# Abstract

## The dissertation work of Karyukin Vladislav Igorevich on the topic “The research and development of a module for an intelligent system for analyzing and evaluating the social mood of society in the media space of the Republic of Kazakhstan” submitted for the degree of Doctor of Philosophy (Ph.D.), specialty 6D070300 - Information systems

**Relevance of the work**. The development of Internet technologies has contributed to a significant increase in the number of news sites and social networks describing various events in the world. Posting opinions, thoughts, and ideas about ongoing local and global events on social media has become common. Many social networks, such as Twitter, Facebook, YouTube, and others, remain popular and attract many users. In addition, new platforms like TikTok, Instagram, Pinterest, and others are gaining popularity in social media, covering a massive number of events in the world.

Since the number of news topics and user opinions is growing at an incredibly fast pace, there is a significant need to keep track of the most important topics in various areas of life (for example, politics, economics, civil society, education, healthcare, ecology, culture, and sports, etc.). The volume of facts and opinions shared on social media makes such tracking impossible without automated methods, which has made analytics platforms especially important. The main element of these platforms is the sentiment analysis tool. Although algorithms are not able to fully understand human feelings, emotions, culture, and mentality, they allow you to determine the general trend of public opinion on certain events using analytical tools. Manual analysis is a very long and resource-intensive process, leaving uncertainties and ambiguities. The use of algorithms makes it possible to quickly obtain operational analytics and implement various hybrid approaches: vocabulary, machine learning algorithms, and neural networks.

Currently, there are a number of foreign analytical platforms. Among them, such applications as Sproutsocial, Hubspot, Buzzsumo, Hootsuite, IQBuzz, Brandmention, and Snaplytics stand out. Despite their focus on the business sector, these platforms have similar features, which makes the analysis of socio-political and economic aspects of life underrepresented. These platforms also mainly work with resource-rich languages such as English, Spanish, Italian, French, and others. Texts in Russian and Kazakh have a very limited representation. Despite their diverse functionality, these systems are similar to each other in that they are mainly focused on business goals, leaving important social, economic, and political problems little represented by complex analysis. In addition, most existing platforms focus on resource-rich languages such as English, German, French, Italian, Spanish, and Portuguese. In contrast, texts and comments in resource-poor languages such as Russian and Kazakh are not presented well enough. Therefore, a new information system for monitoring public opinion called the Opinion monitoring system (OMSystem) [29,30], which pays great attention to various topics taking place in the country, was developed to implement the monitoring of the social media space of Kazakhstan. The OMSystem supports leading Kazakh news portals and popular social networks such as Facebook, VKontakte, Instagram, Twitter, and YouTube. A key element of OMSystem is the social sentiment analysis module, which utilizes a data analysis method with the use of sentiment dictionaries, machine learning models, neural networks, and social marketing indicators.

The collected database of 132000 texts from news portals and social networks of Kazakhstan was used during the development of machine learning models for automatic sentiment detection. The texts underwent preprocessing, stemming, the feature extraction using the *tf-idf* metric and the FastText word embedding method and class balancing to obtain the best classification results. At the classification stage, a number of the most popular machine learning algorithms were used (Support vector machine – SVM, Logistic regression – LR, Decision tree – DT, Random forest –

RF, Naive Bayes – NB, k-nearest neighbors – k-NN, and XGBoost) and neural networks (Deep neural networks – DNN, Convolutional neural networks – CNN, and Recurrent neural networks – RNN). The data classification results are presented as summary tables of metrics for evaluating the effectiveness of algorithms: accuracy, precision, recall, and F1-score, curve plots (Area under curve – Receiver operating characteristics – AUC–ROC), and confusion matrices. To analyze the social mood of society, models have been developed using marketing indicators in social networks: the level of interest in the topic in society, the level of activity of the topic's discussion, and the level of social mood.

The effectiveness of the developed models was evaluated by conducting an experiment on the topic of vaccination against Covid-19. The summary analysis presented data on public attitudes toward the vaccination campaign, vaccination policy, and government actions and methods to combat the pandemic. The next step was the development of the electronic Social Mood (eSM) module, which is an application that analyzes data obtained using the OMSystem platform.

**The purpose of the dissertation work** is to develop a method for assessing the social mood of society in the media space of the Republic of Kazakhstan using machine learning models, neural networks, and marketing technologies.

**Research Objectives**:

1. Analysis of architecture and functionality of the intelligent OMSystem.

2. Development of a module for analysis and assessment of the social mood of society in the media space of the Republic of Kazakhstan using machine learning models, neural networks, and marketing technologies of the OMSystem.

3. Evaluation of the developed module using the analysis of the theme of vaccination against Covid-19.

4. Development of an electronic Social Mood (eSM) module that analyzes data obtained using the OMSystem system and evaluates the social mood of society.

**The object of research**: text data, publications, news resources, and social media space of the Republic of Kazakhstan.

**Research methods**: Data mining, Web mining, Natural language processing (NLP), Sentiment analysis (SA), Machine learning, Neural networks, and Marketing technologies of social analytics.

**The theoretical significance of the study**: analysis of the architecture and development of the functionality of the intelligent OMSystem, evaluation of the effectiveness of the social mood assessment module of society.

**The practical significance of the study**: analysis of the social mood of society using the developed module of data processing and analysis of the intelligent system OMSystem.

**The scientific novelty of the research conducted and the results obtained**:

1. A social mood analysis method, characterized by machine learning models and marketing indicators of the user interest in the topic, topic discussion activity, and level of social mood, has been developed.

2. An integrated model of evaluating social mood, including seven attribute machine learning and four deep learning models, has been developed.

3. A sentiment dictionary for the Kazakh language, which is used for an integrated model of social mood analysis, has been developed.

**Decrees for defense**:

1. A developed model of analysis of the social mood of society using machine learning methods and marketing indicators allows the evaluation of various socio-political topics and user responses to governmental campaigns.

2. A developed integrated model of evaluating social mood that includes seven attribute machine learning and four deep learning models.

3. Experimental results on the topic of vaccination against Covid-19 demonstrate public attitudes and government activities through a model of social mood analysis.

**The structure of work**. The dissertation work consists of 156 pages and includes 69 figures and 30 tables. The content includes 6 sections.

**The introduction** describes the relevance, novelty, and main purpose of the dissertation work. A list of the main tasks and objects of the study was given, as well as the theoretical and practical significance of the study.

**The first section** describes the main aspects of information and analytical systems for monitoring social networks, examines in detail foreign and domestic platforms for monitoring and analyzing the social media space, and highlights their advantages and disadvantages. The description of the main methods for determining the sentiment of texts is given: lexicon-based, machine learning-based, and deep learning-based.

**The second section** presents the developed analytical OMSystem platform in detail. It specializes in advanced analysis and monitoring of social networks and news portals of the media space of the Republic of Kazakhstan. In addition, OMSystem includes Russian and Kazakh sentiment dictionaries, machine learning and neural network models, and tools for modeling and determining the social mood and well-being of society. This section also presents the development of models for binary and multiclass classification of texts, which is an essential part of the dissertation work. The results are presented as graphs, summary tables, and conclusions. The development of models based on marketing management methods in social networks, which allows to determine the indicators of the social mood of the society on given topics, is presented.

**In the third section**, an experiment was carried out to analyze the social mood of society regarding vaccination against Covid-19. This topic has gained particular popularity due to the rapid spread of the pandemic in the world. It was actively discussed in news resources and social networks, and thousands of comments were written under posts devoted to this topic. The evaluation of user opinions was performed with the use of sentiment dictionaries, machine learning models, neural networks, and marketing technologies.

**The fourth section** presents the electronic Social Mood (eSM) module developed on the Django Python framework, which is an application that analyzes data obtained using the OMSystem platform. This module performs the following main functions: creating the main categories of topics for analyzing the social mood of society, extracting quantitative data on each of the topics from the OMSystem database, calculating the level of topic activity in society, the level of interest in the topic in society and the level of social mood, visual presentation of the results obtained in the form charts and tables.

**In conclusion**, the theoretical and practical results of this dissertation work are summarized, and its most significant aspects in the analysis of the mood of the society using machine and deep learning methods and indicators of social mood are given.

**Personal contribution of the researcher**. As a result of the work, a detailed analysis of existing platforms for monitoring social networks was carried out. A detailed description of the architecture, functionality, and features of the analytical OMSystem platform, where the study in this work was carried out, was also given. Experimental studies were conducted on developing ML models and NN to determine the sentiment of text data obtained from the work of the web crawler of the analytical system. The eSM module for assessing social mood was also fully developed.

**The degree of validity and reliability of scientific results**. The results of the dissertation were presented in 12 scientific papers, of which 2 articles and 1 chapter in the book were published in journals and book series peer-reviewed in the Scopus database, 4 articles in journals recommended by the Committee for Quality Assurance in Education and Science of the Ministry of

Education and Science Republic of Kazakhstan, and 2 articles in scientific conferences, peer-reviewed in the Scopus database, and 3 articles in the materials of international conferences:

1. Karyukin, V., Mutanov, G., Mamykova, Z. *et al.* On the development of an information system for monitoring user opinion and its role for the public. *Journal of Big Data* 9, 110 (2022). https://doi.org/10.1186/s40537-022-00660-w.
2. G. Mutanov, V. Karyukin and Z. Mamykova, "Multi-class sentiment analysis of social media data with machine learning algorithms," *Computers, Materials & Continua*, vol. 69, no.1, pp. 913–930, 2021. https://doi.org/10.32604/cmc.2021.017827.
3. Mutanov, G., Mamykova, Z., Karyukin, V., Yessenzhanova, S. The Approach to Building a Context-Dependent Sentiment Dictionary. In: Mutanov, G., Serikbekuly, A. (eds) Digital Transformation in Sustainable Value Chains and Innovative Infrastructures. Studies in Systems, Decision and Control, vol 443, 2022. Springer, Cham. https://doi.org/10.1007/978-3-031-07067-9_1.
4. Mutanov G.M., Mamykova Zh. D., Karyukin V.I., Zhaksykeldi A.Zh. Development of a machine-learning algorithm for determining the sentiment of user perception of content. Bulletin of KazNITU Series Technical Sciences, Kazakhstan, 135 (5), 2019.
5. Alimzhanova L.M. Karyukin V.I. A classification model based on decision-making processes. Bulletin of KazNITU Series Technical Sciences, Kazakhstan, 138 (2), 2020.
6. Rakhimova D.R., Turarbek A.T., Karyukin V.I., Karibayeva A.S., Turganbayeva A.O. Overview of modern machine translation technologies for the Kazakh language. Bulletin of KazNITU Series Technical Sciences, Kazakhstan, 141 (5), 2020.
7. Karibayeva A., Karyukin V.I., Turganbayeva A., Turarbek A. The translation quality problems of machine translation systems for the Kazakh language. Journal of Mathematics, Mechanics and Computer Science, Kazakhstan, vol. 111, n. 3, 2021.
8. Vladislav Karyukin, Aidana Zhumabekova, and Sandugash Yessenzhanova. 2020. Machine Learning And Neural Network Methodologies of Analyzing Social Media. In Proceedings of the 6th International Conference on Engineering & MIS 2020 (ICEMIS'20). Association for Computing Machinery, New York, NY, USA, Article 9, 1–7. https://doi.org/10.1145/3410352.3410739.
9. Diana Rakhimova, Vladislav Karyukin, Aidana Karibayeva, Assem Turarbek, and Aliya Turganbayeva. 2021. The Development of the Light Post-editing Module for English-Kazakh Translation. In The 7th International Conference on Engineering & MIS 2021 (ICEMIS'21). Association for Computing Machinery, New York, NY, USA, Article 69, 1–5. https://doi.org/10.1145/3492547.3492651.
10. Karyukin V., Yesenzhanova S. Construction of a context-sensitive sentiment dictionary. International scientific conference of students and young scientists "FARABI ALEMI," Almaty, Kazakhstan, 2020.
11. Karyukin V. An approach to building an eSM application. International scientific conference of students and young scientists "FARABI ALEMI," Almaty, Kazakhstan, 2020.
12. Karyukin V. Multiclass classification with the use of machine learning algorithms. International scientific conference of students and young scientists "FARABI LEMI," Almaty, Kazakhstan, 2021.

**Connection of the dissertation with research work**. This study was carried out as part of the project for the commercialization of the results of scientific and (or) scientific and technical activities "Opinion Monitoring Information System OMSystem (Opinion monitor system)," 0101-18-GK. (The main role of the PhD student was to develop a module for analyzing and evaluating social mood, machine learning models, and neural networks, conducting an experiment on analyzing social mood on the topic of vaccination against Covid-19, and developing an electronic Social Mood module).