

Are University Students Independent: Twitter Sentiment Analysis of Independent Learning in Independent Campus Using RoBERTa Base IndoLEM Sentiment Classifier Model

Lidia Sandra¹, Marcel², Gatot Gunarso³, Fredicia⁴ and Olfien W Riruma⁵

¹Psychology Department, Faculty of Psychology

²Information System Department, ⁴Informatics, Faculty of Engineering and Computer Science

³Management Department, Faculty of Economics and Business

Krida Wacana Christian University, Jakarta 11470, Indonesia

Email: ¹lidia.sandra@ukrida.ac.id, ²marcel@ukrida.ac.id, ³gatot.gunarso@ukrida.ac.id, ⁴fredi.cia@ukrida.a.c.id,
⁵olfien.riruma@ukrida.ac.id

Abstract— Higher education is one of the fields that is affected by the COVID-19 pandemic. One of the regulations in Indonesian higher education is Merdeka Learning - Independent Campus (MB-KM). In this research article we are aiming to do an analysis of the sentiment using Twitter as the dataset in order to find out the sentiment of university students towards the program. The analysis are being done using RoBERTa Base IndoLEM Sentiment Classifier Model. With the total number of more than 30 thousands positive sentiments and almost 20 thousands negative sentiments found. The result of the model shows that it achieved an accuracy of 91.73%, with a precision of 83.33%, recall 89.74%, and a harmonic mean of the two (F1 score) of 86.42%. Based on the analysis, it also shows the distribution of the sentiment is 63.0% of positive sentiment and 37.0% of negative sentiments. This paper shows that there are more positive sentiments than the negative one.

Keywords— sentiment analysis; classification; twitter; merdeka learning; independent learning; MB-KM

I. INTRODUCTION

The industrial revolution 4.0, Society 5.0 and the COVID-19 pandemic are a strong combination that

drives the acceleration of digital transformation both nationally and at the institutional level [1]. Indonesia as one of the countries with great potential in terms of economy and socio-cultural diversity does not remain silent and plays an active role in the world's digital ecosystem, this can be seen on Table 1 referenced from the country ranking where Indonesia occupies the fifth position in the world from the number of national startups [2].

Table 1. Top 5 Startups per Country by Startup Ranking [2]

No.	Country	Startups
1	United States	70.591
2	India	12.377
3	United Kingdom	6.133
4	Canada	3.214
5	Indonesia	2.322

Temasek, Google, and Baik & Company in their report e-Conomy SEA 2021 stated that the potential value of Indonesia's digital economy in 2021 reached US\$ 70 billion (the largest in the ASEAN region), a significant increase compared to 2020 which was only US\$ 47 billion [3]. Data from the Ministry of Finance and the Ministry of Trade also shows the potential for

even greater growth, up to eight times by 2030, worth around IDR 4,531 trillion [4]. The results of the latest study conducted by East Ventures through the Digital Competitiveness Index 2021 also shows that the pandemic condition is one of the factors that drives the acceleration of digital transformation in Indonesia and there is a tendency to improve in terms of equity (reducing digital disparities between provinces) [5]. These are: first, the development of the national communication network infrastructure; second, there is an increase in spending on ICT-based services or products both among households and institutions.

Efforts to realize the potential of the national digital economy as a result of digital transformation are not without obstacles, data from the World Bank shows that Indonesia lacks human resources with ICT talents, at least 600 thousand quality digital talents are needed every year [6]. Data for 2020 released by UNDP for the Human Development Index shows that Indonesia is in fifth position for the ASEAN region, although in terms of the index it shows that Indonesia is already in the high category, but there are still many improvements that need to be made, especially at the national education level [7].

Seeing these conditions, the Ministry of Education and Culture released a strategic program called Merdeka Learning - Independent Campus (MB-KM) which gives students the right to study for up to 3 semesters outside of their study program. The key to the MB-KM program is not about the right to study for 3 semesters, but about efforts to narrow the gap in the quality of graduates with the real needs in the industrial world [8]. This MB-KM program also offers different kinds of learning activities that the students can choose depending on their own preference [9]. These learning activities are student exchange (*pertukaran pelajar*), internship (*magang*), teaching assistance in lower education (*asistensi mengajar di satuan pendidikan*), research (*penelitian*), humanity project (*proyek kemanusiaan*), entrepreneur activity (*kegiatan wirausaha*), independent study or project (*studi proyek independen*), and a thematic community service program (*membangun desa kuliah kerja nyata tematik*) [10].



Figure 1. Learning Activities of MB-KM

The key to the success of implementing MB-KM in universities lies in the combination of 2 strategies, namely a combination of collaboration and digital transformation [11]. Collaboration here is also expanded not only with lecturers and students, but also with the industrial world; while digital transformation talks about how universities are able to create a digital ecosystem that supports a flexible and collaborative teaching-learning process [12]. From the combination of these two strategies, the learning experience received by students will lead to student-centered and experiential through participatory and collaborative approaches that emphasize the discussion of case studies and projects that simulate real problems in the field.

So it is also a question of how universities are required to be able to produce students who are ready to face the era of digital transformation, but universities as institutions do not try to carry out digital transformation? Does digital transformation just stop being just a jargon in lectures? Considering the role of universities as a 'gateway' of knowledge and being one of the main pillars to produce talent with quality digital competencies, then transformation is clearly not an option but a necessity in the long term to maintain the existence and sustainability of the university itself [13]. The initial strategy that can be done is to adopt a framework or framework to determine the dimensions of the scope and priorities of the digital transformation that will be carried out.

II. METHOD

The sentiment analysis are being used to understand the nodes' responses in higher education about the implementation of Merdeka Learning - Independent Campus (MB-KM) during the pandemic. Presently, sentiment analysis is a matter of great interest and growth since it has some useful applications [14]. Since publicly and privately accessible knowledge on the Internet is always developing, a large number of text expressing opinions are available at review websites, forums, blogs, and social media [15]. Tools that are being used are Pandas to gather data from Twitter to collect the dataset. In this paper, Pandas are being used because it is a very useful tool especially for reading and writing data into data structures or files, it is also providing a powerful aggregation function that can be used to manipulate the data that we have [16]. We are also using Tweepy to gather the data from Twitter by connecting to Twitter API to collect the dataset also. Tweepy itself is a library for Python that we can use to access Twitter API, with an access to it we can collect the needed data from Twitter [17]. The first thing that we do is defining the functions to do the text pre-processing and also the scrapping first, and then deciding on the keywords, the period of time which the Tweet is posted, and a total number of Tweet. The process of developing this paper in executing the analysis of sentiment is to do Data Collection, Pre-processing, continued with feature extraction, and lastly creating the model. We can see the flowchart of the study from the Figure 2 below.

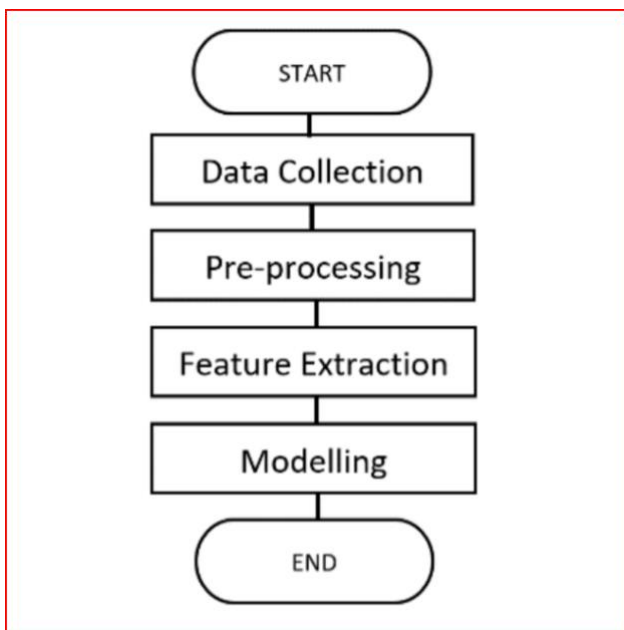


Figure 2. Flowchart

A. Data Collection

Data collection is the systematic way to collect and evaluate data from a variety of sources to make a full and precise picture of the area of interest [18]. Data collection enables us as an individual or even organisation with a group of people in this research to tell pertinent questions, assess results and make predictions about future probabilities, trends, and answering this paper's goal [19]. There are two tools that are employed to collect the dataset from Twitter. These tools are Tweepy and Pandas. First, we request Twitter API to get permission on the database and also general data. After that, it comes to defining the function for text pre-processing later. This function includes defining the keywords we need, the period of the Tweet posted and the number of the Tweet data gathered. There are indicators that we set for the dataset, which are the content of the Tweet itself, completed with each Tweet ID and how many retweets it has, also the number of likes, and the coordinate location of the author. The dataset that was scraped will be guaranteed to contain the keywords that consist of: "kampus merdeka", "MBKM", "merdeka belajar", "magang", "magang bersertifikat", "kampus mengajar", "proyek independent", "pertukaran mahasiswa", Merdeka Learning - Independent Campus (MB-KM). The total of the sentiments gathered are 50.881 sentiments which can be represented in the Figure 3 below.

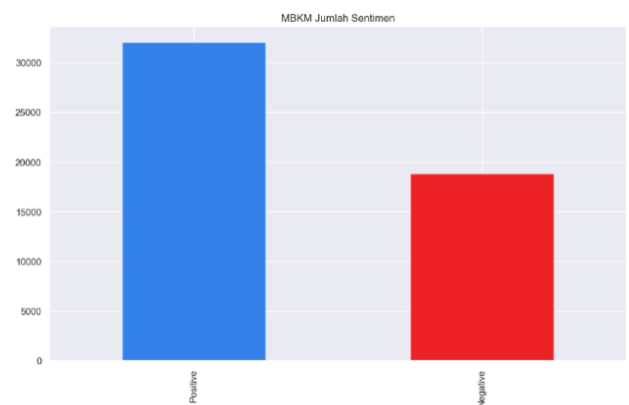


Figure 3. Total Number of Sentiments Gathered

B. Pre-processing

Information or data pre-processing is the data mining method that involves translating raw data into understandable information [20]. Real-world information is often incomplete, contradictory, and or lacking in specific behaviours or tendencies, and is expected to be some mistakes [21]. When the dataset are collected, pre-processing processes are to be done to make sure that the extraction process of the feature on the database can be gathered successfully. Then,

we load the list that contains the words with a positive sentiment characteristics and negative sentiment, also the stop words are applied in Bahasa. The data is then randomly selected for a data training and modelling test, with the previously uploaded list of phrases being utilized to filter these datasets. Based on the stop words we submitted, this will filter out characters and words. These also resulting in analysing of the amount to total retweets from each data. The number of retweets gained from positive sentiments are 30.546 as for the negative sentiments garnered 14.071 retweets. The number of retweets can be seen in Figure 4.

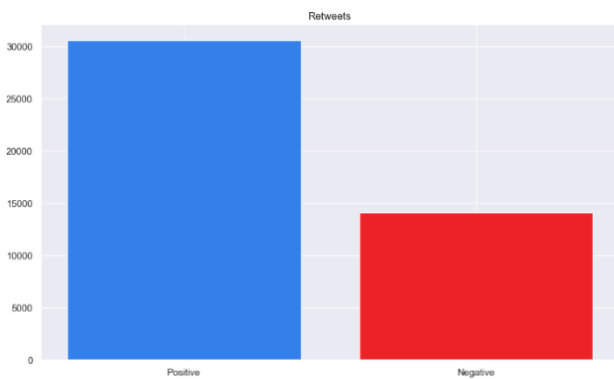


Figure 4. Number of Retweet in Each Sentiments

The next step is to examine the number of negative and positive terms in the list we previously uploaded, and then count the remaining positive and negative thoughts. Negative sentiments will be represented by 0 and positive sentiments by 1. The method reveals that a total of 50.881 data points is collected., 32.011 is a positive sentiment, while the rest which is 18.800 are negative sentiments. The next step after that is to analyse the number of likes each of the data has. For the positive sentiments we have 30.546 total likes, while the negative mass an amount of 14.071 likes for the sentiments. Number of likes can also be seen in the Figure 5.

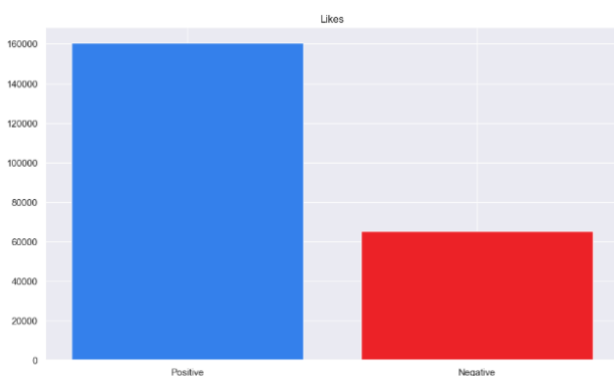


Figure 5. Number of Like for Each Sentiments

C. Feature Extraction & Modelling

The Indonesian RoBERTa Base IndoLEM Sentiment Classifier model made by Wilson Wongso was used for the MBKM sentiment analysis. It is a model based on the previously created Indonesian sentiment analysis RoBERTa Base model trained on the OSCAR dataset. The model is subsequently fine-tuned using the IndoLEM's sentiment analysis dataset. The model was trained and evaluated using a 5-fold cross-validation procedure, with the model used being trained on the first fold. Overall, it achieved an accuracy of 91.73%, with a precision of 83.33%, recall 89.74%, and a harmonic mean of the two (F1 score) of 86.42%.

To conduct sentiment and word cloud analysis on the MBKM program, several processes were done. Firstly, tweets regarding the program were gathered from the first of July till 25th of December 2021, with the following keywords: “kampus merdeka”, “MBKM”, “merdeka belajar”, “magang”, “magang bersertifikat”, “kampus mengajar”, “proyek independent”, “pertukaran mahasiswa”. Then, each tweet's sentiment was classified using the Indonesian RoBERTa Base IndoLEM Sentiment Classifier model. Each tweet is then also pre-processed, by standardizing the character case for every tweet, the removal of unnecessary characters (e.g., https, numbers, @usernames), slang normalization, and stop word removal was done to ensure that the wordcloud used only meaningful words.

III. RESULTS

The sentiment analysis results for the keyword "Merdeka Learning - Independent Campus (MB-KM)". Sentiment analysis data analysis is meticulously carried out as follows. The results of sentiment analysis using the MB-KM keyword revealed that the majority of sentiment (63.0 percent) was positive, with the remainder being negative (37.0 percent). Figure 6 shows that positive sentiments appear to outnumber negative sentiments.

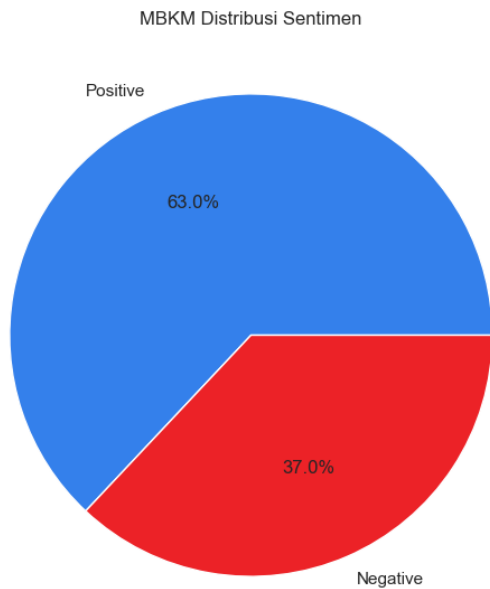


Figure 6. Sentiment Distribution

From the word cloud on MBKM's positive sentiment, a few words like thank you, wonderful Indonesia, the Indonesian president Joko Widodo also appears quite often, culture and research, as well as the word independence for learning, "yuk" as an expression of invitation from participants who are satisfied with the various activities of BKP MBKM and also pocket money which seems to be the driving force for student participation in BKP MBKM. The frequency of terms that occur the most in each intriguing sentiment is analyzed, and it is discovered that the key words that exist in all categories nevertheless include negative sentiments such as: kampus merdeka (freedom campus), magang (internship), kampus mengajar (teaching campus), MBKM, program kampus (campus program). All of which can be represented in a word cloud in Figure 7 below.



Figure 7. Negative Sentiments

While in negative word cloud, several difficulties arose such as, value conversion, tired, sad (when they did not pass the MBKM flagship), paid, date, which stated some delays in payment of incentive money, difficult, lacking which indicated that some felt the existing program had not really been implemented maturely and afraid to express the worries and doubts of some students to take part in the Merdeka Learning program at the Merdeka Campus. The phrase "kampus merdeka," which translates as "freedom campus," appears the most frequently. The most discussed fascinating terms in the area of positive sentiment are merdeka belajar (freedom of learning) and program kampus (campus program). People who are directly touched by the implementation of Merdeka Learning may express positive opinions. - Independent Campus (MB-KM). Magang (internship) is the most basic of MB-KM's influence now. These can be represented in the Figure 8 below.

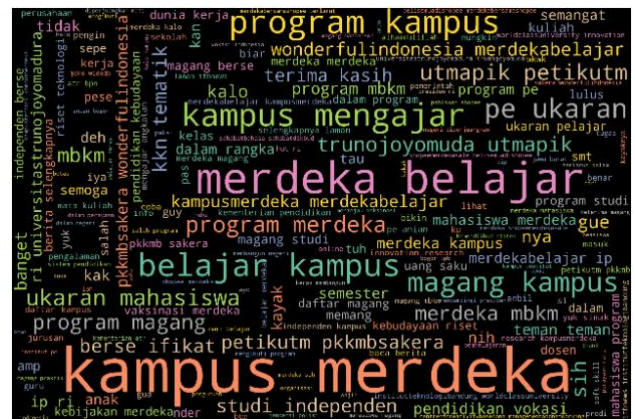


Figure 8. Positive Sentiments

IV. DISCUSSION & CONCLUSION

After doing a sentiment analysis using RoBERTa Base IndoLEM Sentiment Classifier Model It is possible to deduce based on the data set that. Overall, the model obtained an accuracy of 91.73%, with a precision of 83.33%, recall 89.74%, and a harmonic mean of the two (F1 score) of 86.42%. The data set also reveals that the most common sentiment is connected to Merdeka Learning - Independent Campus (MB-KM) has a positive sentiment.

The result is a representative of what the university students thinks about the whole MB-KM program. Looking at the word cloud we can see that participants who are satisfied with the various activities of BKP MB-KM discussing about pocket money which seems to be the driving force for student to participate in BKP MB-KM. Also from the word cloud we can conclude that some felt the existing program had not

really been implemented maturely and afraid to express the worries and doubts of some students to take part in the BKP MB-KM program.

V. ACKNOWLEDGEMENT

Funding assistance for Merdeka Learning - Independent Campus (MB-KM) policy research programs and community service based on research results and PTS – Perguruan Tinggi Swasta (Private Higher Education Institution) prototypes in collaboration with UKRIDA and the Director General of Higher Education, Research and Technology in 2021.

REFERENCES

- [1] Javaid, M., Haleem, A., Vaishya, R., Bahl, S., Suman, R., & Vaish, A. (2020). Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), 419-422.
- [2] Startup ranking: Find the top and new startups worldwide. StartupRanking. (n.d.). Retrieved December 26, 2021, from <https://www.startupranking.com/countries>
- [3] Google. (n.d.). E-conomy sea 2021: Post-pandemic marketing - think with google apac. Google. Retrieved December 28, 2021, from <https://www.thinkwithgoogle.com/intl/en-apac/consumer-insights/consumer-trends/e-conomy-sea-2021/>
- [4] Kemenkeu. (n.d.). Menuju 2030: Transformasi Kekuatan ekonomi natural resources ke human capital. Kementerian Keuangan. Retrieved December 28, 2021, from <https://www.kemenkeu.go.id/publikasi/berita/menuju-2030-transformasi-kekuatan-ekonomi-natural-resources-ke-human-capital/>
- [5] East Ventures indonesia digital competitiveness index 2021. East Ventures. (2021, March 23). Retrieved December 28, 2021, from <https://east.vc/dci/>
- [6] Indonesia. Data. (n.d.). Retrieved December 28, 2021, from <https://data.worldbank.org/country/ID>
- [7] Human development reports. Human Development Index (HDI) | Human Development Reports. (n.d.). Retrieved December 28, 2021, from <http://hdr.undp.org/en/content/human-development-index-hdi>
- [8] Kampus Merdeka. (n.d.). Retrieved December 28, 2021, from <https://kampusmerdeka.kemdikbud.go.id/>
- [9] Nyoto, N. (2021). Perception of PGSD FKIP UPR Students on the Independent Campus Learning Program. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4(4).
- [10] Effendi, D. I., Lukman, D., Eryanti, D., & Garnita, A. (2021). *Dokumen Kurikulum MBKM-KKNI BKI FDK UIN SGD Bandung*.
- [11] Fuadi, T. M., & Aswita, D. (2021). Merdeka Belajar Kampus Merdeka (MBKM): Bagaimana Penerapan dan Kedala Yang Dihadapi oleh Perguruan Tinggi Swasta di Aceh. *Jurnal Dedikasi Pendidikan*, 5(2), 603-614.
- [12] Baharuddin, M. R. (2021). Adaptasi Kurikulum Merdeka Belajar Kampus Merdeka (Fokus: Model MBKM Program Studi). *Jurnal Studi Guru dan Pembelajaran*, 4(1), 195-205.
- [13] Krishnapatria, K. (2021). Merdeka Belajar-Kampus Merdeka (MBKM) Curriculum in English Studies Program: Challenges and Opportunities. *ELT in Focus*, 4(1), 12-19.
- [14] Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (Eds.). (2017). *A practical guide to sentiment analysis* (pp. 1-196). Cham, Switzerland: Springer International Publishing.
- [15] Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013, October). Comparing and combining sentiment analysis methods. In *Proceedings of the first ACM conference on Online social networks* (pp. 27-38).
- [16] Al Bashaireh, R., Zohdy, M., & Sabeeh, V. (2020, May). Twitter Data Collection and Extraction: A Method and a New Dataset, the UTD-MI. In *Proceedings of the 2020 the 4th International Conference on Information System and Data Mining* (pp. 71-76).
- [17] Shelar, A., & Huang, C. Y. (2018, December). Sentiment analysis of twitter data. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1301-1302). IEEE.
- [18] Burkard, A. W., Knox, S., & Hill, C. E. (2012). *Data collection*.
- [19] Sapsford, R., & Jupp, V. (Eds.). (1996). *Data collection and analysis*. Sage.
- [20] Famili, A., Shen, W. M., Weber, R., & Simoudis, E. (1997). *Data preprocessing and intelligent data analysis*. *Intelligent data analysis*, 1(1), 3-23.
- [21] García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Cham, Switzerland: Springer International Publishing.