Wasted Pumpkins: A Real Halloween Horror Story

Dr Ebru SURUCU-BALCI

School of Management, University of Bradford, Bradford, UK, e.balci@bradford.ac.uk

Bercim BERBEROGLU

Manisa Celal Bayar University, Manisa, Turkey, bercim.berberoglu@cbu.edu.tr

Purpose: This study aims to understand pumpkin waste awareness among people by converting unstructured quantitative data into insightful information to understand the public's awareness of pumpkin waste during Halloween.

Design/methodology/approach: To fulfil the study's purpose, we extracted Halloween-related tweets by employing #halloween and #pumpkin hashtags and then investigated Halloween-related tweets via a topic modelling approach, specifically Latent Dirichlet Allocation (LDA). The tweets were collected from the UK between October 25th and November 7th, 2020. The analysis was completed with 11,744 tweets.

Findings: The topic modelling results revealed that people are aware of the pumpkin waste during Halloween. Furthermore, people tweet to reduce pumpkin waste by sharing recipes for using leftover pumpkins.

Originality/value: The study offers a novel approach to convert social media data into meaningful knowledge about public perception of food waste. This paper contributes to food waste literature by revealing people's awareness of pumpkin waste during Halloween using social media analytics. Norm activation model and communicative ecology theory used for the theoretical underpinning of topic modelling.

Keywords: Communicative ecology theory, Norm activation model, Food waste, Halloween, LDA Model, Public awareness, Twitter

1. Introduction

Sustainable production and consumption are of significant importance since natural resources' utilisation exceed the planetary limits (Steffen et al., 2015). This situation jeopardises meeting not only the needs of future generations but also the needs of humanity today (United Nations, 1987). Various efforts exert to overcome these challenges in several industries such as transportation, travelling, and energy, and the food industry is not exempted. Unconscious consumption of resources and excessive waste in the food sector put pressure on the planetary systems and endanger sustainable development. Each year, 1.3 billion tons of food becomes wasted (FAO, 2021). However, considering the energy, water, and minerals used to create food, it can be acknowledged that the actual loss is much higher.

Food waste refers to the decrease in the quantity or quality of food resulting from the behaviours and actions of individuals (SOFA, 2019). Food waste jeopardises sustainable production and consumption in food supply chains (FSCs) (Surucu-Balci and Tuna, 2021; Gimenez et al., 2021). To reach sustainable production and consumption in FSCs, United Nations released Sustainable Development Goals (SDGs). Specifically, SDG 12.3 sets the target about food waste as *"halve the per capita global food waste by 2030 at the retail and consumer levels and reduce food loss and waste along the production and supply chains"* (United Nations, 2015). While some of the food waste is unavoidable, such as bones, peels, and coffee grounds, some of the food waste is avoidable, meaning that all edible parts of the food, which the consumer intends typically to eat when purchasing the food but somehow ended as waste (Leverenz et al., 2019). According to the Waste and Resource Action Program (WRAP), UK households generate five million tons of avoidable food waste (WRAP, 2020). Furthermore, during celebrations such as festivals, religious celebrations, and holidays the avoidable food waste amount increases (Min et al., 2020; Aktas et al., 2018; Pool, 2012)

Halloween, which is celebrated in many countries worldwide, is no exception to these special occasions. Each year, similar news shows how millions of pumpkins are wasted during Halloween (Theguardian, 2020; 2019). In fact, 12.8 million pumpkins are wasted in the UK yearly; approximately 18,000 tons of pumpkin, which is in good condition to be consumed by people, is wasted during and after Halloween at the households (Theguardian, 2020). Environmental non-governmental organisations (eNGOs) (i.e., Hubbub) have run online campaigns on social media to raise awareness and reduce pumpkin waste amount by sharing tips and suggestions (Hubbub, 2019). Awareness is crucial while tackling food waste because awareness triggers knowledge and beliefs about food waste, resulting in specific goals and actions to reduce the waste (Martin-Rios et al., 2018; Aschemann-Witzel et al., 2017). Furthermore, awareness helps establish environmentally conscious actions, in other words, pro-environmental behaviours (Carmi, 2013).

Social media (SM) has become an inseparable part of our lives. In 2020, 3.6 billion people actively used SM and generated 2.5 quintillion data bytes per day (Bulao, 2021). The data includes the users' daily routines (i.e., sharing Instagram stories), opinions about different aspects of occasions (i.e., tweeting about recent events), working (i.e., sending emails) and entertainment activities (i.e., watching Netflix). He et al. (2018) suggested that SM data can be more reliable and trustworthy than companies marketing research data since SM data reflects people's opinions and decisions. However, 90% of the SM data is qualitative, unstructured, and includes emoticons, lexical or syntactic difficulties (Kumar et al., 2020). Therefore, social media analytics (SMA) has gained popularity to turn unstructured data into meaningful information.

Researchers have started to use SMA to understand perceptions, attitudes, and behaviours on pro-environmental behaviours, such as the impact of SM posts on the consumption behaviour of plastic products (Rapada et al., 2021), and interpret public environmental concerns about air pollution (Yang et al., 2021). However, social media analytics regarding food waste is a paucity (Jiang et al., 2021; Ventura et al., 2021). Although survey-based research exists to understand people's awareness (Richter, 2017; Qi and Roe, 2016; Principato et al., 2015), none of the studies focused on food waste awareness during special occasions. In addition, none of the studies examined the awareness by using SMA. Thus, this study aims to understand pumpkin waste awareness among people by converting unstructured quantitative data into insightful information to understand the public's awareness of pumpkin waste. To achieve this aim, we have three objectives (1) to identify shared and discussed topics among people regarding Halloween; (2) to demonstrate whether people have an awareness about the pumpkin waste that occurs during Halloween; (3) to demonstrate whether people can utilise pumpkin differently to avoid waste. The theoretical underpinning of this study was ensured by combining two theories: norm activation model (NAM) and communicative ecology theory (CET). While NAM is used to explain pro-environmental behaviours, CET is utilised to understand communication between people and groups while ensuring a holistic point-of-view (Stern and Dietz, 1994; Foth and Hearn, 2007).

In this study, we selected Twitter because it is one of the biggest SMs with more than 350 million active users who generate 6000 tweets per second (Zote, 2021). In addition, Twitter Application Programming Interface (API) provides access to tweets. After collecting tweets using identified hashtags, we employed Latent Dirichlet Allocation (LDA) to determine the topics. The topic modelling results revealed that people in the UK have a certain level of awareness about pumpkin waste during Halloween. In addition, people share suggestions to tackle pumpkin waste.

This study's structure is as follows: Section 2 provides the study's theoretical background, ensured by NAM and CET. Section 3 explains the interpretation of methodology. Section 4 reports the findings, while Section 5 discusses theoretical and practical implications, and . research limitations, and future study suggestions.

2. Theoretical Background

2.1.Norm activation model and awareness of food waste during special occasions

Food wastage occurs in the pre-consumption and consumption stages (McCarthy et al., 2020). The consumption stage includes out-of-home consumption points and households, and the pre-consumption stage includes production, manufacturing, transportation, and retail stages. The waste that occurs at the consumption stages is defined as food waste, while the waste that occurs at the pre-consumption stages is referred food loss (FAO, 2013). Food waste is mainly related to consumer behaviour, habits, and awareness, while food loss is mainly related to lack of technology investments and infrastructure systems for transportation and storing (Principato et al., 2021; Surucu-Balci and Tuna, 2021; Canali et al., 2017; Gustavsson et al., 2011). A recent WRAP report indicated that 70% of the food waste occurs at the household stage, 16% at manufacturing, 12% at out-of-home consumption points and 3% at retail in the UK (WRAP, 2021). Therefore, it is essential to focus on the reduction of food waste at the household stage.

The majority of the research found that psychological factors, situational factors, demographics and socioeconomic factors, and habits (i.e., purchasing, storing, cooking, and eating) affect the amounts of food waste (Aktas et al., 2018; Lanfranchi et al., 2016; Setti et al., 2016; Secondi et al., 2015). Although the amount of wasted food increases during special occasions (Theguardian, 2020; Pool, 2012), only Liang et al. (2021) and Aktas et al. (2018) found that people are aware of this waste. Liang et al. (2021) investigated household attitudes towards food waste in Macau and revealed that business parties, wedding banquets and bereavements events are the special occasions where the amount of food waste becomes higher. Aktas et al. (2018) examined food waste behaviour in Qatar and found that the amount of food waste increases during Ramadan (religious festival).

Norm activation model, which is used to explain pro-environmental behaviours, was proposed by Schwartz (1973). Pro-environmental behaviour is an individual or group’s action that encourages or leads to sustainable utilisation of natural resources (Ramkissoon et al., 2013). The norm activation model articulates that awareness is necessary for pro-environmental behaviour (Schwartz, 1973), meaning individuals should know about environmental issues to engage in environmental behaviours (Stern and Dietz, 1994). Furthermore, the norm activation model infers that individuals will be encouraged to engage in an environmental dilemma if they believe that significant problems require immediate solutions and that such solutions are heavily reliant on their cooperative behaviours and decisions (Adel et al., 2021).

Food waste awareness is referred to knowledge and beliefs about food waste issues that lead to specific goals and actions (Martin-Rios et al., 2018). Chen (2019) stated that individuals' food waste awareness influences their behaviour and effort to reduce household food waste. Engaging in activities that reduce food waste is regarded as pro-environmental activity (de Groot et al., 2021). Following the norm activation model, to reduce food waste (behave pro-environmentally), people need to be aware of the impact and consequences of food waste because awareness of the consequences of food waste can affect the motivation to act and waste less (van Geffen et al., 2020). Therefore, raising awareness of food waste during special occasions can diminish food waste during this time.

2.2.Communicative ecology theory and Halloween-related tweets

Communicative ecology is a conceptual framework in the media and communication field to interpret the dynamic relationship among technology, content, and social factors (Jin et al., 2019). Communicative ecology theory (CET) adopts a holistic point-of-view which ensures understanding the communication within the group or between different groups without focusing solely on a person or single communication channel. The theory suggests that communication behaviours are a result of the combination of three different but interrelated layers of the communicative ecology; the technological layer, discursive (content) layer, and people (social) layer (Forth and Hearn, 2007). The technological layer includes information technology and devices which connect people and enable communication. The discursive layer is the content that involves communication themes or ideas. The social layer consists of people and communication process structures such as social networks and community organisations.

In the last decade, SM and sharing content, which includes people's daily activities, life happenings, and feelings, on SM became people's routines (He et al., 2018; Guo et al., 2017; Seol et al., 2016). For instance, 500 million tweets are sent daily (Zote, 2021). Furthermore, O'Shea (2018) and Hutchinson (2016) indicated that the tendency to share content increases during holidays such as Christmas, Halloween, and Easter. Although Halloween's origin is based on religion, it is currently celebrated globally by people from different backgrounds on October 31st. Halloween has its traditions, such as trick-or-treating, attending costume parties, and carving pumpkins. While people do what is necessary for these traditions, they frequently post on their SM. Thus, each year during Halloween, the number of Halloween-related posts increases.

In this study, Twitter is chosen as a technological layer since it is the most commonly used micro-blogging site with 350 million users. As one of the most popular micro-blogging sites, Twitter allows users to share messages and information with 280 characters or less. Gathering almost real-time posts via API makes Twitter attractive for collecting data for identifying themes and understanding public opinion. Thus, we utilise Twitter as the primary data source for this study. The hashtag is identified as the social layer in this current research. Hashtags act as a tool to disseminate specific information to related parties. To improve visibility and reach related people, users can utilise more than one hashtags. Since we will identify Halloween-related themes and topics discussed during Halloween, we have decided to focus on *#halloween* and *#pumpkin* hashtags while collecting data.

User-generated contents (UGCs) consist of the discursive layer of this study. On SM, the communication proceeds on UGCs, taking various forms such as tweets, Facebook updates, YouTube videos, or product reviews on e-commerce sites. These UGCs contain essential clues about the users since they share their daily routines, perceptions, and understandings. However, UGCs are considered unstructured data, which means that they are qualitative and include noises such as emoticons. To extract meaningful information, UGCs require pre-treatment, such as removing emoticons or removing unnecessary words before obtaining meaningful data. Therefore, we need to collect Halloween-related tweets by utilising pre-determined hashtags to reach the study purpose. Then we need to implement pre-processing stage on the collected tweets. After these stages, we can analyse the data, identify Halloween-related themes and topics, and understand whether people are aware of the pumpkin waste during Halloween. Further information about the methodology is provided in the next section.

3. Research Methodology

The research process of this study includes four steps which are presented in Figure 1. First, we collected tweets related to Halloween by employing pre-determined hashtags. Then to remove unnecessary noise, we pre-processed the tweets by removing unnecessary noise. We employed LDA to find topics. Lastly, we identified topics. Each step is further explained below.

--Figure 1 insert here--

3.1.Data collection

Twitter generates massive data with about 500 million tweets each day that correspond to nearly 6,000 tweets every second (Zote, 2021). Since it is challenging to analyse all data, a prevalent way for the scholars is to specify the data collecting stage with specific parameters such as search terms, location, language, and period (Sivarajah et al., 2017). We utilise Twitter search streaming API to retrieve publicly available Twitter data in this research. To obtain relevant tweets, we fed our algorithm using various parameters that enabled us not to deal with unrelated sharing to our focal point. Moreover, using parameters such as text query and geocode, we improved the potential of the representation accuracy (Stieglitz et al., 2018).

The data collection process started with finding the most appropriate keywords to capture our research topic. In this respect, we determined the most relevant keywords as *#halloween* and *#pumpkin*. We utilised geocode to specify the tweets' area since we restricted our research to the UK. We filtered retweets in order to obtain unique tweets. Lastly, we adjusted the algorithm only to collect tweets in the English language. As a result, we obtain unstructured Halloween-related tweets sent by users residing in the UK. The data was collected between October 25th and November 7th, 2020, which corresponds to one week before and after Halloween. A total of 11,744 tweets were collected for this period. The collected tweets included the features such as coordinates, hashtags, user names, URLs, retweets, favourites and followers count, screen name, and many others besides the tweets. Thus, before analysing, we had to apply pre-processing steps to prepare raw data for extracting meaningful information, which is explained next.

3.2.Pre-processing

Data cleansing practices should be carefully executed to ensure the textual data analysis quality because the collected data contains noises (Chae, 2015). As there is no optimal pipeline for pre-processing, scholars utilise some typical applications and heuristics for this step (Singh et al., 2018). We applied four pre-processing operations to prepare unstructured Twitter data for analysis: data cleansing, removing stop words, tokenisation, and stemming.

The data cleansing step includes converting lower case, filtering, removing punctuations, numbers, URLs, user names, and emoticons; however, we excluded the stop words such as "a," "the," "with" so that these words do not affect the meanings of the text (Singh et al., 2018). Second, to remove stop words, we adopted the stop words of NLTK (Bird et al., 2009), which is Python's one of the most utilised libraries' for natural language processing tasks. Then we expanded the stop words using GENSIM (Rehurek and Sojka, 2010) library's stop words list. Third, we split text data into meaningful elements called tokens during the tokenisation step (Xiang et al., 2017). Lastly, in the stemming step, we transformed words to their roots. For instance, the words such as "celebrating," “celebrates”, and “celebrated” turned into “celebrate”. Once the data cleansing process was completed, we continued with topic modelling.

3.3.Topic Modelling

LDA was first introduced by Blei, Ng, and Jordan in 2003 and has become one of the most commonly utilised methods in topic modelling since then. LDA is a generative probabilistic topic model used to examine the most salient topics in textual data. Topic models are unsupervised machine-learning algorithms to reveal the latent structures in documents. Thus, labelling or annotating data is unnecessary (Syed and Spruit, 2017). By considering the statistical properties of the documents, LDA allows discovering the underlying topics of the large bodies of unstructured data. The patterns, themes of the social media posts, and interconnections between the themes can effectively be investigated. Recently, LDA has received increasing attention in social sciences, and it is utilised for various applications, including tourist satisfaction analysis (Guo et al., 2017), detecting people's discourse and psychological reactions to the Covid-19 pandemic (Xue et al., 2020), investigating the impact of natural disasters (Zhou et al., 2021), discovering the main themes of researches (Zhou et al., 2022), online accommodation reviews (Sutherland et al., 2020), and household waste management (Jiang et al., 2021).

This study employed the LDA model to discover the most salient themes during Halloween. In this respect, we captured the tweets' underlying topics during Halloween, the importance of each topic, and dominant words belonging to these topics. Our choice of the LDA model instead of other text analysis methods in the literature lay on several grounds. The mechanism of LDA is as follows (Blei et al., 2003). Each document is considered a collection of words. The document is symbolised as a combination of the hidden topics distinguished by a distribution over words.

In this study, words represent the elements people utilise to show their emotions and opinions via tweets, while documents (d) symbolise the tweets. Furthermore, tweet collections establish corpus, and LDA is the generative probabilistic corpus model, including the tweets (M) formed as a random proportion over K hidden topics.

Taking into account the explanations, the LDA model assumes the following generative process for each document d including N_d words (d ∈ 1, …, M) (Jelodar et al., 2019; Blei et al., 2003):

(1) Choose a multinomial distribution φ t for topic t (t ∈ {1, …, T }) from a Dirichlet distribution with parameter β,

(2) Choose a multinomial distribution θd for document d (d ∈ {1, … , M }) from a Dirichlet distribution with parameter α,

(3) For a word w n (n ∈{1,...,Nd}) in document d,

a. Choose a topic z_n from θ_d.

b. Choose a word w_n from φ_{zn}.

The probability of a corpus is obtained as follows:

$$p(D|\alpha,\beta)=\prod_{d=1}^{M}\int p(\theta_d|\,\alpha)\left(\prod_{n=1}^{N_d}\sum_{z_{dn}} p(z_{dn}|\theta_d)p(w_{dn}|z_{dn},\beta)\right)\mathrm{d}\theta_d$$

In the formula, w represents observable variables while others mean latent variables (φ and θ). Alongside being hyperparameters, α and β can be derived by Gibbs sampling, Expectation-maximization, or Variational Bayes inference method (Jelodar et al., 2019; Guo et al., 2017). The variables θ_d are document-level variables accepted to be sampled once per tweet. The variables z_{dn} , w_{dn} are word-level variables and sampled once for each word in each tweet.

4. Results

4.1.Descriptive results

After implementing pre-processing steps, we reach 11,744 unique tweets, consisting of 266,558 words in total for October 25th and November 7th, 2020. The days between October 25th and October 30th represent the earlier week of Halloween, while days between November 1st and November 7th show the week after Halloween. Figure 2 illustrates the number of tweets by date.

--Figure 2 insert here--

In general, preparations for Halloween start one week in advance, and as October 31 approaches, preparations intensify, and when Halloween is over, the preparations made will be removed. This situation is also prominent in the number of twits. Throughout the first week, the number of tweets increased gradually. A total of 9850 tweets were posted in the first week. The number of tweets reached its highest point on Halloween -October 31st-, with 3009 tweets on a single day. After Halloween, the number of tweets decreased sharply. However, collecting the data from the week after Halloween is essential because this period can provide insight into how the leftover pumpkins are utilised. A total of 1895 tweets were posted in the second week.

Figure 3 shows the most common ten words in collected tweets. The most pronounced words are pumpkin (n=12,603) and Halloween (n=11,976). This result is expected because we employed these two words as hashtags while collecting Halloween-related tweets. They are followed by the following words as respectively; carving (n=5,785), happy (n=4,663), year (n=3,698), like (n=1,973), today (n=1,185), love (n=992), make (n=601), and great (n=599).

--Figure 3 insert here--

The "Carving" word is the most used third word in the tweets. This result is expected because pumpkin carving is one of the most applied Halloween traditions. The other frequently used words are "happy", "like", "love", and "great", which reflect the jovial ambience of the Halloween period. As a result of the examination of the tweets, it can be said that keywords such as “year” and “today”, which are among the high-frequency words, express that this year's or today's Halloween will be different due to the pandemic. We associate the “make” word, with the re-evaluation of wasted pumpkins and people’s use for different purposes behaviour.

4.2. Topic Modelling Results

We utilised the LDA model to extract the most salient themes from tweets that were sent from the UK during the Halloween period. We used the hyperparameters in their standard configurations in constructing the model except for the topic number. Before reaching the ultimate LDA model, it is crucial to determine the optimal number of topics. Several metrics can be used in literature to determine the optimal topic number, such as coherence score (Xue et al., 2020), perplexity (Jiang et al., 2021), and log-likelihood scores. We chose the appropriate topic number according to the coherence score. While seeking the optimal number of topics, we restricted topic numbers between two and fifty and used the LDA model's default configurations. Figure 4 shows the coherence scores and the topic numbers.

--Figure 4 insert here--

As a result, we reached the optimal topic number with eleven topics with the highest coherence score. Namely, the model with eleven topics achieves the maximum coherence score of around

0.5, demonstrating that the determined model is more appropriate than the other models with different topic numbers.

Figure 5 demonstrates the total probability of the top 100 words in each topic. Each line represents the 11 topics selected based on the coherence score in the figure. As seen, the weights fall very sharply as the rank of the most important words decrease. This sharp weight decrease implies that the words up to rank around 20 have greater importance than the remaining ones. These words can be seen in word clouds.

--Figure 5 insert here--

Figure 6 illustrates the total probability of the top ten words in each topic. The top ten words account mainly for Topic 1 and Topic 2 with nearly 20% probability and explain the other topics mostly around 5%. We interpret that these words explain the particular proportion of their topic's total probability ranging from nearly 25% to 5%, considering the LDA model. Although each topic consists of more than ten words, we used the top ten words to identify and name the topics.

--Figure 6 insert here--

In this study, 11 topics were obtained using the LDA method. Table 1 shows the 11 topics according to their descending topic weights order. Topic weights imply the distribution of words with the highest proportion in the related topic. Thus, they illustrate the importance of the captured topics by the LDA model. Visualisation of each topic is ensured via word clouds, including keywords with high probabilities based on their proportion of probability in the topic.

The topics were manually named and classified under the themes based on human judgment (Guo et al., 2016). With the utilisation of the top 10 words, we named the topics. Two people completed the naming procedure; one is the researcher of this study, and the other is a different researcher who did not know the research objectives. Once each researcher completed the naming, they came together and compared the name of the topics. If a disagreement occurred while revealing the topic names, researchers discussed their topic names and finalised the topic name once a consensus was reached. Researchers disagreed on two (topic nine and topic three) out of eleven topics and finalised the two topics' names after reaching a consensus. Furthermore, we checked from original tweets that contained these top 10 words to verify whether the topic names mirror the original meanings of tweets.

--Table 1 insert here--

After finalising topic labelling, we developed themes by bringing similar topics together. Accordingly, four themes emerged: Halloween tradition, food waste, food waste minimisation, and miscellaneous. Halloween tradition theme includes *pumpkin carving contest* (topic 1), *Halloween celebration* (topic 9), *trick or treat in Covid-19* (topic 4), *winning carving contest* (topic 7), and *pumpkin carving kit* (topic 8). The food waste theme consists of *pumpkin waste* (topic 2) and *resident food loss* (topic 5). The food waste minimisation theme involves *recipes for pumpkin leftovers* (topic 11), *pumpkin ornaments* (topic 10), and *re-use of carving leftovers* (topic 6). Lastly, the miscellaneous theme includes *online communication* (topic 3).

The *pumpkin carving contest* has the highest weight of all topics among the identified topics (0.9547). Statistically, this means that people's primary interest in Halloween was related to pumpkin carving. *Pumpkin waste* is the second-ranked topic on the list (0.9336). Although other Halloween tradition-related topics still occur in the list, the appearance of the food waste topic as second-ranked shows that people are aware of the pumpkin waste that occurs each year. Furthermore, this awareness secured its position with the third-ranked topic because the third-highest weighted topic is also about food waste and named *resident food loss,* which implies the food loss at the personal level during Halloween (0.9228).

Halloween celebration is the fourth-ranked topic and represents the positive feelings and festive times about Halloween (0.9202). *Recipes for pumpkin leftovers* have the fifth-highest weight (0.9128). Having such a topic on the list pinpoints that people are not only aware of the pumpkin waste but also tackling the waste. Specifically, users share recipes for soups and cakes. Moreover, according to results, sharing recipes is not the only way to deal with pumpkin waste. The topic which has the sixth-highest weight is *pumpkin ornaments* (0.9125). This topic demonstrates that users share tips and clues for ornaments such as pumpkin flower arrangements used for decorating homes during Halloween alongside food recipes.

Trick or treat in Covid-19 is another Halloween tradition-related topic on the seventh-ranked topic (0.9093). Since trick or treat is a typical Halloween custom for children, such a topic is expected to appear. Nevertheless, this topic also implies that people adjusted to trick-or-treat tradition due to Covid-19 restrictions and social distancing measures. Instead of trick or treat, a pumpkin spotting game was played to give candies to children after they spotted pumpkin in the neighbourhood. *Online communication* (no. 3) is ranked eighth (0.9011). The *winning carving contest* is the ninth topic (0.8945), another Halloween tradition theme. *Re-use of carving leftovers,* ranked tenth, is another topic related to food waste minimisation (0.8668). The *pumpkin carving kit* is the last identified topic in the Halloween tradition theme (0.8652).

Since pumpkin carving is one of Halloween's most essential activities, tools used for this purpose can gain attention.

5. Conclusion and discussion

Food waste, which is a threat to achieving United Nations' SDG target, is mainly based on the behaviours and habits of individuals (Canali et al., 2017; Gustavsson et al., 2011). Literature states that the amount of avoidable food waste increases during special occasions (Theguardian, 2020; Pool, 2012). Therefore, to tackle avoidable food waste during special occasions, we need to understand whether the individuals are aware of the food waste because awareness enables acts that lead to change. It is almost impossible not to see a food-related post while scrolling down on social media. Zhang et al. (2019) stated that SM is considered a creative and tempting instrument for gathering and sharing information about food-related issues. Parallel to the increase in food-related content in SM, the utilisation of SMA to understand people's awareness, understanding, and opinions have gained momentum in recent years (Sutherland et al., 2020; Xue et al., 2020; Guo et al., 2017). However, using SMA to understand people's awareness about food waste is a paucity (Jiang et al., 2021; Ventura et al., 2021). Thus, this study aims to fulfil this gap by investigating the awareness of pumpkin waste during Halloween using SMA.

The theoretical underpinning of the study was ensured by using NAM and CET. NAM is used to explain pro-environmental behaviour, while CET is used to understand the dynamic relationship between technology, context, and social layers. We collected 11,744 tweets from the UK between October 25th and November 7th, 2020. Collected tweets were scrutinised via the LDA model. The results revealed that people are aware of pumpkin waste during Halloween. The emergence of pumpkin waste (topic 2) and resident food loss (topic 5) shows a certain level of awareness of pumpkin waste in society. Furthermore, this study also showed that people share ideas to tackle pumpkin waste. LDA model revealed that different tackling methods are utilised, such as sharing recipes for pumpkin leftovers (topic 11), making pumpkin ornaments (topic 10) and re-using of carving leftovers (topic 6).

5.1. Implications for practice

Although results reveal that pumpkin waste awareness occurs among people, the precaution and tackling methods are not enough to reduce the pumpkin waste. Therefore, there is still a need to increase awareness among people. Knowing that the environment will be affected negatively due to an individual's behaviour leads people to act pro-environmental behaviour

(Adel et al., 2021). For this reason, it is necessary to share and increase the number of posts showing how the pumpkin waste affects the environment on SM.

It was revealed that the leftover pumpkins are most frequently re-used by making food. However, awareness should be increased regarding the use of leftovers for different do-it-yourself. SM can be used to disseminate information and run campaigns (Han and Cheng, 2020). Campaigns can be run to encourage people to use pumpkin leftovers in do-it-yourself projects such as soap and candles. Many hashtags can be created for these campaigns, and recipes can be disseminated through hashtags. Leftover pumpkins can not only be used for baking but also for making cocktails or mocktails. Therefore, similar to sharing food recipes, cocktails and mocktail recipes can be shared on SM.

Having contests on Halloween is a tradition. The results of this study also approved this situation in topic 1 (pumpkin carving contest), topic 7 (pumpkin ornaments) and topic 8 (pumpkin carving kit). Like the pumpkin carving contest, baking contests in which leftover pumpkins are used can be held in the regions, and people are encouraged to utilise the leftover pumpkin in this way.

Without adopting circular economy principles in society and supply chains, we cannot have a waste-free Halloween. Therefore, no matter how hard we try to reduce it, waste will be generated. Nevertheless, how we dispose of the waste impacts the environment (Papargyropoulou et al., 2014). The current study did not reveal the disposal habits of leftover pumpkins. However, we can infer that leftover pumpkins were sent to landfills, where leftover pumpkins produce methane gas, which is 25 times more hazardous than CO_2 (WorldEconomicForum,2019; EPA, 2016). This fact shows that the disposal type of pumpkin is also essential. People need to be directed to more environmentally friendly ways to eliminate this while throwing away the carved pumpkin. Therefore, campaigns need to be run to pivot people to use leftover pumpkins for animal feeds, industrial uses or composting.

5.2.Implications for theory

This study makes several contributions to the literature. First of all, this is the first study focused solely on understanding the people's perception of food waste during a specific occasion, Halloween. Although literature states that the amount of food waste increases during special occasions, none of the previous studies focused solely on special occasions. Second, the study contributes to the literature by revealing people's food waste awareness using social media analytics. Earlier studies evaluated people's awareness about food waste via questionnaires

(Richter, 2017; Qi and Roe, 2016; Principato et al., 2015). Nevertheless, this study shows that we can understand people's awareness about food waste by utilising LDA.

Third, this is one of the first attempts to underpin topic modelling with the utilisation of theory. Earlier studies, which employed LDA, did not underpin their research via any theory (Jiang et al., 2021; Han et al., 2020; Samuel et al., 2020; Xue et al., 2020). To the best of our knowledge, this is the first study to use NAM and CET theory together. While NAM is used to explain pro-environmental behaviour, CET is utilised to understand communication between people and groups while ensuring a holistic point-of-view.

5.3.Limitations and future research

This study is subject to several limitations. First, the data were collected using two hashtags. A future study can be conducted by using more related hashtags. Second, the data collection period was short. A future study can be conducted by collecting yearly data to understand whether the awareness of loss has changed over the years and see what attempts have been made to reduce the pumpkin loss. Third, this study only focused on the British people's awareness about pumpkin waste during Halloween, yet Halloween is celebrated worldwide. Therefore, a similar study can be conducted by collecting data from different countries to reveal worldwide awareness and demonstrate the similarities and differences among people's awareness.

Although food waste increases during special occasions, this study only focused on Halloween. Therefore, future studies can be conducted for other special occasions like Thanksgiving, Easter or Ramadan by adopting a similar approach. Lastly, although LDA topic modelling is based on data and mathematical implementations, the topics were determined using human judgments subject to bias (Gandomi and Haider, 2015). Therefore, a future study can focus on this limitation and use advanced machine learning while determining topics.

References:

Adel, A.M., Dai, X. and Roshdy, R.S. (2021). Investigating consumers' behavioural intentions toward suboptimal produce: An extended theory of planned behaviour – a cross-cultural study. *British Food Journal.*

Aktas, E., Sahin, H., Toplaoglu, Z., Oledinma, A., Huda, A.K.S, Irani, Z., Sharif, A.M., Wout, T.and Kamrava, M. (2018). A consumer behavioural approach to food waste. *Journal of Enterprise Information Management,* 31(5): 658-673.

Aschemann-Witzel, J., de Hooge, I.E., Rohm, H., Normann, A., Bossle, M.B., Gronhoj, A. and Oostindjer, M. (2017). Key characteristics and success factors of supply chain initiatives tackling consumer-related food waste- A multiple case study. *Journal of Cleaner Production,* 155: 33-45.

Bamberg, S., Hunecke, M. and Blöbaum, A. (2007). Social context, personal norms and the use of public transportation: Two field studies. *Journal of Environmental Psychology,* 23(3): 190-203.

Bird, S., Edward L and Ewan K. (2009). *Natural Language Processing with Python.* O'Reilly Media Inc.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research,* 3, 993-1022.

Bulao, J. (November 1, 2021). How much data is created every day in 2021? https://techjury.net/blog/how-much-data-is-created-every-day/#gref (19/11/2021)

Canali, M., Amani, P., Aramyan, L., Gheoldus, M., Moates, G., Östergren, K., Silvennoinen, K., Waldron, K. and Vittuari, M. (2017). Food waste drivers in Europe, from identification to possible interventions. *Sustainability,* 9(1):37.

Carmi, N. (2013). Caring about tomorrow: Future orientation, environmental attitudes and behaviours. *Environmental Education Research,* 19(4): 430-444.

Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247-259.

Chen, H. S. (2019). Environmental concerns and food consumption: What drives consumers' Actions to reduce food waste? *Journal of International Food & Agribusiness Marketing*, 31(3), 273–292

D'Ambrosi, L. (2018). Pilot study on food sharing and social media in Italy. *British Food Journal,* 120 (5): 1046-1058.

de Groot, J. I., Bondy, K., & Schuitema, G. (2021). Listen to others or yourself? The role of personal norms on the effectiveness of social norm interventions to change pro-environmental behaviour. *Journal of Environmental Psychology*, 78, 101688.

EPA (United States Environmental Protection Agency). (2016). Overview of Greenhouse Gases. https://www.epa.gov/ghgemissions/overview-greenhouse-gases#methane (Accessed Date: 12/12/2021)

FAO. (2013). Food wastage footprint: Impacts on natural resources. https://www.fao.org/3/i3347e/i3347e.pdf (23 November 2021)

FAO. (2021). Food wastage: Key facts and figures. http://www.fao.org/news/story/pt/item/196402/icode/#:~:text=FAO's%20new%20report%20is%20the,amounts%20to%201.3%20billion%20tonnes (accessed date April 15, 2021)

Foth, M. and Hearn, G. (2007). Networked individualism of urban residents: Discovering the communicative ecology in inner-city apartment building. *Information, Communication & Society,* 10 (5): 749-772.

Gandomi, A. and Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics, *International Journal of Information Management*, 35 (2):137–144, https://doi.org/10.1016/j.ijinfomgt.

Gimenez, A., Aschemann-Witzel, J. and Ares, G. (2021). Exploring barriers to consuming suboptimal foods: A consumer perspective. *Food Research International,* 14: 110106.

Guo, I., Vargo, C.J., Pan, Z., Ding, W. and Ishwar, P. (2016). Big social data analytics in journalism and mass communication. *Journal of Mass Communication,* 93(2): 322-359.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467-483.

Gustavsson, J., Cederberg, C. and Sonesson, U. (2011). Global food losses and food waste. https://www.madr.ro/docs/ind-alimentara/risipa_alimentara/presentation_food_waste.pdf (23 November 2021)

Han, X., Wang, J., Zhang, M. and Wang, X. (2020). Using social media to mine and analyse public opinion related to Covid-19 in China. *International Journal of Environmental Research and Public Health,* 17: 1-22.

He, W., Tian, X., Hung, A., Akula, V. and Zhang, W. (2018). Measuring and comparing service quality metrics through social media analytics: a case study. *Information Systems E-Business Management,* 16 (3), 579–600.

Hubbub. (2019). Join the pumpkin rescue. https://www.hubbub.org.uk/pumpkin-rescue (accessed date Aril 15, 2021)

Hutchinson, A. (October 15, 2016). Facebook releases new data on holiday season trends and usage patterns. https://www.socialmediatoday.com/social-networks/facebook-releases-new-data-holiday-season-trends-and-usage-patterns-infographic (accessed date April 15, 2021)

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modelling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211.

Jiang, P., Fan, Y.V. and Klemes, J.J. (2021). Data analytics of social media publicity to enhance household waste management, *Resources, Conservation & Recycling,* 164: 105146.

Jin, X.L., Zhou, Z. and Yu, X. (2019). Predicting users' willingness to diffuse healthcare knowledge in social media: A communicative ecology perspective. *Information Technology & People,* 32(4): 1044-1064.

Kumar, A., Dabas, V. and Hooda, P. (2020). Text classification algorithms for mining unstructured data: a Swot analysis. *International Journal of Information Technology,* 12: 1159-1169.

Lanfranchi, M., Calabro, G., De Pascale, A., Fazio, A. and Giannetto, C. (2016). Household food waste and eating behavior: Empirical survey. *British Food Journal,* 118(12): 3059-3072.

Leverenz, D., Moussawel, S., Maurer, C., Hafner, G., Schneider, F., Schmidt, T. and Kranert, M. (2019). Quantifying the prevention potential of avoidable food waste in households using a self-reporting approach. *Resources, Conservation & Recycling,* 150: 104417.

Liang, Y., Song, Q., Liu, G. and Li, J. (2021). Uncovering residents and restaurants' attitude and willingness toward effective food waste management: A case study. *Waste Management,* 130: 107-116.

Martin-Rios, C., Demen-Meier, C., Gössling, S., and Cornuz, C. (2018). Food waste management innovations in the foodservice industry. *Waste Management,* 79: 196-206.

McCarthy, B., Kapetanaki, A.B. and Wang, P. (2020). Completing the food waste management loop: Is there market potential for value-added surplus products (VASP)? *Journal of Cleaner Production,* 256: 120435.

Min, S., Wang, X. and Yu, X. (2020). Does dietary knowledge affect household food waste in the developing economy of China? *Food Policy,* article in press.

O'Shea, D. (November 21, 2018). Retailers to see 75% increase in social media during the holidays. https://www.retaildive.com/news/retailers-to-see-75-increase-in-social-media-during-the-holidays/542777/ (accessed date April 15, 2021)

Papargyropoulou, E., Lozano, R. Steinberger, J.K., Wright, N. and bin Ujang, Z. (2014). The food waste hierarchy as a framework for the management of food surplus and food waste. *Journal of Cleaner Production,* 76: 106-115.

Pool, E. (January 2012). The nightmare after Christmas. https://ieeexplore-ieee-org.brad.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=6156550 (accessed date April 15, 2021)

Principato, L., Secondi, L. and Pratesi, C.A. (2015). Reducing food waste: an investigation on the behaviour of Italian youths. *British Food Journal,* 117(2): 731-748.

Qi, D. and Roe, B. (2016). Household food waste: multivariate regression and principal component analyses of awareness and attitudes among U.S. consumers. *PLoS One,* 11(7): 1-19.

Ramkissoon, H., Smith, L. D. G., & Weiler, B. (2013). Relationships between place attachment, place satisfaction and pro environmental behaviour in an Australian national park. *Journal of Sustainable Tourism*, 21(3), 434–457.

Rapada, M.Z., Yu, D.E. and Yu, K.D. (2021). Do social media posts influence consumption behavior towards plastic pollution? *Sustainability,* 13(22), 12334.

Rehurek, R., and Sojka, P. (2010). Software framework for topic modelling with large corpora. In In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks.

Richter, B. (2017). Knowledge and perception of food waste among German consumers. *Journal of Cleaner Production,* 166: 641-648.

Samuel, J., Ali, G.G.N., Rahman, M., Esawi, E. and Samuel, Y. (2020). Covid-19 public sentiment analysis insights and machine learning for Tweets classification. *Information,* 11: 1-22.

Schwartz, S.H. (1973). Normative explanations of helping behaviour: A critique, proposal and empirical test. *Journal of Experimental Social Psychology,* 9(4): 349-364.

Schwartz, S.H. and Howard, J.A. (1981). A normative decision-making model of altruism. In: Rushton, J.P., Sorrentino, R.M. (Eds.), Altruism and Helping Behaviour. Erlbaum, Hillsdale, pp. 89-211.

Secondi, L., Principato, L. and Laureti, T. (2015). Household food waste behaviour in EU-27 countries: A multilevel analysis. *Food Policy,* 56: 25-40.

Seol, S., Lee, H., Yu, J. and Zo, H. (2016). Continuance usage of corporate SNS pages: A communicative ecology perspective. *Information & Management,* 53: 740-751.

Setti, M., Falasconi, L., Segre, A., Cusano, I. and Vittuari, M. (2016). Italian consumers' income and food waste behaviour. *British Food Journal,* 118(7): 1731-1746.

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research,* 70:263-286.

Singh, A., Shukla, N. and Mishra, N. (2018). Social media data analytics to improve supply chain management in food industries. *Transportation Research Part E*, 114: 398-415.

Steffen, W., Richardson, K., Rockstrom, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., carpenter, S.R., Vries de, W., Wit de, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, K.M., Ramanathan, V., Reyers, B. and Sorlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science,* 347, 736.

Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection and data preparation. *Int. J. of Inf. Mgnmt,* 39: 156-168.

Surucu-Balci, E. and Tuna, O. (2021). Investigating logistics-related food loss drivers: A study on fresh fruit and vegetable supply chain. *Journal of Cleaner Production,* 318: 128561.

Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via latent dirichlet allocation. *Sustainability*, 12(5), 1821.

Syed, S. and Spruit, M. (October, 2017). Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation. In 2017 IEEE International conference on data science and advanced analytics (DSAA) (pp. 165-174). IEEE.

The Sate of Food and Agriculture (SOFA). (2019). Moving forward on food loss and waste reduction. http://www.fao.org/3/ca6030en/ca6030en.pdf (accessed date April 15, 2021)

theguardian. (October 23, 2019). Pumpkin waste in UK predicted to hit scary heights this Halloween. https://www.theguardian.com/environment/2019/oct/23/pumpkin-waste-uk-halloween-lanterns (accessed date April 15, 2021).

theguardian. (October 8, 2020). Over half UK's 24m Halloween pumpkins destined for food waste. https://www.theguardian.com/environment/2020/oct/08/over-half-uks-24m-halloween-pumpkins-destined-for-food-waste (accessed date April 15, 2021).

United Nations. (1987). Our Common Future. Report of the UN World Commission on Environment and Development. Annex to document A/42/427. http://www.undocuments.net/our-common-future.pdf (accessed April 15, 2021).

United Nations, 2015. Transforming Our World: the 2030 Agenda for Sustainable Development. A/RES/70/1. New York. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E (accessed 18/12/2021).

van Geffen, L., van Herpen, E., Sijtsema, S., & van Trijp, H. (2020). Food waste as the consequence of competing motivations, lack of opportunities, and insufficient abilities. *Resources, Conservation & Recycling:* X, 5, 100026.

Ventura, V., Cavaliere, A. and Ianno, B. (2021). #Socialfood: Virtuous or vicious? A systematic review. *Trends in Food Science & Technology,* 110: 647-686.

Wang, S., Wang, J., Zhao, S. and Yang, S. (2019). Information publicity and resident's waste separation behavior: An empirical study based on the norm activation model. *Waste Management,* 87: 33-42.

WorldEconomic Forum (October 30, 2019). Are Halloween pumpkins a problem for the planet? https://www.weforum.org/agenda/2019/10/halloween-pumpkins-food-waste-energy/ (Accessed date 24/12/2021)

WRAP. (2020). UK progress against Courtauld 2025 targets and UN sustainable development goal 12.3. https://wrap.org.uk/sites/files/wrap/Progress_against_Courtauld_2025_targets_and_UN_SDG_123.pdf (accessed date April 15, 2021)

WRAP. (2021). Food surplus and waste in the UK-key facts. https://wrap.org.uk/sites/default/files/2021-10/food-%20surplus-and-%20waste-in-the-%20uk-key-facts-oct-21.pdf (23 November 2021)

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management,* 58: 51-65.

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., and Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modelling on Twitter. *PloS one*, 15(9), e0239441.

Zhang, C., Yue, Z., Zhou, Q., Ma, S. and Zhang, Z.K. (2019). Using social media to explore regional cuisine preferences in China. *Online Information Review,* 43(7): 1098-1114.

Zhou, S., Kan, P., Huang, Q. and Silbernagel, J. (2021). A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura. *Journal of Information Science*, 01655515211007724.

Zhou, H., Yip, W. S., Ren, J., & To, S. (2022). Thematic analysis of sustainable ultra-precision machining by using text mining and unsupervised learning method. *Journal of Manufacturing Systems*, 62, 218-233.

Zote, J. (March 11, 2021). 12 essential Twitter stats to guide your strategy in 2021. https://sproutsocial.com/insights/twitter-statistics/ (03 December 2021)