

Л. Ю. Емалетдинова, М. С. Пищулин

**КЛАССИФИКАЦИИ ТОНАЛЬНОСТИ ТЕКСТА НА ОСНОВЕ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ***Ключевые слова: LSTM, рекуррентные нейронные сети, тональность текста.*

*В данной работе предлагается методика автоматического определения тональности текстовых отзывов покупателей интернет-магазинов с использованием рекуррентных нейронных сетей, оснащённых двумя LSTM-блоками с 1024 и 128 нейронами, каждый из которых использует ReLU активацию. Для кодирования текста в векторы была взята модель FastText, позволяющая учитывать морфологические особенности русского языка и извлекать информацию из частей слов. При этом два последовательно соединенных LSTM-слоя обеспечивают моделирование долгосрочных контекстуальных зависимостей, важных для анализа текстовых данных. Исследование направлено на преодоление ограничений существующих методов, позволяющих лишь частично учитывать семантические и морфологические особенности языка. В основе предлагаемого подхода лежит автоматическое выделение контекстуальных зависимостей, что особенно актуально при анализе неструктурированных данных. Экспериментальная часть исследования включает разработку модели, обученной на данных 60 тысяч размеченных отзывов, разделенных на тренировочную и тестовую выборки в соотношении 80 на 20, а также сравнительный анализ полученных на ее основе результатов классификации с использованием логистической регрессии и моделей машинного обучения с одним LSTM блоком. Полученные результаты демонстрируют повышение точности классификации (до 90 % на тестовой выборке) за счёт применения двух LSTM-слоёв. Сравнительный анализ показал, что предлагаемая схема превосходит классический алгоритм логистической регрессии и однослойные LSTM-модели по ключевым метрикам (Accuracy  $\approx 0,89$ ; F1-Score  $\approx 0,9$ ; AUC  $\approx 0,95$ ). Методика, предложенная в работе, позволяет качественно анализировать эмоциональную окраску текстов при обработке больших объемов данных. Предложенный подход демонстрирует высокую практическую значимость для масштабируемого мониторинга обратной связи и может быть расширен за счёт интеграции механизмов внимания (self-attention) и гибридных архитектур, объединяющих сильные стороны RNN и трансформеров.*

L. Yu. Emaletdinova, M. S. Pishchulin

**TEXT SENTIMENT CLASSIFICATION BASED ON NEURAL NETWORK MODELING***Keywords: LSTM, recurrent neural networks, text tone.*

*In this work, we propose a method for automatically determining the sentiment of text reviews posted by online store customers using recurrent neural networks equipped with two LSTM blocks with 1024 and 128 neurons, each employing ReLU activation. For text encoding into vectors, we use the FastText model, which captures the morphological nuances of the Russian language by extracting information from word sub-units. The two successive LSTM layers enable modeling of long-term contextual dependencies, crucial for analyzing textual data. This study aims to overcome the limitations of existing methods that only partially account for semantic and morphological language features. At the core of our approach is the automatic extraction of contextual dependencies, which is especially relevant when processing unstructured data. The experimental part of the research consists of developing a model trained on a dataset of 60000 annotated reviews, split into training and test sets in an 80:20 ratio, and conducting a comparative analysis of its classification results against those obtained with logistic regression and single-layer LSTM models. The results demonstrate an increase in classification accuracy – up to 90 % on the test set – through the use of two LSTM layers. Comparative evaluation shows that our two-layer LSTM architecture outperforms the classical logistic regression algorithm and single-layer LSTM models across key metrics (Accuracy  $\approx 0.89$ ; F1-Score  $\approx 0.90$ ; AUC  $\approx 0.95$ ). The proposed methodology enables effective sentiment analysis of large volumes of text data. Moreover, it exhibits high practical relevance for scalable monitoring of user feedback and can be extended by integrating attention mechanisms (self-attention) and hybrid architectures that combine the strengths of RNNs and Transformers.*

**Введение**

Анализ тональности отзывов покупателей интернет-магазинов представляет собой актуальную задачу в условиях стремительного развития электронной коммерции. Ежедневно в интернет-среде генерируется огромное количество неструктурированных текстовых данных, отражающих мнения пользователей о товарах и услугах. Корректная интерпретация эмоциональной окраски этих отзывов позволяет предприятиям оперативно реагировать на критику, улучшать качество обслуживания и выстраивать эффективную маркетинговую стратегию.

**Анализ публикаций**

Традиционные методы анализа текста, основанные на статистических и лингвистических

подходах, часто оказываются недостаточно гибкими для учета сложных семантических и синтаксических зависимостей в естественном языке [1-4].

В работе [2] используются статистические методы для оценки сходства текстовых документов. Используемый подход предполагает ручную настройку значимости отдельных слов, что может привести к ошибкам и затруднить использование на больших объемах текстовых данных. Более того, использование статистических мер не позволяет полностью учитывать контекст: предложения с одинаковым набором слов в различном порядке могут оцениваться как несхожие, а полисемия (наличие у слова двух и более значений) не будет учитываться. Нейронные сети, в свою очередь, позволяют избежать данных проблем, так как

автоматически закладывают значимость и контекст слов при их кодировании во время обучения.

В работе [4] для тематической классификации текстов применяются предварительная обработка текста и классические алгоритмы машинного обучения, такие как логистическая регрессия, деревья решений, метод k-ближайших соседей и другие. Текст кодируется методами Bag Of Words и TF-IDF, что приводит к потере семантических и контекстуальных связей — критически важного аспекта для смыслового анализа. Нейронные сети, напротив, извлекают зависимости из исходных данных, основываясь на глобальном контексте, и не требуют сложной предобработки, снижая риск утраты полезной информации.

### Подход на основе нейронных сетей

Современные методы глубокого обучения, в частности рекуррентные нейронные сети (Recurrent neural network – RNN), открывают новые возможности для анализа тональности текстов. Благодаря своей способности учитывать последовательность входных данных, RNN хорошо справляются с задачами, где контекст играет ключевую роль. Однако стандартные RNN подвержены проблемам исчезающего и взрывающегося градиента [5], что затрудняет обучение на длинных последовательностях. Для преодоления этих трудностей были разработаны специализированные архитектуры, такие как блоки долгой краткосрочной памяти (Long short-term memory – LSTM) [6], которые эффективно сохраняют и передают релевантную информацию на протяжении всей последовательности.

В данной работе предлагается метод определения тональности отзывов покупателей на основе рекуррентных нейронных сетей с использованием LSTM-блоков. Важным компонентом предлагаемого подхода является применение модели FastText [7-8] для преобразования текстовых данных в векторное представление. Модель FastText учитывает составляющие части слова, что позволяет получать более точные векторы даже для редких или морфологически сложных слов, характерных для русского языка.

### Алгоритм построения модели

Для решения поставленной задачи был сформирован набор отзывов, собранных с различных интернет-платформ. Каждый отзыв предварительно размечался экспертами, что обеспечивало наличие двух классов тональности.

На рис.1. представлена схема работы предлагаемого алгоритма классификации тональности текста с использованием нейронной сети. На начальном этапе проводилась предварительная обработка текстов: приведение к нижнему регистру, удаление пунктуации, специальных символов и «стоп-слов», не обладающих важной информацией (предлоги, союзы и т.п.), что способствовало снижению шума в данных. После чего происходит разбиение текста на

слова (токенизация), а также приведение к нормальной форме (лемматизация).

Ключевым этапом обработки является векторизация текстовых данных. Для этой цели использовалась модель FastText, предварительно обученная на обширном наборе русскоязычных текстов. Преимущество модели FastText заключается в его способности представлять слова не только как отдельные единицы, но и учитывать фрагментную структуру через n-граммы, что существенно повышает качество кодирования слов для языков с богатой морфологией. Полученные векторы использовались для формирования входных последовательностей модели фиксированной длины, при этом короткие тексты дополнялись специальными токенами (PAD), которые отвечают за «пустой» текст.

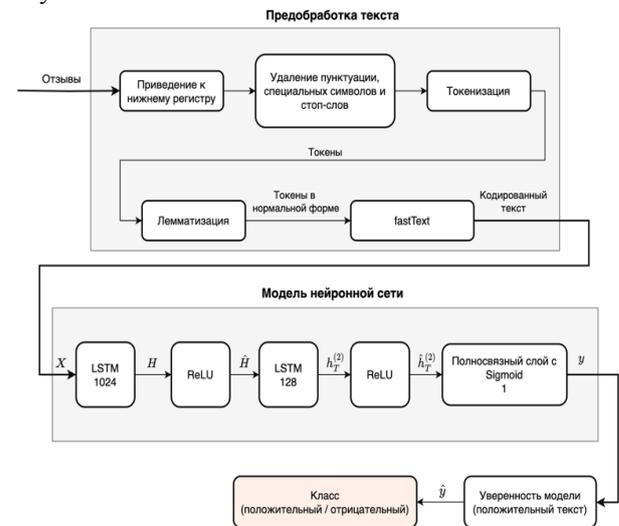


Рис. 1 – Схема работы алгоритма классификации тональности текста

Fig. 1 – Schematic representation of the text tone classification algorithm

Архитектура модели нейронной сети представляет собой рекуррентную нейронную сеть с LSTM-блоками. На вход модели подается кодированный текст – тензор фиксированной длины  $X = (x_1, x_2, \dots, x_T)$ ,  $x_t \in \mathbb{R}^d$ , где  $x_t$  – закодированное слово на шаге  $t$ ,  $T$  – длина отзыва,  $d$  – размерность векторов слов.

Первый LSTM-слой на каждом шаге  $t$  принимает вектор  $x_t$  и на выходе формирует тензор  $H = (h_1^{(1)}, h_2^{(1)}, \dots, h_T^{(1)})$ ,  $h_t^{(1)} \in \mathbb{R}^q$ , где  $h_t^{(1)}$  – скрытое состояние слоя после обработки  $t$  слов, а  $q$  – количество элементов вектора скрытого состояния. Далее к каждому  $h_t^{(1)}$  применяется функция активации ReLU [9], в результате чего получаем тензор  $\hat{H}$ . Таким образом, первый LSTM-слой формирует карты признаков для токенов, которые отражают важные семантические и синтаксические зависимости.

Второй LSTM слой получает на вход тензор  $\hat{H}$ . Далее работа этого слоя аналогична первому, за исключением того, что выходом будет являться не вся последовательность скрытых состояний, а только

последний вектор  $h_r^{(2)}$ . Этот вектор можно интерпретировать, как вектор, содержащий в себе ключевые признаки всего текста, важные для определения тональности отзыва.

Финальным этапом является использование полносвязного выходного слоя с функцией Sigmoid [10], формирующий вероятностное распределение по классам тональности. Обучение модели проводилось с использованием функции потерь Binary Cross-entropy и метода стохастической оптимизации Adam (Adaptive Momentum) [11-12].

Стоит отметить, что в ряде исследований по анализу тональности текстов [14-16] применяется векторизация с использованием Word2Vec [17], а архитектура модели ограничивается одним слоем LSTM, после которого результаты передаются на полносвязный слой с сигмоидальной функцией активации для формирования окончательного распределения по классам. В отличие от такого подхода, в данной работе используется модель FastText для векторизации текста, что позволяет более точно учитывать морфологические особенности и редкие слова за счёт использования n-грамм. Дополнительное применение двух слоев LSTM даёт возможность лучше моделировать сложные и длительные зависимости в последовательности, а включение слоев с функцией активации ReLU способствует более эффективному нелинейному преобразованию извлечённых признаков. Таким образом, предложенные модификации позволят привести к повышению качества анализа тональности отзывов благодаря более точному извлечению контекстной информации.

### Экспериментальная часть

Параметры модели (количество токенов для отзывов, размеры обучаемых параметров в LSTM-блоках, количество LSTM слоев) подбирались эмпирически. Обучение и тестирование рассмотренных архитектур моделей проводилось на размеченной выборке, состоящей из 60 тысяч отзывов, с разделением данных на обучающую и тестовую выборки в соотношении 80% и 20%. Для сравнения с классическими методами анализа текстов проводился эксперимент с использованием алгоритма логистической регрессии и моделей машинного обучения с одним LSTM блоком (таблица 1).

В табл. 1 рассматриваются модели:

- 1 – Логистическая регрессия
- 2 – Один LSTM-блок с векторизацией Word2Vec
- 3 – Один LSTM-блок с векторизацией FastText
- 4 – Два LSTM-блока со слоями ReLU и векторизацией FastText.

Метрика Accuracy характеризует точность и вычисляется как:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

где TP – число правильно предсказанных «положительных» классов; TN – число правильно предсказанных «отрицательных» классов; FP – число неверно предсказанных «положительных» классов;

FN – число неверно предсказанных «отрицательных» классов.

Таблица 1 – Сравнение результатов тестирования обученных моделей

Table 1 – Comparison of test results of trained models

Модели	Метрика Accuracy	Метрика F1-Score	Метрика AUC
1	0.72	0.77	0.71
2	0.8	0.81	0.87
3	0.82	0.83	0.9
4	0.89	0.9	0.95

Для расчета метрики F1-score, характеризующей соотношение полноты и точности классификации, используется выражение:

$$F1\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

где  $Precision = \frac{TP}{TP+FP}$   $Recall = \frac{TP}{TP+FN}$

Метрика AUC (area under curve) характеризует прогностическую силу модели и вычисляется как площадь под ROC-кривой [18]. Сама ROC-кривая отражает зависимость чувствительности TPR от специфичности FPR, которые вычисляются как:

$$TPR = \frac{TP}{TP+FN} \text{ от } FPR = \frac{FP}{FP+TN}$$

Полученные результаты показали, что модель на базе LSTM и FastText превосходит традиционные методы по качеству классификации текста, что обусловлено способностью рекуррентной архитектуры учитывать контекстуальные зависимости. При этом использование модели FastText способствовало не только улучшению качества классификации, но и ускорению процесса обработки данных за счет более быстрой ее работы.

На рис.2 представлен график обучения модели с двумя слоями LSTM с ReLU и векторизацией FastText. Из рисунка видно, что точность классификации 87% достигается на 8-10 эпохах. Анализ показателей выявил, что наилучшие результаты достигаются при различении явно выраженных положительных и отрицательных отзывов, в то время как нейтральные отзывы иногда классифицируются с меньшей уверенностью, что связано с их неоднозначной языковой структурой.

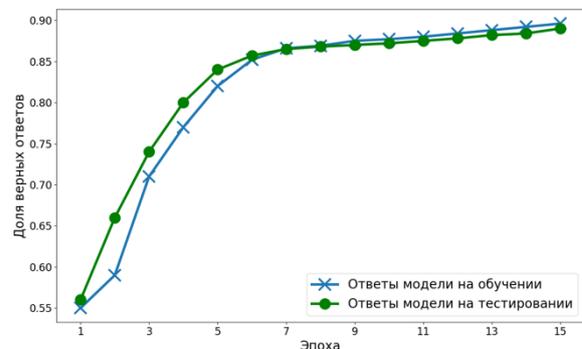


Рис. 2 – График обучения модели с двумя слоями LSTM с ReLU и векторизацией FastText

Fig. 2 – Training schedule for the model with two LSTM layers with ReLU and FastText vectorization

Добавление третьего слоя LSTM не давало значительного прироста в качестве, но сильно замедляло процесс обучения. Таким образом, наилучшим решением оказалось использование размера отзывов в 300 элементов, совместно с двумя LSTM-слоями с параметрами размера 1024x300 для первого слоя и 128x1024 для второго слоя.

Качественный анализ ошибок классификации выявил, что модель испытывает затруднения при интерпретации саркастических высказываний и отзывов, содержащих сложные синтаксические конструкции. Эти случаи часто связаны с тем, что тональность такого текста может зависеть от контекста, подразумеваемого за пределами явного смыслового содержания. В будущем возможно интегрирование механизмов внимания (self-attention), что позволит модели более точно фокусироваться на ключевых элементах текста и, таким образом, повысить качество классификации сложных отзывов.

### Заключение

Предложенный подход с использованием рекуррентных нейронных сетей, оснащенных двумя LSTM-блоками, в сочетании с моделью FastText для векторизации текста, демонстрирует высокую эффективность при решении задачи анализа тональности отзывов. Полученные экспериментальные результаты подтверждают конкурентоспособность данного метода по сравнению с традиционным алгоритмом, что делает его перспективным для практического применения в системах мониторинга обратной связи и улучшения клиентского сервиса. Дальнейшие исследования могут быть направлены на расширение обучающей выборки, внедрение дополнительных механизмов для обработки сложных языковых конструкций, а также интеграцию с другими моделями глубокого обучения (например, CNN и RNN или Transformer), что позволит еще больше повысить точность и устойчивость системы.

### Литература

1. Янников, И. М. Методы и алгоритмы для поиска сходства между текстами / И. М. Янников, М. В. Ершова, А. Н. Исенбаев // Интеллектуальные системы в производстве. – 2024. – Т. 22, № 2. – С. 103-113. – DOI 10.22213/2410-9304-2024-2-103-113. – EDN BZJCSI.
2. Chandrasekaran, D. Evolution of Semantic Similarity—A Survey / D. Chandrasekaran, V. Mago // ACM Computing Surveys. — 2022. — Т. 54, № 2. — С. 1–37. — DOI: 10.1145/3440755.
3. Хорошилов А.А., Кан А.В., Евдокимова Е.А., Пицхелаури С.Г. Установление сходства текстовых документов // Моделирование и анализ данных. 2023. Том 13. № 4. С. 45–58. DOI: 10.17759/mda.2023130403.
4. Денисов, Д. В. Анализ методов машинного обучения для тематической классификации текстов / Д. В. Денисов // Международный журнал информационных технологий и энергоэффективности. – 2024. – Т. 9, № 4(42). – С. 5-11. – EDN IBXFCW.
5. R. Pascanu, T. Mikolov, Y. Bengio. On the difficulty of training Recurrent Neural Networks / R. Pascanu, T. Mikolov, Y. Bengio [Электронный ресурс] // arXiv.org : [сайт]. — URL: <https://arxiv.org/abs/1211.5063> (дата обращения: 23.03.2025).

6. Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-Term Memory. Neural Computation. 9. 1735-1780. 10.1162/neco.1997.9.8.1735.
7. A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov. FastText.zip: Compressing text classification models / A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov. [Электронный ресурс] // arXiv.org : [сайт]. — URL: <https://arxiv.org/abs/1612.03651> (дата обращения: 24.03.2025).
8. A. Joulin, E. Grave, P. Bojanowski, T. Mikolov. Bag of Tricks for Efficient Text Classification / A. Joulin, E. Grave, P. Bojanowski, T. Mikolov. [Электронный ресурс] // arXiv.org : [сайт]. — URL: <https://arxiv.org/abs/1607.01759> (дата обращения: 24.03.2025).
9. Соснин, А. С. Функции активации нейросети: сигмоида, линейная, ступенчатая, relu, tahn / А. С. Соснин, И. А. Сулова // Наука. Информатизация. Технологии. Образование : Материалы XII международной научно-практической конференции, Екатеринбург, 25 февраля – 01 2019 года. – Екатеринбург: Российский государственный профессионально-педагогический университет, 2019. – С. 237-246. – EDN VUZYBA.
10. Бунина, Л. В. Функции активации нейронных сетей / Л. В. Бунина, М. А. Лихачев, А. П. Титов // Проблемы информатики в образовании, управлении, экономике и технике : сборник статей XXII Международной научно-технической конференции, Пенза, 09–10 декабря 2022 года / Пензенский государственный университет. – Пенза: Автономная некоммерческая научно-образовательная организация «Приволжский Дом знаний», 2022. – С. 58-62. – EDN LVDNFG.
11. Градиентные алгоритмы оптимизации в подборе гиперпараметров моделей машинного обучения scikit-learn / К. В. Парфентьев, Д. Д. Пронин, В. К. Борисов, В. А. Кабаков // Современные наукоемкие технологии. – 2023. – № 12-1. – С. 67-74. – DOI 10.17513/snt.39862. – EDN GZUKJM.
12. D. P. Kingma, J. Ba. Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba [Электронный ресурс] // arXiv.org : [сайт]. — URL: <https://arxiv.org/abs/1412.6980> (дата обращения: 24.03.2025).
13. Оптимизация в ML // Яндекс.Образование URL: <https://education.yandex.ru/handbook/ml/article/optimizaciya-v-ml> (дата обращения: 05.05.2025).
14. Лыченко, Н. М. Сравнение эффективности методов векторного представления слов для определения тональности текстов / Н. М. Лыченко, А. В. Сорокина // Математические структуры и моделирование. – 2019. – № 4(52). – С. 97-110. – DOI 10.24147/2222-8772.2019.4.97-110. – EDN JCKZBV.
15. Khattak, F. K. A survey of word embeddings for clinical text / F. K. Khattak, S. Jeblee, Ch. Pou-Prom, M. Abdalla, C. Meaney, F. Rudzicz // Journal of Biomedical Informatics. — 2019. — Т. 100, № S. — С. 1–18. — DOI: 10.1016/j.yjbinx.2019.100057.
16. Самигулин, Т. Р. Реализация алгоритма анализа тональности текста в чате видеоконференцсвязи / Т. Р. Самигулин // Сборник трудов X Конгресса молодых ученых : Материалы Конгресса, Санкт-Петербург, 14–17 апреля 2021 года. Том 2. – Санкт-Петербург: федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет ИТМО", 2021. – С. 266-269. – EDN OTLCSB.
17. T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space / T. Mikolov, K. Chen, G. Corrado, J. Dean [Электронный ресурс] // arXiv.org : [сайт]. — URL: <https://arxiv.org/abs/1301.3781> (дата обращения: 02.04.2025).

18. AUC ROC (площадь под кривой ошибок)// Анализ малых данных URL: <https://alexanderdyakonov.wordpress.com/2017/07/28/auc-roc-площадь-под-кривой-ошибок> (дата обращения: 05.05.2025).

### References

1. Yannikov, I. M. Methods and algorithms for searching similarity between texts / I. M. Yannikov, M. V. Ershova, A. N. Isenbaev // Intellectual Systems in Production. - 2024. - Т. 22, № 2. - С. 103-113. - DOI 10.22213/2410-9304-2024-2-103-113. - EDN BZJCSI.
2. Chandrasekaran, D. Evolution of Semantic Similarity-A Survey / D. Chandrasekaran, V. Mago // ACM Computing Surveys. - 2022. - Т. 54, № 2. - С. 1-37. - DOI: 10.1145/3440755.
3. Khoroshilov, A.A.; Kahn, A.V.; Evdokimova, E.A.; Pitskhelauri, S.G. Similarity detection of text documents // Data modeling and analysis. 2023. Vol. 13. No. 4. С. 45-58. DOI: 10.17759/mda.2023130403
4. Denisov, D. V. Analysis of machine learning methods for thematic classification of texts / D. V. Denisov // International Journal of Information Technologies and Energy Efficiency. - 2024. - Т. 9, № 4(42). - С. 5-11. - EDN IBXFCW.
5. R. Pascanu, T. Mikolov, Y. Bengio. On the difficulty of training Recurrent Neural Networks / R. Pascanu, T. Mikolov, Y. Bengio. Bengio [Electronic resource] // arXiv.org : [website]. - URL: <https://arxiv.org/abs/1211.5063> (accessed 23.03.2025).
6. Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-Term Memory. Neural Computation. 9. 1735-1780. 10.1162/neco.1997.9.8.1735.
7. A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov. FastText.zip: Compressing text classification models / A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov. [Electronic resource] // arXiv.org : [website]. - URL: <https://arxiv.org/abs/1612.03651> (date of reference: 24.03.2025).
8. A. Joulin, E. Grave, P. Bojanowski, T. Mikolov. Bag of Tricks for Efficient Text Classification / A. Joulin, E. Grave, P. Bojanowski, T. Mikolov. [Electronic resource] // arXiv.org : [website]. - URL: <https://arxiv.org/abs/1607.01759> (date of reference: 24.03.2025).
9. Sosnin, A. S. Neural network activation functions: sigmoid, linear, step, relu, tahn / A. S. Sosnin, I. A. Suslova // Nauka. Informatization. Technologies. Education: Proceedings of the XII International Scientific and Practical Conference, Ekaterinburg, February 25 - 01, 2019. - Yekaterinburg: Russian State Professional and Pedagogical University, 2019. - С. 237-246. - EDN VUZYBA.
10. Bunina, L. V. Activation functions of neural networks / L. V. Bunina, M. A. Likhachev, A. P. Titov // Problems of informatics in education, management, economics and technology : proceedings of the XXII International Scientific and Technical Conference, Penza, December 09-10, 2022 / Penza State University. - Penza: Autonomous non-profit scientific and educational organization "Volga Region House of Knowledge", 2022. - С. 58-62. - EDN LVDNFG.
11. Gradient optimization algorithms in hyperparameter selection of scikit-learn machine learning models / K. V. Parfentyev, D. D. Pronin, V. K. Borisov, V. A. Kabakov // Modern knowledge-intensive technologies. - 2023. - № 12-1. - С. 67-74. - DOI 10.17513/snt.39862. - EDN GZUKJM.
12. D. P. Kingma, J. Ba. Adam: A Method for Stochastic Optimization / D. P. Kingma, J. Ba [Electronic resource] // arXiv.org : [website]. - URL: <https://arxiv.org/abs/1412.6980> (date of reference: 24.03.2025).
13. Optimization in ML // Yandex.Education URL: <https://education.yandex.ru/handbook/ml/article/optimizaciy-a-v-ml> (date of address: 05.05.2025).
14. Lychenko, N. M. Comparison of the effectiveness of methods of vector representation of words for determining the tone of texts / N. M. Lychenko, A. V. Sorokova // Mathematical Structures and Modeling. - 2019. - № 4(52). - С. 97-110. - DOI 10.24147/2222-8772.2019.4.97-110. - EDN JCKZBV.
15. Khattak, F. K. A survey of word embeddings for clinical text / F. K. Khattak, S. Jeblee, Ch. Pou-Prom, M. Abdalla, C. Meaney, F. Rudzicz // Journal of Biomedical Informatics. - 2019. - VOL. 100, NO. S. - С. 1-18. - DOI: 10.1016/j.yjbinx.2019.100057.
16. Samigulin, T. R. Implementation of the algorithm of text tone analysis in a videoconference chat room / T. R. Samigulin // Proceedings of the X Congress of Young Scientists : Proceedings of the Congress, St. Petersburg, April 14-17, 2021. Volume 2. - St. Petersburg: Federal State Autonomous Educational Institution of Higher Education "National Research University ITMO", 2021. - С. 266-269. - EDN OTLCSB.
17. T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space / T. Mikolov, K. Chen, G. Corrado, J. Dean. Mikolov, K. Chen, G. Corrado, J. Dean [Electronic resource] // arXiv.org : [website]. - URL: <https://arxiv.org/abs/1301.3781> (date of reference: 02.04.2025).
18. AUC ROC (area under the error curve)// Small Data Analysis URL: <https://alexanderdyakonov.wordpress.com/2017/07/28/auc-roc-площадь-под-кривой-ошибок> (accessed 05.05.2025).

© Л. Ю. Емалетдинова – д-р техн. наук, профессор кафедры Прикладной математики и информатики (ПМИ), Казанский национальный исследовательский технический университет им. А.Н. Туполева (КНИТУ им. А.Н. Туполева), Казань, Россия, [lilia@stcline.ru](mailto:lilia@stcline.ru); М. С. Пищулин – бакалавр кафедры ПМИ, КНИТУ им. А.Н. Туполева, [renwok22@gmail.com](mailto:renwok22@gmail.com).

© L. U. Emaletdinova – Doctor of Sciences (Technical Sci.), Professor of the Department of Applied Mathematics and Informatics (AMI), Kazan National Research Technical University named after A.N. Tupolev (KNRTU named after A.N. Tupolev), Kazan, Russia, [lilia@stcline.ru](mailto:lilia@stcline.ru); M. S. Pishchulin – Bachelor-Student of the AMI department, KNRTU named after A.N. Tupolev, [renwok22@gmail.com](mailto:renwok22@gmail.com).

Дата поступления рукописи в редакцию – 27.05.25.

Дата принятия рукописи в печать – 10.06.25.