Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 197 (2022) 76–81

Procedia Computer Science

www.elsevier.com/locate/procedia

Sixth Information Systems International Conference (ISICO 2021)

Twitter stance detection towards Job Creation Bill

Arif Hamied Nababan, Rahmad Mahendra, Indra Budi*

Faculty of Computer Science, University of Indonesia, Depok 16424, Indonesia

Abstract

The formation of Job Creation Bill has raised the polemics in Indonesia. This study aims to identify the public's stance on the Job Creation Bill on Twitter social media. We collect tweets using keywords related to the Job Creation Bill and annotate nearly 10K tweets with a class label describing stance position. The experiments were conducted using the Naïve Bayes, Support Vector Machine, and Logistic Regression, with unigram and bigram features. The best performance in our experiment achieved by the Logistic Regression classifier using the unigram feature obtains a micro F-1 score of 71.8%.

Keywords: Job Creation Bill; stance detection; text classification; Twitter

1. Introduction

In the 2020-2024 National Legislation Program (in Indonesian: Prolegnas), the People's Representative Council (in Indonesian: Dewan Perwakilan Rakyat) approved 248 bills (in Indonesian: Rancangan Undang-Undang, RUU), and as many as 50 bills were agreed as priorities in 2020 [1]. However, out of the 50 priority bills, several bills are considered highly controversial because the drafting or discussion process of the bill itself is deemed incompatible with the regulations and aspirations of the people. One of them is the Job Creation Bill (in Indonesian: RUU Cipta Kerja). According to the Confederation of Indonesian Workers Unions (KSPI), the Job Creation Bill adversely affects workers. Said Iqbal, the President of KSPI, conveyed nine reasons for rejecting the Job Creation Bill, which was deemed not to include three labor principles [2]. The rejection of the Work Creation Bill caught the public's attention

* Corresponding author.
E-mail address: indra@cs.ui.ac.id

1877-0509 © 2021 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2021.12.120

and various national print and electronic media. The public continues to voice their rejection on social media or in the form of strikes and demonstrations in various regions in Indonesia throughout 2020.

The masses against the Job Creation Bill organize several campaigns using hashtags and keywords on Twitter and other social media. Tweets containing hashtags #tolakomnibuslaw and #MosiTidakPercaya became trending topics on Twitter in Indonesia [3]. On the other hand, people who support the controversial bill also express their stance on Twitter.

Indeed, Twitter's social media is currently a growing platform in Indonesia. With the total number of users reached by ads to 10.65 million, Twitter is in the five most active social media in Indonesia. Twitter's users say that they can get the latest information faster, social media that saves quotas, and is freer to express themselves. Twitter can use pictures or short videos, but the main feature that is considered to provide these advantages is a “Tweet” in the form of 280 characters long text. Among the various data formats exchanged on social media, most of them are texts. Twitter allows users to post text to communicate the latest news, share information, and participate in events. This emerging medium has become a powerful channel of communication.

Text mining technique can be beneficial in capturing public opinion on government regulations or policies aimed at increasing community participation in decision making [4]. In this research, we aim to examine how Twitter conversation can capture community’s stance toward proposed legislation using text mining. We work on stance detection task. More specifically, we collect and annotate the tweets matching the keywords related to Job Creation Bill, and implement machine learning models to automatically detect the stance label for those tweets.

2. Related work

Text mining (also known as data mining in the form of text or knowledge discovery in textual databases) is a semi-automatic process of extracting patterns (useful information and knowledge) from a large number of unstructured data sources [5]. A text mining approach usually follows a general framework, beginning with data retrieval to form a text corpus, followed by preprocessing, representation, and knowledge discovery. One of the text mining tasks on social media is stance detection. Stance detection is a task to automatically determine whether the author of a text supports, opposes, or is neutral to the proposition or target in the test [6].

Stance detection is also known as stance classification, stance identification, stance prediction, debate-side classification and debate stance classification [7]. Stance detection is a classification problem where the alignment of the author of the text is sought in the form of category labels from the set: {Favor, Against, Neither}. Occasionally, the Neutral category label is also added to the category set. The target of writing text can be mentioned or not stated implicitly or explicitly.

A survey [7] found that in the feature-based machine learning approach using Support Vector Machine is by far the most commonly employed feature-based machine learning approach for stance detection. Support Vector Machine is used in more than 40 studies on stance detection, either as the main best-scoring approach or as the baseline approach against which other approaches are compared. Logistic Regression and Naïve Bayes are the second and third most frequent classifiers used for stance detection.

Stance detection task has been studies in several area of application. Considerable effort is devoted to the creation of stance-annotated datasets. Somasundaran and Wiebe [8] constructed a dataset of stance-annotated posts from online forum debating the product brands among competitors. Mohammad et.al. [9] crowdsourced the three label annotation of stance (favor, against, and no stance) corresponding to six different targets of interest (i.e. ‘Atheism’, ‘Climate Change is a Real Concern’, ‘Feminist Movement’, ‘Hillary Clinton’, ‘Legalization of Abortion’, and ‘Donald Trump’) on English tweet data. Jannati, et.al. [10] modeled the stance of the blog author whether he support or deny a political figure to compete and win a general election event, for example a candidate of President in the Presidential election. They experimented with feature engineering, i.e. n-gram, sentiment lexicon, ortography, and word embeddings.

Kaunang, et.al. [11] identifies public opinions on the use of e-cigarette in Indonesia by detemining the stance of tweets. They proposed pipeline model in order to handle the tweet with no stance. Several other works on stance detection determined the support or opposition to proposed legislation [12], 2016 referendum on reform of the Italian Constitution [13], and the bill/issue of the speech under consideration in Indian parliamentary [14].

3. Methodology

We extract public stance on the Job Creation Bill using text mining approach. The type of data used in this study is secondary data from Twitter. We follow the methodology: data collection and annotation, data pre-processing, feature extraction, model building, and evaluation.

3.1. Data collection

We crawl tweets using Python tools and library. The queries used are as follows:
"ruu cipta kerja" OR "#ruuciptakerja" OR "#ciptakerja" OR "ruu ciptaker" OR "#ruuciptaker" OR "#ciptaker" OR "uu cipta kerja" OR "#uuciptakerja" OR "uu ciptaker" OR "#uuciptaker" OR "omnibus law" OR "#omnibuslaw" OR omnilaw OR "#omnilaw" OR omnibuslaw OR "ruu cilaka" OR "#ruucilaka" OR "uu cilaka" OR "#uucilaka" OR "ruu cipta lapangan kerja" OR "#ruuciptalapangankerja"

Tweets taken are from the period of 25 October 2019 to 25 October 2020. This period is the beginning of the creation of the Job Creation Bill until its ratification. Data is retrieved and stored in XLSX format documents. These documents contain Tweet data per month, so there are 13 documents. The data collected amounted to 1,114,773 data. Deduplication, removing URLs, and changing each mention to @Anon were carried out to clarify the contents of the Tweet text for the annotation process.

3.2. Annotation

We apply stratified random sampling to sample the data. We select 1.79% data for each month during period and have total of 10,050 tweets being annotated. The tweet is annotated with single label of four options.

Table 1 exemplifies the annotated tweets. The example of PRO tweet contains information about someone who supports the bill by ordering his subordinates to accelerate the production of the bill and also believes that the bill is needed by the community. The example of ANTI tweet contains explicit rejection of the Job Creation Bill because the author believes the bill is detrimental to the community. In the example of ABS tweet, the author does not show support or disapproval of the bill. On the other hand, the example of IRR tweet contains the keyword but aim to discuss things other than the Job Creation Bill; the focus of the author in this Tweet is on his/her Twitter timeline which happens to be full of Tweets related to the bill.

Table 1. Example of annotated tweets.

Labels	Tweet example
PRO	*Jokowi Perintahkan Menteri Kebut Omnibus Law Cipta Lapangan Kerja "Goal besar pekerjaan kita adalah cipta lapangan kerja. Karena ini yang dibutuhkan, masyarakat. Jangan sampai ada kementerian, provinsi, kabupaten, kota yang tidak mengerti" kata @Anon*
ANTI	*Kami bagian dari buruh jakarta baik aliansi gebrak dan gerakan buruh di seluruh indonesia menyatakan menolak adanya omnibus law dan RUU Cipta lapangan kerja yg jrlas merugikan kaum pekerja dsn rakyat indonesia.*
ABS	*Masih nunggu dari sudut pandang yg lebih ahli soal RUU Cipta Kerja*
IRR	*Bnr2 isi tl omnibuslaw semua*

Each tweet is initially annotated by two independent annotators. We compute an inter-agreement measurement on data that had been annotated by the two independent annotators The coefficient of Cohen's Kappa [15] for annotation of our data is 0.54 which is based on [16] including a moderate level of agreement. For a subset of tweets that are labeled differently between the two annotators, another annotator assign the third label. In total, we have 9,440 labeled data. The label distribution of our data is as follows. There are 1,827 PRO tweets (19.35%), 4,417 ANTI tweets (46.79%), 1,723 ABS tweets (18.25%), and 1,473 IRR tweets (15.60%).

3.3. Preprocessing

Specific characters, i.e. punctuations, digits, multiple whitespaces, mentions, and non alphanumeric characters, are removed from the text. All capital letters are transformed into lowercase. The elongated words are normalized in order that no more than two consecutive letters are retained. Lexical normalization is applied using colloquial dictionary [17]. Text preprocessing also includes stemming and stopword removal using Sastrawi tool.

3.4. Model building and evaluation

We represent the tweet as the features of the word vector. In this study, vectorization using TF-IDF weighting which produces unigram, bigram, and unigram+bigram vectors. Two scenarios of feature extraction are performed, each using the parameters max_feature = 5000 and max_feature = 10000. We experiment with three classification algorithms: Multinomial Naïve Bayes, Support Vector Machine, and Logistic Regression.

We split the data into test-train partition (7,552 instances in training set and 1,888 in testing set). We evaluate the model using a micro F1-score metric.

4. Result and analysis

The micro F1-score of the models using 5000 and 1000 features can be seen in Table 2 and Table 3, respectively.

Table 2. Micro F1-score for each model using 5000 features.

Experiment	Unigram (%)	Bigram (%)	Unigram+Bigram (%)
Multinomial Naïve Bayes	66	63	68,9
Support Vector Machine	70,1	61,8	70,6
Logistic Regression	71,1	62,4	71,1

Table 3. Micro F1-score for each model using 10000 features.

Experiment	Unigram (%)	Bigram (%)	Unigram+Bigram (%)
Multinomial Naïve Bayes	63,5	62,8	68,3
Support Vector Machine	70,4	64	71,1
Logistic Regression	71,3	64,2	71,8

Using 5000 features, the Logistic Regression classifier with the unigram feature performs the best with F1 score of 71.13%, slightly better than Logistic Regression classifier with unigram + bigram features (71.08%). As seen in Table 3, using bigram features tend to have lower evaluation performance compared to using unigram and unigram + bigram features. Support Vector Machine classifier with bigram features only scores 61.8%.

On the other hand, for the classification with 10,000 features, the Logistic Regression classifier also outperform the others. The model using unigram + bigram features achieves a micro F1-score of 71.8%. The experiments with 5,000 features shows the similar pattern with 10,000 features, in which the model with bigram features has lower performance compared to other models. Multinomial Naive Bayes model with bigram feature becomes the model with the lowest performance with a micro F1-score of 62.8%.

Table 4 breakdown the precision, recall, and F1-score of Logistic Regression model that is considered as the best classifier in our experiment. While the model capability to predict PRO and ANTI labels is relatively satisfying, it still finds problem when dealing with ABS and IRR labels.

Table 4. Precision, recall and F1-score for best sodels.

Models	Metrics	Labels			
		PRO	ANTI	ABS	IRR
5000 Features (LR Classifier with Unigram Feature)	Precision	0.746	0.751	0.541	0.685
	Recall	0.726	0.883	0.441	0.495
	F1-score	0.736	0.812	0.486	0.575
10000 Features (LR Classifier with Unigram+Bigram feature)	Precision	0.760	0.756	0.546	0.695
	Recall	0.748	0.889	0.461	0.471
	F1-score	0.754	0.817	0.500	0.562

Misclassification mostly occurs due to multi-targets or unclear targets in the Tweet. Other misclassification occurs because half of the text uses a language other than Indonesian or due annotation error. Annotation error may happen because the bill covers a fairly wide scope.

For example, the following tweet is an ABS labeled Tweet but predicted as ANTI by the model because it contains multiple targets.
kalo enggak salah surat isi perintah jawa barat tolak omnibuslaw serikat kerja jadi perintah jabar cuma jadi jembatan enggak jamin batal omnibuslaw tolak jalan dalam omnibuslaw memang laku

On the other hand, the following tweet is an ANTI labeled Tweet but predicted as IRR because of half of its text is in foreign language (English).
perintah tuh sogok berapa sih sama usaha pabrik asing uu cipta kerja keluar ajhgbsh i remember when all vote jokowi all didn t sign for this right

5. Conclusion

This study built the stance detection model that can identify a tweet content into groups that support, reject, and abstain from the Job Creation Bill. We experimented with three classifiers, i.e. Multinomial Naive Bayes, Logistic Regression, and Support Vector Machine by implementing a combination of n-gram features. The Logistic Regression classifier with Unigram + Bigram feature achieved the best performance in our experiment with an average micro F1-score of 71.8%.

Further research may consider using advanced model, e.g. deep learning architecture. Modeling topics that become the center of attention in the discussion of the Job Creation Bill can be interesting direction. Exploring the stance detection task for other bills may be useful case studies in accommodating people's aspirations and in making decisions regarding the formation of future bills.

Acknowledgment

We would like to thank Universitas Indonesia for their support through grant “Publikasi Terindeks Internasional (PUTI) Q2: NKB-4060/UN2.RST/HKP.05.00/2020”.

References

[1] DPR RI. (2020) “Program Legislasi Nasional [Title in English: National Legislation Program].” [Online]. Available: https://www.dpr.go.id/uu/prolegnas-long-list. [Accessed October 30, 2020].

[2] Purnamasari, D.M. (2020) “Ini 9 Alasan KSPI Menolak Omnibus Law Cipta Kerja [Title in English: These are 9 reasons KSPI rejects the omnibus law on job creation].” [Online]. Available: https://nasional.kompas.com/read/2020/02/16/15353591/ini-9-alasan-kspi-menolak-omnibus-law-cipta-kerja?page=all#:~:text=Ketiga%20hal%20itu%20adalah%20job,atau%20jaminan%20sosial%20terhadap%20pekerjaan. [Accessed November 16, 2020].

[3] Siregar, H. (2020) “Trending Topik di Twitter Diborong Tagar Penolakan UU Cipta Kerja, Akibat RUU Omnibuslaw Disetujui DPR [Title in English: Trending Topics on Twitter Purchased the Rejection of the Job Creation Law, Due to the Omnibus Law Bill being Approved by the

DPR].” [Online]. Available: https://daulatkepri.co.id/2020/10/05/trending-topic-populer-di-twitter-diborong-tagar-penolakan-uu-cipta-kerja-akibat-ruu-omnibus-law-disetujui-dpr/. [Accessed January 4, 2021].

[4] Stylios, George, D. Christodoulakis, Jeries Besharat, Maria-Alexandra Vonitsanou, I. Kotrotsos, Athanasia Koumpouri, and Sofia Stamou. (2010) “Public Opinion Mining for Governmental Decisions.” in *Electronic Journal of e-Government* **8**:83–92.

[5] Sharda, Ramesh, Dursun Delen, and Efraim Turban. (2015) *Business Intelligence and Analytics : Systems for Decision Support* (Tenth), New Jersey, Pearson Education.

[6] Mohammad, Saif M., P. Sobhani, and S. Kiritchenko. (2017) “Stance and Sentiment in Tweets.” *ACM Transactions on Internet Technology* **17 (3)**.

[7] Küçük, D., and Fazli Can. (2020) “Stance Detection : A Survey.” *ACM Computing Surveys* **53 (1).**

[8] Somasundaran, S., and J. Wiebe. (2009) “Recognizing stances in online debates.” *Proceedings of the Joint Conferenceof the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* **1**: 226–234.

[9] Mohammad, Saif M., S. Kiritchenko, P. Sobhani, Xiaodan Zhu, and Colin Cherry. (2016) “A dataset for detecting stance in tweets.” *Proceedings of the Language Resources and Evaluation Conference:* 3945–3952.

[10] Jannati, R., Rahmad Mahendra, C. W. Wardhana, and Mirna Adriani. (2018) “Stance Classification towards Political Figures on Blog Writing.” *2018 International Conference on Asian Language Processing (IALP)*: 96–101.

[11] Kaunang, C. P. S., F. Amastini, and Rahmad Mahendra. (2021) “Analyzing Stance and Topic of E-Cigarette Conversations on Twitter: Case Study in Indonesia.” *IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*: 304-310.

[12] Thomas, M., Bo Pang, and Lillian Lee. (2006) “Get out the vote: Determining support or opposition from congressional floor-debate transcripts.” *In Proceedings of the Conference on Empirical Methods in Natural Language Processing*: 327–335.

[13] Lai, M., V. Patti, G. Ruffo, and P. Rosso. (2018) “Stance evolution and Twitter interactions in an Italian political debate.” in *Proceedings of 23rd International Conference on Natural Language and Information Systems.*

[14] Rohit, S. V. K., and N. Singh. (2018) “Analysis of speeches in Indian parliamentary debates.” *arXivpreprint arXiv:1808.06834.*

[15] Cohen, J. (1960) “A Coefficient of Agreement for Nominal Scales.” **XX (1)**: 37–46.

[16] Mchugh, M. L. (2012) “Lessons in biostatistics Interrater reliability: the kappa statistic.” *Biochemia Medica*: 276–282.

[17] Salsabila, N. A., Yosef Ardhito W., Ali Akbar S., and Ade Jamal. (2018) “Colloquial Indonesian Lexicon.” *International Conference on Asian Language Processing (IALP)*: 226-229.