Food Quality and Preference 104 (2023) 104714

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Exploring the nexus between food and veg*n lifestyle via text mining-based online community analytics

Ra Yoo [a], Seo-Young Kim [b], Do-Hee Kim [c], Jiyoung Kim [d], Ye Ji Jeon [d], Jung Han Yoon Park [e], Ki Won Lee [a,b,e,f,*], Hee Yang [e,1,*]

[a] Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Republic of Korea
[b] Advanced Institutes of Convergence Technology, Suwon 16229, Republic of Korea
[c] Kim Jaechul Graduate School of Artificial Intelligence, KAIST, Daejeon 34141, Republic of Korea
[d] Division of Computing, Data Science, and Society, University of California, Berkeley, CA 94720-1786, USA
[e] Bio-MAX Institute, Seoul National University, Seoul 08826, Republic of Korea
[f] Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea

ARTICLE INFO

Keywords:
Vegan
Vegetarian
Online communities
Consumer
Perception
Text mining

ABSTRACT

Attention towards veg*nism is increasing as the impact of food choices on health and sustainability as well as ethical concerns regarding animal welfare emerge. Although online user analysis is an effective tool to obtain practical insights without geographical constraints, implementation on a large veg*n population has been carried out within a limited scope. This study investigated two veg*n subreddits, r/Vegan and r/Vegetarian, using multiple text mining techniques to classify users' interests and preferences. Based on K-means and term frequency-inverse document frequency, six clusters were identified: Food, Perception, Health, Altruism, Emotion, and Situation. The proportion of each cluster and keywords representing the clusters were obtained. Being a major sector, further assessment of the Food cluster was conducted using Latent Dirichlet Allocation topic modeling technique. Confusion was observed in relation to being pressured with sudden changes in dietary patterns, including meal composition, preparation, and shopping routines. The results also revealed barriers to transition for individuals who have recently started veg*n diets, and those wishing to switch to stricter dietary patterns. In addition to difficulties relating to economic and social aspects, our findings suggest that the establishment of detailed guidelines may help accommodate the various dietary compositions across the veg*n spectrum, and clear information relating to veg*n food products is needed from manufacturers.

1. Introduction

1.1. Global veg*n trends

As the impact of food choices on environmental sustainability gains increasing attention, interest in veg*nism is rapidly increasing in parallel (Beverland, 2014; Contini et al., 2020). Environmental issues arising from the meat production process (Graça et al., 2019; Rosi et al., 2017), as well as awareness of reducing animal consumption and suffering are resonating for many consumers (Grossmann & McClements, 2021; Scherer et al., 2018). In addition, although there are ongoing debates regarding nutritional adequacy, veg*n diets are being recognized as attractive options that can offer health benefits (Godfray et al., 2018; Sadler, 2004; Scherer et al., 2019). The variety of veg*n friendly foods available are also gradually diversifying into beverages, meat and dairy alternatives, and synthetic eggs (Grossmann & McClements, 2021; He et al., 2020; McClements et al., 2021).

Veg*nism is a collective term that encompasses veganism and the varying types of vegetarianism (MacInnis & Hodson, 2021; Rosenfeld & Burrow, 2018). There are several types of veg*n diets that incorporate different extents of animal-derived food consumption, from vegan (does not allow any animal foods), to lacto-vegetarian (allows dairy), ovo-vegetarian (allows eggs), lacto-ovo-vegetarian (allows dairy and eggs), pesco-vegetarian (allows seafood), pollo-vegetarian (allows poultry),

* Corresponding authors at: Department of Agricultural Biotechnology, Seoul National University, Seoul, 08826, Republic of Korea (K.W. Lee), Bio-MAX Institute, Seoul National University, Seoul 08826, Republic of Korea (H. Yang).
E-mail addresses: kiwon@snu.ac.kr (K.W. Lee), yhee6106@kookmin.ac.kr (H. Yang).
[1] Present address: Department of Food and Nutrition, Kookmin University, Seoul 02707, Republic of Korea.

https://doi.org/10.1016/j.foodqual.2022.104714
Received 31 March 2022; Received in revised form 4 August 2022; Accepted 30 August 2022
Available online 6 September 2022

and flexitarian (allows reduced consumption of red meat) (Beardsworth & Keil, 1991; Forestell et al., 2012; Meister, 1997; Raphaely & Marinova, 2014). Researchers have presented various motives that lead to a dietary shift in individuals who practice veg*nism by examining dietary choices and behavioral patterns, such as environmental preservation, animal welfare, human rights, and personal health (Cherry, 2015; Plante et al., 2019). The notion of food in veg*nism is highlighted as a way to represent commitment and identity (Campbell, 2017).

1.2. Application of text mining techniques in online veg*n communities

The important roles of communities for veg*ns have included offering social support and extending positive influence on the maintenance of lifestyles and dietary commitments (Barr & Chapman, 2002; Bresnahan et al., 2016; Rosenfeld et al., 2020). Different veg*n societies and communities have been examined to further understand the characteristics of these groups (Michalak et al., 2012; Tordjman et al., 2016). Studies of veg*n population are often conducted through formats such as surveys or focus group discussions (Potts & White, 2008; Thane & Bates, 2000; van Loo et al., 2020). Although these traditional methods offer the advantages of being able to obtain information on the personal backgrounds of participants in accordance with the purpose of a study, there exist limitations in observing communication and trends that occur simultaneously within groups of a large population. Community analyses have also been carried out in relatively small proportions of the population with limited group sizes and questionnaires.

Text mining technologies have the advantage of being able to analyze vast amounts of data from modern channels and formats, including online communications (Hong et al., 2011). In addition, online community analysis can be useful in preparing adequate procedures by examining the impact and reactions over specific timelines, such as pandemics and elections (Song et al., 2014), as well as offering the advantage of being able to obtain information without the limitations of geographical boundaries (Guan et al., 2018; Mesch & Talmud, 2006; O'sullivan, 2000) and securing expressions of honest feelings that may not be expressed in real life (Felix, 2004; Shi et al., 2009). In accordance with these benefits, it is an effective marketing tool for generating product development insights (Christensen et al., 2017), understanding consumer trends (Bodendorf & Kaiser, 2010; Shen et al., 2019), and collecting opinions on brands and products (Mostafa, 2013). The food industry also utilizes online data for consumer analysis, including dietary pattern characterization and food knowledge discovery (Tao et al., 2020). Technologies such as K-means clustering, term frequency-inverse document frequency (TF-IDF), and Latent Dirichlet Allocation (LDA) topic modeling are useful tools to identify market insights that characterize consumers' perceptions and needs (Lin et al., 2017; Tleis et al., 2017). However, despite the strengths that such analysis may offer as a marketing strategy, there have been few studies analyzing the online veg*n community.

We sought to investigate topics being discussed by online veg*n community users to better understand existing demands and difficulties in practicing veg*nism, as well as the various components of the veg*n lifestyle. For this purpose, we identified content by K-means clustering, TF-IDF, and LDA topic modeling and analyzed two veg*n subreddits on Reddit.com, r/Vegetarian and r/Vegan.

2. Methods

2.1. Data collection

Reddit.com is an online community platform consisting of 2.8 million micro-communities and 430 million users around the world (Lin, 2021). It is a widely used source in multispectral research, as it enables researchers to collect linguistic data such as posts and comments written in "subreddits" focusing on various subjects (Blackburn et al., 2018; Jamnik & Lane, 2017). We extracted the posts from r/Vegan and r/Vegetarian uploaded between January 1, 2016, and December 30, 2020. Pushshift API Wrapper was used to crawl the data from the two subreddits at once. After preprocessing the dataset by eliminating the posts with no subtitles or bodies and deleted posts, the total number of posts used in our analysis was 109,274. The annual number of posts from the dataset is exhibited in Table 1.

2.2. Data preprocessing

We used the spaCy library to convert the posts into lowercase English and to remove special characters (e.g. @,!), and URLs. We then conducted tokenization, which converts the extracted texts into token-level of nouns, verbs, and adjectives (Ding et al., 2019). Following this, tokens were lemmatized, which refers to a preprocessing method that transforms a token into a lemma. A lemma is a root form that represents a lexeme, which is a set of all the forms with the same meanings (Jivani, 2011). Lemmatization is an essential step in Natural Language Processing (NLP) (Zhao et al., 2021), because it enhances information retrieval and reduces computation time by removing suffixes (Hickman et al., 2022). Stop words (e.g. the, is) provided by the Python library NLTK were eliminated, and the lemmatized tokens with a length of 3 characters or longer were extracted for data analyses.

2.3. Data analyses

2.3.1. Word embedding

Word embedding is a fundamental procedure that manages data within NLP (Li & Yang, 2018), and allows for the contextual similarity of words to be maintained (Naili et al., 2017). Since the crawled data was in natural language, Word2Vec was imported from the Gensim library for word embedding purposes. As we aimed to analyze a specific domain of vegans and vegetarians, we conducted word embedding based on the Word2Vec function from scratch with a vector dimensionality of 100 and a window size of 5 (Liu et al., 2018). Skip-gram, a word embedding architecture that predicts context words based on a center word (Al-Saqqa & Awajan, 2019) was selected as an underlying model as it offers advantages in presenting rare words or phrases (Giatsoglou et al., 2017). Through this process, a dataset containing vector conversion results from the textual data was obtained as a set of tokens of 8,853.

2.3.2. K-means

Clustering was performed to group contextually similar words and get a better understanding of the topic ratio in the overall communication with the data extracted from r/Vegan and r/Vegetarian. K-means clustering algorithm from the scikit-learn library was used to partition the embedded set of tokens into K number of clusters and assign the centroid coordinate (Kapil & Chawla, 2016). Since the cosine distance metric was implemented with K-means, the distances between each centroid and words were obtained. With a vector dimensionality of 100, we used k-d tree, a binary tree structure that represents a division of a k-dimensional space (Bentley, 1990). As it offers the nearest neighbor search by finding the closest data point to the target point, we used k-d tree to acquire the keywords of each cluster in the order of shortest distance from the centroids by setting the cluster centroid as a query (Li et al., 2020). The optimal number of K-means clusters was determined

Table 1
Number of posts per year used for analysis.

Year	Number of posts
2016	13,985
2017	19,516
2018	22,486
2019	26,666
2020	26,621
Total	109,274

based on the implementation of both elbow method and qualitative assessment. We confirmed the potential range of cluster numbers with elbow method and experimented with various cluster solutions. For the overall analysis of r/Vegan and r/Vegetarian, six clusters were taken by consensus based on the probable keywords (Boussalis & Coan, 2016; Park et al., 2018).

2.3.3. Document clustering via TF-IDF

TF-IDF is a widely used weighting scheme that reflects the importance of a word in a document collection or corpus (Bafna et al., 2016). TF-IDF is composed of term frequency (TF), measuring the number of occurrences of a word in a document collection, and inversed document frequency (IDF), measuring the informativeness of a word based on how rare the word appears in the document (Machuca et al., 2021). We implemented TF-IDF to obtain the document clusters based on the K-means word clusters acquired in *2.3.2* (Wu et al., 2008; Zhang et al., 2018). We first computed the TF-IDF scores to determine the importance (weight) of the tokens within the corresponding document sets based on the word frequency. Then the computed TF-IDF scores were combined with the word clusters through an inner-product between the matrix of TF-IDF scores (the number of words * the number of TF-IDF scores), and the results of the K-means word cluster matrix (the number of words * the number of word cluster). In Fig. 1, Step 4 is highlighted with each matrix involved in document clustering. The embedded word matrix (a) was combined with the cluster centroid matrix (b), which produced the word cluster matrix (c). We also computed TF-IDF scores of the embedded words throughout the initial document dataset, and the TF-IDF document matrix was generated (d). As we combined word cluster matrix (c) with the TF-IDF document matrix (d), the TF-IDF-based document cluster matrix (e) was produced based on the scores associated between the documents and six clusters. The relevant posts were extracted in the order of the size of the gap between the highest score and the second highest score. In other words, the bigger the gap between the two highest scores, the clearer the association between the documents and a specific cluster. In this regard, a document was clustered into the cluster where the highest score took a place.

Through this document matching process, we obtained the document clusters derived from K-means word clustering that considers both word semantics and word frequency. As Park et al. repeated the clustering process to validate their clustering results (Park et al., 2018), we validated the cluster solution (K = 6) by repetitively conducting the following procedures 15 times in an unsupervised setting: K-means word clustering based on the cosine distance metric, extraction of the cluster keywords by k-d tree, and relevant document matching with the TF-IDF-based document clusters. The clustering results can be found in Supplementary Table 1 with the rates column indicating the appropriateness based on the percentage of all documents partitioned into specific clusters. One of the document clusters (Cluster 1) was used as the input data for the LDA analysis.

2.3.4. LDA topic modeling

Clusters with a big number of data frequently require further partitioning into smaller clusters (Yu et al., 2018), because investigating subcategories can potentially offer additional insights (Wen et al., 2017). Among the six clusters we generated from previous procedures, Cluster 1 was analyzed more in-depth as it accounted for 56.15% of the total data. We proceeded with a hybrid approach by applying topic modeling, because combining clustering and topic modeling can offer mutual assistance to one another as the output of clustering can be further characterized by main themes through applying topic modeling techniques (Shotorbani et al., 2016). In this study, LDA was selected, because it is a powerful topic model for revealing the hidden thematic structures within a collection of documents (Blei, 2012; Luyi & William, 2016) by considering each document as a composition of several topics produced by a mixture of words (Gangadharan & Gupta, 2020).

LDA was imported from the Gensim library, and the TF-IDF-based

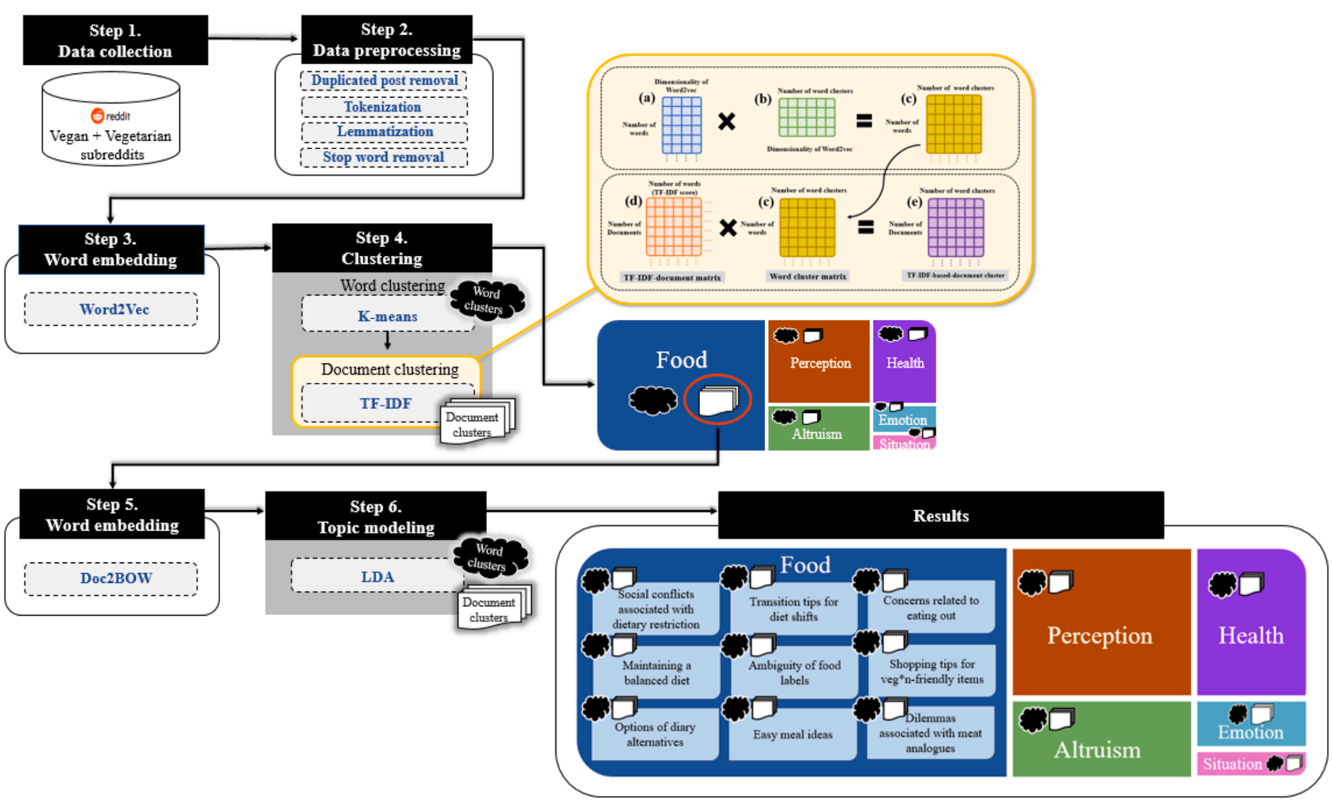

Fig. 1. Flowchart of r/Vegan and r/Vegetarian subreddit analyses. The nature of the clusters is illustrated with icons; black cloud icon symbolizes word clusters and white document icon indicates document clusters. The matrixes affiliated with the document clustering procedure is specified as (a) embedded word matrix; (b) cluster centroid matrix; (c) word cluster matrix; (d) TF-IDF document matrix; (e) TF-IDF-based document cluster matrix.

documents of Cluster 1 acquired in *2.3.3.* was used as the input data of LDA topic modeling. Since the format of the input data was in documents, word embedding was conducted through Doc2BOW provided within the Gensim LDA module. As the result, a bag of words (BOW) was produced, and a set of 6,785 tokens were obtained. The number of topics clusters was determined by the U_mass coherence score and qualitative assessment. The U_mass coherence score is based on the measure of pairwise distributional similarities over the sets of topic words, and is widely used for validation of topic numbers (Stevens et al., 2012; Zuo et al., 2016). The coherence score exceeding 0.5 indicated a reliable value (Fang et al., 2020) for nine topic clusters in Supplementary Fig. 1. In addition to the nine topics suggested by the coherence metric, we also experimented with different numbers of topics and decided on the optimal number of topics as nine based on consensus through the keywords and relevant posts. The final outcome of topic-word and document-topic distributions was obtained through LDA, and the topic cluster sizes were obtained by mapping each document to the most prominent topic. The representative documents were selected based on the contribution score of each document to a specific topic (Yang et al., 2020). In order to establish the labels of topics, four researchers reviewed the labels generated based on the keywords of each topic and the final labels were determined.

3. Results

3.1. Composition of clusters within online veg*n communities

We obtained six word clusters that represent the online veg*n communities (Fig. 2) and the proportion of each cluster is exhibited in Fig. 3. In this section, the findings are presented with the results of the word clusters, and the label of each K-means cluster was designated to reflect the nature of the top 20 keywords in Table 2.

Cluster 1 contained keywords for individual ingredients used including *onion, pepper, rice, garlic, potato, tomato, chickpea, mushroom,* and *butter*, and keywords that serve as the main ingredient to enhance taste were also identified. Such keywords that play a role in improving the taste of dishes included *sauce, spice,* and *seasoning*. As types of dishes, *pasta, curry, soup,* and *salad* appeared, along with *homemade* and *creamy* to describe the characteristics of cooking. *Bake* and *mix* were seen as keywords related to cooking methods. Considering these groups of emergent keywords, we labeled Cluster 1 as Food.

Keywords accompanied by negative nuances such as *preachy, angry, push, hypocrisy, insult,* and *ignorant* appeared in Cluster 2. In addition, terms representing an individual's perspective, such as *attitude, belief, view,* and *judge* could also be observed. *Defend* and *defensive* appeared at the front of the list together, indicating that it was an important keyword for the characteristics of the cluster. Therefore, Cluster 2 was also labeled reflecting the terms encompassing perception.

The most noticeable keywords in Cluster 3 were symptoms such as *fatigue, migraine, chronic, deficiency, inflammation, anemia, construction, headache,* and *blotting*, as well as the clinical terms *doctor* and *diagnose*. Keywords for health-affecting factors, such as *vitamins, excise, intake*, and *diet* also appeared, which indicated characteristics of health.

Cluster 4 included *human, livestock, animal*, and *breed*, along with keywords referring to the Earth, such as *ecosystem* and *land*. Terms indicating a method of practice such as *farming, agriculture*, and *production* were noted, as well as keywords indicating outcomes, including *unethical, destruction*, and *exploitation*. Terms presented in Cluster 4 covered altruistic values such as animal production and agricultural practices, which are closely related to ethical motivators of veg*nism

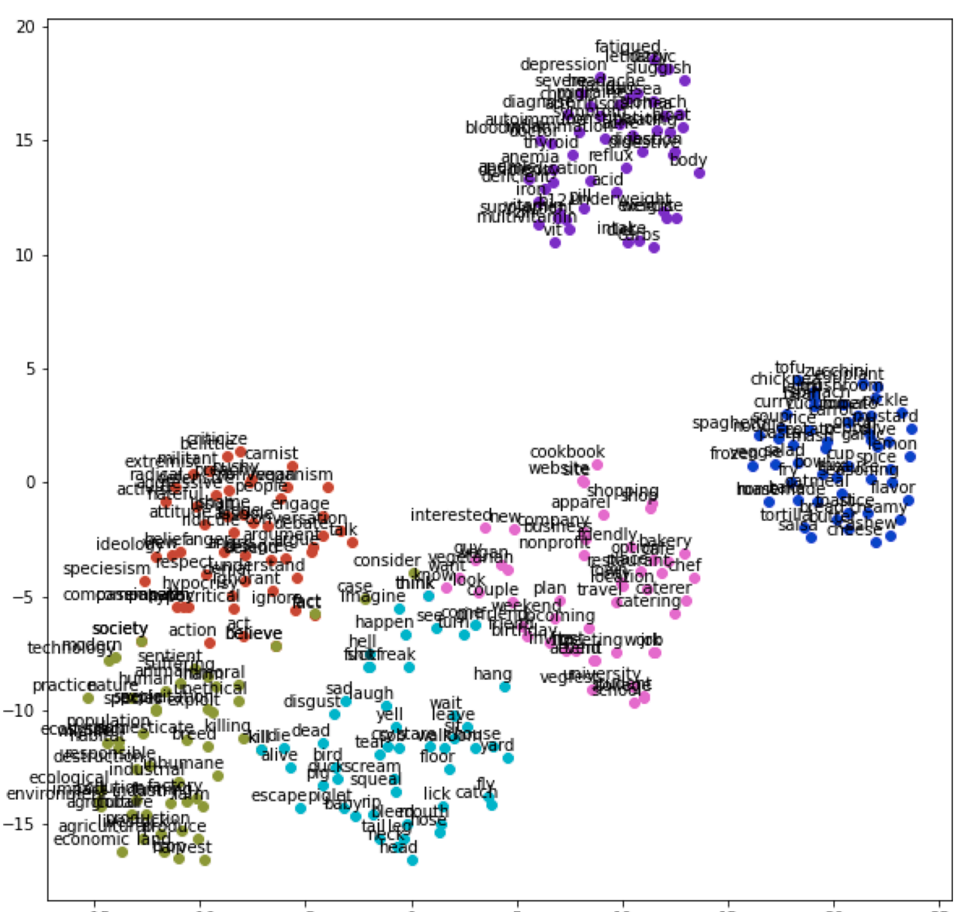

Fig. 2. K-means clusters of r/Vegan and r/Vegetarian. Words belonging to each cluster are expressed in colored dots.

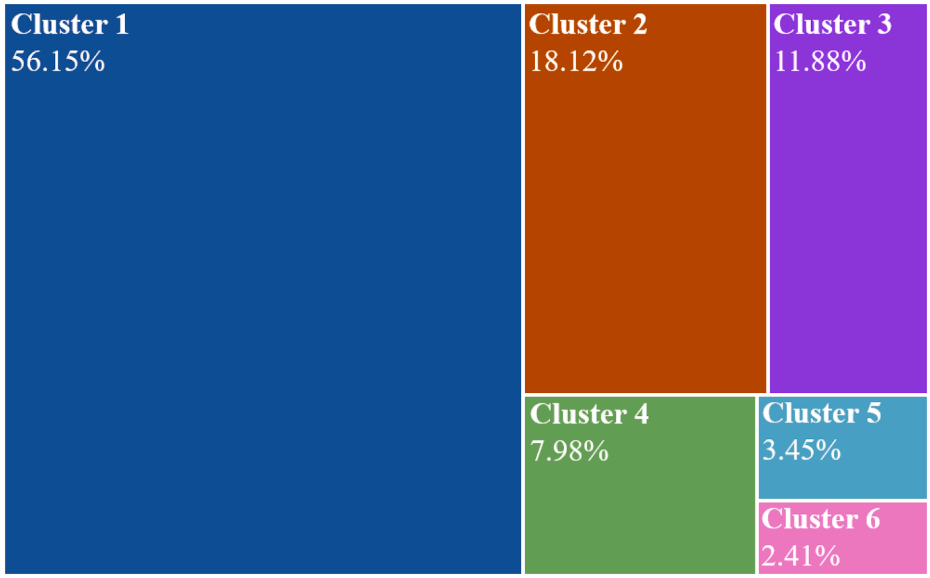

Fig. 3. Ratios of the main clusters from the online veg*n communities. Treemap areas are expressed in the same colors as the matching clusters in Fig. 2.

Table 2
Representative keywords of 6 clusters in the online veg*n communities.[a]

Label	Keywords
Food	sauce, mix, onion, pepper, homemade, rice, garlic, potato, pasta, tomato, pasta, tomato, spice, seasoning, chickpea, curry, soup, mushroom, salad, creamy
Perception	people, defend, attitude, preachy, angry, defensive, extremist, belief, view, judge, understand, pushy, hypocrisy, act, veganism, insult, ignorant, aggressive, conversation, radical
Health	fatigue, symptom, migraine, chronic, digestion, doctor, vitamin, deficiency, inflammation, anemia, exercise, intake, constipation, weight, anemic, headache, diet, diagnose, thyroid, bloating
Altruism	human, species, livestock, production, species, ecosystem, ecological, farming, agriculture, believe, crop, agricultural, destruction, animal, breed, exploitation, land, suffering, exploit, unethical
Emotion	floor, cry, sit, scream, leave, bird, dead, walk, piglet, think, laugh, tear, mouth, pig, leg, kill, house, rip, head, hang
Situation	think, vegan, university, vegetarian, restaurant, know, city, plan, event, look, want, new, work, attend, travel, invite, catering, interested, chef, college

[a] The enlisted words are the top 20 keywords obtained from k-d tree.

(Cherry, 2015; Plante et al., 2019).

In Cluster 5, terms related to emotion included *cry, scream, laugh*, and *tear*. Additionally, *mouth, leg, head* representing body parts, and *dead, kill, rip*, and *hang* reminiscent of death appeared. Animal related keywords that appeared included *bird, piglet*, and *pig*.

Keywords describing situation such as, *university, work*, and *college* appeared in Cluster 6. In reference to events, *plan, event, attend*, and *invite* were observed. Keywords indicating geological settings of *city* and *travel* were also found, as well as the venue-related terms *restaurant, catering*, and *chef*.

Each cluster was labeled by a theme that describes the characteristics most effectively: Cluster 1 as Food, Cluster 2 as Perception, Cluster 3 as Health, Cluster 4 as Altruism, Cluster 5 as Emotion, and Cluster 6 as Situation. The proportion of each cluster is presented in Fig. 3 denoting the Food cluster as the major section at 56.15%, followed by Perception (18.12%), Health (11.88%), Altruism (7.98%), Emotion (3.45%), and Situation (2.41%) (Table 2, Fig. 3).

3.2. Further analysis of the notion of food for veg*ns

Numerous studies have highlighted the significance of food for individuals who practice veg*nism (Boyle, 2011; Rosenfeld, 2019; Stiles, 1998). The meaning of food for veg*ns has been described as a gateway that offers self-realization and a part of the healing process (Costa et al., 2019). As can be seen in Fig. 3, food accounted for over 50% of the total dataset, emphasizing the importance of the cluster. Thus, we conducted an additional analysis of the Food cluster using the LDA topic modeling technique to investigate the contents further. Within the documents from the Food cluster, nine topics were identified as subcategories. The ratio of each topic cluster and affiliated keywords are presented in Fig. 4. Since there were nine topics derived from one cluster, the findings are conveyed with example posts to illustrate the details. The Example posts have been paraphrased and partially masked in order to maintain the privacy of the users.

In Topic 1, terms such as *decision*, *guilty*, *environmental*, *impact*, *moral* as well as *pig*, and *farming* were seen as motives for individuals to follow veg*n dietary patterns. However, keywords describing family, such as *parent* and *mum* were noticed at the same time. Also, there were terms illustrating conflicts including *push*, *quit*, *cry*, *argument*, and *anxiety*. These resultant terms and example posts of Topic 1 indicated restricted dietary patterns may cause conflicts between the users and omnivores around them, especially when the user is the only person practicing veg*nism in the social circle. Thereby, Topic 1 was labeled as Social conflicts associated with dietary restriction. Examples of those posts are shown below.

> *"I have been vegetarian for less than three years and want to go vegan for environmental and ethical reasons. However, I am a teen living with my parents and my mom cooks almost every night. Every time I ask her for vegan dinner, she tells me that is not going to happen because preparing vegan dinner every night is extremely hard and it is not fair to the rest of the family."*
> *"I am 15 years old and had been cutting back on meat slowly. When I finally refused to eat meat tonight, my parents got furious and told me going plant-based is unhealthy and I will have to eat meat as long as I live with them."*

Topic 2 exhibited keywords of information sources including *helpful*, *cookbook*, *blog*, *recipe*, *channel*, and *youtube*. Terms such as *beginner* and *welcome* indicated the users' recent diet shifts, and the potential characteristics of required culinary skills were illustrated as *adventurous*, and

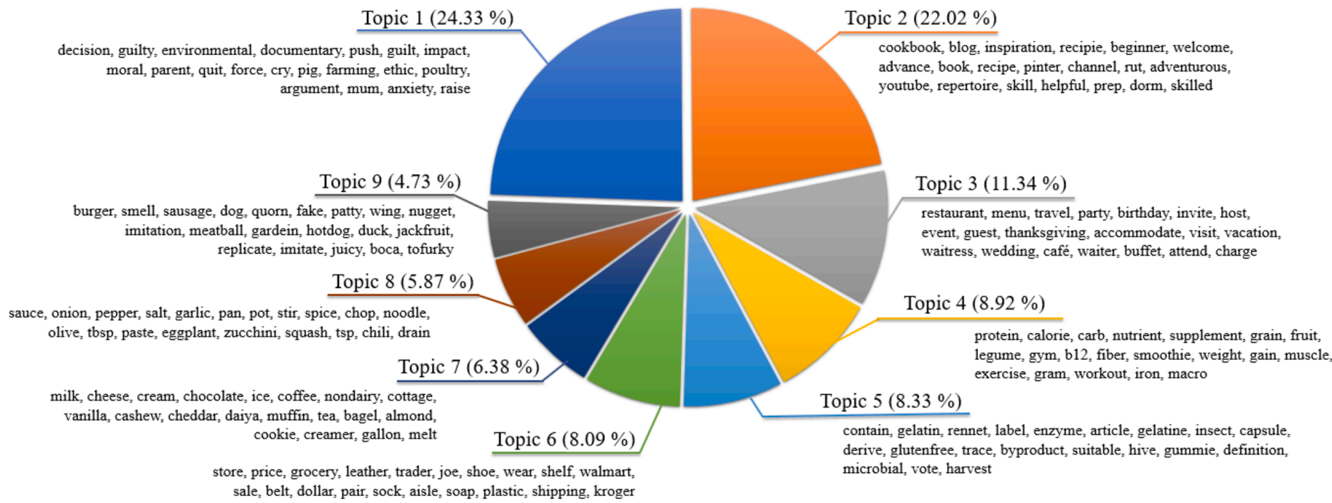

Fig. 4. Specified food subcategories via LDA topic modeling. Duplicated terms that appeared in more than one topic were removed to provide a concise list of representative keywords of each topic.

skilled. The occurrence of *prep* and *dorm* suggested that the users may have limited cooking resources. We also observed that challenges of dietary transition can apply to beginners as well as to individuals who have been practicing another form of a veg*n diet. Considering the keywords and the posts where the users sought to find solutions by requesting recipes and channels of information, Topic 2 was labeled as Transition tips for diet shifts. Examples of those posts are shown below.

"I realized transitioning from vegetarian to vegan is tougher than I imagined, because my usual meals include eggs or dairy. Can someone recommend a collection of recipes or blogs I can look into?"
"I'm trying to find the best way for me to go vegan. I live in a college dorm, so don't really have the resources for time-consuming recipes. Is there any meal prep I can do to make things easier?"

Keywords affiliated with eateries were noticed in Topic 3 through *restaurant*, *menu*, *waitress*, *café*, *waiter*, *buffet*, and *charge*. Terms such as *travel*, *party*, *birthday*, *event*, *thanksgiving*, *vacation*, and *wedding* indicated the occasions. Not only did we find that the users shared their unpleasant experiences related to eateries, we also confirmed that the users were often concerned with the occasions that included dining. Based on these observations, Topic 3 was labeled as Concerns related to eating out. Examples of those posts are shown below.

"Today I went to a restaurant that serves only one vegan meal with an omni friend. As we tried to place an order, the waitress told us they were out of the item. We went right after they started to serve dinner, and I still don't understand how a restaurant can run out of a menu so fast."
"I'm invited to a wedding and the meal options are chicken or pork. When I asked about vegetarian options, they told me there will be cake and maybe a salad. Would it be rude if I brought my own food to the wedding?"

Terms related to nutritional intake were observed in Topic 4 by *calorie* and *nutrient*, as well as specific nutrients including *protein*, *carb*, *b12*, *fiber*, *iron*, and *macro*. Keywords such as *supplement*, *fruit*, *legume*, and *smoothie* indicated the nutrient sources, and *gym*, *weight*, *gain*, *muscle*, *exercise*, and *workout* indicated users' interests in energy balance. The posts indicated that many users were keen on inspecting the nutritional values of the food they consumed. Thus, Topic 4 was labeled as Maintaining a balanced diet. Examples of those posts are shown below.

"I want some recommendations for mid-day smoothies before a workout. I am thinking of spinach and kale with protein powder and some bananas for calories. Do I need to add any other vegetables or stuff for calories?"
"I just started weight training and am consistently hungry! I eat over 2000 calories with tofu, vegetables, fruits, nuts, beans, and grains. The goal is to have 120 g of protein per day."

In Topic 5, *gelatin*, *rennet*, *insect*, and *hive* denoted animal-based ingredients. Keywords including *derive*, *trace*, and *byproduct* indicated non-veg*n contaminants as well as *enzyme* and *microbial*. *Capsule* and *gummie* specified the product applications, and *label*, *article*, and *suitable* indicated the criteria that the users check for suitability. We noticed that users mostly referred to the food labels to confirm the products do not contain animal ingredients. Also, users were often confused by vague statements of allergen information and ingredients enlisted within food labels. In this regard, Topic 5 was labeled as Ambiguity of food labels. Examples of those posts are shown below.

"I bought this candy which I've been eating for a while. When I looked at the ingredient list on the back, I didn't find any animal products, and there were no statements saying it contained animal products either. However, when I looked up the ingredients today on their website, it said their candy contains gelatin."

"I just found out I bought a spaghetti with a label saying "May contain traces of eggs". Does it mean it contains some traces of animal products? It's kind of confusing to me.".

Purchase-related keywords were predominant in Topic 6 presented by *grocery*, *store*, *price*, *shelf*, *sale*, *dollar*, *aisle*, and *shipping*. Also, supermarket franchise businesses including *trader*, *joe*, *walmart*, and *kroger* were observed. The users posted questions regarding stores that carry specific food items they required preferably at lower prices. Considering these components, Topic 6 was labeled as Shopping tips for veg*n-friendly items. Examples of those posts are shown below.

"I've been wanting to try seitan. Can someone tell me which brand to buy? I live around Walmart, Aldi, and Trader Joe's. I prefer other places than Trader Joe's if not necessary, because their prices are very high in my city."
"Does anybody have a guide to vegan sugar? Ever since Veganuary I've been only using maple syrup and brown sugar I bought at a local grocery store, which was expensive but vegan and organic certified. I haven't baked anything with regular sugar since because I can't tell which brands of sugar are vegan or not."

Topic 7 contained *milk*, *cheese*, *cream*, *nondairy*, *cottage*, *cheddar*, *creamer*, and *gallon* indicating dairy alternatives. *Cashew* and *almond* were seen as the ingredients as well as *chocolate* and *vanilla* for the flavors. *ice*, *coffee*, *muffin*, *tea*, *bagel*, and *cookie* suggested foods that the users may apply dairy alternatives to. The users often posted reviews or experiences related to various dairy substitutes, and also categorized the alternatives by the purpose of usage. Thus, Topic 7 was labeled as Options of dairy alternatives. Examples of those posts are shown below.

"I just finished a pint of mint chocolate cookie and it was amazing! here's my rating for some of the ice cream flavors I've tried. Mint chocolate cookie (9.5 out of 10), banana foster (9 out of 10), coffee caramel fudge (8 out of 10), and caramel almond brittle (7.5 out of 10). Haven't tried boom chocolatta, but I think mint chocolate cookie will stay number one."

"Almond and soy milks were great in coffee and cereal, but they didn't go too well with tea. So, I've been thinking about trying oat or cashew milks.".

In Topic 8, keywords related to individual ingredients of recipes were confirmed including *sauce*, *onion*, *pepper*, *salt*, *garlic*, *spice*, *noodle*, *olive*, *paste*, *eggplant*, *zucchini*, *squash*, and *chili* were observed in Topic 8. Keywords describing instructions such as *stir*, *chop*, *tbsp*, *tsp*, and *drain* were noticed with kitchen utensils including *pan* and *pot*. The recipes shared by the users were often brief and seldom time-consuming. Based on these observations, Topic 8 was labeled as Easy meal ideas. Examples of those posts are shown below.

"I love roasted veggie salads, because you can change the ingredients any way you want. I prefer using a slow cooker, but sheet pan recipes are as good. First, you roast large diced vegetables with some bread (you can add anything like mushrooms, eggplant, or zucchini). Then drizzle with some olive oil or balsamic vinegar. Go ahead if you want to add some spices and herbs. Roasting temperature and time will depend on the types of vegetables used."

"I just perfected my breakfast recipe and wanted to share! 1 tbsp olive oil, 1 10 oz canned chickpeas (drained and rinsed), 1 yellow onion (diced), 1/2 scallions (chopped), 2 tbsp lemon juice, 2 tbsp soy sauce, 1 tsp of sriracha, 1 clove garlic (minced). Place chickpeas in a bowl and make a chunky mixture using a fork. Add lemon juice, soy sauce, sriracha, and garlic and mix well. Sauté onion until it turns translucent over medium heat (around 5 min). Then add the chickpea chunk mixture and stir until the mixture turns brown and crispy (around 10 min). Place scallions to garnish and serve hot."

Topic 9 represented keywords such as *fake*, *imitation*, and *replicate*, as well as *burger*, *sausage*, *patty*, *wing*, *nugget*, *meatball*, and *hotdog*, which indicated an association with meat analogues. Sensory-related terms including *smell* and *juicy* were observed. Product brands were identified with *quorn*, *gardein*, *boca*, and *tofurky*. The users often shared their thoughts on meat substitutes including cultured meat, and the general opinion on the smell was rather negative. Also, many users showed rejection of meat analogues that mimic the texture of animal products. Therefore, Topic 9 was labeled as Dilemmas associated with meat analogues.

"I am about to order some meat substitutes, but don't want to buy something that smells like dog food like the last time I tried. What are some meat substitutes that don't smell awful?"

"I really don't like fake ground beef, turkey slices, and sausages. One of the reasons I decided to go vegan was that I never actually liked the texture and taste of meat products. I prefer stuff made with beans and vegetables."

The nine subcategories of the Food cluster were labeled with themes that best describe the characteristics: Topic 1 as Social conflicts associated with dietary restriction (24.33%), Topic 2 as Transition tips for diet shifts (22.02%), Topic 3 as Concerns related to eating out (11.34%), Topic 4 as Maintaining a balanced diet (8.92%), Topic 5 as Ambiguity of food labels (8.33%), Topic 6 as Shopping tips for veg*n friendly items (8.09%), Topic 7 as Options of dairy alternatives (6.38%), Topic 8 as Easy meal ideas (5.87%), and Topic 9 as Dilemmas associated with meat analogues (4.73%).

4. Discussion

With rising attention across the globe, many people are considering a transition to some form of veg*nism (Aschemann-Witzel et al., 2021; Saari et al., 2021; Tziva et al., 2020) for various reasons including the environment, animal and human rights, and health (Fox & Ward, 2008; Phillips, 2005; Plante et al., 2019; Ruby, 2012). However, discourse on the aspects and interests of veg*ns has only been partially observed and within a limited scope.

We sought to investigate the veg*n ecosystem by text mining-based analytics of the online communities to identify various hardships that veg*ns encounter. As can be seen in Fig. 1, we confirmed six main categories by exploring the keywords and posts from the r/Vegan and r/Vegetarian subreddits via K-means and TF-IDF. Additionally, the Food cluster was further sectioned into nine topics by implementing LDA topic modeling. By examining the veg*n communities, we found that the notion of food for veg*ns is not bound to energy intake alone, but also informs the basis of a broader identity.

As previous studies have reported, the main concerns of transition included sudden changes in dietary patterns and meal preparation (Lea et al., 2006; Pohjolainen et al., 2015), which is related to the barriers to sustainable dietary patterns presented by researchers (Cherrier et al., 2012; Lea et al., 2006). In our study, related posts often inquired about simple recipes with minimal ingredients and kitchenware, supported by terms and posts in transition tips for diet shifts (Food cluster, Topic 2). This regard applied not only to the users who recently started veg*n diets, but also to those aiming to transition to a stricter veg*n type across the spectrum. We also noticed events such as Veganuary and 30-day challenges affect the inflow of users. New users tended to collect information on veg*n friendly products by asking existing members for favorite foods and recipes to enable a smooth transition (Food cluster, Topic 6). Associatively, there were users who introduced simple recipes requiring a short list of ingredients (Food cluster, Topic 8).

Additionally, concerns relating to the nutritional sufficiency of veg*n diets (Pohjolainen et al., 2015) were also found among the veg*n population. Many users wanted assurance by sharing their food compositions to be evaluated with personal *calorie* and *nutrient* intake goals (Food cluster, Topic 4). Also, keywords such as *fatigue, migraine, anemia, vitamin*, and *deficiency* appearing in the Health cluster of this study emphasized the nutritional factors of interest covered in previous studies (Craig & Mangels, 2009; Foster et al., 2013; Remer et al., 1999). This indicates that users are aware of the adverse effects may follow restricted dietary patterns.

Although there have been studies showing that consumers are willing to pay higher prices for sustainable products (Forbes et al., 2009; Sánchez-Bravo et al., 2021), our findings indicate there are many veg*n customers who are sensitive toward prices, as described by the terms *price*, *sale*, *dollar*, and posts where users sought to find moderate-priced products (Food cluster, Topic 6). The general perception of veg*n food being expensive (Paslakis et al., 2020; Rivera & Shani, 2013) was also confirmed.

Various combinations of plant-based foods and alternatives have been adopted due to reduced choices arising from the avoidance of animal-derived products (Leitzmann, 2014; Rizzo et al., 2013; Turner-McGrievy et al., 2015). We confirmed a wide variety of dairy products that are currently available to veg*ns as plant-based forms on the market and interests in related products, including *milk*, *cheese*, and *cream* (Food cluster, Topic 7). Furthermore, an expanding lineup of products now offers alternatives to meat, fish, and egg products including those made of soy, almond, oats, cashew, and coconut (Cleveland et al., 2021; Curtain & Grafenauer, 2019; Jeske et al., 2018; Tziva et al., 2020). The rising consumption of these emerging food categories was also observed in our study (Food cluster, Topic 9).

Considering such rapid market growth and changing consumer preferences, associated certification systems have been implemented by various institutions and organizations (Wrenn, 2011), acknowledging that veg*n consumers also tend to be more cautious with labels when purchasing manufactured products (Kumar & Kapoor, 2017). We confirmed that product labels play an important role for making purchase decisions in veg*n population (Food cluster, Topic 5). However, there also existed cases where the users accidentally purchased animal-derived products due to being misled by product labels. Incidents of purchasing products containing animal derived ingredients often involved *gelatin* and *rennet*. The confusion derived from selecting animal product alternatives and miscommunication in product labels were also shared among users.

Although food is an essential component, veg*nism refers to a more complex concept beyond the diet, which centers upon dietary commitments as a way to express an identity (Greenebaum, 2012; Jabs et al., 2000). It was confirmed that the various motives of veg*ns mentioned in previous studies (e.g. environment, health, animal rights, and human rights) remain important topics within the online communities as well (Altruism cluster, Health cluster). In the process of practicing alternative dietary patterns, many veg*ns face difficulties in social relationship aspects (Hirschler, 2011). According to previous studies, cynical labels such as *preachy* and *angry* are often given to veg*ns (Earle & Hodson, 2017; Minson & Monin, 2012), as we confirmed in the Perception cluster. In accordance, a negative undertone was widespread in the Emotion cluster as well including the terms of *cry*, *scream*, and *tear*. Furthermore, posts illustrating experiences of impacted relationships with family members were prevalent (Markowski & Roxburgh, 2019). Conflicts may arise within households, as users' dietary patterns were frequently dismissed or ignored when eating with omnivorous family members (Food cluster, Topic 1). In many cases, the grievances felt by members of the omnivore family were expressed by minors who did not have control over meal preparation, and there were posts where frustration was expressed about other family members' food choices (Cherry, 2015). Although the general perception of appropriate meals including meat dishes may influence those who prepare them (Holm & Møhl, 2000), we observed that it can also lead users towards feelings of disconnection within close social circles. Perhaps households are one of many places where veg*ns may feel isolated, as we confirmed various settings in the Situation cluster with keywords including *university* and *work*. In addition, veg*ns often face difficulties when dining out, mainly due to limited options and a lack of employee training (Choi et al., 2022). The users of the online veg*n communities presented similar patterns, indicated by the episodes they shared related to dining experiences (Food cluster, Topic 3). The significant topic proportion of Social conflicts associated with dietary restriction may indicate individuals' alternative diets can potentially influence their social relations with omnivores (Nezlek et al., 2021).

In this study, various interests and opinions of users were investigated through online veg*n community analysis, and the flow of dietary hardships and information sharing closely related to daily life was observed. However, there exist several limitations to our findings. First, only posts were treated as data for the analyses, and the interactions within comments between the author and the respondent were not assessed. Thus, obtaining a more in-depth understanding of communication between users might have been prevented. Second, detailed information on the users was not obtained, which left factors regarding personal dietary preferences unaccounted for. Third, although there are various spectra within the category of veg*nism, analysis by veg*n type could not be performed because the users' types were not necessarily indicated. In addition, conducting an additional investigation on one cluster may have prevented us from interpreting the veg*n lifestyle in every aspect. Therefore, other clusters of relatively smaller sizes (e.g., Perception, Health, and Altruism) require further elucidation. As new techniques continuously emerge, the implementations of various clustering and topic modeling approaches may contribute to further discovery with multi-faceted analysis. Despite these limitations, it was self-evident that communities play a significant supporting role for veg*ns, and the study was able to record their lively first-hand experiences.

5. Conclusions

This study analyzed a population of online veg*n communities with the aim to record the content and characteristics expressed by users in unsupervised settings. By using text mining techniques, we noted various situations that individuals encounter while practicing veg*nism and especially verified the notion that food for veg*ns is not bound to energy intake, but rather constitutes an aspect of their lifestyle. Moreover, food is a major aspect that encompasses changes in dietary patterns and detailed components that contribute to each individual's daily life, by influencing routines including grocery shopping, food intake, meal preparation, and social interactions.

We can conclude that adequate dietary guidelines are needed to prevent compromised nutrition due to sudden dietary changes, as well as social acceptance. For those wishing to transition to veg*nism, we suggest acknowledging several changes to be expected in daily routines following a new dietary regime. Arranging available cooking resources and having to alternate recipes might be challenging for some, while changes in shopping patterns, and practical and social pressures associated with preparing daily meals require additional effort to secure adequate nutrient intake. In addition, adopting veg*n diets may influence each individual's personal paradigm outside of food choices and lead to new social and environmental changes. Thus, careful due diligence with veg*n nutrition is required and gradual replacement of animal-based products may help the transition process. With a more comprehensive understanding of veg*ns based on our community analyses, we hope our findings can be leveraged for better communication to the manufacturers, health educators, and nutritionists targeting the veg*n population, as well as provide insight to individuals who are planning a dietary shift.

Funding

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the High Value-added Food Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (321029-05), and also supported by the K-BIO KIURI Center Program through the Ministry of Science and ICT (2020M3H1A1073304), and the BK21 FOUR Program of the Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea.

CRediT authorship contribution statement

Ra Yoo: Conceptualization, Investigation, Writing – original draft, Visualization, Project administration. **Seo-Young Kim:** Resources. **Do-Hee Kim:** Methodology. **Jiyoung Kim:** Data curation. **Ye Ji Jeon:** Validation. **Jung Han Yoon Park:** Supervision. **Ki Won Lee:** Supervision, Funding acquisition. **Hee Yang:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodqual.2022.104714.

Reference

Al-Saqqa, S., & Awajan, A. (2019). The use of word2vec model in sentiment analysis. In *Proceedings of the 2019 international conference on artificial intelligence, robotics and control* (pp. 39–43). https://doi.org/10.1145/3388218.3388229

Aschemann-Witzel, J., Gantriis, R. F., Fraga, P., & Perez-Cueto, F. J. A. (2021). Plant-based food and protein trend from a business perspective: Markets, consumers, and the challenges and opportunities in the future. *Critical Reviews in Food Science and Nutrition, 61*(18), 3119–3128. https://doi.org/10.1080/10408398.2020.1793730

Bafna, P., Pramod, D., & Vaidya, A. (2016). Document clustering: TF-IDF approach. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (pp. 61–66). https://doi.org/10.1109/ICEEOT.2016.7754750

Barr, S. I., & Chapman, G. E. (2002). Perceptions and practices of self-defined current vegetarian, former vegetarian, and nonvegetarian women. *Journal of the American Dietetic Association, 102*(3), 354–360. https://doi.org/10.1016/S0002-8223(02)90083-0

Beardsworth, A. D., & Keil, E. T. (1991). Vegetarianism, veganism and meat avoidance: Recent trends and findings. *British Food Journal, 93*(4), 19–24.

Bentley, J. L. (1990). k-d trees for semidynamic point sets. *Proceedings of the Sixth Annual Symposium on Computational Geometry - SCG '90*, 187–197. https://doi.org/10.1145/98524.98564.

Beverland, M. B. (2014). Sustainable eating: Mainstreaming plant-based diets in developed economies. *Journal of Macromarketing, 34*(3), 369–382. https://doi.org/10.1177/0276146714526410

Blackburn, K. G., Yilmaz, G., & Boyd, R. L. (2018). Food for thought: Exploring how people think and talk about food online. *Appetite, 123*, 390–401. https://doi.org/10.1016/j.appet.2018.01.022

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Bodendorf, F., & Kaiser, C. (2010). Detecting opinion leaders and trends in online communities. *Fourth International Conference on Digital Society, 2010*, 124–129. https://doi.org/10.1109/ICDS.2010.29

Boussalis, C., & Coan, T. G. (2016). Text-mining the signals of climate change doubt. *Global Environmental Change, 36*, 89–100. https://doi.org/10.1016/j.gloenvcha.2015.12.001

Boyle, J. E. (2011). Becoming vegetarian: The eating patterns and accounts of newly practicing vegetarians. *Food and Foodways, 19*(4), 314–333. https://doi.org/10.1080/07409710.2011.630620

Bresnahan, M., Zhuang, J., & Zhu, X. (2016). Why is the vegan line in the dining hall always the shortest? *Understanding vegan stigma. Stigma and Health, 1*(1), 3–15. https://doi.org/10.1037/sah0000011

Campbell, T. C. (2017). Cancer prevention and treatment by wholistic nutrition. *Journal of Nature and Science, 3*(10).

Cherrier, H., Szuba, M., & Özçağlar-Toulouse, N. (2012). Barriers to downward carbon emission: Exploring sustainable consumption in the face of the glass floor. *Journal of Marketing Management, 28*(3–4), 397–419. https://doi.org/10.1080/0267257X.2012.658835

Cherry, E. (2015). I was a teenage Vegan: Motivation and maintenance of lifestyle movements. *Sociological Inquiry, 85*(1), 55–74. https://doi.org/10.1111/soin.12061

Choi, H., Joung, H.-W., Choi, E.-K., & Kim, H.-S. (2022). Understanding vegetarian customers: The effects of restaurant attributes on customer satisfaction and behavioral intentions. *Journal of Foodservice Business Research, 25*(3), 353–376. https://doi.org/10.1080/15378020.2021.1948296

Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2017). In search of new product ideas: Identifying ideas in online communities by machine learning and text mining. *Creativity and Innovation Management, 26*(1), 17–30. https://doi.org/10.1111/caim.12202

Cleveland, D. A., Gee, Q., Horn, A., Weichert, L., & Blancho, M. (2021). How many chickens does it take to make an egg? Animal welfare and environmental benefits of replacing eggs with plant foods at the University of California, and beyond. *Agriculture and Human Values, 38*(1), 157–174. https://doi.org/10.1007/s10460-020-10148-z

Contini, C., Boncinelli, F., Marone, E., Scozzafava, G., & Casini, L. (2020). Drivers of plant-based convenience foods consumption: Results of a multicomponent extension of the theory of planned behaviour. *Food Quality and Preference, 84*, Article 103931. https://doi.org/10.1016/j.foodqual.2020.103931

Costa, I., Gill, P. R., Morda, R., & Ali, L. (2019). "More than a diet": A qualitative investigation of young vegan Women's relationship to food. *Appetite, 143*. https://doi.org/10.1016/j.appet.2019.104418

Craig, W. J., & Mangels, A. R. (2009). Position of the American dietetic association: Vegetarian diets. *Journal of the American Dietetic Association, 109*, 1266–1282. https://doi.org/10.1016/j.jada.2009.05.027

Curtain, F., & Grafenauer, S. (2019). Plant-based meat substitutes in the flexitarian age: An audit of products on supermarket shelves. *Nutrients, 11*(11), 2603. https://doi.org/10.3390/nu11112603

Ding, C., Utiyama, M., & Sumita, E. (2019). NOVA: A feasible and flexible annotation system for joint tokenization and part-of-speech tagging. *ACM Transactions on Asian and Low-Resource Language Information Processing, 18*(2), 1–18. https://doi.org/10.1145/3276773

Earle, M., & Hodson, G. (2017). What's your beef with vegetarians? Predicting anti-vegetarian prejudice from pro-beef attitudes across cultures. *Personality and Individual Differences, 119*, 52–55. https://doi.org/10.1016/j.paid.2017.06.034

Fang, Y., Zhao, C., Huang, C., & Liu, L. (2020). SankeyVis: Visualizing active relationship from emails based on multiple dimensions and topic classification methods. *Forensic Science International: Digital Investigation, 35*, Article 300981. https://doi.org/10.1016/j.fsidi.2020.300981

Felix, U. (2004). Performing beyond the comfort zone: Giving a voice to online communication. In *Proceedings of the 21st ASCILITE Conference* (pp. 284–293).

Forbes, S. L., Cohen, D. A., Cullen, R., Wratten, S. D., & Fountain, J. (2009). Consumer attitudes regarding environmentally sustainable wine: An exploratory study of the New Zealand marketplace. *Journal of Cleaner Production, 17*(13), 1195–1199. https://doi.org/10.1016/j.jclepro.2009.04.008

Forestell, C. A., Spaeth, A. M., & Kane, S. A. (2012). To eat or not to eat red meat. A closer look at the relationship between restrained eating and vegetarianism in college females. *Appetite, 58*(1), 319–325. https://doi.org/10.1016/j.appet.2011.10.015

Foster, M., Chu, A., Petocz, P., & Samman, S. (2013). Effect of vegetarian diets on zinc status: A systematic review and meta-analysis of studies in humans. *Journal of the Science of Food and Agriculture, 93*(10), 2362–2371. https://doi.org/10.1002/jsfa.6179

Fox, N., & Ward, K. (2008). Health, ethics and environment: A qualitative study of vegetarian motivations. *Appetite, 50*(2–3), 422–429. https://doi.org/10.1016/j.appet.2007.09.007

Gangadharan, V., & Gupta, D. (2020). Recognizing named entities in agriculture documents using LDA based topic modelling techniques. *Procedia Computer Science, 171*, 1337–1345. https://doi.org/10.1016/j.procs.2020.04.143

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications, 69*, 214–224. https://doi.org/10.1016/j.eswa.2016.10.043

Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., Pierrehumbert, R. T., Scarborough, P., Springmann, M., & Jebb, S. A. (2018). Meat consumption, health, and the environment. *Science, 361*(6399). https://doi.org/10.1126/science.aam5324

Graça, J., Godinho, C. A., & Truninger, M. (2019). Reducing meat consumption and following plant-based diets: Current evidence and future directions to inform integrated transitions. *Trends in Food Science & Technology, 91*, 380–390. https://doi.org/10.1016/j.tifs.2019.07.046

Greenebaum, J. (2012). Veganism, identity and the quest for Authenticity. *Food, Culture and Society, 15*(1), 129–144. https://doi.org/10.2752/175174412X13190510222101

Grossmann, L., & McClements, D. J. (2021). The science of plant-based foods: Approaches to create nutritious and sustainable plant-based cheese analogs. *Trends in Food Science & Technology, 118*, 207–229. https://doi.org/10.1016/j.tifs.2021.10.004

Guan, T., Wang, L., Jin, J., & Song, X. (2018). Knowledge contribution behavior in online Q&A communities: An empirical investigation. *Computers in Human Behavior, 81*, 137–147. https://doi.org/10.1016/j.chb.2017.12.023

He, J., Evans, N. M., Liu, H., & Shao, S. (2020). A review of research on plant-based meat alternatives: Driving forces, history, manufacturing, and consumer attitudes. *Comprehensive Reviews in Food Science and Food Safety, 19*(5), 2639–2656. https://doi.org/10.1111/1541-4337.12610

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods, 25*(1), 114–146. https://doi.org/10.1177/1094428120971683

Hirschler, C. A. (2011). "What pushed me over the edge was a deer hunter": Being vegan in North America. *Society and Animals, 19*(2), 156–174. https://doi.org/10.1163/156853011X562999

Holm, L., & Møhl, M. (2000). The role of meat in everyday food culture: An analysis of an interview study in Copenhagen. *Appetite, 34*(3), 277–283. https://doi.org/10.1006/appe.2000.0324

Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in Twitter. *Proceedings of the 20th International Conference Companion on World Wide Web - WWW '11*, 57. https://doi.org/10.1145/1963192.1963222.

Jabs, J., Sobal, J., & Devine, C. M. (2000). Managing vegetarianism: Identities, norms and interactions. *Ecology of Food Nutrition, 39*(5), 375–394. https://doi.org/10.1080/03670244.2000.9991625

Jamnik, M. R., & Lane, D. J. (2017). The use of Reddit as an inexpensive source for high-quality data. *Practical Assessment, Research and Evaluation, 22*(5), 1–10. https://doi.org/10.7275/J18T-C009

Jeske, S., Zannini, E., & Arendt, E. K. (2018). Past, present and future: The strength of plant-based dairy substitutes based on gluten-free raw materials. *Food Research International, 110*, 42–51. https://doi.org/10.1016/j.foodres.2017.03.045

Kapil, S., & Chawla, M. (2016). Performance evaluation of K-means clustering algorithm with various distance metrics. In *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)* (pp. 1–4). https://doi.org/10.1109/ICPEICES.2016.7853264

Kumar, N., & Kapoor, S. (2017). Do labels influence purchase decisions of food products? Study of young consumers of an emerging market. *British Food Journal, 119*(2), 218–229. https://doi.org/10.1108/BFJ-06-2016-0249

Lea, E. J., Crawford, D., & Worsley, A. (2006). Public views of the benefits and barriers to the consumption of a plant-based diet. *European Journal of Clinical Nutrition, 60*(7), 828–837. https://doi.org/10.1038/sj.ejcn.1602387

Leitzmann, C. (2014). Vegetarian nutrition: Past, present, future. *American Journal of Clinical Nutrition, 100*(SUPPL. 1), 496S–502S. https://doi.org/10.3945/ajcn.113.071365

Li, L., Wang, J., & Li, X. (2020). Efficiency Analysis of Machine Learning Intelligent Investment Based on K-Means Algorithm. *IEEE Access, 8*, 147463–147470. https://doi.org/10.1109/ACCESS.2020.3011366

Lin, K.-P., Shen, C.-Y., Chang, T.-L., & Chang, T.-M. (2017). A Consumer Review-Driven Recommender Service for Web E-Commerce. In *2017 IEEE 10th Conference on*

Service-Oriented Computing and Applications (SOCA) (pp. 206–210). https://doi.org/10.1109/SOCA.2017.35
Lin, Y. (2021). in 2020 (infographic). Oberlo *Reddit Statistics Every Marketer Should Know, May 11). 10* https://www.oberlo.com/blog/reddit-statistics.
Liu, H., Wu, L., Zhang, D., Jian, M., & Zhang, X. (2018). Multi-perspective User2Vec: Exploiting re-pin activity for user representation learning in content curation social network. *Signal Processing, 142*, 450–456. https://doi.org/10.1016/j.sigpro.2017.07.002
Li, Y., & Yang, T. (2018). Word Embedding for Understanding Natural Language: A Survey. In S. S (Ed.), *Guide to Big Data Applications* (Vol. 26, pp. 83–104). Springer, Cham. https://doi.org/10.1007/978-3-319-53817-4_4.
Luyi, Z., & William, S. W. (2016). LDA-TM: A two-step approach to Twitter topic data clustering. *IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2016*, 342–347. https://doi.org/10.1109/ICCCBDA.2016.7529581
Machuca, C. R., Gallardo, C., & Toasa, R. M. (2021). Twitter Sentiment Analysis on Coronavirus: Machine Learning Approach. *Journal of Physics: Conference Series, 1828* (1), Article 012104. https://doi.org/10.1088/1742-6596/1828/1/012104
MacInnis, C. C., & Hodson, G. (2021). Tensions within and between vegans and vegetarians: Meat-free motivations matter. *Appetite, 164*. https://doi.org/10.1016/j.appet.2021.105246
Markowski, K. L., & Roxburgh, S. (2019). "If I became a vegan, my family and friends would hate me:" Anticipating vegan stigma as a barrier to plant-based diets. *Appetite, 135*, 1–9. https://doi.org/10.1016/j.appet.2018.12.040
McClements, D. J., Weiss, J., Kinchla, A. J., Nolden, A. A., & Grossmann, L. (2021). Methods for Testing the Quality Attributes of Plant-Based Foods: Meat- and Processed-Meat Analogs. *Foods, 10*(2), 260. https://doi.org/10.3390/foods10020260
Meister, K. (1997). *Vegetarianism*.
Mesch, G. S., & Talmud, I. (2006). Online Friendship Formation, Communication Channels, and Social Closeness. *International Journal of Internet Science, 1*(1), 29–44. https://www.researchgate.net/publication/26495342.
Michalak, J., Zhang, X., & Jacobi, F. (2012). Vegetarian diet and mental disorders: Results from a representative community survey. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 67. https://doi.org/10.1186/1479-5868-9-67
Minson, J. A., & Monin, B. (2012). Do-Gooder Derogation. *Social Psychological and Personality Science, 3*(2), 200–207. https://doi.org/10.1177/1948550611415695
Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications, 40*(10), 4241–4251. https://doi.org/10.1016/j.eswa.2013.01.019
Naili, M., Chaibi, A. H., & ben Ghezala, H. H. (2017). Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, *112*, 340–349. https://doi.org/10.1016/j.procs.2017.08.009.
Nezlek, J. B., Cypryanska, M., & Forestell, C. A. (2021). Dietary similarity of friends and lovers: Vegetarianism, omnivorism, and personal relationships. *The Journal of Social Psychology, 161*(5), 519–525. https://doi.org/10.1080/00224545.2020.1867042
O'sullivan, P. B. (2000). What You Don't Know Won't Hurt Me: Impression Management Functions of Communication Channels in Relationships. *Human Communication Research, 26*(3), 403–431.
Park, A., Conway, M., & Chen, A. T. (2018). Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. *Computers in Human Behavior, 78*, 98–112. https://doi.org/10.1016/j.chb.2017.09.001
Paslakis, G., Richardson, C., Nöhre, M., Brähler, E., Holzapfel, C., Hilbert, A., & de Zwaan, M. (2020). Prevalence and psychopathology of vegetarians and vegans – Results from a representative survey in Germany. *Scientific Reports, 10*(1), 6840. https://doi.org/10.1038/s41598-020-63910-y
Phillips, F. (2005). Vegetarian nutrition. *British Nutrition Foundation Nutrition Bulletin, 30*, 132–167.
Plante, C. N., Rosenfeld, D. L., Plante, M., & Reysen, S. (2019). The role of social identity motivation in dietary attitudes and behaviors among vegetarians. *Appetite, 141*. https://doi.org/10.1016/j.appet.2019.05.038
Pohjolainen, P., Vinnari, M., & Jokinen, P. (2015). Consumers' perceived barriers to following a plant-based diet. *British Food Journal, 117*(3), 1150–1167. https://doi.org/10.1108/BFJ-09-2013-0252
Potts, A., & White, M. (2008). New Zealand Vegetarians: At odds with their nation. *Society and Animals, 16*(4), 336–353. https://doi.org/10.1163/156853008X357667
Raphaely, T., & Marinova, D. (2014). Flexitarianism: Decarbonising through flexible vegetarianism. *Renewable Energy, 67*, 90–96. https://doi.org/10.1016/j.renene.2013.11.030
Remer, T., Neubert, A., & Manz, F. (1999). Increased risk of iodine deficiency with vegetarian nutrition. *British Journal of Nutrition, 81*(1), 45–49. https://doi.org/10.1017/S0007114599000136
Rivera, M., & Shani, A. (2013). Attitudes and orientation toward vegetarian food in the restaurant industry. *International Journal of Contemporary Hospitality Management, 25* (7), 1049–1065. https://doi.org/10.1108/IJCHM-07-2012-0116
Rizzo, N. S., Jaceldo-Siegl, K., Sabate, J., & Fraser, G. E. (2013). Nutrient Profiles of Vegetarian and Nonvegetarian Dietary Patterns. *Journal of the Academy of Nutrition and Dietetics, 113*(12), 1610–1619. https://doi.org/10.1016/j.jand.2013.06.349
Rosenfeld, D. L. (2019). A comparison of dietarian identity profiles between vegetarians and vegans. *Food Quality and Preference, 72*, 40–44. https://doi.org/10.1016/j.foodqual.2018.09.008
Rosenfeld, D. L., & Burrow, A. L. (2018). Development and validation of the Dietarian Identity Questionnaire: Assessing self-perceptions of animal-product consumption. *Appetite, 127*, 182–194. https://doi.org/10.1016/j.appet.2018.05.003
Rosenfeld, D. L., Rothgerber, H., & Janet Tomiyama, A. (2020). From mostly vegetarian to fully vegetarian: Meat avoidance and the expression of social identity. *Food Quality and Preference, 85*, Article 103963. https://doi.org/10.1016/j.foodqual.2020.103963
Rosi, A., Mena, P., Pellegrini, N., Turroni, S., Neviani, E., Ferrocino, I., di Cagno, R., Ruini, L., Ciati, R., Angelino, D., Maddock, J., Gobbetti, M., Brighenti, F., del Rio, D., & Scazzina, F. (2017). Environmental impact of omnivorous, ovo-lacto-vegetarian, and vegan diet. *Scientific Reports, 7*(1), 6105. https://doi.org/10.1038/s41598-017-06466-8
Ruby, M. B. (2012). Vegetarianism. A blossoming field of study. *Appetite, 58*(1), 141–150. https://doi.org/10.1016/j.appet.2011.09.019
Saari, U. A., Herstatt, C., Tiwari, R., Dedehayir, O., & Mäkinen, S. J. (2021). The vegan trend and the microfoundations of institutional change: A commentary on food producers' sustainable innovation journeys in Europe. In *Trends in Food Science and Technology* (Vol. 107, pp. 161–167). Elsevier Ltd. https://doi.org/10.1016/j.tifs.2020.10.003.
Sadler, M. J. (2004). Meat alternatives — market developments and health benefits. *Trends in Food Science & Technology, 15*(5), 250–260. https://doi.org/10.1016/j.tifs.2003.09.003
Sánchez-Bravo, P., Chambers V, E., Noguera-Artiaga, L., Sendra, E., Chambers IV, E., & Carbonell-Barrachina, Á. A. (2021). Consumer understanding of sustainability concept in agricultural products. *Food Quality and Preference*, *89*, 104136. https://doi.org/10.1016/j.foodqual.2020.104136.
Scherer, L., Behrens, P., & Tukker, A. (2019). Opportunity for a dietary win-win-win in nutrition, environment, and animal welfare. *One Earth, 1*(3), 349–360. https://doi.org/10.1016/j.oneear.2019.10.020
Scherer, L., Tomasik, B., Rueda, O., & Pfister, S. (2018). Framework for integrating animal welfare into life cycle sustainability assessment. *The International Journal of Life Cycle Assessment, 23*(7), 1476–1490. https://doi.org/10.1007/s11367-017-1420-x
Shen, C., Chen, M., & Wang, C. (2019). Analyzing the trend of O2O commerce by bilingual text mining on social media. *Computers in Human Behavior, 101*, 474–483. https://doi.org/10.1016/j.chb.2018.09.031
Shi, X., Zhu, J., Cai, R., & Zhang, L. (2009). User grouping behavior in online forums. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 777–785).
Shotorbani, P. Y., Ameri, F., Kulvatunyou, B., & Ivezic, N. (2016). A hybrid method for manufacturing text mining based on document clustering and topic modeling techniques. In *IFIP advances in information and communication technology* (Vol. 488, pp. 777–786). https://doi.org/10.1007/978-3-319-51133-7_91.
Song, M., Kim, M. C., & Jeong, Y. K. (2014). Analyzing the political landscape of 2012 Korean presidential election in twitter. *IEEE Intelligent Systems, 29*(2), 18–26. https://doi.org/10.1109/MIS.2014.20
Stevens, K., Philip, K., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning* (pp. 952–961).
Stiles, B. (1998). Vegetarianism: Identity and experiences as factors in food selection. *Free Inquiry in Creative Sociology, 26*(2), 213–226.
Tao, D., Yang, P., & Feng, H. (2020). Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety, 19*(2), 875–894. https://doi.org/10.1111/1541-4337.12540
Thane, C. W., & Bates, C. J. (2000). Dietary intakes and nutrient status of vegetarian preschool children from a British national survey. *Journal of Human Nutrition and Dietetics, 13*(3), 149–162. https://doi.org/10.1046/j.1365-277x.2000.00227.x
Tleis, M., Callieris, R., & Roma, R. (2017). Segmenting the organic food market in Lebanon: An application of k-means cluster analysis. *British Food Journal, 119*(7), 1423–1441. https://doi.org/10.1108/BFJ-08-2016-0354
Tordjman, K., Grinshpan, L., Novack, L., Göen, T., Segev, D., Beacher, L., Stern, N., & Berman, T. (2016). Exposure to endocrine disrupting chemicals among residents of a rural vegetarian/vegan community. *Environment International, 97*, 68–75. https://doi.org/10.1016/j.envint.2016.10.018
Turner-McGrievy, G. M., Davidson, C. R., Wingard, E. E., Wilcox, S., & Frongillo, E. A. (2015). Comparative effectiveness of plant-based diets for weight loss: A randomized controlled trial of five different diets. *Nutrition, 31*(2), 350–358. https://doi.org/10.1016/j.nut.2014.09.002
Tziva, M., Negro, S. O., Kalfagianni, A., & Hekkert, M. P. (2020). Understanding the protein transition: The rise of plant-based meat substitutes. *Environmental Innovation and Societal Transitions, 35*, 217–231. https://doi.org/10.1016/j.eist.2019.09.004
van Loo, E. J., Caputo, V., & Lusk, J. L. (2020). Consumer preferences for farm-raised meat, lab-grown meat, and plant-based meat alternatives: Does information or brand matter? *Food Policy, 95*, Article 101931. https://doi.org/10.1016/j.foodpol.2020.101931
Wen, B., Qiao, Y., Lin, W., & Yang, J. (2017). Connecting mobility, online behavior and urban structure from cellular network data. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 1–7). https://doi.org/10.1109/PIMRC.2017.8292739
Wrenn, C. L. (2011). Resisting the globalization of speciesism: Vegan abolitionism as a site for consumer-based social change. *Journal for Critical Animal Studies, 9*(3), 9–27. https://www.wellbeingintlstudiesrepository.org/glosmov.
Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting TF-IDF term weights as making relevance decisions. *ACM Transactions on Information Systems, 26* (3), 1–37. https://doi.org/10.1145/1361684.1361686
Yang, P., Yao, Y., & Zhou, H. (2020). Leveraging global and local topic popularities for LDA-based document clustering. *IEEE Access, 8*, 24734–24745. https://doi.org/10.1109/ACCESS.2020.2969525

Yu, S.-S., Chu, S.-W., Wang, C.-M., Chan, Y.-K., & Chang, T.-C. (2018). Two improved k-means algorithms. *Applied Soft Computing, 68*, 747–755. https://doi.org/10.1016/j.asoc.2017.08.032

Zhang, C., Tao, F., Chen, X., Shen, J., Jiang, M., Sadler, B., Vanni, M., & Han, J. (2018). TaxoGen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2701–2709). https://doi.org/10.1145/3219819.3220064

Zhao, H., Liu, Z., Yao, X., & Yang, Q. (2021). A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Information Processing & Management, 58*(5), Article 102656. https://doi.org/10.1016/j.ipm.2021.102656

Zuo, Y., Zhao, J., & Xu, K. (2016). Word network topic model: A simple but general solution for short and imbalanced texts. *Knowledge and Information Systems, 48*(2), 379–398. https://doi.org/10.1007/s10115-015-0882-z