Makara Human Behavior Studies in Asia

Volume 27 | Number 2 Article 5

12-19-2023

Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach

Fery F.
Faculty of Economics and Business, Universitas Padjadjaran, Jatinangor 45363, Indonesia, fery19001@mail.unpad.ac.id

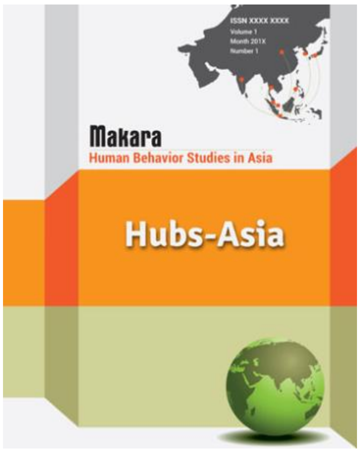
Sunu Widianto
Faculty of Economics and Business, Universitas Padjadjaran, Jatinangor 45363, Indonesia, sunu.widianto@unpad.ac.id

Recommended Citation
F., F., & Widianto, S. (2023). Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach. *Makara Human Behavior Studies in Asia, 27*(2). https://doi.org/10.7454/hubs.asia.1191223

MAKARA
Human Behavior Studies in Asia

eISSN: 2406-9183
pISSN: 2355-794X

Author's Accepted Manuscript



Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach

Fery & Sunu Widianto

DOI : https://doi.org/10.7454/hubs.asia.1191223
To appear in : Makara Human Behavior Studies in Asia

Received: 15 September 2023
Revised: 16 December 2023
Accepted: 19 December 2023

Cite this article as: Fery, F., & Widianto, S. (2023). Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach. *Makara Human Behavior Studies in Asia, 1-13.* https://doi.org/10.7454/hubs.asia.1191223

MAKARA
Human Behavior Studies in Asia

eISSN: 2406-9183
pISSN: 2355-794X

Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach

Analisis Karakteristik Pengguna Twitter dalam Penggunaan Teks dan *Hashtag* selama *Tech Winter Layoff* Pasca COVID-19 dengan Menggunakan Pendekatan LDA dan K-*Means Clustering*

Fery* & Sunu Widianto
Faculty of Economics and Business, Universitas Padjadjaran, Jatinangor 45363, Indonesia

ABSTRACT
Post-COVID-19 pandemic has significantly impacted the global economy, resulting in a surge of job losses and layoffs across various industries, including the technology sector. The pandemic has led to changes in consumer behavior, supply chain disruptions, and an overall decrease in demand, all of which have contributed to the current economic situation. With the rise of social media platforms, individuals have been using Twitter to express their thoughts and opinions on the impact of the pandemic on the technology industry, including the increase in job losses and layoffs. In this study, we analyze the characteristics of Twitter users and their text and hashtag usage in the context of the pandemic's impact on the technology industry. We employ topic modeling and K-means clustering to a preprocessed dataset of tweets related to tech layoffs to identify common themes or topics in Twitter users' responses to tech winter layoffs in Indonesia. The analysis revealed a high number of negative tweets expressing anger and sadness. The use of predetermined keywords did not provide a comprehensive understanding of the phenomenon as other topics such as politics, religion, news, and advertisements were prevalent. This study revealed three thematic clusters. Cluster 1 centers on technology, Cluster 2 on politics and social aspects, and Cluster 3 on healthcare and employment. Latent Dirichlet Allocation (LDA) further refines the understanding, uncovering key themes within the broader discourse on tech layoffs.

ABSTRAK
Pasca pandemi COVID-19 telah berdampak signifikan pada ekonomi global, mengakibatkan peningkatan pemutusan hubungan kerja di berbagai industri, termasuk sektor teknologi. Pandemi ini telah menyebabkan perubahan perilaku konsumen, gangguan rantai pasokan, dan penurunan permintaan secara keseluruhan, yang semuanya telah berkontribusi pada situasi ekonomi saat ini. Dengan munculnya platform media sosial, individu menggunakan Twitter untuk mengungkapkan pemikiran dan pendapat mereka tentang dampak pandemi pada industri teknologi, termasuk peningkatan pemutusan hubungan kerja. Dalam studi ini, peneliti menganalisis karakteristik pengguna Twitter dan penggunaan teks dan hashtag mereka dalam konteks dampak pandemi terhadap industri teknologi. Peneliti menggunakan topic modeling dan k-means clustering untuk kumpulan data tweet yang telah diproses terkait pemutusan hubungan kerja untuk mengidentifikasi tema atau topik umum dalam tanggapan pengguna Twitter terhadap tech winter layoff di Indonesia. Analisis ini mengungkapkan jumlah tweet negatif yang tinggi yang menyuarakan kemarahan dan kesedihan. Penggunaan kata kunci yang telah ditentukan sebelumnya tidak memberikan pemahaman yang komprehensif tentang fenomena tersebut karena topik lain seperti politik, agama, berita, dan iklan lebih mendominasi. Studi ini mengungkapkan tiga klaster tematik. Klaster 1 berfokus pada teknologi, Klaster 2 pada politik dan aspek sosial, dan Klaster 3 pada kesehatan dan ketenagakerjaan. Latent Dirichlet Allocation (LDA) lebih mempertajam pemahaman dengan mengidentifikasi tema-tema kunci dalam wacana

Original Article

***Correspondence Author:**
Fery
E-mail: fery19001@mail.unpad.ac.id

Received: 15 September 2023
Revised: 16 December 2023
Accepted: 19 December 2023

Keyword: clustering, post-COVID-19, layoffs, Twitter, topic modeling

Cite this article: Fery, F., & Widianto, S. (2023). Examining Characteristics on Twitter Users' Text and Hashtag Utilization During Tech Winter Layoff Post-COVID-19 Using LDA and K-Means Clustering Approach. *Makara Human Behavior Studies in Asia*, 1-13. https://doi.org/10.7454/hubs.asia.1191223

lebih luas tentang pemutusan hubungan kerja di bidang teknologi.

1. Introduction

The impact of the COVID-19 pandemic on the economy has been extremely harsh in Indonesia. In Q1 2020, there was a noticeable decrease in the year-on-year gross domestic product (GDP) growth, dropping from 5% in Q4 2019 to 3% and even more visible in Q2 2020 which contracted the economy by 5.3%. People in Indonesia are restricting their consumption, which is having a significant impact on the economy. Household consumption expenditure contributes around 57% of the total GDP, and as a result of reduced consumption, the sectoral GDP is reflecting this decline as presented in Figure 1 of Appendix (Ing & Basri, 2022). The negative trend has had a severe impact, particularly on agriculture, property, food and beverage, tourism, finance, and other industries, affecting everyone from micro, small, and medium-sized enterprises to companies listed on the Indonesia Stock Exchange (IDX). Numerous companies in these sectors are losing money and even closing down (Rediyono, 2022). This resulted in job losses that affected approximately 29 million people.

According to Keynes' theory, during times of economic downturn and low confidence, governments often choose to incur debt to boost the economy through increased spending (Crouch, 2009). The government spent large sums of money in terms of stimulus to rescue struggling industries and small businesses, as well as to increase unemployment assistance. The total expenditure on these initiatives amounted to billions of dollars (Sooreea & Sooreea, 2022). Ahmad Suaidy (Ahmad Suaidy, 2020) categorized Indonesian's governments' policy response to COVID-19 into three fiscal stimulus volumes:

1. Stimulus Volume 1: Primarily aimed to assist the tourism, accommodation, and transportation sectors and was part of the policy response. The allocated funds for this stimulus amounted to IDR 10.2 trillion.
2. Stimulus Volume 2: Included four fiscal and taxation-related policies, as well as four non-fiscal policies. The total amount of funds designated for this stimulus was IDR 22.9 trillion.
3. Stimulus Volume 3: Centers around three main objectives. Firstly, it focuses on enhancing the health sector by increasing the availability of medical equipment, personal protective gear, masks, and hand sanitizers, as well as expanding hospital capacity to combat the coronavirus. Secondly, it aims to provide social safety nets for the community by increasing the value of legal aid benefits, offering discounted 450VA and 900VA electricity rates, and issuing pre-employment cards. Thirdly, the government is preparing a stimulus to support affected business industries to ensure their continuity amidst economic pressures. The total amount of stimulus reaches IDR 405.1 trillion.

The broad scope and comprehensive nature of the government's stimulus initiatives were instrumental in minimizing Indonesia's economic decline to a rate of −2.07%, a figure that stands notably below the worldwide average of −3.5% (Ing & Basri, 2022).

The implementation of expansive monetary and fiscal policies by the government resulted in a significant surge in the money supply (Gharehgozli & Lee, 2022). Figure 2 of Appendix illustrates a positive trend of the yearly Indonesia M2 money supply, measured in billions of Indonesian Rupiah from 2020-2023. In terms of the data, elevated rates of monetary expansion may cause fluctuations in inflation. This can be interpreted as meaning that an augmentation in the money supply corresponds with a rise in the rate of inflation, or that regulation of the money supply can have a moderating effect on inflationary pressures (Klotz, 2007). Figure 3 of Appendix indicates more on how the inflation rate in Indonesia continued to fluctuate over the COVID-19 pandemic period.

When the rate of inflation rises, the government needs to act. Through the central bank (Bank Indonesia), the government will augment interest rates to manage the movement of money in society, ensuring that the flow of money is regulated (Rachmad, 2022). Interest rate in Indonesia has begun to spike in the second half of 2022 as shown in Figure 4 of Appendix. By increasing interest rates, the government aims to slow down economic growth and discourage spending by both consumers and businesses. This reduction in demand helps to lower prices over time. Rising interest rates have a direct impact on how much companies are willing to borrow, as it becomes more expensive to do so. This directly affects venture capitalists (VCs) and other sources of funding for startups. Companies tend to be more cautious about investing in riskier ventures when the future of the economy is uncertain. Uncertainty surrounding economic conditions often causes companies to rethink their hiring and growth strategies (Hetler, 2023).

As revenues begin to decline, investors seek companies to reduce expenses. Venture capitalists (VCs) are concerned about the possibility of lower profits following a period of significant growth. One approach to lowering expenses is to implement layoffs, which can result in a decrease in overhead costs (Brookman et al., 2007). Due to the limited availability of skilled

Information Technology (IT) human resources in Indonesia, costs tend to be high. Furthermore, the scarcity of available talent means that financing offers are often steep. This poses a significant challenge for the thriving startup ecosystem, as the rapid growth of these ventures is not being matched by the availability of qualified personnel (Ridho & Azizah, 2022).

Another part of the rise in layoffs is due to over-hiring. When the pandemic was at its peak, the use of technology grew exponentially as everything went digital. With people in quarantine and encouraged to stay at home, online activities skyrocketed, leading tech companies to record-breaking profits and a hiring frenzy to keep up with demand. These companies anticipated this trend to continue and expanded their teams rapidly. However, as work shifts back to pre-pandemic methods and people are spending less time online, the demand for tech services has decreased, leading to a reduction in the need for new hires.

As a result of the impact of the recent layoffs, numerous online discussions have taken place, primarily on social media platforms. While it predominantly represents a health and economic rather than a social crisis, discoveries from studies on online human behavior remain highly pertinent (Jaya, 2020). Thus, this research aims to examine the textual and hashtag attributes of social media users during the tech layoff post-COVID-19 period when coping with this phenomenon and how it is being perceived in the wider community, specifically Twitter. The observation will focus on identifying and examining the text and hashtags used by users about the layoffs. The researcher would like to understand the characteristics of Twitter users' text and hashtag utilization during tech winter layoffs post-COVID-19 in Indonesia. Besides that, the researcher wanted to find out if there were any trends in Twitter users' sentiment during the tech winter post-COVID-19 in Indonesia. The researcher will use Latent Dirichlet Allocation (LDA) & K-Means Clustering approach and would like to know whether the approaches able to identify common themes or topics in Twitter users' responses to tech winter layoffs post-COVID-19 in Indonesia. In addition, the research can provide valuable insights into how Twitter users are responding to the current economic situation in Indonesia. These insights can help inform policies and interventions to mitigate the negative effects of tech winter layoffs and support individuals in dealing with the challenges they face.

Cruickshank & Carley (Cruickshank & Carley, 2020) conducted prior research that examined the discourse of COVID-19-related topics on Twitter. The study employed a clustering methodology that grouped text and hashtags, highlighting the evolution of these discussions during the pandemic. They introduced an innovative clustering technique that utilized diverse data sets, which demonstrated how users interacted with hashtags and text. The research revealed the existence of both topical and temporal patterns within the COVID-19 Twitter discussion. Specifically, certain clusters of hashtags shifted over time, while others remained constant throughout the pandemic. Furthermore, the study discovered clear trends in the usage of hashtags during different periods. Wahyudi (Wahyudi et al., 2021) has also conducted prior research on Topic Modeling of Online Media News Headlines during the COVID-19 Emergency Response in Indonesia, utilizing the LDA Algorithm. Through modeling LDA topics from www.detik.com's news headlines over eight months period (March-October 2020), the outcomes demonstrated that the optimal number of monthly topics produced were three, with corona cases, positive corona, positive COVID, and COVID-19 news dominating each topic.

Although previous research thoroughly investigated the clustering of hashtags and text, as well as topic modeling, there exists a potential gap in comprehending the intricacies of user engagement and sentiment within these clusters. Additionally, both studies are confined to specific timeframes during the COVID-19 era, potentially leaving a research gap in understanding the post-COVID-19 evolution of discourse. This study aims to fill these gaps by conducting a more nuanced analysis, delving into user sentiment, and providing insights into post-COVID-19.

2. Methodology

In the next section, we will provide details about the terms and methods that the researcher uses.

Tech Winter Layoffs

Tech winter itself is a phrase used to refer to the reduced enthusiasm and investment in the technology sector or firms with sophisticated technologies, and as a result, the condition of startups or technology-based companies is starting to decline, and they are gradually failing (Spacey, 2016). This decrease in interest and investment necessitated that tech firms modify their priorities, business strategies, objectives, and interests to address the future uncertainties confronting the sector (Bahurekso, 2022).

News has surfaced about the recent massive job layoffs carried out by a major technology company (Brier, 2022). In Indonesia itself, the technology sector is currently undergoing an unparalleled surge in growth, largely attributable to the COVID-19 pandemic. Over the last couple of years, the country's digital economy has increased significantly, reaching $77 billion in 2022 from $63 billion in 2021 (Google et al., 2022). However, a significant transformation occurred in the global startup industry in 2022. Y Combinator, a distinguished Silicon Valley accelerator that has financed over 3,500 companies, advised its portfolio startups to brace for a worst-case scenario by decreasing expenditure and prolonging their financial runways (Singh, 2022). This advice made it challenging for startups to secure investment from investors.

The primary reason for the startup industry's decline in 2022 was the weakening of the worldwide economy. Several countries encountered a sharp increase in inflation rates, which hurt their economies (Rasyid et al., 2023). Therefore, in the present struggling economy, layoffs have become frequent among high-tech company employees who were once considered secure in their positions (Blanchette, 2003).

Twitter

Due to the recent layoffs, numerous workers have turned to social media to express their dissatisfaction with the unfair treatment they have received (Brier, 2022). One of the social media that is being used is Twitter. Twitter is a micro-blogging social media platform that allows users to easily connect through retweets, follows, hashtags, trending topics, and likes, making it easy to highlight everything that's happening in real-time (Kwak et al., 2010). With Twitter, users can stay up to date with the latest news and trends, share their thoughts and opinions, and connect with people from all over the world. By using hashtags and trending topics, users can join in on important conversations and contribute to larger discussions. In essence, Twitter offers a wide spectrum of communication channels, spanning from personal and intimate conversations to large-scale discussions that can reach a vast audience (Wu et al., 2011).

Oftentimes, user reactions and engagement on Twitter are analyzed (Srivastava et al., 2019). In this study, we chose Twitter as one of the popular micro-blogging social media platforms to be the primary source of data to develop a model that identifies clusters and topics of social media conversations related to the impact of COVID-19 on the technology industry, in the context of tech winter layoffs.

LDA

To determine the distances between textual documents, an Information Retrieval (IR) system must utilize semantic information. In situations where extracting semantic information is the primary objective, the use of Latent Dirichlet Allocation (LDA) is highly advantageous. LDA is widely recognized as one of the most prevalent topic models (Kaiser & Editors, 2020). Topic Modeling is a technique used to identify patterns of topic recognition by analyzing text data and extracting themes based on statistical analysis. The term "Latent" in Latent Dirichlet Allocation refers to hidden features within the data. "Dirichlet" refers to the distribution of topics in documents and words in topics, while "Allocation" refers to assigning topics to documents and selecting appropriate wording for topics. As a result, the algorithm is named Latent Dirichlet Allocation (LDA).

LDA is a generative probabilistic model used to analyze a collection of documents known as a corpus (Wahyudi et al., 2021). The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. The LDA model is represented as a probabilistic graphical model in Figure 5 of Appendix. This formula demonstrates the probability of a corpus (Blei et al., 2003).

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

As depicted in the diagram, the LDA representation comprises three levels. The parameters α and β are considered corpus-level parameters and are assumed to be sampled only once during the process of generating a corpus. The variables θ_d are document-level variables and are sampled once for each document, whereas the variables z_{dn} and wdn are word-level variables and are sampled once for every word in each document.

K-Means Clustering

K-means clustering, also known as K-means, is a type of unsupervised learning algorithm. This algorithm is particularly useful when working with unlabeled data sets. Essentially, the K-means algorithm groups similar data points into K categories based on their proximity to each other. This is achieved through a distance-based or centroid-based approach, where distances between data points are calculated to assign them to their respective clusters. The value of K is determined in advance, and the data is then divided into K clusters. This process results in higher similarity within each cluster, making it easier to differentiate between them (Cui, 2020).

To run the algorithm, we need to provide a matrix of M points in N dimensions and a matrix of K initial cluster centers, also in N dimensions. The size of each cluster is represented by the variable *NC*(L), while *D*(I, L) measures the Euclidean distance between point I and cluster L. The algorithm works by searching for a K-partition with the best possible sum of squares within each cluster. This is done by moving points between clusters until an optimal configuration is achieved (Society et al., 2012). The steps are as follows:

The first step is to choose K objects that will serve as the initial cluster centers based on the research objective. From there, the algorithm calculates the Euclidean distance between each data point and the cluster centers. Next, the data points are divided into categories based on their proximity to the nearest cluster center.

$$S_i^{(t)} = \left\{ x_p : |x_p - m_i^{(t)}|^2 \le |x_p - m_j^{(t)}|^2 \forall j, 1 \le j \le k \right\}$$

After the initial categorization, the next step involves recalculating the cluster centers of the newly formed clusters. Once the new centers are established, the data points are then re-categorized based on their distance to the new cluster centers, using the same approach as before.

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j$$

The third step involves iterating through the process until the cluster centers no longer change, at which point the algorithm can be stopped. There are three criteria for stopping the algorithm: first, the newly formed cluster's centroid remains the same; second, the data points remain in the same cluster as before; and third, the maximum number of iterations has been reached. If the newly formed cluster's centroid remains unchanged, it is a sign that the algorithm has learned all possible patterns, and training can be stopped. Similarly, if the data points remain in the same cluster after multiple iterations, it's a sign that training has reached a plateau, and it's time to stop. Finally, if the maximum number of iterations is reached, training should be stopped. For example, if we set the number of iterations to 200, the algorithm will repeat the process up to 200 times before stopping.

The choice of using LDA and K-Means Clustering rather than other techniques is driven by their complementary strengths in uncovering latent topics, identifying sentiment trends, and handling the dynamic and unstructured nature of social media data. This combination is expected to provide a robust analytical framework for extracting valuable insights from Twitter users' responses to tech layoffs in the post-COVID-19 tech winter in Indonesia.

This study uses text and hashtag tweets data on Twitter that contain specific keywords. The stages of the study can be seen in Figure 6 of Appendix.

Dataset Collection & Description

Social media research has become an increasingly valuable source of data in recent years, transitioning from the periphery to a more integral role. In addition to traditional methods such as interviews and surveys, social media research is now frequently employed, with online communities examining themselves as a subject of study. Given the prevalence of this research, Twitter data has been utilized from various viewpoints, yielding numerous innovative findings (Ahmed, 2021).

The study utilized a dataset from Twitter academic research Application Programming Interface (API) consisting of 311.625 tweets authored by 145.428 unique users, collected over for approximately 15 months from 01 January 2022 to 31 March 2023. The data taken from that period is due to the tech layoffs that occurred in Indonesia, which began to occur massively during that time range. The tech layoff data in Indonesia can be seen in Figure 7 of Appendix.

The study collected data from Twitter using a set of relevant keywords, including "phk" (*layoff*), "layoff", "tech layoff", "tech winter", "pemutusan hubungan kerja" (*layoff*), "startup layoff", "startup phk" (*startup layoff*) as well as several relevant hashtags, such as "#phk" (#*layoff*), "#layoff", "#layoffs", and "#techwinter". The data were filtered to include only tweets in the Indonesian language. Table 1 provides an example of the collected initial tweet data. Username is obfuscated to maintain privacy.

[Table 1 about here]

The dataset itself originally consisted of 4 categorical features which are text, hashtags, created at, and username. Text and hashtags are the most dominant features that are being used in providing the analysis. Features description can be found in Table 1 of Appendix.

Data Preprocessing

Preprocessing is an essential step in the data preparation process that involves cleaning and transforming raw data into a format that is suitable for analysis. To achieve this, several steps are taken. The first step involves removing duplicate records to reduce redundancy and improve the accuracy of the analysis. Next, non-textual elements such as URLs (Uniform Resource Locator), emojis, non-language strings, and mentions are removed to ensure that the analysis is not affected by irrelevant information.

After removing non-textual elements, stop words are removed from the text data. Stop words are commonly used words in the language that do not carry significant meaning and can interfere with the analysis results. Removing stop words helps to reduce noise and improve the accuracy of the analysis. Finally, stemming is applied to the text data, which involves reducing words to their base form by removing prefixes, suffixes, and other affixes. This step helps to consolidate similar words and reduce the dimensionality of the text data, making the analysis more manageable.

Feature Extraction (TF-IDF)

Feature extraction refers to the process of extracting relevant entities from the text data. This stage involves calculating the Term Frequency (TF) and Inverse Document Frequency (IDF) values of the collected text data. By computing the TF-IDF weight value using the equation below, the significance of tokens in the corpus can be determined.

$$w_{t,d} = tf_{t,d} \times \log \frac{N}{df_t}$$

Where:

$tf_{t,d}$ = the weight of a term t in document d

N = the total number of documents

df_t = the number of documents containing term t

Unsupervised Modeling

Twitter introduced the use of hashtags as a means of categorizing messages (tweets) based on their topic, enabling users to conveniently search for particular content and share information about it (Caleffi, 2015) while text tweets are frequently employed to express one's emotions, opinions on products, and position on issues (Mohammad & Bravo-Marquez, 2017). By considering the different characteristics of the text and hashtags on a tweet, we proposed using LDA to the textual attributes of tweets and then involves applying TF-IDF and K-means on the numerical attributes of hashtags.

3. Results

In this section, we describe the results of the clustering and topic modeling on tech layoff post-COVID-19 data. In the first section, we provide an overview of the exploratory analysis of the data. In the second section, we detail the results of feature extraction on the hashtag and text data. In the third section, we analyzed the K-means clustering on hashtags results. In the fourth section, we analyzed the topic modeling on text results.

Exploratory Analysis

Daily Statistics. In Figure 8, 9 & 10 of Appendix, the daily statistics of tweets and hashtag usage within the dataset are presented. It is important to note that, as a data preprocessing step, only tweets with hashtags and hashtags that were used in at least 3 tweets were included in the analysis which brought 10.813 data to the analysis. Hashtags that had a lower frequency of usage were omitted, as they often represent misspellings or less commonly used variants of more popular hashtags.

Sentiment & Emotion Analysis. In connection with data analysis, we carry out sentiment and emotion analysis. We are using Indonesian BERT Base Sentiment Classifier (Mdhugol, n.d.) as a sentiment-text-classification model which was originally the pre-trained IndoBERT Base Model (phase1 - uncased) model using Prosa sentiment dataset and Indo RoBERTa Emotion Classifier (StevenLimcorn, n.d.) as an emotion classifier based on Indo-roberta model which was trained on IndoNLU EmoT dataset and successfully achieve an f1-macro of 72.05%, accuracy of 71.81%, precision of 72.47% and recall of 71.94%. 98.743 data that has filtered out all duplicate tweets are used as the basis for analysis. The results of the analysis are shown in Table 2.

[Table 2 about here]

Terms Frequency

Before performing K-means clustering, we initially examine the term frequency of every hashtag in the text that has been collected. Based on the keyword that has been mentioned in the Dataset Collection & Description section, we show 10 hashtags with the appearances as well as their sums of frequency (rank), as seen in Table 2 of Appendix. After getting the hashtags with their term's frequency, the next step is to cluster them using K-means clustering.

K-means Clustering on Hashtag Results

As the K-means algorithm will group similar data points into K categories, we first determine the value K which will be then defined as the number of clusters. In this instance, we utilized the elbow method as a visual tool to identify the ideal value of K. This method involves calculating WCSS (Within-Cluster Sum of Squares), which represents the sum of the squared distance between the points in a cluster and the centroid of that cluster. Figure 11 of Appendix shows the distortion score elbow for the K-means clustering. By looking at the figure, we concluded that selecting K = 3 was appropriate for the number of clusters. To fit the model, we established a maximum of five iterations for the k-means algorithm during a single run. The initial score from the first iteration was 15.290, and subsequent iterations led to continued improvement until the fourth iteration, where the score stabilized at 8.155,9.

The cluster analysis yielded the following results: Cluster 1 had 1.051 hashtags, Cluster 2 had 7.730 hashtags, and Cluster 3 had 1.487 hashtags. The key features for each cluster are outlined in Table 3 of Appendix. The results of this clustering are then evaluated for their level of accuracy using Twin-Sample Validation. When using Twin-Sample validation, we use train test split as a model selection and input the test data to evaluate the accuracy then compare the resulting clusters to those from the training data. We can see the results in the report in Table 3.

[Table 3 about here]

Metrics that were used above in Table 3 can be used to quantify the similarity between the clusters. Accuracy F1-Score in the first row shows that the model made a correct prediction with 0.79 as a score across the entire dataset.

Topic Modeling on Text Results

The results of the topic modeling are carried out by implementing LDA algorithm to understand the topics that people find most understandable. The results of this process will be shown in tables or diagrams to make it easier to understand. Before applying the topic modeling, we first need to go through the evaluation of topic modeling by introducing Coherence Values. Coherence Values are a measure of how clear and understandable a group of words is when we use topic modeling (Wahyudi et al., 2021). We will look for the highest number and biggest difference in the coherence score graph. A higher coherence value means that people can understand the topics better. Figure 12 of Appendix shows the coherence scores.

As we want to have the highest number and biggest difference, in this case, we picked K = 4. Next, we want to select the optimal alpha and beta parameters by choosing the values that yielded maximum coherence scores for K = 4. Here we choose alpha = 0.01 and beta = 0.61 as they yielded the maximum coherence score.

[Table 5 about here]

Once we have determined the optimal number of topics based on the coherence score, the next step is to apply the LDA method. The result will be a set of words that represent the distinctive features of each topic, and these will be used as a reference for assigning the tweets to a particular topic shown in Table 6.

[Table 6 about here]

The significance (weights) of keywords is crucial in determining the topics. Moreover, the frequency of word occurrence in the documents is also worth considering. Words that appear in multiple topics and have a higher relative frequency compared to their weight often turn out to be less important. In Figure 13 of Appendix, we can see the result of including such words in the stop words list before re-running the training process.

Finally, in Figure 14 of Appendix, we can see a topic bubble that represents the distribution of topics in a 2-dimensional space. The size of each bubble corresponds to the frequency of the topic in the documents. In our topic model, the bubbles are big and non-overlapping, scattered across the chart, indicating a small number of topics. The distance between the bubbles approximates the semantic relationship between the topics. Moreover, the lack of overlap between the bubbles in the figure suggests that these topics do not share common words, as they are not closely placed together.

4. Discussion

The tweets and hashtags from Q1 2022 to Q3 2022 show stable numbers with slight spikes at some moments. However, starting from Q4 2022 to Q1 2023, there is an increase in the number of tweets and hashtags, with sharp spikes at certain time periods. The increase in tweet and hashtag numbers may indicate a trending increase in layoff incidents during that period, prompting people to share comments and express their opinions on Twitter social media. Also, as the scope of layoffs post-COVID-19 expanded, it is conceivable that tweets and hashtags originally unrelated to the topic may have become included in the conversation. Some examples of the topic that occurred on the tweet and hashtag spikes during the layoff events are:

1. The increase in fuel prices on 9th September 2022. The topic surrounds fuel price (*bahan bakar minyak/BBM)* of around 475 tweets and 506 hashtags.
2. GoTo lays off 1,300 employees on 18th November 2022. The topic surrounds GoTo layoff of around 193 tweets and 457 hashtags.
3. No event happened on 24th December 2022, but topic discussion varies around 1489 tweets and 2964 hashtags.

There are other few spikes that are not mentioned in the above example. While the reason for these temporal spikes remains unclear, they are likely linked to the evolving behavior of Twitter users during various stages of the post-COVID-19 pandemic layoffs.

The sentiment analysis categorizes tweets into three main sentiments: negative, neutral, and positive. The majority of the tweets, a significant count of 51,881, fall into the negative sentiment category, indicating a prevailing sense of dissatisfaction or concern among Twitter users. In contrast, 37,407 tweets are categorized as neutral, suggesting that a considerable portion of the discourse doesn't strongly lean towards positive or negative sentiments. The positive sentiment category comprises 9,455 tweets, reflecting a smaller but noteworthy proportion of the discussions characterized by optimism or approval. The emotion analysis delves deeper into the nuanced expressions within these sentiments. A predominant emotion in the analyzed tweets is anger, with 45,685 instances, emphasizing a substantial undercurrent of frustration or discontent among users. Sadness is the second most expressed emotion, with 31,594 tweets, indicating a significant presence of sentiments associated with sorrow or disappointment. In contrast, happiness is expressed in

12,722 tweets, suggesting moments of joy or positivity within the discussions. Fear, represented by 8,561 tweets, signifies a notable level of anxiety or apprehension among users. Interestingly, love is the least expressed emotion, with 181 tweets, indicating a relatively smaller but still present aspect of positive sentiment within the discussions. These findings collectively paint a nuanced picture of the emotional dimensions of Twitter users' responses to tech layoffs. The prevalence of negative sentiments, coupled with the detailed breakdown of specific emotions, provides valuable insights into the complex and multifaceted nature of the discussions on social media. Furthermore, other study indicates that aside from concerns about COVID-19 & its impact, individuals facing mental health challenges, including anxiety and depression which contributes to individuals being overwhelmed by negative emotions (Darmayanti et al., 2020). Understanding the interplay of sentiments and emotions is crucial for comprehensively grasping the public's reactions and perceptions during challenging economic conditions, aiding policymakers and organizations in tailoring appropriate interventions and strategies to address the diverse emotional needs of the community.

The results of the cluster analysis offer valuable insights into the distinct themes and topics emerging from the hashtag usage within the examined Twitter discussions related to tech layoffs during the post-COVID-19 tech winter in Indonesia. Three distinct clusters were identified, each characterized by a unique set of hashtags representing specific themes or discussions within the larger conversation. Cluster 1, with a total of 1,051 hashtags, seems to revolve around a mix of technology-related terms. This cluster suggests a focus on the technological landscape, encompassing discussions on layoffs, startups, and major players in the tech industry. Cluster 2, the largest cluster with 7,730 hashtags, point towards a thematic concentration on political and social aspects, including discussions related to Islam, political figures like Jokowi, and national development for the people. Cluster 3, comprising 1,487 hashtags, exhibits a different thematic focus which suggests a cluster centered around healthcare, employment, and possibly the hospitality industry. This cluster indicates a diversified discourse, extending beyond tech layoffs to include discussions on health, job opportunities, and the hotel industry. This cluster analysis is instrumental in revealing the multifaceted nature of the Twitter discussions, showcasing the diverse topics that intertwine with the central theme of tech layoffs. These findings can guide a deeper understanding of the varied concerns and interests within the online discourse during this critical period.

The application of Latent Dirichlet Allocation (LDA) in topic modeling has yielded a set of distinctive words that encapsulate the key themes within the Twitter discussions on tech layoffs during the post-COVID-19 tech winter in Indonesia. Each topic is represented by a list of words along with their corresponding weights, indicating their significance in characterizing the content of tweets associated with that particular topic. The significance of these keywords lies in their weighted representation, indicating their importance in characterizing the topics. The weights provide insights into the relevance and prominence of specific terms within each theme. Additionally, the frequency of word occurrence across topics and their relative frequency compared to their weights is a critical consideration. Words that appear in multiple topics but have a higher relative frequency compared to their weight may be less distinctive and may not contribute significantly to topic differentiation. Understanding these topics allows for a nuanced exploration of the various dimensions of the tech layoffs discourse on Twitter. The identified themes shed light on the multifaceted nature of the discussions, ranging from labor rights and corporate actions to broader economic challenges and personal experiences.

5. Conclusion

In this study, we have successfully analyzed the characteristics of tweets and hashtags from Twitter data. To make the data clean and easy to analyze, we conducted pre-processing on the data. The tweets and hashtags from Q1 2022 to Q3 2022 showed stable numbers with slight spikes at some moments. This increase may indicate a trending increase in layoff incidents during that period, prompting people to share comments and express their opinions on Twitter. There are several examples of topic discussion around the spikes in tweet and hashtag numbers, including the increase in fuel prices and layoffs at GoTo. The reason for these spikes remains unclear, but they are likely linked to the evolving behavior of Twitter users during various stages of the post-COVID-19 pandemic layoffs.

We also conducted sentiment and emotion analysis using the Indonesian BERT Base Sentiment Classifier and Indo RoBERTa Emotion Classifier. The sentiment analysis revealed that 51.881 tweets were negative, 37.407 were neutral, and 9.455 were positive. The emotion analysis showed that 45.685 tweets expressed anger, 31.594 expressed sadness, 12.722 expressed happiness, 8.561 expressed fear, and 181 expressed love. The classifiers used were trained on the Prosa sentiment dataset and IndoNLU EmoT dataset and achieved good accuracy, precision, and recall. The analysis of sentiment and emotion offers a glimpse into the collective emotional atmosphere and content of the examined tweets. The notable frequency of negative sentiments, including anger and sadness, could signify a prevailing sense of dissatisfaction, frustration, or concern among Twitter users throughout the studied timeframe. Recognizing this provides valuable insights for assessing public responses, pinpointing areas of concern, and shaping strategies for communication or intervention.

As mentioned by Habibi & Cahyo (Habibi & Cahyo, 2019), TF-IDF can serve as a valuable feature to aid in the identification of hashtag occurrences during searches. Additionally, the K-Means method is a viable clustering technique that can be used to determine Twitter user attributes based on hashtags. This is proven by the successful modeling using K-means clustering to cluster the tweets into 3 categories based on the elbow method to identify the ideal value of K. We also employed Latent Dirichlet Allocation for Topic Modeling to gain insights into the most comprehensible topics among individuals. This method was also utilized by Wahyudi et al., (Wahyudi et al., 2021) to discover that there is a tendency for three news topics to emerge each month with a reasonably even distribution of the number of news articles on each topic, averaging at 33% per topic. The models effectively explained the topical discussions of tweets during the tech winter layoff period after COVID-19. During the tech winter layoff post-COVID-19 condition, we observed that the use of text and hashtags related to predetermined keywords does not exclusively pertain to the researched phenomenon. Popular hashtags associated with tech layoffs are used even in tweets unrelated to this topic. A significant finding is the prevalent use of text and hashtags containing political and religious elements in a substantial portion (75%) of the tweets. This suggests a multifaceted discussion that intertwines tech layoffs with broader societal and ideological considerations. The tweets on the layoff topic come from various sources, including news portals, advertisements, user opinions, and other topical text and hashtags (14%). This diversity indicates that discussions about tech layoffs extend beyond individual perspectives to encompass media, advertising, and other contextual elements.

Based on the findings of the study, several recommendations emerge to enhance the analysis and understanding of the discourse surrounding tech layoffs in the post-COVID-19 tech winter. First and foremost, there is a need to refine monitoring criteria to improve the precision of sentiment and emotion analysis. Given the observed use of popular hashtags in unrelated tweets, establishing more specific criteria for identifying and categorizing tech layoff-related discussions can contribute to a more accurate representation of public sentiments. Additionally, acknowledging the multifaceted nature of the discourse, as evidenced by the inclusion of political and religious elements, suggests the importance of considering a broader context in the analysis. This implies a shift towards a more comprehensive perspective that encompasses societal and ideological dimensions intertwined with discussions on tech layoffs. Furthermore, recognizing the diversity of sources contributing to these discussions, including news portals, advertisements, and individual opinions, underscores the necessity of engaging with a wide array of stakeholders. Collaborating with these diverse groups can provide valuable insights into their perspectives and concerns related to tech layoffs, ultimately contributing to a more holistic understanding of the phenomenon. These recommendations collectively aim to refine analysis methodologies, consider broader contextual elements, and engage with diverse stakeholders to enhance the depth and accuracy of insights derived from the study.

The implications drawn from this study cast a significant light on the dynamics and repercussions of discussions surrounding tech layoffs in the post-COVID-19 tech winter. Foremost, the identification of prevailing negative sentiments and emotional expressions provides a crucial understanding of the public's response to these events. This insight is invaluable for companies and policymakers as they navigate through the challenges associated with workforce reductions, enabling them to tailor strategies that address and mitigate the widespread dissatisfaction and concerns. Moreover, the recognition of the inclusion of political and religious elements in these discussions highlights a broader societal and ideological context within which tech layoffs are perceived. This finding underscores the need to contextualize these events, emphasizing the interconnectedness of economic shifts with larger societal narratives. As a result, businesses and policymakers can better appreciate the multifaceted nature of these discussions and craft responses that consider the diverse perspectives and concerns of the public. Furthermore, the study implies that effective communication strategies should be crafted to manage negative sentiments and foster a more nuanced dialogue. By acknowledging and addressing the emotional aspects revealed in the discourse, organizations can work towards rebuilding trust and understanding among stakeholders. Lastly, the study's implications guide future research by establishing a foundation for exploring the intersections of technology, societal values, and individual experiences during challenging economic conditions, fostering a more comprehensive understanding of the evolving landscape of tech layoffs in the post-COVID-19 era.

The strength of our research lies in the thorough pre-processing of the data, ensuring its cleanliness and accessibility for analysis. The stability of tweet and hashtag numbers throughout Q1 2022 to Q3 2022, with noticeable spikes, indicates the potential emergence of trends and heightened discussions during certain periods. These spikes coincide with notable events such as fuel price increases and layoffs at GoTo, suggesting a correlation between external events and online discussions. The sentiment and emotion analysis, employing advanced classifiers trained on relevant datasets, offers a nuanced understanding of the collective emotional tone expressed in tweets. The notable frequency of negative sentiments, especially anger and sadness, provides a valuable lens into the prevailing dissatisfaction and concerns among Twitter users. The clustering methods, including K-means clustering and Latent Dirichlet Allocation for Topic Modeling, effectively categorized and identified meaningful patterns within the vast Twitter dataset. The integration of diverse methodologies enhances the robustness of our findings, contributing to a holistic understanding of the multifaceted discussions surrounding tech layoffs on Twitter.

Despite these strengths, several limitations need to be addressed in future research. First, our analysis may not accurately represent the sentiments and emotions expressed in other languages as this phenomenon is not exclusive to Indonesia and we were using tweets in the Indonesian language. It would be beneficial for future studies to include tweets in other languages. Second, for future research, supplementary modeling techniques such as Multi-View Modularity Clustering (MVMC) could be employed to authenticate the precision of our results. This method has been demonstrated to have the potential to expand clustering to tasks such as clustering hashtags in extensive social media data (Cruickshank & Carley, 2020). Last, future research could investigate how layoffs and other economic conditions are discussed on other social media platforms and how such discussions are shaped by broader factors.

References

Ahmad Suaidy. (2020). Stimulus Fiskal di Tengah Badai Pandemi. *Anggaran.Kemenkeu.Go.Id, April*.

Ahmed, W. (2021). Using Twitter as a data source: an overview of social media research tools (2021). *Lse*, 1–4.

Bahurekso, P. R. (2022). *What is Tech Winter?* https://journalradar.my.id/what-is-tech-winter/

Blanchette, S. (2003). Surviving unemployment in the high-tech downturn. *Computer*, *36*(7), 24–28. https://doi.org/10.1109/MC.2003.1212686

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet Allocation*. *3*, 993–1022.

Brier, D. (2022, November 24). *Tech winter? How cold is it really*.

Brookman, J. T., Chang, S., & Rennie, C. G. (2007). CEO cash and stock-based compensation changes, layoff decisions, and shareholder value. *Financial Review*, *42*(1), 99–119. https://doi.org/10.1111/j.1540-6288.2007.00163.x

Caleffi, P.-M. (2015). The "hashtag": A new word or a new rule? *SKASE Journal of Theoretical Linguistics*, *12*(2), 46–69.

Crouch, C. (2009). Privatised Keynesianism: An unacknowledged policy regime. *British Journal of Politics and International Relations*, *11*(3), 382–399. https://doi.org/10.1111/j.1467-856X.2009.00377.x

Cruickshank, I. J., & Carley, K. M. (2020). Characterizing communities of hashtag usage on twitter during the 2020 COVID-19 pandemic by multi-view clustering. *Applied Network Science*, *5*(1). https://doi.org/10.1007/s41109-020-00317-8

Cui, M. (2020). *Introduction to the K-Means Clustering Algorithm Based on the Elbow Method*. 5–8. https://doi.org/10.23977/accaf.2020.010102

Darmayanti, K. K. H., Winata, E. Y., & Anggraini, E. (2020). Why Can Other People Live Normally While I Cannot?: An Application of Telecounseling Due to COVID-19. *Makara Human Behavior Studies in Asia*, *24*(2), 109. https://doi.org/10.7454/hubs.asia.1140920

Ecommurz. (2023). *Tech Layoffs in Indonesia*. https://ecommurz.com/layoff/

Gharehgozli, O., & Lee, S. (2022). Money Supply and Inflation after COVID-19. *Economies*, *10*(5), 1–14. https://doi.org/10.3390/economies10050101

Google, Temasek, & Bain & Company. (2022). *Indonesia E-conomy SEA 2022 Report, Through the waves, towards a sea of opportunity*. 1–16.

Habibi, M., & Cahyo, P. W. (2019). Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *13*(4), 399. https://doi.org/10.22146/ijccs.50574

Hetler, A. (2023). *Tech sector layoffs explained: What you need to know*.

Ing, L. Y., & Basri, M. C. (2022). COVID-19 in Indonesia: Impacts on the Economy and Ways to Recovery. In *COVID-19 in Indonesia: Impacts on the Economy and Ways to Recovery*. Routledge. https://doi.org/10.4324/9781003243670

Jaya, E. S. (2020). Editorial Note: Human Behavior and COVID-19. *Makara Human Behavior Studies in Asia*, *24*(1), 1. https://doi.org/10.7454/hubs.asia.1270720

Kaiser, M. S., & Editors, K. R. (2020). *Advances in Intelligent Systems and Computing 1309 of International Conference on Trends in Computational and Cognitive Engineering*.

Klotz, B. (2007). Monetary policy and the causality between inflation and money supply in Indonesia. *Serials Librarian*, *53*(1–2), 191–201. https://doi.org/10.1300/J123v53n01_15

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 591–600. https://doi.org/10.1145/1772690.1772751

Mdhugol. (n.d.). *Indonesian BERT Base Sentiment Classifier*.

Mohammad, S. M., & Bravo-Marquez, F. (2017). Emotion intensities in tweets. **SEM 2017 - 6th Joint Conference on Lexical and Computational Semantics, Proceedings*, 65–77. https://doi.org/10.18653/v1/s17-1007

Rachmad, Y. E. (2022). The Influence And Impact of The Money Burning Strategy on The Future of Startups. *Proceedings of the 1st Adpebi International Conference on Management, Education, Social Science, Economics and Technology (AICMEST)*, 1–5.

Rasyid, M. F., Sumirat, E., & Rahadi, R. A. (2023). *QUANTITATIVE STUDY OF STARTUP VALUATION AND STRATEGY POST-2022*. *2*(2), 384–393. https://doi.org/10.58344/jws.v2i2.226

Rediyono, R. (2022). How Does Investment Progress in Indonesia During the COVID-19 Pandemic? *International Journal of Finance & Banking Studies (2147-4486)*, *11*(2), 16–24. https://doi.org/10.20525/ijfbs.v11i2.1637

Ridho, W. F., & Azizah, N. (2022). Factor Analysis of the Phenomenon of Mass Layoffs At Startups: Mixed Method Approach With Structural Equation Modeling. *Jurnal MEBIS (Manajemen Dan Bisnis)*, *7*(2), 195–208. https://doi.org/10.33005/mebis.v7i1.373

Singh, M. (2022, May 19). *YC advises founders to 'plan for the worst' amid market teardown*.

Society, R. S., Society, R. S., & Statistics, A. (2012). *Algorithm AS 136 A K-Means Clustering Algorithm*. *28*(1), 100–108.

Sooreea, R., & Sooreea, B. (2022). The Impacts of COVID-19 on Business Practice: Some Key Insights. *Advances in Social Sciences Research Journal*, *8*(12), 366–376. https://doi.org/10.14738/assrj.812.11521

Spacey, J. (2016, October 1). *What is a Technology Winter?*

Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment Analysis of Twitter Data. *International Journal of Healthcare Information Systems and Informatics*, *14*(2), 1–16. https://doi.org/10.4018/ijhisi.2019040101

StevenLimcorn. (n.d.). *Indo RoBERTa Emotion Classifier*.

Tradingeconomics.com. (n.d.-a). *Indonesia Inflation Rate*.

Tradingeconomics.com. (n.d.-b). *Indonesia Interest Rate*.

Tradingeconomics.com. (n.d.-c). *Indonesia Money Supply M2*.

Wahyudi, M. D. R., Fatwanto, A., Kiftiyani, U., & Wonoseto, M. G. (2021). *Telematika Topic Modeling of Online Media News Titles during COVID-19 Emergency Response in Indonesia Using the Latent Dirichlet Allocation (LDA) Algorithm*. *14*(2), 101–111.

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on twitter. *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, 705–714. https://doi.org/10.1145/1963405.1963504

Table 1. Sample collected tweets data

Text	Hashtags	Created At	Username
gue sedih kalo denger ada berita layoff lagi tech winter bakalan sampe berapa lama ya (*I feel sad when I hear news about layoffs again in this tech winter. I wonder how long it will last.*)		20-07-2022 14:48:30	dx…ya
bener ga yakalau valuasi teknologi itu seperti nilai sayuran dan buahan ketika masa panen buahan dan sayuran itu fresh tp ketika udah melewati batas waktu jd layu dan ngga fresh lgcepat masa usangnya (*Isn't it true that the valuation of technology is like the value of vegetables and fruits during the harvest season? When the fruits and vegetables are fresh, but once they exceed their time limit, they wither and no longer stay fresh. The aging process happens quickly.*)	'startupbubble', 'layoffs'	26-07-2022 14:56:05	HD…06
dari pada terlalu khawatir soal phk lebih baik upgrade diri kita agar bisa diperhitungkan lebih dan dihargai oleh perusahaan semangat ya untuk siapa pun yang sedang berjuang (*Rather than being overly worried about layoffs, it's better to upgrade ourselves so that we can be more considered and valued by companies. Keep up the spirit for anyone who is currently striving.*)	'investasi' (*investment*), 'startup', 'saham' (*stocks*), 'bisnis' (*business*), 'fyp'	10-11-2022 12:23:00	da…23

Table 2. Sentiment & Emotion Analysis Results

Sentiment	Tweets Count
Negative	51.881
Neutral	37.407
Positive	9.455

Emotion	Tweets Count
Anger	45.685
Sadness	31.594
Happy	12.722
Fear	8.561
Love	181

Table 3. Evaluation report

	Precision	Recall	F1-Score	Support
Accuracy			0.79	514
Weighted Average	0.69	0.79	0.74	514

Table 5. Coherence Values for K = 4

Topics	Alpha	Beta	Coherence
4	0.01	0.01	0.446157375
4	0.01	0.31	0.521444147
4	**0.01**	**0.61**	**0.541881818**
4	0.01	0.91	0.539503313
4	0.31	0.01	0.435804217
4	0.31	0.31	0.496271702
4	0.31	0.61	0.537934758
4	0.31	0.91	0.518448915
4	0.61	0.01	0.445193222
4	0.61	0.31	0.49657328
4	0.61	0.61	0.467351553
4	0.61	0.91	0.494705138
4	0.91	0.01	0.467749733
4	0.91	0.31	0.479713267
4	0.91	0.61	0.510960669
4	0.91	0.91	0.533005962

Table 6. List of Words Representing Topic

Topic	List of Words
1	0.020*"buruh" (*labor*) + 0.010*"pabrik" (*factory*) + 0.008*"uuciptakerja" (*job creation law*) + 0.007*"pengusaha" (*entrepreneur*) + 0.007*"pemerintah" (*government*) + 0.006*"korban" (*victim*) + 0.006*"jokowi" + 0.005*"pekerja" (*labor*) + 0.005*"nasib" (*fate*) + 0.004*"demo" (*demonstration*)
2	0.058*"karyawan" (*labor*) + 0.026*"perusahaan" (*company*) + 0.013*"phk" (*layoff*) + 0.009*"pesangon" (*severance pay*) + 0.008*"goto" + 0.007*"twitter" + 0.005*"meta" + 0.005*"google" + 0.004*"shopee" + 0.004*"teknologi" (*technology*)
3	0.025*"startup" + 0.020*"badai" (*storm*) + 0.015*"ekonomi" (*economy*) + 0.011*"islam" + 0.009*"rakyat" (*people*) + 0.008*"pengangguran" (*unemployed*) + 0.008*"industri" (*industry*) + 0.008*"indonesia" + 0.007*"kapitalisme" (*capitalism*) + 0.005*"muslimah" (*muslim woman*)
4	0.033*"kena" (*hit*) + 0.031*"layoff" + 0.010*"orang" (*person*) + 0.006*"habis" (*gone*) + 0.006*"gaji" (*salary*) + 0.005*"dapet" (*get*) + 0.005*"anak" (*child*) + 0.004*"semoga" (*hopefully*) + 0.004*"resign" + 0.003*"susah" (*difficult*)

Appendices

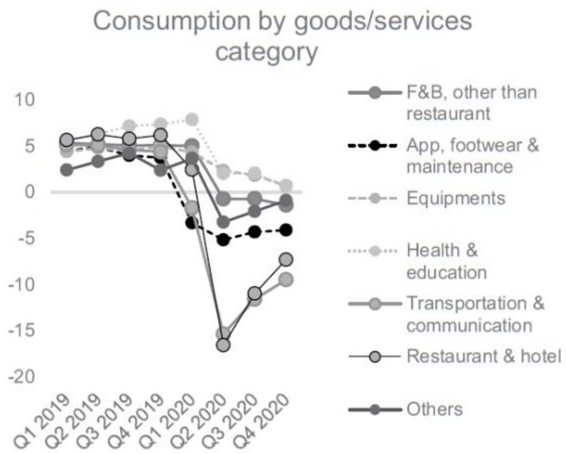


Figure 1. GDP Consumption Composition

Figure 2. Indonesia's Money Supply (M2) in IDR Billion, yearly from 2020-2023. (Tradingeconomics.com, n.d.-c)

Figure 3. Indonesia's Inflation Rate in Percentage, yearly from 2020-2023. (Tradingeconomics.com, n.d.-a)

Figure 4. Indonesia's Interest Rate in Percentage, April 2022 to March 2023. (Tradingeconomics.com, n.d.-b)

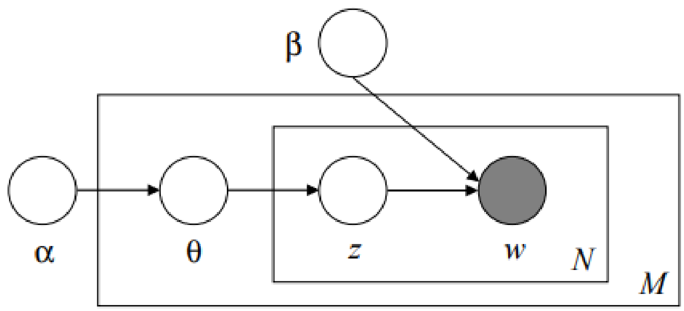


Figure 5. Graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

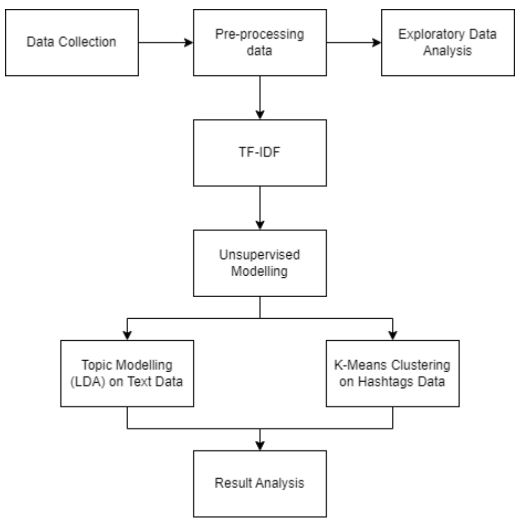


Figure 6. Research Stages

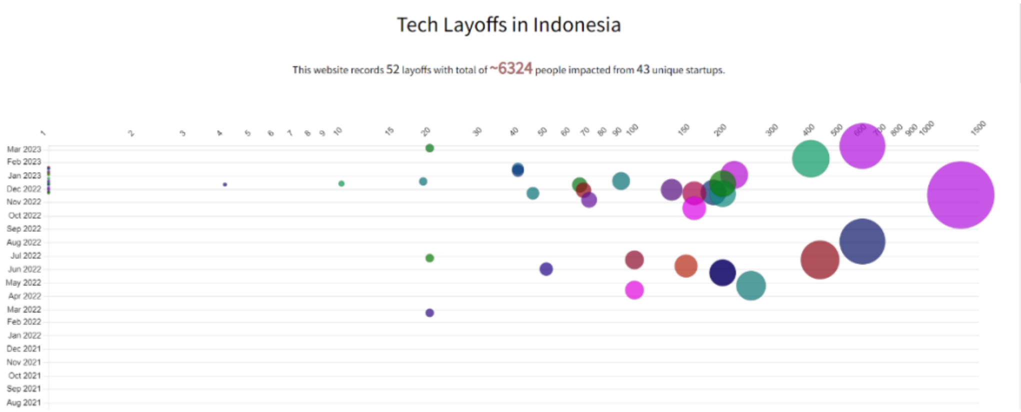


Figure 7. Tech Layoffs in Indonesia (Ecommurz, 2023)

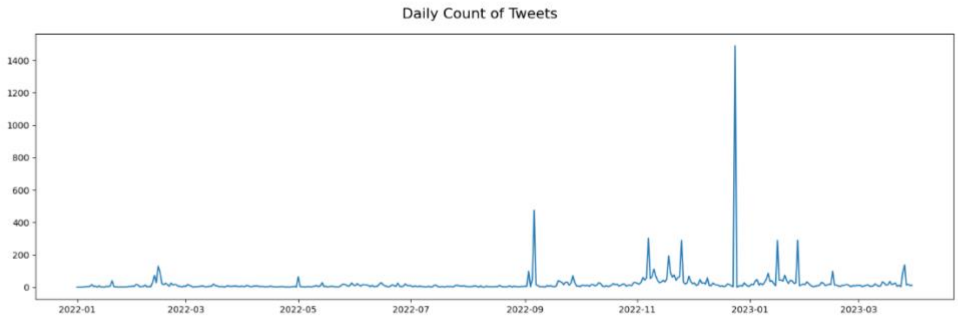


Figure 8. Daily Count of Tweets

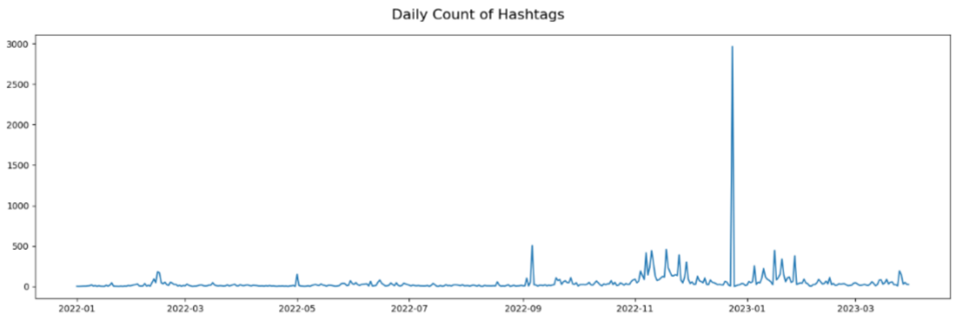


Figure 9. Daily Count of Hashtags

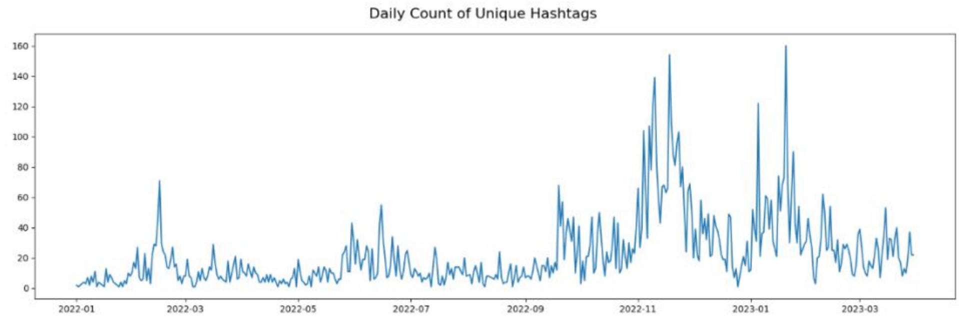


Figure 10. Daily Count of Unique Hashtags

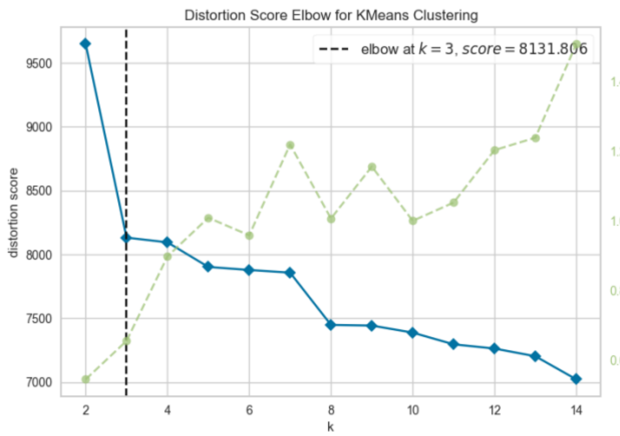


Figure 11. Distortion Score Elbow for K-means Clustering

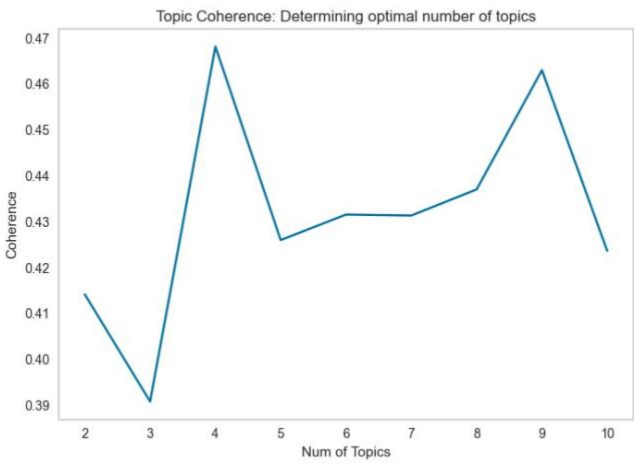


Figure 12. Coherence Scores

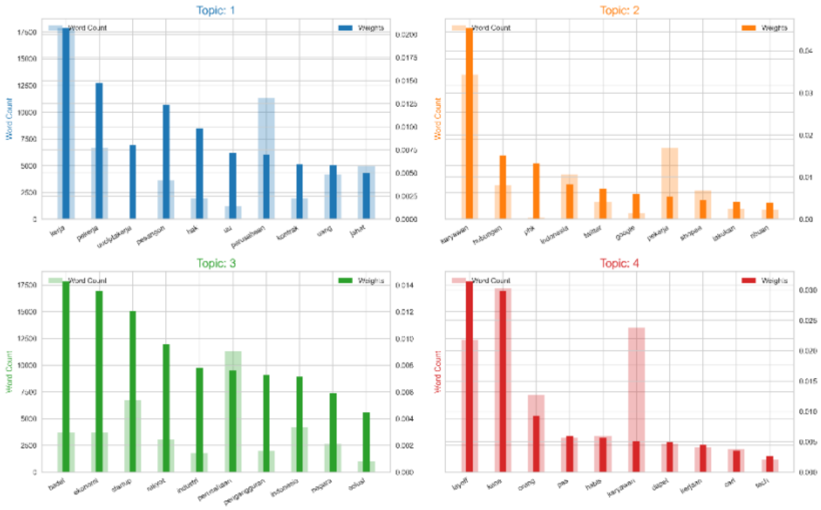


Figure 13. Word Count and Importance of Topic Keywords

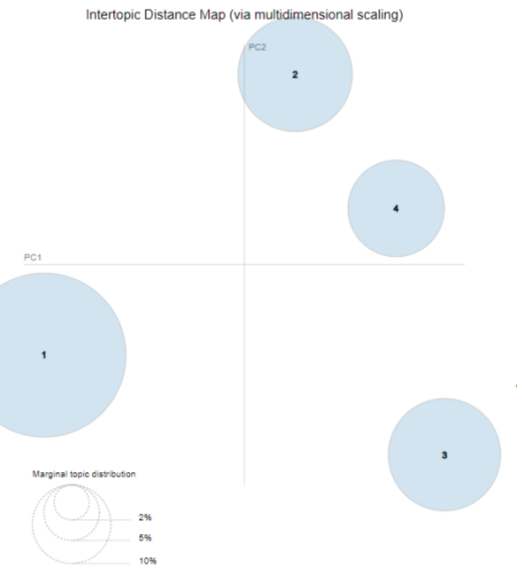


Figure 14. Intertopic Distance Map

Table 1. Dataset features that are being used

Feature Name	Feature Description
Text	Tweets text that was being created by a user
Hashtags	Hashtags text that was being created by a user
Created At	Date time which specifies when a user was tweeting the text or hashtags
Username	The user who was creating the text or hashtags tweet

Table 2. Hashtag with the appearances

Hashtag	Rank

islamkaffah (*concept related to Islam*)	4.918
phk (*layoff*)	1.448
cabutpermenjht56tahun (*revoke Ministerial Regulation on the 56-year Pension Fund*)	703
penyaluranbltbbm (*aid cash disbursement*)	535
layoffs	406
startup	313
phkmassal (*large scale layoff*)	231
batalkanpermenaker2_2022	207
goto	203
shopee	187

Table 3. Key features for each cluster

Cluster	**Key Features**
1	phk (*layoff*), startup, shopee, twitter, goto, elonmusk, amazon, meta, google, tokopedia
2	islamkaffah (*concept related to Islam*), cabutpermenjht56tahun (*revoke Ministerial Regulation on the 56-year Pension Fund*), penyaluranbltbbm (*aid cash disbursement)*, jokowipresidenku (*jokowi my president*), jokowimajubersamaindonesia (*jokowi stands with indonesia*), pembangunanuntukrakyat (*development for the people)*, islammenyejahterakanburuh (*islam prospers the workers*), tolakkenaikancukairokok (*rejecting the increase in cigarette tax*)
3	healthcare, headline, harianposkota (*a news magazine company*), harianjabar (*a news magazine company*), hiring, hotelrumahluwih, hotelindonesianatour, hotelibisyogyakarta, hotelbali