DOI: 10.37943/21KUXZ6354

Gulnur Kazbekova
Candidate of Technical Sciences, Associate Professor, Head of the Department of Computer Engineering
gulnur.kazbekova@ayu.edu.kz, orcid.org/0000-0002-2756-7926
Khoja Akhmet Yassawi International Kazakh-Turkish University, Kazakhstan
Arypzhan Aben
Master of Science, Department of Computer Engineering
arypzhan.aben@ayu.edu.kz, orcid.org/0000-0001-8534-3288
Khoja Akhmet Yassawi International Kazakh-Turkish University, Kazakhstan
Anuarbek Amanov
PhD, Senior Lecturer, Department of Computer Engineering,
anuarbek.amanov@ayu.edu.kz, orcid.org/0000-0003-0638-6859
Nurseit Zhunissov
PhD, Senior Lecturer, Department of Computer Engineering
nurseit.zhunissov@ayu.edu.kz, orcid.org/0000-0001-6531-9408
Khoja Akhmet Yassawi International Kazakh-Turkish University, Kazakhstan
Aiman Abibullayeva
PhD, Senior Lecturer, Department of Computer Engineering
aiman.abibullayeva@ayu.edu.kz, orcid.org/0000-0003-2449-2540
Khoja Akhmet Yassawi International Kazakh-Turkish University, Kazakhstan

EFFECTIVENESS OF MACHINE LEARNING METHODS IN DETERMINING EARTHQUAKE PROBABLE AREAS: EXAMPLE OF KAZAKHSTAN

Abstract: This study investigates the effectiveness of machine learning methods in identifying earthquake-prone areas in Kazakhstan and its neighboring regions. By leveraging a comprehensive dataset encompassing significant earthquake data from 1900 to 2023, various machine learning algorithms were employed, including RandomForest, GradientBoosting, Logistic Regression, Support Vector Classification (SVC), K-Nearest Neighbors (KNeighbors), Decision Tree, XGBoost, LightGBM, AdaBoost, and MLPClassifier. The primary objective was to analyze and compare the performance of these models in predicting earthquake magnitudes and frequencies. The results reveal that certain algorithms significantly outperformed others in terms of accuracy, underscoring the potential of machine learning techniques to enhance earthquake prediction capabilities. Notably, XGBoost and RandomForest demonstrated the highest predictive accuracy, suggesting their suitability for application in seismic risk assessment. These findings offer valuable insights for governmental agencies engaged in disaster management and prevention planning, highlighting the practical implications of integrating advanced analytical techniques in their strategies. In addition to model performance analysis, a visual heatmap was generated to illustrate the geographical distribution of earthquake occurrences across the studied regions. This visual representation effectively identifies high-risk areas, serving as a crucial tool for local authorities and researchers in making informed decisions regarding safety measures and emergency preparedness. This research contributes to the expanding body of knowledge on earthquake prediction utilizing machine learning, emphasizing the necessity for continuous improvement in predictive models by incorporating additional environmental and geological factors. The implications of these findings extend beyond academic discourse, holding significant potential for enhancing public safety in regions vulnerable to seismic activity. As such, this study advocates for the integration of machine learning methodologies in disaster management frameworks to mitigate risks and enhance preparedness in earthquake-prone regions.

Keywords: earthquake prediction, machine learning, Kazakhstan, Seismic Risk Assessment, predictive modeling, disaster management, Geographical Heatmap, algorithm performance.

Introduction

An earthquake is a natural phenomenon that directly affects the safety and economic stability of humanity. The earthquake that occurred in Turkey and Syria in February 2023 [1] and the earthquake in Almaty on January 23, 2024 [2] highlighted the need for thorough research in this field. As the Republic of Kazakhstan is located in seismically active regions, earthquakes pose a significant threat to society. Since 2010, over 1,000 earthquakes have been recorded in the country, with the majority observed in Almaty, East Kazakhstan, and the Zhetysu regions. In this context, identifying and predicting potential earthquake zones is of particular importance.

In recent years, machine learning methods have been widely used in the field of seismology. The advantages of machine learning in data analysis, forecasting, and modeling processes are evident. Processing large volumes of data, identifying patterns, and making effective predictions are the main features of these methods [3]. Through machine learning techniques, effective solutions can be proposed for identifying potential earthquake zones, predicting their timing, and planning safety measures.

Research conducted in Kazakhstan has demonstrated the potential of machine learning methods in earthquake prediction [4]. Seismic data, geophysical information, and climate change are the key components for building machine learning models. The issues of identifying and predicting potential earthquake zones are not only scientifically significant but also critically important in vital sectors. The practical benefits of this research are especially reflected in construction, emergency management, and environmental protection.

In the construction sector, accurately predicting potential earthquake-prone areas will significantly improve the planning and construction processes of buildings and infrastructure. The use of modern machine learning methods allows for the preliminary identification of regional seismic activity, which in turn facilitates the introduction of high seismic resilience construction standards. Especially in frequently shaken areas like Almaty, tightening construction norms will help prevent future losses and ensure the safety of people [5].

The direct impact of this research on human life safety is also substantial. By predicting earthquakes, early warning systems for emergencies can be enhanced. Such systems provide additional minutes for people to move to safer locations by timely notifying them, which plays a crucial role in saving lives, especially in densely populated areas. Furthermore, the effectiveness of emergency services and evacuation measures increases, and the recovery process after disasters can be executed more quickly and systematically.

Additionally, this research has a significant impact on urban planning and development. By identifying earthquake-sensitive areas in advance, it is possible to promote sustainable urban development, strengthen safety measures, and reduce construction activities in high seismic zones. This approach will become an essential tool for managing settlements that are prone to seismic activity in the future.

Regarding the environment, identifying frequently occurring earthquake areas can help prevent ecological disasters and natural calamities. Earthquakes pose threats not only to humanity but also to the natural landscapes, forests, and rivers of the earth [6]. The findings of this research will enable the implementation of measures that can prevent natural disasters and mitigate their impact on the environment. In this regard, this study examines the effectiveness of machine learning methods in identifying potential earthquake zones in Kazakhstan. The main goal of the article is to assess seismic hazards through machine learning techniques, identify potential zones, and predict future earthquakes.

Utilizing machine learning methods allows for a comprehensive analysis of the country's seismic situation, the identification of risks, and the automation of prediction processes. This, in turn, is essential for protecting the population and infrastructure and effectively organizing safety measures by the government [7]. Additionally, this research aims to systematically review the results of scientific work conducted to identify potential earthquake zones in Kazakhstan and explore ways to utilize the collected data effectively.

The hypothesis underlying this research posits that advanced ensemble learning models, such as XGBoost and LightGBM, will demonstrate higher predictive accuracy and robustness in identifying earthquake-prone areas compared to other machine learning methods, given their ability to handle complex, non-linear relationships within the data. If validated, this hypothesis could support the application of sophisticated ML models in geophysical research and emergency preparedness.

The primary contributions of this study are threefold. First, it offers a comparative analysis of multiple machine learning models specifically applied to seismic prediction in Kazakhstan, providing insights into model performance in a unique geographical context. Second, it contributes to the field by demonstrating how machine learning can be used to map potential earthquake zones, thereby enhancing predictive modeling techniques in seismology. Third, the research provides practical insights for regional policymakers and engineers, who can use these predictive models to improve building codes, develop early warning systems, and establish more resilient urban infrastructure. By combining machine learning with traditional seismological data, this study lays the groundwork for more sophisticated, data-driven approaches to earthquake risk mitigation in Kazakhstan and similar seismically active regions.

Literature Review

The research on earthquakes in Kazakhstan has significantly increased in recent years, particularly with the noteworthy application of machine learning methods. Researchers are employing various methods and technologies to identify and predict potential earthquake-prone areas. Karmenova and colleagues [8] proposed effective ways to process real-time data using machine learning algorithms for clustering seismic events. This research introduced a new perspective on the accumulation and analysis of data in the field of seismology. The data analysis resulted in the proposal of effective models for determining seismic hazards. Nurtas and others [9] explored the use of volumetric statistical data for earthquake prediction in their work. Their results particularly demonstrated the effectiveness of machine learning methods compared to traditional approaches. This method is vital for predicting the frequency of seismic events and their potential impacts. Turarbek and colleagues [10] conducted a study using deep convolutional neural networks (CNN) to predict the intensity of earthquakes. This technology opened new possibilities for analyzing the dynamics of earthquakes and their geographical distribution. Baktibayev and colleagues [11] investigated models for earthquake prevention and prediction using natural language processing (NLP) methods. Their work highlighted the importance of using modern technologies in managing the subsequent effects of earthquakes. Karmenova and others [12] performed a seismic assessment of urban buildings using data analysis methods. This study indicated the need to introduce new standards in the construction sector for assessing the potential consequences of earthquakes. Amey and colleagues [13] revealed data about buried faults beneath the city of Almaty using high-resolution satellite DEM. This research provides crucial information for assessing seismic safety in the city. Turarbek and colleagues [14] achieved more accurate predictions by employing deep convolutional neural networks for earthquake prevention. Their research results underscore the necessity of enhancing data quality and the accuracy of algorithms. Yavuz and colleagues [15] used discriminant functions and tree-based machine learning algorithms to identify the sources of seismic events in Turkey's Soma region. The results of this work demonstrated the effectiveness of new algorithms in predicting the occurrence of earthquakes. Li and others [16] studied the interaction between earthquakes and landslides, raising issues of seismic risk in Central Asia. Their findings were significant in identifying the links between seismic hazards and environmental changes. Ahmed-Zaki and colleagues [17] proposed the development of a web application for visualizing urban disasters. This work highlighted the importance of utilizing visual tools in understanding the consequences of earthquakes. Kim and colleagues [18] introduced new methods for classifying signal-to-noise ratios in microseismic data using machine learning. Their work significantly contributed to improving the prediction capabilities for earthquakes. Overall, machine learning methods are effective in identifying potential earthquake-prone areas in Kazakhstan. Each study addresses relevant issues and positively impacts the development of the scientific community.

Objectives and Tasks of the Research

The aim of the research is to assess the effectiveness of applying machine learning methods to identify and predict earthquake-prone areas in Kazakhstan. The study is directed towards reducing the consequences of earthquakes and improving seismic safety measures.

The research tasks include:

1. Investigating seismic-active areas in Kazakhstan and collecting data.
2. Developing and testing machine learning models for earthquake prediction.
3. Comparing different machine learning methods and identifying the most effective ones.
4. Visualizing the model results to create a map of areas at risk of earthquakes.

Methods and Materials

This section of the research consists mainly of five stages: data collection, data preprocessing, data transformation, performance evaluation of models, and creating a visual map of potential earthquake-prone areas. Below, Figure 1 provides a description of the research stages.

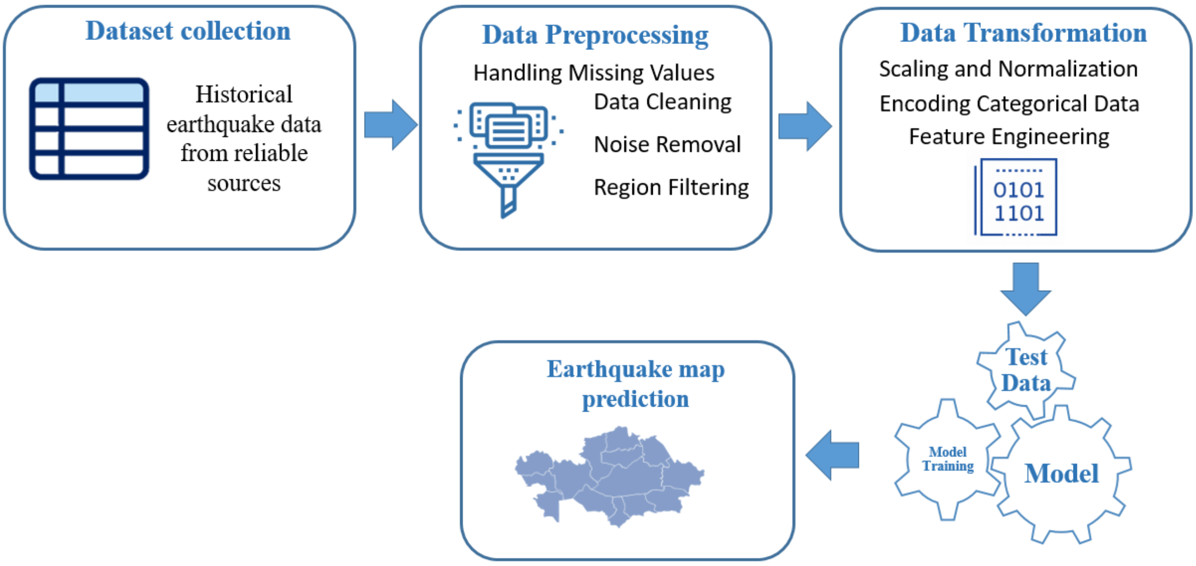

Figure 1. Research Stages

A. Data Collection

The dataset used in this study comprises significant earthquake data from 1900 to 2023, representing a comprehensive collection of major earthquakes worldwide over the past 123 years. This dataset is processed and maintained by the United States Geological Survey (USGS) National Earthquake Information Center (NEIC) and is continuously updated to provide the most accurate and current information on earthquake events.

Each record in the dataset, which includes over 37,000 earthquakes, contains information such as the date, time, location, magnitude, and depth of the earthquake [19].

This dataset serves as an invaluable resource for seismologists, geologists, and other researchers studying earthquakes, as well as for emergency management personnel and professionals engaged in disaster response and preparedness[20], [21]. Additionally, it acts as a valuable tool for the general public interested in understanding the history of earthquakes and their impacts on human society. The dataset is illustrated in Table 1 below.

Table 1. Dataset of Significant Earthquakes from 1900 to 2023

Place	Latitude	Longitude	Depth	Mag	gap	rms	Updated
116 km SSE of Sosnovka, Kyrgyzstan	41.7307	74.6053	21.53	5.5	41.0	0.72	2022-08-01T22:15:44.480Z

38 km ESE of Osh, Kyrgyzstan	40.3762	73.2038	18.0	5.6	18.0	0.73	2022-08-01T18:23:18.294Z
0 km NNE of Lugovoy, Kazakhstan	40.081	71.41	20.0	6.1	32.7	0.9	2022-04-08T21:55:10.868Z
31 km SSE of Osh, Kyrgyzstan	40.288	72.985	6.0	5.6	35.1	0.94	2022-07-13T23:20:45.583Z

B. Machine Learning Models

In this study, the following machine learning algorithms were employed for earthquake prediction:

1. *Random Forest:* This model constructs multiple decision trees and makes predictions based on their collective outcomes. It primarily ensures maximum accuracy by utilizing various subsets of the data.
2. *Gradient Boosting:* An ensemble learning method where each new model attempts to reduce the errors of the previous models. This model improves results by gradually correcting errors.
3. *Logistic Regression*: A simple and interpretable model used to predict relationships between two or more classes. In this study, it is applied to predict the probability of an earthquake.
4. *SVC (Support Vector Classifier):* One of the vector methods aimed at identifying boundaries between different classes. This method aids in predicting potential earthquake regions by finding the boundary between data.
5. *KNeighbors:* This algorithm analyzes each object by comparing it to its nearest neighbors. It is one of the suitable models for predicting areas prone to earthquakes.
6. *Decision Tree:* The decision tree makes decisions based on several potential features at each point, which is utilized for earthquake prediction.
7. *XGBoost:* One of the precise and efficient models that analyzes and predicts the complex structure of the data using boosting methods.
8. *LightGBM:* A lightweight and fast model optimized for handling large volumes of data. This model is also effective in earthquake prediction tasks.
9. *AdaBoost:* An adaptive boosting method where each new model attempts to correct errors from previous models.
10. *MLPClassifier:* A multilayer perceptron model capable of learning and predicting complex relationships in data using neural networks.

All models are utilized to evaluate the effectiveness of earthquake prediction, and their results are comparatively analyzed.

Heat Map Visualization of Seismic Activity in Kazakhstan Using Folium Library

The heat map, developed based on seismic data from earthquakes within Kazakhstan, enables the visualization of regional seismic activity. Python's Folium library was utilized for this visualization, with the underlying mathematical model based on geographic coordinates and earthquake magnitude data. This approach allows for evaluating the spatial distribution of energy and the impact level across affected areas.

Mathematical Model for Earthquake Prediction

Earthquake prediction is a critical challenge in geophysics and disaster management. This study aims to forecast earthquake magnitudes and locations using machine learning techniques, leveraging prior research on geospatial modeling [23]. The problem is formulated as a classification task, where the model predicts the seismic class (i.e., magnitude categories) based on geological and environmental features such as latitude, longitude, depth, and historical seismic activity. Following the methodology outlined by Bisarinova, define the dataset as:

$$D = \{(x_i, y_i)\}_{i=1}^{N} \tag{1}$$

- $x_i = R^d$ is the feature vector for the i-th earthquake instance, which includes parameters such as *latitude (lat), longitude (lon), depth (depth),* and other environmental features.
- $y_i = R^m$ is the target variable, representing the earthquake magnitude class (e.g., small, medium, large) or risk level.
- *N* is the total number of instances in the dataset, and ddd is the number of features used.

Feature Representation

Let the feature vector x_i be defined as:

$$x_i = [lat_i, lon_i, depth_i, mag_{i-1}] \tag{2}$$

where each component represents the seismic and geospatial characteristics of an earthquake, aligning with the geospatial modeling approach discussed in Bisarinova. This structure enables the model to capture temporal dependencies and spatial distribution patterns that are crucial for accurate earthquake forecasting.

The integration of these features follows principles from geoinformatics, as highlighted in Bisarinova's dissertation, which emphasizes the importance of multi-parameter data analysis in predictive modeling. By leveraging such a framework, the proposed machine learning model aims to improve earthquake prediction accuracy and contribute to early warning systems

Model Formulation

The machine learning model f maps the input feature vector x_i to the target variable y_i , i.e., the magnitude class or risk level:

$$y_i = f(x_i, \theta) \tag{3}$$

where θ represents the parameters of the model (which are learned during training). The choice of model f depends on the machine learning algorithm used. For instance:

1. **Random Forest / Decision Trees**: These algorithms model the decision process using tree-like structures. The output is obtained by traversing the tree based on the feature values.
$$y_i = f_{RF}(x_i) \tag{4}$$
2. **Gradient Boosting / XGBoost**: These are ensemble methods that use boosting techniques to combine multiple weak learners (typically decision trees) into a strong learner.
$$y_i = f_{GB}(x_i) \tag{5}$$
3. **Support Vector Machine (SVC)**: The decision boundary is constructed in a high-dimensional feature space using the kernel trick.
$$y_i = sign(w^T \phi(x_i) + b) \tag{6}$$
where $\phi(x_i)$ is the feature mapping, w is the weight vector, and b is the bias term.

Training Objective

To train the model, the objective is to minimize a loss function L that measures the difference between the predicted values and the true labels across all instances. Common loss functions include:

For classification (e.g., earthquake magnitude classes):

Cross-entropy loss (used for categorical classification):

$$L_{CE}(y_i, \hat{y}_i) = -\sum_{k=1}^{K} y_{i,k} \log(\hat{y}_{i,k}) \tag{7}$$

where $\hat{y}_{i,k}$ is the predicted probability of class k, and $y_{i,k}$ is the true label (one-hot encoded).

For regression (e.g., predicting continuous magnitude):

Mean squared error (MSE):

$$L_{MSE}(y_i, \hat{y}_i) = \frac{1}{N} \sum_{i=1}^{N} (y_i, \hat{y}_i)^2 \tag{8}$$

The performance of the trained model is evaluated using various metrics, which include:

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^{N} I(y_i = \hat{y}_i) \tag{9}$$

where I is the indicator function that equals 1 if $y_i = \hat{y}_i$, and 0 otherwise.

Precision, Recall, and F1-Score*:* These metrics are used to evaluate the balance between false positives and false negatives for each class[24].

$$\text{Precision} = \frac{TP}{TP+FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{11}$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

where TP, FP, and FN refer to true positives, false positives, and false negatives, respectively.

This mathematical formulation demonstrates the steps involved in predicting earthquake magnitudes and locations using machine learning models. The models use features such as geographic coordinates and previous earthquake data to classify or predict seismic events. The model parameters are optimized to minimize the loss function, and various evaluation metrics are used to assess the model's performance.

Each seismic event is identified by its coordinates latitude (φ) and longitude (λ) along with its respective magnitude. The magnitude value determines the radius of influence on the heat map, where higher-magnitude events are represented with a wider radius of effect.

Geographic coordinates are projected onto a 2D plane, allowing operations to be performed in the WGS84 coordinate system, which supports computational convenience for the map. This projection makes it easier to calculate energy distribution at each point on the heat map.

The intensity at each coordinate is calculated using a Gaussian Kernel function, often applied for such visualizations to model heat intensity as follows:

$$I(x, y) = e^{-\frac{d^2}{2\sigma^2}} \tag{13}$$

where d is the distance to the seismic event location, and σ represents an influence radius proportional to the event's magnitude. This function enables the intensity of energy across the heat map to be evaluated at each point [25].

Through overlapping the influence of each point, the model calculates the cumulative heat intensity across all coordinates. The resulting heat map uses a color scale to illustrate areas of high and low seismic activity: regions with more events appear brighter, while sparser regions appear darker. Folium's HeatMap function processes this model and saves the visualization as an HTML file.

The resulting heat map serves as a basis for analyzing and forecasting seismic activity within Kazakhstan, highlighting high-risk areas for further investigation.

Results

This section encompasses data collection, data preprocessing, data transformation, model performance evaluation, and the creation of a visual map of potential earthquake regions. First, the results of the machine learning models will determine the appropriate model for the research, while the section on the visual map of potential earthquake regions will create a visual representation by analyzing earthquake epicenters in Kazakhstan and nearby areas.

Results of Machine Learning Models

During the research, several machine learning models were selected, namely Random Forest, Gradient Boosting, Logistic Regression, SVC, KNeighbors, Decision Tree, XGBoost, LightGBM, AdaBoost, and MLPClassifier. The stages of data collection, preprocessing, transformation, and model performance evaluation were detailed in the methods section.

Initially, during the data collection phase, data related to earthquakes in Kazakhstan and nearby regions was compiled. Subsequently, in the data preprocessing phase, defective data was removed, and usable data was obtained. In the data transformation phase, the information was formatted to be compatible with the models.

As a result, the performance metrics of the models are presented in Table 2. This table includes performance indicators such as accuracy, precision, recall, and F1-score for each model. The performance of the models plays a crucial role in assessing the research outcomes and selecting prediction methods.

Table 2. Results of machine learning models.

Model	Accuracy	Precision	Recall	F1-Score
RandomForest	0.731800	0.724498	0.731800	0.727133
GradientBoosting	0.732070	0.716564	0.732070	0.714987
LogisticRegression	0.681055	0.634783	0.681055	0.619052
SVC	0.723410	0.708056	0.723410	0.709615
KNeighbors	0.700812	0.691851	0.700812	0.695277
DecisionTree	0.681732	0.687200	0.681732	0.684205
XGBoost	0.745737	0.736242	0.745737	0.738447
LightGBM	0.739783	0.728956	0.739783	0.731135
AdaBoost	0.716103	0.697694	0.716103	0.698002
MLPClassifier	0.729635	0.717127	0.729635	0.719530

If we pay attention to the performance of the models shown in the table, the XGBoost model showed the best results, where the accuracy (Accuracy) was 0.745737, the precision (Precision) was 0.736242, the recall (Recall) was 0.745737, and the F1-calculator was 0.738447. This shows that the model processes the data effectively and is suitable for solving the problem under study.

The LightGBM model was also distinguished by high performance indicators, but slightly lower than the XGBoost model. This refers to the speed and efficiency of the model in working with data. The GradientBoosting model also offered a high level of performance, but was slightly behind the LightGBM and XGBoost models.

The SVC, RandomForest, and MLPClassifier models also perform well, but their results are lower than those of the XGBoost and LightGBM models. The KNeighbors and AdaBoost models showed relatively average performance. Finally, the DecisionTree and LogisticRegression models showed the lowest results, indicating that they did not fit the complexity of the data.
Thus, the XGBoost and LightGBM models were selected as proposed algorithms for research purposes, as they provide efficient and accurate data processing. The results of the study demonstrate the effectiveness of machine learning algorithms in earthquake prediction, as well as open opportunities for further research and model optimization in this field in the future.

Visual map of earthquake prone areas.

The geographical location of Kazakhstan is of particular importance for the study of seismic activity. The territory of Kazakhstan is located approximately between 40°-55° latitude and 47°-87° longitude coordinates. Taking into account these geographical boundaries, there is a need to determine the territory of Kazakhstan and the adjacent regions using the coordinates (latitude and longitude) of the earthquakes.

In the course of the research, the coordinates of earthquakes are used from the dataset for the purpose of studying the seismic activity of Kazakhstan and nearby regions. As a result, the number of earthquakes belonging to this territory is 281.

This indicator is an important data for determining the seismic safety of Kazakhstan and potential earthquake zones. Analysis of the number and geographical location of earthquakes also allows predicting the possible influence of earthquakes in these regions in the future by using machine learning methods [22].

For a deeper understanding of the seismic activity of Kazakhstan, it is necessary to pay attention to the magnitude and frequency of earthquakes. Figure 2 presents a diagram showing earthquakes by magnitude and frequency in Kazakhstan and nearby regions.

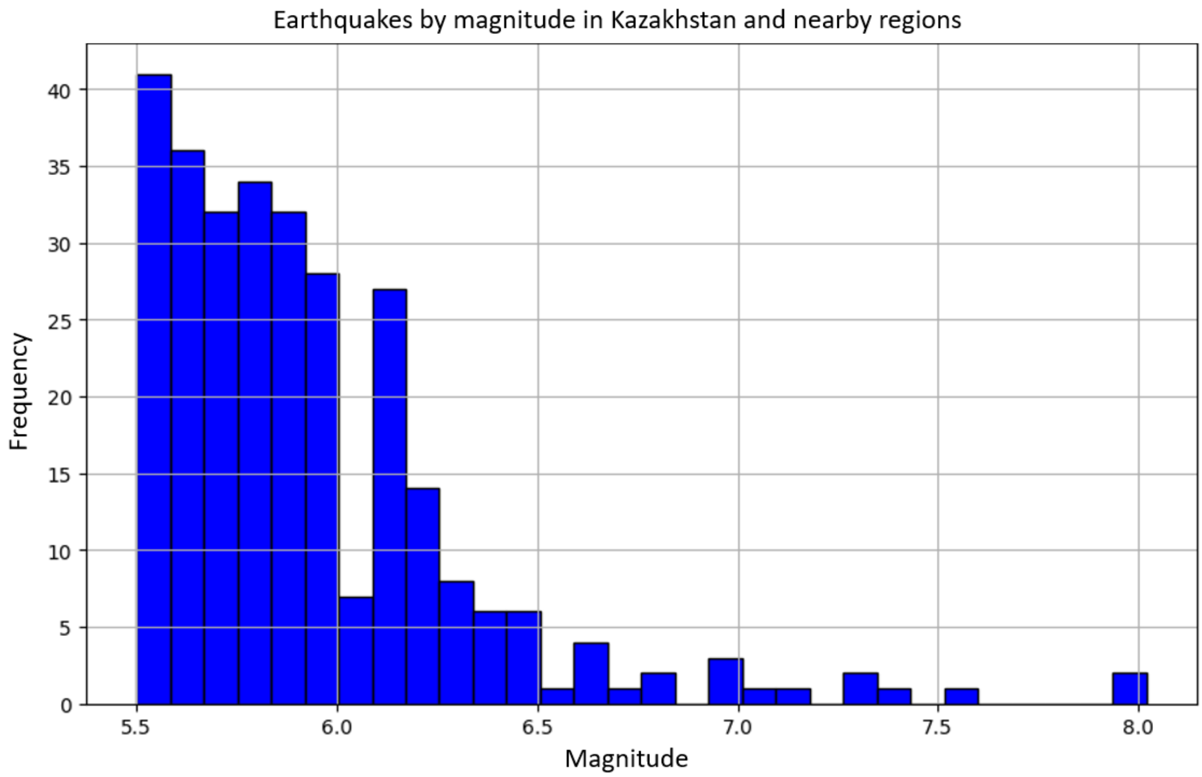

Figure 2. Earthquakes by magnitude in Kazakhstan and nearby regions.

In this diagram, you can see the relationship between the different magnitude levels of earthquakes and their frequency. The frequency diagram shows the number and distribution of high-magnitude seismic events, as well as the frequency of low-magnitude earthquakes.

This relationship between magnitude and frequency plays an important role in estimating and predicting seismic activity. The results of the research will be used to effectively assess the impact of earthquakes in Kazakhstan, to plan preventive measures and to improve standards in the field of construction.

The graph showing the number of earthquakes over time in Kazakhstan and nearby regions also showed the change in the number of earthquakes by year. Figure 3 (specify the figure number) presents a chart showing the frequency of seismic events by year.

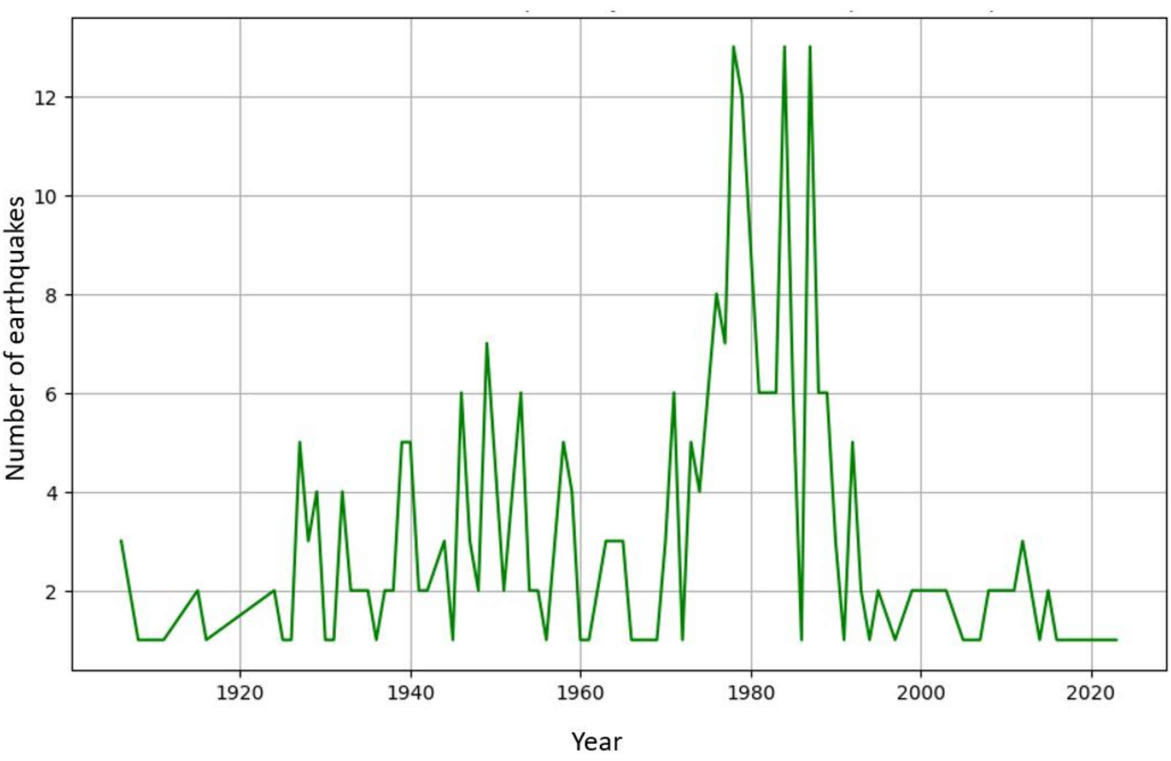

Figure 3. Number of earthquakes over time in Kazakhstan and nearby regions.

This chart clearly shows the number of earthquakes recorded during the time period. The annual dynamics of the number of earthquakes shows changes in seismic activity over time. The number of recorded earthquakes in each year also reflects changes in their frequency and magnitude.

The results of the research allow to understand the distribution of earthquakes, as well as to strengthen seismic safety measures. Studying the dynamics of earthquakes over time is important information in planning preventive measures to ensure seismic safety of the country, as well as to protect life and property of citizens.

Figure 4 shows the relationship between magnitude and depth in Kazakhstan and nearby regions. This diagram allows to analyze the correlation between magnitude and depth of seismic events.

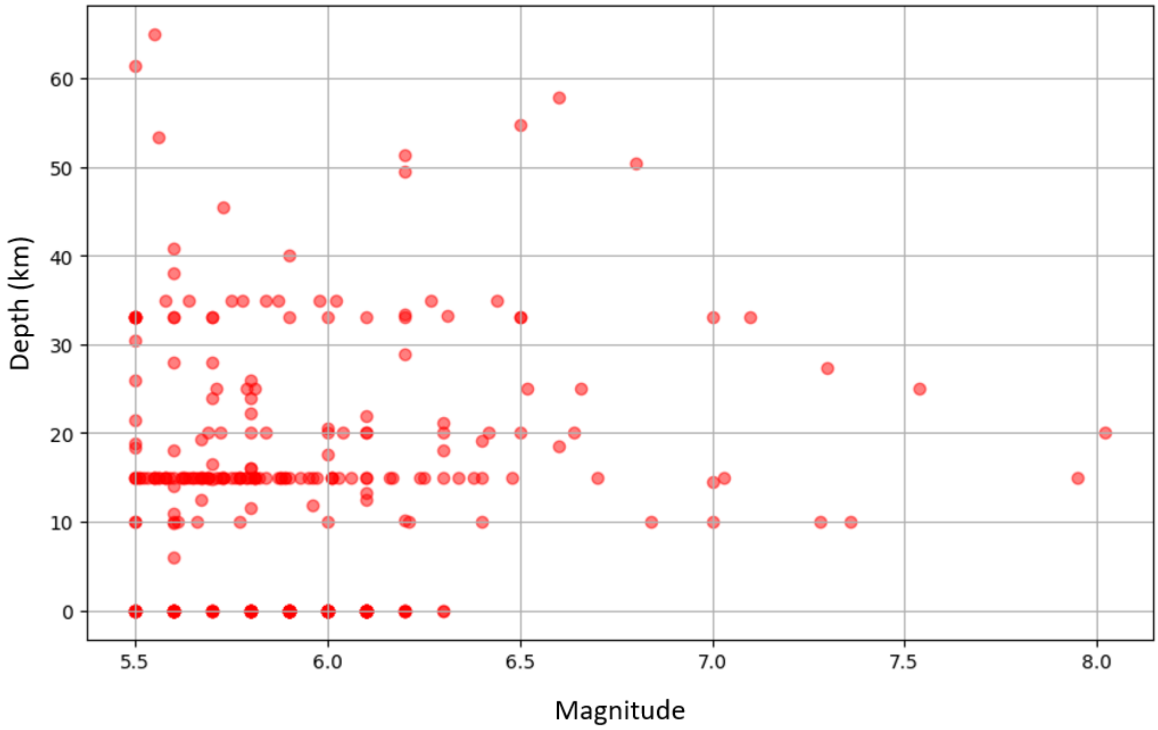

Figure 4. Relationship between magnitude and depth in Kazakhstan and nearby regions.

Magnitude and depth values are graphically represented in the diagram, which helps to better understand the nature of seismic activity. The relationship between magnitude and depth is important for determining the mechanics of earthquakes and their causes.

Correlation between the depth and magnitude of earthquakes plays a crucial role in studying the dynamics of movements and tectonic processes in the earth's crust. This information can also be useful in the development of seismological models and forecasting systems. Seismologists and geophysicists try to predict the nature of earthquakes and their effects through the relationship between magnitude and depth.

Figure 5 shows the foci of earthquakes in Kazakhstan and nearby regions. This map shows areas marked in light blue, yellow and red according to the level of earthquake risk. Areas marked in red are high risk, while light blue and yellow are low risk.

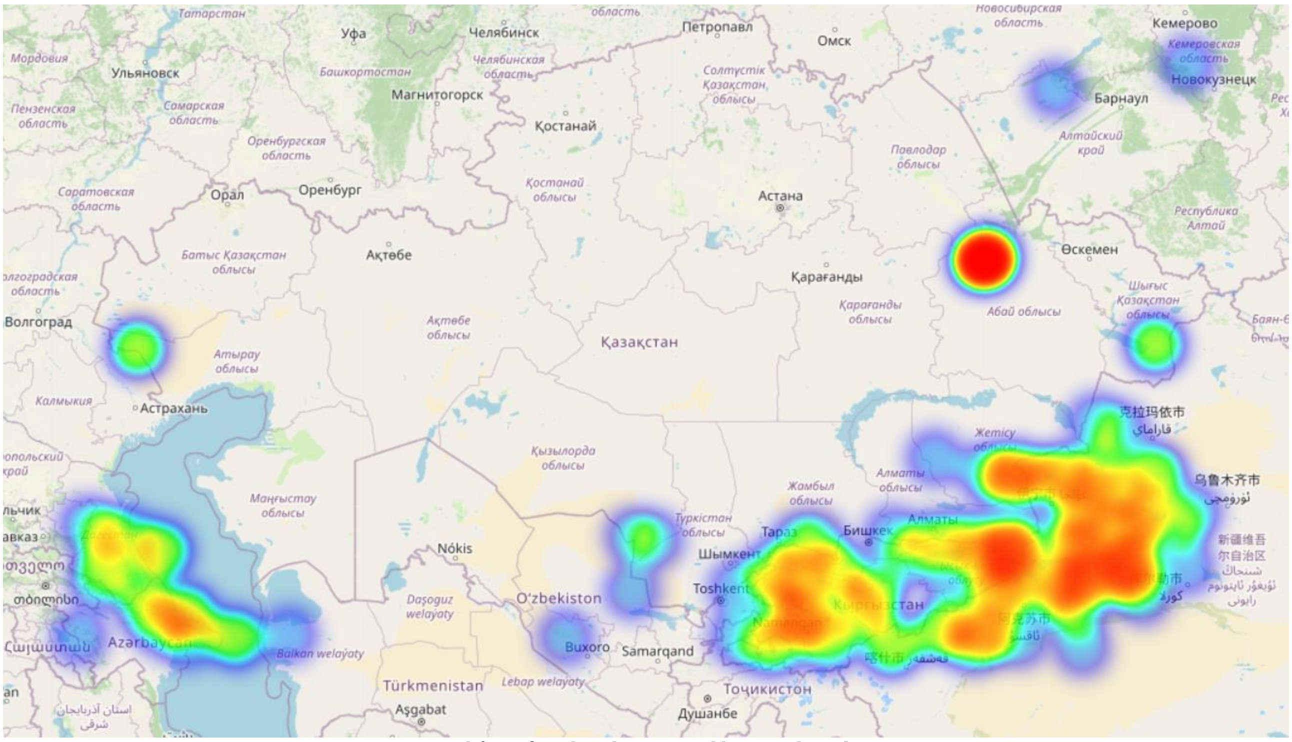

Figure 5. Potential foci of earthquakes in Kazakhstan and nearby regions.

Visualization of possible foci of earthquakes in Kazakhstan and nearby regions plays an important role in the research. The heat map created on the basis of the coordinates of earthquakes makes it possible to identify dangerous areas. With the help of this map, it is possible to estimate the possible frequency of earthquakes in Shymkent, Kyzylorda, Atyrau, and East Kazakhstan regions.

The indicators of the map provide important information in determining the dangerous areas that may be exposed to the risk of earthquakes and planning safety measures in these areas. For example, in regions such as Abay region, Almaty city, and Zhambyl region, the risk may be high, which indicates the need to develop preventive measures to protect residents and infrastructure. Research results can be considered as useful information in the field of construction, civil defense and emergency management systems.

The findings of this study present significant contributions to seismic risk management, urban planning, and emergency preparedness through the application of advanced machine learning techniques. By identifying XGBoost and LightGBM as the most effective models for earthquake-prone area prediction, this research demonstrates an enhancement in predictive accuracy for seismic risk assessment. The improved model performance supports earlier and more reliable detection of high-risk zones, providing critical data for stakeholders.

In the context of urban and infrastructure planning, the precise delineation of high-risk seismic zones offers valuable insights for policymakers, engineers, and urban developers. Data-driven assessments derived from these models enable authorities to prioritize building codes, plan infrastructure resilience, and enhance construction standards within vulnerable areas. This is particularly relevant for resource allocation in seismically active regions like Kazakhstan, where optimized planning can significantly mitigate the potential impacts of earthquakes.

This study underpins improved resource allocation for emergency response and preparedness initiatives. By highlighting high-risk areas, the results empower emergency management teams to concentrate resources effectively, facilitating targeted risk management and response strategies. Such optimized preparedness is crucial for minimizing both immediate and long-term impacts of seismic events on affected communities.

The methodological advancements presented in this research contribute to the broader field of earthquake prediction and risk assessment, providing a framework that can be adapted and applied in other seismically active regions globally. The successful integration of XGBoost and LightGBM in this context underscores the potential of machine learning algorithms to address complex geospatial prediction challenges, marking a promising direction for future studies.

The study enhances public awareness of earthquake risk, promoting a culture of safety and proactive preparedness. Through precise identification of high-risk zones, communities are better informed and motivated to engage in preventative measures, ultimately supporting a resilient and informed society. Collectively, these contributions underscore the potential of machine learning in enhancing the safety, preparedness, and resilience of earthquake-prone regions.

Discussion of results

The machine learning models used in the study, namely XGBoost and LightGBM, showed the highest performance indicators, the accuracy was 0.745737 and 0.739783. These models have proven to provide effective solutions in earthquake prediction.

In addition, an interactive heat map was created to visualize the risk of earthquakes in Kazakhstan and nearby regions. On the map, foci of earthquakes, dangerous areas are marked in bright blue, yellow, and red colors. This visualization makes it possible to identify high-risk regions and plan preventive measures during earthquakes.

Thus, the results of the research show the practical application of machine learning methods and the usefulness of cartographic visualization, which is the basis for the development of earthquake research from a scientific and practical point of view.

Conclusion

During the research, various machine learning methods were used to detect earthquakes in Kazakhstan and nearby regions. In order to predict the magnitude and frequency of earthquakes, the research included RandomForest, GradientBoosting, LogisticRegression, SVC, KNeighbors, DecisionTree, XGBoost, LightGBM, AdaBoost, and MLPClassifier models.

As the results show in the table, the XGBoost model achieved the highest result with 74.57% accuracy. LightGBM and GradientBoosting models showed 73.98% and 73.21% accuracy, respectively. These results are important data for effective analysis and prediction of earthquake dynamics. The results serve as an effective source of information for the state bodies of the Republic of Kazakhstan in planning earthquake prevention measures and for scientific research institutes.

During the research, a visual map was also created, which was used to show the possible foci of earthquakes. This map serves as a useful tool in identifying dangerous areas and planning preventive measures to ensure the safety of local residents. In particular, Shymkent, Kyzylorda, Atyrau, and East Kazakhstan regions, as well as Abay, Almaty, and Zhambyl regions, were identified as possible centers of earthquakes.

The conducted research confirmed the effectiveness of machine learning methods in predicting earthquakes. In the future, it is important to increase the accuracy of forecasting systems by expanding the scope of research and introducing additional environmental and geological factors. The results of the research are relevant not only from a scientific point of view, but also for practical use.

Future research in earthquake risk assessment should focus on several key areas to enhance the impact of machine learning in seismology. Expanding datasets to include more geospatial and temporal data sources will improve model robustness and generalization. Hybrid models combining traditional machine learning with deep learning techniques could capture nonlinear relationships and enhance predictive accuracy.

Additionally, developing real-time data processing pipelines will enable faster response times and improve early-warning systems. Incorporating interpretable machine learning methods, such as SHAP or LIME, would help translate complex predictions into actionable insights for policymakers.

Testing these models across different earthquake-prone regions will help assess their generalizability, while integrating socioeconomic data could provide a more comprehensive risk assessment. These advancements will refine predictive capabilities and support disaster resilience efforts, ensuring that machine learning models are not only accurate but also practical for real-world applications.

References

[1] National Earthquake Information Center. (2023, February 6). *M 7.8 – Pazarcik earthquake, Kahramanmaras earthquake sequence*. United States Geological Survey. https://www.usgs.gov/earthquakes/eventpage/us7000kz0d/region-info

[2] KGU «Upravlenie obshhestvennogo razvitija g. Almaty». (2024, October 2). Almatyda tagy da zher silkindi. TOO «Alataý Aqparat». https://aqshamnews.kz/kz/article/almatyda-tagy-da-jer-silkindi.html

[3] Ridzwan, N. S. M., & Yusoff, S. H. M. (2023). Machine learning for earthquake prediction: a review (2017–2021). Earth Science Informatics, 16(2), 1133-1149.

[4] Abdalzaher, M. S., Elsayed, H. A., Fouda, M. M., & Salim, M. M. (2023). Employing machine learning and iot for earthquake early warning system in smart cities. Energies, 16(1), 495.

[5] Joshi, A., Raman, B., Mohan, C. K., & Cenkeramaddi, L. R. (2024). Application of a new machine learning model to improve earthquake ground motion predictions. Natural Hazards, 120(1), 729-753.

[6] Pourghasemi, H. R., Pouyan, S., Bordbar, M., Golkar, F., & Clague, J. J. (2023). Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination. Natural Hazards, 116(3), 3797-3816.

[7] Pyakurel, A., Dahal, B. K., & Gautam, D. (2023). Does machine learning adequately predict earthquake induced landslides?. Soil Dynamics and Earthquake Engineering, 171, 107994.

[8] Karmenova, M., Tlebaldinova, A., Krak, I., Denissova, N., Popova, G., Zhantassova, Z., & Györök, G. (2022). An approach for clustering of seismic events using unsupervised machine learning. *Acta Polytechnica Hungarica*, *19*(5), 7-22.

[9] Nurtas, M., Zhantaev, Z., Altaibek, A., Nurakynov, S., Mekebayev, N., Shiyapov, K., ... & Ydyrys, A. (2023). Predicting the Likelihood of an Earthquake by Leveraging Volumetric Statistical Data through Machine Learning Techniques. *Engineered Science*, *26*, 1031.

[10] Turarbek, A., Adetbekov, Y., & Bektemesov, M. (2023). 2-d deep convolutional neural network for predicting the intensity of seismic events. *International Journal of Advanced Computer Science and Applications*, *14*(1).

[11] Baktibayev, D., Baigozha, B., Akhmetov, I., Mussabayev, R., Krassovitskiy, A., & Toleu, A. (2024). Literature review on aftershock and earthquake prediction models aided by NLP summarization and ontology extraction techniques. *Procedia Computer Science*, *238*, 579-586.

[12] Karmenova, M., Nugumanova, A., Tlebaldinova, A., Beldeubaev, A., Popova, G., & Sedchenko, A. (2020, April). Seismic assessment of urban buildings using data mining methods. In *Proceedings of the 2020 6th International Conference on Computer and Technology Applications* (pp. 154-159).

[13] Amey, R. M., Elliott, J. R., Hussain, E., Walker, R., Pagani, M., Silva, V., ... & Watson, C. S. (2021). Significant seismic risk potential from buried faults beneath Almaty City, Kazakhstan, revealed from high-resolution satellite DEMs. *Earth and Space Science*, *8*(9), e2021EA001664.

[14] Turarbek, A., Bektemesov, M., Ongarbayeva, A., Orazbayeva, A., Koishybekova, A., & Adetbekov, Y. (2023). Deep Convolutional Neural Network for Accurate Prediction of Seismic Events. *International Journal of Advanced Computer Science and Applications*, *14*(10).

[15] Yavuz, E., Iban, M. C., & Arpaz, E. (2023). Identifying the source types of the seismic events using discriminant functions and tree-based machine learning algorithms at Soma Region, Turkey. *Environmental Earth Sciences*, *82*(11), 265.

[16] Li, F., Torgoev, I., Zaredinov, D., Li, M., Talipov, B., Belousova, A., ... & Schneider, P. (2021). Influence of earthquakes on landslide susceptibility in a seismic prone catchment in central Asia. *Applied Sciences*, *11*(9), 3768.

[17] Akhmed-Zaki, D., Mansurova, M., Yertuyak, A., & Chikibayeva, D. (2021, April). Development of Web Application for Visualizing City Emergencies. In *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)* (pp. 1-5). IEEE.

[18] Kim, S., Yoon, B., Lim, J. T., & Kim, M. (2021). Data-driven signal–noise classification for microseismic data using machine learning. *Energies*, *14*(5), 1499.
[19] *The ultimate earthquake dataset from 1990-2023* [Data set] . (2023, February 6). Kaggle. https://www.kaggle.com/datasets/alessandrolobello/the-ultimate-earthquake-dataset-from-1990-2023
[20] Zhao, B., Su, L., Xu, Q., Li, W., Xu, C., & Wang, Y. (2023). A review of recent earthquake-induced landslides on the Tibetan Plateau. Earth-Science Reviews, 104534.
[21] Sunte, J. (2023). The Controlling Measures and Solution to Problems of Earthquake. IJSRMME, 7 (1).
[22] Martínez-Garzón, P., & Poli, P. (2024). Cascade and pre-slip models oversimplify the complexity of earthquake preparation in nature. Communications Earth & Environment, 5(1), 120.
[23] Bısarınova A. (2021). Megapolistin jekologijalyk zhagdajyn baкylauga arnalgan geoakparattyk zhujenin kurylymy (Doktorlyk dissertacija). K.I.Satbaev atyndagy Kazak Ulttyk tehnikalyk zertteu universiteti.
[24] OpenAI. (2024, December 10). Precision, Recall, and F1-Score: Evaluating Classification Performance. OpenAI. Retrieved from https://openai.com
[25] Wikipedia contributors. (n.d.). Gaussian function. Wikipedia, The Free Encyclopedia. Retrieved February 27, 2025, from https://en.wikipedia.org/wiki/Gaussian_function