# Data and Method

The data used in this study originated from several sources, i.e.: (1) Twitter; (2) Indonesian Change petition website; (3) Google Playstore; (4) Online mass media; and (5) Interviews. In addition to interviews, data was obtained by Semi-Automatic Scrapping (SAS) method, which is by extracting content from downloaded HTML files (Fahmi, Wibowo, and Yudanto 2018). Although the method takes much time and a lot of computer memory the results obtained can provide a comprehensive understanding throughout the time the existence of the parameters to get the data used. This comprehensive data can overcome the typical limitations of online research, which is the time representation of data (Marwick 2014).

Twitter data is used to find out topics that are popular among netizens as well as knowing the growing discourse behind the issues discussed. We first make observations to get parameters which can be used to obtain the Twitter data needed [(Table 1)](#Table 1. Tweets used in the study). The keywords "Teluk Benoa/Benoa Bay" and "Gojek" are used to collect data from the Change Indonesia website [(Table 2)](#Table 2. Data from the Change Petition Page). Mass media data were obtained using the parameters "Teluk Benoa/Benoa Bay" and "transportasi online/online transportation" from the selected online mass media pages (Kompas, Tempo, Detik, Berita Bali, and Balipost) [(Table 3)](#Table 3. News from online mass media). Finally, we also took comments for two applications (Gojek and Gojek Driver) published by PT Gojek Indonesia on Google Playstore [(Table 4)](#Table 4. Comments from Google Playstore).

Similar to some previous studies on the same field (See: Hotho, Nürnberger and Paaß, 2005; Adedoyin-Olowe, Gaber and Stahl, 2014; Bao *et al.*, 2014), the pre-processing stage is done before the analysis is carried out. At this stage, stop words, URLs, usernames (text from Twitter) are removed, numbers are converted to English texts, and some words are normalised to improve the modelling result interpretability. As suggested in several previous similar studies [references], data analysis in this study was conducted using a mixed method of text mining techniques and Critical Discourse Analysis (CDA). Chronologically the analysis is done by applying Social Network Analysis (SNA) to find out the important actors in the network at the first stage. Then Topic Modeling is used to find topics from the data used in the research automatically and avoid cherry picking in describing the present discourse. Finally, Critical Discourse Analysis (CDA) is applied to interpret the topics obtained.

SNA in this study was used to find out the accounts that became influencers and determine the research participants. Therefore, we use the Eigen centrality concept as the primary parameter to determine the position of a username in the network. Eigen centrality is the centrality of network-defined nodes determined by the relationship with all the nodes and other influential nodes (Bonacich, 2007). SNA is carried out only for data from Twitter using Gephi network visualisation software.

By combining CDA and Topic Modelling, this research can also be said to use a new method that is currently developing rapidly, namely the Corpus-Assisted Discourse Studies (CADS). Simply put, CADS is an analytical method derived from the linguistic corpus approach and utilises computer assistance and algorithms to handle large amounts of text data (Partington, Duguid, and Taylor 2013, 10). Therefore, CDA is preceded by utilising the Latent Dirichlet Allocation (LDA) algorithm developed by Blei et al. (2003) to extract the hidden topics of a corpus automatically. Topic modelling is done using packages topicmodels (Grün and Hornik 2011) in the R programming language environment. The number of topics is assessed by using perplexity, which is a way to find goodness-of-fit and evaluate LDA (Chen and Wang 2003; Koltsova and Koltcov 2013). Here is one example of perplexity from one of the corpora.

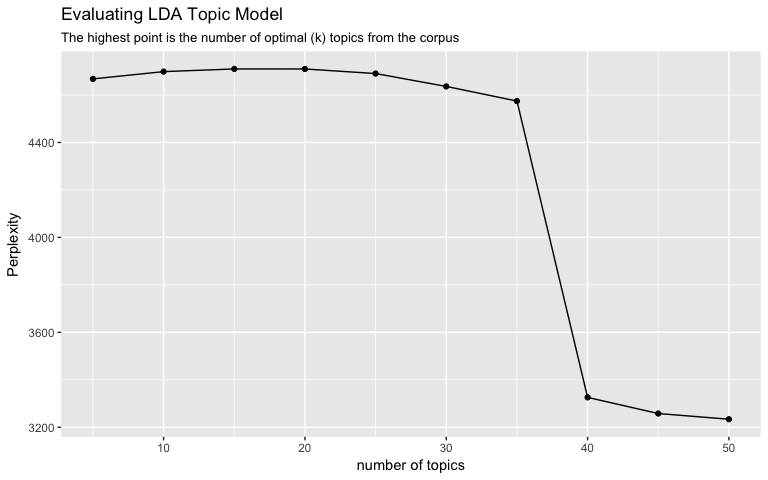


Figure 1. The perplexity of tweets containing #balitolakreklamasi and #balinotforsale hashtag

Based on the figure above, the number of topics that can be generated from tweets containing the #balitolakreklamasi and #balinotforsale hashtag is between 5 - 35 topics. Then, we performed an in-depth validation by reviewing the document containing the highest probability term to create topic labels as suggested by Maier et al. (2018) to validate the topic modelling results. The labelled topic is then analysed by utilising discourse analysis method. Here, we use the intertextuality technique, an attempt to comprehend the context of a text from another text source (Gee 2011, 54).

# References:

Adedoyin-Olowe, Mariam, Mohamed Medhat Gaber, and Frederic Stahl. 2014. ‘A Survey of Data Mining Techniques for Social Network Analysis’. *International Journal of Research in Computer Engineering and Electronics* 3(6): 1–25. http://jdmdh.episciences.org/18/pdf%5Cnhttp://jdmdh.episciences.org/18/.

Bao, Yanwei, Changqin Quan, Lijuan Wang, and Fuji Ren. 2014. ‘The Role of Pre-Processing in Twitter Sentiment Analysis’. In *Intelligent Computing Methodologies. ICIC 2014. Lecture Notes in Computer Science*, eds. DS. Huang, KH Jo, and L Wang. Springer, 615–24.

Blei, David M., Andrew Y.Ng, and Michael I. Jordan. 2003. ‘Latent Dirichlet Allocation’. *Journal of Machine Learning Research* 3: 993–1022.

Bonacich, Phillip. 2007. ‘Some Unique Properties of Eigenvector Centrality’. *Social Networks* 29(4): 555–64.

Chen, Si, and Yufei Wang. 2003. ‘Latent Dirichlet Allocation’. *Journal of Machine Learning Research* 3(1): 993–1022.

Fahmi, Ujang, Canggih Puspo Wibowo, and Faturahman Yudanto. 2018. ‘Hybrid Method in Revealing Facts behind Texts: A Combination of Text Mining and Qualitative Approach’. In *Reinventing Public Administration in a Globalized World: A Non-Western Perspective*, Atlantis Press, 132–48. http://sci-hub.tw/http://www.atlantis-press.com/proceedings/aapa-18/25896109.

Gee, James Paul. 2011. *An Introduction to Discourse Analysis: Theory and Method*. 3d ed. London: Routledge.

Grün, Bettina, and Kurt Hornik. 2011. ‘Topicmodels: An R Package for Fitting Topic Models’. *Journal of Statistical Software* 40(13): 1–30. http://www.jstatsoft.org/v40/i13/.

Hotho, Andreas, Andreas Nürnberger, and Gerhard Paaß. 2005. ‘A Brief Survey of Text Mining’. *Journal for Computational Linguistics and Language Technology* 20: 19–62.

Koltsova, Olessia, and Sergei Koltcov. 2013. ‘Mapping the Public Agenda with Topic Modeling: The Case of the Russian LiveJournal’. *Policy and Internet* 5(2): 207–27. http://dx.doi.org/10.1002/1944-2866.POI331 (September 8, 2017).

Maier, Daniel et al. 2018. ‘Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology’. *Communication Methods and Measures* 00(00): 1–26. https://www.tandfonline.com/doi/full/10.1080/19312458.2018.1430754.

Marwick, Alice E. 2014. ‘Ethnographic and Qualitative Research on Twitter’. In *Twitter and Society*, eds. Katrin Weller et al. New York: Peter Lang, 109–22.

Partington, Alan, Alison Duguid, and Charlotte Taylor. 2013. 2 Studies in corpus linguistics *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies*. Philadelphia: John Benjamins Publishing Company. http://0-site.ebrary.com.adrastea.ugr.es/lib/univgranada /Doc?id=5000133.

LAMPIRAN

## Table 1. Tweets used in the study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Kasus | Parameter | Jumlah | Unique Users | Keterangan |
| 1 | Change | Tweet yang memention akun @changeOrg\_ID | 38.283 | 24.629 | Warga net cenderung menggunakan akun ini untuk menyebarluaskan petisi yang mereka buat |
| #balinotforsale | 9.975 | 5.404 | Tagar gerakan sosial di Bali |
| #balitolakreklamasi | 72.641 | 23.498 | Tagar gerakan penolak reklamasi Teluk Benua di Bali |
| 2 | Gojek | #savegojek | 1.414 | 1.124 | Tagar yang digunakan untuk menyuarakan perlawanan terhadap adanya upaya pelarangan keberadaan ojek daring |
| #saveojekonline | 195 | 111 |
| #savedrivergojek | 1.069 | 425 | Tagar yang digunakan untuk mendukung pengemudi ojek daring saat mereka dianggap dirugikan oleh penyedia layanan |
| Tweet yang menyebut akun @CeritaTranspOL | 738 | 415 | Akun yang mewadahi cerita seputar ojek daring |

## Table 2. Data from the Change Petition Page

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Kasus | Parameter | Jumlah Petisi | Pendukung | Jumlah Komentar\* |
| 1 | Change | Teluk Benua | 12 | 162.394 | 5.798 |
| 2 | Gojek | Gojek | 41 | 109.003 | 1.037 |

\*komentar diambil dan dihitung dari petisi yang didukung oleh minimal 1000 penandatanganan/pendukung

## Table 3. News from online mass media

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Kasus | Media | Parameter | Jumlah | Keterangan |
| 1 | Change | [www.balipost.com](http://www.balipost.com) | Reklamasi Teluk Benua | 95 | Koran lokal Bali |
| [www.beritabali.com](http://www.beritabali.com) | 75 |
| [www.kompas.com](http://www.kompas.com) | 57 | Media massa nasional |
| [www.tempo.co](http://www.tempo.co) | 73 |
| 2 | Gojek | [www.detik.com](http://www.detik.com) | Transportasi Online | 1.847 | Media massa nasional |
| [www.kompas.com](http://www.kompas.com) | 97 |
| [www.tempo.co](http://www.tempo.co) | 97 |

## Table 4. Comments from Google Playstore

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Kasus | Aplikasi | Jumlah | Keterangan |
| 1 | Gojek | Gojek | 6.890 | Aplikasi yang digunakan oleh pengguna jasa layanan Gojek |
| Gojek Driver | 2.548 | Aplikasi yang digunakan oleh pengemudi Gojek (Mitra PT Gojek Indonesia) |