 29]. Recent years have brought increased consumers' interest in organic food. Its global market is valued at 96.7 billion USD and a growing number of countries adopts distinct regulations for its cultivation [30]. Organic food production and trading are among the fastest-growing industries in the food market in Europe, South America, Oceania, and Japan [31].

Customers buy organic foods due to rising food safety concerns [17] and health consciousness [18], but also because of its taste [32–34]. Consumer motivations for choosing organic foods are categorized as related to health, environment, and social consciousness [16]. A rising level of concern for environmental and ecological welfare supports organic food consumption as animal friendly [20, 35] and as a counter-option to conventional food production using chemical, synthetic and genetically modified means [24, 30, 36]. Other non-egotistical factors of organic food consumption include supporting local producers [37] and reflect concern for society and consumer social surroundings [19, 38].

Recent studies suggest that perceived values (utilitarian and hedonic) influence consumer willingness to buy organic foods [39, 40]. It is the perceived value of a product from the multisensory and emotive aspects. The utilitarian value is the assessment of functional benefits and drawbacks of a product or service [41, 42]. In the context of organic foods, the utilitarian value refers to the nutritional quality, purity, safety, and healthy attributes [43]. The hedonic value of organic food refers to the pleasure felt from the taste and the freshness and purity [44].

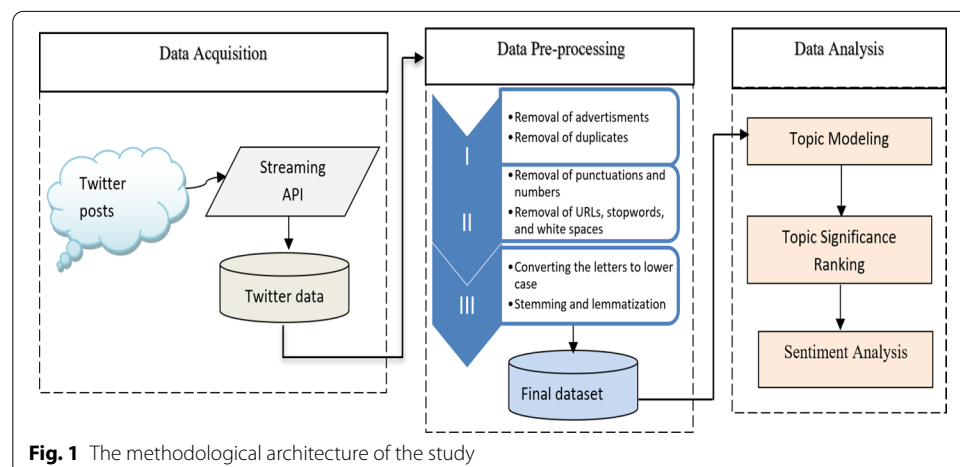
Concerning emotions influencing sustainable customer behavior, the picture is also very complex. Literature delivers evidence for the consequences of both negative and positive emotions in this regard [45]. Among negative emotions, e.g., guilt is argued to influence sustainable behaviors because of consumer feelings of moral responsibility for the environment [45, 46]. Also, sadness was shown to affect more pro-environmental behaviors, such as giving higher monetary donations to sustainable causes [47]. On the other hand, consumers are likely to engage in sustainable behaviors when they derive positive emotions from the behavior [45, 48]. For example, joy and pride influence intentions to decrease plastic water bottle usage, while optimism can motivate to maintain such behaviors over time [45, 49]. However, positive emotions can also encourage

unsustainable consumer behaviors, such as driving gas-powered automobiles [50], so probably also purchasing unhealthy food. Associations between food and emotions influencing food acceptance and preference gained attention in consumer and sensory research e.g. [51, 52].

The review of the literature suggests that the big data and text mining approaches have been implemented by the researchers in different contexts. For example, Wang et al. [53] studied the attitude of the Chinese public toward off-site construction using Sina Weibo data, while Wu et al. [7] explored the attitude of the Chinese public toward municipal solid waste sorting policy; Feizollah et al. [8] studied halal tourism using Twitter data; Philander and Zhong [54] utilized text mining techniques to explore hospitality customers' attitudes and perceptions using Twitter data; Prameswari et al. [55] applied text mining techniques on online hotel reviews to identify priority tourist destinations in Indonesia; Oliveira et al. [56] used Twitter data to predict stock market behavior. In these studies, it has been argued that text mining techniques used to explore the comments posted, for example, on Twitter, give an insight into public attitudes in the more objective way than traditional research methods. Therefore, addressing the calls for the application of big data and text mining in different fields of research, the present study uses big data and text mining approach to investigate public attitudes about organic foods.

## Research method

This section introduces the methods used in the study, followed by the subsections outlining each stage in detail. Figure 1 depicts the methodological architecture of the study. As shown in Fig. 1, the first step involved data acquisition. Tweets are collected via Twitter APIs using the keyword. In the second step, the tasks of data preprocessing were performed (please refer to the “Data collection” and “Pre-processing” sections for more details). In the third step, the final data obtained from the second step was analyzed. Topic modeling was applied for topic extraction followed by topic significance ranking. In the last step, emotion-based sentiment analysis was performed, wherein the underlying emotions in the tweets were calculated and visualized the overall results.



### Data collection

Twitter posts were extracted for this study. Twitter is a very popular online social network for mass communication and information dissemination [57] and referred to as the gold mine of data [58, 59]. People find Social Networking Sites (SNS) a platform to share their views, opinions, information, and ideas at any time, thereby leaving the scopes for scientific research. The data is easily accessible and is useful for those studying public attitudes, opinions, beliefs, and behaviors [60]. In this study, tweets were obtained by using Twitter Streaming Application Programming Interface (API). R 4.0.4 and Python 3.9 were used for the data extraction, storage, pre-processing, data mining, and result visualization. *Tweepy*, a popular Python library, was used for accessing the Twitter API. The tweets were streamed for the keyword “Organic Food”. Only tweets written in English were considered, with no geographical restriction. The extracted dataset has the Tweet object which includes fundamental attributes such as ‘created\_at’ (the time when this tweet was created), and ‘text’ (tweets), screen\_name, and geo-tagged location. However, we have taken only tweets in our analysis.

### Data pre-processing

Social media data is interesting but is unstructured. A piece of meaningful information can be drawn if the quality data is ensured. The unstructured and semi-structured data need pruning to get an insight from it. The quality of Twitter data in this study was ascertained through the framework of accuracy proposed by [61], followed by extensive data pre-processing and cleaning. The accuracy was ascertained keeping in mind the query error, interpretation error, and coverage error during the data collection phase as suggest in the study of [61].

We used Microsoft Excel and R Programming Language for data pre-processing. We adopted the following three steps for data pre-processing. Denoising of data was done in the first step. In here, advertisements, duplicate data, and irrelevant posts were deleted. In the second step, stop words, URLs, symbols, and the words that appear frequently and do not contribute to sentiment analysis were also removed. We employed the ‘tm’ package to remove such contents from the text data. Stop words, such as “is”, “the”, “an”, “how”, and “what”, were eliminated as these words do not contain relevant information. Similarly, URLs and numbers also do not contain any information, thus were eliminated. Finally, in step 3, words were converted into the lower case before stemming and lemmatization was done. Thus, the cleaned data obtained was considered in the final analyses of the study.

### Topic modeling

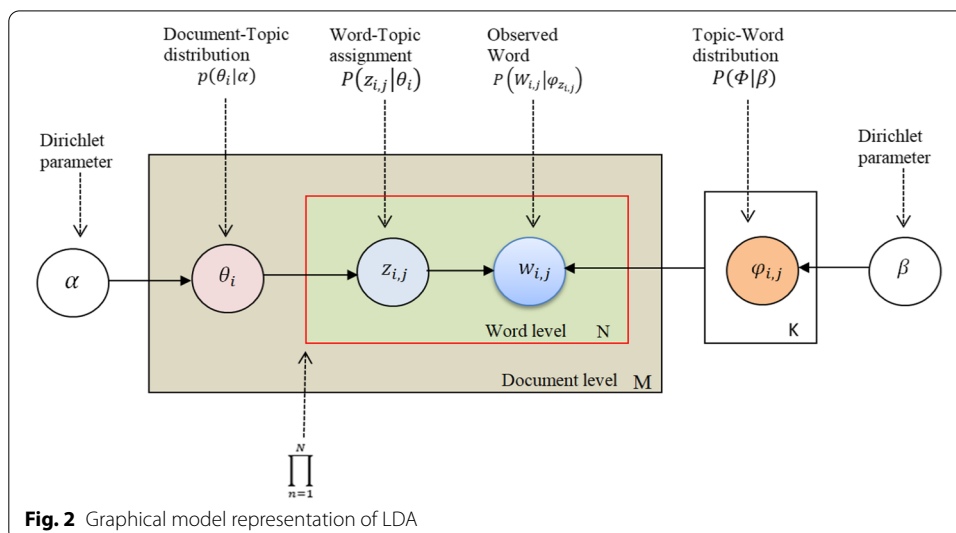
Latent semantic analysis (LSA), probabilistic latent semantic analysis (pLSA), and LDA are the three most commonly used topic modeling techniques [62]. We adopted LDA in this study for the following reasons. LSA is an effective dimensionality reduction technique, which can achieve significant compression in large collections while capturing some aspects of basic linguistic notion. However, the drawback of LSA is its inability of fitting a model to data to represent the documents into multiple topics. Although the pLSA approach solves the representation challenge in LSA, it cannot

generalize to unseen documents due to the lack of a well-defined generative model. The method is also prone to overfitting [63], which is a serious issue, whereas LDA is capable of performing dimensionality reduction, semantic annotation, mixture modeling, and generalization ability without such major issues [62]. Since this study is based on the cross-sectional research design, analysis of the temporal pattern of associated topics and topic correlation is out of the scope of this paper. Therefore, the topic modeling techniques like ‘Topics over Time’, ‘Dynamic Topic Model’ and ‘Pachinko Allocation Model’ are ruled out. Thus, LDA-based topic modeling suits well for achieving the objectives of this study.

LDA is a sophisticated unsupervised machine learning technique for identifying the key topics from large textual data [63]. It has been widely used in the field of natural language processing (NLP) among several topic modeling algorithms [7, 64]. It is a three-stage Bayesian probabilistic model. The model gives the topic of each document in the documents (D) with a distribution over the topics (K), and each topic is a multinomial distribution over words (W) in the corpus. The statistical model adopts the following three stages generative process (Fig. 2) for a corpus D of M documents with length  $N_i$  of each;  $\alpha$  is the Dirichlet parameter of per-document topic distributions, and  $\beta$  is the Dirichlet parameter of per-topic word distribution. And,  $\theta_i$  and  $\phi_k$  denote the topic distribution for the document  $i$  and the word distribution for topic  $k$ , respectively.

Based on the definition, LDA assumes the following generative processes for each tweet ( $w$ ) in a corpus ( $D$ ):

1. Choose  $N \sim \text{Poisson}(\xi)$ , where  $N$  denotes the length of documents;
2. Draw a topic distribution, choose  $\theta_d \sim \text{Dir}(\alpha)$ , where  $\alpha$  is the parameter of the Dirichlet prior on the per-tweet topic distributions; and
3. For each of  $N$  words  $w_n$ :
  - (a) Choose a topic  $z_n \sim \text{Multinomial}(\theta)$ ; and



- (b) Choose a word  $w_n$  from  $p(w_n|z_n, \beta)$ , a multinomial probability conditioned on the topic  $z_n$ .

Step 1 indicates the selection of number of words  $N$  for a document, based on a Poisson distribution. The dimensionality  $k$  of the Dirichlet distribution is assumed known and fixed. Step 2 refers to the probability that the dimension (topic) occurs and is represented as a  $k$  dimensional random variable  $\theta$  in a given document (tweet). According to probability density function, it can be expressed as follows:

$$p(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^k \alpha_i\right)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k \theta_i^{\alpha_i-1} \quad (1)$$

The variable  $\theta$  can take values in the  $(k-1)$ -simplex  $\theta_1, \dots, \theta_{k-1} > 0, \sum_{i=1}^{k-1} \theta_i < 1, \sum_{i=1}^k \theta_i = 1$ , where the parameter  $\alpha$  is a  $k$ -vector with components  $\alpha_i > 1$ , and  $\Gamma(x)$  is the gamma function. Equation 1, a probability density function of Dirichlet distribution, is a conjugate prior for the multinomial distribution, which facilitates the development of inference for Eq. (2).

In step 3, the given parameters  $\alpha$  and  $\beta$ , the joint distribution of a topic mixture  $\theta$ , a set of  $N$  topics  $z$ , and a set of  $N$  words  $w$  is expressed as:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta) p(w_n|z_n, \beta) \quad (2)$$

where  $p(\theta|\alpha)$  refers to tweet-level (i.e., document-level) parameters, while the other terms refer to word-level parameters. And,  $p(z_n|\theta)$  is  $\theta_i$  for the unique  $i$  where  $z_n^i = 1$ .

We can obtain the marginal distribution of a tweet in our corpus by using Eq. (3):

$$P(w|\alpha, \beta) = \int p(\theta|\alpha) \left( \prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta \quad (3)$$

Finally, the probability of a corpus can be obtained by taking the product of the marginal probabilities of single documents (i.e., tweets, in our case):

$$P(D|\alpha, \beta) = \prod_{d=1}^M \int P(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{d_n}} P(z_{d_n}|\theta_d) P(w_{d_n}|z_{d_n}, \beta) \right) d\theta_d \quad (4)$$

where the term  $\theta_d$  is document-level variable assumed to be sampled per tweet. The variables  $z_{d_n}$  and  $w_{d_n}$  are word-level variables and are sampled once for each word in each tweet. The graphical representation is illustrated by Fig. 2 [63].

The model statistically discovers the group of topics that best explains the observed words in the entire documents. We utilized 'topicmodels' library of the R package for this task. *LDavis*, a web-based interactive visualization tool, was also utilized for the visualization of topic models.

### Sentiment analysis

Sentiment analysis (also known as opinion mining or emotion AI) is located at the heart of NLP, text mining, and computational linguistics. It measures the subjective



information extracted from textual data. In this era of the internet and social media, sentiment analysis has gained much popularity among researchers from different domains for its applicability in a wide range of fields such as psychology, linguistics, public health, and business management [65]. In the current study, the word-level sentiment analysis was performed. NRC Emotion Lexicon was used to list the words and their association with eight basic emotions (trust, fear, surprise, sadness, disgust, anger, anticipation, and joy) and two sentiments (negative and positive). *Syuzhet* package was utilized to analyze the underlying sentiments. The algorithm calculates the presence of eight different emotions and sentiments in the text dataset. It assigns each token (i.e., word) in the corpus a score corresponding to each emotion and sentiments; '1' if it is associated, otherwise '0'. Thus, a data frame containing sentiment scores is obtained, where each row represents a word and columns represent eight different emotions and two sentiments.

## Results

The data for this study was collected from Twitter. The tweets posted between January 10, 2021 and March 7, 2021 was extracted using Python. A total of 43,724 tweets were collected. However, after removing the duplicates, 41,009 tweets are processed for further analysis. Before the topic analysis, it is necessary to evaluate the models.

### Topic models' evaluations

LDA topic modeling is based on probabilistic inference; hence, requires a huge amount of data and tuning to get reasonable results [66]. We used two measures- perplexity and topic coherence- to evaluate the best underlying topics in the corpus. Perplexity is defined as the inverse probability of the test set, normalized by the number of words. It is an intrinsic evaluation metric and is measured as the normalized log-likelihood of a test set data. It is a general quality measure for the entire model [63], and captures how well a probability model predicts the sample for new data it has not seen before.

For training and validation of the model, the data was divided into two sets. 80% of data (in-sample set) was employed to train the LDA model and the remaining held-out set (validation set) of 20% data was employed for model validation. Table 1 shows the results of perplexity scores of 15 topics. The model with nine topics scores the minimal validation perplexity of 237.5, indicating that the model with eight topics is more reliable than the other models. The perplexity validation does not test the semantic similarity between words in a topic. Therefore, perplexity optimization cannot yield human

**Table 1** Validation perplexities under the different numbers of topics

Number of topics	Validation perplexity	Number of topics	Validation perplexity	Number of topics	Validation perplexity
1	290.5	6	247.9	11	244.2
2	269.1	7	241.1	12	256.9
3	276.4	8	237.5	13	251.9
4	253.7	9	239.4	14	270.7
5	245.6	10	243.3	15	267.2

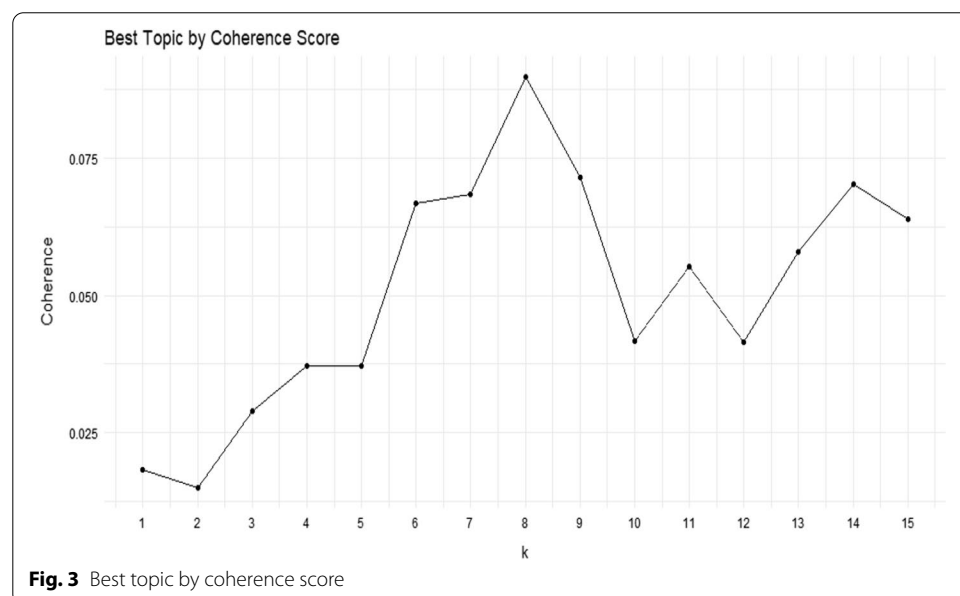


interpretable topics [67]. Thus, this test alone may not be an adequate measure for predicting the best model. This limitation of perplexity warrants a test for topic coherence.

Coherence is another method of evaluation of a topic model. It measures the degree of semantic similarity between high-scoring words in the topic. The coherence score is calculated based on pairwise scores of the top  $n$ -frequently occurring words in a topic. There are different coherence measures discussed in the literature. However, we used the CV measure, the best measure to compare different topic models based on their human interpretability [68]. The measurement helps to distinguish between topics that are an artifact of statistical inferences, and semantically interpretable. Figure 3 shows the coherence score for the different number of topics. A model with the highest coherence score before flattening out or a major drop is considered as the best model. To our surprise, both the tests confirm the presence of eight topics in our data. Therefore, we picked a model with eight topics ( $k=8$ ).

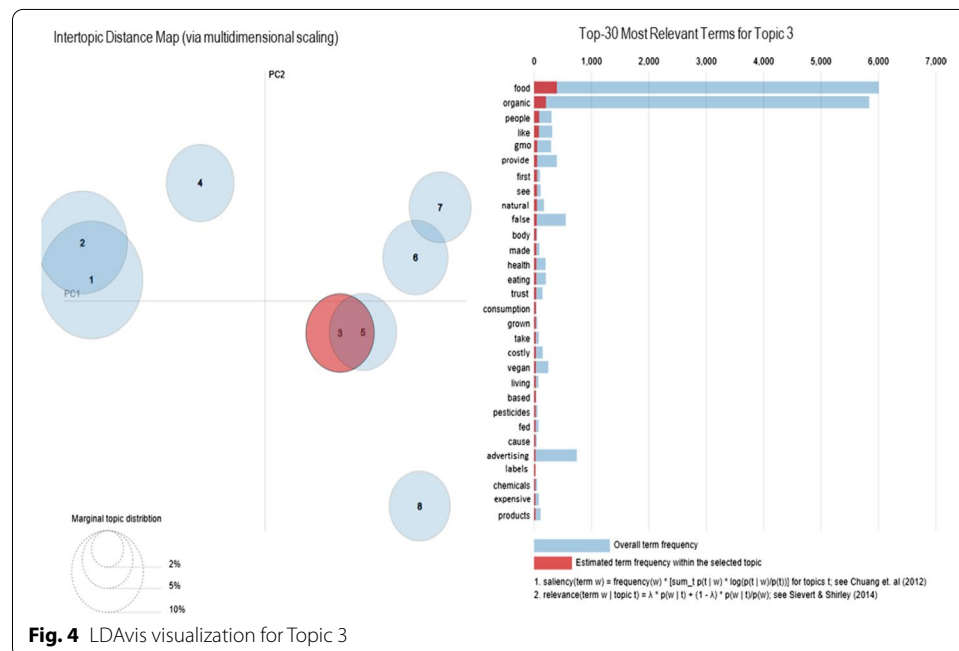
### LDA model

Based on the results of perplexity and coherence scores, we ran an LDA model of 8 topics with 12 terms. Topics are labeled based on the identification of a logical connection between most the frequent terms of the topic. Table 2 presents the topic labels with their key terms. Topic 1 and Topic 2 are about the United States (US) politics as tweets were collected between January 10 and March 7, 2021, which is very close to an incidence in the US. Jacob Chansley, a Capitol Hill attacker, demands all-organic foods inside the prison [69]. This incident has led to a huge inflow of Twitter posts related to US politics and organic foods. Topic 3 discusses organic food in connection with food additives such as pesticides and chemicals. The topic is composed of both positive perceptions (i.e., organic food has no additives) and negative or skeptical views (i.e., organic food is not free from additives). It can be understood with the tweets: *“Deceptive Labels on Food. Making sure you understand the terms that are used”, “what if they are not using*



**Table 2** Topics listing top 12 key terms extracted from LDA model

Topic id	Terms	Label
Topic 1	food, organic, jail, judge, Capitol, horns, get, hunger, riot, ordered, fur, strike	US politics (Attack on Capitol Hill)
Topic 2	organic, food, get, jail, man, white, prison, guy, judge, horns, privilege, vacation	US politics (Capitol attacker's demand for organic food)
Topic 3	organic, food, people, consumption, health, vegan, grown, eating, fed, pesticides, chemicals	Authenticity
Topic 4	food, organic, month, seasonal, three, spring, package, survival, vitality, bundle, chakra, better	Seasonality
Topic 5	organic, food, plant-based, vegan, diet, ecological, refreshing, vegetables, health, future, good	Plant-based diet
Topic 6	organic, food, plant, local, fresh, natural, farming, agriculture, save, planet, impact, world	Saving the planet
Topic 7	farming, choose, harvest, latest, technology, crops, labels, standards, agriculture, supply-chain, organic, focus	Organic farming and standardization
Topic 8	online, grocery, demand, local, store, orders, delivery, affordable, save, COVID19, planet, markets	Food delivery



*organic eggs feeding the chickens GMO FOOD*. The key terms are “people”, “health”, “grown”, “eating”, “chemicals”, etc. Topic 4 is about the seasonality of organic foods. It emphasizes on benefits of choosing seasonal organic foods. The key terms are “seasonal”, “package”, “vitality”, and “better”.

Topic 5 discusses the importance of shifting toward plant-based foods. Some of the key terms are “plant-based”, “vegan”, “diet”, and “vegetables”. Topic 6 is about saving the planet by natural and local farming which is suggested by terms such as “natural”, “planet”, “save”, “impact” and “world”. Topic 7 highlights the need for technology and standardization in organic farming and supply-chain. The key terms are “farming”, “organic”, “technology”, and “standards”, etc. The last topic discusses the surge in online

orders of organic foods and delivery systems, presumably caused by the COVID-19 pandemic.

The visual display map of topic models produced using LDAvis highlights the topic with its top 30 most relevant terms. The display map of Topic 3 is shown below as Fig. 4, and the visual display map of the remaining five topics on organic foods is appended at the end of the paper. The visualization provides the global view of topics with the terms highly associated with the individual topic.

In the visual display map on the left side, the distance between the circles indicates the topic similarity (or dissimilarity), i.e., the closer the circles, the more similar the themes. And, the size of the circle depicts its prevalence. The horizontal bar chart on the right side of the visualization depicts the terms most relevant to the selected topic.

As per the display map (Fig. 4), Topic 1 and Topic 2 are similar to each other but differs much from other topics. Similarly, Topic 3 and Topic 5 are close to each other. Topic 8 is away from other topics and refers the unrelatedness as it discusses organic food demands and online deliveries during the COVID-19 pandemic.

### Topic significance ranking

We assessed topic quality through topic significance ranking (TSR) scores based on the criteria proposed by [70]. TSR is a novel unsupervised analysis of Probabilistic Topic Models (PTM). We applied a 4-phase Weighted Combination decision strategy to evaluate the inferred topics of PTM to evaluate their semantic significance. Topic significance is obtained by combining the information from the “multi-criteria measures” to form a single index of evaluation based on a “Weighted Linear Combination” (WLC) decision strategy. WLC is a widely used technique in the area of multi-criteria decision analysis. WLC assesses each topic by the following formula:

$$A_k = \sum_{m=1}^{N_m} \Psi_m S_{m,k} \quad (5)$$

where  $N_m$  is number of different measures,  $\Psi_m$  refers the weight of the measure in the total score, and  $S_{m,k}$  is the score of  $k$  th topic for the  $m$  th measure.

A 4-phase weighted combination approach was implemented.  $S_{m,k}$  and  $\Psi_m$  scores were to be computed on different scales and criteria; therefore, measures are standardized before combination.

### Phase 1: Standardization procedure

In the first phase, “standardization procedures” are performed to transfer each distance measure from its true value into two standardized scores (relative score of the distances and weight value).

To compute both the scores  $S_{m,k}$  and weights  $\Psi_m$ , two standardized procedures are used for each of the criteria. The first, score is re-scaled based on the weight of each score against the total score overall topics and is expressed in Eq. (6). The second, score range procedure is followed based on the minimum and maximum values as scaling points for standardization and is given by Eq. (7).

$$C_{1k}^m = C_k^m \times \frac{\sum_{j=1, j \neq k}^k C_j^m}{\sum_{j=1}^k C_j^m} \quad (6)$$

$$C_{2k}^m = \frac{C_k^m - C_{\min}^m}{C_{\max}^m - C_{\min}^m} \quad (7)$$

where  $C_{\min}^m$  and  $C_{\max}^m$  are the minimum and maximum distance values measured by the distance measure  $m$  under the criterion  $C$ , and  $m \in \{KL, COR, COS\}$ ; KL-Kullback–Leibler (KL) Divergence, COR-Correlation Coefficient, and COS- Cosine Dissimilarity are three distance measures.

### Phase 2: Intra-criterion weighted linear combination

To compute the rank significance, different measures are combined into a single figure. In this phase, topic ranking performs a WLC of the standardized score of the distance measures. The WLC score of criteria  $C$  for topic  $k$  is given by Eq. (8).

$$s_k^C = \frac{C_k^{KL} + C_k^{COR} + C_k^{COS}}{3} \quad (8)$$

where  $C_k^{KL}$ ,  $C_k^{COR}$ , and  $C_k^{COS}$  are the standardized scores of the three distance measures.

### Phase 3: Inter-criterion Weighted Combination

In phase 3, a Weighted Combination (WC) is done over the scores computed in phase two. Two different WC approaches are used to combine the scores and weights. The first approach is based on Eq. (6) and applies the standardized score of the background criterion as a weight for the Uniformity ( $u$ ) and Vacuousness ( $v$ ).

$$\hat{S}_k = \hat{s}_{1k}^b \left( \Psi_u S_{1k}^u + \Psi_v S_{1k}^v \right) \quad (9)$$

where  $\Psi_u$  and  $\Psi_v$  are the weights of the Uniformity and Vacuousness criteria in the score, respectively;  $\hat{s}_{1k}^b$  is the weight of the topic background.

The second approach is performed based on the score range standardization procedure (Eq. 7). This is done by using the WLC in Eq. (5) as below:

$$\Psi_k = \Psi_u S_{2k}^u + \Psi_v S_{2k}^v + \Psi_b s_{2k}^b \quad (10)$$

### Phase 4: Topic Significance Score

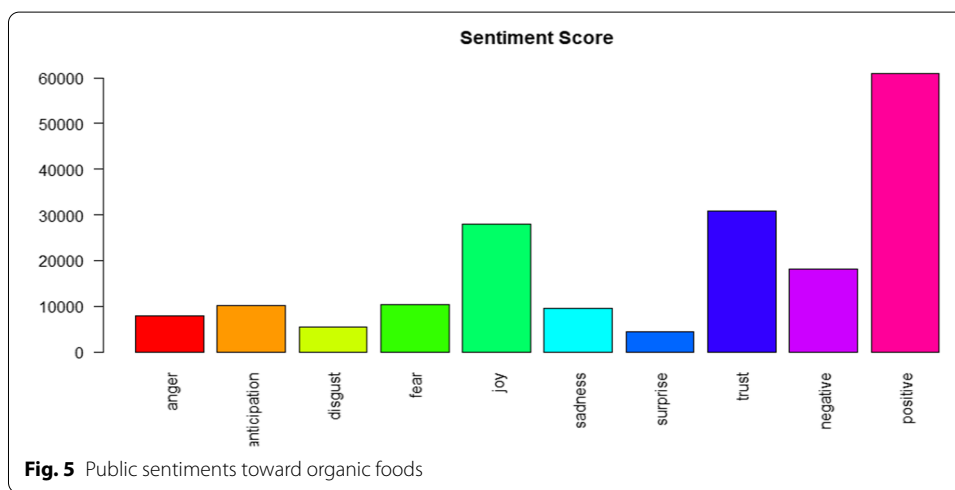
The final rank is computed using the total topic score from Eq. (9) and the normalized weight from Eq. (10). Thus, the final rank of the topic is computed by:

$$TSR_k = \Psi_k \times \hat{S}_k \quad (11)$$

A significance ranking algorithm was applied on the topics discovered by LDA with  $K$  set to 8. This technique has the potential to properly rank the topic based on its true significance. The rank of a topic depends on its distribution of background words. A topic with higher TSR describes the underlining themes better and presents the

**Table 3** Topic significance rank

Topic id	Topic label	$\hat{\zeta}$	$\hat{\psi}$	TSR	Rank
Topic 5	Plant-based diet	21.3	0.8	17.0	1
Topic 3	Authenticity	20.7	0.8	16.6	2
Topic 4	Seasonality	19.8	0.8	15.8	3
Topic 7	Organic farming and standardization	18.5	0.8	14.8	4
Topic 6	Saving the planet	16.3	0.8	13.0	5
Topic 2	US politics (Capitol attacker's demand for organic food)	15.7	0.8	12.6	6
Topic 8	Food delivery	14.8	0.7	10.4	7
Topic 1	US politics (Attack on Capitol Hill)	14.5	0.7	10.2	8

**Fig. 5** Public sentiments toward organic foods

meaningful class semantics. Table 3 presents the topic significance rank index. The topics are ordered by their rank.

According to our *LDavis* results, Topic 1 and Topic 2 are the most prevalent topics. However, these two topics receive low significance ranks. Topic 5 (Plant-based diet), Topic 3 (Authenticity), and Topic 4 (Seasonality) are the three most significant topics.

### Sentiment results

NRC lexicon enabled us to classify eight different emotions in addition to positive and negative sentiments. The negative emotions include fear, anger, disgust, and sadness, while positive emotions include joy, trust, anticipation, and surprise. The results of sentiment analysis show that public views toward organic foods are mostly positive (see Fig. 5). The results also show the presence of negative emotions but in a lower proportion as compared to positive ones. The frequent positive words for organic food are “healthy,” “safe,” “fascinating,” “eat,” “perfect,” “vitality,” “aroma,” and “delicious,” etc. Joy and trust are two prominent emotions expressing positive emotions about organic foods with a total sentiment score of more than 30,000 for each.

The frequent negative words are “unnatural,” “waste,” “doubt,” “junk,” “ridiculous,” “fake,” and “unbelievable,” etc. The keywords suggest that people are suspicious about

the authenticity of organic foods, and express negative emotions. Some of the tweets are: *“Only 4–6% of the food is organic. The rest is produced by a few major food processors and full of chemicals,” “Can I sue the food place for false advertising?” and “All food is organic. Inorganic matter can’t be digested.”* Therefore, the fact cannot be ignored that not everyone is convinced about organic claims, i.e., people are still skeptical about organic foods available in the markets being natural, healthy, and chemical or pesticide-free. However, the entire negative emotions in the results are not caused by the public distrust alone. It contains the negative emotions about the US riot and politics over organic foods (“insurrection,” “disgraceful,” “riot,” “attack,” etc.).

## Discussion

The quality of output of a machine learning model depends on the purity of the data and the training of the model. In the present study, the data passes through rigorous data cleaning and pre-processing followed by model tuning and evaluations. Topic evaluation results indicated the presence of eight topics reflecting public attitude on organic foods. Except for Topic 1 and Topic 2, which discuss organic foods linking a specific incidence in the USA, all other topics highlight various aspects related to organic foods.

The majority of previous studies on organic food consumption were based on traditional, self-reported methods. These studies focused, for example, on consumers’ motivations to buy organic foods, such as health, environment, and social consciousness [16]. The only study based on content analysis of online comments about organic foods [10] revealed 65 beliefs that the authors structured into 22 themes of a wide breadth, from ‘price’ to ‘food security’. The authors structured them into four main categories: ‘product’, ‘food system’, ‘authenticity’, and ‘production’. Our study contributes to this line of research by providing further structuring of public attitudes on organic foods based on objective techniques of big data and text mining. Moreover, based on the topic significance rank test, we have identified top the three most significant topics, i.e., plant-based diet, authenticity, and seasonality. The other significant themes related to organic food are organic farming and standardization, saving the planet, and food delivery. Our findings suggest also that people are quite enthusiastic about shifting to plant-based food. They believe that a plant-based diet can contribute to reducing greenhouse gas emissions. Dedicating natural resources such as land and water for growing plant-based foods can save the planet (Topic 6). Authenticity is another interesting theme. Under this theme, tweets indicated that a large group of people advocate organic food for being natural, healthy, non-GMO, and chemical and pesticide-free, while another group is skeptical about these claimed features of organic foods. On the ‘seasonality’, people tweeted about the benefits of eating seasonal foods. Organic food eaters find seasonal foods fresh, tasty, and healthy [71].

People in their tweets also highlighted the need for organic farming using the latest technology and standardization of food products and processes (Topic 7). The standards also aim to avoid “false organic claims.” Unfortunately, there are differences in standards across the countries. Further, Topic 8 discusses online delivery services and the availability of organic foods in local stores and markets. Online food retailers have witnessed “unprecedented demand” for organic products during COVID19 lockdowns [72, 73]. Whenever there is a health scare (like SARS/COVID19), people look for ways to disease

prevention and boost their health. According to a study by Tastewise, an AI-powered intelligence startup, consumers' mention of 'immunity' in online food searches rose 27% between February 2019 and March 2020 [73]. As people, especially in the outbreaks, search for healthy and safe foods for their at-home families, organic foods are proving to be the best food choice [74].

Surprisingly, our results do not report any theme on price. It is contrary to the findings of most of the survey research [34, 75–78], which reports organic food is more expensive than conventional food, which is a major consumer concern. However, results of [10], based on content analysis of online comments about organic food posted on news websites and forums, report two positive price beliefs- namely, 'organic products are worth the price premium' and 'organic products are not (much) more expensive than conventional products.' Thus, when it comes to health and environmental benefits from the products, consumers pay less attention to the price [79]. This might be a likely reason for the absence of a prominent theme on the price in our results. Whereas, in line with our findings, 'authenticity' was found to be a most discussed theme in the study of [10].

Our sentiment analysis results show that public sentiment is predominantly positive toward organic foods. However, a small proportion of people has negative sentiments too. This result corroborates the findings of [80]. In their study, [80] also found a high relevance for positive sentiment valence than the negative valence analyzed from comments made on organic food videos on YouTube. Positive sentiments can be attributed to hedonic (taste, freshness, appearance, aroma, etc.) and utilitarian (nutritional contents, healthfulness, ecological welfare, healthfulness, and purity, etc.) values. The research from a consumer cognition perspective also suggests the emotions about organic foods are the outcome of belief held for its healthiness and safety, ecological consequences, as well as a role in supporting local economies. As per our findings, negative emotions about organic foods are mainly due to public skepticism about organic claims.

## Conclusion

Since the consumer behavior towards organic food is changing rapidly due to awareness of environmental degradation, health hazards, and other related issues, this research started with an aim to answer the same existing question- how does the public view organic foods? Numerous studies have been conducted in the past. The majority of them are based on surveys and interviews, which are subject to various limitations such as social desirability bias, a response bias, and common method bias. In this study, we collected public posts from Twitter and utilized text mining techniques to extract the most discussed topics and sentiments towards organic foods. People willingly publish their opinions online on Twitter, which is free of an external bias. The findings obtained by utilizing the big data from social media platforms for text mining, provide relatively objective and verifiable results.

The present study makes academic as well as practical contributions. From an academic perspective, this is the first of early researches to study the public's attitude and sentiments towards organic foods using big data and text mining approaches. Also, this study adds to the existing literature by providing a broader and richer understanding of public perspectives on organic food consumption. The study explored public attitude on organic foods and identified underlying different emotions as opposed to just positive or negative sentiments.

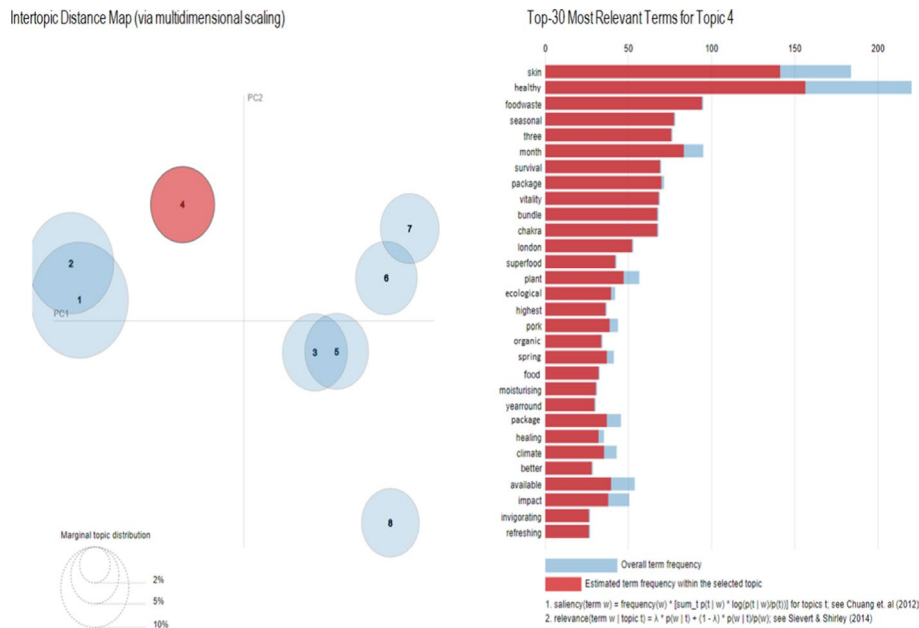


The results indicated the presence of negativity among the public mainly due to distrust in the authenticity of organic claims. Therefore, eliminating the distrust among the consumers may boost the growth of the organic food market. Thus, from a practical perspective, our findings can help policy-makers and certification bodies enhancing the communications about organic foods, thereby developing trust and positivity.

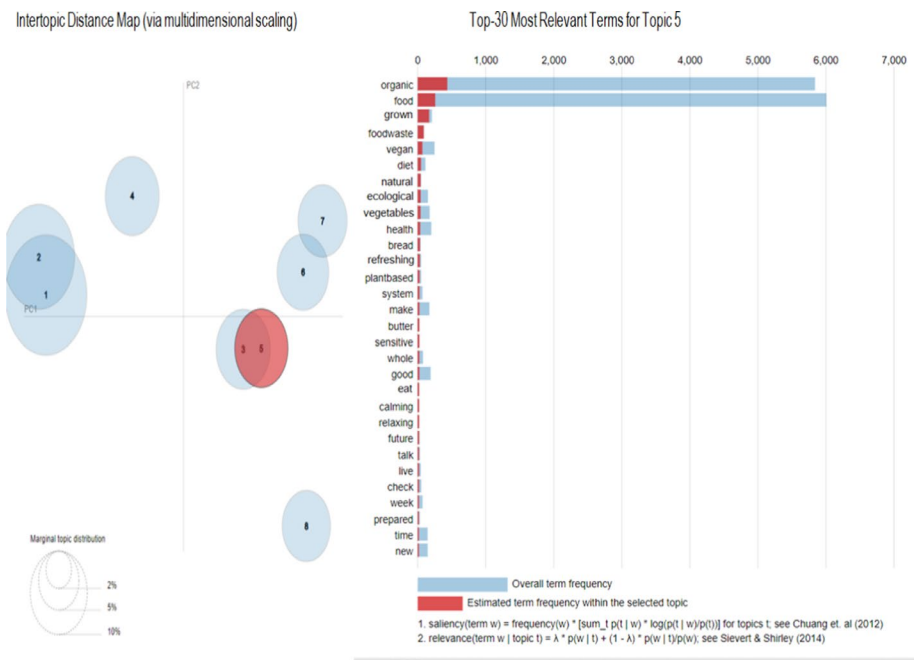
This research has some limitations. First, the public posts were extracted from Twitter during a specific time period. Second, users of Twitter may not cover all age groups as the elderly people may not be familiar with the internet and social networking sites. Third, only posts in English were analyzed. Therefore, for large and diverse sample data, future research should collect the data from different social networking sites for a larger time span or in different time intervals. Future research, in this direction, is encouraged to perform a topic modeling and sentiment analysis of texts in other languages to validate our findings. Last, a lexicon-based approach was used in the present study, which may fail to identify some human expressions such as sarcasm and irony.

Appendix A. Topic visualizations

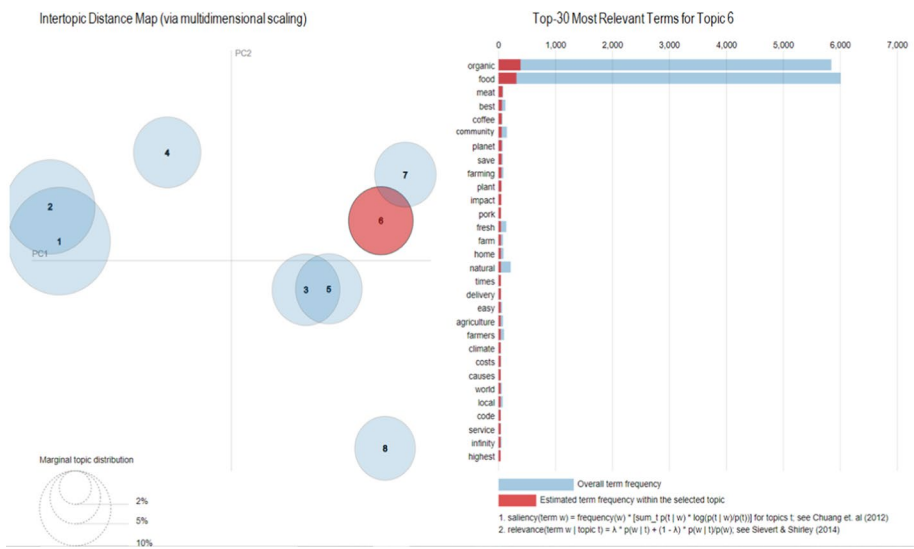
A1: LDAvis visualization for Topic 4.



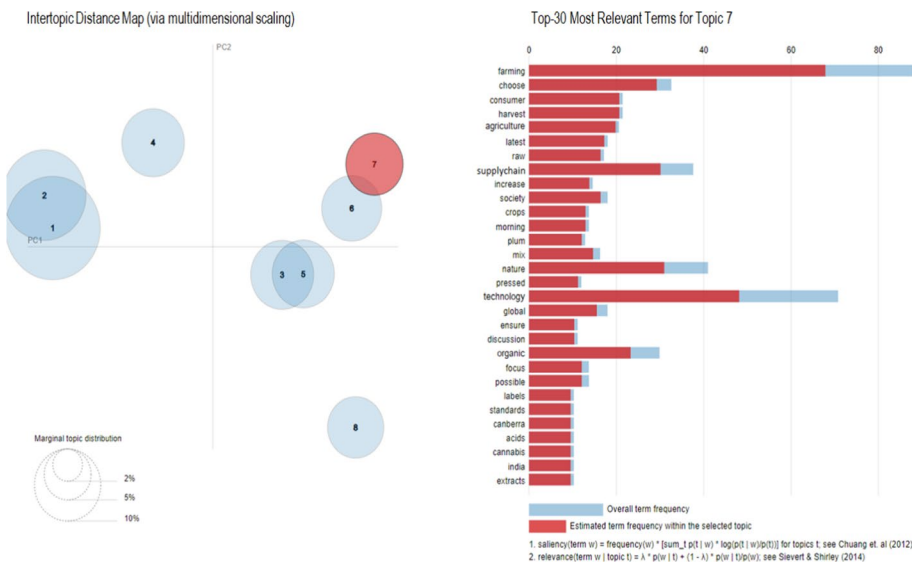
A2: LDAvis visualization for Topic 5.



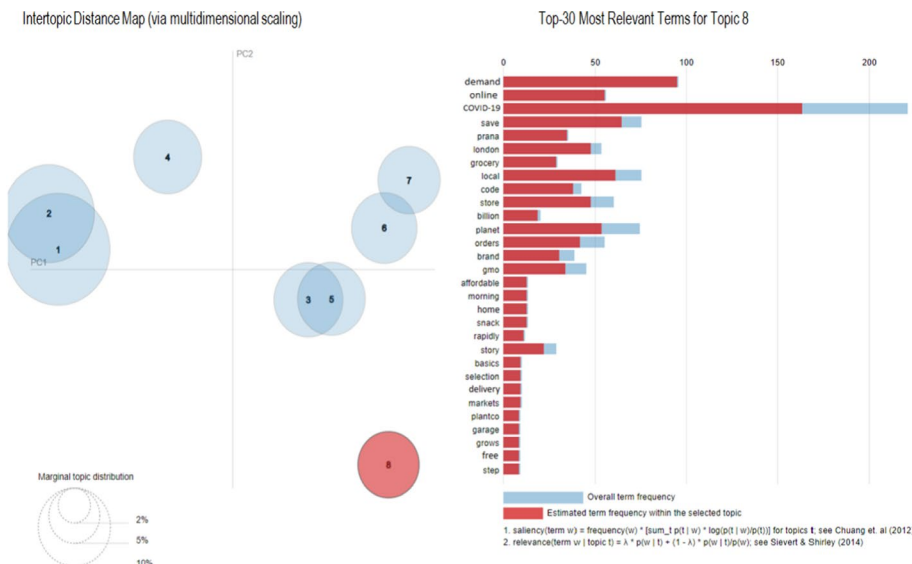
A3: LDAvis visualization for Topic 6.



A4: LDAvis visualization for Topic 7.



A5: LDAvis visualization for Topic 8.



**Abbreviations**

LDA: Latent Dirichlet Allocation; SNS: Social Networking Site; API: Application Programming Interface; NLP: Natural language processing; TSR: Topic significance ranking.

**Acknowledgements**

None.

**Authors' contributions**

Conceptualization: AS, and AG; data curation: AS; formal analysis: AS; methodology: AS; resources: AG; software: AS; validation: AS, and AG; writing—original draft: AS, and AG; writing—review and editing: AS, and AG. Both authors read and approved the final manuscript.

**Funding**

None.

**Availability of data and materials**

The dataset used in this study can be obtained from the corresponding author on reasonable request.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 12 July 2021 Accepted: 19 December 2021

Published online: 06 January 2022

**References**

- Meiselman HL. A review of the current state of emotion research in product development. *Food Res Int*. 2015;76:192–9.
- Ismael D, Ploeger A. Consumers' emotion attitudes towards organic and conventional food: a comparison study of emotional profiling and self-reported method. *Foods*. 2020;9:1.
- Demetriou C, Ozer BU, Essau CA. Self-Report Questionnaires. *Encycl. Clin Psychol*. 2015;1:6.
- Podsakoff PM, MacKenzie SB, Podsakoff NP. Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol*. 2012;63:539–69.
- Oatley K, Keltner D, Jenkins J. Understanding emotions. Malden: Wiley-Blackwell; 2006.
- Weiss HM. No Title. In: Emotions in the workplace: Understanding the structure and role of emotions in organizational behavior. 2002. p. 20–63.
- Wu Z, Zhang Y, Chen Q, Wang H. Attitude of Chinese public towards municipal solid waste sorting policy: a text mining study. *Sci Total Environ*. 2021;756:889.
- Feizollah A, Mostafa MM, Sulaiman A, Zakaria Z, Firdaus A. Exploring halal tourism tweets on social media. *J Big Data*. 2021. <https://doi.org/10.1186/s40537-021-00463-5>.
- Hemmerling S, Hamm U, Spiller A. Consumption behaviour regarding organic food from a marketing perspective—a literature review. *Org Agric*. 2015;5:277–313.
- Danner H, Menapace L. Using online comments to explore consumer beliefs regarding organic food in German-speaking countries and the United States. *Food Qual Prefer*. 2020;83:9.
- Robertson JL, Barling J. Greening organizations through leaders' influence on employees' pro-environmental behaviors. *Wiley Online Libr*. 2012;34(2):176–94.
- Stern PC. Toward a Coherent Theory of Environmentally Significant Behavior. Vol. 56, *J Soc Issues*. 2000. <https://okapi.texas.pdf>.
- Poore J, Nemecek T. Reducing food's environmental impacts through producers and consumers. *Science*. 2018;360(6392):987–92.
- Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. *Proc Natl Acad Sci U S A*. 2018;115(25):6506–11.
- European Union. The EU's organic food market: facts and rules. 2020. <https://www.europarl.europa.eu/news/en/headlines/society/20180404STO00909/the-eu-s-organic-food-market-facts-and-rules-infographic>.
- Hansen T, Sørensen MI, Eriksen MLR. How the interplay between consumer motivations and values influences organic food identity and behavior. *Food Policy*. 2018;74:39–52.
- Molinillo S, Vidal-Branco M, Japutra A. Understanding the drivers of organic foods purchasing of millennials: Evidence from Brazil and Spain. *J Retail Consum Serv*. 2020;52:9.
- Pham TH, Nguyen TN, Phan TTH, Nguyen NT. Evaluating the purchase behaviour of organic food by young consumers in an emerging market economy. *J Strateg Mark*. 2019;27(6):540–56. <https://doi.org/10.1080/0965254X.2018.1447984>.
- Iyer P, Davari A, Paswan A. Green products: Altruism, economics, price fairness and purchase intention. *Soc Bus*. 2016;6(1):39–64.
- Hasselbach JL, Roosen J. Motivations behind Preferences for Local or Organic Food. *J Int Consum Mark*. 2015;27(4):295–306.
- Kareklas I, Carlson JR, Muehling DD. "I eat organic for my benefit and yours": Egoistic and altruistic considerations for purchasing organic food and their implications for advertising strategists. *J Advert*. 2014;43(1):18–32.
- Chekima B, Oswald AI, Wafa SA, Chekima K. Narrowing the gap: Factors driving organic food consumption. *J Clean Prod*. 2017;166:1438–47. <https://doi.org/10.1016/j.jclepro.2017.08.086>.
- Sultan P, Tarafder T, Pearson D, Henryks J. Intention-behaviour gap and perceived behavioural control-behaviour gap in theory of planned behaviour: moderating roles of communication, satisfaction and trust in organic food consumption. *Food Qual Prefer*. 2020;81:103838. <https://doi.org/10.1016/j.foodqual.2019.103838>.
- Tandon A, Jabben F, Talwar S, Sakashita M, Dhir A. Facilitators and inhibitors of organic food buying behavior. *Food Qual Prefer*. 2021;88: 104077. <https://doi.org/10.1016/j.foodqual.2020.104077>.

25. Dowd K, Burke KJ. The influence of ethical values and food choice motivations on intentions to purchase sustainably sourced foods. *Appetite*. 2013;69:137–44.
26. D'Amico M, Vita G Di, Production LM-J of C, 2016 undefined. Exploring environmental consciousness and consumer preferences for organic wines without sulfites. Elsevier, New York
27. Van Loo EJ, Hoefkens C, Verbeke W. Healthy, sustainable and plant-based eating: Perceived (mis)match and involvement-based consumer segments as targets for future policy. *Food Policy*. 2017;69:46–57.
28. Padel S, Röcklinsberg H, Schmid O. The implementation of organic principles and values in the European Regulation for organic food. *Food Policy*. 2009;34(3):245–51.
29. De Marchi E, Caputo V, Nayga RM, Banterle A. Time preferences and food choices: Evidence from a choice experiment. *Food Policy*. 2016;62:99–109.
30. Willer H, Schlatter B, Trávníček J, Kemper L, Julia L (Eds). The World Of Organic Agriculture. Statistics & emerging trends 2020. Research Institute of Organic Agriculture (FiBL) & IFOAM - Organic International. 2020. p. 1–337. <https://orgprints.org/id/eprint/37222/9/willer-et-al-2020-full-document-2020-02-28-4th-corrigenda.pdf>.
31. Nguyen DT, Truong DC. The Impact of Psychological and Environmental Factors on Consumers' Purchase Intention toward Organic Food: Evidence from Vietnam First Author and Corresponding. *J Asian Financ*. 2021;8(1):915–25.
32. Denver S, Jensen JD. Consumer preferences for organically and locally produced apples. *Food Qual Prefer*. 2014;31(1):129–34.
33. Rousseau S, Vranken L. Green market expansion by reducing information asymmetries: Evidence for labeled organic food products. *Food Policy*. 2013;40:31–43.
34. Singh A, Verma P. Factors influencing Indian consumers' actual buying behaviour towards organic food products. *J Clean Prod*. 2017;167:473–83. <https://doi.org/10.1016/j.jclepro.2017.08.106>.
35. Aertsens J, Verbeke W, Mondelaers K, van Huylenbroeck G. Personal determinants of organic food consumption: A review. *Br Food J*. 2009;111(10):1140–67.
36. Talwar S, Jabeen F, Tandon A, Sakashita M, Dhir A. What drives willingness to purchase and stated buying behavior toward organic food? A Stimulus–Organism–Behavior–Consequence (SOBC) perspective. *J Clean Prod*. 2021;293:125882. <https://doi.org/10.1016/j.jclepro.2021.125882>.
37. Janssen M. Determinants of organic food purchases: Evidence from household panel data. *Food Qual Prefer*. 2018;68:19–28.
38. Atkinson L. Buying In to Social Change: How Private Consumption Choices Engender Concern for the Collective. *Ann Am Acad Pol Soc Sci*. 2012;644(1):191–206.
39. Ghali-Zinoubi Z. On linking socioeconomic status to consumer willingness to buy and pay for organic food. *J Food Sci Technol*. 2021;58(3):1042–50.
40. Mohammed AA. What motivates consumers to purchase organic food in an emerging market? An empirical study from Saudi Arabia. *Br Food J*. 2020;123(5):1758–75.
41. Fuljahn A, Moosmayer DC. The myth of guilt: a replication study on the suitability of hedonic and utilitarian products for cause related marketing campaigns in Germany. *Int J Bus Res*. 2011;11(1):85–92.
42. Zeithaml VA. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *J Mark*. 1988;52(3):2–22.
43. Ditlevsen K, Sandøe P, Lassen J. Healthy food is nutritious, but organic food is healthy because it is pure: The negotiation of healthy food choices by Danish consumers of organic food. *Food Qual Prefer*. 2019;71:46–53.
44. Maehle N, Iversen N, Hem L, Otnes C. Exploring consumer preferences for hedonic and utilitarian food attributes. *Br Food J*. 2015;117(12):3039–63.
45. White K, Habib R, Hardisty DJ. How to SHIFT Consumer Behaviors to be More Sustainable: A Literature Review and Guiding Framework. *journals.sagepub.com*. 2019;83(3):22–49. <https://doi.org/10.1177/0022242919825649>
46. Kaiser FG, Shimoda TA. Responsibility as a predictor of ecological behaviour. *J Environ Psychol*. 1999;19(3):243–53.
47. Schwartz D, Loewenstein G. The Chill of the Moment: Emotions and Proenvironmental Behavior. *J Public Policy Mark*. 2017;36(2):255–68. <https://doi.org/10.1509/jppm.16.132>.
48. Corral-Verdugo V, Bonnes M, Tapia-Fonlem C, Fraijo-Sing B, Frias-Armenta M, Carrus G. Correlates of pro-sustainability orientation: The affinity towards diversity. *J Environ Psychol*. 2009;29(1):34–43.
49. Peter PC, Honea H. Targeting social messages with emotions of change: The call for optimism. *J Public Policy Mark*. 2012;31(2):269–83. <https://doi.org/10.1509/jppm.11.098>.
50. Steg L. Car use: Lust and must Instrumental, symbolic and affective motives for car use. *Transp Res Part A Policy Pract*. 2005;39(2):147–62.
51. Desmet PMA, Schifferstein HNJ. Sources of positive and negative emotions in food experience. *Appetite*. 2008;50(2–3):290–301.
52. King SC, Meiselman HL, Carr BT. Measuring emotions associated with foods in consumer testing. *Food Qual Prefer*. 2010;21(8):1114–6. <https://doi.org/10.1016/j.foodqual.2010.08.004>.
53. Wang Y, Li H, Wu Z. Attitude of the Chinese public toward off-site construction: A text mining study. *J Clean Prod*. 2019;238: 117926. <https://doi.org/10.1016/j.jclepro.2019.117926>.
54. Philander K, Zhong YY. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *Int J Hosp Manag*. 2016;55:16–24. <https://doi.org/10.1016/j.ijhm.2016.02.001>.
55. Prameswari P, Zulkarnain, Surjandari I, Laoh E. Mining online reviews in Indonesia's priority tourist destinations using sentiment analysis and text summarization approach. *Proc 2017 IEEE 8th Int Conf Aware Sci Technol iCAST 2017*;8:121–6.
56. Oliveira N, Cortez P, Areal N. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Syst Appl*. 2017;73:125–44. <https://doi.org/10.1016/j.eswa.2016.12.036>.
57. Antonakaki D, Fragopoulou P, Ioannidis S. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Syst Appl*. 2021;164:8.
58. Singh S, Chauhan A, Dhir S. Analyzing the startup ecosystem of India: a Twitter analytics perspective. *J Adv Manag Res*. 2020;17(2):262–81.

59. Lakhiwal A, Kar AK. Insights from twitter analytics: Modeling social media personality dimensions and impact of breakthrough events. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). New York: Springer; 2016. p. 533–44.
60. Kang GJ, Ewing-Nelson SR, Mackey L, Schlitt JT, Marathe A, Abbas KM, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine*. 2017;35(29):3621–38.
61. Salvatore C, Biffignandi S, Bianchi A. Social Media and Twitter Data Quality for New Social Indicators. *Soc Indic Res*. 2021;156(2–3):601–30. <https://doi.org/10.1007/s11205-020-02296-w>.
62. Bastani K, Namavari H, Shaffer J. Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Syst Appl*. 2019;127:256–71.
63. Blei DM, Ng AY, Edu JB. Latent Dirichlet Allocation Michael I. Jordan. Vol. 3, *Journal of Machine Learning Research*. 2003. [https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?TB\\_iframe=true&width=370.8&height=658.8](https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?TB_iframe=true&width=370.8&height=658.8)
64. Chiru C, Rebedea T, Ciotec S. Comparison between LSA-LDA-Lexical Chains. *scitepress.org*. 2014;(2):255–62. <http://wordnet.princeton.edu/>
65. Pang B, Lee L. Opinion mining and sentiment analysis. *Found Trends Inf Retr*. 2008;2(1–2):1–135. <https://doi.org/10.1561/1500000011>.
66. Nguyen T, Ngoc T, Nguyen T, Thu H, Nguyen VA. Mining aspects of customer's review on the social network. *J Big Data*. 2019;70:654.
67. Chang J, Boyd-Graber J, Gerrish S, Wang C, Blei DM. Reading Tea Leaves: How Humans Interpret Topic Models. *papers.nips.cc*. 2009 [cited 2021 Jul 5]. <http://rexa.info>
68. Ray SK, Ahmad A, Kumar CA. Review and Implementation of Topic Modeling in Hindi. *Appl Artif Intell*. 2019;33(11):979–1007. <https://doi.org/10.1080/08839514.2019.1661576>.
69. Milman O. I'm facing a prison sentence': US Capitol rioters plead with Trump for pardons. *The Guardian*. 2021. <https://www.theguardian.com/us-news/2021/jan/16/us-capitol-rioters-donald-trump-pardons>
70. AlSumait L, Barabara D, Gentle J, Domeniconi C. Topic significance ranking of LDA generative models. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2009. p. 67–82.
71. Tobler C, Visschers VHM, Siegrist M. Eating green Consumers' willingness to adopt ecological food consumption behaviors. *Appetite*. 2011;57(3):674–82.
72. Aday S, Aday MS. Impact of COVID-19 on the food supply chain. *Food Qual Saf*. 2020;4(4):167–80.
73. Askew K. Organic food gets coronavirus boost. 2020 [cited 2021 Jul 10]. <https://www.foodnavigator.com/Article/2020/05/06/Organic-food-gets-coronavirus-boost>
74. Rodríguez-Pérez C, Molina-Montes E, Verardo V, Artacho R, García-Villanova B, Guerra-Hernández EJ, et al. Changes in dietary behaviours during the COVID-19 outbreak confinement in the Spanish COVIDiet study. *Nutrients*. 2020;12(6):1–19.
75. Thøgersen J, de Barcellos MD, Perin MG, Zhou Y. Consumer buying motives and attitudes towards organic food in two emerging markets: China and Brazil. *Int Mark Rev*. 2015;32(3–4):389–413.
76. Lee H-J. Individual and Situational Determinants of U.S. Consumers' Buying Behavior of Organic Foods.pdf. 2016. p. 117–31.
77. Lee HJ, Hwang J. The driving role of consumers' perceived credence attributes in organic food purchase decisions: A comparison of two groups of consumers. *Food Qual Prefer*. 2016;54:141–51.
78. Lee HJ, Yun ZS. Consumers' perceptions of organic food attributes and cognitive and affective attitudes as determinants of their purchase intentions toward organic food. *Food Qual Prefer*. 2015;39:259–67.
79. Royne MB, Levy M, Martinez J. The public health implications of consumers' environmental concern and their willingness to pay for an eco-friendly product. *J Consum Aff*. 2011;45(2):329–43.
80. Meza XV, Yamanaka T. Food communication and its related sentiment in local and organic food videos on youtube. *J Med Internet Res*. 2020;22:8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---