JAMIA Open, 2(3), 2019, 330–338
doi: 10.1093/jamiaopen/ooz030
Advance Access Publication Date: 5 August 2019
Research and Applications

Research and Applications

Detecting reports of unsafe foods in consumer product reviews

Adyasha Maharana,[1] Kunlin Cai,[2] Joseph Hellerstein,[3] Yulin Hswen,[4,5,6] Michael Munsell,[7] Valentina Staneva,[3] Miki Verma,[8] Cynthia Vint,[9] Derry Wijaya,[2] and Elaine O. Nsoesie,[10,11]

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, [2]Department of Computer Science, Boston University, Boston, Massachusetts, USA, [3]eScience Institute, University of Washington, Seattle, Washington, USA, [4]Computational Epidemiology Lab, Harvard Medical School, Boston MA, USA, [5]Innovation Program, Boston Children's Hospital, Boston MA, USA, [6]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston MA, USA, [7]Department of Economics, Brandeis University, Waltham, Massachusetts, USA, [8]Applied and Computational Mathematical Sciences, University of Washington, Seattle, Washington, USA, [9]Computer and Information Systems, Boston University, Boston, Massachusetts, USA, [10]Department of Global Health, Boston University School of Public Health, Boston, Massachusetts, USA, and [11]Institute for Health Metrics and Evaluation, University of Washington, Seattle, Washington, USA

Authors 2 to 9 are arranged alphabetically.
Corresponding Author: Elaine Nsoesie, Department of Global Health, Boston University School of Public Health, 801 Massachusetts Ave, Crosstown Center 3rd Floor, Boston, MA 02119, USA; onelaine@bu.edu

Received 2 May 2019; Revised 27 June 2019; Editorial Decision 4 July 2019; Accepted 6 July 2019

ABSTRACT

Objectives: Access to safe and nutritious food is essential for good health. However, food can become unsafe due to contamination with pathogens, chemicals or toxins, or mislabeling of allergens. Illness resulting from the consumption of unsafe foods is a global health problem. Here, we develop a machine learning approach for detecting reports of unsafe food products in consumer product reviews from Amazon.com.
Materials and Methods: We linked Amazon.com food product reviews to Food and Drug Administration (FDA) food recalls from 2012 to 2014 using text matching approaches in a PostGres relational database. We applied machine learning methods and over- and under-sampling methods to the linked data to automate the detection of reports of unsafe food products.
Results: Our data consisted of 1 297 156 product reviews from Amazon.com. Only 5149 (0.4%) were linked to recalled food products. Bidirectional Encoder Representation from Transformations performed best in identifying unsafe food reviews, achieving an F1 score, precision and recall of 0.74, 0.78, and 0.71, respectively. We also identified synonyms for terms associated with FDA recalls in more than 20 000 reviews, most of which were associated with nonrecalled products. This might suggest that many more products should have been recalled or investigated.
Discussion and Conclusion: Challenges to improving food safety include, urbanization which has led to a longer food chain, underreporting of illness and difficulty in linking contaminated food to illness. Our approach can improve food safety by enabling early identification of unsafe foods which can lead to timely recall thereby limiting the health and economic impact on the public.

Key words: food and drug administration, food safety, consumer product safety, machine learning, artificial intelligence

BACKGROUND AND SIGNIFICANCE

Unsafe products in the marketplace can have major health consequences that include injury, illness, and death, as well as economic burden to these product markets and individuals affected. These product categories include food, cosmetics, appliances, toys, and many other industries.[1–4] Undeclared food allergens on labels are the number one reason for food recalls in the United States.[5,6] Data from 1999, and from 2002 to 2012, indicated that eggs[5] and milk were the most frequently undeclared food allergens,[7,8] respectively. And bakery products were the most frequently implicated food product.[7,8] The prevalence of self-reported food allergies is about 9.1%,[5] therefore food labels need to be accurate and truthful to prevent serious health harms.

Another major reason for recall is the contamination of meat, poultry and produce with pathogens such as, Listeria, Salmonella, and Escherichia coli (*E. coli*).[9–11] Examples of food recalls in recent years have included, romaine lettuce[12] and spinach due to *E. coli* contamination,[13,14] and peanut[15,16] and eggs due to Salmonella.[17–19] Illnesses resulting from unsafe foods produce significant disease burden. The Centers for Disease Control and Prevention estimates that 76 million foodborne illnesses, including 325 000 hospitalizations and 5000 deaths, occur each year in the United States.[20] According to the United States Department of Agriculture (USDA), foodborne illness costs the US economy $10–83 billion per year with more detailed aggregated models finding it to be $77.7 billion annually.[20,21]

In recent years, the number of foodborne disease outbreaks and concomitant has increased.[22] Traditional forms of detecting unsafe food products require a significant amount of time and resources, which leads to delays in detection and increases [in] the number of infections, illnesses, and deaths.[16] In the United States, the Food Safety and Inspection Service of the USDA and the Food and Drug Administration (FDA) of the United States Department of Health and Human Services are responsible for regulating the food supply.[23,24] These agencies are also authorized to take administrative actions to conduct recalls of unsafe food products (such as those that are mislabeled, contaminated, or spoiled) reported by the public or public health agencies. However, issuing a recall requires thorough investigation which can take several months, meanwhile the contaminated product is being consumed by the public who are unaware of the unsafe food.[24] Furthermore, this rarely occurs due to limitations in reporting and validation. Food recalls are almost always a voluntary action initiated by the food manufacturer to remove a food from commerce which are a result from public complaints on the impure, unsafe, mislabeled, or pathogen contaminated food product.[24] Therefore, it is essential to have a rapid and reliable food surveillance system for early detection of unsafe foods in order to prevent the onset of outbreaks and serious harm to the public.

This project explores how consumer product reviews can be used to aid the existing system of detection and reporting of unsafe foods by applying machine learning to data from the FDA, and an online retailer (ie, Amazon). The two aims of this project are: (1) mine and integrate a large corpus of data posted online to understand trends and features in unsafe food product reports, and (2) develop a machine-learning/informatics approach for early identification of unsafe food products. Early identification of unsafe foods can lead to timely recall thereby curbing the health and economic impact on the public.

METHODS

Data gathering

Our data consisted of Amazon reviews of Grocery and Gourmet Food products and enforcement reports from the FDA. We manually downloaded FDA enforcement reports which were available as weekly CSV files from 2012 to 2017. The Amazon reviews were downloaded from a public repository[25,26] and were available for the years 1996–2014. The dataset is a collection of product data across multiple categories with a comprehensive gathering of product reviews for each product. The data consisted of the following information: reviewer ID, the Amazon Standard Identification Number (ASIN) which Amazon uses to identify products, reviewer name, helpfulness of rating, review text, overall rating (1–5 stars), summary of review, and review time.

Our first project aim was to create a database linking Amazon products and reviews to product recall data from the FDA. The most reliable way to match recalled products present in the FDA data with Amazon reviews was by using the item's Universal Product Code (UPC), which is the number that appears on a barcode and uniquely identifies that particular product. The FDA enforcement reports often (but not always) contained the UPC or UPCs of the product(s) being recalled within a larger text field. We used regular expressions to extract these codes and, in some cases, where partial UPCs were provided, generated lists of the possible complete UPCs from the partial codes. We used a publicly available conversion tool—UPCtoASIN.com—to convert UPCs to ASINs. The integrated data was placed in a PostGres relational database that links FDA product recalled data to Amazon reviews of the same product (Figure 1). FDA reports that did not contain the UPC were discarded from the joint database, but included in our analysis of FDA product recalls. See our GitHub repository[27] for additional information about the database and how you can gain access for research purposes.

Data processing and analysis

Reasons for FDA product recall

Our second project aim was to develop a framework for early identification of unsafe food products. In order to identify relevant product reviews, we needed to identify major reasons why the FDA issues product recalls. We used unsupervised topic modeling through non-negative matrix factorization[28] and cluster analysis to broadly categorized recall reasons from FDA reports into seven categories. Topic Modeling is a traditional text mining approach which represents each document (in our case each review) by a weighted combination of a few topics (each topic is represented by a small set of keywords). The unstructured text in the reviews is vectorized by calculating the Term Frequency-Inverse Document Frequency (TF-IDF) matrix, which in practice represents each review by the histogram of words appearing in them, scaled by the frequency of those words across documents. Then performing non-negative matrix factorization on that matrix yields two new matrices: one containing the representation of the topics, and one representing the topic weights for each review. Under certain conditions, this approach is equivalent to the probabilistic framework of topic modeling through Latent Dirichlet Allocation.[28] The topics identified were used to generate synonyms to identify relevant product reviews in the Amazon data, which was then annotated via crowdsourcing and used for training machine learning classifiers.

Tagging Amazon product reviews

In order to train a classification algorithm, we assigned tags (yes/no) that reference whether or not a product review in the PostGres relational database was published within the chronological vicinity of the time that an FDA recall corresponding to the product was

Figure 1. Database for linking Amazon reviews to FDA recalls.

released. This is not a very straightforward task, because the timeframe between a product needing to be recalled and a product actually being recalled varies greatly. Since it was unclear the best way to define the recall/review relationship, we used all reviews submitted prior to the date of product recall.

Crowdsourced annotation

To train machine learning classifiers to identify reviews indicating illness or product mislabeling, we needed a manually annotated dataset. We randomly sampled 6000 reviews from our PostGres relational database that contained words related to FDA recall reasons (such as, "sick," "label," "ill," "foul," "rotten," and etc.), along with metadata (ie, title of review, rating etc.). We employed workers on Mechanical Turk to annotate the data and provided detailed instructions including what to look for in the product reviews. Workers were compensated 0.15 USD for each product review and each review was annotated by two Mechanical Turk workers into one of four categories:

1. Review implies that consumer fell sick/had allergic reactions or has labeling errors
2. Review implies that the product expired or looks/tastes foul and should be inspected
3. Review does not imply that the product is unsafe
4. Review cannot be categorized to the above three categories

The final dataset contained 352 reviews which directly implied that the product was unsafe (ie, belonging to Category 1).

Identifying reports of unsafe food products

We applied several machine learning classification methods to the review text, title and consumer rating to identify relevant reviews. Since food recalls only occur for a small percentage of products and usually for specific product batches, the dataset is highly skewed; the classification task is akin finding a needle in the haystack. We explored the following standard machine learning classifiers for identifying unsafe food products: linear Support Vector Machine (SVM), multinomial Naive Bayes, and weighted logistic regression,[29] with various

approaches for addressing data imbalance. The multinomial Naive Bayes is a very simple classifier which relies on the assumption that the features are independent. Its main advantage is that it is efficient, scales well to massive datasets and can be easily used in real-time systems. The weighted logistic regression allows for the incorporation of the class imbalance directly into the method as a prior on the class distribution. The linear SVM involves finding the hyperplane that maximizes separation between annotated classes, while minimizing misclassification between classes. We also coupled these classifiers with feature selection and Synthetic Minority Over-sampling Technique (SMOTE)[30] to address the imbalanced data problem. SMOTE combines subsampling [of] the majority class with oversampling of the minority class and creates synthetic samples from the minority class by adding random perturbations to attributes of similar instances (according to a distance measure).

Additionally, we explored a deep learning method; bidirectional Encoder Representation from Transformations (BERT)[31]—an unsupervised language model that is trained on large text corpus such as articles in English Wikipedia. Using the pretrained BERT language model, we learn from small datasets by fine tuning BERT for the specific task. BERT uses the "masked language model" as a pretraining objective: it randomly deletes some words from a sentence, and then trains a bidirectional transformer to predict the identity of the removed words thus fuzing both the left- and right-hand contexts of words together in the same vector space. This innovation creates better performing word and sentence embeddings that once trained, can be quickly and easily fine-tuned by adding one layer to the end of the network to continue training on a specific task. We used the pretrained BERT-base uncased model to fine-tune to our unsafe food classification task by adding a prediction layer (safe vs. unsafe) on top of the model. We trained this network for 10 epochs with a batch size of 4 and 2e-5 learning rate.

Feature selection

As noted, both shallow and deep text classifier models were trained for classifying unsafe food reviews. Extensive feature engineering was performed to get the best possible performance from shallow classifiers. Preprocessing steps like lowercasing, removal of special characters, editing repetitive characters in social media expressions, and replacing numbers were applied to the words before extraction of n-gram features.

Every product was accompanied by several categories of metadata information that can be utilized for identifying patterns that are associated with recalled products. Extensive feature engineering was performed to utilize this information and integrate into feature vectors for classification. Features included average review length (both word and character count), total number of reviews, percentage of 1–5 star reviews, product category, product description, *n*-grams from reviews, and etc. Further, Chi-squared and mutual information-based feature selection methods were used to retain only the most informative features for classification.

We hypothesized that the sale of a contaminated batch of a product might lead to a temporary uptick in the number of reviews appearing on the e-commerce platform. To account for this pattern, we built feature variables like maximum and minimum weekly/monthly review count and weekly/monthly increase in reviews. From our analysis of FDA recall reasons, we identified several words such as "pesticide," "warning," "contamination," for which we built explicit indicator variables. Consequently, the final feature vector used in the shallow classifiers was a mix of categorical, continuous, and textual variables.

RESULTS

Data summary

The Amazon data contained 171 760 products. The top five product categories were beverages (3925 products), cooking and baking (2434), tea (1791), chocolate (1043), and snack foods (983). The most reviewed product categories were beverages, cooking and baking, tea, sugar, and vinegars with 23 570, 20 297, 9737, 5793, and 4792 reviews, respectively. There were vastly more reviews for products that have not been recalled than there were for recalled products, since most food products are never recalled. Over 1.2 million (99.6%) reviews were for nonrecalled products, and 5149 (0.4%) were for recalled products (Figure 2). The number of reviews for both recalled and nonrecalled products increased over time, likely reflecting Amazon's popularity as an e-commerce website and/or the number of food products on the site (see Figure 3a). The average and median number of reviews for recalled products was 24.75 and 5.0, respectively. While the average and median number of nonrecalled products was 9.29 and 2.0, respectively. The disparities in these figures might be due to the significantly larger number of products in the nonrecalled class (ie, 171 552 vs. 208) and the fact that many products had few reviews. For both recalled and nonrecalled products, five-star reviews were the most common (see Figure 3b). Reviews for recalled products are distributed across all ratings categories, likely due to the fact that products are recalled in batches.

Reasons for FDA product recalls

Of the nearly 3000 FDA products recalled, most were due to errors in the labels and undeclared ingredients (Figure 4). The second most frequent reason for recall was contamination with Salmonella and Fungi, with mold being the most frequently cited Fungi. The remaining five categories are listed in Figure 3. Broadly reasons for recall can be divided into three categories—contamination, presence of foreign objects and undeclared ingredients (or labeling/packaging errors). Contamination is usually accompanied by the mention of bacteria like Clostridium Botulinum, Listeria, or Salmonella. Foreign objects are contained in packaged products due to factory mishaps or errors in the production line. These are usually metal or glass fragments. Due to labeling or packaging errors, some ingredients such as milk, nuts, and wheat go undeclared. This is a problem because these items are potential allergens and may pose risk to people's health. Other items that are classified as undeclared ingredients include chemicals such as food colors, and preservatives. In addition to the above mentioned broad categories of recalls, many other reasons were found to occur at a lesser frequency (see Table 1). We found synonyms for terms associated with FDA recalls in more than 20 000 reviews, most of which were associated with nonrecalled products, suggesting that further investigation might have led to more recalls than those noted in our data.

Prediction of unsafe food reports

Of the 6000 manually annotated reviews, the F-measure for inter-annotator agreement was 0.44, 0.33, 0.79, and 0.1 for categories 1, 2, 3, and 4, respectively. These included 352 unsafe food reviews. The reviews with disagreeing annotations were reannotated by a third person in our team resulting in 5642, after excluding reviews annotated as ambiguous. However, all the new reviews were associated with nonrecalled products.

The 5-fold stratified cross validation performance (macroaveraged measures since cross-validation folds were equal sized) of various supervised machine learning classifiers under different

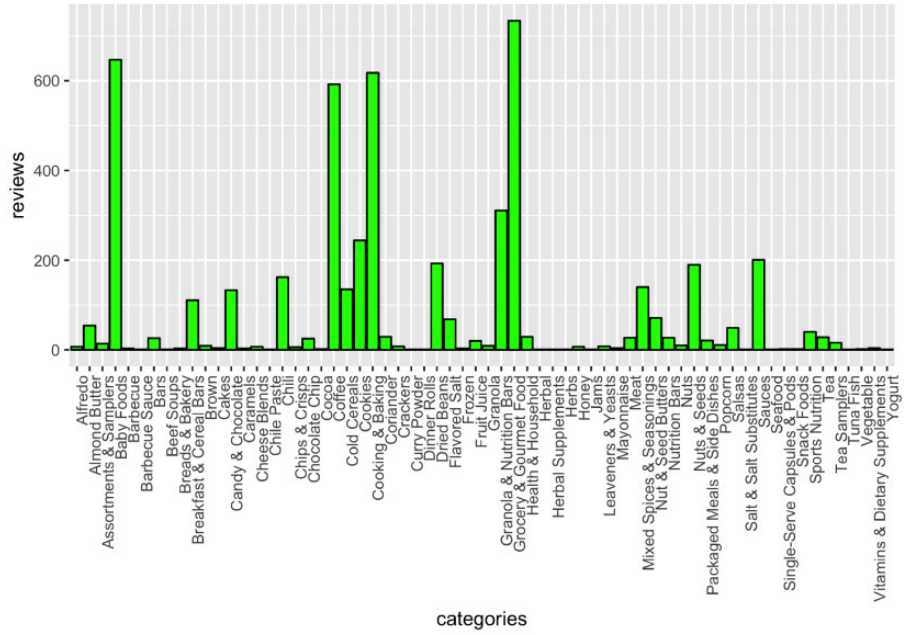

Figure 2. Distribution of consumer reviews across recalled product categories.

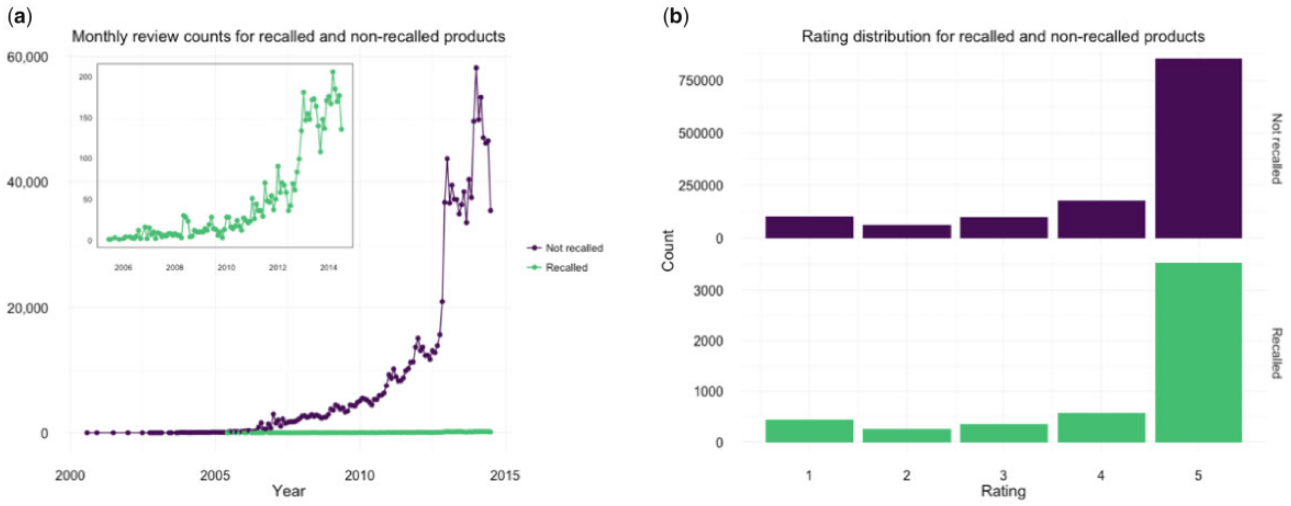

Figure 3. Features of Amazon reviews for the study period. Temporal trends (a) and distribution of customer ratings (b) of Amazon reviews.

conditions is described in Table 2. The performance of all three shallow machine learning methods ranged in F1 scores between .6 and .7. The simple Naive Bayes classifier did not significantly underperform, which suggests that the selected features were not redundant.

The best performing classifier was the deep learning classifier, BERT with an F1 score of 0.74 (precision and recall of 0.78 and 0.71, respectively). Similar to the typical approach for formulating the task of classification of severely imbalanced data to that of anomaly detection, we trained an autoencoder neural network on BERT vector representation of sentences from safe product reviews with the goal of compressing the vectors into their lower-dimensional representations and then reconstructing the original vectors. The autoencoder was trained to capture the meaning of sentences from safe product reviews and therefore when it was fed vectors of sentences from unsafe product reviews, it should have resulted in larger re-construction errors if the sentences of unsafe product reviews were very different from the sentences of safe product reviews (in other words if the unsafe reviews were indeed anomalies). However, in our data we find that the distribution of reconstruction errors of sentences from safe and unsafe product reviews were very similar, with the reconstruction errors of sentences from safe and unsafe product reviews having the same mean and median values (0.26 and 0.25, respectively) as shown in Figure 5.

DISCUSSION

In this study, we developed an approach for identifying unsafe food products. The deep learning classifier, BERT, was the best performing and achieved a reasonable balance of accuracy and recall. Our findings are important for several reasons. First, reports of mislabeled products even by a single consumer can be important and can provide a warning to companies. Second, identification of unsafe

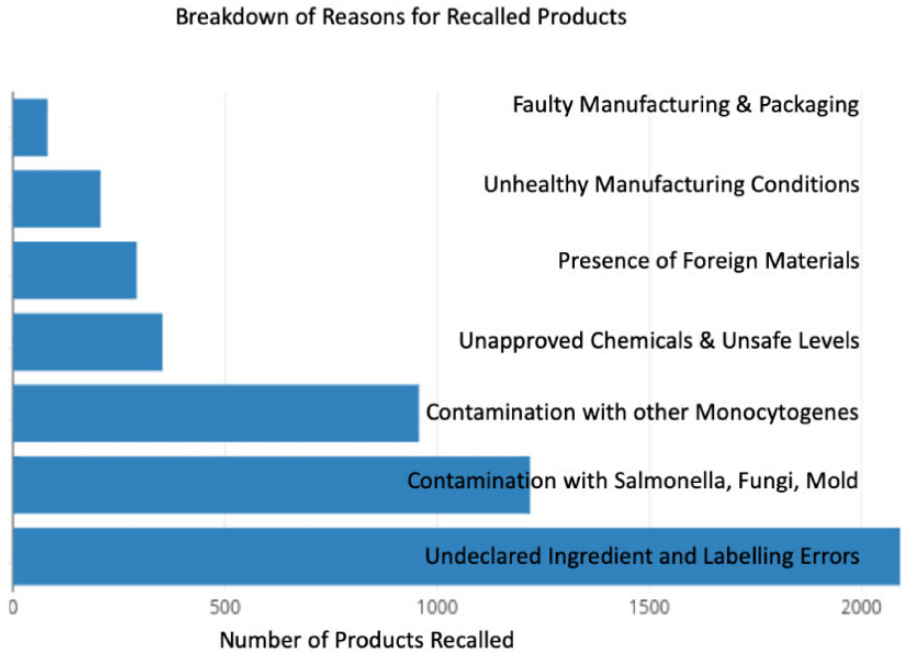

Figure 4. Reasons for FDA food product recalls.

Table 1. Recall reasons not captured in Figure 3

Contamination (others)
contains a raw material that may contain >0.3 ppb chloramphenicol
contaminated with undeclared steroids
products in vacuum packages were undercooked.
Illegal
levels of Aflatoxin above legal limit
does not meet pH standard of 10 for boiled/preserved eggs
found a chemical which does not have a tolerance level in US
FDA testing found unapproved pesticides/not permitted in US
pesticide not allowed in US but approved for usage in EU
violative levels of lead.
Issues with manufacturing/transport
liquid containing vessel may leach lead
firm was manufacturing acidified foods without license
may not have been transported at a safe temperature
recalling firm lacked adequate Good Manufacturing Practices
packet may have an incomplete seal which could allow air to enter the packet causing
oxidation
improperly pasteurized
faulty screen at flour mill.
Voluntary recall
notification of opportunity to initiate voluntary recall - letter from FDA

reviews could lead to timely investigation and recall of products, thereby limiting the health and economic impact.

Here, we defined unsafe foods, as foods products that were recalled by the FDA. However, our analysis revealed that over 20 000 reviews contained terms typically used by the FDA in enforcement reports but most of these were for products that were not recalled. This suggest that these data can be useful for monitoring reports of unsafe foods to improve food safety in the United States.

Due to a severely unbalanced dataset, and minimal difference between reviews of safe and unsafe food products, the task of identifying unsafe food reviews and products proved to be difficult despite comprehensive feature engineering. We found that it is easier to automate the identification of unsafe food reviews than it is to predict whether a product will be recalled or not. This is because products tend to be recalled in batches because a sample of products might become contaminated during processing or distribution. Nevertheless, examination of the most informative features reveals that the classifiers can capture relevant information for identifying recalled products. Review classifiers assign higher importance to the presence of negative sentiments like "gross," "terrible," "strange," and "unacceptable." Words like "manufacturer" and "factory" also feature as informative features because of customer reviews that discuss about manufacturing defects like labeling errors.

Data augmentation using SMOTE technique did not lead to substantial improvement in review classification performance. This could be because SMOTE randomly selects feature attributes (*n*-grams, in our case) and perturbs them to create synthetic samples. Targeted feature perturbation such as replacing words associated with sentiment/opinion with words that are close in the semantic space, might lead to more effective data augmentation. Another way to improve classification for imbalanced datasets is to implement cost sensitive learning—random forest and decision tree classifiers are the most widely used methods for implementing this idea. But, they did not achieve better results than weighted logistic regression.

Weighted logistic regression with feature selection using Chi-squared information measure outperforms other algorithms in terms of recall, however, text classification using BERT embeddings gives the highest F-score for this dataset. The use of SMOTE leads to marginal gains in the performance of unsafe product classification. The best performing shallow algorithm is weighted logistic regression along with SMOTE and feature selection, resulting in F-score of 0.59. It is harder to imagine a better data augmentation scenario for the highly unbalanced product classification dataset. Effective data augmentation will seek to replicate

Table 2. Performance of the various machine learning approaches employed for identifying unsafe food products

Classifier description	Precision	Recall	F1 score
Linear SVM (Feature selection using Chi^2 $k = 500$)	0.61	0.64	0.62
Multinomial Naive Bayes (Feature selection using Chi^2, $k = 500$)	0.66	0.66	0.66
Weighted logistic regression (Feature selection using Chi^2, $k = 500$)	0.58	0.74	0.65
Weighted logistic regression (Feature selection using Chi^2, $k = 1000$)	0.64	0.71	0.67
Weighted logistic regression (Feature selection using mutual information, $k = 1000$)	0.60	0.68	0.64
Weighted logistic regression with SMOTE (ratio = 1: 5) (tested on real data points only)	0.62	0.68	0.65
Weighted logistic regression with SMOTE (ratio = 1: 3) (tested on real data points only)	0.62	0.71	0.66
Weighted logistic regression with SMOTE (ratio = 1: 2) (tested on real data points only)	0.62	0.70	0.66
Weighted logistic regression with SMOTE (ratio = 1: 1) (tested on real data points only)	0.63	0.66	0.64
BERT (epoch = 10, max sequence length = 128)	0.76	0.67	0.71
BERT (epoch = 10, max sequence length = 128) with focal loss for dealing with imbalanced data ($\alpha = 0.915$, $\gamma = 5$)	0.75	0.74	0.73
BERT (epoch = 20, max sequence length = 256)	0.79	0.67	0.72
BERT (epoch = 30, max sequence length = 256)	0.78	0.71	0.74
BERT (epoch = 30, max sequence length = 256) with focal loss for dealing with imbalanced data ($\alpha = 0.915$, $\gamma = 5$)	0.77	0.71	0.74

BERT is the best performing classifier. Chi^2 refers to Chi-square. The accuracy ([true positives + true negative]/total reviews), precision (also known as positive predictive value = true positives/predicted positive condition), recall (also known as sensitivity = [true positive/[true positives + false negatives]), and F1-score (the harmonic mean of the precision and recall) are discussed.

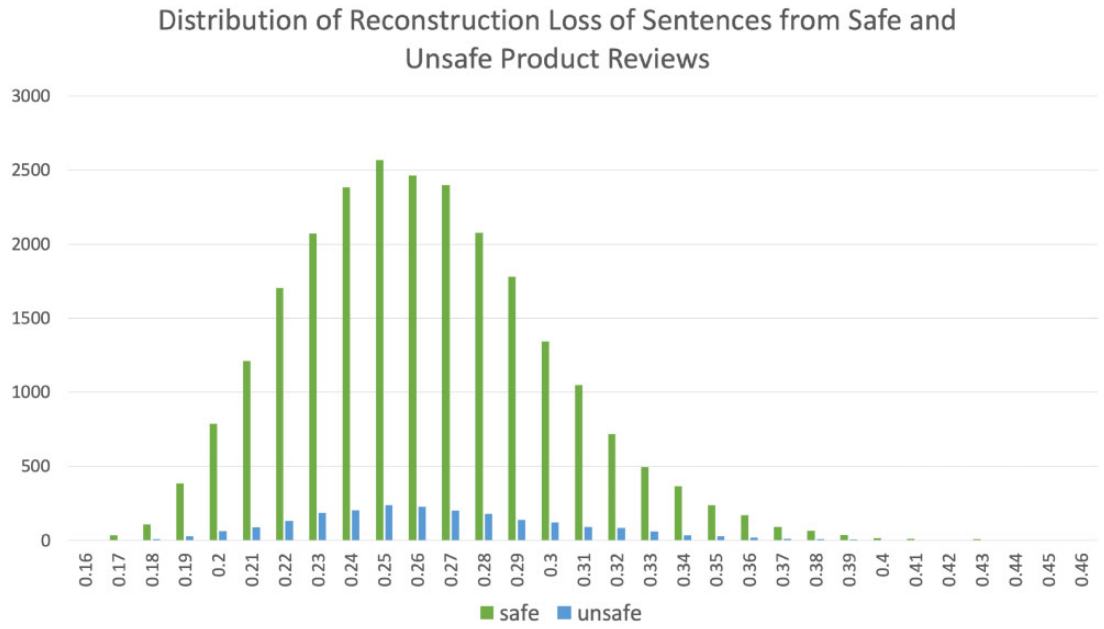

Figure 5. Plot of the reconstruction error. This shows that sentences from unsafe and safe product reviews are not significantly different. This might explain the difficulty in the classification process.

and synthesize the consumer response that follows consumption of an unsafe product, however, there aren't any studies that analyze such scenarios. In future work, we plan to extend this dataset and use predictions from the review-based classifier to predict product recall.

Unsafe foods are a global public health problem.[32] There are more than 200 diseases associated with unsafe foods that can cause illness and death.[33] While foodborne illness affects individuals of all ages, young children, elderly and sick individuals, tend to be more severely affected. The World Health Organization estimates that in 2010, 600 million people experienced illness due to contaminated food, globally.[34] Data from social media, crowdsourced websites, Google searches and business review websites, such as, Yelp have been shown to be useful in monitoring reports of foodborne illness.[35–42] Health departments in cities such as, New York City, St. Louis, Chicago, and Las Vegas have demonstrated that data from these sources can be used for targeted restaurant inspections and identification of outbreaks not reported through traditional surveillance systems.[35–37,42]

Our approach complements these foodborne illness surveillance efforts by exploring how we can improve early detection of unsafe foods to reduce morbidity and mortality resulting from the consumption of these foods. If successful, our approach can be implemented for early detection of unsafe food products based on consumer reviews submitted on Internet-based platforms such as, e-commerce websites, forums, and social media. Early identification of unsafe food products would have important implications not only in the United States but globally. By identifying unsafe food prod-

ucts early, companies can take appropriate actions to stop the sale of these products. This would also limit the occurrence of large foodborne disease outbreaks, thereby preventing illness and deaths, and reducing the health and economic impact on households, businesses and the food industry.

FUNDING

This project is supported by a fellowship from the Boston University Data Science Initiative through the Rafik B. Hariri Institute for Computing and Computational Science.

ACKNOWLEDGMENTS

We thank Kara H. Woo for contributing to data extraction, processing and preliminary analysis. We also thank the University of Washington eScience Institute for supporting this project through the Data Science for Social Good (DSSG) program.

COMPETING INTEREST

The authors declare no competing interests.

REFERENCES

1. Zhang P, Penner K, Johnston J. Prevalence of selected unsafe food-consumption practices and their associated factors in kansas. *J Food Saf* 1999; 19 (4): 289–97.
2. Story L. Lead paint prompts Mattel to recall 967, 000 toys. *The New York Times*. 2007: 2.
3. Story L, Barboza D. Mattel recalls 19 million toys sent from China. *The New York Times*. 2007: 15.
4. Teagarden MB, Hinrichs MA. Learning from toys: reflections on the 2007 recall crisis. *Thunderbird Int Bus Rev* 2009; 51 (1): 5–17.
5. Vierk KA, Koehler KM, Fein SB, *et al*. Prevalence of self-reported food allergy in American adults and use of food labels. *J Allergy Clin Immunol* 2007; 119 (6): 1504–10.
6. FARE. *Food Labeling Issues*. 2018. https://www.foodallergy.org/life-with-food-allergies/newly-diagnosed/food-labeling-issues. Accessed December 14, 2018.
7. Malyukova I, Gendel S, Luccioli S. Milk is the predominant undeclared allergen in US food product recalls. *J Allergy Clin Immunol* 2012; 129 (2): AB234.
8. Gendel SM, Zhu J. Analysis of US Food and Drug Administration food allergen recalls after implementation of the food allergen labeling and consumer protection act. *J Food Prot* 2013; 76 (11): 1933–8.
9. Teratanavat R, Hooker NH. Understanding the characteristics of US meat and poultry recalls: 1994–2002. *Food Control* 2004; 15 (5): 359–67.
10. Gorton A, Stasiewicz MJ. Twenty-two years of US meat and poultry product recalls: implications for food safety and food waste. *J Food Prot* 2017; 80 (4): 674–84.
11. Bennett SD, Sodha SV, Ayers TL, *et al*. Produce-associated foodborne disease outbreaks, USA, 1998–2013. *Epidemiol Infect* 2018; 146: 1397–406.
12. Centers for Disease Control and Prevention (CDC). Multistate Outbreak of E. coli O157:H7 Infections Linked to Romaine Lettuce (FINAL UPDATE). March 23, 2012. https://www.cdc.gov/ecoli/2011/romaine-lettace-3-23-12.html (Accessed: July 17, 2019).
13. Todd ECD, Harris CK, Knight AJ. Spinach and the media: how we learn about a major outbreak. *Food Prot Trends* 2007; 27: 314–21.
14. Lang S. Effects of food recalls on retailersã stock price. 2018. Cornell University (Thesis). https://ecommons.cornell.edu/handle/1813/60163. (Accessed July 17, 2019).
15. Cavallaro E, Date K, Medus C, *et al*. Salmonella typhimurium infections associated with peanut products. *N Engl J Med* 2011; 365 (7): 601–10.
16. Powell DA, Jacob CJ, Chapman BJ. Enhancing food safety culture to reduce rates of foodborne illness. *Food Control* 2011; 22 (6): 817–22.
17. Centers for Disease Control and Prevention (CDC). Multistate Outbreak of Human Salmonella Enteritidis Infections Associated with Shell Eggs. August 27, 2010. https://www.cdc.gov/salmonella/enteritidis/archive/082710.html (Accessed July 17, 2019).
18. Kuehn BM. Salmonella cases traced to egg producers. *JAMA* 2010; 304 (12): 1316.
19. Laestadius LI, Lagasse LP, Smith KC, *et al*. Print news coverage of the 2010 Iowa egg recall: addressing bad eggs and poor oversight. *Food Policy* 2012; 37 (6): 751–9.
20. Nyachuba DG. Foodborne illness: is it on the rise? *Nutr Rev* 2010; 68 (5): 257–69.
21. Scharff RL. Economic burden from health losses due to foodborne illness in the United States. *J Food Prot* 2012; 75 (1): 123–31.
22. Doyle MP, Erickson MC, Alali W, *et al*. The food industry's current and future role in preventing microbial foodborne illness within the United States. *Clin Infect Dis* 2015; 61: 252–9.
23. Kramer MN, Coto D, Weidner JD. The science of recalls. *Meat Sci* 2005; 71 (1): 158–63.
24. U.S. Department of Agriculture (USDA). Responding to a Food Recall Procedures for Recalls of USDA Foods. 06/24/2014. https://www.fns.usda.gov/responding-food-recall-procedures-recalls-usda-foods (Accessed July 17, 2019)
25. Amazon data source. http://jmcauley.ucsd.edu/data/amazon/links.html. (Accessed July 17, 2019)
26. McAuley J, Pandey R, Leskovec J. Inferring networks of substitutable and complementary products. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015; ACM: 785–94.
27. Mining Online Data for Early Identification of Unsafe Food Products. GitHub Repository. https://github.com/uwescience/DSSG2016-Unsafe-Foods. (Accessed July 17, 2019)
28. Ding C, Li T, Peng W. Nonnegative matrix factorization and probabilistic latent semantic indexing: equivalence chi-square statistic, and a hybrid method. In: AAAI. 2006;42:137–43.
29. Friedman J, Hastie T, Tibshirani R, 2001. The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
30. Chawla NV, Bowyer KW, Hall LO, *et al*. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16: 321–57.
31. Devlin J, Chang M-W, Lee K, *et al*. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv: 181004805; 2018.
32. World Health Organization. WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007–2015 [cited 2016 Feb 8]. World Health Organization, 2015.
33. WHO. *WHO Food Safety. World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/food-safety. Accessed February 6, 2019.
34. Kirk MD, Pires SM, Black RE, *et al*. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLoS Med* 2015; 12: 1–21.
35. Harris JK, Mansour R, Choucair B, *et al*. Health department use of social media to identify foodborne illness-Chicago, Illinois, 2013-2014. *Morb Mortal Wkly Rep* 2014; 63: 681–685.
36. Jenine K, Harris JBH, Nguyen L, *et al*. Using twitter to identify and respond to food poisoning: The Food Safety STL Project. *J Public Health Manag Pract* 2017; 23: 577–580.
37. Harrison C, Jorder M, Stern H, *et al*. Using online reviews by restaurant patrons to identify unreported cases of foodborne illness-New York City, 2012-2013. *MMWR Morb Mortal Wkly Rep* 2014; 63: 441–445.
38. Henly S, Tuli G, Kluberg SA, *et al*. Disparities in digital reporting of illness: a demographic and socioeconomic assessment. *Prev Med* 2017; 101: 18–22.

39. Cesare N, Grant C, Hawkins JB, *et al*. demographics in social media data for public health research: does it matter? arXiv preprint arXiv: 171011048; 2017.
40. Nsoesie EO, Kluberg SA, Brownstein JS. Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports. *Prev Med* 2014; 67: 264–9.
41. Quade P, Nsoesie EO. A platform for crowdsourced foodborne illness surveillance: description of users and reports. *JMIR Public Health Surveill* 2017; 3 (3): e42.
42. Sadilek A, Caty S, DiPrete L, *et al*. Machine-learned epidemiology: real-time detection of foodborne illness at scale. *Npj Digital Med* 2018; 1 (1): 36.