DOI: 10.55648/1998-6920-2025-19-3-52-60 УДК 004.852

Классификатор речевой агрессии в русскоязычных неструктурированных текстах

И. Е. Воронина, М. К. Пастревич

Воронежский государственный университет (ФГБОУ ВО ВГУ)

Аннотация: рассматриваются проблемы классификации языковой агрессии в неструктурированных текстах информационного пространства. Анализируются различные подходы к определению понятия «языковая агрессия», а также методы её выявления и оценки. Особое внимание уделяется вопросам автоматической обработки текстов для обнаружения агрессивных высказываний. Описываются существующие алгоритмы и модели машинного обучения, применяемые для решения этой задачи. Помимо этого, обсуждаются перспективы развития технологий анализа текста и их применения в борьбе с распространением вербальной агрессии в интернете.

Ключевые слова: классификация, модель, обучение, языковая агрессия, нейронная сеть.

Для цитирования: Воронина И. Е., Пастревич М. К. Классификатор речевой агрессии в русскоязычных неструктурированных текстах // Вестник СибГУТИ. 2025. Т. 19, № 2. С. 52–60. https://doi.org/ 10.55648/1998-6920-2025-19-3-52-60.

Статья поступила в редакцию 22.03.2025;
переработанный вариант – 03.05.2025;
принята к публикации 14.05.2025.

1. Введение

Современный мир характеризуется стремительным ростом объемов информации, циркулирующей в цифровой среде. Социальные сети, форумы, чаты и другие формы коммуникации стали неотъемлемой частью повседневной жизни миллионов людей. Вместе с тем, рост общения в информационном пространстве приводит к увеличению проявлений речевой агрессии, которая может принимать самые разные формы: от явных оскорблений до скрытого сарказма и манипуляций. Языковая агрессия оказывает значительное воздействие на психоэмоциональное состояние людей, способствует возникновению конфликтов и может служить катализатором негативных социальных явлений.

Классификация речевой агрессии является важным шагом для понимания природы этого явления и разработки методов его выявления и предотвращения. Однако до сих пор остается множество нерешенных вопросов относительно того, какие именно признаки следует учитывать при определении уровня агрессии в неструктурированном тексте, а также каким образом автоматизировать этот процесс.

Поскольку вербальная агрессия – многоаспектное социокультурное явление, имеющее целью унижение оппонента, создание дисгармоничной психологической среды, его следует рассматривать как с лингвистической, так и с психологической точек зрения. Это позволяет не только понять, какие средства используются для достижения агрессивного эффекта, но и как они воспринимаются и интерпретируются участниками коммуникации. Такой комплексный подход важен для разработки эффективной автоматической классификации вербальной агрессии в неструктурированных текстах.

Поскольку явление речевой агрессии затрагивает пользователей разных стран, исследования в этой области проводились как зарубежными, так и отечественными учеными. Особое внимание стоит уделить русскоязычному контенту, так как именно многоаспектность понятий в русскоязычном языке представляет определенную трудность в области автоматической классификации вербальной агрессии.

Вопросами классификации языковой агрессии занимались многие лингвисты, например, Горбачева Е. Ю. [1], Денисова А. В. [2], Щербинина Ю. В. [3], Шейгал Е. И. [4],Бобр А. Д. [17]. В работах [1, 2, 3] представлены классификации, которые в полном объеме подходят для анализа структурированного текста. В исследовании [17] расширено понятие коммуникативного пространства социальных сетей, характеризующееся как конфликтогенная среда, а также расширены маркеры вербальной агрессии, используемые в социальных сетях. Следует учесть, что лингвистами не решались вопросы создания автоматической классификации.

Среди психологов вопросами классификации вербальной агрессии занимались такие ученые, как Басс А. [5], Выготский Л. С. [6], Ениколопов С. Н. [7], Фромм Э. [8], но они не ставили перед собой задачи классификации неструктурированных текстов и автоматизации данного процесса.

Следует отметить, что исследования в области вербальной агрессии и кибербуллинга ведутся учеными разных стран, например, Чуо Ху [18], Акхтер Т. М. [19], Кхан У. [20], Алматарнех С.[21], Миок К. [22]. В исследовании [18] автор, работая с набором данных, состоящим из 20 категорий и 18828 непустых документов, приходит к выводу, что на таком наборе оптимально использование мультиномиального байесовского классификатора с параметрами Дирихле имеет аналогичную производительность с обычным мультиномиальным байесовским классификатором. В работе [19] автор с помощью различных моделей (NB,SVM и т. д.) классифицируют агрессию на языке урду и латинский урду, в результате приходят к выводу, что сверточные нейросети больше подходят для решения поставленной задачи. В исследовании [20] автор классифицирует англоязычные твиты по восьми эмоциональным признакам. Для решения поставленной задачи им была использована модель DNN. В работе [21] исследователь классифицирует вербальную агрессию в английском и испанском языках с использованием различных классификаторов. В процессе работы лучшие результаты показали SVM (опорная векторная машина), RF (случайный лес) и CNB (комплементарный наивный байесовский алгоритм). В исследовании [21] авторы рассматриваются англоязычные твиты, содержащие агрессивные высказывания, но для решения задачи классификации предлагают BERT.

Классификацией неструктурированных текстов занимались отечественные исследователи, например, Макарова Е. А.[9], Котельников Е. В. [10]. В исследовании [9] помимо повышения эффективности линейной классификации неструктурированных текстов (в основу которых были положены новостные статьи), но и решается задача автоматического распознавания террористического и радикального контента. В работе [10] решается задача повышения точности при исследовании разнообразных мнений в текстах.

В связи с вышеизложенным, разработка программной реализации для классификации языковой агрессии в русскоязычных неструктурированных текстах остается актуальной.

2. Материалы и методы

Классификация агрессии с точки зрения психологии имеет следующий вид [5]:

1) физическая – активная – прямая, представляется в нанесении различных ударов;
2) физическая – активная – непрямая, чаще всего проявляется в различных разговорах с врагом;
3) физическая – пассивная – прямая, выражающаяся в стремлении физически не позволять другому человеку достигнуть поставленной цели или заниматься желаемой деятельностью;

4) физическая – пассивная – непрямая, заключающаяся в отказе от выполнения поставленных задач;

5) вербальная – активная – прямая, проявляется в вербальном оскорблении или унижении собеседника;

6) вербальная – активная – непрямая, выражается в намеренном распространении клеветы или сплетен о другом индивидууме;

7) вербальная – пассивная – прямая, заключающаяся в отказе от общения с оппонентом;

8) вербальная – пассивная – непрямая, проявляется в отказе от предоставления различных словесных объяснений оппоненту.

С учетом многообразия классификаций речевой агрессии [11, 12, 13, 14] для машинного обучения наиболее подходящей является [4]:

– Эксплицитная (оскорбления, прямые угрозы, угрожающие заявления, например: «Я тебе сломаю руку», «Лучше бы ты держал рот закрытым, иначе пожалеешь»).

– Манипулятивная (газлайтинг, например: «Ты просто преувеличиваешь, ничего такого не было»).

– Имплицитная (ирония или сарказм, например: «Ой, конечно, ты такой умный, раз решил это сделать!»).

Помимо этого, для корректной работы необходимо расширить выбранную классификацию и добавить еще одно условие – «нейтральная лексика».

Поскольку готовые предобученные модели (`S-nlp/russian_toxicity_classifier`, `Cointegrated/rubert-tiny-toxicity`, `Sismetanin/rubert-toxic-pikabu-2ch`, `IlyaGusev/rubertconv_toxic_clf`, `Tinkoff-ai/response-toxicity-classifier-base`) в ходе проведенных экспериментов [23, 24] оказались непригодны для реализации выбранной классификации, было принято решение о создании собственного классификатора.

В ходе разработки программной реализации классификации языковой агрессии в русскоязычных неструктурированных текстах в информационном пространстве при анализе многообразных алгоритмов классификации, был выбран мультиномиальный наивный байесовский классификатор (MNB). Следует отметить, что данный метод особенно эффективен для работы с текстовыми данными, где каждый документ представлен в виде набора слов.

В основе мультиномиального байесовского классификатора формула Байеса принимает следующий вид:

$$P(y|x_1,...,x_n)=\frac{P(y)P(x_1,...,x_n|y)}{P(x_1,...,x_n)}, \quad (1)$$

где y – класс документа; $x_1,\ldots,x_n$ – слова в документе; P(y) является вероятностью того, что случайно выбранный документ будет относиться к классу y без учета содержания документа; $P(x_1,...,x_n|y)$ – условная вероятность появления слов $x_1,\ldots,x_n$ в документе, который относится к классу y; $P(x_1,\ldots,x_n)$ – общая вероятность появления этих слов в любом документе.

Распределение вектора параметризуется следующим образом: $\theta_y = (\theta_y,\ldots,\theta_{yn})$ для каждого класса y, где n – размер словаря. Параметры θ_y учитываются с помощью сглаженной версии метода максимального правдоподобия:

$$\widehat{\theta_{yi}}=\frac{N_{yi}+\alpha}{N_y+\alpha n}, \quad (2)$$

где $N_{yi}=\sum x \in T$, T – тренировочный набор, N_{yi} является общим количеством всех функций класса y. Если $\alpha \geq 0$, тогда учитываются функции, которые отсутствуют в обучающих выборках, предотвращая затем нулевую вероятность в последующих вычислениях. Если $\alpha=1$, то параметр является сглаживанием Лапласа [15].

Корпус данных, реализуемый в программном комплексе, имеет следующий вид:

$$(x_i, y_i)_i^{l.} = 0, \quad (3)$$

где $x_i \in R^n$ – i-й комментарий пользователя, $y_i \in \{0,1,2,3\}$ является метками класса. Для каждого комментария устанавливает соответствующую метку функция F:

$$F(x_i) = y_i. \quad (4)$$

При разработке программной реализации были использованы следующие метрики качества: Accuracy – доля верных ответов модели среди всех предсказаний; Precision– доля истинно положительных ответов среди всех положительных ответов; Recall – доля истинно положительных ответов среди всех правильных ответов.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (5)$$

где *TP* является истинно положительным решением; *FP* – ложноположительное решение, *FN* представляет собой ложноотрицательное решение, *TN* – истинно отрицательное решение.

$$Precision = \frac{TP}{TP + FP}, \quad (6)$$

$$Recall = \frac{TP}{TP + FN}, \quad (7)$$

Помимо этого, используется метрика F-мера, являющаяся гармоничным средним между точностью и полнотой.

$$F_1 = \frac{2 * TP}{2 * TP + FP + FN}. \quad (8)$$

Для работы был взят корпус данных, состоящий из 120000 комментариев пользователей социальных сетей (*Одноклассники, Вконтакте, Пикабу*); мессенджеров (*Telegram, WhatsApp*). В целях более точной классификации было уменьшено количество комментариев, содержащих нейтральную и эксплицитную лексику, поскольку объем именно этих данных в несколько раз превышал объем других категорий, что приводило к неточностям в работе классификатора. Для машинного анализа были представлены данные в равных долях, содержащие различные типы языковой агрессии. Распределение комментариев представлено на рис. 1, где синим цветом представлена нейтральная лексика, оранжевым – имплицитная, серым – эксплицитная и желтым – манипулятивная.

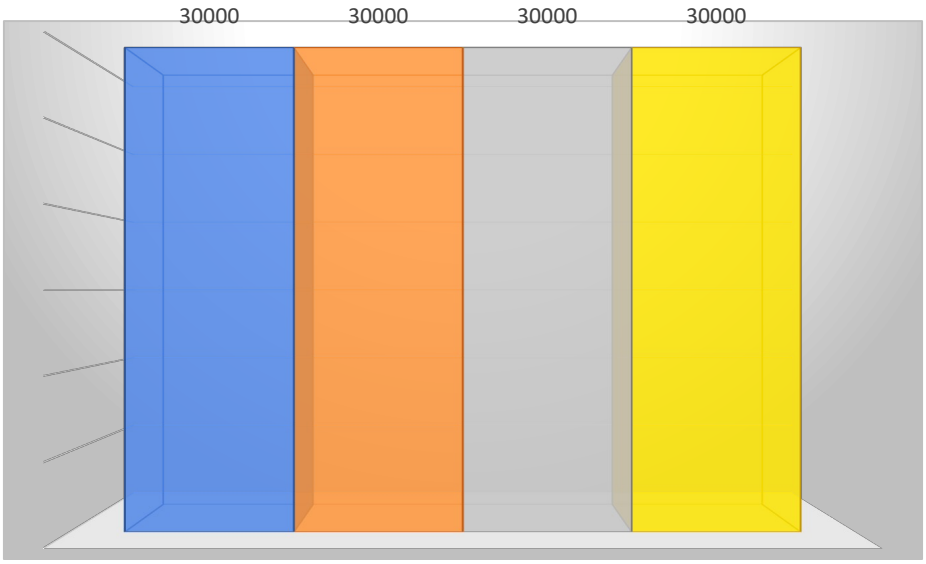

Рис.1. Распределение комментариев

При разработке программного комплекса для классификации речевой агрессии, помимо мультиномиального байесовского классификатора, был использован векторизатор TF-IDF, показавший более высокую точность.

Сравнительный анализ скоростей векторизаторов представлен в табл. 1.

Таблица 1. Сравнительный анализ точностей векторизаторов

Векторизатор	Точность
Word2Vec	0.260
TF-IDF	0.877
Bag of Words	0.314
HashVectorizer	0.864

На рис.2. представлена точность программной реализации, где наблюдается, что доля истинно положительных ответов (Recall) для всех четырех условий классификации довольно высока.

	0	1	2	3	accuracy	macro avg	weighted avg
precision	0.835172	0.961170	0.998367	0.706879	0.877333	0.875397	0.889734
recall	0.847067	0.851933	0.939874	0.865856	0.877333	0.876183	0.877333
f1-score	0.841077	0.903261	0.968238	0.778333	0.877333	0.872727	0.880751
support	5898.000000	6828.000000	6503.000000	4771.000000	0.877333	24000.000000	24000.000000

Рис.2. Точность программной реализации

На рис.3. представлена матрица ошибок, где 0 – эксплицитная лексика, 1 – имплицитная, 2 – манипулятивная и 3 – нейтральная.

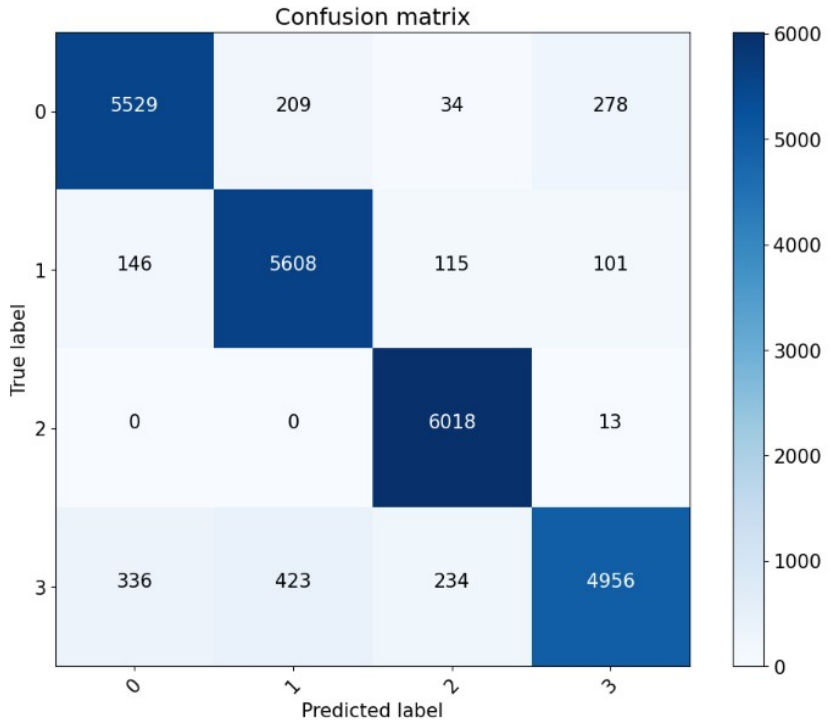

Рис.3. Матрица ошибок

3. Результаты исследования

Разработки алгоритма классификации языковой агрессии состоит из следующих этапов:

1. формирование корпуса данных;
2. ручная разметка данных (для каждой записи указывается «верный» ответ, т. е. тип речи, который является нейтральным или агрессивным видом лексики);
3. представление записей в виде векторов;
4. выбор необходимого векторизатора и дальнейшая классификация;
5. применение классификации.

Программная реализация была осуществлена с помощью Python и Java и представлена в виде плагина для Google Chrome.

В настоящее время данная программная реализация внедрена в качестве средств обучения студентов ФГБОУ ВО «ВГУ» по дисциплине «Искусственный интеллект».

Заключение

Данная программная реализация в перспективе может использоваться как администраторами социальных сетей, так и пользователями с целью выявления языковой агрессии, а также практикующими психологами при работе с людьми, обладающими низким эмоциональным интеллектом.

Программная реализация на корпусе данных из 120000 комментариев показала точность в 0.865.

В настоящий момент создан плагин для браузера Google Chrome, в будущем планируется создание расширений и на другие web-браузеры.

Литература

1. *Горбачева Е. Ю.* Агрессивная коннотация в текстах СМИ / Е.Ю. Горбачева, К.Э. Омельченко // Молодой ученый. – 2016. – № 28.1 (132.1). – С. 8–11.
2. *Денисова А. В.* Лексико-семантические способы выражения речевой агрессии в английском и русском языках / А.В. Денисова // Вестник Воронежского государственного университета. – 2021. – № 1. – С. 48–56.
3. *Щербинина Ю. В.* Русский язык. Речевая агрессия и пути её преодоления: учебное пособие/ Ю. В. Щербинина. – Москва: Флинта-Наука, 2018. – 225 с.
4. *Шейгал Е. И.* Семиотика политического дискурса: монография / Е. И.Шейгал. – Волгоград: Перемена, 2000. – 367 с.
5. *Buss A. H.* The Psychology of Aggression / A.H. Buss. – Michigan : Wiley, 1961. – 307 с.
6. *Выготский Л. С.* Проблемы развития психики / Л.С. Выготский. – Москва : Педагогика, 1982. – 368 с.
7. *Ениколопов С. Н.* Агрессивность и имплицитная теория насилия / С. Н. Ениколопов, Н. В. Чудова // Прикладная юридическая психология. – 2017. – № 2. – С. 18–26.
8. *Фромм Э.* Анатомия человеческой деструктивности / Э. Фромм. –Москва : Республика, 1994. – 261 с.
9. *Макарова Е. А.* Модели и алгоритмы обработки слабоструктурированных текстовых данных на основе методов искусственного интеллекта : специальность 2.3.1 Системный анализ, управление и обработка информации, статистика : диссертация на соискание ученой степени кандидата технических наук / Макарова Елена Андреевна. – Брянск, 2023. – 166 с. .
10. *Котельников Е. В.* Методология интеллектуального анализа мнений при обработке текстовой информации на основе правдоподобного вывода : специальность 05.13.17 Теоретические основы информатики : автореферат диссертации на соискание ученой

степени кандидата технических наук / Котельников Евгений Вячеславович. – Нижний Новгород, 2019. – 39 с. .
11. *Енина Л. В.* Современные российские лозунги как сверхтекст: специальность 10.02.01: автореферат диссертации на соискание кандидата филологических наук / Енина Лидия Владимировна. – Екатеренбург, 1999. – 18 с.
12. *Михальская А. К.* Основы риторики: Мысль и слово: учебное пособие для учащихся 10–11 классов / А.К. Михальская. – Москва : Просвещение, 1996. – 416 с.
13. *Паламарчук Н. А.* Способы выражения агрессии в текстах Интернет– комментариев / Н.А. Паламарчук // Актуальные проблемы теоретической и прикладной лингвистики. – 2011. С. 19–28.
14. *Щербинина Ю. В.* Русский язык. Речевая агрессия и пути её преодоления: учебное пособие/ Ю.В. Щербинина. – Москва: Флинта–Наука, 2018. – 225 с.
15. *Фромм Э.* Анатомия человеческой деструктивности / Э. Фромм. –Москва : Республика, 1994. – 261 с.
16. Scikit-learn [Электронный ресурс]. – Режим доступа: https://scikitlearn.org/stable/modules/naive_bayes.html#multinomial-naive-bayes (дата обращения:20.11.2024).
17. *Бобр А. Д.* Методы вербальной агрессии в комментариях пользователей социальных сетей : специальность 5.9.8. –Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки): автореферат диссертации на соискание ученой степени доктора филологических наук / Бобр Арина Дмитриевна. – Москва, 2024. – 21 с.
18. *Xu Shuo* Bayesian Naive Bayes classifiers to text classification / Shuo Xu // JOURNAL OF INFORMATION SCIENCE, 2018. – С. 48–59.
19. Abusive language detection from social media comments using conventional machine learning and deep learning approaches / M. P. Akhter, J. B. Zheng, I. R. Naqvi [и др.] // MULTIMEDIA SYSTEMS, 2022. – С. 1925–1940.
20. Aggression Detection in Social Media from Textual Data Using Deep Learning Models / U. Khan, S. Khan, A. Rizwan [и др.] // APPLIED SCIENCES-BASEL, 2022. – С. 1–16.
21. Supervised Classifiers to Identify Hate Speech on English and Spanish Tweets / S. Almatarneh, P. Gamallo, FJR Pena, A. Alexeev // DIGITAL LIBRARIES AT THE CROSSROADS OF DIGITAL INFORMATION FOR THE FUTURE: ICADL, 2019. – С. 23–30.
22. To BAN or Not to BAN: Bayesian Attention Networks for Reliable Hate Speech Detection / K. Miok, B. Skrlj, D. Zaharie, M.. Robnik-Sikonja // COGNITIVE COMPUTATION, 2022. – С. 353–371.
23. Пастревич М. К. Выбор подхода для решения задачи программной классификации вербальной агрессии в социальных сетях / И.Е. Воронина, М.К. Пастревич // Информационные системы и технологии. – 2023. – № 3. – С. 52–58.
24. Пастревич М. К. Анализ моделей для классификации неструктурированных текстов в информационном пространстве / И.Е. Воронина, М.К. Пастревич // Информационные системы и технологии. – 2024. – № 1(141). – С. 31–36. – ISSN 2072-8964

Воронина Ирина Евгеньевна

д.т.н., профессор кафедры программного обеспечения и администрирования информационных систем, Воронежский государственный университет (ФГБОУ ВО ВГУ, 394018, Россия, г. Воронеж, Университетская площадь, 1), +7 (473) 220-82-66, e-mail: irina.voronina@gmail.com.

Пастревич Марина Константиновна
преподаватель кафедры программного обеспечения и администрирования информационных систем, Воронежский государственный университет (ФГБОУ ВО ВГУ, 394018, Россия, г. Воронеж, Университетская площадь, 1), +7 (473) 220-82-66, e-mail: mirstat@mail.ru.

Авторы прочитали и одобрили окончательный вариант рукописи.
Авторы заявляют об отсутствии конфликта интересов.
Вклад соавторов: *Каждый автор внес равную долю участия как во все этапы проводимого теоретического исследования, так и при написании разделов данной статьи.*

Classifier of speech aggression in Russian-language unstructured texts

Irina E. Voronina, Marina K. Pastrevich

Voronezh State University (VSU)

Abstract: The problems of classification of linguistic aggression in unstructured texts of the information space are considered. Various approaches to defining the concept of "linguistic aggression" are analyzed, as well as methods for its detection and assessment. Particular attention is paid to issues of automatic text processing for detecting aggressive statements. Existing algorithms and machine learning models used to solve this problem are described. In addition, the prospects for the development of text analysis technologies and their application in the fight against the spread of verbal aggression on the Internet are discussed.

Keywords: classification, model, learning, language aggression, neural network.

For citation: Voronina I. E., Pastrevich M. K. Klassifikator rechevoj agressii v russkoyazychnyh nestrukturirovannyh tekstah [Classifier of speech aggression in Russian-language unstructured texts] // *Vestnik SIBGUTI,* 2025, vol. 19, no. 3, pp. 52–60. https://doi.org/10.55648/1998-6920-2025-19-2-52-60.

The article has been received by the editors 22.03.2025;
revised version – 03.05.2025;
accepted for publication 14.05.2025.

References

1. Gorbacheva E. Yu. Aggressive connotation in media texts / E.Yu. Gorbacheva, K.E. Omelchenko // Young scientist. - 2016. - No. 28.1 (132.1). - P. 8-11.
2. Denisova A. V. Lexical and semantic ways of expressing speech aggression in English and Russian / A. V. Denisova // Bulletin of the Voronezh State University. - 2021. - No. 1. - P. 48-56.
3. Shcherbinina Yu. V. Russian language. Speech aggression and ways to overcome it: a tutorial / Yu. V. Shcherbinina. - Moscow: Flinta-Nauka, 2018. - 225 p.
4. Sheigal E. I. Semiotics of political discourse: monograph / E.I.Sheigal. - Volgograd: Peremena, 2000. - 367 p.
5. Buss A. H. The Psychology of Aggression / A.H. Buss. – Michigan : Wiley, 1961. – 307 p.
6. Vygotsky, L.S. Problems of development of the psyche / L.S. Vygotsky. – Moscow : Peda-gogy, 1982. – 368 p.
7. Enikolopov S. N. Aggression and implicit theory of violence / S.N. Enikolopov, N.V. Chudova // Applied legal psychology. - 2017. - No. 2. - P. 18-26.

8. Fromm E. Anatomy of human destructiveness / E. Fromm. - Moscow: Respublika, 1994. - 261 p.
9. Makarova E. A. Models and algorithms for processing weakly structured text data based on artificial intelligence methods: specialty 2.3.1 Systems analysis, management and processing of information, statistics: dissertation for the degree of candidate of technical sciences / Makarova Elena Andreevna. - Bryansk, 2023. - 166 p. .
10. Kotelnikov E. V. Methodology of intelligent analysis of opinions in processing text information based on plausible inference: specialty 05.13.17 Theoretical foundations of computer science: abstract of a dissertation for the degree of candidate of technical sciences / Kotelnikov Evgeny Vyacheslavovich. – Nizhny Novgorod, 2019. – 39 p. .
11. Enina L. V. Modern Russian slogans as a supertext: specialty 10.02.01: abstract of a dissertation for a candidate of philological sciences / Enina Lidiya Vladimirovna. – Ekaterinburg, 1999. – 18 p.
12. Mikhalskaya A. K. Fundamentals of rhetoric: Thought and word: a teaching aid for students in grades 10-11 / A.K. Mikhalskaya. – Moscow: Prosveshchenie, 1996. – 416 p.
13. Palamarchuk, N.A. Ways of expressing aggression in the texts of Internet comments / N.A. Palamarchuk // Actual problems of theoretical and applied linguistics. – 2011. – No. -. – P. 19-28.
14. Shcherbinina Yu.V. Russian language. Speech aggression and ways to overcome it: a tutorial / Yu.V. Shcherbinina. - Moscow: Flinta-Nauka, 2018. - 225 p.
15. Fromm, E. Anatomy of human destructiveness / E. Fromm. – Moscow: Respublika, 1994. – 261 p.
16. Scikit-learn [Electronic resource]. – Access mode: https://scikitlearn.org/stable/modules/naive_bayes.html#multinomial-naive-bayes (accessed: 20.11.2024).
17. Bobr A. D. Methods of verbal aggression in comments of social network users: specialty 5.9.8. - Theoretical, applied and comparative linguistics (philological sciences): abstract of a dissertation for the degree of Doctor of Philological Sciences / Bobr Arina Dmitrievna. - Moscow, 2024. - 21 p.
18. Xu Shuo Bayesian Naive Bayes classifiers to text classification / Shuo Xu // JOURNAL OF INFORMATION SCIENCE, 2018. – С. 48-59.
19. Abusive language detection from social media comments using conventional machine learning and deep learning approaches / M.P. Akhter, J.B. Zheng, I.R. Naqvi [и др.] // MULTIMEDIA SYSTEMS, 2022. – С. 1925-1940.
20. Aggression Detection in Social Media from Textual Data Using Deep Learning Models / U. Khan, S. Khan, A. Rizwan [и др.] // APPLIED SCIENCES-BASEL, 2022. – С. 1-16.
21. Supervised Classifiers to Identify Hate Speech on English and Spanish Tweets / S. Almatarneh, P. Gamallo, FJR Pena, A. Alexeev // DIGITAL LIBRARIES AT THE CROSSROADS OF DIGITAL INFORMATION FOR THE FUTURE: ICADL, 2019. – С. 23-30.
22. To BAN or Not to BAN: Bayesian Attention Networks for Reliable Hate Speech Detection / K. Miok, B. Skrlj, D. Zaharie, M.. Robnik-Sikonja // COGNITIVE COMPUTATION, 2022. – С. 353-371.
23. Pastrevich, M. K. Selecting an approach to solving the problem of software classification of verbal aggression in social networks / I.E. Voronina, M.K. Pastrevich // Information systems and technologies. - 2023. - No. 3. - P. 52-58.
24. Pastrevich, M. K. Analysis of models for classifying unstructured texts in the information space / I.E. Voronina, M.K. Pastrevich // Information systems and technologies. - 2024. - No. 1 (141). - P. 31-36. - ISSN 2072-8964

Irina E. Voronina

Doctor of Engineering Sciences, Professor of the Department of Software and Information Systems Administration, Voronezh State University (VSU, 394018, Russia, Voronezh, University Square, 1), +7 (473) 220-82-66, e-mail: irina.voronina@gmail.com.

Marina K. Pastrevich

Lecturer, Department of Software and Information Systems Administration, Voronezh State University (VSU, 394018, Russia, Voronezh, University Square, 1), +7 (473) 220-82-66, e-mail: mirstat@mail.ru.