ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ
АКАДЕМИЯСЫ» РҚБ

«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ

ХАБАРЛАРЫ

ИЗВЕСТИЯ

РОО «НАЦИОНАЛЬНОЙ
АКАДЕМИИ НАУК РЕСПУБЛИКИ
КАЗАХСТАН»

N E W S

OF THE NATIONAL ACADEMY
OF SCIENCES OF THE REPUBLIC
OF KAZAKHSTAN

SERIES OF PHYSICS AND MATHEMATICS

1 (353)

JANUARY – MARCH 2025

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

«ҚР ҰҒА Хабарлары. Физика-математика сериясы».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **28.02.2025** ж. берген **№KZ20VPY00113741** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

«Известия НАН РК. Серия физико-математическая».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания **№ KZ20VPY00113741.** Дата выдачи **28.02.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКСНВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

News of the National Academy of Sciences of the Republic of Kazakhstan.
Series of Physics and Mathematics
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty).
Certificate **No. KZ20VPY00113741** on the re-registration of the periodical printed and online publication of the information agency, issued on **28.02.2025** by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan
Subject area*: information and communication technologies.*
Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies».*
Periodicity: *4 times a year.*
Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

NEWS OF THE NATIONAL ACADEMY OF SCIENCESOF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 1. Namber 353 (2025). 270–282

https://doi.org/10.32014/2025.2518-1726.339

УДК: 004.4

K. Rabbany[1], A. Bekarystankyzy[1], A. Shoiynbek[1], D. Kuanyshbay[2], A. Mukhametzhanov[2], 2025.
[1]Narxoz University, Almaty, Kazakhstan;
[2]Suleyman Demirel University, Kaskelen, Kazakhstan.
E-mail: kazi_golam.rabbany@narxoz.kz

DETECTION OF SUICIDAL TENDENCIES IN REDDIT POSTS USING MACHINE LEARNING

Kazi Golam Rabbany – Narxoz University, Almaty, Kazakhstan, E-mail: kazi_golam.rabbany@narxoz.kz, https://orcid.org/0009-0007-4549-0815;
Bekarystankyzy Akbayan – senior lecturer, PhD in Computer Science, Narxoz University, Almaty, Kazakhstan, E-mail: akbayan.b@gmail.com, https://orcid.org/0000-0003-3984-2718;
Shoiynbek Aisultan – professor, PhD in Computer Science, Narxoz University, Almaty, Kazakhstan, E-mail: aisultan.shoiynbek@gmail.com, https://orcid.org/0000-0002-9328-8300;
Kuanyshbay Darkhan – professor, PhD in Computer Science, Suleyman Demirel University, Kaskelen, Kazakhstan, E-mail: darkhan.kuanyshbay@sdu.edu.kz, https://orcid.org/0000-0001-5952-8609;
Mukhametzhanov Assylbek – Suleyman Demirel University, Kaskelen, Kazakhstan, E-mail: 221107046@stu.sdu.edu.kz, https://orcid.org/0009-0009-8528-9985.

Abstract: Suicide remains a critical public health concern worldwide, with many cases preceded by identifiable warning signs. Social media platforms, particularly Reddit, serve as valuable resources for detecting suicidal ideation through user-generated text. This study proposes a machine learning model utilizing Bidirectional Long Short-Term Memory (BiLSTM) networks with an attention mechanism to automate the detection of suicidal tendencies in Reddit posts.

A dataset comprising posts from the "SuicideWatch" and "depression" subreddits was collected via the Pushshift API and labeled as either suicide-related or non-suicidal. The preprocessing pipeline included case folding, HTML tag stripping, tokenization, stopword removal, and sequence padding. The dataset was balanced to ensure an equal distribution of suicide and non-suicide posts and was split into training (60%), validation (20%), and test (20%) sets.

The model architecture integrated an embedding layer, BiLSTM units, an attention mechanism, and fully connected layers. Training was conducted using the Adam optimizer and binary cross-entropy loss, with early stopping applied. Evaluation metrics indicated strong model performance, achieving an F1 score of

0.91 on the validation set and 0.95 on the test set, with an AUC-ROC of 0.98 and 0.989, respectively.

The findings demonstrate the effectiveness of BiLSTM with attention for suicidal ideation detection, outperforming many existing models. Early detection of suicidal tendencies in online communications can significantly improve mental health outcomes and save lives.

Key words: machine learning, natural language processing, bidirectional LSTM, attention mechanism, suicidal ideation detection

Funding. *This research has been/was/is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP22786670).*

К. Раббани[1], А. Бекарыстанқызы[1], Д. Қуанышбай[2], А. Шойынбек[1], А. Мұхаметжанов[2], 2025.

[1]«Нархоз» Университеті, Алматы, Қазақстан;

[2]Сулейман Демирель атындағы Университеті, Қаскелең, Қазақстан.

E-mail: kazi_golam.rabbany@narxoz.kz

МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНУ АРҚЫЛЫ REDDIT ПОСТТАРЫНДАҒЫ СУИЦИДТІК ТЕНДЕНЦИЯЛАРЫН АНЫҚТАУ

Кази Голам Раббани – «Нархоз» Университеті, Алматы, Қазақстан, E-mail: kazi_golam.rabbany@narxoz.kz, https://orcid.org/0009-0007-4549-0815;

Бекарыстанқызы Ақбаян – аға оқытушы, «Нархоз» Университеті, Алматы, Қазақстан, E-mail: akbayan.b@gmail.com, https://orcid.org/0000-0003-3984-2718;

Шойынбек Айсұлтан – профессор, «Нархоз» Университеті, Алматы, Қазақстан, E-mail: aisultan.shoiynbek@gmail.com, https://orcid.org/0000-0002-9328-8300;

Қуанышбай Дархан – профессор, Сулейман Демирель атындағы Университеті, Қаскелең, Қазақстан, E-mail: darkhan.kuanyshbay@sdu.edu.kz, https://orcid.org/0000-0001-5952-8609;

Мұхаметжанов Асылбек – Сулейман Демирель атындағы Университеті, Қаскелең, Қазақстан, E-mail: 221107046@stu.sdu.edu.kz, https://orcid.org/0009-0009-8528-9985.

Аннотация: Суицид бүкіл әлемде маңызды қоғамдық денсаулық мәселесі болып қала беруде және көптеген жағдайлар алдын ала анықталатын ескерту белгілерімен жүреді. Әлеуметтік медиа платформалары, әсіресе Reddit, пайдаланушылар жазған мәтіндер арқылы суицидтік ойларды анықтаудың құнды ресурсы болып табылады. Бұл зерттеу Reddit жазбаларында суицидтік бейімділікті автоматты түрде анықтау үшін назар механизмі бар екібағытты ұзақ қысқа мерзімді жады (BiLSTM) желілерін пайдаланатын машиналық оқыту моделін ұсынады.

«SuicideWatch» және «depression» субреддиттерінен алынған жазбалардан тұратын деректер жиынтығы Pushshift API арқылы жиналып, суицидпен байланысты немесе байланысты емес деп таңбаланды. Алдын ала өңдеу кезеңі мәтінді кіші әріпке ауыстыру, HTML тегтерін жою, токенизация, мағыналық

жүк көтермейтін сөздерді алып тастау және тізбектерді толықтыру сияқты қадамдарды қамтыды. Деректер жиынтығы теңестіріліп, суицидтік және суицидтік емес жазбалар тең пропорцияда сақталды және олар оқуту (60%), валидация (20%) және тестілеу (20%) жиынтықтарына бөлінді.

Модель архитектурасы ендіру (embedding) қабатын, BiLSTM блоктарын, назар механизмін және толық байланысқан қабаттарды қамтиды. Оқыту процесі Adam оптимизаторы мен бинарлық кроссэнтропиялық шығын функциясын қолдана отырып жүргізілді, ерте тоқтату әдісі қолданылды. Бағалау метрикалары модельдің жоғары өнімділігін көрсетті: валидация жиынтығында F1 көрсеткіші 0.91, тест жиынтығында 0.95, ал AUC-ROC сәйкесінше 0.98 және 0.989 мәндерін көрсетті.

Зерттеу нәтижелері назар механизмі бар BiLSTM-нің суицидтік ойларды анықтаудағы тиімділігін дәлелдеді және көптеген қолданыстағы модельдерден асып түсті. Интернеттегі коммуникацияларда суицидтік бейімділікті ерте анықтау психикалық денсаулық көрсеткіштерін едәуір жақсартып, адамдардың өмірін сақтап қалуға көмектесе алады.

Түйін сөздер: машиналық оқыту, табиғи тілді өңдеу, екі бағытты LSTM, назар аудару механизмі, суицидті анықтау.

К. Раббани[1], А. Бекарыстанкызы[1], Д. Куанышбай[2], А. Шойынбек[1], А. Мухаметжанов[2], 2025.

[1]Университет «Нархоз», Алматы, Казахстан;

[2]Университет имени Сулеймана Демиреля, Каскелен, Казахстан.

E-mail: kazi_golam.rabbany@narxoz.kz

ОБНАРУЖЕНИЕ СУИЦИДАЛЬНЫХ ТЕНДЕНЦИЙ В ПУБЛИКАЦИЯХ НА REDDIT С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Кази Голам Раббани – Университет «Нархоз», Алматы, Казахстан, E-mail: kazi_golam.rabbany@narxoz.kz, https://orcid.org/0009-0007-4549-0815;

Бекарыстанкызы Акбаян – старший преподаватель, Университет «Нархоз», Алматы, Казахстан, E-mail: akbayan.b@gmail.com, https://orcid.org/0000-0003-3984-2718;

Шойынбек Айсултан – профессор, Университет «Нархоз», Алматы, Казахстан, E-mail: aisultan.shoiynbek@gmail.com, https://orcid.org/0000-0002-9328-8300;

Куанышбай Дархан – профессор, Университет имени Сулеймана Демиреля, Каскелен, Казахстан, E-mail: darkhan.kuanyshbay@sdu.edu.kz, https://orcid.org/0000-0001-5952-8609;

Мухаметжанов Асылбек – Университет имени Сулеймана Демиреля, Каскелен, Казахстан, E-mail: 221107046@stu.sdu.edu.kz, https://orcid.org/0009-0009-8528-9985.

Аннотация. Суицид остается серьезной проблемой общественного здравоохранения во всем мире, причем во многих случаях ему предшествуют идентифицируемые предупредительные сигналы. Социальные медиа-платформы, в частности Reddit, являются ценным ресурсом для выявления суицидальных мыслей на основе пользовательских текстов. В данном

исследовании предлагается модель машинного обучения, использующая двунаправленные сети долгой краткосрочной памяти (BiLSTM) с механизмом внимания для автоматического обнаружения суицидальных наклонностей в сообщениях Reddit.

Был собран набор данных, включающий публикации из субреддитов «SuicideWatch» и «depression», с использованием Pushshift API. Сообщения были помечены как связанные с самоубийством или не относящиеся к нему. Предварительная обработка включала приведение к нижнему регистру, удаление HTML-тегов, токенизацию, удаление стоп-слов и дополнение последовательностей. Данные были сбалансированы таким образом, чтобы в них было равное количество суицидальных и несуицидальных сообщений, после чего они были разделены на обучающую (60%), валидационную (20%) и тестовую (20%) выборки.

Архитектура модели включала слой встраивания (embedding), BiLSTM-ячейки, механизм внимания и полностью связанные слои. Обучение проводилось с использованием оптимизатора Adam и бинарной кроссэнтропийной функции потерь с применением механизма ранней остановки. Метрики оценки показали высокую эффективность модели: F1-оценка составила 0.91 на валидационном наборе и 0.95 на тестовом наборе, а AUC-ROC достиг значений 0.98 и 0.989 соответственно.

Результаты исследования демонстрируют эффективность использования BiLSTM с механизмом внимания для выявления суицидальных мыслей, превосходя многие существующие модели. Ранняя идентификация суицидальных наклонностей в онлайн-коммуникациях может значительно улучшить психическое здоровье и спасти жизни.

Ключевые слова: машинное обучение, обработка естественного языка, двунаправленный LSTM, механизм внимания, распознавание суицида.

Introduction. Suicide remains one of the leading causes of death globally and is a significant public health challenge. Suicidal ideation, within the framework of social media, refers to thoughts about ending one's life, encompassing a spectrum that ranges from passive reflections to active planning (Su, 2020). The global burden of suicide is particularly concerning given that many instances of suicidal behavior are preceded by warning signs that could be detected and addressed on time.

Social media platforms like Reddit serve as powerful tools for individuals to share their personal experiences, struggles, and emotions. These platforms offer a unique opportunity for real-time monitoring of mental health conditions, including suicidal ideation. Users often disclose sensitive information in online communities, making social media a fertile ground for identifying individuals at risk. Early detection of suicidal tendencies can facilitate timely interventions, potentially saving lives.

Reddit, in particular, hosts numerous subreddits dedicated to mental health discussions, such as "SuicideWatch" and "depression." These communities provide a space for individuals to seek support and share their experiences. By analyzing

the content of these posts, researchers can identify patterns and linguistic markers associated with suicidal ideation. This approach not only aids in the development of automated detection systems but also enhances the capacity of mental health professionals to respond effectively.

This research aims to develop a robust machine learning model capable of accurately detecting Reddit posts that indicate a serious risk of suicide. By leveraging Bidirectional Long-Short-Term Memory (BiLSTM) networks and attention mechanisms, we seek to enhance the accuracy and reliability of detection systems and automate the detection of suicidal tendencies in online posts.

Related Work. The utilization of social media data for mental health monitoring has gained considerable attention in recent years. Studies have primarily focused on platforms like Twitter and Reddit to predict suicidal ideation and related mental health issues.

One notable study by (Sakib, et al., 2021) analyzed suicidal tweets using ensemble classifiers, achieving an F1 score of 0.90. However, the study also highlighted challenges in verifying the accuracy and effectiveness of such models due to limitations in publicly available datasets, raising concerns about the reliability of social media data for predictive analysis. In a similar vein, researchers conducted a comprehensive study (Satya, et al., 2022) which examined various machine learning approaches for identifying suicidal tendencies on social media platforms such as Facebook, Twitter, and Reddit. They utlized various ML models using features, like TF-IDF and N-gram level TF-IDF, and LSTM. However they didn't calculate important metrics like F1 and AUC-ROC.

Expanding on the challenges of dataset reliability, researchers (Fodeh, et al., 2019) proposed a comprehensive machine learning framework aimed at identifying suicide risk factors on Twitter. The study utilized 12,066 public tweets from 3,873 users, categorizing them into "HighRisk" or "AtRisk" labels based on the usage of suicidal ideation terms and used algorithms such as Decision Trees and K-means Clustering. Notably, the topic discovery approach successfully identified seven out of twelve proposed suicide risk factors. The Decision Tree classification model achieved a precision of 0.844, sensitivity of 0.912, and specificity of 0.829 in distinguishing between "HighRisk" and "AtRisk" users (Fodeh, et al., 2019).

Addressing the issue of dataset size and quality, researchers (Liu, et al., 2022) developed a framework for annotating a mental-health-related textual dataset from Reddit. The focus was on identifying posts and comments that indicated suicide attempts and ideations. Various classifiers achieved commendable performance, with the Suicide Ideation classification reaching an accuracy of 0.86, sensitivity of 0.93, and specificity of 0.77, while Suicide Attempt classification achieved an accuracy of 0.87, sensitivity of 0.68, and specificity of 0.91. Despite the promising results, the relatively small dataset size poses challenges for developing highly effective machine learning models. Additionally, the reliance on expert input for annotating ambiguous samples introduces potential subjective biases, necessitating careful evaluation of its impact on model performance (Liu, et al., 2022).

Another study (Lin, et al., 2024) proposed a RoBERTa-CNN model, achieving 98% accuracy on the Suicide and Depression Detection dataset, emphasizing the critical role of data quality in model training. Reddit communities like r/depression were analyzed for depression and suicide signals, achieving F1 score of 77%. Another study (Oliveira, et al., 2022) explored model interpretability using ELI5, revealing the influence of terms like "suicide" and "sadness" on predictions.

Researchers (Squires, et al., 2024) introduced a semi-supervised deep label smoothing method to improve classification accuracy in suicide risk detection, achieving 52% accuracy on the Reddit C-SSRS dataset, with lower precision than accuracy.

A recent study introduced a hybrid text representation approach designed for explainable suicide risk identification on social media (Naseem, et al., 2024). In this work, they proposed a hybrid text representation approach, achieving an F1 score of 0.79 on a public suicide dataset.

Additionally, another research explored the use of machine learning and natural language processing for depression and suicide analysis (Jain, et al., 2022). In this study, researchers applied ML and NLP techniques to predict posts indicative of depression, specifically analyzing data from Reddit. Their work identified that communities such as r/depression and r/SuicideWatch are pivotal in tracking depressive and suicidal ideation signals. The researchers achieved F1 scores of around 77%. Another study (Yao, et al., 2020) focused on opioid users, achieving an F1 score of 96.6% using CNN to detect suicidality, highlighting the effectiveness of neural networks in specific subgroups.

Another study (Ao, et al., 2021) trained various models, including LSTM, GRU, bidirectional LSTM and GRU, logistic classifier, SVM, XGBoost, and LGBM on a dataset of 232,074 Reddit posts labeled as suicide or non-suicide. Their results demonstrated that one-way LSTM achieved the highest accuracy rate, however lacking F1 and AUC-ROC scores. In a related study (Mirtaheri, et al., 2024), the researchers proposed an ensemble LSTM-TCN model with a self-attention mechanism for detecting suicidal ideation on Twitter and Reddit. Their model, AL-BTCN, achieved 95% F1-score on Reddit dataset of 200,000 posts.

These investigations—ranging from ensemble methods on Twitter posts, to active learning frameworks on Reddit data, and interpretability studies—demonstrate the broad scope of current research in suicidal ideation detection. Given the proven efficacy of LSTM models in capturing temporal dependencies in text data, as highlighted in prior studies, we opted for a Bidirectional LSTM with attention mechanism. Despite methodological differences, all underscore the immense potential and complexity of applying NLP and ML/DL models to user-generated text.

Methodology

Dataset Description. The dataset is a collection of posts from the "SuicideWatch" and "depression" subreddits of the Reddit platform. The posts are collected using Pushshift API. All posts that were made to "SuicideWatch" from Dec 16, 2008 (creation) till Jan 2, 2021, were collected while "depression" posts were collected

from Jan 1, 2009, to Jan 2, 2021. All posts collected from SuicideWatch are labeled as suicide, while posts collected from the depression subreddit are labeled as depression. Non-suicide posts are collected from r/teenagers. It has suicide & non-suicide labels.

Data Preprocessing. Before preprocessing, we checked the class distribution, which was balanced at 50% for suicide and 50% for non-suicidal classes. This ensured that our dataset did not suffer from class imbalance issues.

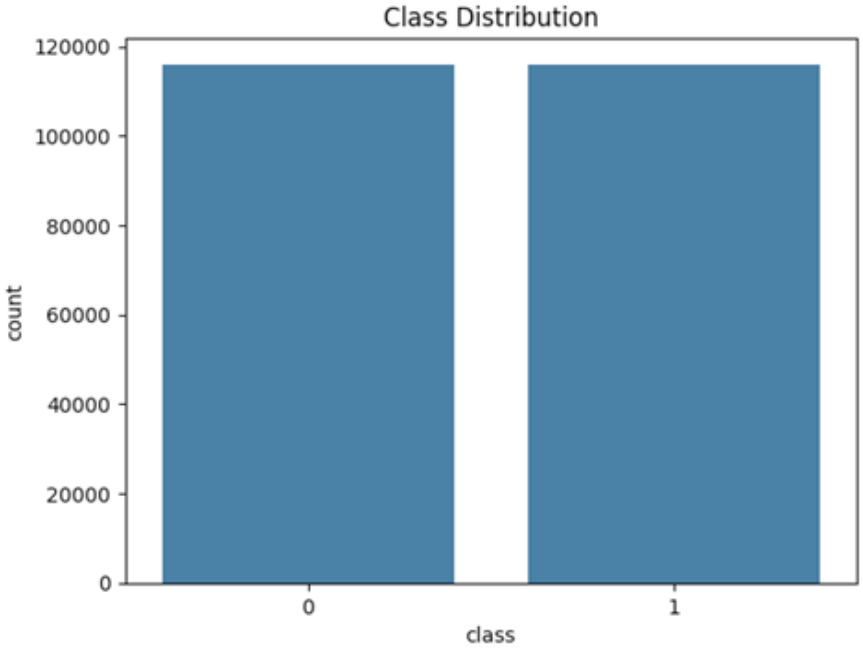

Figure 1. Class distribution

Effective preprocessing is crucial for enhancing the quality of textual data and improving model performance. The preprocessing pipeline implemented in this study included the following steps:

1. Case Folding: All text was converted to lowercase to ensure consistency and reduce variability due to casing differences.

2. HTML Tag Stripping: HTML tags, including iframes and scripts, were removed using BeautifulSoup's strip_html_tags() function to eliminate irrelevant markup content.

3. Accent Removal: Accented characters were normalized to their ASCII equivalents using the remove_accented_chars() function, facilitating uniform text representation.

4. Contraction Expansion: Common contractions were expanded (e.g., "you're" to "you are") using the contractions.fix() method to maintain consistency in word forms.

5. Special Character Removal: Special characters and excess whitespace were removed using regular expressions and the doc.translate() method, simplifying the text.

6. Tokenization: Text was tokenized into individual words using NLTK's word_tokenize(), enabling granular analysis of word usage.

7. Stopwords Removal: Common stopwords were filtered out using the stopwords_removal() function based on NLTK's stopwords list, reducing noise in the data.

8. Normalization: Extra whitespace was trimmed, and leading/trailing spaces were removed to standardize text formatting.

The final normalized posts were recombined into strings and stored as preprocessed text. This preprocessing pipeline ensured that only the most relevant linguistic features were retained, improving the quality of input data for the models.

After preprocessing, the dataset was split into training, validation, and test sets as follows:

The final distribution of samples was:

- Train Samples: 139,244 (60%)
- Validation Samples: 46,415 (20%)
- Test Samples: 46,415 (20%)

These splits ensured a balanced and representative dataset for training, validating, and testing the models.

Tokenization and Vocabulary Building:

Following data splitting, tokenization was performed using Keras' Tokenizer. This step transformed textual data into numerical sequences suitable for model ingestion:

· Tokenizer Configuration: The Tokenizer was instantiated with an out-of-vocabulary token (<UNK>) to handle words not present in the training vocabulary.

· Vocabulary Construction: The tokenizer was fitted on the validation set (X_val), building a vocabulary of 81,536 unique words derived from 46,415 validation documents.

· Sequence Padding: All sequences were padded to a maximum length of 150 words using Keras' pad_sequences(), ensuring uniform input lengths across all samples.

This tokenization process facilitated the conversion of textual data into a format amenable to deep learning models, capturing the essential linguistic features while maintaining computational efficiency.

Padding and Label Encoding:

After tokenization, the sequences were padded to ensure consistent input lengths:

• Padding Sequences: We set a maximum sequence length of 150 words to capture the most significant parts of each post while maintaining manageable input size. Padding was applied to make all sequences uniform in length.

• Label Encoding: The target labels were encoded into numeric form using `LabelEncoder`:

This process ensures that the labels are represented as integers (e.g., 0 and 1), which is required for classification models.

Model Development. The model architecture used for this research is a deep learning model that leverages LSTM and attention mechanisms to capture the temporal dependencies and emphasize key features in the text data:

Model Architecture:

1. Embedding Layer: Transforms input sequences of tokens into dense vector representations. Embedding dimension set to 300, capturing rich semantic information from the 81,536-word vocabulary.

2. Bidirectional LSTM Layer: LSTM units set to 32, allowing the model to learn complex temporal dependencies.

3. Attention Mechanism: Custom attention layer integrated with the BiLSTM outputs.

4. Dense Layers:

- First Dense Layer: 64 units with ReLU activation, facilitating learning of deeper representations.
- Dropout Layer: Applied with a dropout rate of 0.5 to mitigate overfitting.
- Second Dense Layer: 32 units with ReLU activation, further refining feature representations.

5. Output Layer: Produces a probability score indicating the likelihood of suicidal tendencies. Single neuron with sigmoid activation, outputting values between 0 and 1.

Attention Layer. The attention layer was employed to help the model focus on the most important parts of each post. By learning to assign different importance weights to different words in the sequence, the attention mechanism allows the model to focus on the most relevant parts of the input. This layer essentially helps to emphasize the key parts of each Reddit post that are relevant to the prediction task.

Key Features of the Attention Mechanism:

1. Dynamic Weighting: Unlike static attention mechanisms, this implementation allows the model to learn the importance of each word in the context of the entire sequence during training.

2. Masking Support: The layer can handle masked inputs, ensuring that padding tokens do not influence the attention weights, which is crucial for processing sequences of varying lengths.

3. Trainable Parameters: The weight vector W and optional bias b are trainable, enabling the model to adapt the attention scores based on the data during the learning process.

4. Non-linearity: The use of the tanh activation introduces non-linearity, allowing the model to capture more complex relationships between words in the sequence.

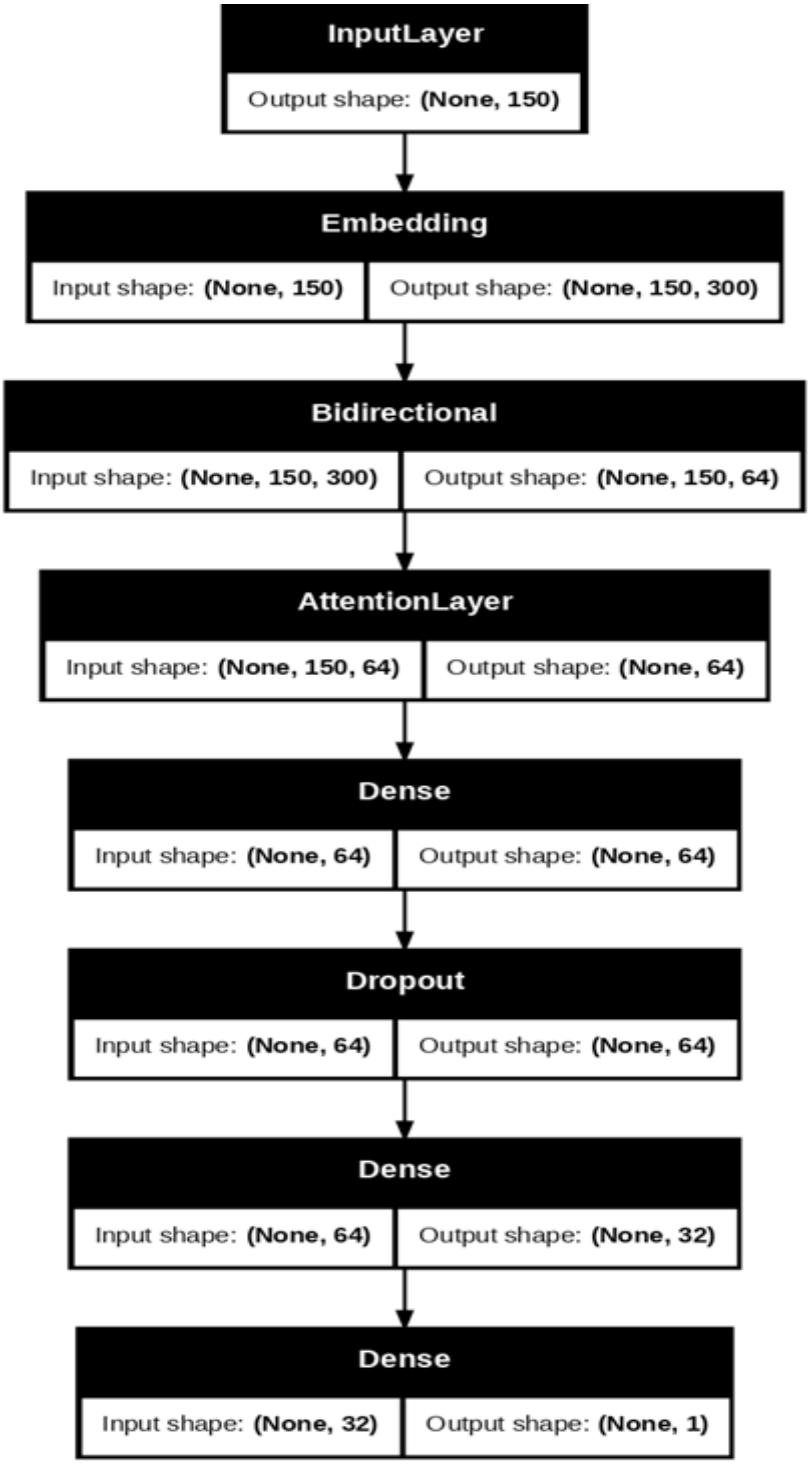

Figure 2. Model structure

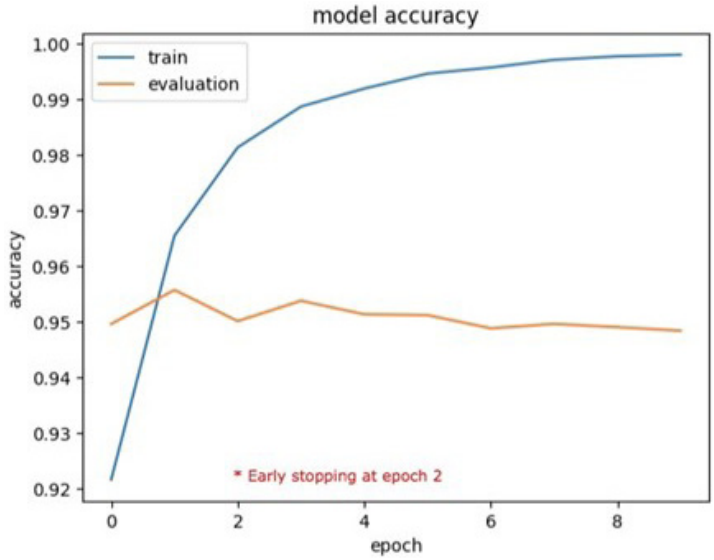

Figure 3. Model accuracy

Training Configuration:

• Loss Function: Binary Cross-Entropy, suitable for binary classification tasks.

• Optimizer: Adam optimizer with a learning rate of 0.001, balancing convergence speed and stability.

• Batch Size: 128 samples per batch, facilitating efficient training without overloading computational resources.

• Epochs: 10, with early stopping based on validation loss to prevent overfitting. The best model was saved at epoch 2 based on the highest accuracy on the evaluation set.

Results. The models were rigorously evaluated using a variety of metrics to assess their performance in detecting suicidal ideation in Reddit posts. The evaluation was conducted on both the validation and test sets to ensure the results' robustness and generalizability. Below are the detailed results for both sets:

Table 2. Long Short-Term Memory (LSTM) Model

Metric	Validation set	Test set
Precision	0.92	0.95
Recall	0.92	0.95
F1 Score	0.91	0.95
AUC-ROC	0.928	0.989

The model achieved an F1 score of 0.91 on the validation set and 0.95 on the test set, reflecting a high degree of accuracy and reliability in detecting suicidal ideation. AUC-ROC provides a measure of the model's ability to distinguish between classes. An AUC-ROC value of 0.98 on the validation set and 0.989 on

the test set indicates that the model has excellent discriminative power, with a near-perfect ability to differentiate between suicidal and non-suicidal posts.

Discussion. Our study demonstrates that integrating a BiLSTM with an attention mechanism is an effective strategy for detecting suicidal ideation in posts from Reddit. Compared to prior research, our model achieved an F1 score that is comparable to or even exceeds previously reported results. In addition, our system's AUC-ROC performance stands out—a key performance indicator that several earlier studies did not consistently include.

Data quality and availability remain crucial. Although our dataset was balanced—50% suicidal posts and 50% non-suicidal posts—this is often not the case in many real-world situations. Researchers frequently rely on balancing techniques, such as oversampling minority classes or undersampling majority classes, which can introduce biases in model performance.

Variations in user-generated text across different platforms introduce additional complexities. Each social media site has its own communication style, vocabulary, and community norms. Twitter users may express ideas in short bursts of text limited by character counts, whereas Reddit posts often contain longer, more detailed narratives.

Conclusion. The high AUC-ROC score achieved by our model underscores its robust discriminative capability, and when considered alongside the strong F1 score, it demonstrates that the combination of BiLSTM and attention mechanisms yields high precision and recall. These outcomes affirm the promise of our approach as a valuable tool for the early detection of suicidal ideation in social media.

For future work, efforts should focus on developing a generic model capable of handling data from multiple social media platforms such as Reddit, Twitter, and others. Leveraging large language models (LLMs) or other methods for deep contextual understanding could enable the creation of a unified system that accommodates the diverse communication styles, vocabulary, and narrative lengths across different platforms. This advancement will enhance the model's generalizability and real-world applicability to improve early detection capabilities for suicidal ideation in a broader online context.

References

Su C., et al. (2020) Machine learning for suicide risk prediction in children and adolescents with electronic health records. Translational Psychiatry. doi: 10.1038/s41398-020-01100-0. (in English)

T.H. Sakib, Md. Ishak, F.F. Jhumu, and Md. A. Ali (2021) Analysis of Suicidal Tweets from Twitter Using Ensemble Machine Learning Methods. International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), IEEE, Jul. 2021, pp. 1–7. doi: 10.1109/ACMI53878.2021.9528252. (in English)

Satya S.S., & Jindal, R. (2022) Social Media Mining For Suicidal Ideation. 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 1466-1471. (in English)

Fodeh, et al. (2019) Using Machine Learning Algorithms to Detect Suicide Risk Factors on Twitter. International Conference on Data Mining Workshops (ICDMW), IEEE, Nov. 2019, pp. 941–948. doi: 10.1109/ICDMW.2019.00137. (in English)

T. Liu, Z. Zheng, Y. Zhou, Y. Yang, and Y. Song (2022) Enriching an online suicidal dataset with active machine learning. Proceedings of the ACM Southeast Conference, New York, NY, USA: ACM, Apr. 2022, pp. 196–200. doi: 10.1145/3476883.3520213. (in English)

Lin E., Sun J., Chen H., & Mahoor M.H. (2024) Data Quality Matters: Suicide Intention Detection on Social Media Posts Using RoBERTa-CNN. 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 1-5. (in English)

A.C. de Oliveira, E.J.S. Diniz, S. Teixeira, and A.S. Teles (2022) How can machine learning identify suicidal ideation from user's texts? Towards the explanation of the Boamente system. Procedia Comput Sci, vol. 206, pp. 141–150. doi: 10.1016/j.procs.2022.09.093. (in English)

Squires M., Tao X., Elangovan S., Acharya R., Gururajan R., Xie H., & Zhou X. (2024) Enhancing Suicide Risk Detection on Social Media through Semi-Supervised Deep Label Smoothing. ArXiv, abs/2405.05795. (in English)

Naseem U., Khushi M., Kim J., & Dunn A.G. (2024) Hybrid Text Representation for Explainable Suicide Risk Identification on Social Media. IEEE Transactions on Computational Social Systems, 11, 4663-4672. (in English)

Jain P., Ram Srinivas K., & Vichare A. (2022) Depression and Suicide Analysis Using Machine Learning and NLP. Journal of Physics: Conference Series, 2161. (in English)

Yao H., Rashidian S., Dong X., Duanmu H., Rosenthal R.N., & Wang F. (2020) Detection of Suicidality Among Opioid Users on Reddit: Machine Learning–Based Approach. Journal of Medical Internet Research, 22. (in English)

Ao Z., Lai J., Lu W., & Mo H. (2021) Comparisons of LSTM and other Machine Learning Approaches on Predicting Suicidal Tendency Regarding to Social Media Posts. International Conference on Computers and Automation (CompAuto), 19-23. (in English)

Mirtaheri S.L., Greco S., & Shahbazian R. (2024) A self-attention TCN-based model for suicidal ideation detection from social media posts. Expert Syst. Appl., 255, 124855. (in English)

CONTENTS

INFORMATION AND COMMUNICATION TECHNOLOGIES

МАЗМҰНЫ

АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР

СОДЕРЖАНИЕ

ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

Publication Ethics and Publication Malpracticein
the journals of the National Academy of Sciences of the Republic of Kazakhstan

For information on Ethics in publishing and Ethicalguidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any otherlanguage, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*
Редакторы: *Д.С. Аленов, Ж.Ш.Әден*
Верстка на компьютере *Г.Д.Жадыранова*

Подписано в печать 20.03.2025.
Формат 60x881/8. Бумага офсетная. Печать –ризограф.
20,0 п.л. Заказ 1.

Национальная академия наук РК
050010, Алматы, ул. Шевченко, 28, т. 272-13-19