Text mining of social media data for enhancing food safety of farmer's market.

Dandan Tao, Hao Feng*

Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, USA

Abstract

This study was conducted to analyze consumers' reviews and/or comments posted on social media to recognize potential food safety issues associated with farmer's markets. Text mining models were built using data from Yelp and Twitter to automatically identify consumers' responses on food safety after visiting a farmer's market. Besides food safety, other aspects such as quality, availability, and environment were also considered in data analysis models. Machine learning tools, including Naïve Bayes (NB), Support Vector Machines (SVM), Logistic Regressions (LR), k-Nearest Neighbor (k-NN), and Random Forests (RF), were used to build the models, and SVM was identified as the optimal model with highest F-1 score, a parameter for evaluating correct identification, on both Twitter (0.68) data and Yelp data (0.75). Based on the SVM models, the most important words used in the classification were identified. For the topic "safety", the important words identified from two datasets were different. On Twitter, words like 'safety', 'coli', 'health', 'recall', 'illness', 'train', 'tip', 'foodborne' were commonly mentioned, indicating people intend to talk about issues of foodborne outbreaks. On Yelp, people tend to comment on the hygiene conditions of a farmer's market with words like 'clean', 'messy', 'safety', 'gross', 'rotten'. The findings could help local public health departments know about the hygiene status or potential food safety issues of a farmer's market based on consumer reviews.

Keywords: Text mining, Food safety, Epidemiology, Social media, Classification models, Consumer responses, Hygiene.

Accepted on May 29, 2021

Introduction

Farmer's markets (FMs) are growing as significant venues for the sale of local produces such as fresh fruits and vegetables. They are direct-to-consumer markets that not only provide consumer access to locally grown produce, they also provide an gathering spot that consumers can meet with farmers who grow their food [1]. The recent few decades have witnessed a rapid increase of the number of FMs in the United States. Despite these advantages, FMs can also encounter challenges such as food safety issues due to the nature of its open environment, lack of refrigeration units and sanitation facilities, and less strict inspection requirements [2,3]. In addition, quality of food, variety, and environment were reported as other factors that could influence consumers' purchase behavior on farmer's market [4-6]. Efficient communications with consumers to discover how they perceive these aspects are of importance to ensure a healthy growth of FMs.

Survey and questionnaire are main tools to know about consumer perceptions. Though those tools are effective in communicating with consumers, they have many disadvantages such as cost, coverage, volume, and timeliness. Recently, social media has been increasingly adopted as an alternative approach for studying small businesses such as farmer's market [5,7,8]. With its ability to allow consumers to express their spontaneous responses online, social media has advantages over surveys in that large amount of data can be collected at almost zero cost. However, previous studies on social media only considered basic problems such as the frequency of posts related to farmer's market, the structure of a farmer's market network, or consumers' feelings when talking about farmer's market. Interesting topics such as consumers' perceptions toward specific farmer's market issues often seen in conventional survey studies are rarely covered in social media studies.

One big challenge in analyzing consumer responses from social media is that the data is usually too large to read line by line as previously done in survey studies dealing with small datasets. The tremendous volume of mostly unstructured text from social media, though can be recognized and processed by humans, is significantly difficult for computers to understand [9]. A number of text mining techniques have been developed for dealing with specific tasks. Text classification is the approach often used to identify specific topics from text documents, often identified as a supervised machine learning (ML) task [10]. Common ML models such as Naïve Bayes (NB) and Support Vector Machines (SVMs) are widely used to preform text classification tasks. More recently, deep learning models such as convolutional neural networks have also been used in text classification.

In this work, the focus was to analyze social media data related to farmer's market on four aspects: quality, safety, environment, and availability. Popular machine learning models were applied to two distinct datasets: Twitter posts (tweets) and Yelp reviews see Figure 1. The purpose was to build language models for automatically identifying these four topics in related to farmer's market and to evaluate their performances. The result would help farmer's market

stakeholders to understand what consumers say or feel about their FM after a visit, so as to improve the communications between farmers and consumers for enhancing food safety and promoting a healthy growth of farmer's market.

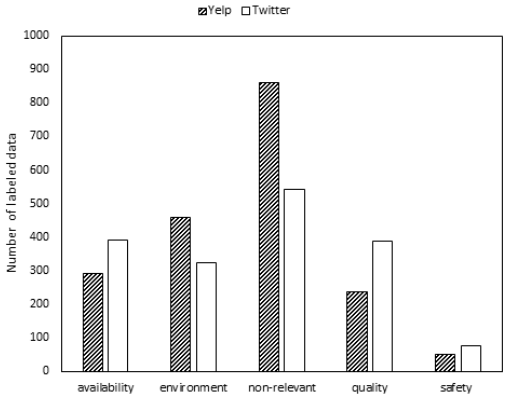

Figure 1. *Distribution of the number of labels in Twitter and Yelp datasets.*

Methods

Data collection

Two social media platforms (Twitter and Yelp) were employed in this study. Twitter provides a Streaming API for developers to collect real-time data. Here, we used a package called Tweepy to collect tweets posted by consumers after visiting a farmer's market. The data collection was conducted over a span of 2019/7/30 through 2019/10/7. To only include information related to farmer's market, we gathered tweets mentioning one of the keywords from "farmers market", "farmersmarket", "#farmers market", or "#farmersmarket". About 8M tweets for analysis were collected through several tries. Then we selected all geo-tagged tweets written in English language and filtered out those that are not posted from the US. The result was a collection of 50K tweets. A uniform sample of 2K tweets was randomly generated for human tagging and building language models. Another part of data was from Yelp, a review website. The data was collected from 2019 Yelp Dataset Challenge, including information about local businesses in 10 metropolitan areas. The dataset included multiple features such as "business_id", "category", "reviews", among which "category" describes the type of business like "restaurant", "grocery" or "farmers market". We downloaded the dataset and collected farmer's market-related reviews using a keyword filter on the "category" feature, which resulted in a collection of 3K reviews. Text segmentation was then applied to the reviews for dividing a paragraph of review into a few sentences. 17K sentences related to farmer's market was then generated from Yelp reviews. Similar to the process of Twitter data, a uniform sample of 2K sentences was randomly selected for human tagging and building classifiers.

Human labeling

Three undergraduate students and one investigator were employed as human taggers to read and label for 2,000 tweets and 2,000 sentences from Yelp reviews related to farmer's market. Each piece of data from Twitter or Yelp is referred to as document hereafter. The task was to assign a class i.e., "quality", "safety", "availability", "environment" or "non-relevant" to a given document. Three taggers were pre-trained so that they could have good understanding of the criteria and label the documents with the following tagging protocol:

Safety: if the text mentioned about sanitation condition in farmer's market or if someone got sick after patronizing farmer's market.

Quality: if the text mentioned about appearance, taste, smell, or overall sensory perception of food purchased from farmer's market.

Availability: if the text mentioned about the food supply condition of the farmer's market such as the variety, the size, the number of vendors.

Environment: if the text mentioned about the people, place, culture, events or any other non-food aspects that could influence consumer's liking toward farmer's market.

Non-relevant: if the text mentioned things not belonging to any of the "quality", "safety", "availability" or "environment" criteria.

For each document, the tagger determined which class it best fit and tagged the label on it. If it described aspects related to multiple classes, the tagger decided to assign it to the preferred one based on their evaluation. The reliability of the authority tagger, the investigator, was cross-checked against the voted opinion of all the other taggers. We observed 86.1% agreement (n=2000; 1722 agreements; 278 disagreements) on Twitter data 94.7% agreement (n=2000; 1894 agreement; 106 disagreement) on Yelp data. Only the document reaching agreement through this procedure was included to build language models, which finally resulted in 1,722 labeled documents from Twitter and 1,894 labeled documents from Yelp.

Data preprocessing

Raw text usually needs to be preprocessed before language analysis. Preprocessing is to remove irrelevant information without changing the basic meaning of the document [9]. Tokenization, normalization and noise removal are three basic preprocessing steps we performed in this section. Tokenization is the key component of preprocessing to break up a sequence into words/phrases (called tokens) for further processing. After tokenization, a series of normalization processes were applied, including converting all text to lowercase, removing punctuation, removing stop-words (a list of words frequently appears in the text without having much content information such as "the", "a", "is"), and eliminating affixes from a word in order to obtain a root through stemming. Specific tasks were performed for removing meaningless elements. For example, we further removed noises such as "rt@" and "https" for the Twitter data whose corpus was noisy than the Yelp data. At last,

term frequency–inverse document frequency (TF-IDF) was applied to transform the documents into vector representations using numerical values such that ML algorithms can process for further classification tasks [10].

Text classification models

The task of text classification is to find a classification model (classifier) *f* which can assign the correct class label to a new document *d* (test data) through a training process with labeled data. Various kinds of text classification models have been developed and publicly available for easy implementation on Python library scikit-learn. Here, we were facing a multi-class classification problem since we had five classes ("quality", "safety", "environment", "availability", and "non-relevant"). Common text classification ML models including Naïve Bayes (NB), Support Vector Machines (SVM), Logistic Regressions (LR), k-Nearest Neighbor (k-NN), and Random Forests (RF) were employed. Below is a description of the features of each model.

NB: The Naïve Bayes classifier is often used as the baseline of text classification as it is fast, easy to implement and relatively effective [11]. The assumption of NB models is that the features (words) are independent. The three common NB models are listed below.

• Bernoulli Naïve Bayes (BNB): in this model a document is represented by a vector of binary features denoting the presence or absence of the words in the document. Thus, the frequency of words is ignored.

• Multinomial Naïve Bayes (MNB): in this model a document is represented by a bag of words capturing the frequency of words. MNB often outperforms BNB for large vocabulary size.

• Complement Naïve Bayes (CNB): An adaptation of MNB is Complement Naïve Bayes (CNB) particularly suited for imbalanced data sets.

SVM: Support Vector Machines are algorithms extensively used in text classification problems which is quite robust in high dimensionality [9]. SVM makes classification decision by finding an optimized hyperplane with the maximum margin from the different classes.

LR: Logistic Regression is also a linear classification algorithm that can be easily generalized to multi-class problems [12].

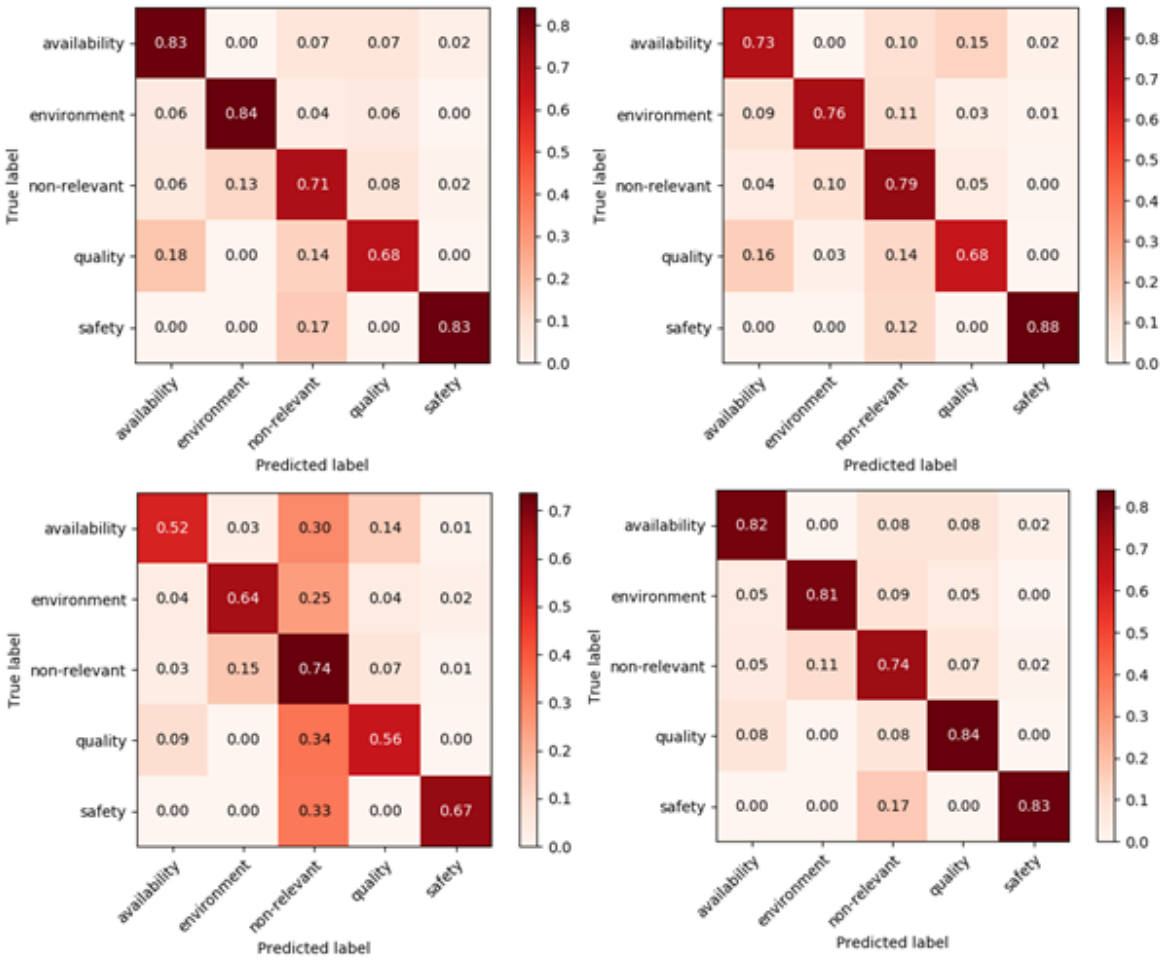

Figure 2. *Confusion matrix of different models in classifying Yelp data. (upper left: Support Vector Machines model; upper-right: Logistic Regression model; lower left: Complement Naïve Bayes model; lower right: Random Forest model).*

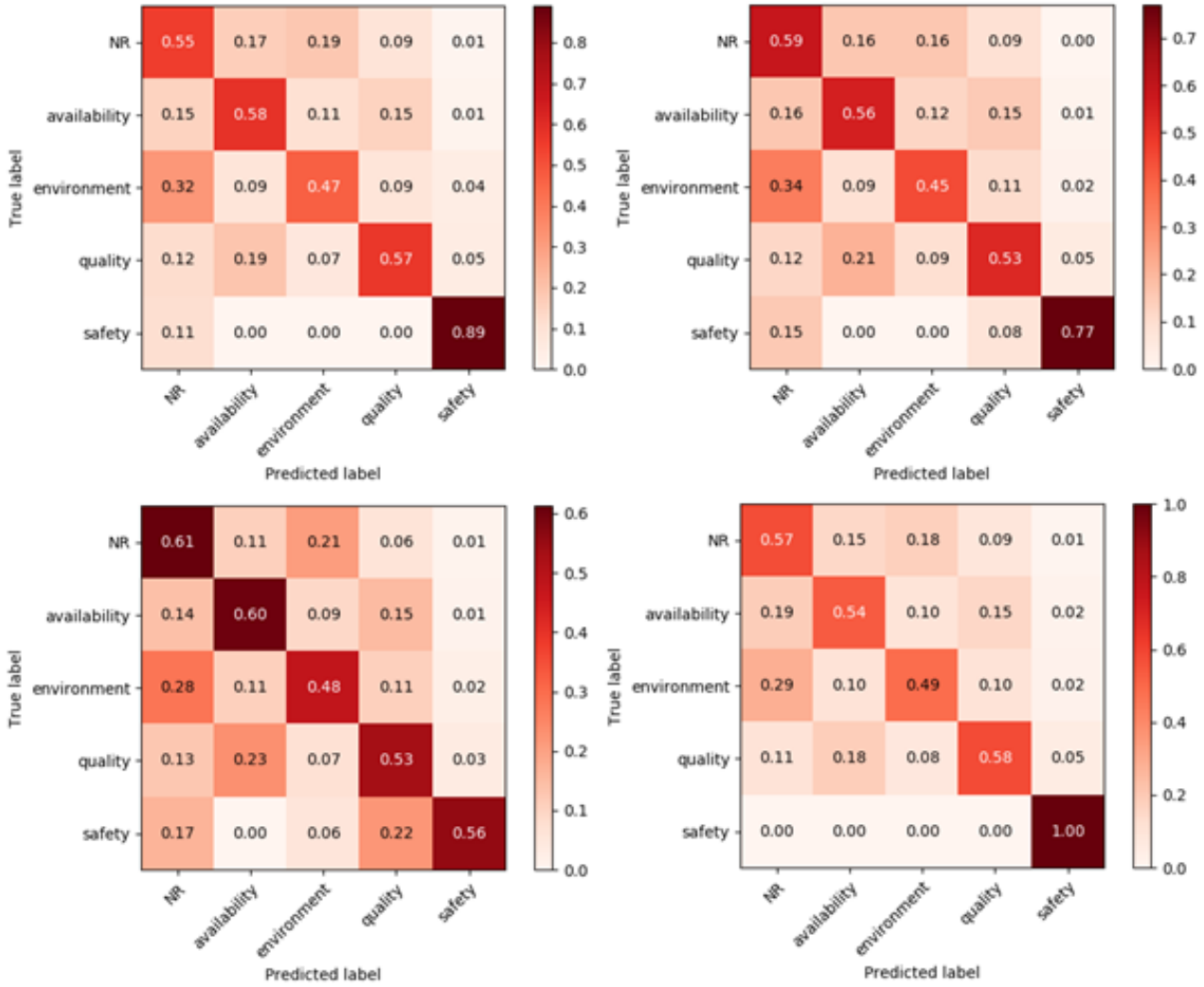

Figure 3. *Confusion matrix of different models in classifying Twitter data (upper left: Support Vector Machines model; upper-right: Logistic Regression model; lower left: Complement Naïve Bayes model; lower right: Random Forest model).*

k-NN: Nearest neighbor classifier is a proximity-based classifier which uses distance-based measures to perform the classification where k is the number of neighbors. The assumption is that documents belonging to one class are more likely similar or close to each other based on similarity measures [9].

RF: Random forest is another text classification algorithm with relatively high accuracy and suitable for imbalanced datasets [13]. It works through boosting of a number of classifiers to identify a class label.

Model evaluation

To evaluate the performance of the classification models, we set aside a random fraction of labeled documents (test set). After training the classifier with training set (80%), we classified the test set (20%) and compared the estimated labels with the true labels and measured the performance. We applied a five-fold cross validation to evaluate the classification models. The portion of correctly classified documents to the total number of documents is called accuracy. To be able to see if the model is mislabeling one class as another, a confusion matrix can allow a good visualization of the performance on each class. In a confusion matrix, each entry (i, j) represents the proportion of predicting the actual class in column j as the class in row i. The common evaluation metrics for text classification are precision, recall, and F-1 score [14]. Precision is the fraction of the correct instances among the identified positive instances. Recall is the correct instances among the identified positive instances. And F-1 score is the geometric mean of precision and recall, which is often used as a preferred matric for evaluating the performance of a classifier than just precision or recall. In a multi-class classification, the performance of a classifier on one individual class C_i is defined by (Accuracy)_i, (Precision)_i, and (Recall)_i shown in Table 1. In this study, the quality of the overall classification was assessed by a measure of the average of the same measures calculated for each class [15]. For the average parameter, weighted average was adopted for that it's more suitable for imbalanced data.

Results and Discussion

Performance evaluation of different models

In this work, we implemented common text classification models to identify important topics (quality, safety, environment, and availability) from Twitter and Yelp data related to farmer's market. The performance of these classification models was evaluated through multiple metrics as shown in Table 1. F-1 score is one of the popular metrics widely adopted for evaluating the performance of different classifiers [16]. Based on the F-1 scores, Support Vector Machines (SVM) was the most effective classifier for both Twitter (0.68) and Yelp data (0.75), in which the same model performed better on Yelp data. Besides, Linear Regression (LR) and Random Forest (RF) also showed relatively high performance in conducting the multi-class classification task. In terms of the classic Naïve Bayes classifiers, Complement Naïve Bayes (CNB) has outperformed Bernoulli Naïve Bayes (BNB) and Multinomial Naïve Bayes (MNB) on both datasets, which might be attributed to the issue of data imbalance (see Figure 1). For both Yelp and Twitter dataset, the largest number of samples were from the class "non-relevant" while the smallest number of samples were from the class "safety." Likewise, k-Nearest Neighbor (k-NN) didn't perform well due to the imbalance issue since it made decisions of a class label based on the majority vote, rather taking the weight of each class into consideration. In addition to the performance metrics, confusion matrix is another method to show more clearly the performance of each classifiers in predicting for each class of topics. The confusion matrix of the best-performed classifiers (SVM, LR, CNB, and RF) are shown in Figure 2 for Yelp datasets and Figure 3 for Twitter datasets, wherein the diagonal values are percentages of the classifier in predicting each class correctly. The color of the entry (i, j) was proportion to the value on that location. In other words, the darker the color, the higher of the percentage of predicting class in column j as the class in row i. In Twitter data analysis, it can be easily seen that "environment" was highly misclassed as "non-relevant" from any classifier, while "quality" or "safety" was misclassed as "availability" most often. In Yelp data

Table 1. *Performance of different classifiers on Twitter dataset and Yelp dataset.*

	Accuracy	F-1 score	Recall	Precision
Performance of models on classifying Twitter dataset				
SVM	0.68 ± 0.03	0.68 ± 0.03	0.68 ± 0.03	0.69 ± 0.03
LR	0.67 ± 0.04	0.67 ± 0.04	0.67 ± 0.04	0.70 ± 0.04
CNB	0.65 ± 0.03	0.65 ± 0.03	0.65 ± 0.03	0.65 ± 0.03
MNB	0.61 ± 0.04	0.59 ± 0.04	0.61 ± 0.04	0.67 ± 0.04
BNB	0.58 ± 0.05	0.58 ± 0.05	0.58 ± 0.05	0.64 ± 0.04
KNN	0.55 ± 0.07	0.54 ± 0.07	0.55 ± 0.07	0.63 ± 0.06
RF	0.66 ± 0.03	0.66 ± 0.03	0.66 ± 0.03	0.67 ± 0.03
Performance of models on classifying Yelp dataset				
SVM	0.75 ± 0.03	0.75 ± 0.03	0.75 ± 0.03	0.77 ± 0.03
LR	0.72 ± 0.03	0.71 ± 0.03	0.72 ± 0.03	0.74 ± 0.03
CNB	0.68 ± 0.02	0.68 ± 0.02	0.68 ± 0.02	0.69 ± 0.02
MNB	0.64 ± 0.02	0.61 ± 0.03	0.64 ± 0.02	0.70 ± 0.03
BNB	0.60 ± 0.02	0.56 ± 0.04	0.60 ± 0.02	0.63 ± 0.06
KNN	0.64 ± 0.01	0.62 ± 0.01	0.64 ± 0.01	0.66 ± 0.02
RF	0.73 ± 0.02	0.72 ± 0.02	0.73 ± 0.02	0.74 ± 0.03

Note. The mean and standard deviation are calculated based on five-fold cross-validation

Table 2. *Top 20 features of the SVM classifier for Twitter data.*

Availability	Environment	Quality	Safety	Non-relevant
Variety	Crowd	Good	Safety	Open
available	Music	Fresh	Health	Guide
Sell	Play	Taste	Coli	Hopefully
Lot	Volunteer	Delicious	Foodsafety	Andersonville
Strawberry	School	Fraud	Train	Elgin
Kale	Student	Organic	Recall	Come
Green	Meet	Best	Tip	Broadcast
Meat	Cruise	Fresher	Illness	Ravenswood
Potato	Band	Yum	Care	Stop
Avocado	Community	Aim	Practice	Grayslake
Tomato	Perform	Smell	Foodborne	Craft
Eat	People	Vegetable	Herb	Champaign
Tell	Intern	Fraudulent	Standard	Saturday
Bundle	Event	Weed	Study	Rain
Breakfast	Popularity	Parmesan	Regret	Lake
Chard	Fat	Better	Layout	Seven
Radish	Customer	Healthy	Amberjack	Tomorrow
Peach	Support	Sweet	Handle	Stand
Clara	Eerience	French	Arctic	Proximity
King	Fun	Amish	Pilot	Central

Table 3. *Top 20 features of SVM classifier for Yelp data.*

Availability	Environment	Quality	Safety	Non-relevant
Variety	Small	Quality	Clean	Stop
Selection	Parking	Taste	Messy	Know
Tomato	Crowd	Delicious	Safety	Day
Plant	Friendly	Organic	Gross	Love
Item	Price	Fresh	Rotten	Time
Fruit	Music	Tasty	Slaughter	Stuff
Assortment	Community	Juicy	Spray	Year
Jewelry	Dog	Corn	Cleaning	Happy
Goody	Cute	Sweet	Mosquito	Guy
Cucumber	Crowded	Produce	Expire	New
August	Fun	Healthy	Swarm	Saturday
Strawberry	Neighborhood	Plus	Organize	Week
Jam	Cash	Die	Hazard	Morning
Peach	Customer	Mind	Port	Experience
Summer	Support	Watermelon	Impeccably	Say
Ton	Smile	Free	Stocked	Open
Vendor	Great	Grow	Cleanly	Frill
Pasta	People	Chicken	Pose	Come
Option	Business	Fantastic	Natural	Hot
Bread	Little	Yummy	Impressed	Yes

analysis, there is a slightly higher accuracy of predicting each class correctly as compared to Twitter data analysis. However, the majority of misclassification was concentrated on predicting "non-relevant" for any of the other classes. This phenomenon might be attributed to the fact that the number of "non-relevant" in trained data was extremely higher than that of other classes.

The issue of data imbalance was addressed through resampling using the method modified from Thakur and co-workers [17]. Specifically, we sampled with replacement from the dataset 276 times on Twitter data and 303 times on Yelp data for each of the five topics, giving much higher chance for topics such as "quality" and "safety" being sampled, so that in expectation it ended up with an even number of each topic. This means that in the final dataset, many of the original instances would not be present and some may also be present more than once. To note that, only the training data was resampled. The overall performance scores of the classifiers trained from the resampled data were not higher than that of the original ones (data not shown). On the other hand, resampling process might introduce a bias by overemphasizing some of the topics. The estimated prediction accuracy for any model learned on the resampled data is therefore not meaningful unless it is estimated independently of the resampling process [17]. The primary purpose of this work was to classify the topics of interest related to farmer's market from real-world data. Therefore, the original models, though not perfect, was adopted for predicting new data.

Based on the evaluation results, SVM was selected as the optimal classifier for both Twitter and Yelp data analysis. The higher accuracy of SVM in classifying Yelp data over Twitter data might be attributed to their purpose of use and the resulting data characteristics. Yelp is a platform designed for consumers to write reviews while Twitter is a platform where anyone can share their ideas with others just like the function of many other social media platforms. Due to this fundamental difference, data on Twitter are more unstructured than data on Yelp. During the process of five-fold training process, a random of 1,378 Twitter data used for training gave a total vocabulary of 3,096 words; a random of 1,515 Yelp data used for training gave a total vocabulary of 2,465 words. It tells us that the vector space for Twitter data was sparser than that of Yelp data (more words and less training samples), which might help to explain the lower performance on Twitter data analysis.

poor hygiene conditions of retail facilities from consumer responses, which provides alternative means for public health departments to conduct regular food safety inspections [18-20]. However, no research has employed social media to investigated food safety topics related to farmer's market. In this research, typical text classification methods have been applied to detect food safety topics as well as three other topics, including food quality, food environment, and food availability, of importance to farmer's market performance. While the SVM classifier performed best among all the models, the accuracy in predicting food safety was still relatively low due to the limited dataset both in labeling and testing. Resampling could improve the performance; however, it would change the distribution of the original dataset. On the other hand, advanced technique such as transfer learning has been proven useful in improving model performance by incorporating more labeled data from other studies on the same topic [21]. Since several studies have explored foodborne illness detection using social media, transfer learning might help improve the model performance in classifying food safety topics (Figures 4 and 5).

Table 4. *Examples of tweets predicted as "safety" topic with the trained SVM classifier.*

Topic	Tweets
Safety	Farmers market food safety tips detective food safety nutrition health
Safety	Sprout's farmers market recalls spinach states listeria test food safety news
Safety	Farmers market vendors need training improve food safety practices
Safety	Food safety tips shopping farmers market
Safety	Bring home bacteria farmers market use tips pick fresh

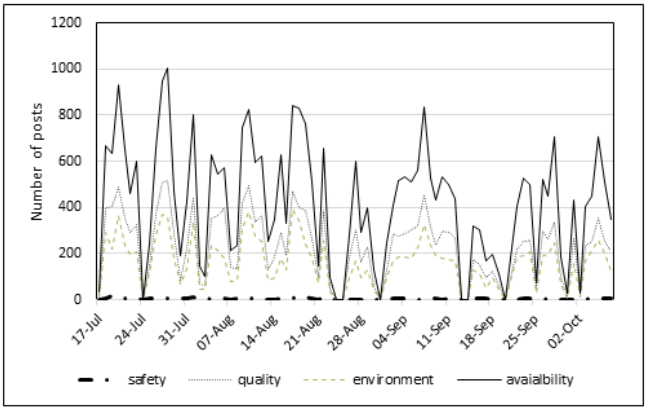

Figure 4. *Trends of farmer's market topics from Twitter streaming data.*

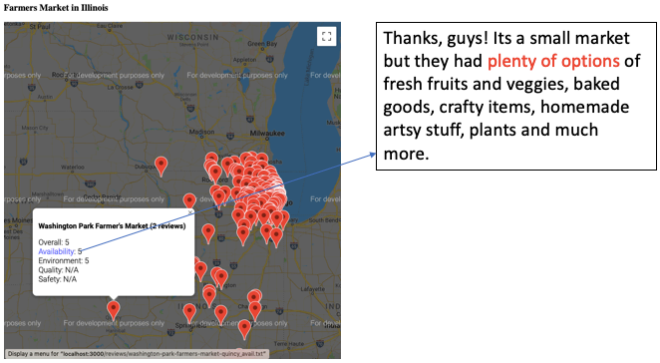

Figure 5. *Map visualization of Farmer's markets and ratings of specific topics.*

Important words for different topics based on the optimal model

Once having fitted the linear SVM classifier it is possible to access the classifier coefficients on the trained model. By looking at the SVM coefficients, the most important words used in the classification were identified. The top-20 words from the SVM classifiers are shown in Table 2 (Twitter) and Table 3 (Yelp). The overall words from the Twitter data were more diverse than that from the Yelp data, including unique words like 'clara', 'eerience', 'foodsafety', 'andersonville', 'elgin', 'ravenswood', 'grayslake', 'champaign'. Some might be part of a farmer's market location names, some were phrases tied together, and some might be common misspellings. However, most of the words gave us a clearer idea of the semantic meaning behind the language model of each topic. For example, important words related to the topic "availability" include 'variety', 'selection', 'lot', 'sell', 'available', 'assortment', 'item' and produce names such as 'strawberry', 'kale', 'potato', 'tomato', etc. Words that people used to mention topic "environment" often include 'crowd', 'parking', 'friendly', 'community', 'price', 'music', 'volunteer', 'customer', 'people', etc. Critical words related to the topic "quality" include 'quality', 'taste', 'fresh', 'organic', 'smell', 'healthy' and a variety of words describing how good the food is, such as 'good', 'tasty', 'delicious', 'juicy', 'sweet', 'fantastic', 'yummy', etc. To note that, "fraud" was found as an important word when people talk about quality of food on farmer's market. It might come from customers' concern around the origin of food there. For example, some vendors may claim that their food was harvested from their farms while the fact was that they purchased the food from supermarkets and sold with higher prices on farmer's market. In terms of the topic "safety", the important words identified from two datasets were different. On Twitter, words like 'safety', 'coli', 'health', 'recall', 'illness', 'train', 'tip', 'foodborne' were commonly mentioned, indicating people intend to talk about issues of foodborne outbreaks. On Yelp, people tend to comment on the hygiene conditions of a farmer's market with words like 'clean', 'messy', 'safety', 'gross', 'rotten' being used as the critical words.

Applications of the trained model

The trained SVM models were applied to both Twitter and Yelp datasets for visualizing the trends of farmer's market topics in both general (on Twitter) and specific scopes (on Yelp). Figure 4 shows the changes of farmer's market topics (quality, safety, environment, and availability) on Twitter within a period spanning from July 17, 2019 to October 7, 2019. Twitter releases its API for people to collect real-time data for research. However, the provided API was disrupted sometimes, which resulted in data missing of some days (see the intervals). Despite the missing data, we were able to observe daily activities of topics related to farmer's market globally. "Availability" turned out to be the most popular topics, followed by "environment", "quality, and "safety." Though not prevailing in number, "safety" was a critical aspect that we most cared about when talking about farmer's market issues. Some examples of safety-related tweets can be seen in Table 4. Due to the limit of number in training process and issue of data imbalance, approximately 54% tweets predicted as "safety" topic were truly relevant. Some examples of tweets classified as "safety" topics are shown in Table 4. The low volume of "safety" related topics on Twitter might indicate that the food safety issues on farmer's market is not significant during the time period observed. However, one challenge in analyzing Twitter data is due to its sparse location information. The results in Figure 4 are obtained without location differentiation. While Twitter provides its API for researchers to access to its data, it only releases sampled data accounting for about 1% of the total dataset related to the research topic. On the other hand, the number of tweets tagged with geolocation are relatively low. Some tools could provide more location information by inference of users' profiles, it is still difficult to know which farmer's market a tweet pertains to when "safety" related tweets are identified. Previous studies often employ official farmer's market account on Facebook or Instagram to analyze the activities inside the community [5,7,8]. However, analysis of specific farmer's markets would miss many consumer's responses of their experiences with farmer's market if they were not commented on the official accounts. Twitter provides an easy-to-report and easy-to-retrieve mechanism that allows consumers to comment on any topics freely and researchers to collect the comments with less restrictions. The major focus of future work should be on developing methods to increase the data resolution, in particular for the location data, which is rather important when significant food safety

issues (e.g., an outbreak) occur. The investigation of food safety issue would otherwise be impossible if little location-related information is provided.

On the other hand, most of the relevant tweets were focused on food safety practices on farmer's market, instead of complains of people getting sick after visiting farmer's market. Difficulty in identifying which farmer's market is associated with the topic is another limitation of Twitter data. On the contrary, Yelp gives a better resolution, in particular the location resolution, of the data to be analyzed. Similar to studies on official farmer's market accounts on Facebook or Instagram, the analysis of specific businesses registered on Yelp also gives the idea of consumer satisfaction of specific farmer's markets. However, the advantage of our work is the ability to process data from multiple farmer's markets automatically and compare the performances between them. For each farmer's market, the language model was able to detect sentences related to each topic, and a summarized score was given through a normalized algorithm based on sentiment analysis (number of positive, negative, and neural, as shown in supplement). As shown in the Figure 5, some farmer's markets registered on Yelp sites were shown on the google map. When clicking a specific farmer's market, one can see the rating of each topic. With a second click on one topic, relevant sentences would come out. This application makes it possible to conduct cross-farmers market comparisons. For example, if the Department of Health in Illinois wants to inspect a farmer's market, the safety score might be an indication for the hygiene status based on consumer reviews.

Conclusion

In this study, we employed social media platforms Yelp and Twitter to investigate the consumer comments on the farmer's market they visited. Different language models were applied for classifying specific topics, including food safety, food quality, food availability, and food environment. Linear classification model SVM was selected as the optimal model for both Twitter data and Yelp data analysis. By analysis of top coefficients of the SVM models for each dataset, it was observed that words most commonly associated with same topic (e.g., food safety) were significant different. On Twitter, people tend to talk about news-related topics such as foodborne outbreaks related to farmer's market while on Yelp people are more likely to comment on the hygiene conditions of the farmer's market. Due to the scarcity of location data from Twitter, it is challenging to identify which farmer's market may have a food safety issue as detected by the language model. On the contrary, Yelp data provided high location resolution, which allowed the connection of negative comments with specific farmer's markets. One limitation of this study is the data imbalance problem due to the uneven distributions of data related to different topics. While methods such as sampling with replacement could improve the performance of the classification models, it would modify the data distribution and won't be applicable to the predictions of new data. The data available is rather limited on safety topic, for which advanced techniques such as transfer learning might be helpful for enhancing the model performances as a number of labeled data related to foodborne illness detection are available from related studies. Through tracking the changes in different topics related to farmer's market on Twitter, potential food safety issues might be noticed when a rapid increase in the volume of a specific topic volume is detected. Also, the application of the language models on Yelp would help local health departments to design their inspection plan of farmer's markets by taking the hygiene status reported by consumers into account.

Acknowledgments

This work was supported by the Agriculture and Food Research Initiative (AFRI) award no. 2020-67021-32459 from the U. S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) and the Illinois Agricultural Experiment Station. The authors thank Amitesh Srivastava, Jae Wook Lee, and Hannah Wang for their assistance in reading and labeling social media data for this study.

References

1. Figueroa-Rodríguez KA, Álvarez-Ávila MDC, Hernández Castillo F, et al. Farmers' Market Actors, Dynamics, and Attributes: A Bibliometric Study Sustainability. 2019;11(3):745.
2. Young I, Thaivalappil A, Reimer D, et al. Food Safety at Farmers' Markets: A Knowledge Synthesis of Published Research. J Food Prot. 2017;80(12):2033-47.
3. Yu H, Gibson KE, Wright KG, et al. Food safety and food quality perceptions of farmers' market consumers in the United States. Fd Ctrl. 2017;79:266-71.
4. Pascucci S, Cicatiello C, Franco S, et al. Back to the future? Understanding change in food habits of farmers' market customers. Int Food Agribusiness Manag. 2011;14(4):105-126.
5. Cui Y. Examining farmers markets' usage of social media: an investigation of a farmers market Facebook Page. J Agri Food Sys. 2014;5(1):87-103.
6. Eberts CE. A Content Analysis of the Dallas Farmers Market Instagram. Master's thesis. 2016.
7. Pilař L, Rojík S, Balcarová T, et al. Farmers' Markets': The Usage of Instagram Posts. ICoM 2016.
8. Pilar L, Balcarova T, Rojik S, et al. Customer experience with farmers' markets: what hashtags can reveal. Int Food Agribusiness Manag. 2018;21(6):755-70.
9. Allahyari M, Pouriyeh S, Assefi M, et al. A brief survey of text mining: Classification, clustering and extraction techniques.
10. Zhai C, Massung S. Text data management and analysis: a practical introduction to information retrieval and text mining. 2016.
11. Rennie JD, Shih L, Teevan J, et al. Tackling the poor assumptions of naive bayes text classifiers. ICML-03. 2003;41:616-23.

12. Indra ST, Wikarsa L, Turang R, et al. Using logistic regression method to classify tweets into the selected topics. ICACSIS. 2016;385-90.

13. More AS, Rana DP. Review of random forest classification techniques to resolve data imbalance. ICISIM. 2017;72-8.

14. Aggarwal CC, Zhai C. Mining text data. Springer Science & Business Media. 2012.

15. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag. 2009;45(4):427-37.

16. Labatut V, Cherifi H. Accuracy measures for the comparison of classifiers. ArXiv. 2012.

17. Thakur M, Olafsson S, Lee JS, et al. Data mining for recognizing patterns in foodborne disease outbreaks. J Food Eng. 2010;97(2):213-27.

18. Park H, Kim J Almanza B. Yelp versus inspection reports: is quality correlated with sanitation in retail food facilities?. J Environ Health. 2016;78(10):8-12.

19. Sadilek A, Kautz H, DiPrete L, et al. Deploying nEmesis: Preventing foodborne illness by data mining social media. AI Mag. 2017;38(1):37-48.

20. Effland T, Lawson A, Balter S, et al. Discovering foodborne illness in online restaurant reviews. J Am Med Inform Assoc. 2018;25(12):1586-92.

21. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010;22(10):1345-59.

22. Nedumaran S, Selvaraj A, Nandi R, et al. Digital integration to enhance market efficiency and inclusion of smallholder farmers: a proposed model for fresh fruit and vegetable supply chain. Int Food Agribusiness Manag. 2020;23(3):319-337.

***Correspondence to:**

Hao Feng
Department of Food Science and Human Nutrition
University of Illinois at Urbana-Champaign
Urbana, IL 6180
USA
E-mail: haofeng@illinois.edu