ISSN 2614-0047

Twitter sentiment analysis about economic recession in indonesia

Fauzan Prasetyo Eka Putra [a,1,*], Fairuz Iqbal Maulana [b,2], Nawawi Muhammad Akbar [c,3], Wicaksono Febriantoro [d,4]

[a] Faculty Of Engineering, informatic program study, Panglegur, Pamekasan, Indonesia, 69313
[b] Department Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
[c] Master of Law Science and Development, Postgraduate School, Airlangga University, Surabaya, Indonesia 60115
[d] Knowledge Lab, University College London (UCL), United Kingdom
[1] prasetyo@unira.ac.id; [2] fairuz.maulana@binus.edu; [3] nawawi.muhammad.akbar-2021@pasca.unair.ac.id; [4] wicaksono.febriantoro.21@ucl.ac.uk
* corresponding author

ARTICLE INFO

Article history
Received January 3, 2023
Revised February 18, 2023
Accepted February 24, 2023

Keywords
Sentiment analysis
Twitter
Perspective
Indonesia
Recession

ABSTRACT

As one of the most popular social media platforms, Twitter enables users to express their opinions on diverse concepts, products, and services. Large quantities of data shared as tweets can be mined for user feedback and used to improve the quality of products and services. Using Twitter data and social media sentiment analysis, tracking how people feel about the recession in real time is possible. As a consequence, relevant organizations or governments can take preventative measures against the disinformation and unlawful conduct caused by the effects of the recession. This study aims to determine if there is a correlation between how people on Twitter feel about the recession. This study's data acquisition utilized "Recession"-tagged Twitter remarks from 2023. This study analyses filtered tweets for sentiment, emotion, word usage, and trends. According to the findings, 94% of tweets had benign sentiments, 4% had positive sentiments, and 2% had negative sentiments. Tweets with moderate subjective valence cluster in the middle of the polarity scale (between 1 and +1), while tweets with strong subjective valence are dispersed throughout the scale.

1. Introduction

Text mining is a data mining technique that can unearth relevant information concealed within text data sets. On social media, text mining techniques, such as mood analysis on Twitter data, can be applied. Text mining techniques include information extraction, document categorization, document retrieval, and clustering [1], [2]. The categorization technique of sentiment analysis is guided learning. Positive, negative, and neutral are the three categories applicable to studies of emotion. Technically, the mood class is determined by determining the orientation value. A positive polarity value represents a positive emotion, while a negative polarity value represents a negative opinion. A polarity number of zero indicates a sense of neutrality [3]. TextBlob is one of the commonly used instruments for sentiment analysis. TextBlob excels at processing mood analysis data in English [4]–[6]. Due to the complexity of the structure of the Indonesian language, text mining with Indonesian data sources still has room for improvement. The Nazief and Adriani Algorithm is one of the text-mining techniques utilized for the analysis of Indonesian data [7]. The modern era has seen a continuous development of text mining (TM) software. Artificial intelligence (AI) techniques have been the subject of numerous research articles in the medical field [8], [9], social survey [10], agriculture [11], including machine learning (ML) [12], online education [13], and predictive analytics [14].

The majority of existing studies employ a technique known as machine learning, in which annotated data is used to analyze sentiment, and the concentration is on improving the performance

https:doi.org/10.31763/businta.v7i1.592

of models [13], [15]. On the other hand, this study proposes a novel method for analyzing sentiment by integrating machine learning (ML) techniques with the TextBlob model. [2], [4], [5] that is based on a dictionary. Before, only a few methods, mostly machine learning models, were used to improve accuracy, and the characteristics of datasets were not taken into account. There is a chance that the annotations are wrong because tweets that are marked as positive could actually be neutral or positive. Because of this, using these kinds of data with machine learning models can hurt how well the models work. Considering this assumption, this study looks at how well TextBlob works for sentiment analysis relative to the initial dataset. Despite previous attempts to improve model performance via hyperparameter settings, model design optimization, preprocessing pipelines, and feature extraction and selection techniques, the models did not substantially or appreciably improve.

Academics are interested in learning more about the effects of the COVID-19-caused economic recession on various socioeconomic phenomena in people's lives. For instance, investigating the effect of the COVID-19 economic recession on aspects of mental health, chronic diseases, life span, and natural mortality [16] or how the recession affected the type of macroeconomic policy that should be implemented to lessen the effects of the recession [17]. Other studies that COVID-19 should investigate predict that the pandemic could cause a worldwide food Recession [18]. Most of the research in this study focused on comprehensive Twitter sentiment analysis, focusing their research on the discussion regarding the Recession. Different topics are associated with various sentiment polarities, implying that sentiment analysis alone cannot disclose much without topic modeling. As a result, the researcher incorporates subject modeling into our study.

2. Method

Researchers obtained Tweet data from Twitter for this study by crawling on social media Twitter. The keywords for crawling used Indonesian input, namely “recession”, and were taken on February 28, 2023. In this study, various aspects of topic modeling and sentiment analysis will be investigated in relation to the Recession. Fig. 1 depicts the method devised by researchers to analyze tweet data and determine the sentiments of Twitter users.

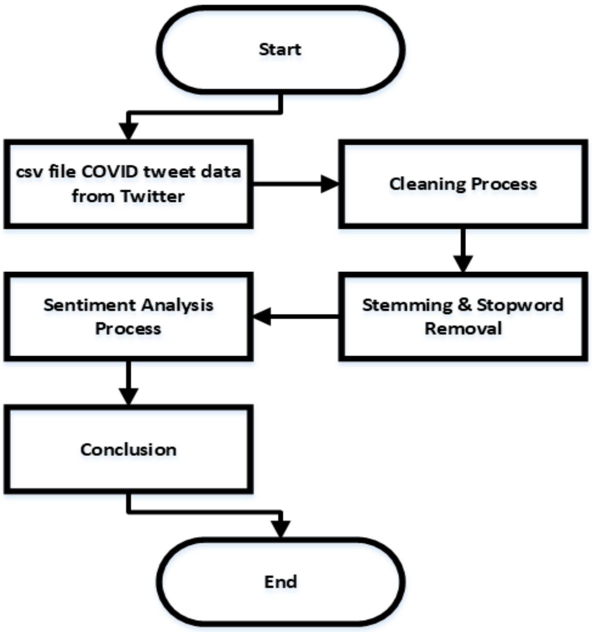

Fig. 1. Workflow of Twitter Opinion with Sentiment Analysis

The data is processed using Google Colab and the Python programming language. The first stage is to scrape data from Twitter, then export it to a CSV file, such as making the data in the CSV file machine-readable through custom programming. Second, the CSV format file data will go through a cleaning process, such as removing “RT, #, \, @, hyperlinks, and emoji, to leading and trailing whitespaces”. Researchers will then ascertain how users feel and categorize their opinions. Next, visualizations can be used to check sentiment on various elements.

2.1. Dataset

The information was acquired by scraping Tweets on Twitter social media using Python programming on Google Colab on February 28, 2023, and exporting as a CSV file. This file contains 1808 tweets from the last few weeks. Table 1 displays a subset of the CSV file data.

Table.1 Sample tweets

DateTime	Tweet ID	Tweet	Username
2023-02-27 22:22:30	1630332575637200896	Jujur ga punya kenangan interaksi secara langsung dengan event Beyond The Summit. Tapi, merasa sangat beruntung bisa interview GodZ di Maret 2021. Masih ga percaya Beyond The Summit kena efek resesi sehebat ini 💔	rosesagaming
2023-02-23 23:15:13	1628896290338914304	indonesia konsumtif tp bagus biar g resesi	adekhoky
2023-02-23 19:15:56	1628836072842199040	ujan ujan rebahan nonton drama uwu ☒ ujan ujan mikirin resesi global ✅ https://t.co/dZJ87kmUS6	iamlilboo

2.2. Data Preprocessing

Ensuring that our data is machine-readable is a crucial aspect of data analysis. Unlike humans, machines can only process binary data., making it challenging to comprehend human language, let alone images and films, and analyzing 1 and 0 as data requires many stages. To be used, the data must first undergo a data cleansing procedure, including turning the raw data into a computer-readable file. The researcher must delete the extensive textual data collection comprising tweets to remove any potential discrepancies. Unwanted information, such as URL connections, user comments, and unescaped HTML characters, is contained in the raw textual data and is not required for mood analysis. We straightforwardly sanitize data. We screen out irrelevant Twitter data during the preprocessing step. We deleted "RT, #, @, hyperlinks, and emoji, to preceding and following whitespaces" because our method of mood analysis does not require verifying extra information.

2.3. Sentiment Analysis

Semantic orientation evaluates texts' polarity and partiality, while sentiment analysis determines their affective tone. In this tweet, directional nouns and adverbs show logical flow. Adverbs divided by words determine emotional direction. Developers use TextBlob to analyze Twitter data and opinions. TextBlob ranks tweets numerically. TextBlob offers text hash tags, polarity, and opinion. Polarity is [-1.1], where -1 indicates negative emotion, and 1 signifies optimism. Negative words become negative figures. The human experience is 0–1. A tweet's biased grade shows how much is the opinion and how much is fact. "Polarity" describes how deeply people feel about views. The outcomes may be positive or negative. Deeply moved by positive emotions such as admiration, faith, or love, a person or entity will adopt a particular worldview. Subjectivity is a person's connection to something. Regardless of others' opinions, emotional commitment and unique contact with the item are key. Sample sentiment analysis based on polarity and subjectivity score as show in Table 2.

Table.2 Sample sentiment analysis based on polarity and subjectivity score

Tweet	Polarity Score	Subjectivity Score	Sentiment
"Jujur ga punya kenangan interaksi secara langsung dengan event Beyond The Summit. Tapi, merasa sangat beruntung bisa interview GodZ di Maret 2021. Masih ga percaya Beyond The Summit kena efek resesi sehebat ini"	0.00	0.00	Neutral
"Terindikasi & disinyalir bahkan banyak yg terpaksa menjual aset2nya (tansh, ruko, toko, kendaraan, ternak, kayu, dll) untuk mendapatkan dana segar. Banyak kredit bank yg macet dibekukan krn Kondisi yg memaksa. Sisi lain sibuk poksan-paksin & propaganda idu resesi ekonomi global"	-0.05	0.05	Negative
"G I truly did the best I could."	1.00	0.03	Positive

3. Results and Discussion

The results show that from the data file used, an average of more than 1699 tweets shows a neutral sentiment. Whereas for positive results, around 72 tweets were obtained and the rest were negative sentiments, as shown in Fig. 2. It describes that 94% of tweets with neutral sentiments, 4% with positive sentiments, and 2% with negative sentiments.

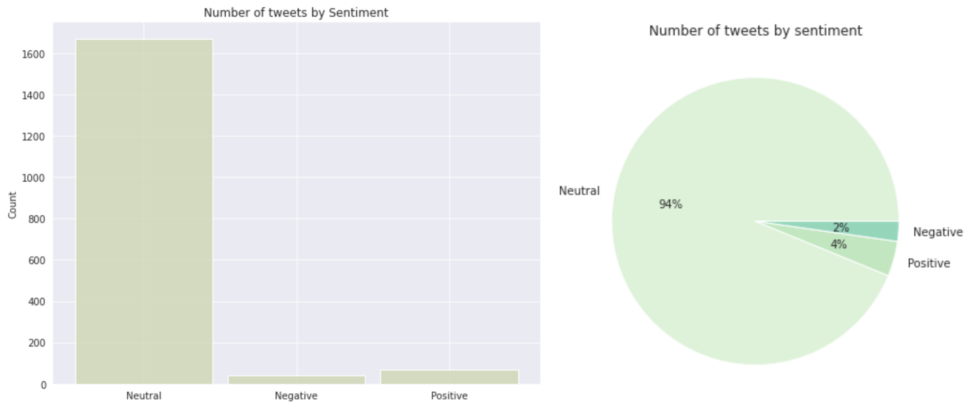

Fig. 2. The Bargraph and Piechart of Sentiment Analysis

Then, the researcher wants to know the word cloud visualization based on the keyword recession and the data files obtained. Fig. 3 depicts a word cloud comprising filtered tweets about the term recession. A word cloud is a graphical depiction of a data collection that emphasizes the regularity with which a specific term or sentence is used.

Fig. 3. Word Cloud of Tweet Posts Related to Recession

As shown in Fig. 4 and Fig. 5, the terms in the cloud collectively reflect the most frequently used words. A bigger font size indicates that the term is stated more frequently than a smaller font size. In summary, the word cloud can wrap up Twitter discussions about the keyword Recession. It can also help us comprehend what a recession is. Our filtered tweet word cloud frequently discovers instances of Indonesian and Turkish that occurred just before this tweet was crawled, as well as many other terms.

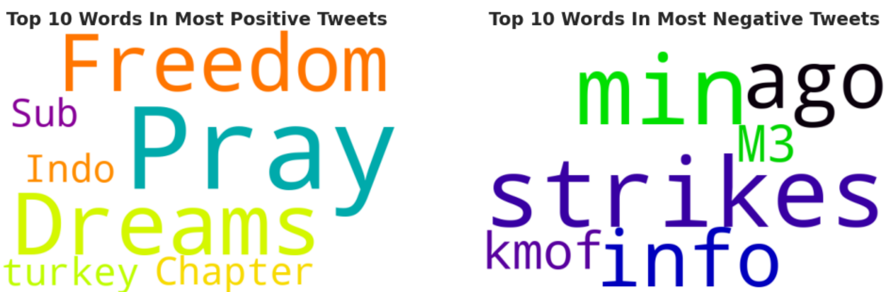

Fig. 4. Word Cloud of Tweet Posts with the Top 10 Words in the Most Positive and Negative Tweets

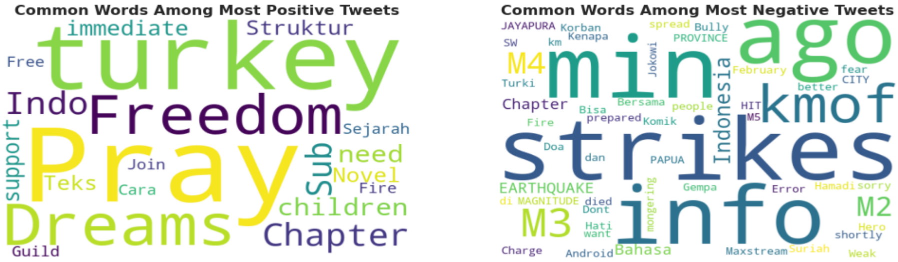

Fig. 5. Word Cloud of Tweet Posts with Common Words among Most Positive and Negative Tweets

Next, the researcher wants to measure polarity and subjectivity, as in Formula 1. The polarity of a tweet is determined by adding the number of the chosen characteristics to the message. A rating is assigned to each tweet based on the following criteria:

$$PolarityScore(tweet) = \sum featurevalue(tweet_i) \quad (1)$$

Based on the subjectivity and polarity values from the data file, Fig. 6 portrays polarity and subjectivity graphs for every tweet ever sent. The color blue represents neutral emotion, while the color green represents negative emotion, and the color yellow represents positive emotion. Positive sentiment is more widespread than negative sentiment around the world. It suggests that our collection of tweets has a broad subjective range, with the majority of tweets falling on the [-1.00, 1.00] polarity scale (negative or positive).

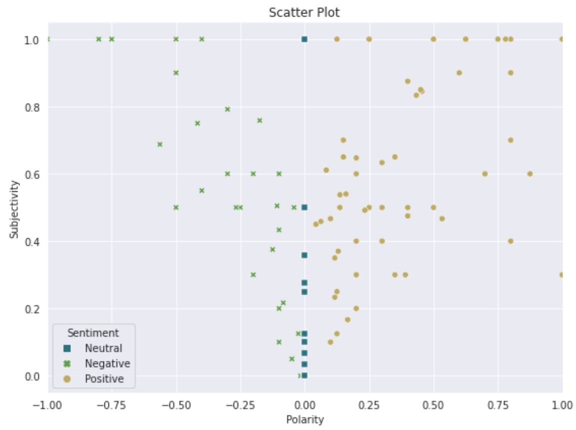

Fig. 6. Polarity and Subjectivity Scatter Graph

A correlation heatmap is a graphical depiction of a data set's correlation matrix between clusters of factors. The association value is represented as a hue in the heat map, with each cell representing a set of factors. The heatmap in Fig 7 represents a symmetric grid with correlation values ranging from -1 to 1, with -1 indicating a perfect negative correlation, 0 indicating no correlation, and 1 indicating a perfect positive correlation. The strength and orientation of the correlation value are reflected in the cell colors, with warmer colors showing positive correlations and colder colors signaling negative correlations (such as blue and green). The darker the color, the stronger the link.

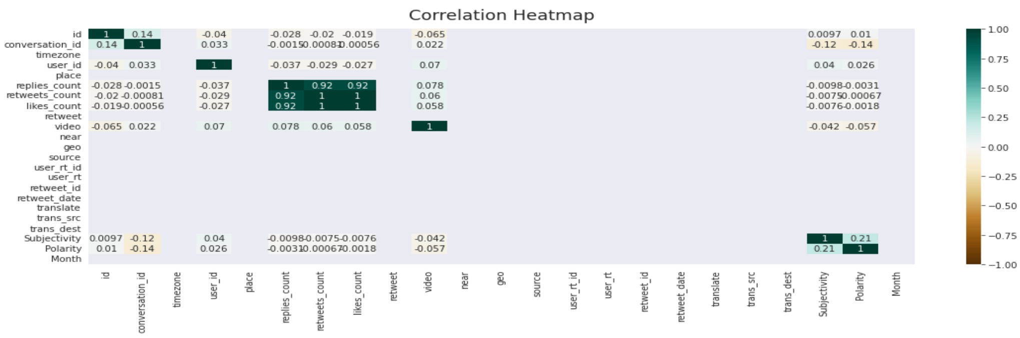

Fig. 7. Heatmap of correlations based on tweets mentioning the recession

Furthermore, the scholar notices that some variables have a strong positive connection, whereas others have a strong negative correlation. Heat maps can help you find patterns and connections in data sets, which can help you make data-driven choices or create predictive models.

For this study, researchers collected and analysed Recession Tweets in Indonesian to determine its prevalence and impact. According to the findings, different individuals have expressed interest in Recession Tweets both directly and through the media. Various social media platforms continue to facilitate the dissemination of Recession Tweet statistics. As a result, the researcher requires cutting-edge computational tools and methods to assess vast quantities of data promptly. In an effort to combat the dissemination of false information on social media, sentiment analysis was developed. Several sectors will benefit from considering consumer insights when developing new policies. Due to the vast amount of data available on social media platforms like Twitter, there is a growing need for a systematic and efficient method of analyzing tweets. This study will benefit numerous industries because it provides a snapshot of the facts on the ground, which is crucial when advocating for legislative or regulatory reform.

4. Conclusion

Based on Twitter statistics in Indonesian, this research examines social media sentiment with the keyword Recession. According to this study, popular opinion on Twitter about the recession is overwhelmingly neutral. This study classified Twitter using text blobs, and the researchers successfully displayed data using a word cloud. With 1808 tweets collected, the sentiment study yielded 94% indifferent sentiment, 4% positive sentiment, and 2% negative sentiment. This study demonstrates that opinion analysis on social media can be used to monitor real-time recession trends. This discovery will presumably aid future mathematical modeling and data analysis research, particularly large-scale data processing using sentiment analysis to monitor responses and public statements via social media, specifically Twitter.

Acknowledgment

The researcher would like to thank everyone who has helped and backed our study “Research on the Perspective of Internet Public Opinion in Social Media Twitter about Recession in Indonesia Using Sentiment Analysis”. Additionally, the researcher would like to express our sincere appreciation to our supervisor for their invaluable guidance and support, to the participants who provided us with data and insights, to the research assistants and technical support team for their expertise and assistance, and to our friends and family for their unwavering support throughout the research process. We are extremely grateful for everyone’s efforts, which made this study feasible.

References

[1] L. Feng, Y. K. Chiam, and S. K. Lo, “Text-Mining Techniques and Tools for Systematic Literature Reviews: A Systematic Literature Review,” in *2017 24th Asia-Pacific Software Engineering Conference (APSEC)*, Dec. 2017, pp. 41–50, doi: 10.1109/APSEC.2017.10.

[2] R. Bose, S. Aithal, and S. Roy, “Sentiment Analysis on the Basis of Tweeter Comments of Application of Drugs by Customary Language Toolkit and TextBlob Opinions of Distinct Countries,” *Int. J. Emerg. Trends Eng. Res.*, vol. 8, pp. 3684–3696, 2020, doi: 10.30534/ijeter/2020/129872020.

[3] M. Arora and V. Kansal, “Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis,” *Soc. Netw. Anal. Min.*, vol. 9, no. 1, p. 12, Dec. 2019, doi: 10.1007/s13278-019-0557-y.

[4] R. Hermansyah and R. Sarno, “Sentiment Analysis about Product and Service Evaluation of PT Telekomunikasi Indonesia Tbk from Tweets Using TextBlob, Naive Bayes & K-NN Method,” in *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2020, pp. 511–516, doi: 10.1109/iSemantic50169.2020.9234238.

[5] W. Aljedaani *et al.*, “Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry,” *Knowledge-Based Syst.*, vol. 255, p. 109780, Nov. 2022, doi: 10.1016/j.knosys.2022.109780.

[6] S. Zahoor and R. Rohilla, “Twitter Sentiment Analysis Using Lexical or Rule Based Approach: A Case Study,” in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization*

(Trends and Future Directions) (ICRITO), Jun. 2020, pp. 537–542, doi: 10.1109/ICRITO48877.2020.9197910.

[7] D. Soyusiawaty, A. H. S. Jones, and N. L. Lestariw, "The Stemming Application on Affixed Javanese Words by using Nazief and Adriani Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 771, no. 1, p. 012026, Mar. 2020, doi: 10.1088/1757-899X/771/1/012026.

[8] S. S. Sohail, M. M. Khan, and M. A. Alam, "An Analysis of Twitter Users From The Perspective of Their Behavior, Language, Region and Development Indices -- A Study of 80 Million Tweets," *Social and Information Networks*, p. 10, May 2021. [Online]. Available at: https://arxiv.org/abs/2105.10245.

[9] S. Das and A. K. Kolya, "Predicting the pandemic: sentiment evaluation and predictive analysis from large-scale tweets on Covid-19 by deep convolutional neural network," *Evol. Intell.*, vol. 15, no. 3, pp. 1913–1934, 2022, doi: 10.1007/s12065-021-00598-7.

[10] M. Lepelaar *et al.*, "Sentiment Analysis of Social Survey Data for Local City Councils," *J. Sens. Actuator Networks*, vol. 11, no. 1, pp.1-13, 2022, doi: 10.3390/jsan11010007.

[11] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100019, 2021, doi: 10.1016/j.jjimei.2021.100019.

[12] N. Leelawat *et al.*, "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," *Heliyon*, vol. 8, no. 10, p. e10894, 2022, doi: 10.1016/j.heliyon.2022.e10894.

[13] M. Mujahid *et al.*, "Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19," *Appl. Sci.*, vol. 11, no. 18, pp. 1-25 2021, doi: 10.3390/app11188438.

[14] N. Chintalapudi, G. Battineni, M. Di Canio, G. G. Sagaro, and F. Amenta, "Text mining with sentiment analysis on seafarers' medical documents," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100005, 2021, doi: 10.1016/j.jjimei.2020.100005.

[15] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model," *Comput. Intell.*, vol. 37, no. 1, pp. 409–434, 2021, doi: 10.1111/coin.12415.

[16] M. H. Brenner, "Will There Be an Epidemic of Corollary Illnesses Linked to a COVID-19-Related Recession?," *American journal of public health*, vol. 110, no. 7. United States, pp. 974–975, Jul. 2020, doi: 10.2105/AJPH.2020.305724.

[17] H. Chernick, D. Copeland, and A. Reschovsky, "The Fiscal Effects of the COVID-19 Pandemic on Cities: An Initial Assessment," *Natl. Tax J.*, vol. 73, pp. 699–732, 2020, doi: 10.17310/ntj.2020.3.04.

[18] S. Goetz, C. Schmidt, L. Chase, and J. Kolodinsky, "Americans' Food Spending Patterns Explain Devastating Impact of COVID-19 Lockdowns on Agriculture," *J. Agric. Food Syst. Community Dev.*, vol. 9, no. 3, pp. 31–33, May 2020, doi: 10.5304/jafscd.2020.093.033.