

# Public Perceptions of Online Learning in Developing Countries: A Study Using The ELK Stack for Sentiment Analysis on Twitter

<https://doi.org/10.3991/ijet.v15i09.11579>

Satria Fadil Persada <sup>(✉)</sup>, Andri Oktavianto, Bobby Ardiansyah Miraja,  
Reny Nadlifatin  
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia  
satriafadil@mb.its.ac.id

Prawira Fajarindra Belgiawan  
Institut Teknologi Bandung, Bandung, Indonesia

A. A. N Perwira Redi  
Universitas Pertamina, Jakarta, Indonesia

**Abstract**—This study explores public perceptions of online learning applications in Indonesia. Many studies about online learning were done in developed countries and only a few in developing countries. Moreover, these studies used a qualitative approach which limits the results to be applied in different settings. While traditional research using a survey to understand people's perceptions towards an entity requires a lot of time and effort, we used efficient and effective manners to gather opinions and then analyzed its sentiments using the Logstash, Kibana, and Python programming language (ELK stack) and Naïve Bayes algorithm. We used the Naïve Bayes algorithm for sentiment analysis and ELK stack for storing & gathering tweets from Twitter. With ELK stack, we successfully collected 133.477 tweets related to online learning. From this study, we understood what kind of words that are sentimentally positive and negative tweets. We also gained some insights regarding Indonesia's student online learning application preferences.

**Keywords**—Online Learning, Sentiment analysis, Twitter, Developing Country, ELK stack

## 1 Introduction

Education, in the past, was typically a one-way knowledge transfer: from teacher to student. There is a downside from this kind of learning method: it fails to emphasize problem-solving and critical thinking while both are considered to be critical learning skills in the 21st century [1]. One-way knowledge transfer also fails to address the knowledge requirements of this rapidly expanding technological society [2]. There have been several studies that successfully explains that, compared to only face-to-

face learning, blending face-to-face classes with online learning technologies could significantly improve students' overall learning process [3, 4]. The use of online learning technologies is called online learning. Furtherly defined, online learning is the use of online mode to deliver at least 80% of the course content [5].

The existence of online learning diminishes a typical process of education that requires students to sit down in the classroom and to listen to materials that they might already be familiar with. Furthermore, a particular advantage of online learning is how teachers and students are not required to be online at the same time; this could reduce by about a third of the total required learning time [6]. However, these advantages are meaningless when students do not have an interest or a good perception of online learning. To the extent, the perception of a students' learning experiences can influence their decision whether to continue with an online course or not [7]. Therefore, it is important to understand students' perceptions better to improve our education system.

Several studies have already investigated students' perception of online learning, and there are varied opinions; some studies show positive student perceptions, while others show concerns for online learning as a learning mode [8-10]. Nevertheless, the majority of students in the studies above show positive opinions towards online learning. However, two of these studies were using quantitative methods, and while quantitative approaches are great in providing descriptive data, sometimes researchers will encounter difficulties when it comes to interpretation. For example, if there is data about the amount of positive and negative sentiment received for each online learning provider in Indonesia, we can know the numbers but it is challenging to know the reasoning behind the positive or negative sentiment because of the absence of their opinions. Also, one of these studies were done using qualitative data analysis, in where a focus group was conducted with the students during one of the weekly class sessions just after the MOOC ended; although there is nothing wrong with using qualitative analysis, it is known for its limits: findings only applicable in specific settings such as specific location or demographic [11]. Other than that, these studies have also been conducted only in developed countries, limiting its practical significance for countries with different economic outlook. The present research also considers the aspect as mentioned above, acknowledging that access to online learning in a developing country has not been the norm, making it is more important to understand students' perception of online learning. For our study, a sample developing country was used: Indonesia. Forecasted to have 20.9 million active Twitter users in 2018 [12], Indonesia is an interesting country to be used as an object for this study. Thus, this study will use Twitter users from Indonesia as the object of this study.

To understand the sentiment of a particular subject better, newer methods are being developed, such as combining data mining with sentiment analysis. Rather than requiring specific settings, this method could capture messages from many and diverse audiences [13]. This is possible by utilizing sentiment analysis and Twitter stream API which many research has successfully done, Kharde and Sonawane use Naïve Bayes algorithm to classify 45.000 tweets from Stanford datasets; while Rosenthal et al. used about 19.000 tweets from public streaming Twitter API and various algorithm for sentiment analysis such as Support Vector Machine and Maximum Entropy [14,

15]. Sentiment analysis could explain people's sentiments, attitudes, opinions, and emotions towards products, individuals, topics, organizations, and services. In many previous studies, researchers had used public sentiment analysis for measuring brand reputation, discovering political campaign insights, and even predicting stock market movements using Twitter data [16-18].

To get the data from Twitter and analyze the sentiment, we used the ELK stack. Python is one of the most popular programming languages used for Sentiment Analysis, while ELK stack is commonly used to collect data from multiple data sources and visualize it in one dashboard.

The content structure of this paper is organized as follows: In this first section, we discussed the current situation and background of this study. In the next section, we reviewed relevant pieces of literature. In section 3, we described our approach and research methods. Section 4 presented the final results from our sentiment analysis using the dataset of Twitter tweets we gathered. Section 5 concludes the paper as well as limitations and some notes for future works.

## **2 Literature Review**

### **2.1 Online learning**

As new technologies become apparent, learning seemed to be differentiated into two methods: traditional and online. As there is still no exact definition of online learning and vague reference to other terms such as online course, web-based learning, web-based training, or distance learning, we will use this generally accepted definition: online learning is a course where most or all of the content is delivered online and typically have no face-to-face meetings [19]. Other terms that are often used to describe other forms of learning consisted of e-learning, digital learning, and distance learning. A piece of research suggested that e-learning is a type of online learning, so it is safe to say that online learning is broader than e-learning [20]. As pointed out by previous studies, when students are not interested or have a negative perception towards online learning, it may lead to boredom and disengagement in students and prevent them from taking full advantage of online learning.

### **2.2 Sentiment analysis**

Sentiment analysis is used to determine whether a sentence or a text expresses a particular sentiment: a positive one or a negative one [21]. In general, there are two approaches to do this, the lexicon-based approach and the machine learning approach.

This study will use machine learning, and the following paragraph will describe the general idea about this approach. The machine learning approach used in this study was supervised learning. It is an approach that utilizes labeled data, which is then used to train the classification algorithm. It means that we must first collect data with the positive, negative, and neutral sentiment as a training set, extract the features from the set and then train the classification algorithm based on the examples. When the classi-

fication algorithm used can classify positive, negative, and neutral text/opinion with high accuracy, then we can use that algorithm to conduct sentiment analysis. This study is using one of the popular classification algorithms, Naïve Bayes.

Naïve Bayes is an algorithm that classifies sentences by looking at its prior probabilities of the individual words and then use that as a basis for determining the future sentiment of sentences. For example, if the word "fortunately" in a dataset has 70% positive occurrences and 30% negative occurrences, the probability of the sentence with "fortunately" in it having positive sentiment will increase. In short, this method treats words separately and counts the number of occurrences of words [22].

Before the algorithm can start classifying text, it must be trained first with a training set and validation set.

1. Training set: Labeled dataset to train the classifier
2. Test/validation set: Labeled dataset to test the classifier accuracy

### **2.3 Developing countries**

There is a classification system developed by the International Monetary Fund (IMF), the United Nations Development Program, and the World Bank to assess a country's development status. In this study, we will adopt the World Bank income classification. Countries will be classified by their level of development as measured by gross national income (GNI) per capita valued annually in US dollars [23]. Accordingly, countries are grouped as high income, upper middle income, lower middle income, and low income. For the current 2020 fiscal year, countries with less than \$1.025 GNI per capita are classified as low-income, those with between \$1.026-3.995 are classified as lower-middle-income, \$3.996-12.375 are classified as upper-middle-income, and those with income more than \$12.376 are classified as high-income [24]. Countries with low and lower-middle-income are considered as developing countries. Indonesia is classified as a lower-middle-income country by the World Bank [25].

### **2.4 Twitter streaming API and text mining**

The Twitter Application Programming Interface (API) currently facilitates a Streaming API. Through the Streaming API, users can obtain real-time access to tweets in a filtered form [26]. "Filtered" means users can get specific tweets by using limitations such as specific keyword ("Manchester United", "Barcelona") and geographical location. By using Twitter Streaming API, we will get public accounts individual tweets in real-time. Retrieved from Twitter Developers website [27], a tweet object has a fundamental structure, as shown in Table 1:

**Table 1.** The structure of a tweet object

Attribute	Description
created_at	UTC when a Tweet is created
id	Integer representation of unique identifier for a tweet
user	The user who posted the tweet
text	The actual UTF-8 text of the tweet status/text

To get access to Twitter Streaming API, users must first apply for access by creating a Twitter app. Users will be required to provide personal details about themselves and the purpose of their app. Then after the application was approved, Twitter will generate consumer keys and access tokens. These keys and access tokens are required to access their Streaming API. There are many ways to access their Streaming API, but in this study, we will use Logstash for the sake of its ease of use.

Logstash is part of the ELK stack. In this study, Logstash was used to capture real-time tweet data and then adding that data to an Elasticsearch index. Elasticsearch index was used to store tweet data that has been retrieved by Logstash.

### 2.5 Online learning provider in Indonesia

Indonesian students are very dependent on outside of school education. In November 2012, Kompas did a survey about tutoring industry in Indonesia with results: 87.8% from 770 respondent agrees that students need to study or take lessons outside of school, with more than half of the respondents claimed that tutoring was needed because the subject taught at school was not sufficiently understood by students [28]. Offline providers previously dominated the tutoring industry in Indonesia. However, with the previously mentioned development of technology, online tutoring is now thriving. Three companies currently dominate the online tutoring industry in Indonesia: Zenius, Ruangguru, and Quipper.

## 3 Methodology

### 3.1 Research design

There are two types of machine learning methods that researchers can use to conduct sentiment analysis: unsupervised learning and supervised learning. In unsupervised learning, the process does not need any labeled data; in order to process unlabeled data, clustering algorithms are used [29]. Unlike unsupervised learning, supervised learning emphasizes dataset labels; hence, the process of supervised learning requires predetermined dataset labels in order for the model to work [30]. Unsupervised learning is proposed when it is not possible to have a training set of labeled opinions to classify words, but since it is possible to have a manually labeled training set, therefore supervised learning selected. This labeled dataset is then trained using the Naïve Bayes algorithm so it can classify texts as positive or negative. Thus, this study will try to investigate whether the result from sentiment analysis is valid for

understanding public opinions for physical and online education using the Naïve Bayes classifier method for sentiment analysis and the ELK stack for data mining. The ELK stack was chosen because of its ease of use.

We used a dataset from Twitter because it contains a considerable number of personal thoughts with public access [31], it became a valuable source to know people's opinions and sentiments towards a variety of topic or current issues. Since many people are voicing their opinion and view towards the education system on Twitter, sentiment analysis might be an appropriate method to discover public perceptions. Sentiment analysis could reveal peoples' opinions towards online learning as positive or negative. Using this method on microblogging sites like Twitter are proven to provide supportive information for decision making and handy tools for consumer research [32], and for this case, public perception.

### 3.2 Data pipeline

The following flow chart models the offered data pipeline for this study:

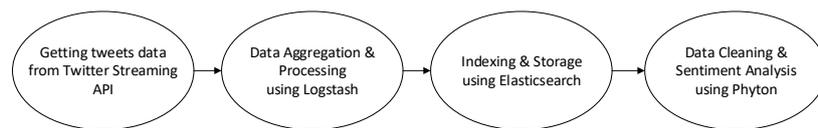


Fig. 1. Data Pipeline

As illustrated in Figure 1, the analysis process required the data from Twitter Streaming API, aggregated using Logstash, stored in Elasticsearch, then cleaned and analyzed using Python.

### 3.3 Pilot test

To make the sentiment analysis result better, we filter out tweets that aren't opinion towards online learning by using sample tweets (these are real tweets that we got from our data pipeline) and then feed those sample tweets to the model which Naïve Bayes algorithm to classify opinion and non-opinion tweets. We select 72 sample tweets (each tweets coming from different users) and manually labeled it as opinion or non-opinion. This filtering method has been used before in similar research by Cham-lertwat et al. [32], in which the researcher manually annotated 600 randomly selected tweets, half of which contain opinion and non-opinion tweets. We use the same method but with a fewer number of sample tweets because the number of collected tweets differs from the research as mentioned above in where they collected about 1.000.000 tweets, in contrast with us only 133.477 tweets.

**Table 2.** Examples of manually labelled tweets

<b>Opinion</b>	<b>Aku pake zenius, ruangguru, rg, quiper translation: I use zenius, ruangguru, rg, quiper</b>
	Zenius jelasin dari dasar, konsep - <i>translation: Zenius explains from the basics, concept</i>
	Aku belajar di Ruangguru <i>translation: - I studied using Ruangguru</i>
	Belajar pake bimbel online - <i>translation: study using online tutoring</i>
	Quipper pernah coba aku - <i>translation: I already tried Quipper</i>
<b>Non-Opinion</b>	Belajar dari buku doang aku - <i>translation: I study from books only</i>
	Lebih cocok aja belajar dari buku - <i>translation: It's more suitable to learn from books</i>
	Aku males mau belajar tp pasti ntar nyesel - <i>translation: I'm lazy to learn, but later I will regret it</i>
	Belajar memisahkan diri kita dengan pikiran dan perasaan - <i>translation: Learn to separate thoughts and feelings</i>
	udah ngasih tau sate klatak yang enak - <i>translation: Already told delicious satay klatak</i>

### 3.4 Keywords

We divided tweets into three category and filtered tweets using the following keywords:

**Table 3.** Keyword Category Division

<b>Category</b>	<b>Keywords*</b>
Online Learning Application	Belajar zenius, pake zenius, belajar ruangguru, belajar RG, pake ruangguru, pake RG, belajar quipper, belajar quiper, pake quipper, pake quiper, belajar youtube, belajar coursera
Online Learning	Kelas online, belajar online
Learning	Mau belajar, belajar buat, belajar pake, kamu belajar, belajar apa, belajar dari

\*: Keywords are in Bahasa Indonesia

We selected these keywords based on pilot test results, which we previously extracted data from sample education-related tweets and count the frequency of each word; educational-related words with high frequency then chosen as keywords. From the pilot test, we found out that one word (unigrams) keywords capture many tweets compared to two words (bigrams) or three words (trigrams), but many of those tweets are not directly related to online learning. Using trigrams, we can capture more specific tweets related to online learning, but the number of tweets that get captured is too small. It is also proven that using bigrams yield a better result of the classification algorithm for subjective texts [33].

### 3.5 Tools used

Twitter Streaming API was used for getting public tweets data from Twitter. Elasticsearch was used as a database for storing tweets' object data. It functioned as a database where we can see, add, or remove an index. Logstash was used as a data aggregator from Twitter Streaming API. With Logstash, we can ingest data from a multitude of sources simultaneously, transforms it, and then sends it to the Elasticsearch index. Kibana is a visualization tool for Elasticsearch data. With Kibana, we can create bar, line and scatter plots, or pie charts and maps on top of large volumes of data. It is also useful to search for specific data using specific keywords. Python is a general-purpose programming language. In this study, we use Python to clean the data and build the Naïve Bayes classification model.

## 4 Findings

### 4.1 Dataset

For this study, we collected 133.477 tweets related to online education with 3.377 unique users. All tweets were collected from March - June 2019 using Twitter Streaming API with keywords mentioned in the second section. Table 4 shows the samples of these tweets.

**Table 4.** Tweets sample

Favorite_count	Hashtag	Is_a_retweet	Created_at	Text
0		FALSE	2019-06-02T18:21:21.000Z	@educationfess aku belajar tpa-nyq dr buku the king + latsol di zenius <i>translation: I study TPA from The King book + test bank from Zenius</i>
0		FALSE	2019-06-03T04:43:33.000Z	Di coba aja satu satu kalo kalian penasaran, tapi saran saya pake yang eduka, zenius atau pantau nilai UTBK - <i>translation: Try one by one if you curious, but I suggest use eduka, zenius, or monitor UTBK score - https://t.co/5kSsB1A2J9</i>
0		FALSE	2019-06-03T16:32:24.000Z	@rinensefalon @educationfess Temenku yg pake ruangguru sih orgnya emg tipe yg suka ada animasi dan ringkas gitu <i>translation: My friend who use ruangguru is the kind of person who likes animation and concise</i>

### 4.2 Processing

We cleaned the tweet content first by separating URL to a website and mention, then removing unnecessary spacing and hashtags to make the data available for sentiment analysis. Although the tweets were cleaner than before, we cannot ignore that not all of them are an opinion towards online learning. So to make the results more accurate, we solve this by separating between tweets that contained opinion towards

online learning or else by using the Naïve Bayes algorithm. The algorithm trained to filter out opinions towards online learning and the opposite using training results from sample educational tweets. The results show that there are only 1.7% tweets marked as opinion towards online learning, which means only 2.244 tweets available for sentiment analysis. The sentiment of each tweet will be further analyzed using different Naïve Bayes algorithm. We train the classifier to classify opinion towards online learning or else using the Naïve Bayes algorithm. Table 5 shows examples of tweets classified by our algorithms as opinion and non-opinion towards online learning.

**Table 5.** Opinion and non-opinion tweets

Opinion	Positive	Educationfess pake rg tp suka pinjem zenius temen. kalo aku personally suka pake zenius cm buat matematika <i>Translation: I used ruangguru but I like to borrow my friend's Zenius subscription. I personally liked zenius for my mathematics subject.</i>
	Negative	AT_USER Aku udah nyoba kelas online tapi ga aku lanjut wkwk semester depan mungkin baru aku lanjut lagi:") - <i>translation: I tried online classes but I stopped, maybe next semester I'll try again?</i>
Non-opinion		Selalu tidak pernah belajar dari kesalahan sebelumnya. Susah - <i>translation: never learned from my past mistakes. Difficult.</i>

Top 5 features or tokenization from our training set that we use to classify tweets into positive and negative are shown in Table 6.

**Table 6.** Top 5 positive and negative features (translation)

Positive Features	Negative Features
Zenius (see 2.5)	Susah (Difficult)
Belajar (Learn)	Enak (Nice)
Online	Gak (Not)
Enak (Nice)	Suka (Like)
Gampang (Easy)	Bimbel (Tutoring)

### 4.3 Results

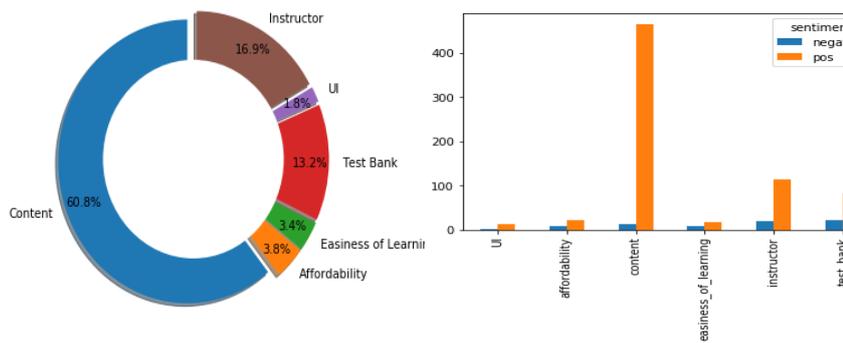
The result shows that 76.6% of the tweets are marked as positive; the rest are marked as negative. This proves that students in Indonesia—who are actively using Twitter—have good perception and experience using online learning applications. Furthermore, 3.7% of students expressed that they used online learning apps for studying tests, with 24.4% of it recommend using Zenius.

Even if 76.6% of the tweets in this study show a positive attitude towards online learning, but there are the not-so-small amount of students that have a negative perception towards online learning. It will be useful for online learning tools provider to knows not only about what students like from online learning, but also what kind of features that need improvement, missing, or annoying from online learning. We can assess that by looking more into the data by exploring negative tweets, and verify whether it is a real problem or not.

In order to make a more detailed analysis, we tried to classify what are people refer to in their tweets and their sentiment towards the topic using six categories: content, instructor, test bank, affordability, easiness of learning, and UI (User Interface). With keywords used for each category:

**Table 7.** Keywords category

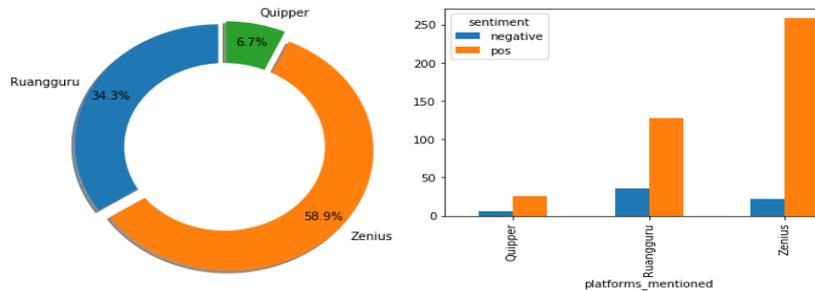
Category	Keywords
Content	materi, bahan belajar, animasi, video, rangkuman <i>translation: material, study material, animation, video, summary</i>
Instructor	guru, ngajar, tutor <i>translation: teacher, teaching, tutor</i>
Test Bank	latihan soal, soal, quiz <i>translation: exercises, question, quiz</i>
Affordability	harga, harganya, murah, mahal <i>translation: price, pricing, cheap, expensive</i>
Easiness of Learning	mudah, gampang, gak bimbel <i>translation: easy, convenient, no tutoring</i>
UI (User Interface)	desain, design, UI <i>translation: design, design, User Interface</i>



**Fig. 2.** What are people refer to in their tweets (left) and the comparisons of sentiments from six categories of features (right)

As shown in Figure 2, people prefer to talk about content, instructor, and test bank of online learning providers. With total tweets about content, instructor, test bank, affordability, easiness of learning, and UI are 477, 133, 104, 30, 27, and 14, respectively. The other 1,484 tweets can't be classified due to not referring to any of the categories.

Content is the number one feature from online learning with the most positive sentiment, followed by an instructor, and test bank. The massive amount of positive sentiment towards content is evident if we observe at how education delivered online: Teaching use easy-to-understand study material, a summary of each topic, and video which contains animation that makes learning topics of interest more comfortable and fun.



**Fig. 3.** How many times online learning providers are mentioned (left) the different sentiments ratio of each brand (right)

During the study period, Zenius is the brand that people mentioned the most, followed by Ruangguru and Quipper, respectively. Zenius, Ruangguru, and Quipper get mentioned in 280, 163, and 32 tweets, respectively. With 1.769 tweets doesn't mention any of those names.

Zenius is the most talked and most impressive one since there are eight to nine times more positive sentiment more than negative sentiment towards them, while Ruangguru has the most negative sentiment compared to other online learning providers.

#### 4.4 Discussion



**Fig. 4.** Bigrams (2-words) Word Cloud for Tweets with Positive Sentiments (left) and Tweets with Negative Sentiment (right)

Positive tweets usually have these keywords (translated from Indonesia to English): “Me studying,” “studying in,” “what is,” “inside a class,” “I remember,” and “use Zenius.” When other online learning applications that operated in Indonesia, such as Ruangguru and Quipper does not get mentioned much in positive tweets, Zenius is.

These keywords also can be used as targeted advertising on Twitter. Below are sample positive tweets:

Sample positive tweets:

“I like Zenius because if we study using concepts, so it’s easier to remember”

“The learning material is more complete in Ruangguru, but the explanation of the materials is better in Zenius, for Quipper I don’t know”

“Zenius teaching is concept-based. If you’re the type of person who likes visual or animated learning materials, I think you’ll find Zenius boring because they only use the text”

“I use Quipper huhu (haha), free to study at any time, I can also lie down on the bed as well wkwkwk (haha)”

Students might like one of the online learning providers over others because of certain aspects such as better explained learning materials, animation on tutor video, complete learning materials, and the existence of their application to access learning materials in mobile phones. On top of that, online learning pricing is cheaper than offline tutoring.

Other positive tweets mentioned they use online learning applications together with offline tutoring to study. It shows that Indonesia’s students feel offline tutoring or online learning only is not enough because both complement each other.

From the negative sentiments Wordcloud, we can see keywords (translated from Indonesia to English) such as online tutoring, youtube channel, and study UTBK (higher education admission test). These keywords can be used for future research to capture negative sentiment towards online learning.

Once again, it is interesting to see what people negatively tweets about online learning. Below are a sample of negative tweets:

Sample negative tweets:

“Honestly, I prefer to study using actual books rather than e-book...”

“I think it’s fine if you want to buy a subscription for online learning apps. However, for the national admission test, I think it’s better you go for offline tutoring...”

From the national choices of sample negative tweets—translated to English, we can know that some students prefer physical books and prefer offline tutoring for the national college admission test.

Compared to other previous research utilizing sentiment analysis, our study opens a new way to capture Twitter data and doing sentiment analysis using ELK stack & Python for online learning topics in developing countries (in this case, Indonesia). On this study we discover two interesting things:

1. Zenius is the most talked and preferred choice of online learning provider on Twitter by Indonesian students, rather than Ruangguru and Quipper.
2. Although lagging behind Zenius in the number of positive tweets, Ruangguru is liked and preferred because their tutor video contains animation, provide a summary for every topic, and provide pretty complete test bank. While the most lagging, Quipper is not really known by Indonesian students.

The findings of this study correlate well with real-world facts. One platform is known for its easy-to-understand tutor video, while other platforms have its attractiveness. Students choose online learning providers based on their preference: Some like to learn fundamental things first and disregard the visual aspect of learning mate-

rials, some favor animation on tutor video to make learning not dull, while others prefer accessible learning materials on their phone.

#### **4.5 Managerial implication**

In this study, we used the ELK stack to collect data from Twitter and the Naïve Bayes algorithm to understand sentiment towards online learning. Results show that 76.6% of the tweets have a positive sentiment towards online learning. It indicated that students in Indonesia—who is actively using Twitter—have positive perceptions towards the usefulness of online learning and even positive experience for those who are already using it. When these students asked to recommend online learning platform to other students on Twitter, 24.4% strongly recommend using Zenius. Because based on their experiences, Zenius makes learning fun and hard-to-learn subject become more comfortable to learn. Understanding what makes Zenius unique than others might be the key to unlock online learning growth potential in developing countries like Indonesia.

### **5 Conclusion**

In this paper, we tried to understand developing countries students' sentiment towards online learning in Twitter using the ELK stack and Naïve Bayes algorithm. The architecture and method used in this paper are generic and can be easily adapted and extended to other domains (for instance, we can use the same method to analyze sentiments towards a brand or political candidate). Our study shows that Indonesian student has a positive attitude towards online learning, and even have a significant willingness to recommend an online learning application when they found it useful to learn subjects taught at school. It is suggested for online tutoring providers to produce educational content with a more straightforward and easier method so that the student will understand the materials faster. Lastly, we established that Twitter is an excellent place to discover opinions towards a brand, product, or person via sentiment analysis.

### **6 Acknowledgement**

For future researchers whom are interested in the data this research used, we make available the present research dataset in the following link: <http://bit.ly/publPercDataset>. Feel free to use the data and contact [oktavianto.andri@gmail.com](mailto:oktavianto.andri@gmail.com) for further questions.

### **7 References**

- [1] B. Trilling and C. Fadel, 21st century skills: Learning for life in our times. John Wiley & Sons, 2009.

- [2] M. J. Hannafin and S. M. Land, "The foundations and assumptions of technology-enhanced student-centered learning environments," *Instructional science*, vol. 25, no. 3, pp. 167-202, 1997.
- [3] S. R. Hiltz and M. Turoff, "The evolution of online learning and the revolution in higher education," *Communications of the ACM*, vol. 48, no. 10, pp. 59-64, 2005. <https://doi.org/10.1145/1089107.1089139>
- [4] M. Ziegler, T. Paulus, and M. Woodside, "Creating a climate of engagement in a blended learning environment," *Journal of Interactive Learning Research*, vol. 17, no. 3, pp. 295-318, 2006.
- [5] I. E. Allen and J. Seaman, *Going the distance: Online education in the United States*, 2011. ERIC, 2011.
- [6] S. Carliner, *An overview of online learning*. Human Resource Development, 2004.
- [7] S. Carr, "As distance education comes of age, the challenge is keeping the students," *Chronicle of higher education*, vol. 46, no. 23, 2000.
- [8] D. O. Bruff, D. H. Fisher, K. E. McEwen, and B. E. Smith, "Wrapping a MOOC: Student perceptions of an experiment in blended learning," *Journal of Online Learning and Teaching*, vol. 9, no. 2, p. 187, 2013.
- [9] A. G. Picciano, "Beyond student perceptions: Issues of interaction, presence, and performance in an online course," *Journal of Asynchronous learning networks*, vol. 6, no. 1, pp. 21-40, 2002.
- [10] M. Gillingham and C. Molinari, "Online courses: Student preferences survey," *Internet learning*, vol. 1, no. 1, pp. 36-45, 2012. <https://doi.org/10.18278/il.1.1.4>
- [11] E. R. Burner, M. D. Menchine, K. Kubicek, M. Robles, and S. Arora, "Perceptions of successful cues to action and opportunities to augment behavioral triggers in diabetes self-management: qualitative analysis of a mobile intervention for low-income Latinos with diabetes," *Journal of medical Internet research*, vol. 16, no. 1, p. e25, 2014. <https://doi.org/10.2196/jmir.2881>
- [12] Statista, "Twitter: Number of users in Indonesia," Available: <https://www.statista.com/statistics/490548/twitter-users-indonesia/>
- [13] Statista, "Leading countries based on number of Twitter users as of January 2019 (in millions)," Statista2019, Available: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.
- [14] S. Rosenthal, P. Nakov, S. Kiritchenko, S. Mohammad, A. Ritter, and V. Stoyanov, "Semeval-2015 task 10: Sentiment analysis in twitter," in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 2015, pp. 451-463. <https://doi.org/10.18653/v1/s15-2078>
- [15] V. Kharde and P. Sonawane, "Sentiment analysis of twitter data: a survey of techniques," arXiv preprint arXiv:1601.06971, 2016.
- [16] N. A. Vidya, M. I. Fanany, and I. Budi, "Twitter sentiment to analyze net brand reputation of mobile phone providers," *Procedia Computer Science*, vol. 72, pp. 519-526, 2015. <https://doi.org/10.1016/j.procs.2015.12.159>
- [17] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," in *Proceedings of the ACL 2012 System Demonstrations*, 2012, pp. 115-120: Association for Computational Linguistics.
- [18] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of twitter data for predicting stock market movements," in *2016 international conference on signal processing, communication, power and embedded system (SCOPEs)*, 2016, pp. 1345-1350: IEEE. <https://doi.org/10.1109/scopes.2016.7955659>

- [19] I. E. Allen and J. Seaman, Online nation: Five years of growth in online learning. ERIC, 2007.
- [20] L. Triacca, D. Bolchini, L. Botturi, and A. Inversini, "Mile: Systematic usability evaluation for e-Learning web applications," in EdMedia+ Innovate Learning, 2004, pp. 4398-4405: Association for the Advancement of Computing in Education (AACE). <https://doi.org/10.1109/wse.2005.6>
- [21] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," HP Laboratories, Technical Report HPL-2011, vol. 89, 2011.
- [22] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment analysis of review datasets using Naïve bayes and k-nn classifier," arXiv preprint arXiv:1610.09982, 2016. <https://doi.org/10.5815/ijieeb.2016.04.07>
- [23] N. Fantom and U. Serajuddin, The World Bank's classification of countries by income. The World Bank, 2016.
- [24] W. Bank, "World Bank Country and Lending Groups," Accessed on: August 9, 2019 Available: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
- [25] W. Bank, "Lower middle income," Accessed on: August 9, 2019 Available: <https://data.worldbank.org/income-level/lower-middle-income>
- [26] Twitter, "Standard Stream Parameter - Twitter Developers," Available: <https://developer.twitter.com/en/docs/tweets/filter-realtime/guides/basic-stream-parameters>.
- [27] Twitter, "Tweet Object," Available: <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object.html>
- [28] L. Kompas, "Survei Litbang Kompas Tentang Bimbingan Belajar," Kompas2012, Available: <https://edukasi.kompas.com/read/2012/12/07/17494989/Ramai-ramai.Les.Pelajaran>.
- [29] P. Beineke, T. Hastie, and S. Vaithyanathan, "The sentimental factor: Improving review classification via human-provided information," in Proceedings of the 42nd annual meeting on association for computational linguistics, 2004, p. 263: Association for Computational Linguistics. <https://doi.org/10.3115/1218955.1218989>
- [30] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in 2014 Seventh International Conference on Contemporary Computing (IC3), 2014, pp. 437-442: IEEE. <https://doi.org/10.1109/ic3.2014.6897213>
- [31] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in LREc, 2010, vol. 10, no. 2010, pp. 1320-1326.
- [32] W. Chamlerwat, P. Bhattarakosol, T. Rungkasiri, and C. Haruechaiyasak, "Discovering Consumer Insight from Twitter via Sentiment Analysis," J. UCS, vol. 18, no. 8, pp. 973-992, 2012.
- [33] S. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification," in Proceedings of the 50th annual meeting of the association for computational linguistics: Short papers-volume 2, 2012, pp. 90-94: Association for Computational Linguistics.

## 8 Authors

**Satria Fadil Persada**, Ph.D (Satria) research interest is studying the behavior of person's in the perspective of consumers and organization. Satria' research interest is also related to pro-environmental behavior.

**Andri Oktavianto** (Andri) is a student in the Business Management Department of Institut Teknologi Sepuluh Nopember, Andri's main research interest is IT-related topics, machine learning, and data mining.

**Bobby Ardiansyah Miraja** (Bobby) is currently studying business management for a bachelor degree in the Department of Business Management, Institut Teknologi Sepuluh Nopember (ITS). Bobby's main research interest is in the technology and human resources area.

**Reny Nadlifatin**, Ph.D (Reny) studies focused on technology acceptance models, specifically on behavior towards information technology.

**Prawira Fajarindra Belgiawan**, Ph.D (Fajar) research interest is in studying whether people's decision can actually be influenced by their significant others. Fajar's other research interest is in digital disruption.

**A. A. N. Perwira Redi**, Ph.D (Redi) research interest is in the field of industrial management.

Article submitted 2019-08-26. Resubmitted 2020-01-30. Final acceptance 2020-01-30. Final version published as submitted by the authors.