ABSTRACT

of the dissertation work by Turarbek Asem Turarbekkyzy on the topic "Research and development of a model for determining the seismic activity of Kazakhstan based on machine learning methods", submitted for the degree of Doctor of Philosophy (PhD) in the specialty 6D070300 – Information systems

The relevance of the work. In the second half of the twentieth century, the number and strength of natural disasters increased. In recent decades alone, a significant number of strong destructive earthquakes have occurred all over the world (in Chile, Peru, Mexico, the USA, Japan, Armenia, Sakhalin, Afghanistan, Turkey, etc.), which caused enormous damage and considerable loss of life. Earthquakes occupy one of the first places among other types of natural disasters. Earthquakes lead to massive casualties and destruction. Earthquakes can cause significant economic damage to countries and regions, especially those that are economically dependent on specific geographical areas. The modern world is closely connected, and a seismic event in one country can have an impact on international supply chains and the economies of other countries.

Natural phenomena, including earthquakes, are inevitable. They cannot be prevented, but their destructive impact can and should be reduced. To do this, you need to know the causes of earthquakes, assess seismic hazards, study the processes associated with their preparation and occurrence, develop methods for predicting these phenomena, identify possible sources of seismic vibrations. Earthquakes account for 13% of the total number of natural disasters. Seismic phenomena such as earthquakes are the result of complex geological and geophysical processes that constantly occur inside the earth's crust. The study of seismic activity and its consequences is of great importance for ensuring occupational safety and employment. In Kazakhstan, as in other Central Asian countries, seismic phenomena pose a serious threat, and their understanding plays a crucial role in preparing and reducing risk. A number of regions of Kazakhstan are subject to seismic risks due to their geological structure and location. In Kazakhstan, by 2023, more than 7 million people live in the most earthquake-prone regions, more than 50% of the industrial potential is concentrated, and more than 400 cities and towns are located.

Seismic activity, which is regularly observed on the territory of Kazakhstan, causes a serious need for the development of effective methods for predicting seismic phenomena.

Seismic activity can pose a threat to human settlements and key sectors of the economy, such as mining, industry and tourism. In the history of the country, cases of devastating earthquakes have been recorded, resulting in significant human and material losses. Forecasting seismic activity through the use of machine learning methods can become a powerful tool for preventing and minimizing the consequences of natural disasters. Forecasting seismic activity can help in the development of building codes and regulations to ensure the safety of buildings and infrastructure. The study of seismic activity can also contribute to the

development of science and technology in the country, as well as the training of specialists in the field of geophysics and seismology.

Currently, there are a number of systems and monitors. Among them, the USGS, EMSC, IRIS stand out, they display information about seismic activity around the world, as well as the National Data Center of Kazakhstan (NDC), the Institute of Seismology in Kazakhstan, Darmen provide information on seismic events in Kazakhstan.

These systems have a similar function, they visualize earthquake epicenters, show a complete description, that is, date, time, magnitude, depth, and indicate the past event on the map. In machine learning models, a collected database of 89629 seismic events from the Institute of Seismology and USGS data was used to determine seismic activity. The data has passed the preprocessing stages. At the classification stage, a number of the most popular machine learning algorithms (Support vector machine – SVM, Logistic regression – LR, Decision tree – DT, Random forest – RF, k-nearest neighbors – k-NN, and XGBoost) and neural networks (Deep neural networks – DNN, Convolutional neural networks – CNN) were used. The classification results are presented in the form of summary tables of metrics for evaluating the effectiveness of algorithms: correctness (accuracy), precision (precision), completeness (recall) and F-measure (F1-score), graphs of curves (Area under curve – Receiver operating characteristics – AUC-ROC). To analyze seismic activity, a model has been developed using historical data from Kazakhsatan seismic events.

Machine learning has developed rapidly in the last decade and promises to significantly change and improve the functions of big data analysis in a wide variety of fields. Compared to traditional methods, machine learning provides significant advantages in solving complex problems, computing performance, spreading uncertainty and processing it, as well as decision support. All the methods used have been trained and tested on real data collected in Kazakhstan over the past 97 years, from 1906 to 2022. The proposed model surpassed other models in terms of precision, recall, accuracy, f-score and AUC-ROC by 63%, 82,4%, 87,4%, 62,7% and 83%, respectively. Based on the results, it can be concluded that the proposed Conv2D model is effective for determining seismic activity.

The purpose of the dissertation is to create a model and an information system for determining the seismic activity of the Republic of Kazakhstan based on seismological data using machine learning methods.

Research objectives:

1. Research of existing information technologies and systems for determining seismic activity and data processing;
2. Research and analysis of algorithms, machine learning methods used to detect seismic activity;
3. Creation of an algorithm for determining the seismic activity of the territory of Kazakhstan;
4. Creation and training of a deep learning model for detecting seismic activity based on features (feature);

5. Development of an information system for detecting seismic activity.

The object of the study is the seismic activity of the territory of Kazakhstan.

Subject of research: Algorithms and machine learning methods for determining seismic activity.

Research methods: Machine learning methods, neural networks, deep neural networks, statistical methods, machine learning algorithms.

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

1. The algorithm of the method for determining the seismic activity of the territory of Kazakhstan has been developed;

2. A deep learning model has been developed to determine seismic activity based on signs;

3. An information system for determining seismic activity has been developed based on the created deep learning model.

Theoretical and practical significance of the study: The study allows to expand the understanding of the possibilities and limitations of using machine learning algorithms in determining seismic activity. The developed model and methodology can be integrated into existing seismic activity monitoring and forecasting systems, increasing their accuracy and timeliness.Объем и структура работы.

The dissertation work consists of an introduction, 4 chapters and a conclusion. The total volume of the dissertation is 102 pages, 36 figures, 15 tables. The list of references consists of 79 titles.

The introduction describes the relevance, purpose, objectives, objects and methods of research, theoretical and practical significance, as well as the novelty of the dissertation work.

The first section describes the seismic regime of Kazakhstan, examines in detail the methods of analysis and prediction of earthquakes. A review and analysis of earthquake monitoring and forecasting systems is presented. An experiment was conducted to determine earthquake damage using the ITRIS GIS.

The second section a description of the main types of machine learning models and neural networks is provided. An analysis and review of the results of other authors using machine learning and neural networks in seismology and determination of seismic activity is presented.

In the third section, the development of a model based on machine learning methods and neural networks is presented in detail, which is the most important part of the dissertation work. An algorithm for developing a model is presented, the target event and attributes are defined, and the stage of collecting and preprocessing seismological data is shown. An assessment was made of the effectiveness of using the proposed model to determine the seismic activity of the territory of Kazakhstan.

The fourth section presents a system developed on the Django Python framework. The system performs the following main functions: creation of the main categories of seismic events, extraction of data by region and year, visual representation of the results in the form of epicenters on the map of Kazakhstan and summary tables.

In conclusion, the theoretical and practical results of this dissertation are summarized, its most significant aspects in the analysis of seismic activity using machine learning models, neural networks and indicators of seismic events are presented.

The main position to be defended. A new model and information system for determining seismic activity based on machine learning. The results of seismic activity prediction experiments confirming the effectiveness of the proposed model.

Personal contribution of the researcher. The dissertation conducted an analysis of existing earthquake monitoring systems and applications; developed an algorithm for determining the seismic activity of the territory of Kazakhstan; developed a deep learning model for determining seismic activity based on signs; developed an information system for determining seismic activity based on the created deep learning model.

The degree of validity and reliability of scientific results. The results of the dissertation were presented in 13 scientific papers, of which 2 articles were published in the journal peer–reviewed in the Scopus database, 3 articles – in journals recommended by the Committee for Quality Assurance in Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan, and 8 articles - in the materials of international conferences.

Scientific articles in journals indexed in the Scopus database:

1. Turarbek Assem, Adetbekov Yeldos, Bektemesov Maktagali 2-D Deep Convolutional Neural Network for Predicting the Intensity of Seismic Events // International Journal of Advanced Computer Science and Applications vol.14(1), pp.788-796, 2023 (IF=0.675, CiteScore rank =44, Q3 Scopus).
2. Turarbek Assem, Bektemesov Maktagali et al. Deep Convolutional Neural Network for Accurate Prediction of Seismic Events // International Journal of Advanced Computer Science and Applications vol.14(10), pp.604-613, 2023 (Q3, Scopus indexed, процентиль 47)

Articles recommended by the Committee for Quality Assurance in Education and Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan:

1) Turarbek A.T. Geoinformation systems in seismology // Bulletin of KazNITU, Series "Earth Science". 2016.No.5.pp.51-53.
2) Turarbek A.T. Using the ITRIS geoinformation system to assess the consequences of earthquakes in Kazakhstan // Bulletin of KazNITU, Series "Technical sciences". –2017. –№3. – Pp.394-399.
3) Turarbek A.T., Sadykova A.B. The current state of earthquake analysis and forecasting in Kazakhstan // Bulletin of KazNITU, Series "Technical Sciences". – 2018. –No.2. – pp.184-191.

Articles in the materials of the international conference:

1) Krivorotko O.I., Kabanikhin S.I., Bektemesov M.A., Sadykova A.B., Turarbek A.T. Geoinformation system of Kazakhstan. Mathematical models of the geoinformation system // Proceedings of the international scientific

conference "Marchukov scientific readings 2017", Russia, Novosibirsk, 2017. – pp.455-462.

2) Turarbek A.T. Application of the ITRIS geoinformation system for modeling and forecasting the seismic regime of Kazakhstan // Proceedings of the international scientific conference "Marchukov scientific readings 2017", Russia, Novosibirsk, 2017. – pp.98-99.

3) Turarbek A.T. On image quality methods for remote monitoring of earthquakes // Materials of the International scientific and methodological Conference "Mathematics in Kazakhstan - past and prospects" dedicated to the 100th anniversary of Ibrashev Hassan Ibrashevich Almaty, 2016, pp.115-117.

4) Turarbek A.T. Applications of remote sensing data when working with Geoinformation systems for earthquake forecasting // Materials of the international scientific conference of students and young scientists "Mir Farabi", Almaty, 2016, p.237.

5) Kenes A., Turarbek A.T. The components of the geoinformation system of earthquakes // Materials of the international scientific conference of students and young scientists "Mir Farabi", Kazakhstan, Almaty, 2017. – P.193.

6) Turarbek A.T., Adetbekov E.N. Kazakstannyn zher silkinisi monitoring zhurgizu usin Python tilin koldanu // Materials VIII International scientific and methodological conference "Mathematical modeling and information technologies in education and science", dedicated to the 90th anniversary of KazNPU named after Abai. Almaty, 2018. –pp.109-111.

7) Turarbek A.T., Adetbekov E.N. Machine learning in earthquake forecasting //Materials of the international scientific conference "Inverse and incorrect problems in natural science", Almaty, 2023 –pp.44-45.

8) Turarbek A.T. Deep machine learning in determining seismic activity//Materials of the international scientific conference "Inverse and incorrect problems in natural science", Almaty, 2024 –pp.56-57.