# Data and Method

The data used in this study came from several sources, namely: (1) Twitter; (2) Page www.change.org; (3) Google Playstore; (4) Online mass media; and (5) Interviews with research participants. In addition to interviews, data was obtained by Semi-Automatic Scrapping (SAS) method, which is by extracting content from HTML files that have been downloaded first (Fahmi, Wibowo, and Yudanto 2018). Although the method takes a lot of time and takes up a lot of computer memory the results obtained can provide a comprehensive picture throughout the time the existence of the parameters to get the data used. This comprehensive data can overcome the general limitations of research that uses data from the internet, which is the time representation of data (Marwick 2014).

Data from Twitter is used to find out topics that are popular and are talked about virtually by netizens as well as knowing the growing discourse behind the issues discussed. To obtain this data we first make observations to get parameters [(Table 1)](#Table 1. Tweets used in the study) which can be used to obtain the data needed. The keywords "teluk benoa" and "gojek" are used to get data from the www.change.org page [(Table 2)](#Table 2. Data from the Change Petition Page). Data from the mass media were obtained using the parameters "teluk benoa" and "online transportation" from the selected online mass media pages, namely, compass, tempo, detik, beritabali, and balipost [(Table 3)](#Table 3. News from online mass media). While the data from Google Playstore was taken from comments in two applications issued by PT GOJEK Indonesia, namely "Gojek" and "Gojek driver" [(Table 4)](#Table 4. Comments from Google Playstore).

Same as some previous similar studies (See: Hotho, Nürnberger and Paaß, 2005; Adedoyin-Olowe, Gaber and Stahl, 2014; Bao *et al.*, 2014), before the analysis is carried out, the data obtained first through the pre-processing stage. At this stage, stop words, URLs, usernames (specifically text from Twitter) are deleted, numbers are converted to text, and some words are normalized to improve the interpretation of the results of the analysis. As suggested in several previous similar studies [references], data analysis in this study was conducted using a mixed method between text mining techniques and Critical Discourse Analysis (CDA). Chronologically the analysis is done by first using Social Network Analysis (SNA) to find out the important actors in the network. Then the topic is Topic Modelling and Critical Discourse Analysis (CDA).

SNA in this study was used to find out the accounts that became influencers and determine the study participants. Therefore, we use the eigen centrality concept as the main 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 visualization software.

Data analysis in this study combines the CDA method and Topic Modelling. Therefore, this research can also be said to use a new method that is currently developing rapidly, namely the Corpus Assisted Discourse Studies (CADS). CADS is discourse analysis with the guidance of algorithms (quantitative/statistical) and computer programming to find out the emerging discourse (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 latent 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.

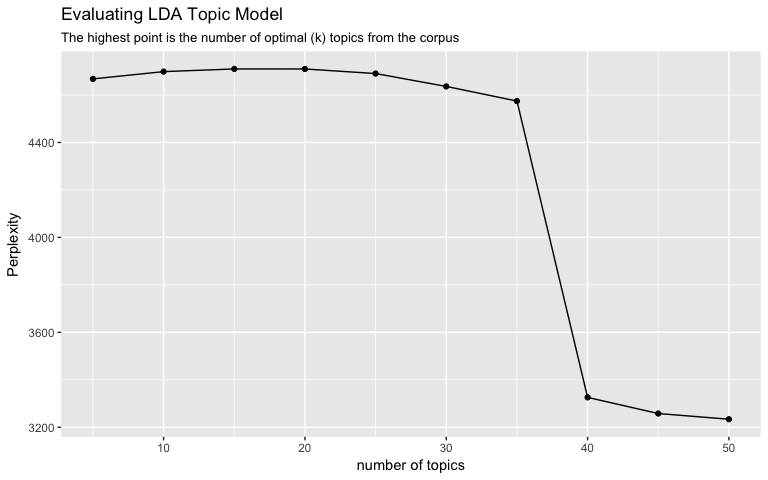


Figure 1. Perplexity of tweets containing #balitolakreklamasi and #balinotforsale hashtag

Based on the picture 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 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).

# Reference:

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 Intoduction 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 Assited 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) |