Original Research Article

Sentiment Analysis for Market Survey in The Framework of The Establishment of The IT Boarding School Vocational School

ABSTRACT

Aims: This study aims to obtain indicators which are then compiled in a questionnaire research instrument which will later be used in further analysis namely sentiment analysis.
Methodology: Sentiment analysis is a method used to identify how a sentiment is expressed using text and how that sentiment can be categorized as positive sentiment or negative sentiment. boarding school. So that researchers can conclude whether the sentiments or issues that arise in the tweets published by the surrounding community on social media are more supportive or even less supportive regarding the boarding school program at a boarding school.
Results: The findings of this study are the variables that can be used in further research, namely marketing communication strategies, marketing mix, and interest in learning.
Conclusion: Based on the sentiment analysis, it was found that society has a neutral tendency regarding public perceptions in Indonesia regarding the boarding school program at a boarding school.

Keywords: Market Survey; Sentiment Analysis; Marketing Communication Strategy; Marketing Mix; Interest to learn

1. INTRODUCTION

One of the causes of the decline in the quality of Indonesian society is the low quality of education, which in this case can be interpreted as a lack of effectiveness in the learning process developed in an institution. The cause is derived from the students themselves, the methods applied, the learning environment and other factors. Students' low motivation interest and limited facilities and infrastructure will cause the learning process to be less effective and efficient which in the end will have a negative impact on the output of the institution concerned.

In the education sector, a student ability survey released by the Program for International Student Assessment (PISA) in Paris, placed Indonesia in 72nd place out of 77 countries. This data puts Indonesia in the sixth lowest rank, still far below neighboring countries such as Malaysia and Brunei Darussalam. The PISA survey is a reference in assessing the quality of education in the world, which assesses reading, math and science skills.

Malang Regency is one of the regencies in Indonesia with a total area of 3,535 km². Based on the 2010 census, the district has a population of 2,451,997, consisting of 1,230,594 males and 1,221,403 females. In the 2015 population census, the population of Malang Regency increased to 2,544,315 people. Malang Regency has 33 sub-districts. The Local Government of Malang Regency as the spearhead of implementing government duties in Malang Regency certainly wants an increase in the welfare felt by the community. As the 2nd largest area in East Java, Malang Regency has enormous potential in terms of economy, health, and education. This potential is also supported by a population that reaches 2.6 million and makes Malang Regency the 2nd largest district in East Java in terms of population. The large number of residents in Malang Regency does not guarantee high welfare either. Based on data from the National Socio-Economic Survey (Susenas), there were around 268,490 poor people in Malang Regency, in March 2018. This is equivalent to 10.37 percent of the total population of Malang Regency. For this reason, special attention is needed in addressing the special character of Malang Regency with programs that support the advancement of its resources.

The quality of Malang's workforce needs to be improved in order to be able to compete with other regions, and meet the specifications required by the business world. Improving the quality of the workforce can be done through improving the education level and skills of the Malang Regency workforce. The amount of APBD funds with a minimum rule of 20% of funds must be allocated for education, is expected to improve the quality of education services and skills which are the basic capital of workforce expertise. BPS data shows that in 2020 only 6.12% of Malang Regency's workforce graduated from university, while elementary school graduates reached 49.29%. This means that the executive and legislative bodies must be more serious in carrying out the functions of planning, controlling, and evaluating the use of APBK funds (Regency Expenditure Budget). Optimizing the function of the institution to ensure that the realized work program has an outcome in the form of increasing human resources in Malang Regency.

The majority of the population of Wajak Subdistrict, Malang Regency work as farmers and traders, besides that there are also many residents who work as craftsmen. Among them are woven mat craftsmen, winnowed woven craftsmen, and there is a village where the majority of the population works as crackers/opak craftsmen, so it is called the opakan village. Various kinds of potential can be increased in order to improve the welfare of the community as well. Improvements are needed regarding human resources and skills that the Wajak District community needs, especially Sumberputih Village as a future service plan. Human Resources training and development are very important things to be considered in the management of an organization, both in the form of institutions and companies. HR training and development activities have objectives that are realized in a series of systematic structured activities. The objectives and benefits of this training and human resource development are useful for increasing the potential of the community based on the field of business/community work as well as improving the skills/skills of the community related to work so that the quality of performance increases. In addition, so that individuals can be competent in dealing with certain situations that can occur in the world of work.

Improving the quality of human resources or workforce is important to do [1]. This can also be done through the distribution of more selective scholarship programs by the government, both for study programs at home and abroad, taking into account the needs of the local area. If an area has potential in the education sector, the scholarship program is directed to studies related to education, so that improving the quality of human resources will be beneficial in the development of local education and science and technology [2]. With qualified and educated human resources, it is hoped that the potential resources of Wajak District can be optimized for the welfare of the community. Therefore, in order to assist with market surveys, explore and find out the educational conditions of the community's families in Sumberputih Village, Wajak District, Malang Regency, East Java.

This market survey assistance was carried out to explore and determine the condition of interest in education in Wajak District, East Java, Indonesia in order to make the right marketing strategy to gain public interest so as to improve the quality of education and the economy in Wajak District, especially Sumberputih Village. This service is expected to provide benefits for various parties, especially for the people of Sumberputih Village.

There are many analyzes in statistical science that can be used to solve problems [3]; [4]; [5]. One of them is sentiment analysis.Sentiment analysis aims to obtain opinions from users on a platform. There have been many previous studies that have used this analysis, including [6]conducted a study entitled "Analysis of Sentiment of Teacher Room Applications on Twitter Using Classification Algorithms" and [7]conducted a study entitled "Analysis of the Pros and Cons of Indonesian Society Sentiments regarding the COVID-19 Vaccine on Twitter Social Media". In this study, sentiment analysis was used to determine public opinion regarding the establishment of an IT Boarding school.

2. LITERATURE REVIEW

2.1 Sentiment Analysis

Sentiment analysis is a method used to identify how a sentiment is expressed using text and how that sentiment can be categorized as positive sentiment or negative sentiment. The results of the prototype system achieve high precision in finding sentiments on web pages and news articles[8].

According to [9]sentiment analysis or opinion mining refers to a broad field of natural language processing, computational linguistics and text mining which has the aim of analyzing opinions, sentiments, evaluations, attitudes, judgments and emotions of a person whether the speaker or writer is pleased with a topic, product, service, organization, individual or activity.

Sentiment analysis aims to group the text into sentences or documents which then determine whether the opinion or opinion is positive, neutral, negative or otherwise [10]. Sentiment analysis is very useful in a process of analyzing opinions or comments and then processing them into something more meaningful [11].

Sentiment is an opinion or view that is based on an exaggerated feeling of something (contrary to thought considerations). Sentiment is found in statements or sentences that have an opinion. Sentiment is used to determine the feeling given to a topic or object [12].

Sentiment analysis can be interpreted as a method used to extract opinion data, understand and process textual data automatically to see the sentiment contained in a sentence whether it has a positive or negative opinion. Sentiment [13] classes are divided into very positive, positive, neutral, negative, and very negative classes. so that users can read and choose to read opinions as desired [14]. There are several names for this study, namely sentiment analysis, opinion mining, opinion extraction, sentiment mining, all of which are now under the realm of sentiment analysis or opinion mining [15].In the industrial world the words “sentiment analysis” and “opinion mining” are often used. The word “sentiment analysis” first appeared in 2003 by Dave Lawrence and Pennock [7].

Sentiment analysis is part of the field of data mining science. The purpose of sentiment analysis is to analyze a person's attitudes, emotions, opinions, and evaluations of certain products, services, or activities [7]. Sentiment Analysis can be distinguished based on the data source, several levels that are often used in Sentiment Analysis research are Sentiment Analysis at the document level and Sentiment Analysis at the sentence level [16]. The following are the types of sentiment analysis:

a. Fine-Grained Sentiment Analysis: is the most commonly used type, the analysis carried out only focuses on the level of opinion polarity, namely by classifying opinions into several categories such as positive and negative.

b. Intent Sentiment Analysis: this type aims to identify and dig deeper than existing opinions, usually used to find out whether what is meant is criticism, suggestions, questions, or awards for certain products.

Aspect-Based Sentiment Analysis: this type focuses on more specific elements, this type allows to relate specific sentiments to various aspects of the desired product.

2.1.1 Data Crawl

Crawling data is a process used to collect user opinions from various websites or certain sites that contain user opinions based on the scope of the proposed research. The process of crawling this data can use the API that is usually provided by the site in question. The data taken is tweet data contained in Twitter using a connection to access the Twitter API. Accessing Twitter tweet data requires access rights to be able to access tweet data in the form of consumer key, consumer secret, access token, and access token secret [17].

2.1.2 Case Folding

Case folding is the process of uniforming all text data into a certain form. Uniform text is done by changing all letters to lowercase (lowercase) and removing characters in the text that do not allow analysis, such as punctuation marks, mention usernames (@), hashtags (#), and web addresses. The case folding process can also overcome the excessive repetition of letters for each word in the text[18].

2.1.3Tokenizing

Tokenization is the process of breaking text into chunks of words called tokens. The token generated can be a phrase or a single word, where each token will be a variable that has a certain value. Tokenizing is the process of cutting the input string based on each word that composes it. Splitting sentences into single words is done by scanning sentences with white space delimiters (spaces, tabs, and newlines) [16].

2.1.4Stemming

Stemming is the process of mapping and parsing the form of a word into its basic word form with certain rules [19]. This process is done by changing the affixed words into basic words.

2.1.5Spelling Normalization

Spelling normalization is a word standardization process that aims to improve the spelling of each word in the text, namely changing abbreviations or words that are not standardized into standard word forms. This process refers to the KBBI (Big Indonesian Dictionary) as a reference for the standard words used[20].

2.1.6Stopwords Disposal

Stopword removal is the process of reducing redundant variables. In the case of data in the form of documents, words that are considered irrelevant are removed and retain the important ones so that the classification process is more effective and accurate [21].

2.1.7Labeling

Data labeling is the process of assigning positive, negative, and neutral labels to text data. This process can be done manually and automatically, with the help of machine learning. Manual data labeling is not effective because it is very subjective. Automatic labeling can use a sentiment lexicon dataset or an emotion dictionary, this method is good for large amounts of text so as to minimize document labeling errors. Labeling is carried out to form training data, and labeling is also carried out on test data to calculate the accuracy of the sentiment results obtained automatically [22].

2.2Boarding School

Boarding school is a word in English which consists of two words, namely boarding and school, boarding has the meaning of riding while school means school, then the word is absorbed into Indonesian into a boarding school. According to[23], "dormitory" is a boarding house for students, employees and so on, while in a dormitory, namely living together in a building or complex. Boarding school is an educational institution where students not only learn, but they live and live together in the institution.

2.3Marketing

Marketing is an organizational function and a set of processes for creating, communicating, and delivering value to customers and for managing customer relationships in ways that benefit the organization and the customer. Marketing as a human activity that is directed to meet and satisfy needs and wants through the exchange process [24]. Marketing (marketing) is a human activity that is directed to meet needs and wants through an exchange process [25]. Strategy is essentially planning, planning and management to achieve goals [26].However, to achieve this goal, the strategy does not function as a journey that only shows the direction of the road, but must be able to show how the operational tactics are [27].

3. METHODOLOGY

The sentiment analysis method in this study uses data from Twitter which was taken on May 18, 2022 and obtained by crawling the data by taking tweets using the keyword, namely boarding school. A total of 1071 tweets used in this study, for the language used to search for issues on Twitter using Indonesian.Crawling data using R software by utilizing the " rtweet " package to retrieve data automatically from twitter. The steps taken in the research using sentiment analysis are crawling data or collecting data on Twitter. Next, do case folding or uniform text data. The steps taken in the research using sentiment analysis are data crawling or data collection on Twitter. Next, do case folding or uniform text data. According to [6], the steps taken before the dataset is entered into the model are (1) transform case to change the letters in the text to all lowercase letters, (2) remove http to remove links contained in tweets, (3) remove @ to remove mentions in tweets, (4) tokenize to separate words in each sentence into separate words while removing unnecessary characters, (5) remove stopwards to remove words that can be ignored. In this study, the next step is tokenizing or breaking sentences into words.Then do stemming or change the affixed words into basic words. Perform spelling normalization or standardization of important words. Remove stopwords or remove meaningless word sets. Followed by labeling into 3 sentiment classes using the sentiment lexicon dataset. It ends by determining the indicators and variables that are formed based on the results of sentiment analysis.

4. RESULTS AND DISCUSSION

4.1. Results of Classification of Community Sentiments Regarding Public Perceptions in Indonesia Regarding the *Boarding School Program* at an Islamic Boarding School

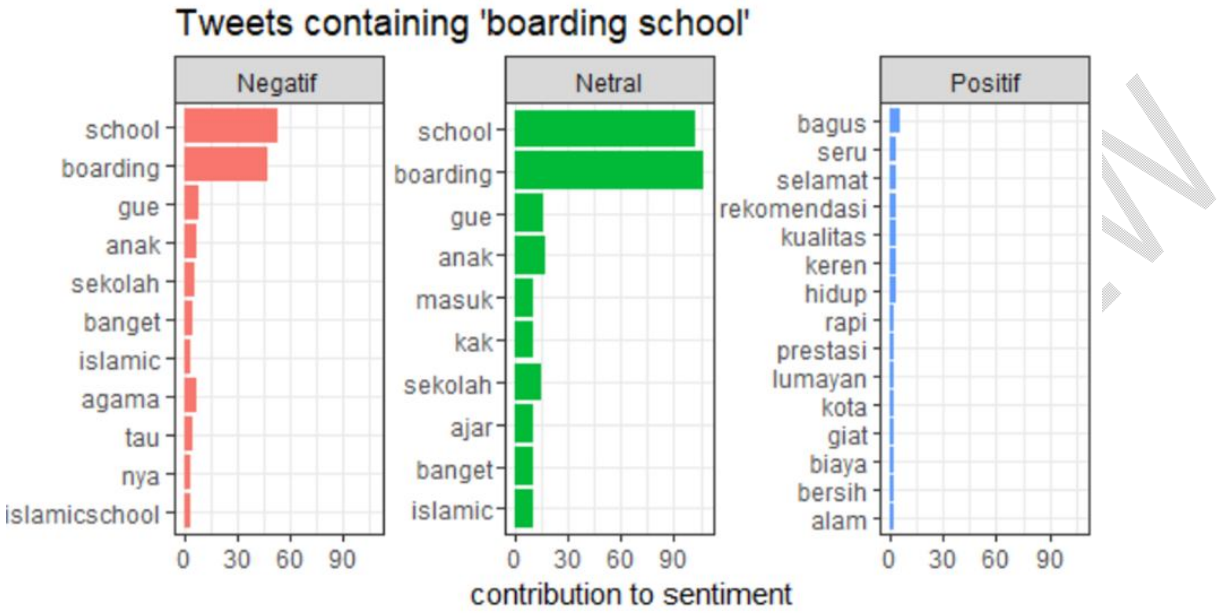

Fig. 1. Proportion of Sentiment Analysis

Figure 1 shows the results of public sentiment regarding boarding schools. Pink bars indicate negative sentences, green bars indicate neutral sentences and blue bars indicate positive sentences. Based on the graph, it can be seen that society has a neutral tendency regarding public perceptions in Indonesia regarding the boarding school program at an Islamic boarding school. From the graph, it can also be seen that the word school occupies the top position on the bar which shows negative sentiment followed by the words boarding and me. The word boarding occupies the top position on the bar which shows neutral sentiment with a total of more than 90 data, then followed by the words school and me. And in positive sentiment data, the word good occupies the highest position with less than 6 data.

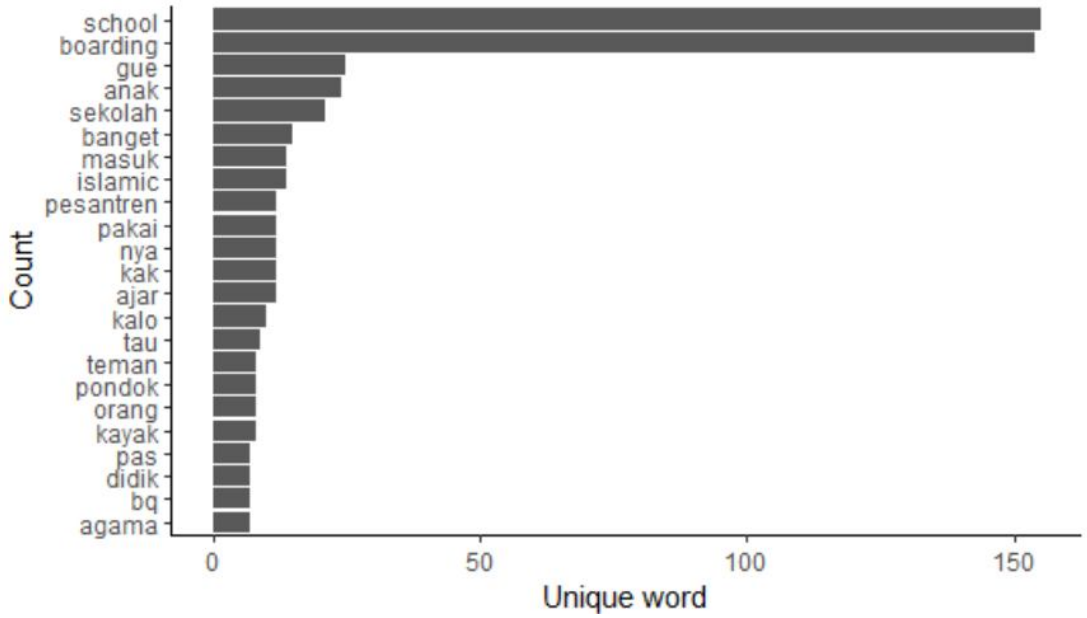

Fig. 2. The results of the distribution of one word on the data

Figure 2 shows the results of the distribution of words in all data without dividing into negative, neutral, or positive sentiments. From these results, it was found that the word school occupies the first position with a frequency of 155 times, followed by the word boarding with a frequency of 154 times and the word I with a frequency of 25 times.

4.2. Identification Of Variables Based on Sentiment Analysis Results

Based on the analysis that has been done, several variables are obtained. These formed variables are mining variables which will then be used in the analysis. The variables obtained are as follows:

1) Marketing Communication Strategy

This variable is inferred from the words obtained from mining results through Twitter social media such as familiar words, saying, connecting, famous, selection, spirit, my friend, language, member, listening, and social media.

2) Marketing mix

After mining through Twitter social media, it can be concluded that the next variable is the marketing mix. This variable is concluded based on mining results words such as facility, form, quality, nuance, recommendation, global, scholarship, info, management, marketing, boarding house, program, and plus.

3) Interest to learn

This variable is concluded based on the results of mining words from twitter such as students, school, boarding, cottage, quality, curriculum, students, adab, technology, assignments, and foundations.

5. CONCLUSION

In this study, a sentiment analysis has been carried out regarding public perceptions in Indonesia regarding the boarding school program at an Islamic boarding school as conveyed through social media, namely Twitter. From these results, it can be concluded that the trend in the use of words in positive sentiment data which is most often used is good with a frequency of 6 times. The trend of using words in neutral sentiment data is the word boarding with a frequency of 105 times. Meanwhile, the trend of word usage in negative sentimental data which is most often used is the word school with a frequency of less than 60 times.

Based on the analysis that has been done, several variables are obtained, namely marketing communication strategy, marketing mix, and interest in learning. These variables were inferred from the words obtained from word mining through the social media twitter.

REFERENCES

[1] Fernandes, A. A. R., Hutahayan, B., Arisoesilaningsih, E., Yanti, I., Astuti, A. B., &Amaliana, L. (2019, June). Comparison of curve estimation of the smoothing spline nonparametric function path based on PLS and PWLS in various levels of heteroscedasticity. In IOP Conference Series: Materials Science and Engineering (Vol. 546, No. 5, p. 052024). IOP Publishing.

[2]Raharjo, K., Nurjannah, N., Solimun, S., & Fernandes, A. A. R. (2018). The influence of organizational culture and job design on job commitment and human resource performance. Journal of Organizational Change Management.

[3]Fernandes, A. A. R. (2017). Moderating effects orientation and innovation strategy on the effect of uncertainty on the performance of business environment. International Journal of Law and Management.

[4]Purbawangsa, I. B. A., Solimun, S., Fernandes, A. A. R., &Rahayu, S. M. (2019). Corporate governance, corporate profitability toward corporate social responsibility disclosure and corporate value (comparative study in Indonesia, China and India stock exchange in 2013-2016). Social Responsibility Journal, 16(7), 983-999.

[5]Fernandes, A. A. R., Budiantara, I. N. I., Otok, B. W., &Suhartono. (2014). Reproducing Kernel Hilbert space for penalized regression multi-predictors: Case in longitudinal data. International Journal of Mathematical Analysis, 8(40), 1951-1961.

[6] Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., &Gata, W. (2020). ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI. JurnalTeknoinfo, 14(2), 115. https://doi.org/10.33365/jti.v14i2.679

[7] Rachman, F. F., &Pramana, S. (2020). AnalisisSentimen Pro dan Kontra Masyarakat Indonesia tentangVaksin COVID-19 pada Media Sosial Twitter. Dalam Health Information Management Journal ISSN (Vol. 8, Issue 2).

[8]Nasukawa, T. and Yi, J. (2003) Sentiment Analysis: Capturing Favorability Using Natural Language Processing. Proceedings of the 2nd International Conference on Knowledge Capture, Florida, 23-25 October 2003, 70-77. http://dx.doi.org/10.1145/945645.945658

[9]Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.

[10]Fernandes, A. A. R., &Taba, I. M. (2018). Welding technology as the moderation variable in the relationships between government policy and quality of human resources and workforce competitiveness. Journal of Science and Technology Policy Management.
[11]Benny Hutahayan, A. A. R. F. S. N. (2020). Comparison of use of Linkage in Integrated Cluster with Discriminal Analysis Approach. International Journal of Advanced Science and Technology, 29(3), 5654 - 5668. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/6191
[12]Gunawan, D., Dwiza, R., Ardiansyah, D., Akba, F., &Alfariz, S. (2020). KomparasiAlgoritma Support Vector Machine Dan Na{\"\i}ve Bayes DenganAlgoritmaGenetika Pada AnalisisSentimen Calon GubernurJabar 2018-2023. J. Tek. Komput. AMIK BSI, 6(1).
[13]Sari, F. V., & Wibowo, A. (2019). AnalisisSentimenPelangganToko Online Jd. Id MenggunakanMetode Na{\"\i}ve Bayes Classifier BerbasisKonversi Ikon Emosi. Simetris: Jurnal Teknik Mesin, Elektro Dan IlmuKomputer, 10(2), 681–686.
[14]Buntoro, G. A., Adji, T. B., &Purnamasari, A. E. (2016). Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute. International Journal of Computer Applications, 975(136.2), 23–29.
[15]Fernandes, S., & Rinaldo, A. A. R. A. A. (2018). The mediating effect of service quality and organizational commitment on the effect of management process alignment on higher education performance in Makassar, Indonesia. Journal of Organizational Change Management.
[16]Fink, C. R., Chou, D. S., Kopecky, J. J., &Llorens, A. J. (2011). Coarse- and Fine-Grained Sentiment Analysis of Social Media Text. Johns Hopkins Apl Technical Digest, 30(1), 22–30.
[17] Sari, D. I., Wati, Y. F., &Widiastuti, W. W. (2020). ANALISIS SENTIMEN DAN KLASIFIKASI TWEETS BERBAHASA INDONESIA TERHADAP TRANSPORTASI UMUM MRT JAKARTA MENGGUNAKAN NAÏVE BAYES CLASSIFIER. JurnalIlmiahInformatikaKomputer, 25(1), 64–75.
[18]Mustaqhfiri, M., Abidin, Z., &Kusumawati, R. (2011). Peringkasanteksotomatisberitaberbahasa Indonesia menggunakanmetode Maximum Marginal Relevance. Matics.
[19]Tala, F. (2003). A study of stemming effects on information retrieval in Bahasa Indonesia.
[20]HONG, Y. (2018). Spelling Normalization of English Student Writings.
[21] Berry, M. W., & Kogan, J. (2010). Text mining: applications and theory. John Wiley \& Sons.
[22]Ardiani, L., Sujaini, H., &Tursina, T. (n.d.). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. JUSTIN (JurnalSistem Dan TeknologiInformasi), 8(2), 183–190.
[23]Khalifah. (2013). Character Education in the Boarding School System. Thesis of the Faculty of Tarbiyah and Teacher Training. UIN SunanKaliaga Yogyakarta Press.
[24] SofjanAssauri, Marketing Management (Jakarta: Rajawali Pers, 2013), p.5.Sumardi, S., & Fernandes, A. A. R. (2020). The influence of quality management on organization performance: service quality and product characteristics as a medium. Property Management, 38(3), 383-403.
[25]Fernandes, A. A. R. (2017). Investigation of instrument validity: Investigate the consistency between criterion and unidimensional in instrument validity (case study in management research). International journal of law and management.
[26]Sumardi, S., & Fernandes, A. A. R. (2020). The influence of quality management on organization performance: service quality and product characteristics as a medium. Property Management, 38(3), 383-403.
[27]OnongUchjanaEfendy, Communication Science Theory and Practice, (Bandung: RemajaKarya, 1984), p. 35.