Creating a Bias-Free Dataset of Food Delivery App Reviews with Data Poisoning Attacks

Hyunmin Lee [1] **SeungYoung Oh** [1] **JinHyun Han** [1 2] **Hyunggu Jung** [1]

Abstract

Artificial Intelligence (AI) models have created many benefits and achievements in our time. However, they also have the potential to cause unexpected consequences if the models are biased. One of the reasons why AI models are biased is due to data poisoning attacks. Therefore, it is important for AI model developers to understand how biased their training data is when preparing a training dataset in order to develop fair AI models. While researchers have reported several data sets for the purpose of training datasets, the existing studies have not taken into account the possibility of data poisoning attacks that the dataset may have due to the bias in the dataset. To address this gap, we created and validated a dataset that reflects the possibility of bias in individual reviews of food delivery apps. This work contributes to the community of AI model developers who aim to create fair AI models by proposing a bias-free dataset of food delivery app reviews with data poisoning attacks as an example.

1. Introduction

Over the past few years, researchers have developed artificial intelligence (AI) models that brought many benefits and achievements (Furman & Seamans, 2019). For example, Google has reduced data center cooling costs by up to 40% using AI models (Gao, 2014). However, AI models tend to have negative effects with biased outputs that can have unexpected consequences. For example, Google Photos once had a problem classifying black people as gorillas (Pan & Rajan, 2022). Google Ads targeted advertising tools argued against gender discrimination because it served fewer ads for high-paying jobs to women than to men (Datta et al., 2014).

[1]University of Seoul, Republic of Korea [2]UST21, Republic of Korea. Correspondence to: Hyunggu Jung <hjung@uos.ac.kr>.

AI & HCI Workshop at the 40^{th} International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA. 2023.

Figure 1. An example of the delivery app restaurant review event.

Two types of attacks lead to the creation of biased AI: the data poisoning attack, which poisons the training data to drive AI model learning in the wrong direction, and the model poisoning attack, which distorts the learning algorithm or model structure and leads to incorrect predictions (Goldblum et al., 2022). For example, crowdsourcing systems are vulnerable to data poisoning attacks, where malicious actors inject fake data to disrupt the analytical results of other applications in the cloud (Li et al., 2019). Model poisoning attacks occur when multiple clients are involved in the training of an AI, such as in federated learning. For example, a few malicious clients arbitrarily change values such as parameters in the local model, or intentionally train the local model incorrectly using incorrect training data (Zhou et al., 2021; Muñoz González et al., 2017). Therefore, it is important for AI model developers to be aware of how biased their training data is when selecting the training dataset in order to develop fair AI models. However, Ryan et al. found that machine learning and human-computer interaction experts find it difficult to collect and curate high-quality data due to data bias (Ryan et al., 2023).

An example of a data poisoning attack is restaurant reviews on delivery apps, because some restaurants run “review events” (see Figure 1) to get positive reviews from customers to increase their sales (Choi, 2015). Review events lead customers to write positive reviews to receive additional services through a review event instead of the primary purpose of evaluating the restaurant, thus creating information bias (Choi & Noseworthy, 1992) in the review dataset (Lee et al., 2022).

Existing studies have created various datasets for use in training AI models. For example, Visalli et al. created a review dataset of consumers' sensory and affective perceptions of red wine, which was used to evaluate the influence of culture and expertise on temporal sensory evaluations (Visalli et al., 2023). Also, Sutoyo et al. collected product reviews written in Indonesian languages to assist AI model developers in building a model for emotion classification tasks (Sutoyo et al., 2022). In addition, Boland et al. and Sorgente et al. created a corpus dataset using reviews to support AI model developers in building a sentiment analysis models (Boland et al., 2013; Sorgente et al., 2014). However, although the researchers reported the dataset for the purpose of model training, these studies did not consider the possibility of data poisoning attack that each data may have, which may lead to the creation of biased AI.

To reduce this gap, we created a bias-free dataset where data with possibility of poisoning attacks are labeled. We created our dataset by collecting reviews of the food delivery app reviews that considers the possibility of poisoning in the delivery app review dataset and provide a label about it. Our study contributes to the community of artificial intelligence and human computer interaction as follows: First, we created a dataset consisting of reviews in restaurants that hold review events from Korean food delivery apps. Second, our data has potential to be used for future research to understand food delivery app reviews and how review events influence restaurant reviews. To the best of our knowledge, this is the first study that attempted to create a AI training dataset by filtering data with possibility of poisoning attacks.

2. Related Work

Previous studies reported datasets on consumer reviews (Visalli et al., 2023; Sutoyo et al., 2022; Manzoor et al., 2022; Restrepo et al., 2022; Plotnikov et al., 2020; Li et al., 2020; Barbopoulos & Johansson, 2017; Boland et al., 2013; Sorgente et al., 2014). Several studies controlled their dataset to minimize the causes that have a negative impact on their dataset (Visalli et al., 2023; Restrepo et al., 2022; Li et al., 2020; Boland et al., 2013). For example, three studies performed dataset balancing to prevent the dataset from being heavily weighted toward one class (Visalli et al., 2023; Restrepo et al., 2022; Boland et al., 2013), while the other preprocessed the collected review text before constructing a dataset to minimize bias (Li et al., 2020).

Meanwhile, other researchers collected their reviews without controlling the dataset. Three out of the five studies performed sentiment analysis to provide a sentiment label based on the review text (Sutoyo et al., 2022; Sorgente et al., 2014; Manzoor et al., 2022). One study analyzed the readability of the review text by calculating the Gunning fog index (Plotnikov et al., 2020), and the other created a Consumer Motivation Scale (CMS) by conducting a survey asking Likert scale questions about the importance of each indicator (Barbopoulos & Johansson, 2017).

Previous studies attempted to create a bias-free dataset by addressing information bias (Alkhaled et al., 2023) and selection bias (Orlando et al., 2016; Curtó et al., 2017). For example, Orlando et al. created a Contact Prediction dataset where the selection bias of 3D structure observation is removed (Orlando et al., 2016), and Alkhaled et al. proposed a metric for estimating social bias and proposed a dataset for training models to detect bias (Alkhaled et al., 2023). In addition, Curto et al. created a bias-free dataset containing human faces from different ethnic groups by balancing each ethnic group (Curtó et al., 2017).

Previous studies have three limitations. First, although some prior studies have attempted to control the dataset to minimize the bias or class imbalance of the dataset, they did not attempt to find the bias in each data of their dataset. Second, none of the studies created a dataset of food delivery app reviews or checked whether each review could be written for review events. Third, some studies that attempted to create a bias-free dataset did not account for the data poisoning attacks. To address this gap, we created a dataset of food delivery app reviews that considers the possibility of being poisoned which is made by the review events of the delivery app restaurants.

3. Method

We have created a dataset of customer reviews of restaurants that indicates whether each review has the possibility of being biased. In this section, we describe a method for collecting and labeling such reviews. First, we selected a food delivery app. Then, we selected restaurants in our target delivery app. Finally, we collected and labeled reviews from the selected restaurants. The workflow of our data collection process is shown on Figure 2. The data collection processes were run on a computer with an AMD Ryzen 7 5700U with a Radeon graphics processor running at 1.80 GHz with 64 GB of RAM, running Windows 11 Home.

3.1. Choosing a Target Delivery App

First, we established eligibility criteria for selecting a delivery app to collect restaurant reviews. The criteria are as follows: Does the delivery app have a PC website, and are you able to access the restaurant and review information through that website? With that, we decided to collect de-

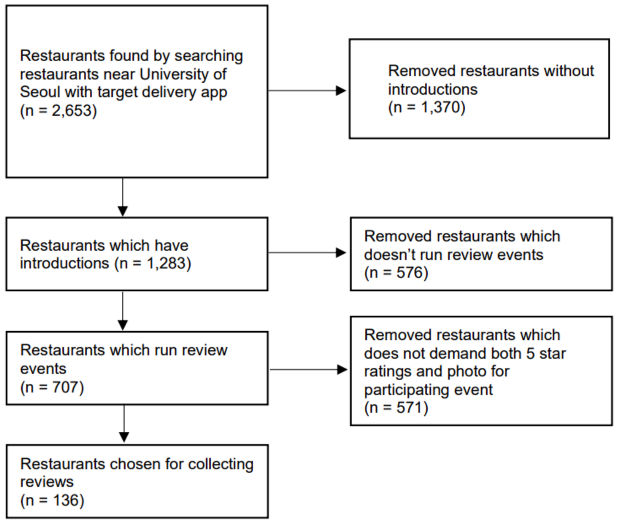

Figure 2. The data collection process of collecting restaurant reviews in the delivery app.

livery app reviews from "Yogiyo"[1], which meets the above criteria.

3.2. Choosing a Target Restaurant

To select target restaurants to collect review data, we first created a list of restaurants by retrieving all restaurants found when the location was set to University of Seoul, and obtained a list of 2,653 restaurants and their IDs. Then, we used the restaurant IDs to collect their reviews.

Second, to determine which of the restaurants on the list had review events, we collected introductions from the restaurants and read them to determine if they had review events. First, we excluded 1299 restaurants with no introductions from the list because we could not determine that the restaurant had review events without introductions. We then checked the text introductions of the remaining 1354 restaurants to see if they were running review events. We found that 643 restaurants had review events. Next, we analyzed the introductions of the restaurants that were found to be running review events and checked the criteria that restaurant owners require from consumers in order to participate in the review event (see Table 1). Some of them are determined by looking at reviews, while others are not. So we decided to consider only the criteria that are determined by looking at reviews (see Table 1).

Based on the results of the review event analysis, we decided to collect review data only from restaurants that demanded every distinguishable criterion. We focused on the restaurants that asked customers to include five-star ratings and a photo of the delivered food item when writing a review of their delivered food. A total of 136 restaurants were selected in this way.

[1] https://www.yogiyo.co.kr/

Table 1. A list of criteria that delivery app users needs to follow to participate in a review event hosted by a restaurant. In order to examine the bias of the review as conservatively as possible, we selected only the two criteria, which were (2) and (6).

Criterion	Selected
Add more than one of the main menu item to your order.	
Give five-star ratings in your review.	✓
Provide identifying information if you wish to participate in the review event.	
Add the restaurant to the user's list of favorite restaurants.	
Indicate that the user wishes to participate in the review event.	
Take a picture of the delivered menu and include it in your review.	✓

3.3. Collecting and Labeling Restaurant Reviews

After selecting the target restaurants, we collected and tagged all the reviews of the restaurants. First, we collected the reviews using metadata that we defined by analyzing the reviews (see Figure 3). See Table 2 for detailed information about the metadata of our dataset. In this way, we collected each review from the target restaurants and represented them as tabular data.

Second, we checked each review data to see if the review met the two criteria which the restaurant required to participate in the review event. We easily checked the "Take a picture of the delivered menu and include it in your review" criterion because we only needed to check whether the review contained a picture. However, in the case of the star rating criterion, although there are four types of star ratings on Yogiyo, restaurants that required a five-star rating did not specify which one they meant. Therefore, we treated the "five star rating" criterion as being met if one or more star ratings were given a value of 5. For reviews that met both criteria, we assigned a value of "0" to the BiasFree column, since these reviews meet the condition for participation in review events, and thus we are not certain that these reviews were not written for review events. For the reviews that do not meet the condition for attending the review event, we gave the BiasFree column a value of "1" because these reviews are not written for review events.

4. Results

The dataset consists of several csv files and is currently available at https://github.com/yunoa64/ICML-Review-Dataset.git. Table 2 and Table 3 describe the specification of the dataset and the metadata used to create the dataset, respectively.

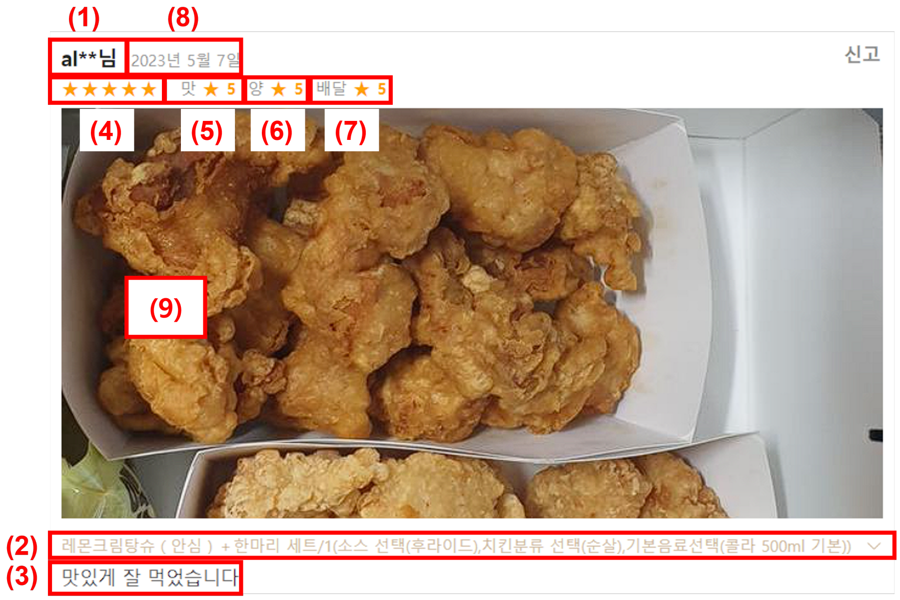

Figure 3. The review example from Yogiyo delivery app. Reviews' metadata are described as red boxes (see Table 3)

.

5. Dataset Validation

In this section, we validate our dataset by creating and comparing two classifiers using a single random restaurant review dataset (1096420.csv[2]) and its original dataset(without the BiasFree column) to show that our dataset is effective in preventing biased datasets. The dataset contains 199 reviews.

Using the unlabeled original review dataset, we built our classifier to classify reviews as positive, negative, and neutral. The review was classified as positive if the average value of star rating value was above 3. And if the average star rating value was below 3, the review was classified as negative. And if the average star rating was equal to 3, the review was classified as neutral. 196 out of the total reviews were classified as positive, two as negative, and the remaining one as neutral (see Figure 4).

When we used the labeled review dataset, we first classified the reviews that were labeled as potentially biased as "not sure". The remaining reviews were classified in the same way as the above. As a result, 146 of the total reviews were classified as "not sure", 50 of them were classified as positive, two of them were classified as negative, and one of them was classified as neutral (see Figure 5).

In summary, we found that about 75% of the reviews that were previously classified as positive should have been classified as "not sure". Although we are not able to know the ground truths about some of those "not sure" reviews, we can say that some of the "not sure" reviews are false positives written for review events, while others are actually true positive reviews. Therefore, if we exclude these "not sure" reviews, it is obvious that this will reduce the number of reviews that the classifier classifies as positive when they are actually negative. Therefore, classifying reviews using our new labeled dataset is more accurate because it excludes uncertain data and makes the number of correctly classified data almost equal to the number of total data, thus increasing the accuracy. This confirms that our proposed dataset is better compared to the original review dataset because it provides AI model developers with safer reviews than the unlabeled, original data (see Table 4).

6. Discussion

We discuss interesting findings from this study. First, we found that none of the previous studies accounted for the poisoning attacks in the data they generated. We wondered if there is a model that considers the bias of the dataset. Previous study have proposed a toolkit called Fairlearn, which helps AI model developers choose the appropriate metrics and mitigation algorithms for their needs (Bird et al., 2020). This response shows that we are able to create a model that takes into account the bias in the data set if we define the appropriate metrics and mitigation algorithms for the models that are created.

[2]https://github.com/yunoa64/ICML-Review-Dataset/blob/master/labeled_reviews/1096420.csv

Table 2. Dataset specifications. Raw data was collected using the Beautiful Soup package. Table columns such as "BiasFree" in the csv file contain the result of data analysis.

Subject	Customer restaurant review
Specific subject type	The data collected relates to the restaurant review data field for a platform specializing in commercial food delivery. The existing restaurant review data has been expanded to include a column indicating whether the restaurant is safe from poisoning attacks.
Data type	Tables
File name	(restaurant ID).csv
How data was acquired	We used data mining techniques – Beautiful Soup is a Python programming language package for parsing HTML documents. And Selenium is the library package in Python that we use to crawl web pages.
Data format	Raw Analyzed
Data collection description	We collect data from the Yogiyo delivery app. Using BeautifulSoup4 and Selenium, we collect reviews by crawling the site and parsing the data we need. There are 136 restaurants with a total of 128668 reviews. The average number of reviews for each restaurant is 964, and the standard deviation is about 1651.
Data source location	Yogiyo delivery app
Data accessibility	Direct URL to data: https://github.com/yunoa64/ICML-Review-Dataset.git

Second, our proposed dataset included four-star ratings and a label indicating whether the review had pictures, which were used to define and identify bias in our dataset. This finding led us to investigate whether restaurants on different delivery apps have events that might bias their reviews, and if so, what part of the review is affected. However, to our knowledge, no previous studies have investigated events from restaurants on food delivery apps other than Yogiyo, but it would be important to clarify how any bias in a review dataset would affect the performance of AI models. It has been reported that it is possible to obtain and reduce bias by trading accuracy, i.e., lowering accuracy to minimize discrimination (Kamiran & Calders, 2012). Thus, if the collected restaurant reviews of the delivery app contain bias, we would lower the accuracy of the trained model to reduce bias. This implies that preparing a bias-free dataset minimizes the bias of the trained model and maximizes the accuracy.

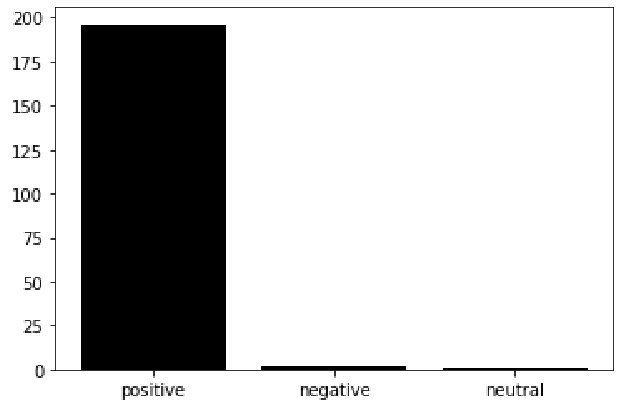

Figure 4. Classification results using an unlabeled review dataset. The bars in the graph indicate the three types of classifications, and the y-axis indicates the number of reviews for each classification result.

7. Limitations and Future Work

This study has several limitations. First, although we collected and filtered the food delivery app reviews, the collected reviews were only from restaurants near University of Seoul, and only Yogiyo app was used. Therefore, the resulting dataset is small and may be biased toward certain locations. Second, though we filtered our dataset using certain criteria, it was not possible to find the ground truth about whether the reviews were written for review events reviews was not possible due to the insufficient information on each review. Also, our method of filtering reviews is critically flawed because it may resul in the removal of organic user data. Third, although we created a dataset, we did not evaluate our dataset by conducting a user study. Fourth, although we created our dataset and made it available to the public, we did not develop a visualization tool to help AI model developers easily obtain the information from our dataset.

To overcome such limitations, we need to expand our dataset by analyzing delivery app restaurants and their reviews from other domains. This will help us understand food delivery app reviews and suggest better criteria for filtering reviews, which will also help us minimize the removal of organic user data and reduce the selection bias of our dataset. In addition, we need to evaluate our dataset by conducting a user study by surveying or interviewing real AI model developers to understand whether our proposed dataset is useful and to get helpful feedback to improve our dataset. As shown in (Lee et al., 2023), it would be helpful to create an interactive visualization tool that would allow AI model developers to understand our dataset and choose which part to use for their model development. It would also be important to evaluate

Table 3. Dataset metadata description.

#	Column Name	Type	Numerical Range	Description
(1)	UserID	object	NaN	The reviewers identifier(ID)
(2)	Menu	object	NaN	The dishes which the customer ordered
(3)	Review	object	NaN	The review text
(4)	Total	numerical	1 ∼5	Star rating to rate the overall service
(5)	Taste	numerical	1 ∼ 5	Star rating to rate the taste of the food delivered
(6)	Quantity	numerical	1 ∼ 5	Star rating to rate the quantity of the food delivered
(7)	Delivery	numerical	1 ∼ 5 or NA	Star rating for rating the delivery services, which has a value of 1∼5 for delivery services, and "NA" for the packing services
(8)	Date	datetime64[ns]	NaN	The date when the review was written
(9)	HasPicture	boolean	0 ∼1	An indicator that indicates whether the food image is included in the review
(10)	BiasFree	boolean	0 ∼1	Boolean value indicating whether the review is free from the risk of review event
(11)	RestaurantID	numerical	NaN	The ID of the source restaurant of the review data

Table 4. The classification results on original dataset and our modified dataset. The column names indicate labels of reviews on each dataset, and values are the number of reviews on each classification results. The results indicate that about 146 out of the 196 reviews were previously classified as positive, which were supposed to be classified as "not sure."

	Positive	Negative	Neutral	Not Sure
Original dataset	196	2	1	0
Our dataset	50	2	1	146

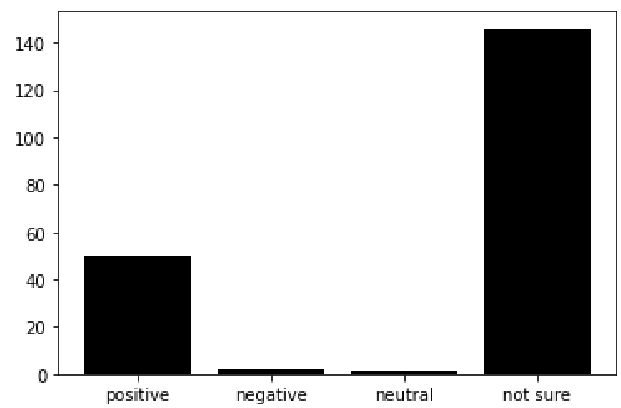

Figure 5. Classification results using a labeled review dataset. The bars in the graph indicate the four types of classifications, and the y-axis indicates the number of reviews for each classification result.

the usability and feasibility of the potential visualization tool with stakeholders that include AI model developers.

8. Conclusion

This study aims to contribute to the communities of AI and HCI researchers by creating a novel dataset that enables them to create a fair AI model. In order to create a bias-free dataset with a food delivery app reviews, we filter out the reviews that have the possibility of being written for the review events. Our proposed dataset is practical for future studies on food delivery app reviews, especially the impact of review events on customer reviews. We hope that our study will inspire AI and HCI researchers to develop fair AI technologies by avoiding data poisoning attacks, using datasets that are safe from bias to prevent data poisoning attacks.

Acknowledgements

We appreciate HCAIL members for their constructive feedback on our initial manuscript.

References

Alkhaled, L., Adewumi, T., and Sabry, S. S. Bipol: A novel multi-axes bias evaluation metric with explainability for nlp. *arXiv preprint arXiv:2304.04029*, 2023.

Barbopoulos, I. and Johansson, L.-O. The consumer motivation scale: A detailed review of item generation, exploration, confirmation, and validation procedures. *Data in Brief*, 13:88–107, 2017. ISSN

2352-3409. doi: https://doi.org/10.1016/j.dib.2017.04.054. URL https://www.sciencedirect.com/science/article/pii/S2352340917301865.

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., and Walker, K. Fairlearn: A toolkit for assessing and improving fairness in ai. *Microsoft, Tech. Rep. MSR-TR-2020-32*, 2020.

Boland, K., Wira-Alam, A., and Messerschmidt, R. *Creating an Annotated Corpus for Sentiment Analysis of German Product Reviews*, volume 2013/05 of *GESIS-Technical Reports*. GESIS - Leibniz-Institut für Sozialwissenschaften, Mannheim, 2013. URL https://nbn-resolving.org/urn:nbn:de:0168-ssoar-339398.

Choi, B. C. and Noseworthy, A. L. Classification, direction, and prevention of bias in epidemiologic research. *Journal of Occupational Medicine*, pp. 265–271, 1992.

Choi, S. Consumers' propensity of post fake online reviews. Master's thesis, The Pennsylvania State University, 2015.

Curtó, J. D., Zarza, I. C., De La Torre, F., King, I., and Lyu, M. R. High-resolution deep convolutional generative adversarial networks. *arXiv preprint arXiv:1711.06491*, 2017.

Datta, A., Tschantz, M. C., and Datta, A. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *arXiv preprint arXiv:1408.6491*, 2014.

Furman, J. and Seamans, R. Ai and the economy. *Innovation policy and the economy*, 19(1):161–191, 2019.

Gao, J. Machine learning applications for data center optimization, 2014.

Goldblum, M., Tsipras, D., Xie, C., Chen, X., Schwarzschild, A., Song, D., Mądry, A., Li, B., and Goldstein, T. Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (2):1563–1580, 2022.

Kamiran, F. and Calders, T. Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1):1–33, Oct 2012. ISSN 0219-3116. doi: 10.1007/s10115-011-0463-8. URL https://doi.org/10.1007/s10115-011-0463-8.

Lee, H., Chun, M., Jeong, H., and Jung, H. Toward fairness in artificial intelligence: Uncovering food delivery app reviews with poisoning attacks. *Communications of the Korean Institute of Information Scientists and Engineers*, 40(11):44–52, 2022. ISSN 1229-6821. URL http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11157439.

Lee, H., Chun, M., and Jung, H. Porde: Explaining data poisoning attacks through visual analytics with food delivery app reviews. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23 Companion, pp. 46–50, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701078. doi: 10.1145/3581754.3584128. URL https://doi.org/10.1145/3581754.3584128.

Li, J., Lowe, D., Wayment, L., and Huang, Q. Text mining datasets of β-hydroxybutyrate (BHB) supplement products' consumer online reviews. *Data in Brief*, 30:105385, 2020. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2020.105385. URL https://www.sciencedirect.com/science/article/pii/S2352340920302791.

Li, M., Sun, Y., Lu, H., Maharjan, S., and Tian, Z. Deep reinforcement learning for partially observable data poisoning attack in crowdsensing systems. *IEEE Internet of Things Journal*, 7(7):6266–6278, 2019.

Manzoor, M. F., Abid, A., Nawaz, N. A., and Alvi, A. Aspect based sentence segregated dataset of hybrid car's consumers online reviews. *Data in Brief*, 42:108293, 2022. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2022.108293. URL https://www.sciencedirect.com/science/article/pii/S2352340922004954.

Muñoz González, L., Biggio, B., Demontis, A., Paudice, A., Wongrassamee, V., Lupu, E. C., and Roli, F. Towards poisoning of deep learning algorithms with back-gradient optimization. In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, AISec '17, pp. 27–38, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450352024. doi: 10.1145/3128572.3140451. URL https://doi.org/10.1145/3128572.3140451.

Orlando, G., Raimondi, D., and Vranken, W. Observation selection bias in contact prediction and its implications for structural bioinformatics. *Scientific Reports*, 6(1): 36679, 2016.

Pan, R. and Rajan, H. Decomposing convolutional neural networks into reusable and replaceable modules. In *Proceedings of the 44th International Conference on Software Engineering*, pp. 524–535, 2022.

Plotnikov, A., Shcheludyakov, A., Cherdantsev, V., Bochkarev, A., and Zagoruiko, I. Data on post bank customer reviews from web. *Data* *in Brief*, 32:106152, 2020. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2020.106152. URL https://www.sciencedirect.com/science/article/pii/S2352340920310465.

Restrepo, F., Mali, N., Sands, L. P., Abrahams, A., Goldberg, D. M., White, J., Prieto, L., Ractham, P., Gruss, R., Zaman, N., and Ehsani, J. P. Injury prevention for older adults: A dataset of safety concern narratives from online reviews of mobility-related products. *Data in Brief*, 42:108044, 2022. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2022.108044. URL https://www.sciencedirect.com/science/article/pii/S2352340922002554.

Ryan, S., Nadal, C., and Doherty, G. Integrating fairness in the software design process: An interview study with hci and ml experts. *IEEE Access*, 11:29296–29313, 2023. doi: 10.1109/ACCESS.2023.3260639.

Sorgente, A., Vettigli, G., and Mele, F. *An Italian corpus for aspect based sentiment analysis of movie reviews*, pp. 349–353. Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014 : 9-11 December 2014, Pisa. Pisa University Press, Pisa, 2014. ISBN 9788867414727. URL http://digital.casalini.it/9788867414727. P. 349-353.

Sutoyo, R., Achmad, S., Chowanda, A., Andangsari, E. W., and Isa, S. M. Prdect-id: Indonesian product reviews dataset for emotions classification tasks. *Data in Brief*, 44:108554, 2022. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2022.108554. URL https://www.sciencedirect.com/science/article/pii/S2352340922007612.

Visalli, M., Dubois, M., Schlich, P., Ric, F., Cardebat, J.-M., and Georgantzis, N. A dataset on the sensory and affective perception of bordeaux and rioja red wines collected from french and spanish consumers at home and international wine students in the lab. *Data in Brief*, 46:108873, 2023. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2022.108873. URL https://www.sciencedirect.com/science/article/pii/S2352340922010769.

Zhou, X., Xu, M., Wu, Y., and Zheng, N. Deep model poisoning attack on federated learning. *Future Internet*, 13(3):73, 2021.