IRSTI 28.23.25 https://doi.org/10.26577/jpcsit20253104

Symbat Kabdrakhova[1,2,*], Zhanelya Assan[1], Jelena Caiko[3], Abilmansur Seilkhan[1]

[1] Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2] Institute of mathematics and mathematical modeling, Almaty, Kazakhstan
[3] Riga Technical University, Riga, Latvia
*e-mail: symbat2909.sks@gmail.com

INVESTIGATION OF EMERGENCY SITUATIONS IN ALMATY USING MACHINE LEARNING METHODS

Abstract. At present, the protection of the population from emergencies that occur daily and cause harm to people and the country's territory necessitates organizational measures for monitoring, research, forecasting, and prevention. This study focuses on different types of emergencies, including natural, social, and man-made disasters. With the increasing volume of information on the Internet, there is a growing need to analyze the continuous flow of data published on news websites. In this study, machine learning-based methods and approaches were utilized. A research analysis of emergency-related data was conducted, identifying the key factors influencing the frequency of incidents. Additionally, emergencies were classified and assessed based on their types. During the evaluation of various algorithms, the most effective machine learning methods were determined. Data was collected from open sources in text format and subsequently processed using natural language preprocessing techniques. By leveraging historical weather data for the city of Almaty, a correlation between emergencies and weather conditions was identified.

Key words: emergencies, machine learning, classification, NLP, preprocessing, KNN, LR, RF, text classification.

1. Introduction

Emergencies are one of the key causes of economic and human losses. These incidents vary in scale and the damage inflicted on society, ranging from small road accidents involving a few citizens to large-scale natural and man-made disasters affecting hundreds of people. Effective prediction and analysis of emergencies play a crucial role in mitigating their negative consequences. The study and management of emergency situations are no less important than other global challenges. In this regard, the study, prevention, and mitigation of the consequences of emergencies through innovative solutions represent a rapidly developing research field [1].

Recently, with the expansion of the Internet, online platforms have become a crucial means of information dissemination. Every day, vast amounts of data in the form of text, audio, and images are generated and published online. This information often includes news about emergencies such as natural disasters, traffic accidents, and fires [2]. However, traditional news analysis methods require significant time and computational resources, making the implementation of automated systems based on machine learning methods necessary. The use of machine learning significantly accelerates data processing and improves analysis accuracy. The relevance of this study is determined not only by the growing number of emergencies but also by the need for rapid data processing for timely response. In Almaty, one of the largest metropolises in Kazakhstan, thousands of emergencies of various types are recorded annually, including earthquakes, floods, fires, and traffic accidents. Thus, the development and implementation of automated emergency analysis systems is a pressing scientific challenge.

Many studies focus on the collection and processing of emergency-related data from social networks. Sara Piscitelli and Edoardo Arnaudo have developed a multilingual tool that automatically classifies tweets based on their informational content. To achieve real-time classification with multilingual support, a deep learning-based classifier is employed. This approach enables the system to work with any language using semantic analysis [3].

The study presented in [4] examines not only various classifiers but also the application of n-grams and vectorization techniques. The findings indicate that machine learning-based classifiers perform well in determining the relevance of news reports. Moreover, these approaches provide real-time assessments of data quality for emergency services.

Several scientific studies focus on specific classes of emergencies. One such study proposes an earthquake management support system based on machine learning and natural language processing. The system demonstrated high performance in identifying citizen-generated messages that contain useful information about the magnitude, scale, and location of an earthquake [5].

During data collection via keyword-based search, a vast amount of information is retrieved. However, this dataset often contains a mix of relevant and irrelevant data, making it difficult or even impossible to compile a balanced training set [6]. In addition to classifying and analyzing emergency situations, the classification of messages requiring transmission to relevant authorities is also considered. In [7], researchers collected and categorized 9,246 microblog messages, revealing that 6.95% of messages were directed to the fire department, 52.97% to the government, 10.71% to ambulance services, and 10.21% to the police. This further underscores the importance of research in emergency management.

Machine learning algorithms and neural networks are widely used for text classification. The study in [8] developed a classification system for BBC news articles. In the "Classifier Implementation" section, the authors selected k-nearest neighbors (KNN), logistic regression (LR), and random forest (RF), followed by a comparative analysis of these algorithms. A comparison based on five parameters–accuracy, precision, recall, F1-score, and confusion matrix–enabled the identification of the most effective classifier for the given dataset.

Among machine learning methods, support vector machines (SVM) are a supervised learning model. This classifier operates by finding an optimal hyperplane that separates a dataset into distinct classes [9].

In logistic regression, the target variable is categorical. The study in [10] outlines the theoretical foundation of the logistic regression algorithm and provides the key formulas used for data classification. Another research paper [11] demonstrates a structured approach to text classification in English. The authors proposed a classification model starting with data collection, as they did not use a pre-existing dataset. Using Python libraries, they performed web scraping to gather data. As in many models, the data collection phase was followed by a preprocessing stage, which included tokenization, stop-word removal, and other steps. After preprocessing, a vector representation of the document was created, followed by machine learning-based classification. Finally, a comparative analysis of different models was conducted using performance metrics to determine the most effective algorithm.

The study in [12] introduced a machine learning-based methodology for classifying texts containing positive classes alongside unlabeled data. The primary goal of this research was to identify positive classes and distinguish unlabeled data within the dataset.

This study not only contributes to the development of intelligent data analysis methods but also offers practical solutions for emergency monitoring and management. It opens opportunities for further research in predictive analytics and the automatic detection of patterns in emergency data.

2. Materials and Methods

2.1 Research area and data

Almaty is the largest city of Kazakhstan of republican significance, located in the center of Eurasia. The megacity of Kazakhstan includes 8 districts: Auezovsky, Bostandyk, Medeu, Turksib, Almaly, Zhetysu, Alatau and Nauryzbay. The territory is more than 170 square kilometers. The population of the city is about 2 million people. Every day there are outbreaks of emergencies on the territory of the megalopolis, as well as the inherent dangers for the population of the city. A huge number of articles on the topic of emergencies are published in information portals, but the selection and sorting of these data is not given due attention. This research paper examines emergency situations in the city of Almaty of a social, man- made, as well as natural nature. The dataset consists of news articles collected from the site https://tengrinews.kz and historical data on the weather conditions of the city of Almaty. The assembled dataset consists of 26 columns and 3290 rows. This dataset is used for research analysis of emergency situations in the city of Almaty. In the future, many predictors will be excluded for text classification.

Table 1 – Database.

№	Column name	Meaning
1	date	date of the emergency incident
2	content	data in the form of text
3	category	emergency category
4	maxtempC	maximum temperature
5	mintempC	minimum temperature
6	totalSnow_cm	snow thickness
7	sunHour	sunny time
8	uvIndex	index of ultraviolet light
9	moon_illumination	the value of the illumination of the moon
10	moonrise	moonrise time
11	moonset	moonset time
12	sunrise	sunrise time
13	sunset	sunset time
14	DewPointC	dew point temperature
15	FeelsLikeC	the value of the temperature feels like
16	HeatIndexC	Thermal index in celsius
17	WindChillC	the value of cold wind in celsius
18	WindGustKmph	wind gust speed
19	cloudcover	cloud cover value
20	humidity	air humidity
21	precipMM	the value of precipitation per day
22	pressure	pressure of the atmosphere
23	tempC	temperature per day
24	visibility	environmental visibility
25	winddirDegree	severity of wind
26	windspeedKmph	wind speed per day

2.1.1 Research data analysis

Research analysis of emergency data gives us an understanding of the causes of these cases. The correlation of natural conditions and emergency situations has been indisputably proved by other research papers.

Figure 1 shows the relationship between the number of breaking news and the days of the month. The diagram shows that in this data set, the second, eighth and fourteenth days of the month are characterized by a large number of emergency reports. These indicators may vary and depend on the amount of news collected.

On a par with the data that include several types of emergencies, a certain type of emergency situations can be distinguished. An example of this is shown in Figure 2. In this case, a certain type of emergency is considered – traffic accidents. Looking at this figure, we can say that the location, namely the initial coordinate and the final coordinate, air humidity and ambient pressure, temperature and wind speed directly or indirectly affect the cases of road accidents. These observations make it clear that you need to pay attention to weather conditions in everyday life.

The amount of news about emergencies is also affected by the seasons, such as autumn, winter, spring, summer. This dataset contains all seasons except spring. The indicators in Figure 3 make it clear that in the autumn the number of news about emergency situations increases.

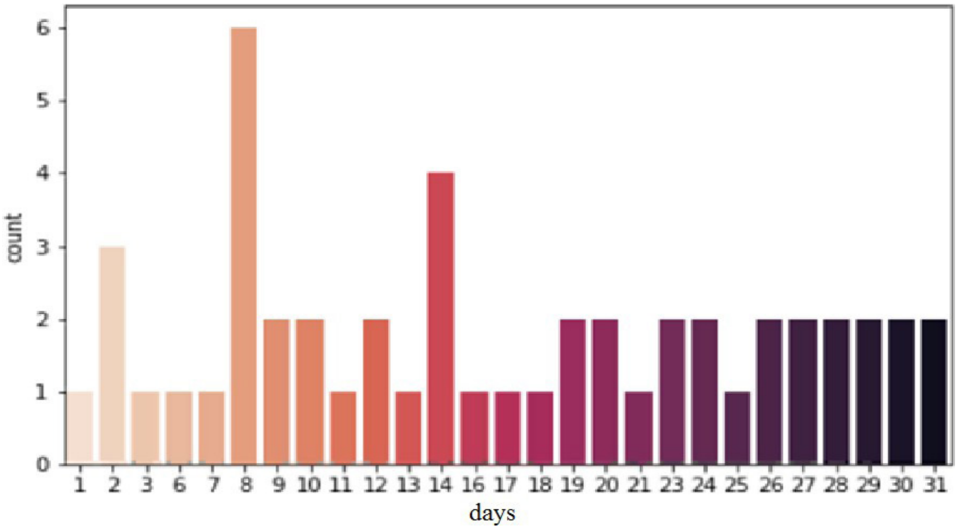

Figure 1 – The relationship between the number of breaking news and the days of the month.

In addition to the influence of the seasons, it is possible to study the influence of the days of the week on the amount of news about emergency situations, the indicators of which can be seen in Figure 4.

For text classification, it is advisable to reduce the data set and remove some predictors that are not involved in text classification. After preprocessing steps, a set consisting of 1712 rows and 3 columns was obtained. Now the data consists of four classes that need to be processed and classified in the future.

The category column is fairly balanced, but if necessary, you need to consider the possibility of using algorithms designed to increase samples.

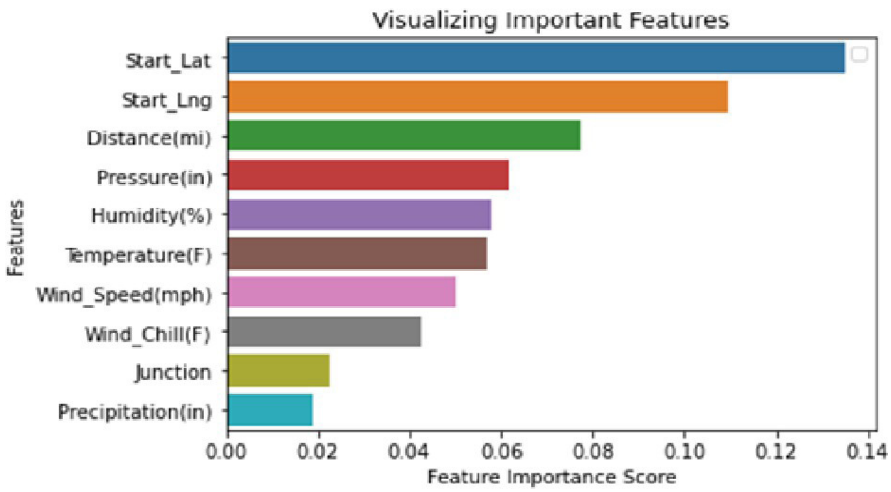

Figure 2 – Factors affecting road traffic accidents.

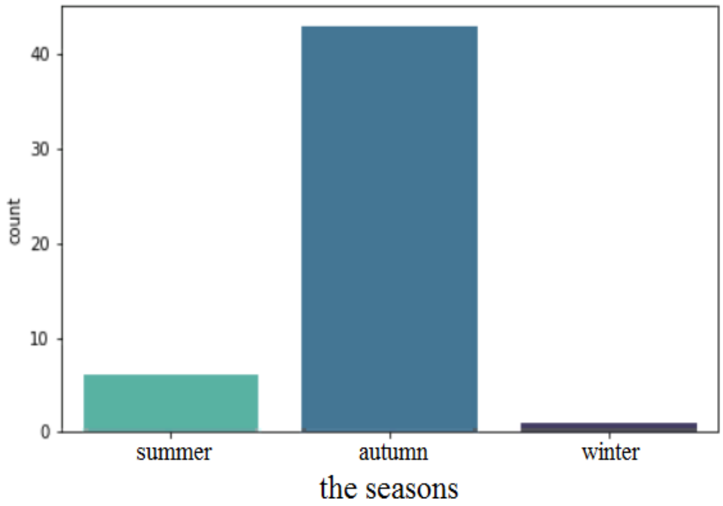

Figure 3 – Dependence of emergency news on the seasons**.**

Table 2 – Percentage indicator of the number of news by category.

Category	**percentage indicator**
fire	27,9 %
road accident	28,9 %
earthquake	27,3 %
flood	15,9 %

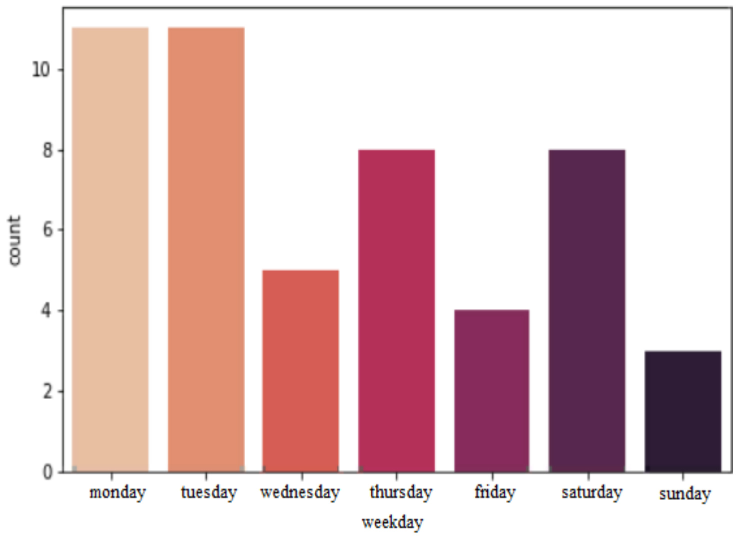

Figure 4 – The influence of the days of the week on the number of emergency news.

2.2. Classification of the text

Text classification is an algorithm for constructing models that classify texts into pre-trained classes. At the moment, this is a well-coordinated process that includes such steps as data preprocessing, increasing or decreasing the dimension, evaluating the quality of data, etc.

Researchers in this field modify existing classification systems or develop new algorithms to achieve higher accuracy rates [13]. Figure 5 shows the main stages that occur before the classification of the text. Special attention is paid to the preprocessing of the text, functions have been created that perform certain actions necessary for further classification. Different approaches and the replacement of some components give different results when evaluating each machine learning algorithm.

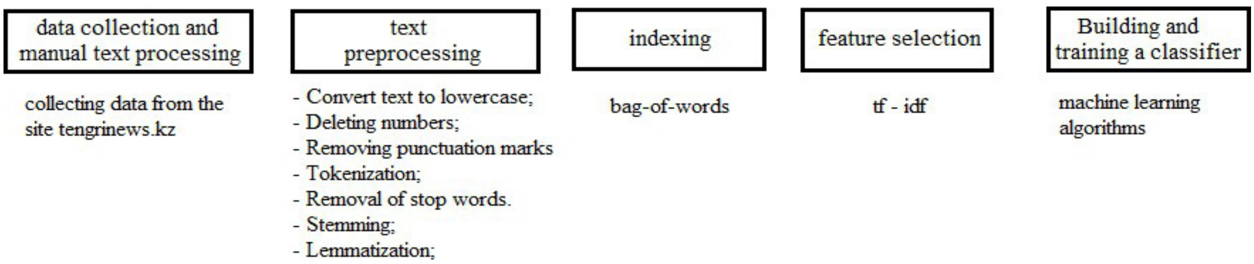

Figure 5 – Stages of text classification.

2.2.1 Data collection and characteristics

The data for this study was collected from the Tengrinews.kz news portal using web scraping techniques with the requests and BeautifulSoup libraries in Python. Tengrinews.kz is one of Kazakhstan's leading news websites, publishing information on various topics, including emergency situations. The data collection process consisted of several steps:

- Identifying relevant information – news articles related to emergency situations were selected.

- Filtering data by categories – only news articles mentioning fires, floods, earthquakes, and traffic accidents were included.

- Automated scraping – GET requests were sent to the website, and the BeautifulSoup library was used to extract headlines, dates, and article content.

- Data storage – the extracted news texts were saved in .xlsx format for further processing.

- Development of a web-based data collection tool – a simplified web interface was created using Flask, allowing users to launch the news collection process and download the final dataset.

Figure 6 shows a site that collects data. As it is written on the website, the first stage is to determine the necessary information on emergency situations. After that, we select the type of news required for collection, this is done by pressing the appropriate button. Then the site displays the process sign. At the final stage, at the output in the download section, we get an excel file with the necessary news.

Figure 6 – Web scraping site.

Many ready-made datasets do not require manual processing and sorting of texts, but since this collection of emergency news was collected for the first time, it was advisable to perform this stage. Manual processing took place in an environment for working with excel spreadsheets. During the analysis and sorting by category, omissions and irrelevant samples were removed, which gave a significant jump in the accuracy of classification.

The collected dataset contains 3,290 records and 26 attributes, which can be grouped into several categories:

Main attributes of news articles:

- date – the date of the incident

- content – the text of the news article
- category – the category of the emergency situation

Meteorological parameters:

- maxtempC, mintempC – maximum and minimum temperature
- humidity – air humidity
- precipMM – daily precipitation amount
- winddirDegree, windspeedKmph – wind direction and speed
- pressure – atmospheric pressure
- cloudcover – cloud coverage
- Other parameters such as UV index, visibility, dew point temperature, etc.

Specifically, irrelevant samples included news articles that mentioned emergencies in a general sense but did not describe specific incidents. Duplicate news items and articles with incomplete or ambiguous information were also removed. After preprocessing (removing duplicates, filtering out irrelevant records, and selecting key data), the final dataset was reduced to 1,712 records, containing three key attributes:

- content (news text)
- category (target variable – emergency category)
- date (auxiliary feature)
- Class Distribution

The emergency categories are distributed as follows:

- Traffic accidents – 28.9%
- Fires – 27.9%
- Earthquakes – 27.3%
- Floods – 15.9%

Since the "Floods" category is underrepresented, techniques such as oversampling (SMOTE) or data augmentation may be considered to balance the dataset.

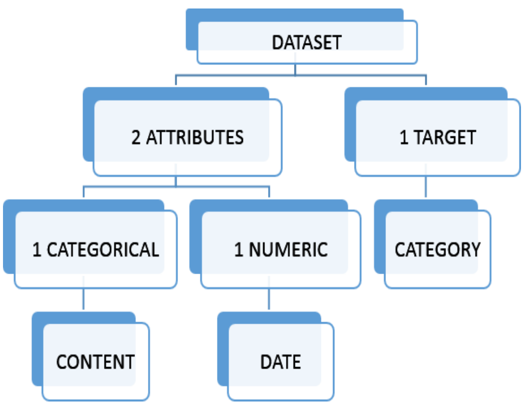

Figure 7 – Data for text classification.

2.2.2 Text preprocessing, indexing, feature selection

Manipulations during text preprocessing in many cases were performed using regular expressions and ready-made functions in Python. Regular expression-based functions were used to remove punctuation and various punctuation marks, as well as to remove numeric characters. Additionally, text normalization included converting all characters to lowercase to ensure uniformity, removing extra spaces, and standardizing date and time formats where applicable. Non-Kazakh and non-Russian language texts were also excluded to maintain dataset consistency. Tokenization and lemmatization were carried out at the expense of ready-made libraries. To remove stop words, a dictionary from the nltk library was used, designed to perform all natural language processing processes Figure 8 shows one of these functions that was used during the text preprocessing process.

```
import re
def prepro(data):
    data = re.sub(r'\d+|[^\w\s]', '', str(data))
    data=data.lower()
    data= data.replace("\n", " ")
    return data
df["Content_new"] = df["Content"].apply(lambda x: peтTey(x))
df.head()
```

Figure 8 – A function designed for data preprocessing.

Indexing is the representation of text in a numeric format, for convenient further processing and working with it. In our research work, the "bag of words"model was used. This model represents a data set in the form of a multidimensional vector of words and, accordingly, their weights. The selection of features is a very important step related to the categorization process. During this process, we make a selection of words that best describe these categories. The accuracy of the classification depends on the selected keywords, as well as the complexity of the computing system depends on the number of these keywords. If the number of signs is large, then the system is inefficient in terms of time. In this research paper, there is a good ratio of categories and words related to them. Table 3 shows the results of feature selection.

Table 3 – Unigrams and bigrams.

<table>
<tr><td>fire</td><td>гореть, возгорание, огонь, пожарный, пожар</td><td rowspan="4">подземный толчок, землетрясение, магнитуда</td></tr>
<tr><td>flood</td><td>тысяча, уровень, река, вода, наводнение</td></tr>
<tr><td>earthquake</td><td>подземный, эпицентр, толчок, магнитуда, землетрясение</td></tr>
<tr><td>road accident</td><td>авария, полиция, автомашина, водитель, дтп</td></tr>
</table>

The choice of preprocessing methods is based on best practices in natural language processing to improve text classification accuracy. Regular expressions and NLP libraries ensure effective text cleaning, reducing noise and improving feature extraction. The Bag of Words model was selected for its simplicity and efficiency in representing text as numerical data, making it suitable for traditional machine learning classifiers. Additionally, stop-word removal, tokenization, and lemmatization help enhance the model's performance by reducing dimensionality and improving the relevance of extracted features.

2.2.3 Evaluation

To assess the performance of the classification algorithms, we use four key metrics: accuracy, precision, recall, and F1-score.

Precision (P) measures the proportion of correctly classified positive instances among all instances predicted as positive.

Recall (R) shows how many actual positive instances were correctly identified by the classifier.

F1-score is the harmonic mean of precision and recall, providing a balanced evaluation of the model's performance.

Accuracy represents the overall percentage of correctly classified instances across all classes.

These metrics help to objectively compare the effectiveness of different machine learning algorithms.

2.2.4 Classification Algorithms

To classify the text data, we applied five machine learning algorithms:

K-Nearest Neighbors (KNN): Classifies text by identifying the closest matches in the training dataset using Euclidean distance.

Support Vector Machine (SVM): Uses a linear kernel and a regularization parameter $C = 0.1$ to separate classes effectively.

Logistic Regression: A statistical model commonly used for binary and multiclass classification.

Gradient Boosting: An ensemble learning method that builds models sequentially to improve classification accuracy.

Multinomial Naïve Bayes: A probabilistic model that estimates class probabilities based on word frequencies.

The results of these models, evaluated using the described metrics, are presented in Table 4 and visualized in the confusion matrices in Figure 9.

3. Results

The data consists of 1712 news articles collected by us from news sites has been pre-processed, consisting of numerous stages. As a result of processing, the data was converted to a numeric format for further classification. During the experiment, for training and testing, the dataset was split into 80% training data (1369 samples) and 20% testing data (343 samples) and apply machine learning algorithms to classify the data. Machine learning algorithms receive data in numerical format. To achieve the transition from a text format to a numeric one, a frequency analysis of words was carried out using the tf-idf method. This method evaluates the importance of each word in relation to categories and the entire dataset. Words that are more often used in the context of one category give a greater chance that an emergency text message refers to it. At the same time, the entire text is used to assess compliance, not just individual words. Thus, the input data of the classifier is a vector of numbers obtained by frequency analysis of words using the tf-idf method, and categorical data converted into numeric labels as target data.

The first machine learning algorithm for text classification is the k nearest neighbors method. In this experiment, the KNN algorithm performs classification by finding close matches in train data. In our case, the Euclidean distance is used to find the distance between the data labels. Each instance of the class has its own location on the Cartesian plane. A data instance that does not have a label is determined by the difference in distances from each class.

The next method is the support vector machine. A linear parameter was selected as the "core" parameter of the classifier. The "C"parameter helps to adjust the classification accuracy. The larger this parameter, the more accurate this algorithm is. In our experiment, the value of "C"is 0.1.

In addition to these two algorithms, the methods of logistic regression, gradient boosting, and multinomial naive bayes were used.

As a result, 5 machine learning algorithms were evaluated by 4 metrics, such as accuracy, recall, precision, f1-score. The classifier based on gradient boosting became the best in terms of accuracy. The classifier based on the naive Bayes method showed the worst result.

Selecting the parameters, we achieved excellent results, which are presented in Table 4, as well as the confusion matrix of algorithms is demonstrated in Figure 9.

Table 4 – Results of classification algorithms.

	Accuracy	Recall	Precision	F1-score
KNN-6	0.92128	0.92062	0.91270	0.91517
KNN-3	0.91253	0.91021	0.90863	0.90879
KNN-8	0.93294	0.92842	0.92488	0.92597
SVM	0.96209	0.96179	0.95253	0.95658
LogisticRegression	0.97667	0.97452	0.97452	0.97343
Gradient Boosting Machine	0.97959	0.97697	0.97511	0.97593
MultinomialNB	0.89795	0.90013	0.88591	0.88475

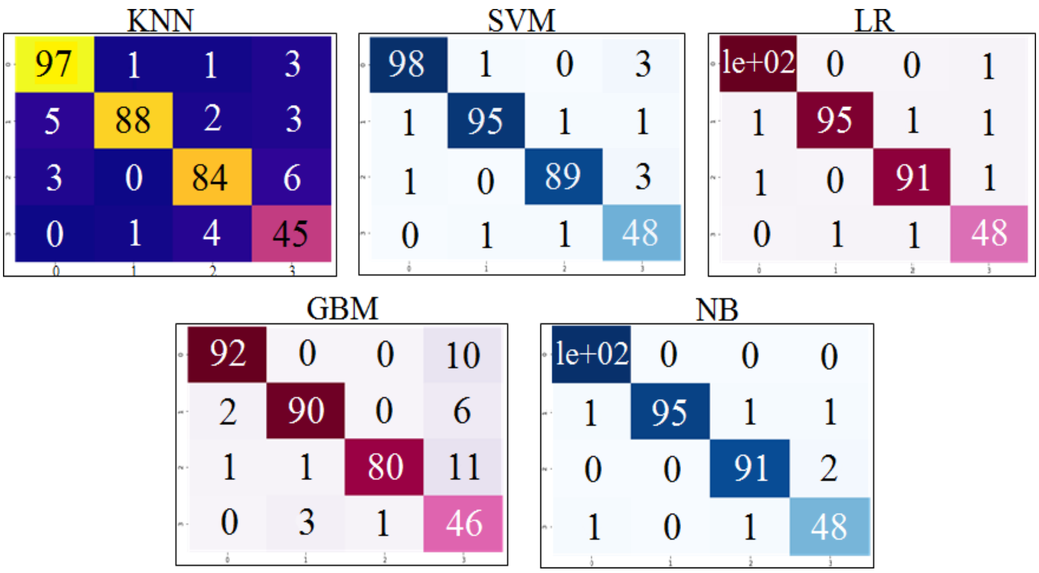

Figure 9 – Confusion matrix of algorithms.

Based on the evaluation of five machine learning algorithms, several key conclusions can be drawn. Gradient boosting demonstrated the highest accuracy among all methods. This can be attributed to its ability to capture complex dependencies in data through iterative model improvement. In contrast, the Naïve Bayes method showed the lowest performance. This is due to its assumption of feature independence, which does not always hold in text classification tasks. In news articles, words are often contextually dependent, which affects the model's effectiveness. The support vector machine (SVM) with a linear kernel produced strong results. The choice of a linear kernel is justified, as text data transformed using the tf-idf method is often well-separated in a linear space. The parameter C = 0.1 helped prevent overfitting. The k-nearest neighbors (KNN) method demonstrated moderate performance. Its effectiveness is highly dependent on the choice of k and the distance metric. The use of Euclidean distance may not fully capture the characteristics of textual data, as it is generally more suited for continuous numerical features.

4. Conclusions

The main objective of this article was to find the most effective classification method. All methods are based on machine learning. The data is collected from the website tengrinews.kz, which in turn is one of the main news sites of Kazakhstan.

- At the initial stage, we performed a text preprocessing process, which includes methods aimed at presenting the text in a form suitable for effective classification.
- Several steps were performed, such as manual and software cleaning of the dataset.
- Manual cleaning and filtering of emergency news texts was performed.
- Methods such as logistic regression, k nearest neighbors, svm, gradient boosting, and multinomial naive bayes were compared. According to the result of the study, all five machine learning methods gave good results, but the gradient boosting method is superior to other machine learning classification algorithms for the collected data set. Comparison of methods was achieved using such indicators as accuracy, precision, recall, f1-score.
- The results of the article can be used to analyze and make decisions on emergency situations in Almaty.

Thus, the choice of algorithm plays a crucial role in solving text classification tasks. Gradient boosting achieved the best results, confirming its ability to identify complex relationships. At the same time, the Naive Bayes method proved less effective due to its strong assumptions about feature independence. A deeper analysis of errors and potential improvements could enhance classification accuracy in future research.

The findings of this study can be useful for analyzing and making decisions related to emergency management in Almaty. Future research could focus on expanding the dataset, applying more complex models, or incorporating additional features such as semantic text analysis. Another promising direction is automatically updating the model with new data, which could enhance its adaptability and accuracy over time.

Author Contributions

Conceptualization, S.K. and Zh.A.; Methodology, S.K. and A.S.; Software, Zh.A. and A.S.; Validation, S.K., Zh.A., and J.C.; Formal Analysis, S.K. and A.S.; Investigation, S.K. and Zh.A.; Resources, S.K.; Data Curation, Zh.A. and A.S.; Writing – Original Draft Preparation, S.K. and Zh.A.; Writing – Review & Editing, J.C. and A.S.; Visualization, Zh.A. and A.S.; Supervision, S.K.; Project Administration, S.K. and J.C.; Funding Acquisition, J.C..

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lopez-Fuentes, L., van de Weijer, J., Gonzalez-Hidalgo, M. et al. "Review on computer vision techniques in emergency situations" , Multimed Tools Appl, -2018, -Vol.77, -P.17069-17107.
2. Kuchment P., Zeng H., "Research on Application of Emergency Text Classification Based on Combination Model" , IOP Conference Series: Materials Science and Engineering. -2020, -Vol.719,-P.1-6.
3. S. Piscitelli, E. Arnaudo and C. Rossi, "Multilingual Text Classification from Twitter during Emergencies" , 2021 IEEE International Conference on Consumer Electronics (ICCE). -2021,-P.1-6.
4. M. Habdank, N. Rodehutskors and R. Koch, "Relevancy assessment of tweets using supervised learning techniques: Mining emergency related tweets for automated relevancy classification" , 2017 4th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). -2017,-P.1-8.
5. Fersini, E., Messina, E. and Pozzi, F.A., "Earthquake management: a decision support system based on natural language processing" , J Ambient Intell Human Comput 8. -2017, -P.37-45.
6. Y. Liu, J. Niu, Q. Zhao, J. Lv and S. Ma, "A Novel Text Classification Method for Emergency Event Detection on Social Media" , 2018 IEEE SmartWorld, Ubiquitous Intelligence, Computing, Advanced, Trusted Computing, Scalable Computing, Communications, Cloud, Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI. -2018, -P.1106-1111.
7. Y. Zhou, L. Yang, B. Van de Walle and C. Han, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification" , Augment Hum Res 5. -2020.

8. Shah, K., Patel, H., Sanghvi, D. et al. , "Classification of Microblogs for Support Emergency Responses: Case Study Yushu Earthquake in China" , 2013 46th Hawaii International Conference on System Sciences. -2013, -P.1553-1562.

9. Mohamed Goudjil and Mouloud Koudil and M. Bedda, "A Novel Active Learning Method Using SVM for Text Classification" , International Journal of Automation and Computing. -2018, Vol.15, -P.290-298.

10. Sadia Zaman Mishu and S. M. Rafiuddin, "Performance Analysis of Supervised Machine Learning Algorithms for Text Classification" , International Conference on Computer and Information Technology. -2016, -Vol.19,-P.409-413.

11. Xiaoyu Luo, "Study on SVM Compared with the other Text Classification Methods" , Alexandria Engineering Journal. -2021, -Vol.60,-P.3401-3409.

12. Ikonomakis, Emmanouil and Kotsiantis and Sotiris and Tampakas, V, "Text Classification Using Machine Learning Techniques" , WSEAS transactions on computers. -2005, -Vol.4, P.966-974.

13. Marcin Michal Mironczuk, Jaroslaw Protasiewicz, "A recent overview of the state-of-the-art elements of text classification" , Expert Systems with Applications. -2018, -Vol.106, P.36-54.

Information About Authors:

Kabdrakhova Symbat, PhD in Physical and Mathematical Sciences, Associate Professor at Al-Farabi Kazakh National University (Almaty, Kazakhstan, symbat2909.sks@gmail.com). ORCID: 0000-0003-0247-5985.

Assan Zhanelya is a first-year PhD student in Computer Engineering at Al-Farabi Kazakh National University (Almaty, Kazakhstan, zh.assanova98@gmail.com). In 2023, he earned a master's degree in technical sciences. Her interests are related to electronics, neural networks, and machine learning. ORCID: 0000-0002-3617-2782.

Jelena Caiko is a Doctor of Engineering Sciences and an Associate Professor at Riga Technical University (RTU) (Riga, Latvia, Jelena.Caiko(at)rtu.lv). ORCID: 0000-0002-1207-1418.

Seilkhan Abilmansur is a first-year PhD student in Computer Science at Al-Farabi Kazakh National University (Almaty, Kazakhstan, seilkhan.mansur@gmail.com). In 2023, he earned a master's degree in technical sciences. His professional interests encompass data analysis, machine learning, and deep learning, while his research interests include neural networks, machine learning, and data analysis. ORCID: 0000-0003-3905-4949.