

Visualization of the Temporal Topic Model on Higher Education Preferences with Higher Education Ranking Indicators

Winda Widya Ariestya¹, Achmad Benny Mutiara², I Made Wiryana³, Setia Wirawan⁴
Gunadarma University, Faculty of Computer Science and Information Technology
Margonda Raya No. 100, Depok, Indonesia

Abstract—Private universities have devised a strategy to counteract the ongoing competition. Private universities can use the appropriate data analysis method to make higher education management decisions. The goal of this research is to find a new approach to data analysis methods in the form of visualization using the TTM (Temporal Topic Model) method to assist private university management. These findings are the two formulas used to generate time-based visualizations and the Temporal Topic Model per month to visually change news topics related to rankings so that management can decide on marketing strategies and policies that are in relation to public opinion.

Keywords—Management decisions; temporal topic model; university; visualization

I. INTRODUCTION

The number of private universities, in Indonesia is around 68 per cent of all universities, is one factor in the existence of competition for private universities in Indonesia. Private universities compete with each other to provide students as their consumers with the best educational services.

The challenge currently facing Indonesian universities is the implementation of the Outcome-Based Education method, where learning focuses not only on the teaching and learning process but also on output [1]. The Accreditation of National and International Higher Education requires a curriculum that is supported by an Outcome-Based Approach [2].

Furthermore, a proper marketing strategy is a requirement for all universities, one of which is to provide services of equal value to the expectations of students, particularly those of stakeholders [3].

News about tertiary education institutions may influence the community in determining the choice of tertiary institutions, as the information presented is one of the contents of the ranking news [4]. Higher education rankings are not widely known to the public, based on the results of research conducted by Gunarto [5] through a survey method with a descriptive analysis of people's perceptions and preferences of the reputation of higher education ratings.

A number of world university rankings, including Webometrics Rangkings Of World University (WRWU), SCIMAGO Institutions Rankings (SIR), Academic Ranking of World Universities (ARWU), Taiwan Higher Education Evaluation and Accreditation Council (HEEACT), THE-QS

World Ranking of Universities, and 4 International Colleges & Universities (4ICU) [6]. Indonesia conducts a National Cluster of Higher Education through the Ministry of Research, Technology and Higher Education, which is released every semester with the objective of mapping Indonesian universities to enhance the standard of higher education under the auspices of the Ministry of Research, Technology and Higher Education, as well as being the basis for the Ministry of Research, Technology and Higher Education [7].

Advances in information technology are currently supporting the management of data needed for higher education management, which is used to obtain user preferences. Information technology can assist management in evaluating information that can offer decision-making options for management. It is not possible to distinguish data analysis from the presentation of enticing data in order to promote the process of analysis. Visual presentation of the data allows management to better understand the summary of the information presented. Information on tertiary institutions visually at a certain time can be used by management to determine policies for making competition decisions that take place.

Several studies on the Temporal Topic Model and the Use of Visualization, including Jeong [8], conducted a time analysis between three sources and two academic fields by conducting a text mining content analysis using LDA techniques. The resulting topic modeling has been declared effective in determining the content and trends of the time series of papers, patents and news articles. Jatmika [9] carried out a data mart visualization design to monitor the performance of the STIKOM Surabaya study programs. Visualization is designed to assist the Head of the Study Program in the academic performance of the Study Program. Ghosh et al [10] introduced a model of time-related issues in news articles to see trends in time-related issues. The time series regression technique in modeling has been found to be able to produce trendy topics efficiently.

This research differs from several previous studies in that it analyzes the community's tendency in choosing universities towards university ranking indicators with a new approach using the Temporal Topic Model and visualizing it in order to apply for higher education management. The resulting visualization is time-based, and the Temporal Topic Model is used once a month to produce shifts in news topics that are

visually related to university rankings, allowing management to make marketing strategies and policies that are in line with community opinions.

II. METHODS

Research commences with the process of collecting data in accordance with the subject under study. The next stage is the processing of data using text mining techniques. There are two stages in data processing, namely pre-processing and modeling using the LDA method and the Temporal Topic Model. Visualization was carried out as a tool for analyzing emerging topics from the results of data processing. The stages of this research can be seen in Fig. 1.

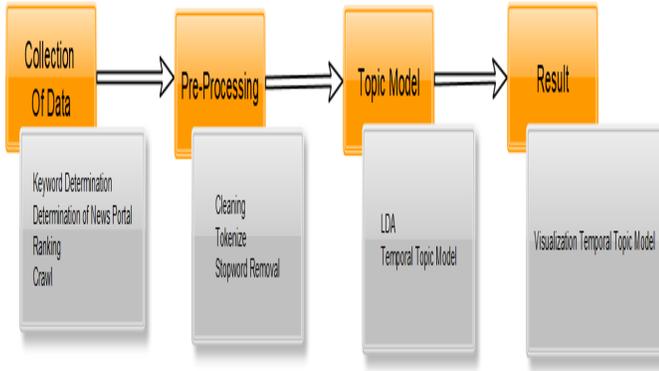


Fig. 1. Methods of Research.

A. Collection of Data

As a data source for this study, Indonesian news about universities was used. Text-only news content consists of text only. Before beginning the text mining process, keywords must be determined. Keywords are expressions that represent a concept, according to the KBBI Big Indonesian Dictionary [11]. The keywords used by Google Keyword Planner are listed in Table I [12].

Webometrics Ranking Of World University (WRWU), Times Higher Education Supplement (THES), Quacquarelli Symonds World University Rankings (QS WUR) are higher education rankings used on the basis of the IREG Ranking Audit [13] and Kemenristekdikti cluster ranking of universities in the world.

TABLE I. KEYWORDS FOR THE GOOGLE KEYWORD PLANNER

Keywords
Most prestigious private university
A private college
The best private university
A private university
There are private colleges
Telkom Institute of Technology
University Positioning
Kopertis-registered colleges
Private college list

In addition, there is a text crawling on the news portal. The news portals used are based on the results of eight national news portals with the highest level of access to alexa rank in 2020, namely the news portal tribunnews.com, detik.com, okezone.com, sindonews.com, kompas.com, liputan6.com, idntimes.com and merdeka.com. For understanding the temporal behaviour, the data is grouped monthly.

B. Text Pre-Processing

The news articles that have been obtained are unstructured text data and require pre-processing of text that is carried out sequentially and connected to each other in order to prepare the data to be more feasible than input in the next process [14]. The results of the preprocessing process become input for the modeling stage of the topic.

Grouped data requires pre-processing that includes the cleaning, tokenization and stopwords removal phases.

- Cleaning is the process of cleaning data, including the removal of links and dual spaces.
- Tokenize is the process of breaking sentences into words.
- Stopword deletion is a process of deleting words that are considered unimportant. This study is based on Sastrawi literary libraries.

C. Topic Model

The next stage is the topic of modeling using the Latent Dirichlet Allocation (LDA) technique introduced by Blei, et al. In the year 2003, this technique is an unsupervised machine learning technique that can be applied to generative probabilistic text data groups. Documents that make use of this technique can be seen as emerging themes from a number of documents [15].

In addition, the temporal subject model technique is used, a new approach to this technique where the subject of the model that has been generated is shown in a period of time per month by searching for relations in the ranking indicator with the formulation of equation 1.

This stage is the novelty of this research, which uses ranking indicators and visualizes with the Temporal Topic Model technique. The Temporal Topic Model (TTM) method is used to display the topic model generated within a certain time period. Each topic will be displayed in the TTM by linking to the ranking indicators.

$$TTM = < T_{k1}^1 \mapsto T_{k2}^2 \mapsto T_{kn}^n > \quad (1)$$

Where to:

TTM is a time-oriented set of consecutive times

$T_1 \dots T_n$ is a topic at times 1 to n

$K_1 \dots k_n$ is a list of topics

After obtaining the TTM, then visualization is carried out using the equation (1). The visualization on the left is a list of topics $T_1 \dots T_{20}$, 20 topics used refer to the research of Al-khairi, Wibisono and Putro [16] which states that the most

optimal number of topics is 20 topics. On the right is a ranking list of $R_1 \dots R_5$ in each month period $W_1 \dots W_n$ as shown in Fig. 2.

Topik	W1					W2					Wn				
	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5	R1	R2	R3	R4	R5
T1															
T2															
T3															
⋮															
⋮															
⋮															
⋮															
⋮															
⋮															
Tn															

Where to:
 R1 is a rank 1.
 R2 is a rank 2.
 R3 is a rank 3.
 R4 is a rank 4.
 R5 is a rank 5.
 W1...Wn is a period of month.

Fig. 2. Design of TTM.

D. Visualization

Visualization is carried out after obtaining the results of the TTM in order to make it easier for the management of private tertiary institutions to see the preferences of individuals with the wording of equation 2.

$$V_i = \begin{cases} 1 & \text{if } i \in \{T_1, T_2, \dots, T_{20}\} \\ 0 & \text{if } i \notin \{T_1, T_2, \dots, T_{20}\} \end{cases} \quad (2)$$

Where to:

V_i is a TTM visualization

$T_1 \dots T_{20}$ is a topic 1 to top 20

1 is a topic related to the ranking indicator

0 is a topic unrelated to the ranking indicator

Each rating is represented by a color that represents each rating as in Fig. 3.

- The red color (W) represents the Webometrics ranking.
- Yellow color (T) represents Thes ranking.
- The green color (Q) represents the QS-WUR ranking.
- Purple (D) represents the Kemenristek Dikti clustering ranking.
- The blue color (O) represents if there are unrelated topics in the ranking.

The color will live if in a month there is a topic that is related to the ranking indicator. The color will not live if the topic is not related to the ranking indicator.



Fig. 3. Color Ranking.

III. RESULT AND DISCUSSION

The national news articles used in this study were 647 articles generated from crawling using predetermined keywords and news portals from 2016 to 2020. Fig. 4 depicts the crawling process for Indonesian-language university news data. They are also seen by time per month to generate data, as shown in Fig. 5.

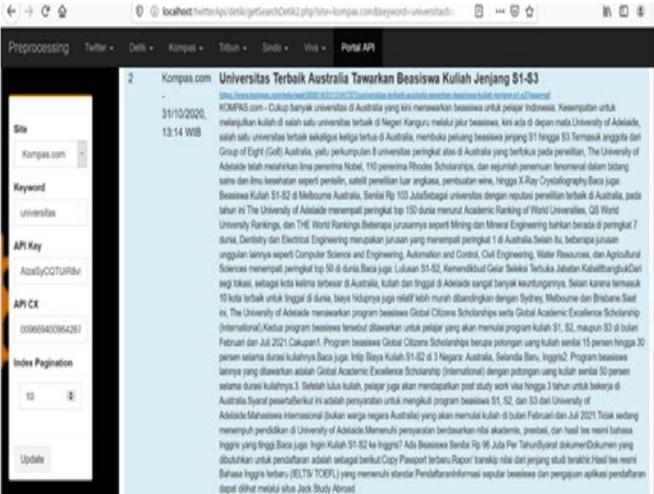


Fig. 4. Crawling Process.

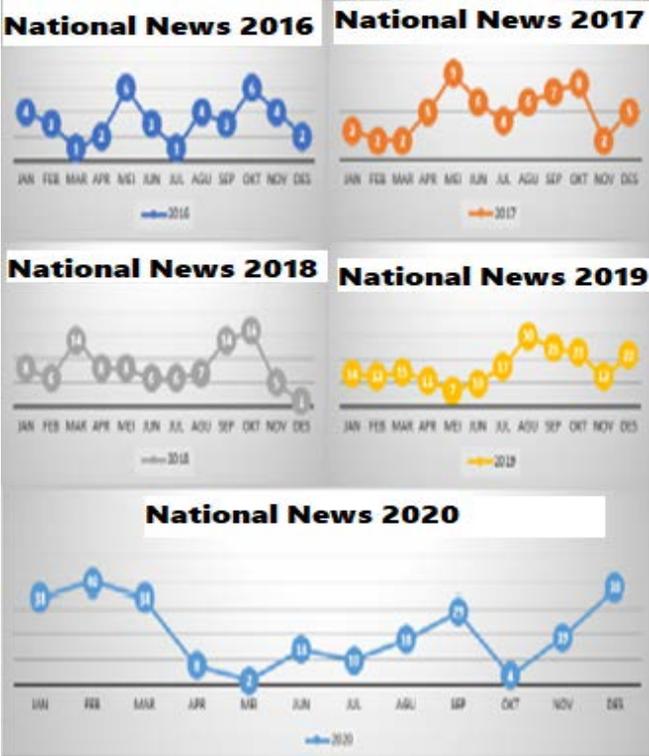


Fig. 5. National News per Month.

The data that has been grouped is then pre-processed to produce better data as input into the process stage of the model theme.

The number of topics to be displayed is determined in advance in the modeling process, and 10 topics are shown in this study. The following is an example of a topic that is produced in December 2020 as shown in Fig. 6.

```
[0,
'0.017*'universitas' + 0.016*'fakultas' + 0.015*'terbaik' + 0.013*'kampus' + '
'0.009*'swasta' + 0.009*'perguruan' + 0.008*'university' + '
'0.007*'pendidikan' + 0.006*'ilmu' + 0.006*'peringkat'),
(1,
'0.021*'universitas' + 0.018*'fakultas' + 0.016*'terbaik' + 0.013*'kampus' + '
'0.008*'perguruan' + 0.007*'university' + 0.007*'swasta' + 0.006*'peringkat' + '
'+ 0.006*'kegiatan' + 0.006*'pendidikan'),
(2,
'0.018*'fakultas' + 0.015*'terbaik' + 0.015*'kampus' + 0.014*'universitas' + '
'0.009*'perguruan' + 0.006*'swasta' + 0.006*'pendidikan' + 0.006*'peringkat' + '
'+ 0.005*'mahasiswa' + 0.005*'ilmu'),
(3,
'0.023*'fakultas' + 0.022*'terbaik' + 0.022*'universitas' + 0.021*'kampus' + '
'0.010*'swasta' + 0.009*'perguruan' + 0.008*'pendidikan' + 0.008*'ilmu' + '
'0.008*'mahasiswa' + 0.008*'peringkat'),
(4,
'0.017*'fakultas' + 0.015*'kampus' + 0.013*'universitas' + 0.011*'terbaik' + '
'0.009*'perguruan' + 0.008*'mahasiswa' + 0.007*'swasta' + 0.006*'mahasiswa' + '
'+ 0.005*'ilmu' + 0.005*'pendidikan'),
(5,
'0.025*'universitas' + 0.021*'fakultas' + 0.019*'terbaik' + 0.017*'kampus' + '
'0.011*'perguruan' + 0.009*'mahasiswa' + 0.007*'pendidikan' + 0.007*'ilmu' + '
'0.007*'peringkat' + 0.006*'swasta'),
(6,
'0.019*'universitas' + 0.019*'fakultas' + 0.018*'kampus' + 0.017*'terbaik' + '
'0.009*'university' + 0.009*'mahasiswa' + 0.008*'perguruan' + 0.007*'swasta' + '
'+ 0.007*'peringkat' + 0.006*'pendidikan'),
(7,
'0.026*'universitas' + 0.025*'fakultas' + 0.022*'kampus' + 0.016*'terbaik' + '
'0.009*'university' + 0.008*'swasta' + 0.007*'pendidikan' + '
'0.007*'perguruan' + 0.007*'ilmu' + 0.006*'peringkat'),
(8,
'0.013*'fakultas' + 0.013*'terbaik' + 0.013*'kampus' + 0.011*'universitas' + '
'0.007*'perguruan' + 0.007*'swasta' + 0.007*'peringkat' + 0.006*'university' + '
'+ 0.006*'ilmu' + 0.006*'jaya'),
(9,
'0.030*'universitas' + 0.020*'fakultas' + 0.015*'kampus' + 0.015*'terbaik' + '
'0.012*'perguruan' + 0.008*'university' + 0.007*'swasta' + '
'0.007*'pendidikan' + 0.006*'kegiatan' + 0.006*'ilmu')]
```

Fig. 6. December 2020 Topic Model.

Topic 9, the ten words that appear at the highest frequency are 'universitas' with a weight of 0.030, followed by the words 'fakultas' with a weight of 0.020, followed by the words 'kampus,' 'terbaik,' 'perguruan,' 'university,' 'swasta,' 'pendidikan,' 'kegiatan' and 'ilmu.' The resulting weight shows the level of importance of the words on the subject.

Visually, words that contain a high frequency of occurrence in one subject are shown in Fig. 7. Using Wordcloud by showing that the word size is larger if the word weight has the highest frequency.

All data for 2020 are shown visually in Fig. 8. Where the left side describes the relationship between one subject and another, while the right side describes the frequency distribution of the word. In the 2020 data set, the top-frequency words are 'fakultas' and 'universitas.' In addition, the Twenty Topics with the highest frequency were used in the Temporal Topic Model process.

The topic model that has been produced only describes the emergence of subjects without knowing the topic shifts that occur every month. For this reason, a Temporary Topic Model and a visualization that is linked to a ranking indicator is needed so that the news topic can be identified as an input for higher education management every month.

Each topic generated per month is linked to the ranking indicator, the indicator is represented in color, so that if a relationship occurs, the color will live on the subject. This determination is based on Eq. 2, so that it is produced as shown in Fig. 9.

From the results of the Visualization of the Temporal Topic Model, it is found that there was a shift in the themes that occurred. For example, from January to May 2016, it is shown in Table II.

Table II shows that there is a shifting subject that happens every month, in January, to news about college scholarships, training, scientific research and scientific publications. February news on scientific research and scientific publications. In March and April, the news is about training, while in May it is about graduates, education and teaching, training and curriculum content.

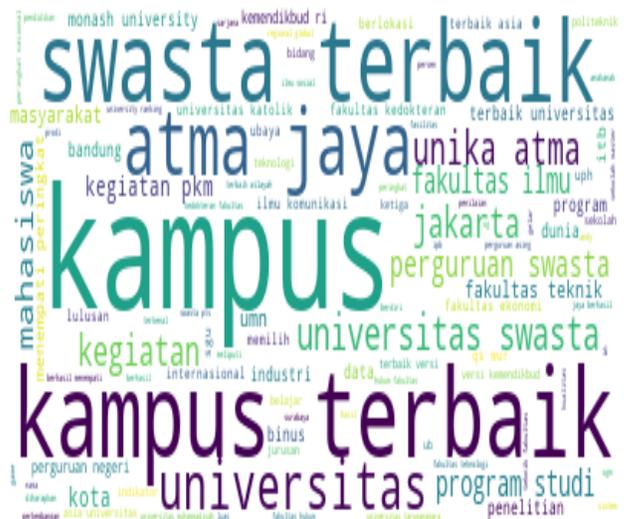


Fig. 7. Wordcloud News December 2020.

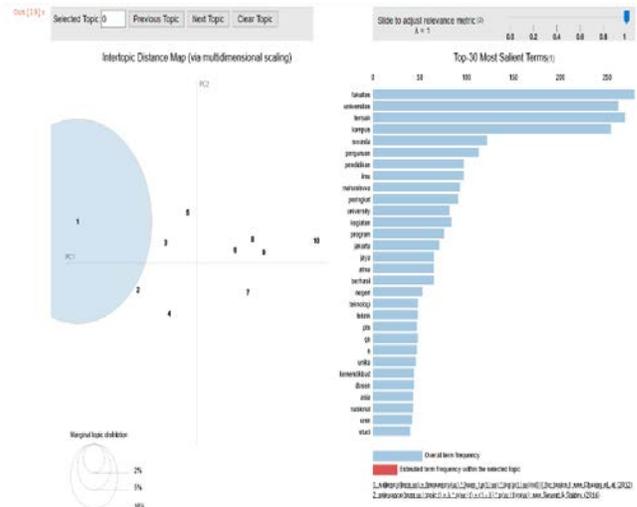


Fig. 8. Frequency of 2020 Data Words.

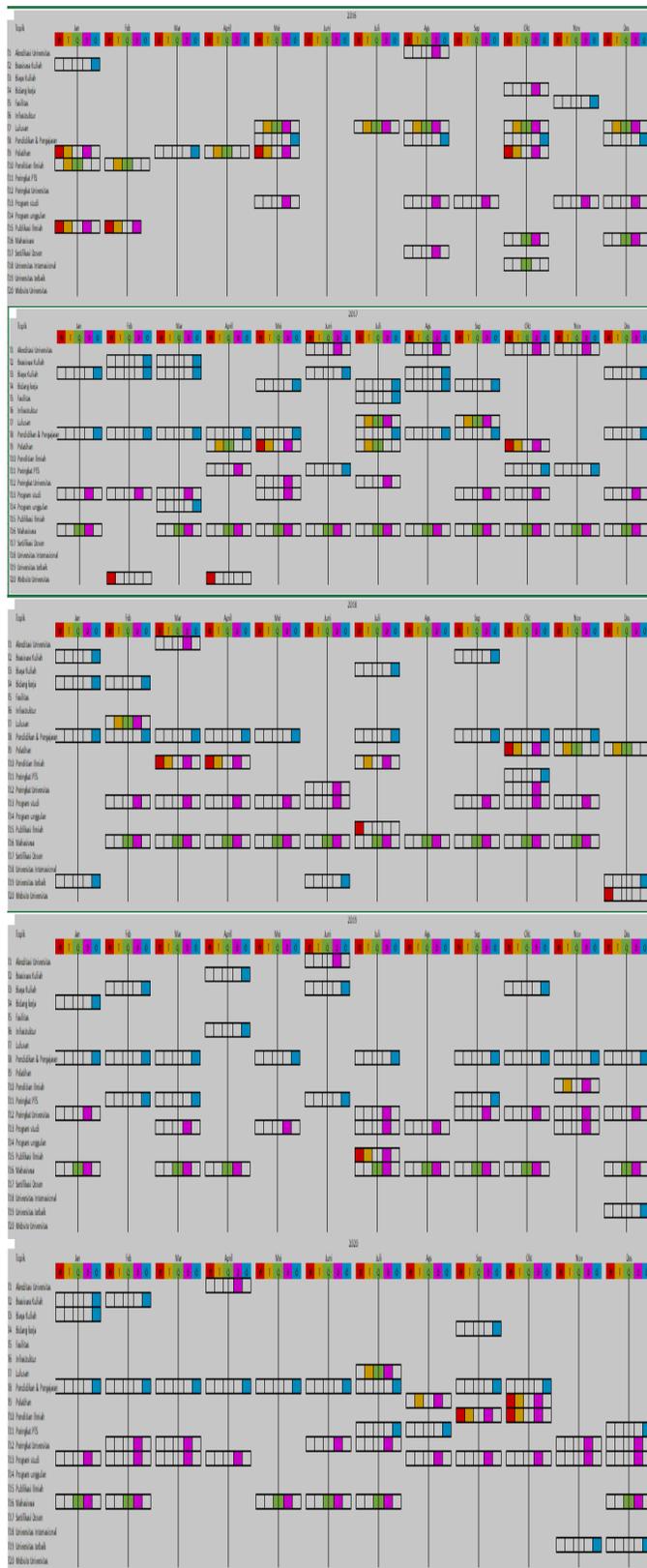


Fig. 9. Visualization of the Temporal Topic Model.

TABLE II. EXAMPLES OF SHIFT TOPIC

January	February	March	April	May
-	-	-	-	graduates
-	-	-	-	education and teaching
-	-	-	-	curriculum content
college scholarships,	-	-	-	-
training,	-	training,	training,	training,
scientific research	scientific research	-	-	-
scientific publications	scientific publications	-	-	-

IV. CONCLUSION

It can provide visual input to the management of private universities to facilitate the analysis of public preferences for higher education, before management decides on marketing strategies and policies that are in accordance with the views of the community, using the findings in the form of two equation formulas that are applied to produce a visualization of the Temporal Topic Model Technique. The resulting visualization is time-based and can be seen changing news topics visually.

The results of the visualization of the TTM obtained class parameters that can be used in the next stage, namely the classification process. The visualization function is carried out by calculating the topics connected to the ranking indicators, then the maximum value of each parameter will be searched, the parameter with the largest value that will be used as a parameter can be used as a feature of the assessment process which is part of the classification process.

ACKNOWLEDGMENT

This work is partially supported by Gunadarma University. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] H. Wahyudi and I.A. Wibowo, "Inovasi dan Implementasi Model Pembelajaran Berorientasi Lulusan (Outcome-Based Education, OBE) dan Washington Accord di Program Studi Teknik Mesin Universitas Mercu Buana", *J. Tek. Mesin*, 7(2), 50, 2018.
- [2] R. Belmawa, "Panduan Penyusunan Kurikulum Pendidikan Tinggi Di Era Industri 4.0", 2018.
- [3] A. Wahyuni, "Kajian Bauran Promosi Di Perguruan Tinggi "X"", *Liquidity*, 1(2), 175-182, 2012. doi: <https://doi.org/10.32546/lq.v1i2.148>
- [4] S. Sastropetro, "Pendapat Publik, Pendapat Umum, dan Pendapat Khalayak Dalam Komunikasi Sosial", Bandung: Remaja Rosdakarya, 1987, unpublished.
- [5] M. Gunarto, "Analisis Persepsi Dan Preferensi Masyarakat Terhadap Reputasi Pemingkat Perguruan Tinggi", *Journal Ilmu Manajemen*, 5(2), 2016.
- [6] O. Alaşehir, "University ranking by academic performance", Master's thesis, 2010. <http://etd.lib.metu.edu.tr/upload/12612484/index.pdf>.

- [7] Nizam, "Direktorat jenderal pendidikan tinggi umumkan klasterisasi perguruan tinggi Indonesia tahun 2020", <https://dikti.kemdikbud.go.id/kabar-dikti/kabar/direktorat-jenderal-pendidikan-tinggi-umumkan-klasterisasi-perguruan-tinggi-ind, 2020>.
- [8] D.H. Jeong and M. Song, "Time gap analysis by the topic model-based temporal technique", *Journal of informetrics*, 8(3), 776-790, 2014. doi: <https://doi.org/10.1016/j.joi.2014.07.005>.
- [9] K. Jatmika and A. Cahyono, "Rancang Bangun Data Mart dan Purwarupa Dashboard untuk Visualisasi Performa Akademik", *SISFO*, 5, 2015.
- [10] S. Ghosh, P. Chakraborty, E.O. Nsoesie, E. Cohn, S.R. Mekar, J.S. Brownstein and N. Ramakrishnan, "Temporal topic modeling to assess associations between news trends and infectious disease outbreaks", *Scientific reports*, 7(1), 1-12, 2017. doi: <https://doi.org/10.1038/srep40841>.
- [11] KBBI, Kamus Besar Bahasa Indonesia (KBBI), Online <https://kbbi.web.id/katakunci>, 2021.
- [12] G.K. Planner, "Google Keyword Planner Perguruan Tinggi", <https://ads.google.com/aw/keywordplanner>, 2020.
- [13] IREG, Ranking Audit, "Rank University", <https://web.archive.org/web/20161029232754/http://ireg-observatory.org/en/information>, 2016.
- [14] W.W. Ariestya, I. Astuti and I.M. Wiryana, "Preprocessing For Crawler Of Short Message Social Media", In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-6), IEEE, 2018, October. doi: 10.1109/IAC.2018.8780451. <https://ieeexplore.ieee.org/abstract/document/8780451>.
- [15] D. M. Blei, A.Y. Ng and M.I. Jordan, "Latent dirichlet allocation", *The Journal of Machine Learning Research*, 3, 993-1022, 2003.
- [16] Y.U. Al-khairi, Y. Wibisono, B.L. Putro, "Deteksi topik fashion pada twitter dengan latent dirichlet allocation". *JATIKOM: Jurnal Aplikasi dan Teori Ilmu Komputer*, 1(1):1-8, 2018. <http://jatikom.cs.upi.edu/index.php/jatikom/article/view/14>.