СЕКЦІЯ XV.
СИСТЕМНИЙ АНАЛІЗ, МОДЕЛЮВАННЯ ТА ОПТИМІЗАЦІЯ

INTELLIGENT SYSTEM FOR SENTIMENT ANALYSIS AND CLUSTERING OF PUBLICATIONS WITHIN THE SOCIAL NETWORK OF TWITTER

Batiuk Taras
ORCID ID: 0000-0001-5797-594X
PhD student of Computer sciences and Information technologies
Lviv Polytechnic National University, Ukraine

The development and implementation of an intelligent system for analyzing the sentiment and organizing publications is a relevant and promising task. In today's world, social networks serve as the primary means of communication, reflecting various situations and circumstances. Each user's message within a social network carries specific semantics, reflecting their thoughts, analysis of the relevant situation, or reactions to specific events. Advanced algorithms and data analysis approaches [1] now enable efficient and straightforward analysis of large volumes of textual data. This enables the determination of the average sentiment expressed by users towards specific events and allows drawing conclusions based on the analyzed content. Consequently, it becomes possible to understand the attitudes of different user groups towards various types of content, offers, goods, and market offerings. Major brands and corporations actively employ these approaches to analyze information, determining the sentiment of user texts and dividing them, along with the users themselves, into clusters for further analysis and engagement.

The authors in the article [2] showed the development of an intelligent vehicle sales forecasting model using Internet public opinion and online search index data. The authors use a dataset of monthly sales of ten major car brands in China from 2012 to 2019. The authors use a time series model and machine learning algorithms to analyze data and forecast future car sales. Overall, the article is well written and the methodology is well explained. The authors provide a clear explanation of the data set and the various models and algorithms used in the analysis. The research results show that online public opinion and online search index data can be useful for vehicle sales forecasting, and the proposed model performs better than traditional time series models.

Article [3] provides valuable information on using public opinion and online search index data to predict vehicle sales. Although the study has some limitations, the results are promising and have important implications for the automotive industry and for the use of online data in predictive modeling. In the article [4], the authors propose a technique for optimizing web pages and increasing their ranking in search engines based on content analysis of keywords of web pages. The authors claim that this approach can be effective in increasing website visibility and traffic, which in turn can contribute to online business success. Overall, the paper provides a comprehensive overview of the proposed methodology, including its theoretical foundations, implementation process, and evaluation results.

The authors in [5] present some case studies to illustrate the application of the approach, including more empirical evidence to demonstrate the effectiveness of the methodology in improving website rankings and attracting traffic. Another shortcoming of the article is that it focuses primarily on the technical aspects of the methodology, neglecting other factors that can contribute to user experience. At the heart of recurrent networks is the concept of memory cells, in which certain points of representation of data are stored. Traditional RNNs have a drawback, which is that at a certain moment of the network's operation, the number of data representation points can increase so much that it will be impossible to remember new data, due to which the final results may be distorted. Convolutional neural networks were used for the work since the amount of data was relatively small.

Thanks to training with the teacher, two different models of user data analysis were created, based on which the comments and posts of users were further processed, and the features of their mental health were determined by analyzing the text's sentiment using a convolutional neural network. The database is not created since the essence of creating an intelligent system consists of developed algorithms for the current analysis of sentiment and clustering of users' publications in the social network Twitter with given keywords. Figure 1 shows the degree of sentiment of the analyzed text instances, the vast majority of publications and comments carry a neutral context.

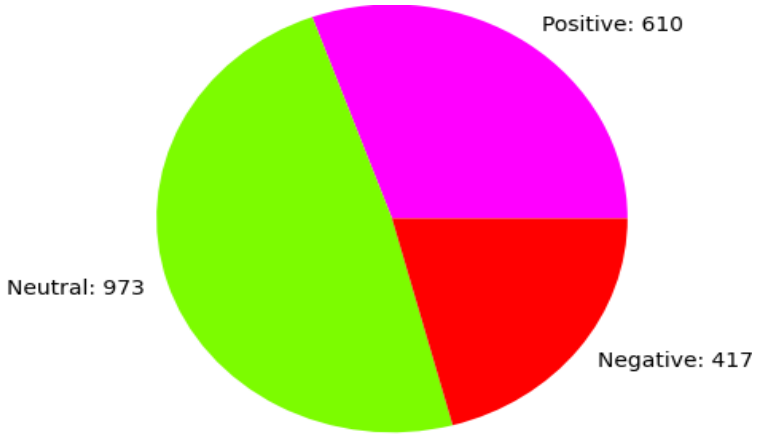

Fig. 1. **General results of the analysis of the sentiment of the text**

After obtaining the value of the initial sentiment, an analysis is carried out, with the help of which we learn the degree of emotionality of the words. The analysis is somewhat more complicated, as it contains three LSTM layers. Accordingly, each of them had to run its activation function. All work will be performed asynchronously for greater efficiency since we have used external resources to execute the code and wait for the result to return from the function. Neutral text instances are depicted in purple, positive ones in orange, and negative ones in green, clearly visible on the graph. We also see that the assessment of the sentiment of positive comments varies in terms of values from 0.1 to 0.4, and the evaluation of the negative distribution of numbers from 0.1 to 0.2, thus received measure of sentiment is shown in Figure 2.

For the correct operation of the intelligent system, the clustering parameters were set, such as model.fit - model settings, max iteration - the number of changes in the position of the cluster center, in our case, it is 300 changes, n_init value - the value indicating how many times clustering will occur and directly depends on the selection of cluster centers, tol or tolerance – checking whether cluster centers move during model training, if the movement is less than one thousandth, then clustering is completed. To determine the number of

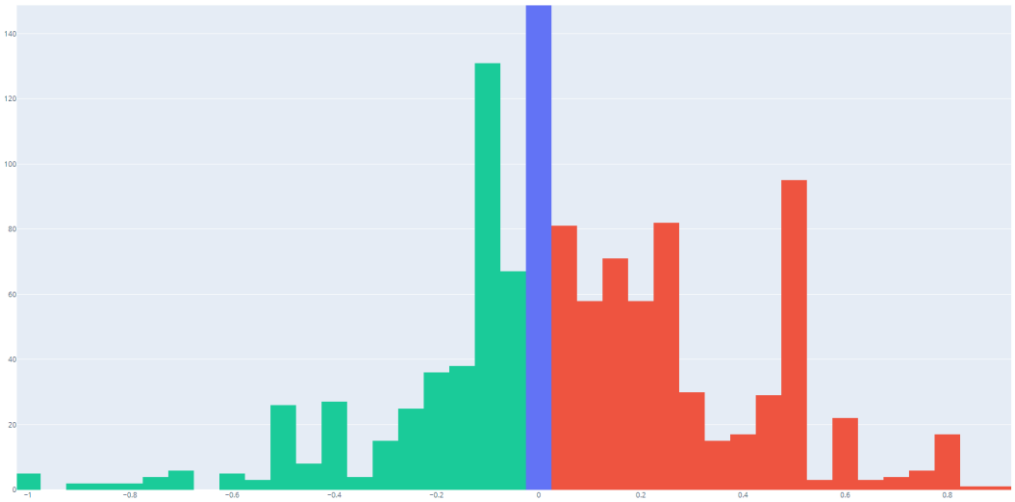

Fig. 2. **Measure of sentiment**

clusters, the "elbow" method was used, which is shown in Figure 3, according to the results, 3 clusters were chosen at the greatest decline.

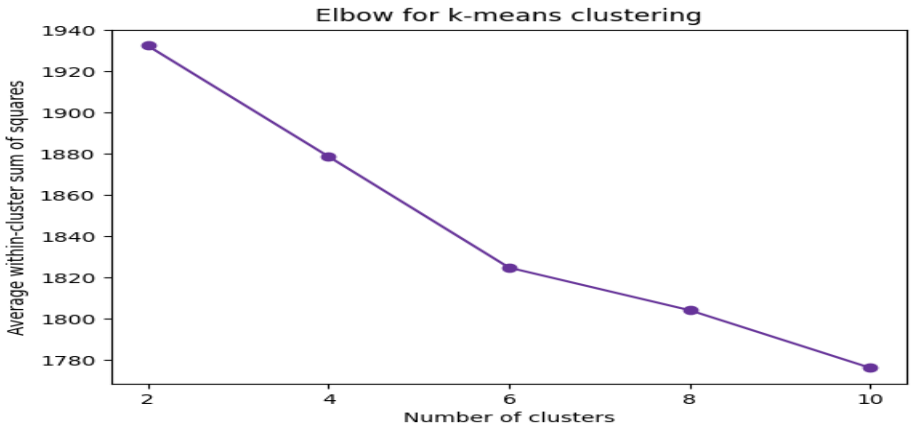

Fig. 3. **The result of the "Elbow" method**

Also, a three-dimensional 3D graph was created for better visualization, which is shown in Figure 4. Based on the graph, it is possible to conclude that all selected text instances belong to a specific cluster in three-dimensional space.

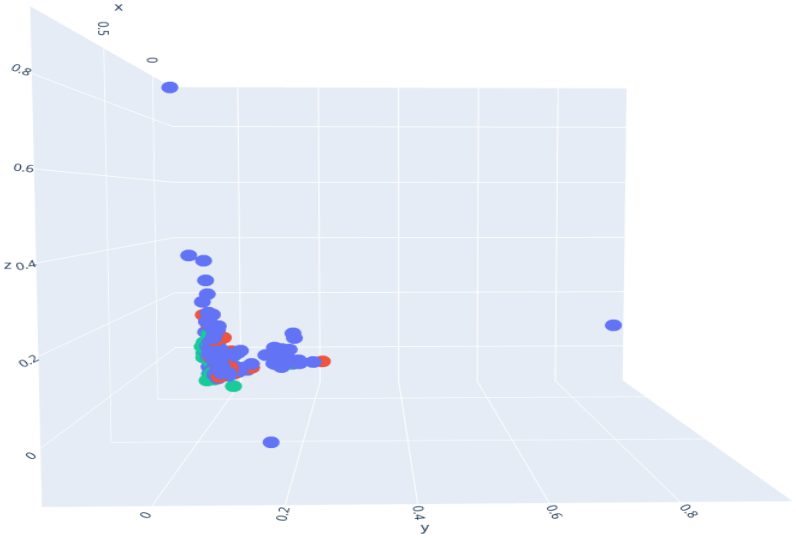

Fig. 4. **Volumetric graph of cluster distribution**

During the project, we successfully developed an intelligent system capable of analyzing the sentiment and clustering of posts within the Twitter social network. This system allows users to input specific parameters and keywords to retrieve relevant data for sentiment analysis, emotional evaluation, and subsequent clustering. We outlined the objectives of the project, emphasizing the necessity of implementing a modern LSTM neural network and the k-means clustering algorithm to enhance system efficiency and improve the accuracy of the final results. Additionally, we provided a detailed description of the system's functionality, including the underlying algorithms and key functional components. Overall, the successful implementation of this intelligent system provides users with a powerful tool for analyzing sentiment, evaluating emotions, and clustering Twitter social network posts. The combination of advanced neural networks and clustering algorithms ensures enhanced system performance and accurate results, enabling valuable insights and data utilization.

References:

1. Kim D., Kim Y., Jeong Y.-S. Graph Convolutional Networks with POS Gate for Aspect-Based Sentiment Analysis. Applied Sciences. 2022. № 12 (19). P. 101–134.
2. Albahli S., Irtaza A., Nazir T., Mehmood A., Ali A., Waleed Albattah W. A Machine Learning Method for Prediction of Stock Market Using Real-Time Twitter Data. Electronics. 2022. № 11 (20). P. 341–363.
3. Opiła J. On Employing of Extended Characteristic Surface Model for Forecasting of Demand in Tourism. Interdisciplinary description of Complex Systems. 2022. № 20 (5). P. 621–639.
4. Li Q., Li X., Du Y., Fan Y., Chen X. A New Sentiment-Enhanced Word Embedding Method for Sentiment Analysis. Applied Sciences. 2022. № 12 (20). P. 712–725.
5. Bagate R. A., Suguna R. Sarcasm Detection with and without #Sarcasm: Data Science Approach. International journal of Information Science and Management. 2022. № 20 (4). P. 1–15.