

Study on Automobile Culture of Developed and Emerging Countries in Asia Using Text Mining Analysis on Social Media

Yongho Joo^{a,1}, Gilsang Yoo^{b,2}

^aDepartment of Business Administration, Graduate School, Sogang University, 35 Beakbeom-ro, Mapo-gu, Seoul, 04107, Republic of Korea

^bCollege of Informatics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea

Corresponding author: ¹jyh@hmtalk.com, ²phd.yoo@gmail.com

Abstract— To be successful in the automobile export business, it is essential to accurately identify consumer preferences in developed and emerging countries. In this study, big data were collected and compared using NodeXL to verify differences in automotive preferences caused by differences in living standards, regional characteristics and culture, and related regulations and policies between developed and emerging countries in Asia. The data to be analyzed were collected from Twitter big data on major Asian and emerging countries and cleaned up, and text mining techniques were applied for frequency analysis. Based on this, we predicted the demand for automobiles in emerging countries in the future and derived a way for automakers to efficiently approach the automobile markets of emerging countries. The analysis results are as follows: First, emerging economies are very fond of foreign vehicles that value their confidence and pride but are more price-sensitive and prefer medium-to-small car sedans rather than SUVs. Second, pride and price should be the first marketing considerations when an automotive business enters an emerging country. Third, in emerging economies, if small- and medium-sized cars that fit lower prices are marketed mainly as sedans rather than as SUVs, local consumers will respond well. This study's results are expected to be used as primary data for the localization sales factors necessary in understanding the car preferences of different countries for overseas marketing.

Keywords— NodeXL; term frequency; text mining; auto industry; comparative analysis; marketing.

Manuscript received 24 Jul. 2020; revised 29 Aug. 2020; accepted 2 Dec. 2020. Date of publication 28 Feb. 2021.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

In modern society, social big data is rapidly generated and spread through blogs and social network services (SNS) in various forms, such as text documents, images, and videos, and emotions, tastes, and preferences for products left by social media users in platforms such as Twitter and Facebook. As various messages in various fields, such as culture, emerged as a source for understanding the emotions of the times, the policy agendas established by mass media can now be understood from social media [1]. As such, social media communication has become a major phenomenon in our society. From such big data on the Internet, including in social media, a large amount of meaning and valuable information can be extracted, making it possible to investigate and predict public perceptions, intentions, and actions [2]. The trend with processing big data to infer meaningful thoughts or actions from the public is the sizes of the data to be analyzed continuously growing because of the generalization of the Internet environment and the revolutionary development of

data storage systems. Companies are now using marketing strategies to quickly respond to products or services based on information collected from SNS [3].

Therefore, numerous comments and large amounts of social log information sent and received by consumers in SNS regarding car preference can be used as key information in establishing management and marketing strategies for automotive and other related businesses.

In the field of management of engineering, text mining is performed using patent documents and big data. The predictive analysis results are applied to conduct the necessary research and development in the enterprise [4]–[7]. Big data analysis is a technique that extracts value and analyzes results from dozens or thousands of terabytes of structured and unstructured data every day [8]. The techniques for analyzing such big data are very diverse and are continuously being developed. Among them, text mining, natural language processing (NLP), and visualization are the main technologies.

Text mining be the process of discovering new meaningful information from the natural language formed from large text populations [9]. Text mining must be preceded by a semantic network analysis process, which analyzes the frequencies of words in the text, based on the recognition that language and knowledge are a network formed by the relationships between words. These relationships are visualized via connections and links to enable the interpretation of cultural phenomena [10]. Research on quantitative analysis using text mining techniques has been actively and successfully conducted in various fields [11]. Whereas traditional research methods obtain opinions from people utilizing questionnaires and manual labor, which consume vast amounts of time and resources, these are now being actively replaced via research into big data analysis using computers.

In this study, we used NodeXL (Twitter) among the text mining techniques to compare car preferences between developed and emerging Asian countries. Based on this, we predicted the demand for automobiles in emerging countries in the future and derived a way for automakers to efficiently approach the automobile markets of emerging countries. A flowchart of the research process for this method is shown in Fig. 1.

This paper is organized as follows. Section II presents an analysis of the current automotive market in major Asian countries, including China and Japan, and describes the term frequency (TF) process via collection and preprocessing of social media data. Section III explains how the hypotheses presented in Section II, were proved through a frequency analysis on major countries. Finally, Section IV outlines the conclusions and implications of this study, based on the frequency analysis results, and reveals future research directions.

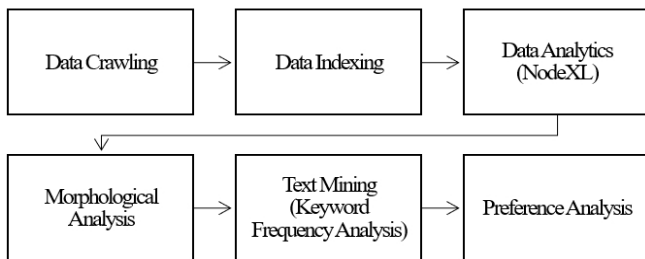


Fig. 1 Research flow

II. MATERIALS AND METHODS

A. Related Work

Because of the development of information and communication technology (ICT) and its ease of use, data exploding in various online media has resulted in a large amount of useful information for corporate marketing. Whereas in the past, there were technical limitations in collecting, processing, and analyzing such data, advanced data analytics professionals in public and private sectors. But now, with big data analysis, it is possible to infer meaningful information by predicting people's behavior patterns or market economic conditions. Trends or insights can be derived, and the new added value can be created. In the evolution of communication via social media services, second-generation SNS was a relationship-based

communication-based activity. It is now evolving from information-delivery to information-sharing activities to digital-curation services that can distribute and communicate advanced qualitative information.

Since the concept of Web 2.0 was introduced, the scale of communication itself has increased enormously as platform services have been changed to be readily available to the public based on the fundamental philosophy of participation, sharing, and openness. This social big data analysis is a series of processes that can find meaningful insights by analyzing huge amounts of data from various social media, such as news, blogs, and Twitter, and reveals time-series patterns of the generation and distribution of stories. Big data analyzed through this technical method are essential to the process of visualization. Like big data visualization, a visualization expression method can deliver an informational message through visualization or a persuasive message.

With the advent of big data, many text data has been created, and research using text mining is being actively conducted [12]–[18]. In text-mining-related studies, researches on topics such as morphological analysis of texts, methodological research related to preprocessing, topic modeling, emotional dictionary construction, and emotional analysis have been reported. These findings are applied to various management decision-making areas, including social marketing, where analysis of real people's opinions is very important.

The importance of the reliability of data used in research has been steadily reported. As research using various kinds of Internet data increases, the importance of the reliability of data used in research also increases. In this regard, there have been qualitative studies that address the importance of data reliability and research model equity in conducting research based on data collected through the Internet. For example, Howison et al. reported on the existing research literature on the necessity of verifying the quality and reliability of data to increase the persuasiveness of research [19].

Regarding the expert system, which is a representative decision support system using qualitative data, the issue of data verification is also emerging. O'Keefe et al. [20] compared the results of expert systems with the results of surveys on real experts. For comparative evaluation, statistical techniques. Such as face test and Bayesian model were used. A study by Liebowitz et al. discussed the use and reliability of expert systems [21]. Our study introduces a method based on statistical techniques for verifying expert systems' performance that deal with qualitative data. A comparative analysis method with survey results is also discussed.

B. Status of the Automotive Market in Asia by Country

1) *China Automotive Market:* In China, the largest automobile producer in the world, the domestic market consumes most of its annual output from the automotive industry. Automobiles in China overtake those in the US and Japan in terms of numbers. Meanwhile, automakers in the US, EU, Japan, and Korea have assembly plants in China. However, the automotive industry in China is characterized by assembly plants in cooperation with Chinese state-owned companies, such as SAIC, under the Chinese government's regulation. Currently, Honda is exporting cars manufactured

in China to Canada, and exports of cars manufactured in China to North America are expected to expand in the future. At present, China has a phenomenal sales volume for the first automobile market in the world. Chinese perceptions of foreign cars can be thought of in connection with globalism, which refers to individual dispositions that are open to diverse cultural experiences that transcend specific cultures or environments. Openness and tolerance are the hallmarks of globalism. Globalist consumers generally value the high quality, reputation, and discriminatory value of foreign brands [22]. Consumer globalism can be influential in developing countries, including China. Globalism has a greater impact on foreign product evaluations among consumers in China than in Korea, with this phenomenon being also seen with automobiles.

TABLE I
JAPANESE AUTOMAKERS IN THE UK

Automakers		Nissan	Toyota	Honda
2014	Ranking	1st	4th	5th
	Number of productions	500,238	172,215	121,799
2015	Ranking	2nd	4th	5th
	Number of productions	476,589	190,161	119,414
2016	Ranking	2nd	4th	5th
	Number of productions	507,000	180,000	134,000

2) *Japan Automotive Market:* Whereas Japanese automobile production continues to decline because of slowing consumption, overseas production is rising as Japanese cars gain popularity due to price competitiveness. The Japanese automotive industry is rushing to renewable energy, such as fuel cell and electric vehicles, as a future strategic model to combat global warming and fossil fuel depletion. Hydrogen cars are also on the rise. After Hyundai Motor Company (HMC) launched the world's first hydrogen car in 2013, Toyota followed in January 2015, followed by Honda in December 2016. German automakers are also involved in the development of hydrogen vehicles. At present, Japanese automakers are focusing on natural gas cars, hybrid cars, electric cars, clean diesel, plug-in hybrids, and fuel cell vehicles like the next-generation cars, and are rushing their development and commercialization. To date, the number of next-generation cars in Japan accounts for 5.4% of the total number of vehicles owned in Japan. To this end, the Japanese government has been providing purchase subsidies since 1998 and subsidizes clean diesel for plug-in hybrid electric vehicles (PHEVs) and electric vehicles (EVs) in consideration of market trends and conditions. Nationwide infrastructure construction is urgently needed to expand the distribution of this next-generation strategic model. To this end, the government, automobile makers, and affiliates are jointly striving to build fuel supplies and charging infrastructures. The three major Japanese automakers have production in the UK and business in Europe. Statistics on Japanese automakers in the UK are listed in Table I [23].

3) *Vietnam Automotive Market:* According to a report by the Korea Customs Service on the Vietnam automotive industry status and outlook for 2019, Vietnam Customs imported 72,650 units in 2018, nearly 20% less than in 2017. Automobile imports in 2018 were reported to have risen 21 percent year over year to over \$ 1.64 billion, with Thailand

and Indonesia being suppliers of imported cars in Vietnam. In 2018, Vietnam imported more than 52,170 cars, worth a total of \$ 1.4 billion, from Thailand. Most of the cars sold in Vietnam are imported as modules and assembled in the country.

4) *Indonesia Automotive Market:* The Indonesian automobile market has been stagnant since 2014 and then recovered slowly until 2018. According to the Indonesian Automobile Manufacturers Association (GAIKINDO), total vehicle sales in Indonesia in 2018 were 1.15 million units, an increase of about 6.8% compared to in 2017. In 2018, the main reason for the rise in vehicle sales in Indonesia was the increase in commercial-vehicle sales. Whereas the increase in passenger-car sales was only 87,660 units, an increase of only 3.82% year on year, that of commercial vehicles was 276,631 units or up 17.77% year on year. In 2018, the number of commercial-vehicle sales had not exceeded those in 2013 and 2014 because of the slowing economic growth. In 2015, as raw material prices fell, the mining and plantation industries' growth, which mainly require commercial vehicles, also slowed down. The Indonesian car-sales market's size and trends in 2013–2018 are shown in Fig. 2.

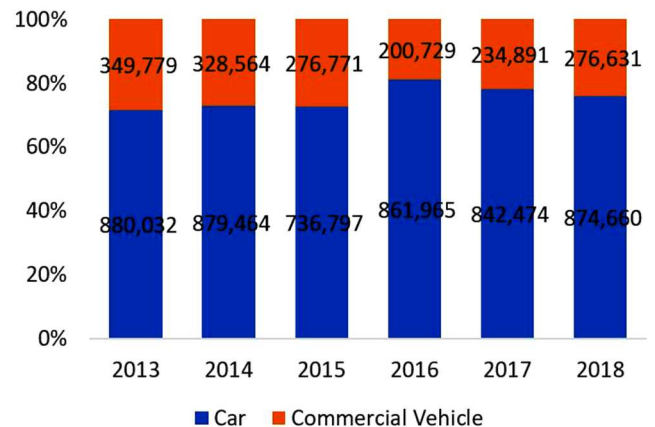


Fig. 2 2013–2018 Indonesia vehicle sales market sizes and trends

C. Hypothesis Settings

The following three hypotheses regarding the automobile preferences of developed and emerging Asian countries were formed:

- *H1:* Emerging Asian countries will be more price-sensitive than in developed countries.
- *H2:* In emerging Asian countries, confidence and foreign awareness will be more important.
- *H3:* Developed Asian countries, compared to emerging *economies*, will prefer gasoline.

1) *Preferred Automotive Designs in Developed and Emerging Countries in Asia:* Developed and emerging economies in Asia will clearly differ in terms of the car brand and design perception because of differences in living standards, regional characteristics and cultures, and related regulations and policies. Developed countries in Asia, unlike emerging countries, will prefer luxury brands and mid- to large-scale luxury designs. The *H1* hypothesis was used to verify what the difference between these two preferences is.

2) *Recognition by Automotive Use in Developed and Emerging Countries in Asia:* Developed and emerging

economies in Asia will obviously have different uses and purposes for using cars due to differences in living standards, regional characteristics and cultures, and related regulations and policies. Developed countries prefer leisure and safety, whereas emerging economies will be more concerned about confidence and public awareness. The *H2* hypothesis was used to verify the difference between the preferred uses and purposes.

3) *Automotive Fuel preferences in Developed and Emerging Asian Countries*: Developed and emerging economies in Asia will certainly differ in their preferred fuels because of differences in living standards, regional characteristics and cultures, and related regulations and policies. The *H3* hypothesis validated the differences in fuel preferences.

D. Research Design and Analysis Methods

This study is about a method of researching automobile consumption culture using big data. This method uses NodeXL (Twitter) to search by car word and country and compare car brands, car favorites, and rankings. First, using Twitter, we collected car preferences by country. The reason why Twitter was chosen among various media, such as portal, newspaper, and TV, reflects the fact that it is composed of implied sentences, unlike Facebook, and is widely used and influenced by the media. A sample of the data collected using NodeXL from Twitter in Japan is shown in Fig. 3.

Fig. 3 Data collected using NodeXL

1) *Search and Collect Car-related Words on Twitter by Country*: SNS has several influences in various fields of domestic society. Table 2 shows the current state of data collection by topic. Two related searches were used for each country, with an average number of crawls of 20,000 sentences.

TABLE II
CAR-RELATED WORDS AND TWITTER SEARCH RESULTS BY COUNTRY

Country	Related Search Terms 1	Related Search Terms 2	Number of Crawls
Korea	자동차	차	13,331
China	汽车	车	19,961
Japan	じどうしゃ	くるま	11,232
Vietnam	xe ô tô	ô tô	30,754
Indonesia	mobil	oto	18,613

2) *Morphological Analysis*: The collected data were processed via unification of synonyms, abbreviations, and spacing, and deletion of unnecessary data. After pure contents were extracted from Twitter, data preprocessing and stemming analysis excluded unnecessary elements. In the data preprocessing step, the words unnecessary for analysis were deleted from the collected text data, and normalization of unifying and similarly expressed words was repeatedly performed. Among the data collected through crawling, stop-words that do not need to be given various special symbols (characters) and meanings were removed. Morphological analysis is the process of recognizing each morpheme constituting a word and restoring the original form from irregular utilization, abbreviation, or dropout. To make the form suitable for morphological analysis, nouns were extracted from Twitter using KoLNP, the R program's Korean natural language analysis package.

3) *Keyword Frequency Analysis with Text Mining*: Keyword frequency analysis was conducted primarily to determine what parts of the keyword analysis of brand preferences by country were of interest. Keyword frequency analysis applied normalized TF by calculating the frequency using the R program and then dividing the word occurrence frequency of the document by the total number of occurrences of all the words, to prevent bias caused by variations in document sizes. To visualize the results of the analysis, we used the word cloud of the R program.

III. RESULTS AND DISCUSSION

Keyword frequency analysis was classified into manufacturers, brand names, and vehicle types (small, medium, large, and SUV). More than 1–20,000 sentences were searched on Twitter, and the frequency of each word was derived. Advanced Asian countries, mainly Korea and Japan, and emerging Asian countries, mainly China, Vietnam, and Indonesia, were analyzed for consumption culture, purpose, and sensitivity. Word cloud analysis, which is an analysis method frequently used in big data analysis, was conducted to intuitively identify key automotive topics, and the result is shown in Fig. 4. This technique visualizes the sizes of words and the number of word correlations without meaning. Table 3 shows the results for 32 high-tech keywords that frequently appeared in the search results on Twitter.

TABLE III
CAR KEYWORD FREQUENCY RESULTS IN TWITTER

Ranking	Extracted Words	Frequency	Translation	Percentage
1	イライラ	486	Frustration	11.27%
2	くるまにあ	398	Kurumania	9.23%
3	プレゼント	395	Gift	9.16%
4	くるまや	384	Tea shop	8.90%
5	はたらくくるま	373	Working car	8.65%
6	くるまさてい	314	Car rating	7.28%
7	くるまかいとり	303	Tea purchase	7.03%
8	法定速度 60	298	Courtroom Speed 60	6.91%
9	なものを	298	item	6.91%

10	kurusoku	167	kurusoku	3.87%
11	トヨタ	117	Toyota	2.71%
12	ホンダ	92	Honda	2.13%
13	ぶつかって	83	Bump	1.92%
14	マツダ	82	Mazda	1.90%
15	くるまほしい	77	Car circle	1.79%
16	おばけに	76	ghost	1.76%
17	開始1行目	74	Start line 1	1.72%
18	ってる	71	Shines	1.65%
19	中古車	60	used car	1.39%
20	車買取	60	Tea purchase	1.39%
21	バイク	58	bicycle	1.34%
22	はたらく	58	work	1.34%
23	温度設定	56	Temperature setting	1.30%
24	ドライブ	56	drive	1.30%
25	クリスマス	53	Christmas	1.23%
26	くるまえび	50	Shrimp	1.16%
27	フォロ	49	Rome	1.14%
28	スタンプ	46	stamp	1.07%
29	ブログ	46	blog	1.07%
30	なんか	46	no wonder	1.07%
31	スバル	44	Subaru	1.02%
32	スズキ	43	Suzuki	1.00%

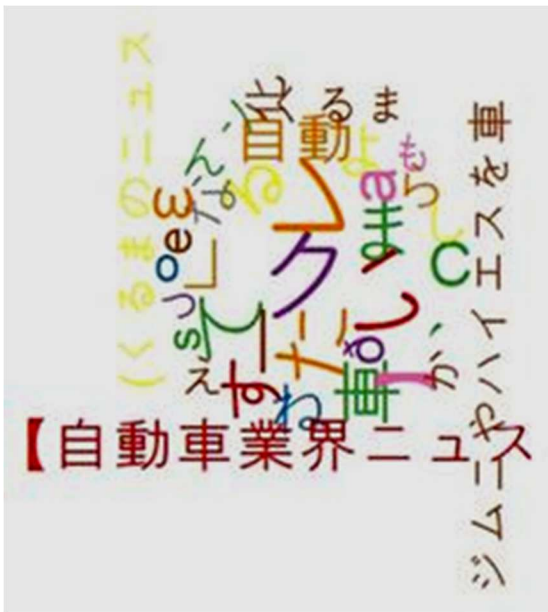


Fig. 4 Word cloud visualization

A. Surveys on Car Brands and Model Types in Developed Countries in Asia

As shown in Table 4, Japan is one of the most advanced countries in Asia, and the lives of Japanese people are inseparable from cars. Preferred automakers are Toyota, Honda, Subaru, Suzuki. Preferred car brands include the Prius, a small car, followed by Lexus and Tomica. In Japan, used car trading seems to be active, and gasoline is preferred as fuel, which corresponds to *H3*.

TABLE IV
COMPARISON OF AUTOMOBILE PREFERENCE IN DEVELOPED AND EMERGING COUNTRIES IN ASIA

	Japan	Indonesia
Price	-	Sensitive
Manufacturer	Toyota>Honda>Subaru>Suzuki	Toyota>Dugati>Ferrari>Nissan>Kawasaki>Honda>Volkswagen>Yamaha>Chrysler> Chevrolet>Suzuki
Brand	Prius>Lexus>Tomica	Nismo>Fiat>G90 (Hyundai)
Makes	-	Sedan>SUV
Fuel	Gasoline preferred	Diesel preferred
Foreign vehicles	Unfriendly	Ho
Used	Active	
Views	Hee-no-rak, Life	Proud

B. Surveys on Car Brands and Model Types in Emerging Asian Countries

As shown in Table 5, Indonesia, one of the emerging economies, has strong car confidence and pride, but the price is important because of economic problems. Preferred automakers are Toyota, Dugati, Ferrari, Nissan, Kawasaki, Honda, Volkswagen, Yamaha, Chrysler, Chevrolet, and Suzuki. Car brand preferences include Nismo, Fiat, and G90 (Hyundai). Indonesians are most likely to love sedans, followed by SUVs. In addition, much attention has been paid to diesel engines (*H1*, *H2* supported).

TABLE V
HYPOTHESIS SYNTHESIS ANALYSIS RESULT

Hypotheses	Results
H 1 Emerging Asian countries will be more price-sensitive than in developed countries.	Accept
H 2 For emerging Asian countries, confidence and foreign awareness will be more important.	Accept
H 3 Developed Asian countries, compared to emerging economies, will prefer gasoline.	Accept

IV. CONCLUSIONS

One of the important characteristics of social media communication is that it allows for constant conversation. In general, these conversations seem to be a series of 1:1 communication, but in real social-media environments, most actually appear in the form of 1:n communication. Processing of this communication should be strategically performed to collect, analyze, and respond to big data. Recently, various methods for discovering meaningful information from big data generated in real-time have been introduced and applied to decision-making for marketing. If a group manages social media, it can thoroughly examine its on- and off-line systems, analyze overall user trends through its platform, and execute systematic big data strategies.

This study aimed to understand the trend of automobile consumption by using social media big data, such as what

keywords are related to automobiles and how consumption behaviors of developed and emerging countries are represented. This study verified and compared the results with hypotheses that automobile preferences would be different because of differences in living standards, regional characteristics, culture, and related regulations and policies between developed and emerging Asian countries.

As demonstrated by the study results, emerging countries are very fond of foreign cars that value their confidence and self-esteem, but they are sensitive to price, and thus they prefer small- and medium-sized cars sedans rather than SUVs. Therefore, the price and marketing should be considered as the top priority when automakers enter these emerging markets. If small- and medium-sized cars that are more affordable are marketed as sedans rather than as SUVs, these markets can be expected to respond more locally.

Through this method, the demand for automobile buyers can be identified, and the necessary direction for activating automobile exports in the future can be established. The results can also be expected to be used as basic data in marketing strategy and product development for automobile export policies.

Despite its significance, this study has some limitations. First, this study analyzed only one keyword and its detailed related keywords. Therefore, future research will need to generalize the results by diversifying subject keywords. Second, the data collection path was limited to Twitter. Because there are more SNSs that can collect text data on the Internet. It is necessary to generalize the method through diversification of collection paths. Third, empirical verification of the results obtained from such data analysis is necessary.

ACKNOWLEDGMENT

Basic Science Research Program supported this research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R111A1A01064580).

REFERENCES

[1] T. A. Al-asadi, A. J. Obaid, R. Hidayat, and A. A. Ramli, "A survey on web mining techniques and applications," *Int J Adv Sci Eng Inf Technol.*, vol. 7, no. 4, pp. 1178–1184, 2017.

[2] P. V. N. Rajeswari, and Kodi Roshan Sai Kumar, "A Design Approach for Efficient Mining Human Activity Patterns Over Big Data," *Indian Journal of Science and Technology*, vol. 12, issue. 3, pp. 1–6, 2019.

[3] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000-2015," *Inf. Process. Manag.* Vol. 53, pp. 122–150, 2017.

[4] S. Malhotra, M. N. Doja, B. Alam, and M. Alam, "Bigdata analysis and comparison of bigdata analytic approaches," in *Proc. ICCCA' 17*, 2017, pp. 309–314.

[5] M. Lyons, N. D. Aksayli, and G. Brewer, "Mental distress and language use: Linguistic analysis of discussion forum posts," *Computers in Human Behavior*, vol. 87, pp. 207–211, 2018.

[6] Z. Rahimi, D. Litman, R. Correnti, E. Wang, and L. C. Matsumura, "Assessing students' use of evidence and organization in response-to-text writing: Using natural language processing for rubric-based automated scoring," *International Journal of Artificial Intelligence in Education*, vol. 27, pp. 694–728, 2017.

[7] G. Yoo, and S. Hong, "Analysis of Autonomic Nervous System of Learners Arousal state and Drowsy state through Photoplethysmogram," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 19, pp. 1901–1913, 2018.

[8] F. B. B. Nasution, N. E. N. Bazin, Dalijusmanto, and A. Zulfikar, "Big data's tools for internet data analytics: modelling of system dynamics," *Int J Adv Sci Eng Inf Technol.*, vol. 7, no. 3, pp. 745–753, 2017.

[9] A. H. Lee, K. N. M. Kamil, and S. Narendranaath, "Beyond Words of Popularization Mining: Reviews on Comic Books Movies," *Indian Journal of Science and Technology*, vol. 11, issue: 25, pp. 1–11, 2018.

[10] A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas, "A survey on big data for network traffic monitoring and analysis," *IEEE Trans Netw Serv Manag.*, vol. 16, no. 3, pp. 800–813, Sept. 2019.

[11] M. Alban1, and D. Mauricio, "Predicting University Dropout through Data Mining: A Systematic Literature," *Indian Journal of Science and Technology*, vol. 12, issue. 4, pp. 1–12, 2019.

[12] D. Buenaño-Fernandez, M. González, D. Gil, and S. Luján-Mora, "Text mining of open-ended questions in self-assessment of university teachers: an LDA topic modeling approach," *IEEE Access*, vol. 8, pp. 35318–35330, Feb. 2020.

[13] C. Rauh, "Validating a sentiment dictionary for German political language—A workbench note.," *J. Inf. Technol. Polit.*, vol. 15, pp. 319–343, 2018.

[14] T. Z. Emara and J. Z. Huang, "A distributed data management system to support large-scale data analysis," *The Journal of Systems and Software*, vol. 148, pp. 105–115, 2019.

[15] S. Salloum, J. Z. Huang, Y. He, and X. Chen, "An asymptotic ensemble learning framework for big data analysis," *IEEE Access*, vol. 7, no. 1, pp. 3675–3693, 2019.

[16] K. Li and G. Li, "Approximate query processing: What is new and where to go?," *Data Science and Engineering*, vol. 3, no. 4, pp. 379–397, 2018.

[17] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Data Mining and Knowledge Discovery*, vol. 8, no. 4, pp. 1–18, 2018.

[18] S. Salloum, J. Z. Huang, and Y. He, "Exploring and cleaning big data with random sample data blocks," *Journal of Big Data*, vol. 6, no. 1, p. 45, 2019.

[19] J. Howison, A. Wiggins, and K. Crowston, "Validity issues in the use of social network analysis with digital trace data," *J Assoc Inf Syst.*, vol. 12, no. 12, p. 767, Dec. 2011.

[20] R. M. O'Keefe and D. E. O'Leary, "Expert system verification and validation: a survey and tutorial," *Artif Intell Rev.*, vol. 7, no. 1, pp. 3–42, Feb. 1993.

[21] J. Liebowitz, "Useful approach for evaluating expert systems," *Expert Syst.*, vol. 3, no. 2, pp. 86–96, Apr. 1986.

[22] R. Hussein and S. Hassan, "Antecedents of global brand purchase likelihood: exploring the mediating effect of quality, prestige and familiarity," *J Int Consum Mark.*, vol. 30, no. 5, pp. 288–303, May 2018.

[23] J. Han and H. Han, "An analysis of competitiveness of Korea and Japan automobile industry due to Japan-EU EPA," *J Korea Res Soc for Cust.*, vol. 19, no. 4, pp. 291–303, Dec. 2018.