



UNLOCKING CUSTOMER SENTIMENT INSIGHTS WITH AZURE SENTIMENT ANALYSIS: A COMPREHENSIVE REVIEW AND ANALYSIS

Cosmina Mihaela Roșca¹

Andy Valentin Ariciu¹

¹ Petroleum-Gas University of Ploiesti, Romania
email: cosmina.rosca@upg-ploiesti.ro

DOI: 10.51865/JPGT.2023.01.15

ABSTRACT

The paper analyzes the accuracy of the Azure Sentiment Analysis service for five languages, namely Romanian, French, Italian, Portuguese, and Spanish. The study generated 300 texts for each language type expressing positive and negative sentiments with varying lengths (less than 100 characters, between 100 and 250 characters, and more than 250 characters). The Azure Sentiment Analysis Review custom-made application was developed using C# language with .NET Framework and Entity Framework for the Microsoft SQL database, and it is used to make a request to the Azure Sentiment Service, and the response sets the label into the database. The expected and Azure labels for each type of analyzed text were described as well. The accuracy of sentiment recognition for different languages and text lengths is presented in the form of statistics, with the global accuracy of the service being 81.8%. The challenges of accurately classifying the sentiment of short and long texts were highlighted. The results suggest that texts of moderate length are easier to classify.

Keywords: Azure Sentiment Analysis, database, algorithm

INTRODUCTION

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language. NLP involves the development of algorithms and computational models that can analyze and understand natural language text or speech, and perform tasks such as language translation, sentiment analysis, chatbot development, text summarization, and more. The goal of NLP is to bridge the gap between human communication and machine understanding, making it possible for humans to interact with computers in a more intuitive and natural way.

NLP sentiment analysis is a subfield of natural language processing that involves analyzing the sentiment or emotion expressed in a piece of text or speech. It is also known as opinion mining, sentiment mining, or emotion Artificial Intelligence (AI). The goal of sentiment analysis is to determine whether a given piece of text expresses a positive, negative, or neutral sentiment. This is typically achieved through the use of machine learning algorithms that are trained on a large corpus of text data annotated with sentiment labels.

Sentiment analysis has a wide range of applications, including:

- **Brand monitoring:** Companies can use sentiment analysis to monitor social media and other online channels to gauge customer sentiment towards their brand and products [1].
- **Customer service:** Sentiment analysis can be used to automatically route customer complaints to the appropriate support agent based on the sentiment expressed in the complaint [2].
- **Product development:** Companies can use sentiment analysis to gain insights into customer opinions about their products and identify areas for improvement [3].
- **Political analysis:** Sentiment analysis can be used to analyze public opinion about political candidates, issues, and policies [4, 5].

Sentiment analysis is a powerful tool that enables businesses and organizations to gain valuable insights into the emotions and attitudes expressed in large volumes of text data.

Azure Sentiment Analysis is a service provided by Microsoft Azure that allows users to perform sentiment analysis on text data using machine learning algorithms. This service uses NLP techniques to analyze the sentiment expressed in a given text, such as a tweet, product review, or customer feedback. The service supports several programming languages, including Python, C#, and Java.

Azure Sentiment Analysis provides three levels of sentiment analysis [6]:

- **Document-level sentiment analysis:** This level of analysis provides a score between 0 and 1 that indicates the overall sentiment of a document, with '0' being the most negative and '1' being the most positive.
- **Sentence-level sentiment analysis:** This level of analysis provides sentiment scores for individual sentences within a document.
- **Aspect-level sentiment analysis:** This level of analysis identifies specific aspects or features within a document and provides sentiment scores for each aspect.

Azure Sentiment Analysis can be used in a variety of applications, such as social media monitoring, customer feedback analysis, and product reviews analysis. By leveraging this service, businesses can gain valuable insights into customer sentiment and improve their products and services accordingly.

In this paper, a review of the Azure Sentiment Analysis service is conducted using the three levels. The paper aims to calculate the degree of confidence in the service.

The accuracy of Azure Sentiment Analysis can vary depending on the quality and complexity of the input text data, as well as the specific use case and application. However, in general, Azure Sentiment Analysis has been found to have a high level of accuracy in identifying the overall sentiment and emotions expressed in text data.

Microsoft has reported that Azure Sentiment Analysis achieves a 90% accuracy rate in identifying positive, negative, and neutral sentiments in English text data. Additionally, the service has been trained on a large volume of data and is continuously updated and improved to enhance its accuracy and performance.

In terms of trust level, Azure Sentiment Analysis is a reliable and trusted service that is widely used by businesses and organizations of all sizes. Microsoft has implemented robust security measures to ensure the confidentiality, integrity, and availability of customer data, and the service is compliant with various industry standards and regulations, including GDPR, HIPAA, and ISO 27001.

Azure Sentiment Analysis supports a wide range of languages including English, Spanish, French, German, Italian, Portuguese, Dutch, Swedish, Norwegian, Danish, Finnish, Chinese (Simplified and Traditional), Japanese, Korean, Arabic, Hindi, and Russian.

However, like any machine learning-based tool, Azure Sentiment Analysis is not perfect and may not always accurately identify the sentiment and emotions expressed in the text. It is important to understand the limitations of the tool and to use it in conjunction with other data sources and analytics tools to gain a more comprehensive understanding of customer sentiment and behavior.

The Azure Sentiment Analysis Service offers several language features that can be purchased in different tiers based on the number of text records used. Summarization features are counted based on the length of both the input document and the corresponding output summary, while other language features are based on the length of the input document. The service offers free usage of up to 5,000 text records per month for features such as sentiment analysis, key phrase extraction, language detection, question answering, named entity recognition, and conversational language understanding. However, some features such as question answering require additional resources such as Azure Cognitive Search.

The service also offers a preview of new features such as conversation summarization and document summarization. Prices for the service range from \$1 to \$5 per 1,000 text records, with additional costs for features such as conversation PII and text analytics for health. Overall, the Azure Sentiment Analysis Service offers a flexible pricing model and a range of language features for businesses and developers to choose from based on their specific needs [7].

Overall, Azure Sentiment Analysis is a powerful and easy-to-use tool that can help businesses gain valuable insights into customer sentiment and emotions. When it is used in conjunction with other data sources and analytics tools, it can be a powerful asset in understanding customer preferences and behavior, and in driving business growth and success.

AZURE SENTIMENT ANALYSIS REVIEW

This paper focuses on the accuracy of the Azure Sentiment Analysis service for five languages: Romanian, French, Italian, Spanish, and Portuguese. For each language, 100 texts were generated with the following types of phrases:

- Phrases with less than 100 characters;
- Phrases with a length between 100 and 250 characters;
- Phrases with more than 250 characters.

For each set of phrases, 50 were constructed to express positive ideas, and 50 for negative ideas. Knowing the expected result that the Azure Sentiment Analysis service should return, the service is called and the number of labels that match what the service returned is computed.

The application, named Azure Sentiment Analysis Review (ASAR), is developed using C# language with .NET Framework and Entity Framework for the Microsoft SQL database. The ASAR makes a request with 10 texts to the Azure Sentiment Service and the responses will set the label into the database.

The generated texts are associated with the psychology of children in relation to their parents. Table 1 presents an example of each type of analyzed text, translated into English. The original text is in the Romanian language. Thus, two texts with less than 100 characters that express positive and negative content are presented. Also, two texts with content of 100-250 characters and two with over 250 characters are described. For each category example, one of the texts is positive and the other is negative. Table 1 depicts the expected label associated with the idea expressed by the sentences and the Azure label represents the label generated by the Sentiment service. The presented examples show that the service returned the same label as the expected one, but it is not mandatory that the service returns as expected. All texts for the five analyzed languages are equivalent and express the same ideas in order to keep the same metric for the accuracy factor computed for the Azure Sentiment Service.

The Service returns three labels:

Positive - the text may contain positive words, phrases, or expressions that convey a positive sentiment.

Negative - the text may contain negative words, phrases, or expressions that convey a negative sentiment.

Neutral - this label indicates that the sentiment of the text is neutral. The text may not contain any words, phrases, or expressions that convey a strong sentiment either way. Alternatively, it could contain equal parts positive and negative sentiment that cancel each other out, resulting in neutral sentiment.

Sentiment analysis is not a perfect science and there can be nuances to language and context that make it difficult to accurately classify sentiments in all cases. However, these labels can still provide a useful high-level overview of the general sentiment conveyed by a piece of text.

In Azure Sentiment Analysis, the neutral label represents sentences that do not express a positive or negative sentiment. These sentences are considered to be objective and have a neutral tone, meaning they do not contain any emotionally charged language or opinions. Neutral sentences are often used to provide factual information, describe a situation or object, or provide a neutral opinion. In the context of sentiment analysis, neutral labels are important because they help distinguish between positive and negative sentiments and provide a more nuanced understanding of the text being analyzed.

The accuracy was computed as a percentage of the number of sentences that were classified correctly by the service, compared to the target sentiment. In this research, it is stated that neutral labels do not contribute to increased data accuracy, as they are considered incorrect classifications of sentiment.

Table 1. Analyzed texts examples translated into English

Sentences	Expected label	Azure label
Love your child with all your heart and be present for them in every important moment.	Positive	Positive
To be a neglectful parent, ignoring the needs and problems of your child.	Negative	Negative
To be tolerant, accepting the unique differences and personalities of your child and teaching them to respect and accept others.	Positive	Positive
An abusive parent can be defined by behaviors and actions that harm their children, such as physical and emotional violence, neglecting their children's needs, excessive control, and domination.	Negative	Negative
Parents who encourage their children regardless of the situation are an essential element in their development. These parents have a positive attitude and a supportive spirit that helps children feel appreciated and valued, regardless of the situations they face. This can have a powerful impact on a child's development and happiness, and can contribute to the formation of a strong and resilient character.	Positive	Positive
Violent parents are characterized by abusive behaviors such as physical and verbal aggression, which have a profoundly negative impact on the child. This type of behavior can have serious consequences on the emotional and physical well-being of the child, such as anxiety, phobia, depression, and behavioral problems.	Negative	Negative

Table 2 presents an example where a positive sentence is classified as neutral, and a negative sentence is classified as neutral as well.

Table 2. Analyzed texts examples translated into English wrong classified

Sentences	Expected label	Azure label
Provide your child with constructive discipline and explain the rules and their consequences.	Positive	Neutral
Not giving emotional and physical neglect to the child.	Negative	Neutral
To be a supporter of your child, encouraging them in their achievements and failures and helping them learn from mistakes.	Positive	Negative
Being a critical parent undermines the child's confidence and self-esteem.	Negative	Positive

Additionally, Table 2 shows a positive sentence classified as negative and a negative sentence classified as positive.

Figure 1 illustrates the block diagram of the ASAR algorithm. Thus, it can be seen that individual text has been generated in order to populate the database with samples. Next, if there are records without an Azure label set, then an Azure Portal request is made for the Azure Sentimental analysis. The service will send the response and the Azure label is set into the database. Finally, the sentences are processed and then different stats are computed.

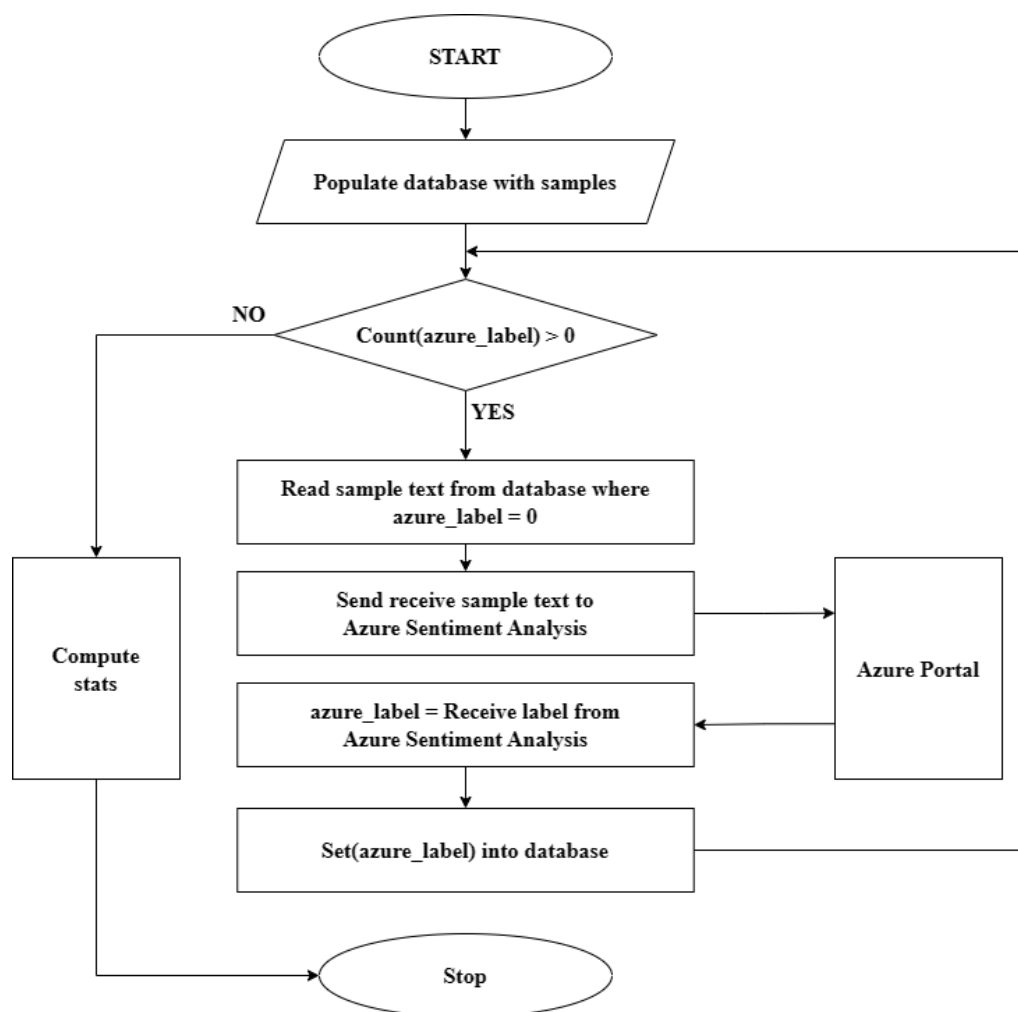


Figure 1. The block diagram of the ASAR algorithm

Figure 2 shows a statistic regarding the accuracy of the service for all five analyzed languages. Thus, the global accuracy of the service was 81.8% regardless of the language in which the analysis is performed. The service recorded a similar accuracy, regardless of the length of the text, noting a slight decrease for texts longer than 250 characters. For sentences with less than 100 characters, the service accuracy was 82.4% (figure 3a), for sentences with 100-250 characters, the obtained accuracy was 82.6% (figure 3b), and for texts with more than 250 characters, the accuracy was 80.2% (figure 3c).

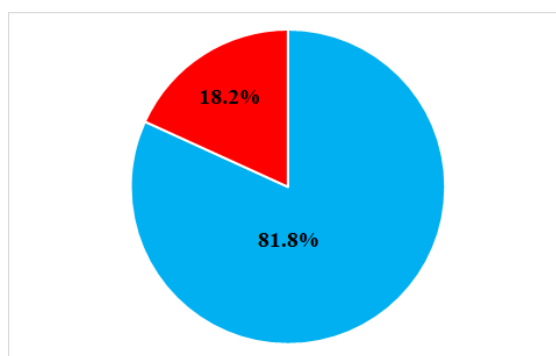


Figure 2. Global accuracy of the Azure Sentiment Analytics for the analyzed countries

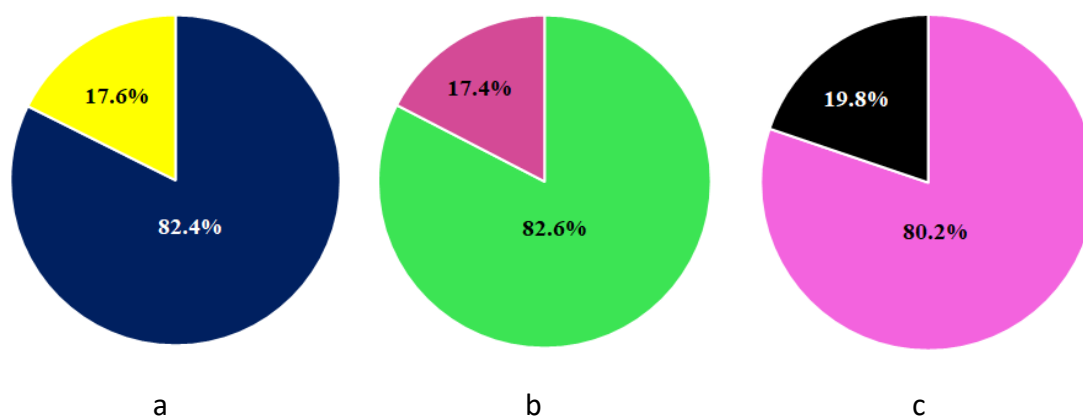


Figure 3. Statistics for Azure Sentiment Analytics: <100 characters - 82.4% accuracy (a); 100 – 250 characters – 82.6% accuracy (b); > 250 characters – 80.2% accuracy (c)

Figure 4 presents the statistics for Azure Sentiment Analytics for the analyzed languages: Romanian (blue marker), French (red marker), Italian (gray marker), Portuguese (yellow marker), and Spanish (green marker). The histogram is divided into three categories: sentences with less than 100 characters, sentences with characters between 100 and 250, and sentences with more than 250 characters. The statistics underscores that:

- For texts with less than 100 characters, the accuracy of sentiment recognition ranges from 76% to 91% among different languages. This suggests that short texts can be challenging to accurately classify in terms of sentiment;
- For texts between 100 and 250 characters, the accuracy of sentiment recognition improves, with a range of 83% to 93% among different languages. This suggests that texts of moderate length are easier to classify in terms of sentiment;
- For texts longer than 250 characters, the accuracy of sentiment recognition decreases slightly, with a range of 77% to 87% among different languages. This suggests that longer texts may become more complex, making it more difficult to accurately classify the sentiment.

The results highlight that there are variations in accuracy among different languages. For example, French has the highest accuracy for all three length categories, while Italian has the lowest accuracy. These variations may be due to differences in language structure, vocabulary, and syntax, among other factors.

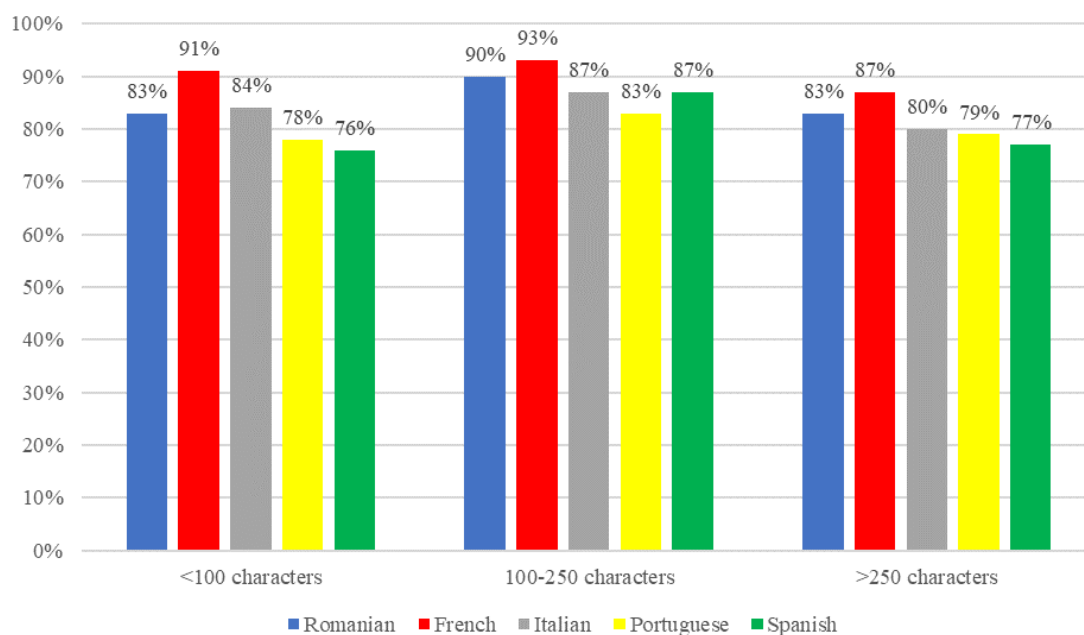


Figure 4. Statistics of Azure Sentiment Analytics for each country, grouped by categories: <100 characters; 100 – 250 characters; >100 characters

There are several factors that can influence the accuracy of sentiment recognition, including:

Language structure: Different languages have unique structures that can impact sentiment analysis. For instance, some languages may have more complex grammatical rules that make it more challenging to determine the sentiment of a text.

Vocabulary: The vocabulary used in different languages can also affect the accuracy of sentiment recognition. Certain languages may have a wider range of emotional words or more subtle nuances of meaning that can be difficult to capture accurately.

Syntax: The order of words in a sentence can also impact sentiment recognition. Some languages may place more emphasis on certain words or use different word orders to convey specific emotions.

While each of the five Romance languages (Romanian, French, Italian, Portuguese, and Spanish) has its own unique characteristics, they all share a complex inflectional system, a rich vocabulary influenced by various languages, and a flexible syntax with a typical Subject-Verb-Object (SVO) pattern, albeit with variations in word order depending on the context.

The statistics for sentiment analysis accuracy varies among the five Romance languages, with French and Italian generally having the highest accuracy rates among all sentence lengths, followed by Portuguese, Spanish, and Romanian, with lower overall accuracy rates, particularly for shorter sentences:

Romanian: For texts with less than 100 characters, the sentiment recognition accuracy is 83%, for texts between 100 and 250 characters, the accuracy increases to 90%, and for texts longer than 250 characters, the accuracy decreases slightly to 83%.

French: For texts with less than 100 characters, the sentiment recognition accuracy is 91%, for texts between 100 and 250 characters, the accuracy increases to 93%, and for texts longer than 250 characters, the accuracy decreases slightly to 87%.

Italian: For texts with less than 100 characters, the sentiment recognition accuracy is 84%, for texts between 100 and 250 characters, the accuracy increases to 87%, and for texts longer than 250 characters, the accuracy decreases to 80%.

Portuguese: For texts with less than 100 characters, the sentiment recognition accuracy is 78%, for texts between 100 and 250 characters, the accuracy increases to 83%, and for texts longer than 250 characters, the accuracy decreases slightly to 79%.

Spanish: For texts with less than 100 characters, the sentiment recognition accuracy is 76%, for texts between 100 and 250 characters, the accuracy increases to 87%, and for texts longer than 250 characters, the accuracy decreases slightly to 77%.

These statistics suggests that sentiment recognition accuracy varies among languages and text lengths. Additionally, it is important to consider factors such as language structure, vocabulary, and syntax when analyzing sentiment in text data.

Despite the variations in accuracy rates among the different languages, it is important to note that the Azure Sentiment Analytics Service still provides a relatively accurate assessment of sentiment in text data. With a minimum accuracy rate of 76% for short sentences and up to 93% for medium-length sentences among all languages, the service can provide useful insights for businesses and organizations looking to extract sentiment-related data from large amounts of text data.

The Azure Sentiment Analytics Service provides a useful tool for businesses and organizations looking to extract sentiment-related data from large amounts of text data among the five Romance languages, i.e. Romanian, French, Italian, Portuguese, and Spanish. While the accuracy rates vary among different languages and sentence lengths, the service still provides a relatively accurate assessment of sentiment in text data. By incorporating more contextual information and using a combination of machine learning algorithms and human input, the service can be further improved to enhance its accuracy rates and provide even more useful insights for businesses and organizations.

CONCLUSION

After analyzing the statistics for sentiment analysis accuracy among the five languages (Romanian, French, Italian, Portuguese, and Spanish), it is obvious that the Azure Sentiment Analytics Service performs relatively well, although with some variations among the different languages and sentence lengths. Overall, the accuracy of the Azure Sentiment Analytics Service is dependent on several factors, including language structure, vocabulary, and syntax, as well as the length and complexity of the input sentences. For instance, languages with complex grammar systems, such as Romanian and Portuguese, tend to have lower accuracy rates compared to languages with simpler grammar systems, such as French and Spanish. Additionally, the accuracy of the service tends to increase as the length of the input sentence increases, with longer sentences providing more context for accurate sentiment analysis.



Languages with rich and diverse vocabularies, such as French and Italian, tend to have higher accuracy rates due to the ability of the service to identify and comprehend more nuanced and complex expressions. Additionally, the flexibility of word order and sentence structure in certain languages, such as French and Italian, also provides more context for the service to perform accurate sentiment analysis.

REFERENCES

- [1] Rahmatulloh A., Shofa R.N., Darmawan I., Ardiansah, Sentiment Analysis of Ojek Online User Satisfaction Based on the Naïve Bayes and Net Brand Reputation Method, 9th International Conference on Information and Communication Technology (ICoICT), IEEE, 2021. doi: 10.1109/ICoICT52021.2021.9527466
- [2] Yunita S., Amaliah Y., Suprianto, Indriani A., Fadlan M., Muhammad, Sentiment Analysis Of Practice Service's Questionnaire Using Naïve Bayes Classifier Method, 3rd International Conference on Cybernetics and Intelligent System (ICORIS), IEEE, 2021. doi: 10.1109/ICORIS52787.2021.9649490
- [3] Uma R., Aafreen Sana H., Jawahar P., Rishitha B.V., Support Vector Machine and Convolutional Neural Network Approach to Customer Review Sentiment Analysis, 1st International Conference on Computational Science and Technology (ICCST), IEEE, 2022. doi: 10.1109/ICCST55948.2022.10040381
- [4] Kusumawardani R.P., Maulidani M.W., Aspect-level Sentiment Analysis for Social Media Data in the Political Domain using Hierarchical Attention and Position Embeddings, International Conference on Data Science and Its Applications (ICoDSA), IEEE, 2020. doi: 10.1109/ICoDSA50139.2020.9212883
- [5] Pota M., Esposito M., Palomino M.A., Masala G.L., A Subword-Based Deep Learning Approach for Sentiment Analysis of Political Tweets, 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE, 2018. doi: 10.1109/WAINA.2018.00162
- [6] <https://learn.microsoft.com/en-us/rest/api/cognitiveservices-textanalytics/3.0/sentiment/sentiment?tabs=HTTP>
- [7] <https://azure.microsoft.com/en-us/pricing/details/cognitive-services/language-service>

Received: April 2023; Revised: May 2023; Accepted: May 2023; Published: May 2023