Mining Social Media (Twitter) Data for Corporate Image Analysis: A Case Study in the Indonesian Mining Industry

D Y Fahmi1,*, Hartoyo1, N Zulbainarni1
1School of Business, IPB University, Bogor, Indonesia

*Corresponding email: dziki.fahmi@gmail.com

Abstract. PT Freeport Indonesia (PTFI) is the largest multinational mining company in Indonesia which frequently becomes trending topics on social media. Since the continuation of the company’s business operations depends on many stakeholders including the regulators, it is important for the company to listen to public opinion in order to develop a solid public relation strategy to maintain its corporate reputation. The objective of this study is to gather insights from social media as input towards a data-driven public relations strategy development. As a case study, we analysed Twitter conversations mentioning PTFI between 20 to 30 August 2020 during which there was a strike in the company. We performed text mining techniques using Drone Emprit system specifically sentiment analysis, bot analysis, and descriptive analysis to get quantitative measures of the corporate image on social media. The findings suggest several improvement opportunities of the current social media strategy of the company including the lack of engagement from local communities, the absence of specific strategy to handle negative sentiment, and the lack of awareness of social media topics not reported by mainstream media. As the implication of the findings, we propose several recommendations including nurturing engagement from employees and local communities, addressing the most influential tweets, and choosing better time periods to get more engagement.

1. Introduction
Social media has been widely used in Indonesia to express opinions, emotions, and spread news, whether in the form of facts or hoaxes. Based on the Digital 2020 report [1], the percentage of active social media users in Indonesia in 2019 is 59% of the total population. From the same report, the average time Indonesian users spend on social media every day is almost 3.5 hours.

As the use of social media has been widely accepted, companies start to utilize social media either as part of their marketing channel or public relation (PR) strategy. As a PR tool, social media is used to increase engagement and build a positive corporate image of the business entities [2]. Corporate image is a mental image of external audience towards an organization at a time. Corporate image forms company reputation which defined as the value of society's perception of the organization in the long term [3]. Trending social media topics about a company with negative polarity can lower the company’s reputation and lead to a construction of public opinion which is not necessarily correct. Therefore, analysis of social media is important for companies to form data driven public relation strategies through digital platforms, especially for managing external communication on developing issues [4]. Specifically for mining industry, the importance of maintaining corporate image on social media was

1 To whom any correspondence should be addressed.
highlighted as one of the top issues facing mining companies because of the existence of organizations, particularly NGOs & opinion makers, that are capable to fully utilize social media to shape opinions of the stakeholders [5] which comprise of government, regulators, investors, organized labors, and communities [6].

In this research, we performed a corporate image analysis of PTFI by examining social media (Twitter) data using Drone Emprit platform [7]. As there are multiple parties involved as stakeholders of mining companies, we also performed bot analysis in our study to validate whether conversations mentioning the company were generated naturally or likely driven by autonomous entities to shape opinions.

Twitter was chosen among other social media platforms because Twitter communication is fast and spontaneous in response to emerging issues generating trending topics [8]. As a use case, we analyzed tweets mentioning keywords PTFI during a strike involving hundreds of PTFI workers demanding ease of COVID-19 lockdown. This issue was covered both in national television, online news [9], and social media.

The objective of this research is: a) to establish quantitative measures of corporate image based on Twitter conversations; and b) to gain insights based on the measures. Our study is guided by the following research questions (RQ):

RQ1a: How do Twitter users communicate in mentioning PTFI during the strike period?
RQ1b: Is the conversation generated naturally (human-like) or having tendency be generated by autonomous entities known as social bots?
RQ2: What insights can be drawn from user activity, sentiments and content generation on Twitter?
RQ3: What implications can companies drawn from the insights?

The paper is structured as follows. First, we provide theoretical background and related works as the foundation of this research. Next, we present our research methodology which detail out case description, research process, data collection and analysis. Finally, the next section comprises the result, discussion, business implications to companies, conclusions and proposed future works.

2. Theoretical Background

2.1. Corporate Reputation in the Mining Industry

The business of the mining industries deals with the extraction of unrenewable resources which raise important environmental, sociopolitical, and governance (ESG) issues that are subject to heightened public attention, and regulatory scrutiny [10]. Mining companies in general are subject to political and reputational risks such as civil unrest, regulatory uncertainty and political instability. The risks come from various groups including local communities, landowners, governments, non-government organizations, and regulators [11]. One of the main strategies employed by mining companies to uphold its reputation and manage the risks is by performing Corporate Social Responsibility (CSR) projects [10]. However, study [11] has shown that CSR only resulted marginal improvement on corporate reputation at best. Deloitte report [12] has identified corporate communication in social media as leading strategies in focus for managing reputational risks in digital era for mining companies.

2.2. Corporate Communication on Social Media

In social media, users can create, search, make interpretations, and share information about a corporation, making a subjective truth become a collective truth which if goes unchecked can quickly lead to a serious reputation risk for the company [13]. The loss of reputation may negatively affect corporations in different areas including business competitiveness, governments and regulators' attitude, the loyalty of stakeholders, also media and pressure groups coverage [14]. Therefore, it is important for corporations to go beyond crises management and perform online reputation management on social media proactively even before any crisis emerge.

In order to proactively and correctly responds to social media conversations, institutions need to get correct signal from social media. It is a complex task because social media platforms are big data
sources. Due to the characteristics of big data, the signal to noise (S/N) ratio from social media is low. Therefore, to increase the S/N ratio, the ideal scenario of engagement between public crowd and institutions is to employ big data analytics [15].

2.3. Analysing Social Media with Text Mining

Text mining is a process of extracting useful patterns or knowledge from texts which are unstructured in nature [16]. It combines techniques from data mining, machine learning, natural language processing, information retrieval, as well as knowledge management [17]. There are various branches of text mining depending on their objectives. In this research, we focused on two of them as follows.

2.3.1. Sentiment Analysis. Also known as “opinion mining”, sentiment analysis is a computational study which aims to analyze people’s sentiments towards entities such as topics, issues, services, organization, and their attributes [18]. Existing sentiment analysis techniques can be categorized into three approaches namely machine learning, lexicon-based and hybrid approach [19]. In the machine learning approach, artificial intelligence algorithms are used to process textual data and automatically classify sentiments into categories such as “positive”, “negative” or “neutral”.

Sentiment analysis has been employed for various purposes including to measure customer satisfaction [20], predict stock market movements [21], and developing crisis management strategy [22]. The use of sentiment analysis can serve as an addition or even replacement to the traditional method of collecting opinions such as surveys and interviews.

2.3.2. Bot Analysis. The increased utilization of social media has been accompanied by efforts to manipulate online conversations using social bots [23]. Social bots are accounts that controlled completely or partially by computer programs. Deceptive social bots are automated social media accounts that are intended to manipulate public opinion by exploiting human nature tendencies of paying attention to trending topics and circle of friends [24]. It was estimated that 9-15% of active Twitter users are social bots [25]. The initial attempt to detect bots was through social crowdsourcing platform with human intervention. However, due to issue with scalability and the growth of social bots, later widely used approach was through supervised machine learning methods [23].

3. Related works

Insights from social media have been studied during certain extreme event, corporate crisis, or government policy reviews. Stieglitz et al. [22] analyzed Twitter to get insights on the VW “Dieselgate” crisis and showed that the lack of response from the official VW account to address emotionality has brought negative impact to the corporation. Another study [26] used Drone Emprit platform to gather insights from social media regarding the increase fee of Indonesian national health insurance policy and found that the official of government account was not visible to address the negative sentiments.

While there were previous studies on social media analysis for various purposes, we found lack of studies specifically addressing mining industries despite its importance due to the nature of its business which rely on corporate reputation to their stakeholders. We also found lack of studies which incorporated bot detection on their research process along with other social media analysis. Therefore, we proposed a new framework on our research methodology to employ bot analysis along with performing corporate image analysis to validate insights. This study provided contributions by focusing on obtaining social media insights specifically for extractive industry, and proposing a framework incorporating bot detection for corporate social media analysis.

4. Research Methodology

4.1 Case Description

PT Freeport Indonesia (PTFI) is Indonesia's largest multinational mining company operating in remote Papua. The nature of its mining business that deals with unrenewable resources, the foreign ownership
of shares (49%), and its operations in complex socioeconomic areas have made PTFI a frequent trending topic and endured negative media coverage [11]. Several cases that led to PTFI being discussed widely in the media were the “Papa Minta Saham” case in 2015, mineral export ban in 2017, and the share transfer to the Indonesian government in 2019.

Due to COVID-19 outbreak in 2020, PTFI has implemented several measures to contain the spread of the virus including limitation of travel from mine site (Tembagapura) to nearby city of Timika, Papua. PTFI has successfully controlled the spread of the virus at the mine site and therefore can keep its operations running. However due to the travel restrictions, some workers demanding the ease of travel during their shift day off to meet their families. This led to a strike involving hundreds of employees blocking main road to the mine [9].

4.2. Research Process
Figure 1 highlights the research process described in this paper.

![Figure 1. Research process](image)

We employed Drone Emprit Academic (DEA) [7], a big data analytic platform from Media Kernels which is able to monitor and analyze social media, online news and other sources in a near-real time basis. Figure 2 represents the architecture of Drone Emprit.

![Figure 2. Drone Emprit Architecture](image)
4.2.1. Data Collection. Twitter API [28] was used on DEA to collect data from Twitter during period 20 – 30 August 2020 when the strike happened using keywords “PTFI”, “PT Freeport Indonesia”, and “Freeport-McMoran”. The collected data includes user profile (user id, username, number of followers), tweet data (user, timestamp, geolocation, retweet count, reply count, and tweet content).

4.2.2. Sentiment Analysis. The DEA uses sentiment analysis with machine-learning approach, specifically the supervised learning adaptive multiple models combining Naïve Bayes and Maximum Entropy methods. As input of sentiment analysis is the collected textual tweet content. Based on established learning model from training data on DEA, the collected tweets are automatically classified as either positive, negative or neutral sentiment [27].

4.2.3. Bot Analysis. We utilized Botometer on DEA to perform Bot detection. Botometer [23] is a scoring system based on machine learning techniques to classify an account as bot or human based on thousands of labelled examples. It works by fetching 1150 features from an account such as its account profile, friends, and temporal activity patterns against labelled examples to determine bot score [25]. As input of bot analysis on DEA is the collected user accounts which then pass on to Botometer API to get bot score. Botometer gives the score between 0 – 5, for 0 as most human-like and 5 as most bot-like or machine generated tweets [27].

4.2.4. Descriptive analysis. We performed several descriptive analyses described below.

4.2.4.1. Mention by hour. Data obtained from Twitter API carries timestamp of the tweet from which we can further analyze to find pattern of mention by day and hour.

4.2.4.2. Location analysis. Some users tagged their location when they tweeted. Twitter API can crawl this data in the form of GPS location (latitude and longitude). The GPS location can be transformed into geographical location and plotted on the map using google maps geolocation API.

4.2.4.3. Analysis of most engagement user. The number of engagements on Twitter is defined as the total number of retweet and reply. Retweet indicates agreement while reply signals discussion.

5. Results and Discussion

During the period of 20 – 30 August 2020, we collected 659 tweets mentioning PTFI on Twitter. The peak was on 29 Aug which generated 118 tweets. We further breakdown the analysis as follows.

5.1. Sentiment Analysis

Figure 3 shows the result of sentiment analysis in pie chart. The share of negative sentiment during the period is 92% (605 tweets), the positive sentiment is 7% (44 tweets), and only 2% is neutral. We expected the issue to have a negative trend during the strike as the narrative of tweets that came with the word ‘strike’ by itself has tendency towards negative opinion.

![Figure 3: The Share of Sentiment](image-url)
To understand how the strike event affected social media conversations, we also analyzed the sentiment 10 days prior to the strike period where no significant event related to the company happened. As shown in Figure 4, the sentiment was initially positive. As the strike started happening, the total mentions increased, and the polarity of sentiment shifted to negative by 20 Aug 2020.

![Figure 4. The Share of Sentiment over Period of Time](image)

The findings highlighted the importance of monitoring social media conversations by companies to maintain its digital reputation. The rise of negative voice can be used as a signal for PR strategist to run countermeasures and the monitoring of the sentiment over time can be utilized to measure the effectivity of the actions.

5.2. Bot Analysis
We identified 292 distinct authors during the period of analysis. The overall average bot-score for all authors is 1.7. By using pareto analysis shown in Figure 5, we found that 90% of the authors were scored < 3. The result indicates that the average of tweets are natural human-like tweets.

![Figure 5. Pareto of Bot Score Analysis](image)

We also analyzed bot score over time during the period and found that bot-like tweets with bot score 3-5 remain the lowest as can be seen in Figure 6. This means that even during the peak period, the conversations were most likely generated by human. The implication for PR strategist is that they can rely on signals from social media analytics knowing that the statistics conversations were not biased and intentionally shaped by social bots.
5.3. Descriptive Analysis

5.3.1. Mentions by hour. Our analysis found that the cycle of tweets followed M-shaped curve as shown in Figure 7 with double peak at 8 am and 3 pm (GMT + 7) during the period of analysis. We expanded the start date of analysis to 1 Jul 2020 to validate the result with more datasets and found that the result was still consistent. This finding indicates the time period for PR strategist to get more engagement on Twitter.

5.3.2. Location analysis. We found that 26% of the tweets shared their location, of which 42% of the tweets were originated from Jakarta and its surrounding (Jabodetabek) areas. While most employees and their families reside in the Mimika Region of Papua province, only two tweets were identified from this location. This analysis provides insight for PR strategist of the company to nurture engagement on social media from their untapped internal resources which are their employees and local communities. Employees and their families particularly know the situation of company first-hand. They can perform fact-check and promote fact-based opinions about the company. Figure 8 shows plotted location of the tweets on the map.
5.3.3. *Most Engagement user (influencer) and tweet.* According to the analysis of the top 5 users with most engagement, @jatamnas with 25003 followers obtained the most engagement (114 tweets). @KompasTV, which an official account of a mainstream media, received second most engagement (101 tweets) even though it has more than 3 million followers.

Of the five tweets with the most engagement during the period, only two of them reported stories about the strike, while the three others which include the most engagement tweet from @jatamnas discussed different rumors about the company that need further fact-check. This means there were issues about the company that were being highlighted in the social media, different from those reported in the mainstream media. It is also important to notice that the company’s official Twitter was not included in the top five. All these findings highlight the importance for companies to have social media analysis tools and a data-driven PR strategy to tackle conversations on social media.

6. Conclusions & Future works

In this study, we focused on the Twitter communication mentioning PTFI specifically during the period of a strike in the company as a case study. We established several quantitative measures for corporate image analysis using a big data analytic platform called Drone Emprit. We gathered insights based on the measures such as the share of sentiments, the nature of conversations, the concentration of tweets location, and the top influencers. We highlighted business implications for PR strategists such as the suggestion to nurture engagement on social media from their internal resources, attention to issues that was brought up by top influencers, and recommended timeframe to get more engagement on social media.

This research analyzed the company’s minor event as a case study with 10-day time frame of analysis. While there were insights that can be obtained in this research, we would recommend further research to analyze the company’s major events that happened in the past such as Indonesia government share acquisition in early 2019 and “Papa minta saham” scandal back in 2015 in order to gather more insights. The limitations of the Twitter API which only allows queries D-7 from the start of the crawling process prevented the authors to conduct such analysis. While analyzing historical tweets is technically possible using another crawling tool such as Twint [22], it is beyond the scope of this research and therefore is left as future works. The insights obtained in this research are envisioned to be the initial steps towards a data-driven corporate social media strategy development based on social media. Situational Crisis Communication Theory (SCCT) [23] is one of the frameworks that can further be applied as future works of strategy development continuing the result of this research.

7. References

[1] We Are Social and Hootsuite, Digital 2020. [Online]. Available: https://wearesocial.com/digital-2020. [Accessed: 13-Sep-2020].
[2] Dijkmans C Kerkhof P and Beukeboom C J, Apr. 2015 A stage to engage: Social media use and corporate reputation *Tourism Management* **47** p. 58–67.

[3] Gray E R and Balmer J M T, Oct. 1998 Managing Corporate Image and Corporate Reputation *Long Range Planning* **31**, 5 p. 695–702.

[4] Zhang B and Vos M, Jan. 2014 Social media monitoring: aims, methods, and challenges for international companies *Corporate Communications: An International Journal* **19**, 4 p. 371–383.

[5] Deloitte, 2015, Tracking the trends 2015 - The top 10 issues mining companies will face in the coming year, Deloitte.

[6] Chamber of Mines of South Africa, 2016, Integrated Annual Review 2016.

[7] Fahmi I, 2018 Drone Emprit Academic: Software for Social Media Monitoring and Analytics.

[8] Stieglitz S Bruns A and Krüger N, 2015 Enterprise-Related Crisis Communication on Twitter in *Smart enterprise engineering: Proceedings of the 12th International Conference on Wirtschaftsinformatik (WI 2015)*, Universität Osnabrück p. 917–932.

[9] Aditri I P, 24-Aug-2020, Karyawan Freeport Demo dan Blokade Jalan Tambang di Tembagapura, Ini Tuntutannya, *KOMPAS.com*. [Online]. Available: https://regional.kompas.com/read/2020/08/24/13484471/karyawan-freeport-demo-dan-blokade-jalan-tambang-di-tembagapura-ini. [Accessed: 13-Sep-2020].

[10] Sethi S P Martell T F and Demir M, 2016 Building Corporate Reputation Through Corporate Social Responsibility (CSR) Reports: The Case of Extractive Industries | SpringerLink *Corporate Reputation Review* **19** p. 219–243.

[11] O’Callaghan T and Spagnoletti B, 2017, Mining, corporate social responsibility, and corporate reputation, .

[12] Deloitte, 2018, Tracking the trends 2019 - The top 10 issues transforming the future of mining.

[13] Aula P, 2010 Social media, reputation risk and ambient publicity management *Strategy & Leadership* **38**, 6 p. 43–49.

[14] Rayner J, 2004 *Managing Reputational Risk, Curbing Threats, Leveraging Opportunities* 6 West Sussex: John Wiley & Sons.

[15] Hutchison K and Kumara S, 2013 Big Data Analytics - Sentiment Analysis of Twitter Data using Clustering Techniques in *IEE Annual Conference Proceedings*.

[16] Tan A-H, 1999 Text mining: The state of the art and the challenges in *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases* **8** p. 65–70.

[17] Feldman R and Sanger J, 2007 *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* Cambridge University Press.

[18] Bing L, 2012 *Sentiment Analysis and Opinion Mining* Morgan & Claypool Publishers.

[19] Medhat W Hassan A and Korashy H, Dec. 2014 Sentiment analysis algorithms and applications: A survey *Ain Shams Engineering Journal* **5**, 4 p. 1093–1113.

[20] Vidya N A Fanany M I and Budi I, 2015 Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers *Procedia Computer Science* **72** p. 519–526.

[21] Pagolu V S Challa K N R Panda G and Majhi B, 2016 Sentiment Analysis of Twitter Data for Predicting Stock Market Movements in *Proceedings of the International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016*.

[22] Stieglitz S Mirbabaie M and Potthoff T, 2018 Crisis Communication on Twitter during a Global Crisis of Volkswagen - The Case of “Dieselgate” in *Proceedings of the 51st Hawaii International Conference on System Sciences*.

[23] Yang K-C Varol O Davis C A Ferrara E Flammini A and Menczer F, 2019 Arming the public with artificial intelligence to counter social bots *Human Behavior and Emerging Technologies* **1**, 1 p. 48–61.

[24] Jun Y Meng R and Johar G V, Jun. 2017 Perceived social presence reduces fact-checking *PNAS* **114**, 23 p. 5976–5981.
[25] Varol O Ferrara E Davis C A Menczer F and Flammini A, 2017 Online Human-Bot Interactions: Detection, Estimation, and Characterization in *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)* p. 280–289.

[26] Laagu M A and Setyo Arifin A, 2020 Analysis the Issue of Increasing National Health Insurance (BPJS Kesehatan) Rates through Community Perspectives on Social Media: A Case Study of Drone Emprit in *2020 International Conference on Smart Technology and Applications (ICoSTA)* p. 1–7.

[27] Fahmi I, 21-May-2020 Jeroan Drone Emprit presented at the Workshop Artificial Intelligence Engines, Live Streaming on Zoom.

[28] Twitter, Twitter API Documentation, Twitter Developer. [Online]. Available: https://developer.twitter.com/en/docs/twitter-api. [Accessed: 13-Sep-2020].

**Acknowledgements**

We would like to thank Drone Emprit Academic (DEA) which enabled us to develop this analysis.

**Disclosure**

The main author of this paper is a PTFI employee currently continuing study on a company-funded scholarship.