Method for Sentiment Analysis of Ukrainian-Language Reviews in E-Commerce Using RoBERTa Neural Network

Olha Zalutska[1], Maryna Molchanova[1], Olena Sobko[1], Olexander Mazurets[1], Oleksandr Pasichnyk[1], Olexander Barmak[1], Iurii Krak[2,3]

[1] *Khmelnytskyi National University, Khmelnytskyi, 11, Instytutska str., 29016, Ukraine*
[2] *Taras Shevchenko National University of Kyiv, Kyiv, 64/13, Volodymyrska str., 01601, Ukraine*
[3] *Glushkov Institute of Cybernetics of NAS of Ukraine, Kyiv, 40, Glushkov ave., 03187, Ukraine*

Abstract

The paper is devoted to the development of a method for sentiment analysis of Ukrainian-language reviews, which will be able to perform binary classification of the tone of e-commerce reviews in everyday Ukrainian. It is proposed to use a modification of the BERT neural network architecture – RoBERTa, which has shown better results in the tasks of classifying short text messages.

In developing the method, were researched: the formation of a labeled dataset for training the neural network, selection and tuning of a neural network classifier, and construction of a semantic model of the language. The developed method allows performing binary classification based on the emotional coloring of reviews written not only in literary Ukrainian but also containing lexical and grammatical elements of different languages and specialized slang, without observing the literary language norms. With bilingual data, the accuracy rate was 92%, which is quite high given the specifics of the language. Further research is aimed at implementing this classifier to evaluate the work of managers when communicating with online store customers, implementing marketing feedback models, and improving the efficiency of classifiers that can work with multiple languages simultaneously.

Keywords

BERT, RoBERTa, sentiment analysis, emotion detection, sentiment classification, reviews in e-commerce, Ukrainian-language, neural network

1. Introduction and literature review

In recent years, the analysis of the emotional tone of text messages [1-4] as a basis for determining their information value [5] and the identification of important user sentiments [6-8], which is part of natural language processing, has attracted the attention of scientists. This is due to the growth of possible areas of application. Text message sentiment analysis is a method of extracting and recognizing user ratings of products and models and has various approaches using machine learning algorithms to classify the emotions behind the text [1]. For example, sentiment analysis of tweets to understand people's perception of certain news, evaluation of human-robot interaction, formation of a recommendation system for choosing products, etc [9, 10].

The problem of determining the emotional tone of text information is currently a widely studied area with numerous approaches [11, 12]. In [9], a framework called the "bidirectional emotional recurrent unit" was proposed by the authors to analyze conversational sentiment. In the proposed system, a generalized neural tensor block is used, followed by a two-channel classifier designed to perform contextual composition and sentiment classification, respectively.

COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems, April 20–21, 2023, Kharkiv, Ukraine
EMAIL: zalutska.olha@gmail.com (O. Zalutska); m.o.molchanova@gmail.com (M. Molchanova); olenasobko.ua@gmail.com (O. Sobko); exe.chong@gmail.com (O. Mazurets); o.a.pasichnyk@gmail.com (O. Pasichnyk); alexander.barmak@gmail.com (O. Barmak); yuri.krak@gmail.com (I. Krak)
ORCID: 0000-0003-1242-3548 (O. Zalutska); 0000-0001-9810-936X (M. Molchanova); 0000-0001-5371-5788 (O. Sobko); 0000-0002-8900-0650 (O. Mazurets); 0000-0002-8760-4688 (O. Pasichnyk); 0000-0003-0739-9678 (O. Barmak); 0000-0002-8043-0785 (I. Krak)

The authors categorize a large number of recent articles and illustrate the latest trends in sentiment analysis research and related areas [13].

The authors [14] found that the combination of machine learning and a lexicon-based method can achieve higher accuracy than any type of sentiment analysis. The authors used a variety of sentiment analysis, machine learning methods, and dictionary-based sentiment analysis to test and compare the effectiveness of user behavior research.

Taking into account the problems of humanity that have arisen recently, such as the coronavirus pandemic, researchers in their works [15-18] analyze the attitude of social network users to the pandemic. Researchers in [19] proposed a dictionary-based method for analyzing sentiment on Twitter, which gave relevant results on sentiment about AstraZeneca/Oxford, Moderna, and Pfizer/BioNTech COVID-19 vaccines for 4 months. Instead, [20] proposes to use TextBlob with TF-IDF vectorization and LinearSVC classification model to assess sentiment, which resulted in an accuracy of 0.96752 for English-language tweets.

Paper [21] shows that modern marketing research has mainly relied on dictionary tools to extract sentiment from text data, which have a clear advantage in terms of interpretation but clearly lose in accuracy. The authors also provide a fairly comprehensive assessment of available sentiment analysis methods and show that machine learning-based methods have higher classification accuracy but lower interpretation.

Also, the authors [22] proposed text classification using bidirectional encoder representations from transformers (BERT) for processing natural language with other variants, and showed that the combination of BERT with CNN, BERT with RNN, and BERT with BiLSTM performs well in terms of accuracy, precision, recall, and F1 score compared to being used with Word2vec. The studies were conducted on a dataset containing the entire English Wikipedia and 11,038 books.

The paper [1] analyzes the use of extended BERT models for sentiment recognition of tweets. For a successful evaluation with Enhanced BERT, the Kaggle SMILE dataset is considered, which is checked for emotions such as "happiness", and "sadness", etc., and classified according to the following categories. Experiments show that this version of the model achieves an accuracy of 0.96.

However, most publications are devoted to the work with English-language texts, since there are a sufficient number of labeled datasets, such as IMDB (a labeled dataset containing more than 50,000 movie reviews) [23] and a set of emotionally labeled reviews from the online store Amazon [24]. As for Ukrainian language research, the first problem scientists face is experimental data [25] and the goal of building a model of the Ukrainian spoken language corpus [26]. Mostly, scientists collect such data by themselves, which is a laborious process, and usually, these data are not labeled, they must be marked "manually". For example, in [27], Python-based software was used to extract comments from the Google Maps service. In this paper, it is proposed to use a combination of support vector machines, logistic regression, and XGBoost in combination with a rule-based algorithm. The practical application of the algorithm allows for analyzing Ukrainian-language text by category with visualization of the research results. The accuracy of the proposed method at worst exceeds 0.88.

The above studies have shown that the area of automatic text emotion recognition is a relevant one, but there are much fewer surveys on Ukrainian than on easily formalized languages such as English. This is due to the insufficient number of datasets and the rather difficult formalization of the language, since the spoken Ukrainian language is characterized by a significant number of borrowings, and in addition to them, it also contains fragments borrowed from other languages (Polish, Russian, etc.) [28, 29].

There are labeled datasets for studying the emotional tint of texts, but most of them are in English, one of the most famous being [23], which has 50K movie reviews for natural language processing or text analytics, and [24], which contains a set of emotionally labeled reviews from Amazon. As for the Ukrainian-language labeled datasets, their number is rather small, and such datasets are also few in number. For example, the TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels is a TBCOV dataset that contains 2014792896 multilingual tweets related to the COVID-19 pandemic. The data in the corpus is presented in 67 international languages, including Ukrainian. The number of Ukrainian-language tweets is 3400. Tweets are labeled by emotional color (negative, neutral, positive) [30].

The purpose of classifying the sentiment of Ukrainian-language texts on the example of e-commerce service reviews can be used both to understand people's perception of certain news and for commercial purposes, such as evaluating the work of a manager, etc.

Thus, the aim of the study is to classify the sentiment of Ukrainian-language reviews of e-commerce services using a neural network method.

The main contributions of this study are as follows:

- a neural network method was developed to classify the sentiment of Ukrainian-language reviews from e-commerce services;
- the developed method was adapted to a bilingual dataset, which achieved a classification accuracy of 92 %.

The structure of this article is as follows: Section 2 presents the experimental data for this research, which is a sample of reviews from the Hotline platform, selects the architecture of the neural network – RoBERTa, builds a classifier based on the semantic language model to solve the problem of binary classification of the tone of e-commerce reviews, and studies its effectiveness. Section 3 presents the results and their discussion, demonstrating that due to the imperfect sample, the neural network begins to use memorization with increasing epochs when it cannot find patterns, which demonstrates an increase in accuracy to 98% for the training sample, and the same 92% for the validation sample.

2. Materials and Method

Based on the purpose of the study, the tone assessment will be conducted in relation to e-commerce reviews. In its turn, e-commerce reviews have the following features:

- limited amount of content (up to 500 words);
- small amount of content (1-3 words);
- the use not only in literary Ukrainian but also containing lexical and grammatical elements of different languages and specialized slang, without observing the literary language norms.

As for the limited amount of content, the vast majority of reviews are less than 100 words, and longer reviews are usually negative.

Another characteristic feature of reviews is that a significant number of them have a small amount of content. Among the positive reviews, the following are very common: "*I recommend*", "*I liked everything*", "*The best store*", and among the negative ones, respectively: "*I don't recommend it*", "*Horrible!*", etc. In addition to the fact that reviews can be quite short, they can also contain a lot of jargon, slang, and words that do not comply with the norms of the Ukrainian literary language (foreign words, distorted words, borrowed words, etc.)., professionalism, product names, etc. An example of a part of a review: "*I needed to bring USB 3.0 to the front of the case, because I have USB 3.0 flash drives, and it's not convenient to go to the back of the computer and insert them, because there is only USB 2.0 in the front. So I ordered a Chieftec USB 3.0 adapter on Rozetka...*". Multilingual content is also quite common in reviews. Here's an example of a review that contains errors and russianisms: "*I ordered a battery from an online store. I ordered it because I checked that they have good reviews*". There are spelling mistakes in this sentence, including those resulting from borrowings from the Russian language.

Given these limitations, there is a need to find experimental data that will satisfy the above criteria.

2.1. Datasets

As shown in the review of the source, based on the above criteria, the word corps under consideration cannot be used for this study. Firstly, their total number is 3400, which is relatively small, and secondly, the specificity of a tweet is always a short message, which is usually one phrase. Therefore, we used the dataset of responses from the "hotline" platform, examples of which are:

- "*Rozetka, do you have a conscience? When the war started, they unilaterally canceled all orders. They promised to return the money within 7 days. In 5 days, I've been waiting for a month. At the same time, operators do not answer, and bots in messengers do not work. There is no* *connection and they are still accepting new orders*” (User rating to the review is “*Do not recommend*”);

- “*I ordered and paid for the goods back on February 11, and since then I have not heard a peep(((is it really so difficult to call and clarify?*” (User rating to the review is “*Do not recommend*”);
- “*I ordered the goods from Rozetka's warehouse (not from partners), they were sent quickly in two days, on March 31, and I am waiting for the operational work of Ukrposhta.*” (User rating to the review is “*Recommend*”).

This choice of experimental data is due to the fact that we are interested in conversational Ukrainian-language content, which should also be labeled. The evaluations will be based on the ratings of customers who write reviews, where "Do not recommend" means negative reviews and "Recommend" means positive reviews. The training set did not include data with other ratings. To extract the reviews, appropriate software based on the Crawlee library [31] was created and further processed using C#, divided into 2 directories – "positive" and "negative". A similar approach was used by the authors in [32].

In total, the dataset consists of 7656 documents, with 6655 documents in the training set, and 1331 of them were used for validation (which is 20% of the training set). The peculiarity of the dataset is that it contains Russianisms, swear words, and partially Russian-language reviews. This is due to the fact that although the Russian language has finally lost its dominant position in social media since the beginning of the war, it still prevails – 37% of posts are in Ukrainian versus 63% in Russian, although the statistics in individual social media differ [33, 34]. In addition, reviews often contain misspelled words. The distribution of reviews in the dataset is illustrated in Figures 1-4.

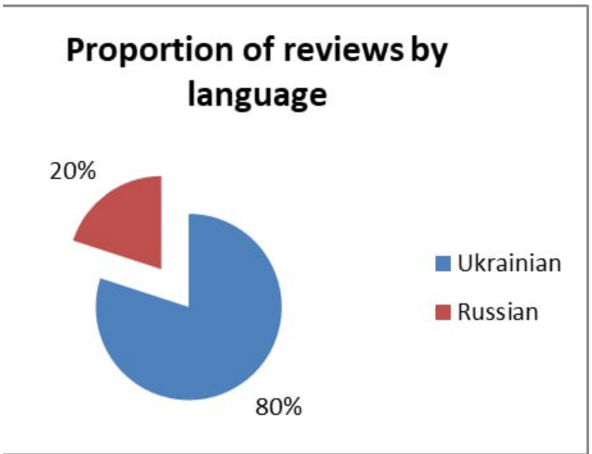

Figure 1: Proportion of reviews by language

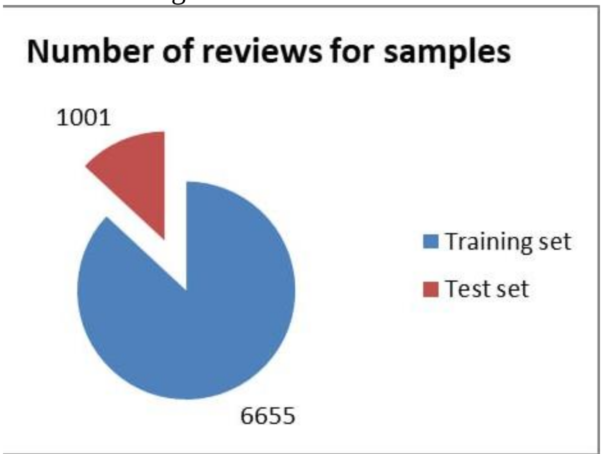

Figure 2: Quantitative distribution of the sample

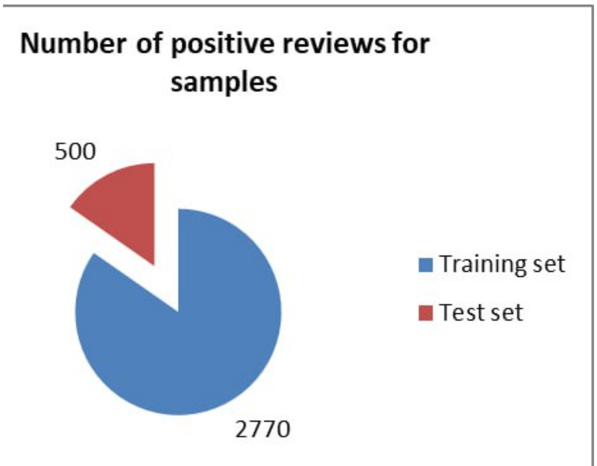

Figure 3: Quantitative distribution of positive reviews

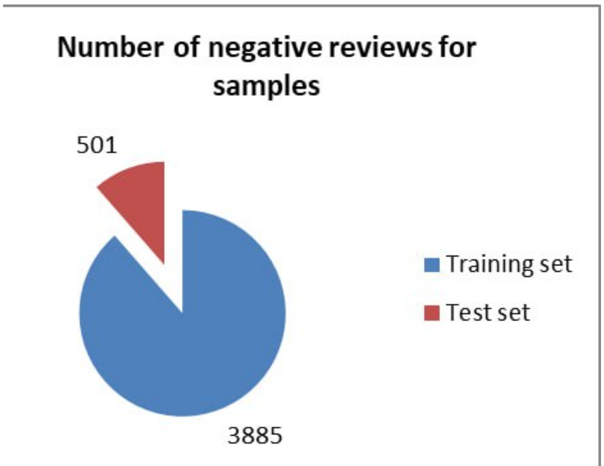

Figure 4: Quantitative distribution of negative reviews

2.2. Choosing a neural network

For binary sentiment classification of Ukrainian-language e-commerce reviews, both neural network options and other options for solving the task were considered. However, based on the analysis of publications, shows that studies that mainly relied on dictionary tools to extract sentiment from text data and have a clear advantage in terms of interpretation, clearly lose accuracy. Among the neural network tools discussed above, BERT-like networks are currently considered the best.

BERT was designed to help computers understand the meaning of ambiguous language in a text by using the surrounding text to understand the context in which the text might have been written [35-37]. However, as already studied by the authors of [25], ukr-RoBERTa, ukr-ELECTRA and XLM-R large tend to perform the best, although XLM-R large and ukr-ELECTRA tend to perform better on longer texts, while ukr-RoBERTa significantly outperforms the other models on shorter sequences. Since the study is conducted on the texts of reviews of the Internet platform "Hotline" [38], which are usually short text messages, and based on the conducted research, it was decided to use the RoBERTa neural network.

2.3. Selecting a semantic language model

The RoBERTa neural network variation (short for "Robustly optimized BERT approach") is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model developed by Facebook AI researchers [39]. Like BERT, RoBERTa is a transformer-based language model that uses self-awareness to process input sequences and create contextualized representations of words in a sentence.

One of the key differences between RoBERTa and BERT is that RoBERTa was trained on a much larger dataset and used a more efficient training procedure. During training, RoBERTa uses a dynamic masking technique that helps the model learn more reliable and generalized word representations.

Since semantic analysis based on a neural network approach is a current area of research, there are also some developments for the Ukrainian language. One of them is a pre-trained multilingual preprocessing model that also works with Ukrainian and more than 50 other languages [40] and embedding [41] by Ukjae Jeong, which is part of the models of the Tensorflow_hub library in Python. Based on these models, it is proposed to create a model that will be trained on the above sample of experimental data. The choice of multilingual models is due to the fact that, as mentioned above, reviews can contain text not only in the literary Ukrainian language.

2.4. Classifier architecture

The neural network configuration based on the selected dataset and neural network type has the structure shown in Figure 5.

The input layer converts the input text information into a Keras tensor, i.e., a symbolic tensor-like object, which is supplemented with attributes that allow building a Keras model based on the input and output data of the model. Subsequently, the tensor is fed to the input of the preprocessing layer, which includes a wrapper of the called object, to be used as a Keras layer based on a pre-trained text preprocessing model [40]. This model uses SentencepieceTokenizer [42], which tokenizes the UTF-8 string tensor and is an unsupervised text tokenizer and detokenizer.

The next layer is the RoBERTa encoder. This layer is based on the pre-trained model "*xlm_roberta_multi_cased_L-12_H-768_A-12*" [41], which is the result of unsupervised cross-language representative training at scale (XLM-RoBERTa) [41] and is pre-trained on 2.5 TB of filtered CommonCrawl data containing 100 languages [43].

The next layer is the dropout layer, which randomly sets the input units to 0 at a rate of speed at each step during training, which helps prevent overtraining [441]. Inputs that are not set to 0 are scaled so that the sum of all inputs does not change.

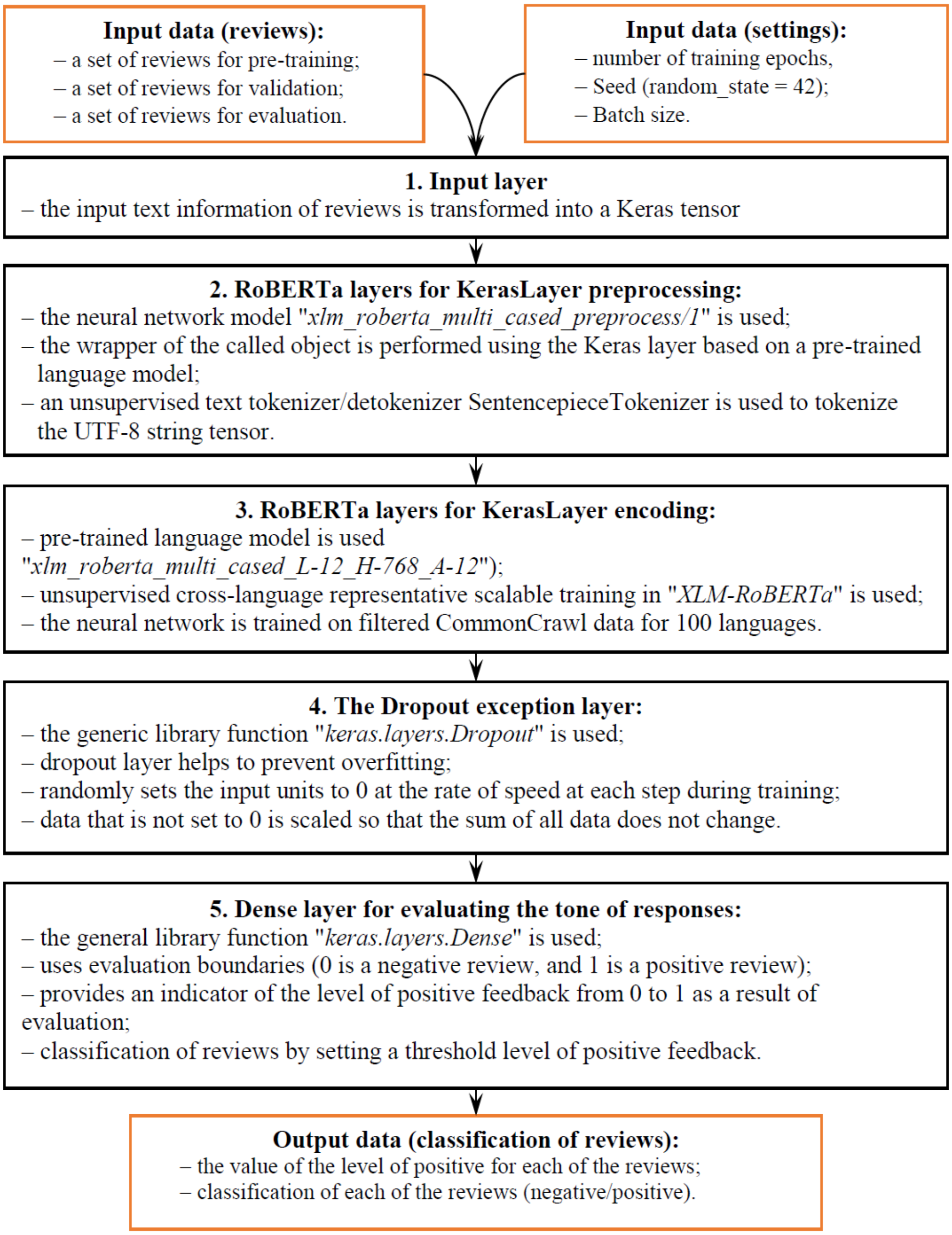

Figure 5: Schema of the RoBERT-based classifier for classifying the tone of e-commerce reviews

The number of training epochs shows how many times the model is to be trained. The Seed parameter will be taken as 42, given [45, 46] that if you do not set random_state to 42, every time the program code is run again, it will create a different test set. Batch size – the number of training examples used within one iteration. It is very difficult to immediately determine what the ideal batch size is for the needs of a particular task [47, 48], so this parameter will be selected experimentally.

2.5. Study of the effectiveness of sentiment classification of Ukrainian-language reviews

According to the selected parameters, the indicators for evaluating the model's functionality were determined, such as training time in seconds, accuracy, and losses. The binary cross-entropic function expressed by the formula [49] was used as a loss function:

$$Loss = -\frac{1}{N}\left[\sum_{j=1}^{N}[t_j \log(p_j + (1-t_j)\log(1-p_j))]\right],$$

where N – is the number of data samples, t_j – is a true value that takes the value 0 or 1, p_j – is the Softmax probability for the i-th data point.

The accuracy of the study is defined as the number of correct answers divided by the total number of answers [50].

3. Result and Discussion

The obtained indicators for evaluating the functionality (training time, accuracy, and losses) of various parameters of the model settings (number of training epochs, seed, batch size) of the neural network classifier are shown in Table 1. The experiment was conducted on the basis of an Intel Core I7 8th gen processor, 16 GB of RAM, and NVIDIA GeForce MX150.

As seen in Table 1, model V1 has the highest accuracy score of 0.92 and the lowest loss function of 0.29, while model V6 also has an accuracy score of 0.92 but a loss function of 0.30 and a much higher training time.

Despite minor deviations in accuracy, almost all versions of the trained models on real-world examples produced results similar to the expert opinions, some of which are shown in Table 2 to compare different versions of the trained models (V1-V6 from Table 1).

Table 1
Classifier retraining parameters

Parameters	V1	V2	V3	V4	V5	V6
Number of training epochs	3	3	4	5	3	10
Seed	42	42	42	42	42	42
Batch size	64	32	32	32	16	64
Training time (sec)	10028	9224	12158	15894	10248	33952
Accuracy	**0.92**	0.91	0.91	0.91	0.91	0.92
Loss	**0.29**	0.31	0.30	0.32	0.31	0.30

Considering that the tested version of RoBERTa is a multilingual transformer trained on bilingual data, the neural network shows no problems with sentiment identification, as illustrated in Table 2.

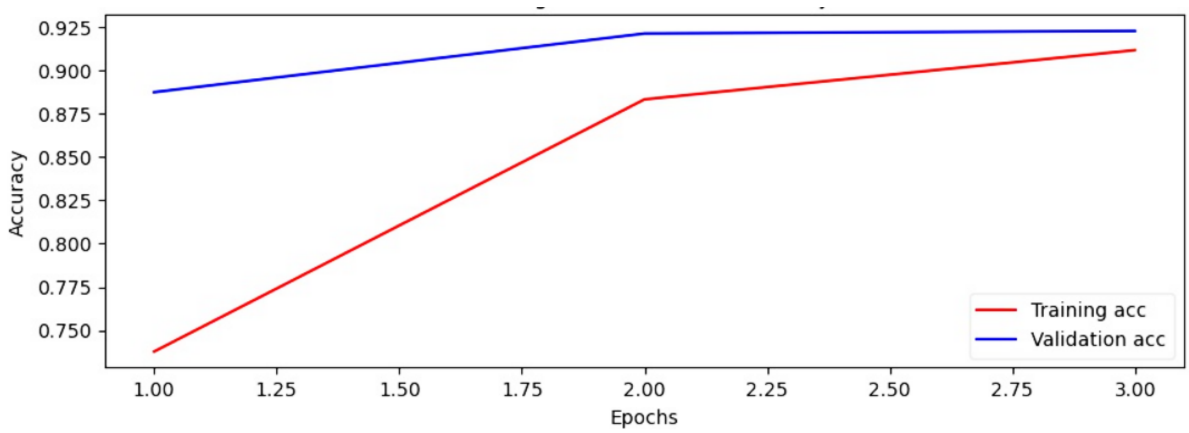

Figure 6: Illustration of the learning process by epochs in terms of accuracy (V1)

By studying the responses that are not present in the training and test samples, the high efficiency of the proposed architecture is shown. The training set was not manually cleaned, so it is possible that there may be a certain percentage of misclassified reviews, but this does not have a significant impact on the final accuracy of the binary classification of the emotional tone of reviews written not only in pure Ukrainian but also containing bilingual data. Figure 6 illustrates the changes in the accuracy parameter depending on the epochs passed, and Figure 7 illustrates the changes in the loss function for the combination of V1 training parameters from Table 1 (3 epochs, 64 batch sizes).

The graph in Figure 6 indicates that the number of training epochs is not enough to stabilize the result, as the Accuracy indicator tended to increase and the loss function indicator tended to decrease, without stabilizing at the same level.

Table 2
Classification of the tone of reviews

Translation Reviews from Ukrainian	Evaluation (V4)	Evaluation (V2)	Evaluation (V6)
Your product is complete shit, you can't find anything worse	0.005181	0.014710	0.000641
We are very satisfied with the purchase, we will come back again	0.997751	0.990176	0.996549
It's good to have such good sellers like you.	0.988962	0.991397	0.995478
Our family buys goods here again and always the service is on top, we recommend	0.948719	0.990182	0.871778
I would never recommend using this service! It's just horrible!	0.002086	0.011778	0.000665
There were no drivers on the computer at all. On 13/01/2023 in the morning, I took the computer to the store for a refund or exchange for another model, as it turned out they could not exchange it, despite the fact that I chose a more expensive model and only issued a refund. We had to sit in the store for 2 hours and wait for the seller to reset the Yepo to factory settings, only then they said they would be able to issue a refund (it was just horrible, we didn't even use it and it was obvious)	0.004561	0.016255	0.001831
As for me, Rozetka is the best store. A big plus is a free delivery to all their branches. There are no questions about the warranty either, so I recommend this store	0.995170	0.988024	0.957222
Rozetka once again pleasantly surprised me with the service! The first time when the router broke after more than a year of work and I was refunded the amount I paid at the time of purchase, not after repairing my own router, and this time I ordered my daughter a set of desk + chair, the price was good, they brought it exactly as specified when ordering. No one blamed us for breaking the lamp, it was mechanical damage and it will not be possible to replace it, this was not even close! Thanks to the outlet for the most adequate solution to our issue!	0.968909	0.836309	0.969624
This is extortion, thievery by prom.ua – there is no other way to describe it. !!!!! Nowhere in the world is there such a thing – that marketplaces take a commission of 10-20% from sellers, and + an annual package of 5700-11500 UAH must be paid in addition to these percentages.	0.003402	0.014268	0.000993

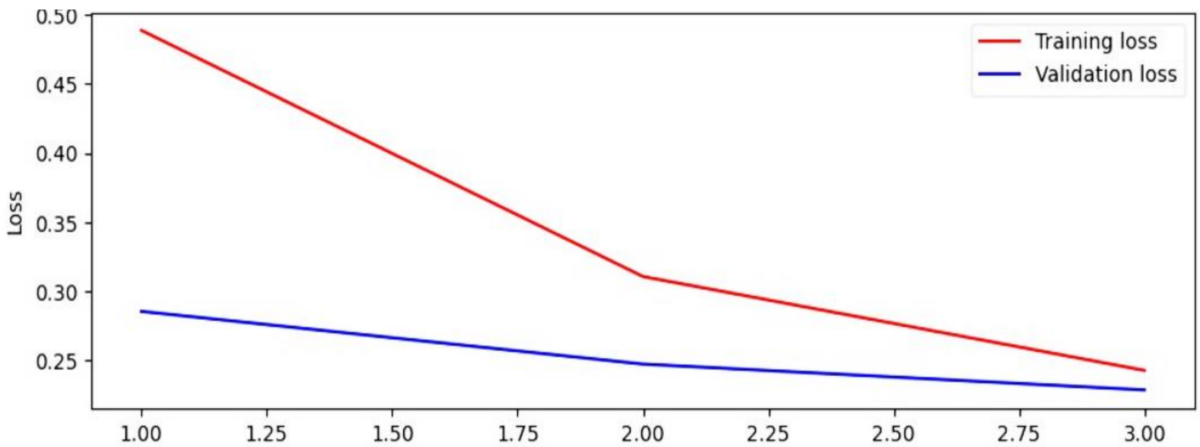

Figure 7: Illustration of the learning process by epochs in terms of the loss function (V1)

However, by continuing the experiment, and changing the number of training epochs to 10, which corresponds to V6 in Table 1, the results illustrated in Figure 8 and Figure 9 were obtained.

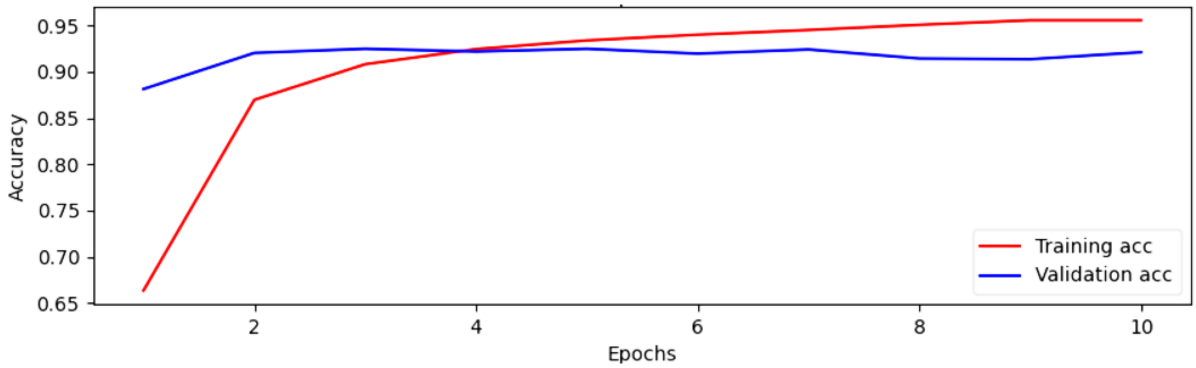

Figure 8: Illustration of the learning process by epochs in terms of accuracy (V6)

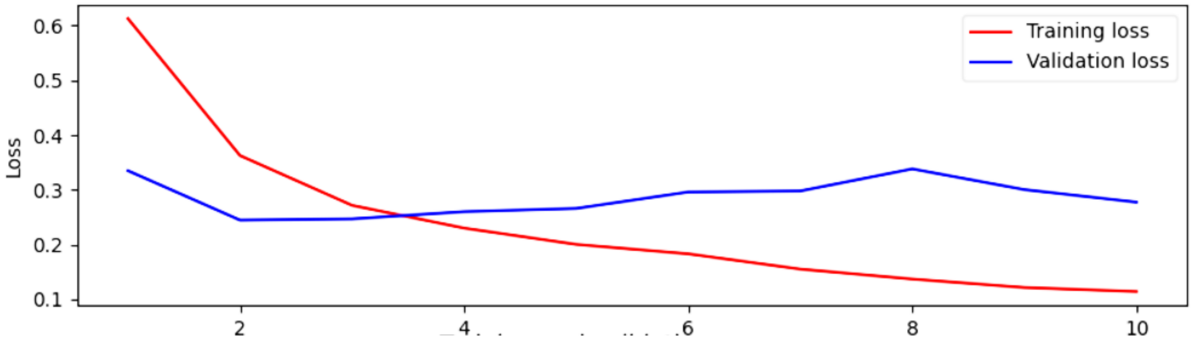

Figure 9: Illustration of the learning process by epochs in terms of the loss function (V6)

The results show that using the validation sample does not increase the classification accuracy. And the loss function generally tended to increase slightly after the 3rd iteration for the validation sample. However, such results may indicate that the samples are not sufficiently filtered. After all, testing the neural network on reviews not contained in the database yielded almost error-free results for 40 reviews that actually contained emotion. The positive sample includes reviews such as: "*Microwave*", "*Bought a computer*", "*Bought headphones*", "*Bought a vacuum cleaner*", etc. However, the same kind of feedback is also found in the negative sample.

The graph illustrating the completion of the retraining process by epochs for V4 of Table 1 is shown in Figure 10 and Figure 11.

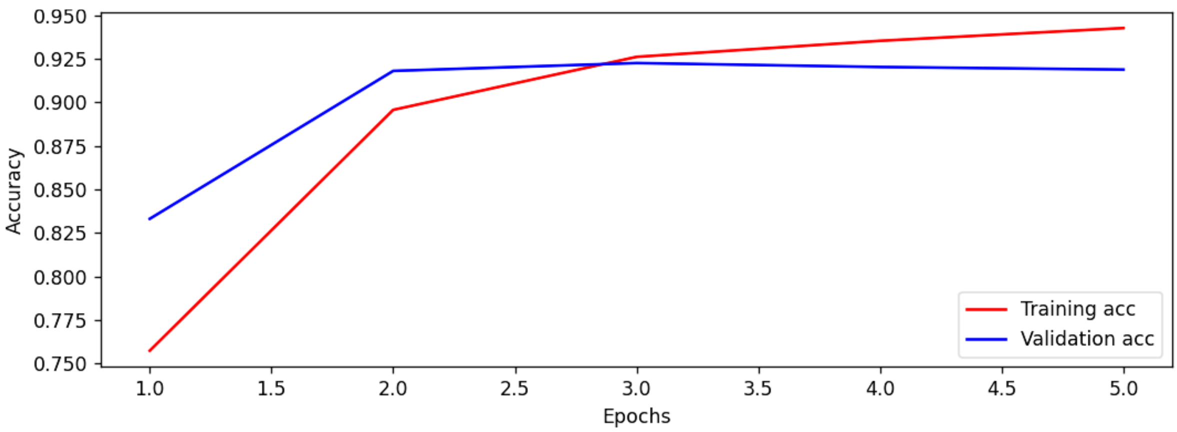

Figure 10: Illustration of the learning process by epochs in terms of accuracy (V4)

The results of this experiment show that the dataset was not manually cleaned. Therefore, as the number of epochs grows, the neural network begins to simply "*remember*" which reviews belong where, as evidenced by the red line in Figures 8-9 and 10-11. Since the loss function is much smaller for the training set, the accuracy is much higher. However, the obtained loss function and precision values are due to the fact that the sample was not manually filtered and contained reviews that included unemotional comments, often consisting of a single word or phrase such as: "*Microwave*", "*bought a computer*", "*bought headphones*", "*bought a vacuum cleaner*" etc.

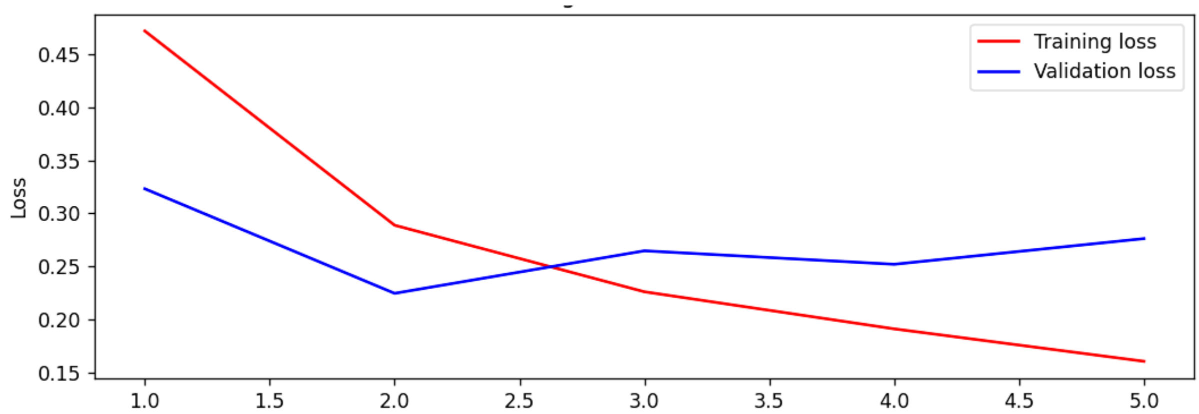

Figure 11: Illustration of the learning process by epochs in terms of the loss function (V4)

In addition, the analysis of tone estimation showed that the neural network coped with the task without any errors out of 40 phrases that were not in either the training or test samples and that had been previously evaluated by an expert, and the feedback contained both stylistic and spelling errors and was represented by multilingual data. Even not-so-unambiguous reviews, such as: "*Delivery in Kyiv on hotline was declared free of charge, but on the store's website there were options for delivery for 100 UAH by courier or 80 UAH by Nova Poshta*" were rated by the neural network at 0.016359, which coincides with the author of the hotline review, who also gave the review a "*Do not recommend*" rating and with the expert's rating. On the other hand, the review "*The seller did not offer unnecessary things, did not impose any additional services or guarantees, did not "sell" accessories I did not need, etc. – everything was quick and clear, he immediately proceeded to place the order and clarify the delivery details. I'm satisfied with the product, I got what I expected.*", which contains words that are responsible for negativity, such as: "*imposed*", "*unnecessary*", "*selling*", the review was identified as positive with a score of 0.808049.

This indicates that the neural network really "*understands*" the context. Some hesitation in the neural network occurs with neutral reviews such as: "*The price is right, so is the availability*". Such a review was written with a rating of "*Recommend*", and the neural network identified it as positive, but with an almost marginal rating of 0.505790. The neural network also handles reviews like this: "*I ordered an Ambrosio Halmar table. Very pleased with the purchase ???? full compliance with the photo and fast delivery (less than two weeks). I recommend ????????"*. The neural network's score for this review is 0.902363, but the expert's understanding of the question marks was ambiguous.

The proposed approach has certain limitations. It is advisable to apply it to determine the tone of short text reviews (up to 500 words long) presented in Ukrainian and may contain not only in literary Ukrainian but also containing lexical and grammatical elements of different languages and specialized slang, without observing the literary language norms. Changing the content of the training dataset affects the result of neural network training, and accordingly affects the efficiency of binary classification of texts. Over time, everyday language may change, which also affects the progress and results of text message sentiment classification.

Further research will be aimed at implementing this classifier to evaluate the work of managers when communicating with online store customers, implementing marketing feedback models, and improving the efficiency of classifiers that can work with multiple languages simultaneously. It is planned to conduct a study with an expanded dataset of responses and removal of ambiguous collocations.

4. Conclusion

The paper considers the current state of the field of semantic text processing, namely, sentiment classification of text messages. The analysis has shown that this area is relevant, in particular, the use of neural networks to classify the sentiment of text documents, which gives a higher classification accuracy than alternative approaches. The BERT architecture was identified as one of the most accurate neural networks, but its modification, RoBERTa, proved to be better for analyzing short documents.

When developing the method, the following issues were researched: the development of a labeled dataset for training the neural network, the selection and tuning of a neural network classifier, and the building of a semantic language model. Since the purpose of the study was to classify the sentiments of Ukrainian-language e-commerce reviews, and such reviews have certain characteristics, an own dataset of 7656 reviews was created to train the selected RoBERTa neural network. The collected reviews were divided into 2 samples – training and testing, each of which had negative comments and positive comments. The accuracy and loss functions were used to evaluate the performance of the proposed architecture. For the combined multilingual reviews, an accuracy of 0.92 was obtained, while the loss function had a value of 0.29.

The proposed approach is advisable to apply it mainly to determine the tone of short text reviews (up to 500 words long) presented in Ukrainian and may contain not only in literary Ukrainian but also containing lexical and grammatical elements of different languages and specialized slang, without observing the literary language norms.

Further research will be aimed at implementing this classifier to evaluate the work of managers when communicating with online store customers, implementing marketing feedback models, and improving the efficiency of classifiers that can work with multiple languages simultaneously.

5. References

[1] S. Mann, J. Arora, M. Bhatia, R. Sharma, R. Taragi, Twitter Sentiment Analysis Using Enhanced BERT, in: A.J. Kulkarni, S. Mirjalili, S.K. Udgata, Intelligent Systems and Applications. Lecture Notes in Electrical Engineering, vol 959, Springer, Singapore, 2023, pp. 263-271. doi:10.1007/978-981-19-6581-4_21.

[2] B. Albadani, R. Shi, J. Dong, A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM, Applied System Innovation, 2022; 5(1):13. doi: 10.3390/asi5010013.

[3] M. Bibi, W. A. Abbasi, W. Aziz, S. Khalil, M. Uddin, C. Iwendi, T. R. Gadekallu, A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis, Pattern Recognition Letters, Volume 158, 2022, pp. 80-86. doi: 10.1016/j.patrec.2022.04.004.

[4] A. P. Rodrigues, R. Fernandes, A. Aakash, B. Abhishek, A. Shetty, K. Atul, K. Lakshmanna, R. M. Shafi, Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques, Computational Intelligence and Neuroscience, vol. 2022 (2022). doi: 10.1155/2022/5211949.

[5] E. A. Manziuk, A. V. Barmak, Y. V. Krak, V. S. Kasianiuk, Definition of information core for documents classification, Journal of Automation and Information Sciences, 50(4), (2018) pp. 25-34. doi:10.1615/JAutomatInfScien.v50.i4.30.

[6] G. C.Huang, J. B.Unger, D. Soto, K. Fujimoto, M. A. Pentz, M. Jordan-Marsh, T. W. Valente, Offline Friendship Networks on Adolescent Smoking and Alcohol Use, doi:10.1016/j.jadohealth.2013.07.001.

[7] R. J. Moreira de Freitas, T. N. Carvalho Oliveira, J. A. Lopes de Melo, J. do V. e Silva, K. C. de Oliveira e Melo, S. Fontes Fernandes, Adolescents' perceptions about the use of social networks and their influence on mental health, 2021. doi:10.6018/eglobal.462631.

[8] B. Dave, Sh. Bhat, P. Majumder, IRNLP DAIICT@DravidianLangTech-EACL2021: Offensive Language identification in Dravidian Languages using TF-IDF Char N-grams and MuRIL, Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, Association for Computational Linguistics, 2021, pp. 266-269.

[9] L. Wei, S. Wei, J. Shaoxiong, E. Cambria, BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis, Neurocomputing (2022), pp. 73-82. doi: 10.1016/j.neucom.2021.09.057.

[10] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, E. Cambria, DialogueRNN: An attentive RNN for emotion detection in conversations, Proceedings of the AAAI Conference on Artificial Intelligence, vol.33, 2019, pp. 6818-6825. doi: https://doi.org/10.48550/arXiv.1811.00405
[11] O. Kovalchuk, V. Slobodzian, O. Sobko, M. Molchanova, O. Mazurets, O. Barmak, I. Krak, N. Savina, Visual Analytics-Based Method for Sentiment Analysis of COVID-19 Ukrainian Tweets, Book Chapter. Lecture Notes on Data Engineering and Communications Technologies, 2023, Vol. 149, pp. 591-607. doi: 10.1007/978-3-031-16203-9_33.
[12] I. Olenych, M. Prytula, O. Sinkevych, O. Khamar, System of Automatic Determination of Ukrainian Text Tone, 2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 2021, pp. 80-83. doi: 10.1109/ELIT53502.2021.9501124.
[13] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal, Vol. 5, Issue 4 (2014), pp. 1093-1113. doi: 10.1016/j.asej.2014.04.011.
[14] H. Li, Q. Chen, Z. Zhong, R. Gong, G. Han, E-word of mouth sentiment analysis for user behavior studies, Information Processing & Management (2022). doi: 10.1016/j.ipm.2021.102784.
[15] L. Lades, K. Laffan, M. Daly, L. Delaney, Daily emotional well-being during the COVID-19 pandemic, British Journal of Health Psychology 25(3) (2020). doi: 10.1111/bjhp.12450
[16] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, A. Hassanien, Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers – A study to show how popularity is affecting accuracy in social media, Applied Soft Computing 97 (2020). doi: 10.1016/j.asoc.2020.106754
[17] M. Mansoor, K. Gurumurthy, R. U. Anantharam, V. R. B. Prasad, Global Sentiment Analysis Of COVID-19 Tweets Over Time, 2020. URL: https://arxiv.org/pdf/2010.14234.
[18] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, G. S. Choi, A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis, PLoS ONE 16(2):e0245909 (2021). doi: 10.1371/journal.pone.0245909.
[19] R. Marcec, R. Likic, Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines, Postgraduate Medical Journal, Volume 98, Issue 1161, (2022), pp. 544-550. doi: 10.1136/postgradmedj-2021-140685.
[20] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, P. Cotae, Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset, Expert Systems with Applications (2023). doi: 10.1016/j.eswa.2022.118715.
[21] J. Hartmann, M. Heitmann, C. Siebert, C. Schamp, More than a Feeling: Accuracy and Application of Sentiment Analysis, International Journal of Research in Marketing (2022). doi: 10.1016/j.ijresmar.2022.05.005.
[22] B. Abayomi, S. Ng, M. Leung, A BERT Framework to Sentiment Analysis of Tweets. Sensors 23 (2023). doi: 10.3390/s23010506.
[23] Kaggle, IMDB Dataset of 50K Movie Reviews, 2019. URL: https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
[24] Kaggle, Amazon Reviews for Sentiment Analysis, 2020. URL: https://www.kaggle.com/datasets/bittlingmayer/amazonreviews.
[25] D. Panchenko, D. Maksymenko, O. Turuta, A. Yerokhin, Y. Daniiel, O. Turuta, Evaluation and Analysis of the NLP Model Zoo for Ukrainian Text Classification, Information and Communication Technologies in Education, Research, and Industrial Applications, ICTERI 2021, Communications in Computer and Information Science, vol 1698, Springer, Cham. doi: 10.1007/978-3-031-20834-8_6.
[26] I. G. Kryvonos, I. V. Krak, O. V. Barmak, R. O. Bagriy, Predictive text typing system for the Ukrainian language, Cybernetics and Systems Analysis, 53(4), (2017), pp. 495-502. doi:10.1007/s10559-017-9951-5.
[27] K. Shakhovska, N. Shakhovska, P. Vesely, The Sentiment Analysis Model of Services Providers' Feedback, Electronics (2020) 9, no. 11, pp. 19-22. doi: 10.3390/electronics9111922.

[28] V. Slobodzian, O. Kovalchuk, M. Molchanova, O. Sobko, O. Mazurets, O. Barmak, I. Krak, Text Data Vectorization Model of Ukrainian-Language Internet Communication Content, CEUR Workshop Proceedings, 2022, vol. 3171, pp. 561-571.
[29] Я. Лазоренко, І. Сініцин, В. Шевченко, Ідентифікація переважної мови спілкування людини, Проблеми програмування, 2022. № 3-4, Спеціальний випуск, с. 271-280. doi: 10.15407/pp2022.03-04.271.
[30] I. Muhammad, U. Qazi, F. Ofli, TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels, in: Data 7, no. 1: 8. doi: 10.3390/data7010008.
[31] Crawlee.Dev, A web scraping and browser automation library. URL: https://crawlee.dev.
[32] G. Maccario, M. Naldi, Alexa, Is My Data Safe? The (Ir)relevance of Privacy in Smart Speakers Reviews, International Journal of Human–Computer Interaction (2022), 13. doi: 10.1080/10447318.2022.2058780.
[33] Speka, What language do Ukrainian social networks speak, 2022. URL: https://speka.media/socialni-merezi/yakoyu-movoyu-govoryat-ukrayinski-socmerezi-v5m019.
[34] BBC. News, Ukrainian has significantly strengthened in all spheres: at home, at work and on the Internet. Poll, 2023. URL: https://www.bbc.com/ukrainian/news-64201995.
[35] Medium, Sentiment Analysis of Movie Reviews with Google's BERT, 2021. URL: https://medium.com/mlearning-ai/sentiment-analysis-of-movie-reviews-with-googles-bert-c2b97f4217f.
[36] F. Sun, H. Xu, Y. Meng, Z. Lu, S. Chen, Q. Wei, C. Bai, BERT and Pareto dominance applied to biological strategy decision for bio-inspired design, Advanced Engineering Informatics, Vol. 55, 2023, doi:10.1016/j.aei.2023.101904.
[37] K. Tian, Y. Jiang, Q. Diao, C. Lin, L. Wang, Z. Yuan, Designing BERT for Convolutional Networks: Sparse and Hierarchical Masked Modeling. arXiv preprint arXiv:2301.03580.
[38] Hotline, Reviews of the store Rozetka, 2023. URL: https://hotline.ua/ua/yp/2476/reviews.
[39] Ai.Facebook, RoBERTa: An optimized method for pretraining self-supervised NLP systems, 2019. URL: https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems.
[40] Tfhub, Text preprocessing model xlm_roberta_multi_cased_preprocess. URL: https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_preprocess/1.
[41] Tfhub, Unsupervised Cross-lingual Representation Learning at Scale. xlm_roberta_multi_cased_L-12_H-768_A-12, 2023. URL: https://tfhub.dev/jeongukjae/xlm_roberta_multi_cased_L-12_H-768_A-12/1.
[42] Tensorflow, Sentence piece Tokenizer, 2023 URL: https://www.tensorflow.org/text/api_docs/python/text/SentencepieceTokenizer.
[43] Huggindface, XLM-RoBERTa (base-sized model). URL: https://huggingface.co/xlm-roberta-base.
[44] Tensorflow, Tf.keras.layers.Dropout, 2023, URL: https://www.tensorflow.org/api_docs/python /tf/keras/layers/Dropout.
[45] The clever programmer, Why Random_state=42 in Machine Learning, 2020. URL: https://thecleverprogrammer.com/2020/12/17/why-random_state42-in-machine-learning.
[46] R. Pramodi, Why do we set a random state in machine learning models? 2022. URL: https://towardsdatascience.com/why-do-we-set-a-random-state-in-machine-learning-models-bb2dc68d8431.
[47] Medium, How does Batch Size impact your model learning, 2022. URL: https://medium.com/geekculture/how-does-batch-size-impact-your-model-learning-2dd34d9fb1fa.
[48] D. Huynh, How to get 4x speedup and better generalization using the right batch size, 2019. URL: https://towardsdatascience.com/implementing-a-batch-size-finder-in-fastai-how-to-get-a-4x-speedup-with-better-generalization-813d686f6bdf.
[49] K. E. Koech, Cross-Entropy Loss Function, 2020. URL: https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e.
[50] Y. Krak, O. Barmak, O. Mazurets, The practice implementation of the information technology for automated definition of semantic terms sets in the content of educational materials, CEUR Workshop Proceedings, 2018, vol. 2139, pp. 245-254. doi:10.15407/pp2018.02.245.