

# Unraveling the Disruptive Innovation in Indonesia's Tutoring Field Using Text Mining

Wachda Yuniar Rochmah<sup>1</sup>, Gadang Ramantoko<sup>2</sup>, and Astri Ghina<sup>3</sup>

{ wachdayr@telkomuniversity.ac.id<sup>1</sup>,

gadangramantoko@student.telkomuniversity.ac.id<sup>2</sup>, aghina@telkomuniversity.ac.id<sup>3</sup>}

School of Economic and Business Telkom University, Bandung, Indonesia<sup>123</sup>

**Abstract.** The emergence of online tutoring innovation which provides a learning service changes the learning model from being conducted face-to-face to an online environment. Moreover, the offline tutoring has given a lot of pain to their customers. In order to understand the customer, some time-consuming methods have been performed. However, these methods are less efficient compared to data analytics using social media. Hence, we conduct a research using text mining to process 6618 tweets. As a case study, we use Ruangguru that has proven its existence by providing 15 million users in 2019 and gain international funding to operate its business. Our objective is to unravel the pain felt by customers of offline tutoring and observe Ruangguru's disruption process in Indonesia which has emerged in the midst of pain while using offline tutoring. As the result, disruption has occurred in the tutoring market by Ruangguru since 2014, where they provide a value that consists of private leaning, filling the pain in the obstacle and reluctant category of pain. As years go by, Ruangguru keep being consistent, even getting better, as shown that more various values continue to answer the pain of offline tutoring. In order to develop the sustainable innovation, pain identification is suggested to every tutoring agency. They should consider every solution to overcome the pain felt by its customers, so that churn rate can be prevented.

**Keywords:** Disruptive Innovation, Ruangguru, Text Mining, Tutoring.

## 1 Introduction

In a process of using a product or service by customer, there will be a distraction either before, in between, or after they finish trying the product; which is known as *pain*. There many categorizations of customer's *pain*, such as unexpected output, obstacle, and risk [1]. Along with the rise of the social media usage, consumption-related behaviors are able to be seen in social media platforms. This kind of behavior can be taken place before, during, or after the consumption [2].

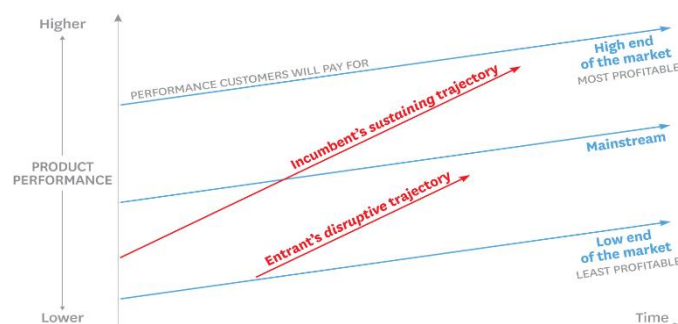
Following the upgrade of technology, an impactful innovation in educational field especially tutoring has emerged, bringing the new concept of learning. It's widely known that online tutoring service provides a new change, from a face-to-face learning class to a class that can be accessed everywhere [3]. This online tutoring business in Indonesia nowadays is dominated by Ruangguru which has passed 15 million users by the end of 2019[4].

There is a concept called *disruption by Clayton Christensen* where a smaller business with less resource is able to successfully challenged the running business in the market. When a common consumer starts to adopt the offer of the newcomers in the high volume, then it's where a *disruption* has existed. This concept is known as *The Disruptive Innovation Model* [5].

Thus, in this research we aim to unravel the pain felt by customers of offline tutoring and observe Ruangguru's disruption process in Indonesia which has emerged in the midst of pain while using offline tutoring. We apply several text mining methodologies which consists of Sentiment Analysis, Topic Modeling, and Multiclass Text Classification.

## 2 Literature Review

An innovation consists of a new idea in a generation and its implementation in a new product, process, or service which leads to the dynamic growth of national economy and job opportunity, along with the profit creation for innovative running businesses. Innovation doesn't only occur in a single time, but in a long cumulative process which consists of many decisions making process in an organization, starting from idea creation to implementation phase [6]. There are 4 types of innovation. They are sustaining innovation, breakthrough innovation, disruptive innovation and basic research [7].



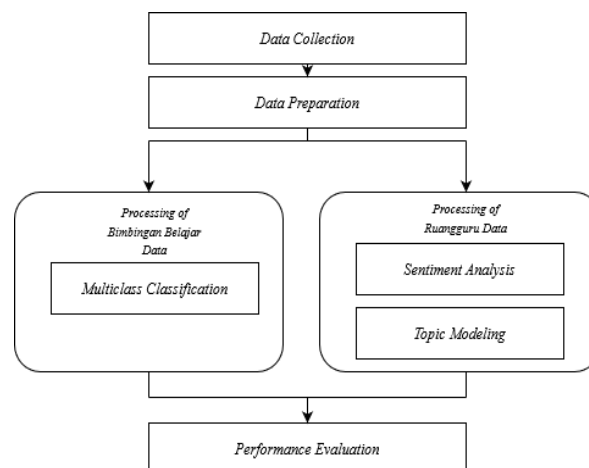
**Fig. 1.** The Disruptive Innovation Model

*Source:* Christensen (2015)

There is a model of disruptive innovation which was introduced as The Disruptive Innovation Model which is illustrated in Figure 1. The graph compares the trajectory of product performance, shown with red lines, with the trajectory of customer demand, shown with blue line. Existing companies introduce high-quality products or services to satisfy the high-end market where the profitability is very high. They exceed the needs of low-end customers and other mainstream customers. Then, this will provide an entry point for new entrants to find a footing in the segment that provides less profit which is ignored by existing companies. Entrants or those who have just entered the disruptive path, shown by the red line below, continue to improve the performance of their offers and move to high-end markets where the highest profitability exists and challenge the dominance of existing companies [5].

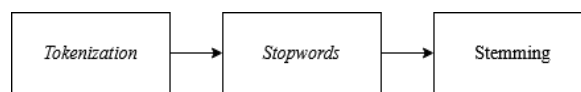
### 3 Methodology

We design the research workflow shown in Figure 2 The first step of the research is data collection. The data used is gathered by Python language by scrapping technique. We collect the data using hashtag bimbel, which means tutoring, and Ruangguru on Twitter from January 1<sup>st</sup> 2014 to December 31<sup>st</sup> 2019. We separate the dataset into two different files for different methods used in the research.



**Fig. 2.** The Research Workflow

After the data collection has finished, we continue to the preparation stage which consists of several stages: data selection, data cleaning and text preprocessing. The data we gather contains many attributes, but in the first stage we select only textual text. After selecting the data, we clean the whole data from unnecessary contents such as spam. Next, we apply text preprocessing which consists of tokenization, stopwords, and stemming.



**Fig. 3.** Text Preprocessing

Tokenization is a step to splitting a text into smaller elements such as terms that contains meaning [8]. The example of the tokenization is shown below.

Sudah bimbel tapi belum keterima juga

Next is stopwords. Stopwords describe the words that frequently occur in the textual document and have the low meaning or value [8]. Therefore, these words will be removed from the document. This is the example of stopwords.

Sudah bimbel    belum keterima

The last step of preprocessing is stemming, which means normalize the text or word into its standard form [8]. The example of a sentence after we apply the stemming step is shown below.

Sudah bimbel    belum    terima

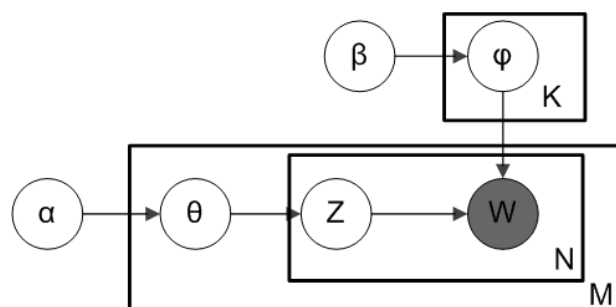
Next, we continue to the third stage of the research which is data processing. In the data processing stage, we use text mining which consist of three methods: Sentiment Analysis, Multiclass Classification, and Topic Modeling. Earlier, we separate the datasets into two different purposes.

The bimbel dataset is used to discover the pain felt by tutoring's customers using Multiclass Classification. Multiclass Text Classification is an activity to classify text into more than 2 classes. Every sample is only classified into a class. There are various algorithms which can be used, i.e. Naïve Bayes, k-Nearest Neighbor, Neural Network, Decision Tree, and Super Vector Machine [9]. In the

pair of offline tutoring dataset, we categorize the training-test using Naïve Bayes classifier into 7 labels which consist of costly, distraction, obstacle, pragmatism, reluctant, tiring, and useless.

The second dataset which contains Ruangguru keyword is used to see the advantages of Ruangguru and whether it becomes a disruptive innovation of tutoring market in Indonesia. We use Sentiment Analysis and Topic Modeling for the process. Sentiment analysis or opinion mining is a field of knowledge that analyzes opinion, sentiment, evaluation, assessment, attitude, and emotion from people towards an entity such as product, service, organization, individual, problem, event, topic, and their attributes. Besides that, sentiment analysis is also known as opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. [10].

Topic modeling can be described as an approach in discovering topics in documents. One of the algorithms is Latent Dirichlet Allocation (LDA), which is described as a learning method among words, topic, and document by assuming a certain probability to model a document [11].



**Fig. 4.** An illustration of Latent Dirichlet Allocation (LDA)  
*Source:* Feldman (2007)

Figure 1 illustrates how Latent Dirichlet Allocation works. The black circle shows the variable which is being analyzed. Latent is the white circle, and the rectangle plays many roles of variables. The variable of this model is explained in the Table 1.

**Table 1. Parameter of Latent Dirichlet Allocation (LDA) Model**

Parameter	Description
K	The number of topics
N	The number of terms in document
M	The numbers of document being analyzed

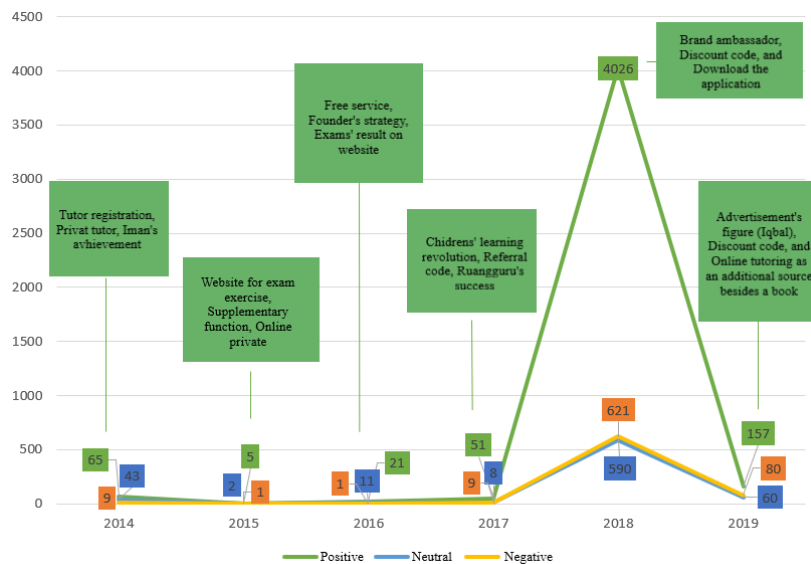
A	Parameter Dirichlet in every topic of document
B	The same parameter from the term distribution in every topic
$\phi(k)$	The distribution of term for topic k
$\theta(i)$	The distribution of topic in document i
$\omega(i,j)$	Term j in document i
$z(i,j)$	The determined topic for (i,j)
$\phi$ and $\theta$	Dirichlet distribution

The results of the topics will be used as the value of Ruangguru. After having the pain classification and topics revealing, we observe the result by comparing the movement of pain every year and the topics of Ruangguru. Here, we want to check whether every value of Ruangguru actually fills the pain of offline tutoring. It is based on The Disruptive Innovation Model, where new entrance starts to grow its business and gain the market that is ignored by offline tutoring agencies. If the value of Ruangguru keeps developing and the pain also appears every year, which means it's not solved, then we are able to conclude that the disruption occurs.

## 4 Results

### 4.1 The Value of Ruangguru

In order to acknowledge the sentiment of Ruangguru in Twitter during six years, we apply Sentiment Analysis to classify into three classes which consists of positive, negative, and neutral. The result contains 4325 positive sentiments 721 negative sentiments, and 714 neutral sentiments. It means that Ruangguru gains more positive sentiments than negative and neutral ones. Then, we provide the dynamic visualization of the sentiment each year along with the value of Ruangguru. The value of Ruangguru can be discovered using Topic Modeling, based on the positive sentiment which has been classified using Sentiment Analysis. This aims to see the advantages of Ruangguru. Hence, we do the measurement during six years, then summarize the topics to be visualized in a dynamic graph to see the changes made every year. The result of sentiment analysis and the value of Ruangguru is shown dynamically in the Figure 5.



**Fig. 5.** The Dynamic Results of Sentiment and Topic of Ruangguru

According to the result, positive sentiment of Ruangguru always has the most amount every year. In Figure 5, positive sentiment of Ruangguru decreases in 2015, then always increases from 2016 to 2018, which also becomes the peak of Ruangguru’s highest achievement so far. This illustrates the good response of citizen towards Ruangguru. On the other hand, the negative and neutral sentiments don’t show any significant changes.

In 2014, we discover many positive sentiments. Several topics being discussed are about the teacher registration for joining Ruangguru, the role of Ruangguru as private learning for student, and the achievement of Ruangguru’s founder, Iman Usman. In fact, Ruangguru is established in 2014, hence it still focuses on introducing Ruangguru in this year.

Next year in 2015, there are several topics being discussed in the positive sentiment of Ruangguru, such as the function of web of Ruangguru as a service provider to exercise, Ruangguru’s supplementary function of current educational institution, and private online program owned by Ruangguru.

In 2016, people talk about free service, the strategic of founder, and the results of tryout on the website. This year, Ruangguru has started to improve their performance by giving more various service.

Moving forward to 2017, the number of positive sentiments keep increasing where people talk about student’s learning revolution, referral code,

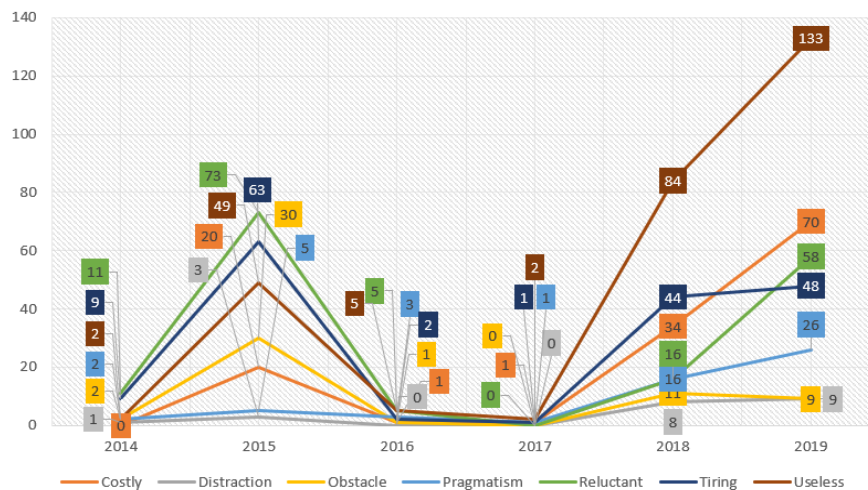
and the success of Ruangguru. This year Ruangguru starts to expand their marketing strategy by giving referral code and discounts.

Next, in 2018 we are able to see a significant change of sentiment has happened here. Based on the topics of positive sentiment, we are able to see that people give good opinions towards the brand ambassador of Ruangguru, the discount code and the application. This year, Ruangguru has once again expanded their marketing by improving engagement with their customers by having collaboration with some youngsters' influencers as their brand ambassadors.

The last year, the value of each sentiment has decreased from the previous year. However, the positive sentiment keeps on leading higher than other sentiments. One of their brand ambassadors, Iqbal, receives positive sentiments. Besides that, people still have positive sentiment towards discount code and the function of Ruangguru as a supplementary of textbook in the formal education.

#### 4.2 Pain of Using Offline Tutoring Service

The pain of using offline tutoring field is illustrated in the Figure 6, which consists of 858 tweets and shows its movement from 2014 to 2019. According to the visualization, we are able to see that each category of pain felt by customers of offline tutoring happens to change every year.



**Fig. 6.** The Dynamic Results of Pain



The first category is costly, illustrated in orange color, which means the high price that customer has to pay for using the service of tutoring. In 2014, there is no one who suffers from its costly fee. In the next year, the number starts raising, even though it decreases again in 2016 and 2017. In 2018, it increases significantly and reached its peak in 2019 with 70 tweets. The second category is distraction. It describes something that distracts customer while using tutorial service, such as the class environment, the tutor's attitude, and many more. There are not many people who express their pain in this category every year. The third category is obstacle, which means the obstacle that may prevent students from going to the class, the example is the weather, transportation, etc. The number of tweets categorized into this category is not much, same as the distraction. The fourth category is pragmatism, which may show customer's idealism. Every year, the number of tweets in this category doesn't show much significant change. The fifth category is reluctant. It illustrates how students mostly be reluctant to go to the offline tutoring. We are able to see that many people express their laziness to go by seeing the number of tweets in this reluctant category each year, especially in 2015 where the number hits 73 tweets. The sixth category is tiring, which student express when they are getting tired and incapable of catching up with school and tutoring schedules every day. This category is shown to have high score in 2018 and 2019. The last category is useless. The way it leads the category shows that many people think that joining tutorial is actually useless, where it starts to show a very significant raising from 2017 to 2019.

### 4.3 How Ruangguru Solves The Pain of Offline Tutoring

According to the result that is shown in Figure 4 and Figure 5, we conduct an analysis to observe whether the pain of offline tutoring field is solved by the value given by Ruangguru. The result is shown in Table 2.

**Table 2. Pain of Offline Tutoring Which is Answered by The Value of Ruangguru**

Year	Value of Ruangguru	Pain						
		Costly	Distraction	Obstacle	Pragmatism	Reluctant	Tiring	Useless
2014	Tutor registration	×	×	×	×	×	×	×
	Private tutor	×	✓	✓	×	✓	×	×

	Founder's achievement	×	×	×	×	×	×	×
2015	Web for exam exercise	×	×	✓	×	✓	✓	×
	Supplementary function	×	×	×	×	×	×	×
	Online private	×	✓	✓	×	✓	✓	×
2016	Free service	✓	×	×	×	×	×	×
	Founder's strategy	×	×	×	×	×	×	×
	Exams' result on website	×	×	×	×	×	×	×
2017	Children's learning revolution	×	×	×	×	×	×	×
	Referral Code	✓	×	×	×	×	×	×
	Ruangguru's success	×	×	×	×	×	×	×
2018	Brand ambassador	×	×	×	×	✓	×	×
	Discount code	✓	×	×	×	×	×	×
	Download the application	×	×	×	×	×	×	×
2019	Advertisement's figure (Iqbal	×	×	×	×	×	×	×
	Discount Code	✓	×	×	×	×	×	×
	Online tutoring as an additional source besides a book	×	×	×	×	×	×	×

According to Table 2, we are able to see the value of Ruangguru every year. It gives solution to every pain in offline tutoring agencies. In 2014, private tutoring that Ruangguru starts to give apparently become the solution for the obstacle and reluctant categories of pain. In 2015, its website as value of Ruangguru gives solution for obstacle, reluctant, reluctant, and tiring categories. In 2016, value of Ruangguru which Is its free service gives solution for costly category of pain. In 2017, Ruangguru's value which becomes the solution of offline tutoring is the referral code. Referral code gives solution for

costly category. In the following year, Ruangguru's value which is their brand ambassadors give solution for reluctant category of pain. The right brand ambassadors are able to give encouragement to students to study, stated by the user of Ruangguru. In 2019, another Ruangguru's value that gives solution for the pain of offline tutoring is discount code.

We find out that people's opinion on twitter match with real world's fact in the term of price, where the price of tutoring keeps raising each year, proven with Ganesha Operation's price, as one of the big agencies, every year. In addition, we are also able to discover that Ruangguru's advantage answers tutoring's pain that hasn't been solved all these years in the costly category by providing some free services and referral code for discount, illustrated by the topics of Ruangguru on Twitter.

#### **4.4 The Disruption Process**

According to Christensen's theory that has been illustrated in Figure 1, when the running business get too focused on chasing profit created by high market and ignore another segment of customer which gives less profit, and keeps on developing their service and finally take over all markets in the industry. This occurs same as Ruangguru, where it starts to gain trust from customer of offline tutoring who has been suffering from the high price of tutoring fee.

### **5 Conclusion**

From 5760 *tweets* about Ruangguru on Twitter, 4325 *tweets* are included as positive sentiments, 721 *tweets* are negative sentiments, and 714 *tweets* are neutral sentiments. It indicates that there are more positive sentiments than negative or neutral ones for Ruangguru in 6-year period. Every value owned by Ruangguru also shows improvement every year and gives solution for the pain felt by traditional tutoring's customers since 2014. The process that occurs between Ruangguru and Traditional tutoring field fits with The Disruptive Innovation Model by Christensen, where Ruangguru as a new entrance in 2014 when the pain keeps increasing. Then this new entrance develops their product and service. This can be illustrated by the number of values that gives solution for offline tutoring in more various categories since 2015 to 2019. Therefore, we conclude that Ruangguru happens to be a disruptive innovation in

educational sector especially in tutoring field. The results of this research are able to be used as a preliminary knowledge to develop innovation according to the pain that has been felt by customers in the social media sphere. By using this knowledge, tutoring agencies are able to understand their customers more and develop customer-based strategic innovation.

## 6 Limitation & Future Research Suggestion

This research provides preliminary knowledge about The Disruptive Innovation in educational field, especially in the tutoring market in Indonesia, using Text Mining. We suggest further research to conduct an expert validation and a validation to offline tutoring agencies about the priority of the pain.

## References

- [1] Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value Proposition Design*. John Wiley & Sons.
- [2] Durukan, T., Bozaci, I., & Hamsioglu, A. B. (2012). An investigation of customer behaviors in social media. *European Journal of Economics, Finance and Administrative Sciences*, 44(44), 148-159.
- [3] Tirtoid. (2017). Retrieved February 16, 2020, from <https://tirto.id/bisnis-industri-pendidikan-yang-makin-diminati-cnRh>
- [4] Ruangguru. (2020). *Tentang kami: Ruangguru*. Retrieved February 16, 2020, from Ruangguru Web site: <https://ruangguru.com/general/about>
- [5] Christensen, C. M., Raynor, M., & McDonald, R. (2015). What Is Disruptive Innovation? *Harvard Business Review*, 93(12), pp. 44-53.
- [6] Rogers, E. M. (2010). *Diffusion of Innovations*. Simon & Schuster.
- [7] Harvard Business Review. (2017). Retrieved April 24, 2020, from Harvard Business Review: <https://hbr.org/2017/06/the-4-types-of-innovation-and-the-problems-they-solve>
- [8] Hofmann, M., & Chisholm, A. (Eds.). (2016). *Text mining and visualization: case studies using open-source tools* (Vol. 40). CRC Press.
- [9] Srivastava, A., & Sahami, M. (2009). *Text Mining: Classification, Clustering, and Applications*. Chapman and Hall/CRC.
- [10] Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1).

- [11] Pasquali, A. R. (2016). Automatic Coherence Evaluation Applied to Topic Models.