

УДК 004.89

DOI: [10.26102/2310-6018/2024.44.1.001](https://doi.org/10.26102/2310-6018/2024.44.1.001)

## Методы отбора признаков в задаче определения авторства в контексте кибербезопасности

А.С. Романов✉

*Томский государственный университет систем управления и радиоэлектроники,  
Томск, Российская Федерация*

**Резюме.** В работе рассмотрены методы определения авторства естественных и искусственно-сгенерированных текстов, важных в контексте кибербезопасности и защиты интеллектуальной собственности с целью предотвращения дезинформации и мошенничества. Использование методов определения автора текста обосновано выводами об эффективности рассмотренных в прошлых исследованиях fastText и метода опорных векторов (SVM). Алгоритм отбора признаков выбран на основе сравнения пяти различных методов – генетического алгоритма, прямого и обратного последовательных методов, регуляризационного отбора и метода Шепли. Рассмотренные алгоритмы отбора включают эвристические методы, элементы теории игр и итерационные алгоритмы. Наиболее эффективным методом признан алгоритм, основанный на регуляризации, в то время как методы, основанные на полном переборе, признаны неэффективными для любого множества авторов. Точность отбора на основе регуляризации и SVM в среднем составила 77 %, что превосходит другие методы от 3 до 10 % при идентичном количестве признаков. При тех же задачах средняя точность fastText – 84 %. Было проведено исследование, направленное на устойчивость разработанного подхода к генеративным образцам. SVM оказался более устойчив к запутыванию модели. Максимальная потеря точности для fastText составила 16 %, а для SVM – 12 %.

**Ключевые слова:** отбор признаков, определение автора, машинное обучение, нейронные сети, анализ текста, информационная безопасность.

**Благодарности:** данная работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках базовой части государственного задания ТУСУРа на 2023–2025 гг. (проект № FEWM-2023-0015).

**Для цитирования:** Романов А.С. Методы отбора признаков в задаче определения авторства в контексте кибербезопасности. *Моделирование, оптимизация и информационные технологии*. 2024;12(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1489> DOI: 10.26102/2310-6018/2024.44.1.001

## Feature selection methods for authorship attribution in cybersecurity context

A.S. Romanov✉

*Tomsk State University of Control Systems and Radioelectronics, Tomsk,  
the Russian Federation*

**Abstract.** This paper considers methods for authorship attribution of natural-language and artificially generated texts, which are important in the context of cybersecurity and intellectual property protection to prevent misinformation and fraud. The use of authorship methods is justified by the findings on the fastText and support vector method (SVM) effectiveness discussed in past studies. The feature selection algorithm is chosen based on the comparison of five different methods: genetic algorithm, forward and backward sequential methods, regularization selection and Shapley's method. The considered selection algorithms include heuristic methods, game theory elements and iterative algorithms. The regularisation-based algorithm is found to be the most efficient method, while methods based on

complete brute-force selection are found to be inefficient for any set of authors. The regularization-based and SVM-based selection accuracy averaged 77 %, outperforming the other methods by between 3 and 10 % for an identical number of features. For the same tasks, the average accuracy of fastText is 84 %. A study was conducted to examine the robustness of the developed approach to generative samples. SVM proved to be more robust to model confounding. The maximum loss of accuracy for fastText was 16 % and for SVM was 12 %.

**Keywords:** feature selection, authorship attribution, machine learning, neural networks, text analysis, information security.

**Acknowledgements:** this research was supported by the Ministry of Science and Higher Education of the Russian Federation the basic part of the state assignment of TUSUR for 2023–2025 (project No. FEWM-2023-0015).

**For citation:** Romanov A.S. Feature selection methods for authorship attribution in cybersecurity context. *Modeling, Optimization and Information Technology*. 2024;12(1). URL: <https://moitvivr.ru/ru/journal/pdf?id=1489> DOI: 10.26102/2310-6018/2024.44.1.001 (In Russ.).

## Введение

В эпоху цифровизации определение авторства текста становится важным и для кибербезопасности. В мире, где тексты и информация мгновенно распространяются через Интернет, возможность точно установить автора становится критически важной для защиты интеллектуальной собственности и предотвращения плагиата. Кроме того, это позволяет противостоять распространению дезинформации и фейковых новостей, выявлять недобросовестные источники, что особенно актуально в условиях информационных войн. В корпоративном секторе такая способность помогает обнаруживать утечки конфиденциальной информации и предотвращать киберпреступления. В социальных сетях продленная аутентификация на основе анализа текста улучшает безопасность, сравнивая стиль новых публикаций пользователя с его старыми текстами. Если обнаруживаются значительные отличия, система может заподозрить взлом и запросить дополнительные подтверждения личности. Это повышает защиту от несанкционированного доступа и киберугроз.

Цель работы состоит в исследовании методов отбора признаков текста для улучшения точности в сложных случаях классификации, когда количество потенциальных авторов-кандидатов составляет 20 и более классов.

Новизна исследования заключается в разработке методов определения автора русскоязычного текста на основе комбинации fastText и SVM с отбором информативных признаков, а также усложнением задачи классификации классических литературных текстов путем проведения атаки на метод запутывающими образцами – искусственно-сгенерированными текстами.

Ранее [1, 2] авторами было установлено, что комбинация метода опорных векторов (SVM) и генетического алгоритма (ГА) с сокращением признакового пространства с 1168 до 500 элементов является наиболее результативным вариантом среди классических методов машинного обучения. При проведении экспериментов с нейронными сетями (НС) хорошие результаты показал fastText, при количестве классов более 20 уступающий в точности глубоким НС до 5 %, но выигрывающий во времени обучения. Однако вопрос об отборе признаков нельзя считать решенным, поскольку с использованием ГА хоть и удалось добиться повышения точности классификации, но не проводилось сравнение с другими методами отбора, чему уделено внимание в настоящей работе.

Авторами статьи [3] предложена Multi-Channel Self-Attention Network (MCSAN). Эксперименты проводились на наборах данных CCAT10, CCAT50 и IMDb62,

содержащих короткие тексты на испанском 10 и 50 авторов, в случае первых двух наборов и отзывы к фильмам, написанные 62 пользователями на английском. в случае третьего. Для сравнения предложенного подхода с другими методами применялись сверточная НС (CNN) и двунаправленная сеть долгой краткосрочной памяти (BiLSTM) как по отдельности, так и в сочетании с MCSAN. Результаты экспериментов показывают, что извлеченные признаки эффективны – их применение улучшает результаты классификации CCAT10 и CCAT50 на 2 % и 3 %, соответственно в сравнении с использованием BiLSTM и CNN без MCSAN.

Касательно русскоязычных текстов, научное сообщество не пришло к консенсусу в вопросе выбора наиболее информативного признакового пространства [4]. Под понятием методов извлечения признаков понимают выделение зависимостей на разных уровнях текста – предложениях, абзацах, главах и текстах целиком; иерархические зависимости в виде графов; семантический анализ; дистрибутивную семантику. Последние представляют наибольший интерес ввиду того, что эти методы агрегировано представляют все остальные.

Ряд исследований направлен на классификацию авторов научных публикаций [5]. В работе [6] представлена система, позволяющая разрешать споры об авторском праве русскоязычных научных статей. При формировании набора данных использовалась база [link.springer.com](http://link.springer.com), а данные научной электронной библиотеки eLIBRARY.ru применялись для получения достоверной информации об авторах. Данные сопоставляются по ключам: ФИО автора, название и дата публикации, соавторы, аффилиации. В случае отсутствия части параметров в одном из источников или их несовпадения применялся алгоритм кластеризации  $k$ -средних для объединения статей в случае превышения заданного порога схожести. Добавление в систему модуля кластеризации статей, не распознанных на этапе сравнения с публикациями из eLIBRARY, позволило улучшить результат идентификации авторства статей до 92 %.

Исследование частот позволяет решать множество задач, связанных с атрибуцией текста, но возникает вопрос в определении множества наиболее информативных и исключении избыточных. Существует как минимум три причины отрицательного влияния избыточных признаков на качество работы классификатора [7]. Во-первых, при увеличении числа признаков, уменьшается статистическая надежность работы алгоритма. Во-вторых, чем больше признаков, тем больше необходимо объектов обучения для надежной классификации. И, в-третьих, чем больше признаков, тем больше время работы алгоритма [8]. Задача отбора состоит в том, чтобы найти оптимальное подмножество признаков, при котором увеличится точность классификации. Такое подмножество уменьшает временные и вычислительные затраты и повышает эффективность алгоритма.

Для отбора признаков используются следующие группы методов:

1. Фильтр [9]. Данные методы основаны на выявлении обобщающих свойств данных и не ориентируются на конкретные алгоритмы интеллектуального анализа. Методы не затратны с точки зрения вычислений, благодаря этому чаще применяются к наборам данных с большим количеством признаков. Однако есть и существенный минус: методы рассматривают каждый признак изолированно.

2. Обертка [10]. Особенность этой группы методов – поиск всех возможных подмножеств признаков и оценка их качества путем прогонки через модель. Метод на основе обертки включается в процесс построения классификатора. За счет этого существенно увеличивается сложность алгоритма и вычислительные затраты.

3. Встроенные. Отбор признаков происходит во время обучения модели, поэтому такие методы зависят от параметров классификатора. Примером может быть отбор, основанный на регуляризации.

На каждом из этапов могут возникнуть неточности, например, из-за зашумленности анализируемого текста, особенностей языка, которые далее приведут к серьезным ошибкам при интерпретации результатов. Кроме того, чем крупнее единицы текста, тем больший объем может потребоваться для того, чтобы статистические показатели стабилизировались. Характеристики уровня символов и слов позволяют моделировать сложные связи, что делает их более перспективными, чем элементы высокого уровня. Однако для определения информативных групп признаков необходимо проведение исследований на всех потенциально возможных признаках.

### Математическая постановка задачи

Математическая постановка задачи включает три конечных множества: тексты с известным авторством  $\mathbf{T}$ , анонимные тексты  $\mathbf{T}'$ , множество авторов  $\mathbf{A}$ . Также определен конечный набор признаков  $\mathbf{X} = \{x_1, x_2, \dots, x_p\}$ . Определена функция  $q: \mathbf{T} \cup \mathbf{T}' \rightarrow \mathbf{X}_i$ , которая отображает каждый текст в вектор признаков. Пусть исходная модель использует  $N$  признаков текста, в таком случае полное признаковое пространство классификационных данных  $\mathbf{F} = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots, \mathbf{X}_n\}$ . Пусть  $\mathbf{X}_i$  содержит  $N=1068$  признаков текста. При решении практических задач не все  $N$  признаки полезны, поэтому мощность конечного множества информативных признаков  $|\mathbf{F}'| = M$ , где  $N > M$ , а подмножество  $\mathbf{L} = N / M = \{l \mid l \in N, l \notin M\}$  является множеством избыточных (неинформативных элементов). Согласно методу отбора определяется целевая функция  $f: R^p \times \mathbf{A} \rightarrow [0, 1]$ , которая возвращает вероятность того, что данный текст написан по данным на вход автором. Для каждого анонимного текста  $t' \in \mathbf{T}'$  вычисляется  $w(q(t'), a_i)$  для каждого  $a_i \in \mathbf{A}$ . Автор с максимальным значением рекомендуется как предполагаемый создатель текста.

### Набор данных

В качестве набора классических литературных текстов использовался ранее описанный датасет [2], включающий 100 авторов и 1100 художественных текстов. Подробные характеристики приведены в работе [2]. В текущем исследовании в дополнение представлен набор сгенерированных с помощью RuGPT-2<sup>1</sup> и RuGPT-3<sup>2</sup> текстов для решения задачи разграничения авторства между профессиональным писателем и генеративной моделью [11, 12] Генерация происходила для выборки из 10 авторов с наибольшей разделяющей способностью. Статистика датасета представлена в Таблице 1.

Таблица 1 – Информация о наборе данных  
Table 1 – Information about the dataset

Характеристика	Художественные тексты	RuGPT-2	RuGPT-3
Число авторов	100	10	10
Число текстов	1000	40	40
Слов в датасете	36911945	2076477	1856746
Ср. длина текста, симв.	836498	631024	619257

<sup>1</sup> <https://github.com/vlarine/ruGPT2>

<sup>2</sup> <https://github.com/ai-forever/ru-gpts>

## Методы отбора информативных признаков

По результатам прошлого исследования особая эффективность SVM была достигнута при совместном использовании с ГА. Исходное множество признаков составляло 1168 элементов. ГА позволил уменьшить размерность признакового пространства и в то же время получить выигрыш в точности. В прошлой работе не рассматривались другие методы отбора, помимо ГА, что стало задачей текущей статьи. Применяемые в текущей работе методы описаны ниже:

1. ГА. Рассмотренный ранее ГА [2] является эвристическим встроенным методом. В текущей статье будет использоваться в качестве базового уровня (baseline). Подробное описание метода и полученные для 2, 5, 10, 20 и 50 авторов результаты представлены в прошлой работе [2].

2. Алгоритм отбора на основе регуляризации (РО) (англ. Regularization Selection (RS)) [13] является эвристическим встроенным методом, основанным на присвоении меры информативности каждому признаку. На вход метода подается линейная модель и устанавливается пороговое значение весов для отбора. Признак признается информативным, если соответствующая мера не ниже заданного порога или соответствует выбранному диапазону значения, например, превышает среднее или медиану. Ключевым для данного метода является параметр регуляризации, поэтому встроенный метод подходит для линейных классификаторов. Процесс регуляризации заключается в добавлении штрафа к параметрам модели. Штраф позволяет избежать настройку модели под решение конкретно заданной задачи, а следовательно, ее неприменимость для других случаев. При  $L1$ -регуляризации штраф применяется к коэффициентам, уменьшающим каждый из предикторов. Было установлено, что в случае SVM параметр  $C$  управляет разреженностью: при меньшем значении параметра отбирается меньшее количество признаков и наоборот.

3. Прямой последовательный отбор (ППО) (англ. Forward Selection (FS)). ППО [14, 15] является фильтрационным итеративным методом и наиболее простым способом уменьшения размерности признакового пространства. В данном методе случайным образом выбирается первый признак, к которому затем добавляется второй на основе лучшей производительности (максимальной точности), затем процесс повторяется до достижения заданного количества признаков. Недостаток метода состоит в том, что при подборе второго и последующих признаков происходит полный перебор, что негативно сказывается на времени отбора, а также может стать причиной переобучения модели – отобранные признаки могут иметь сильную корреляцию, но при этом нести близкую по смыслу информацию.

4. Обратный последовательный отбор (ОПО) (англ. Backward Selection (BS)). ОПО [16] противоположен прямому. В первой итерации алгоритм работает на основе всех признаков. На каждой итерации исключается признак, не несущий ценности для модели. В этом методе тоже присутствует неопределенность при удалении сильно коррелированных признаков.

5. Shapley Additive exPlanations (SHAP) [17] – встроенный метод, который применяется для выделения наиболее информативных признаков, на основе которых в дальнейшем обучался SVM. Этот метод работает так: он использует идеи из теории игр, где разные признаки текста представлены как игроки, работающие вместе. Каждый игрок (признак) вносит свой вклад в общий результат. Метод вычисляет, какой вклад каждый игрок внес в итоговый результат, используя специальный подход, называемый значением Шепли. SHAP – это особая версия этого подхода, которая объясняет вклад каждого признака в общую сумму, как в простой математической модели. Формально это отражено в следующем виде (1):

$$g(z') = \phi_0 + \sum_{j=1}^{\|\mathbf{M}\|} \phi_j z_j, \quad (1)$$

где  $g$  – используемая линейная модель,  $z' \in [0;1]$ ,  $\|\mathbf{M}\|$  – длина вектора коалиции (объединения игроков), в котором запись 1 означает, что соответствующее значение признака присутствует, а 0 – отсутствует,  $\phi_j \in R$  – значения Шепли, выступающие как атрибутивная функция для  $j$ -ого признака,  $\phi_0$  – значение Шепли до включения  $j$ -ого признака. Условие об использовании линейной модели обусловлено возможностью вычисления  $\phi$ . Для случая, в котором коалиционный вектор состоит только из единиц, формула упрощается до следующей (2):

$$g(x') = \phi_0 + \sum_{j=1}^{\|\mathbf{M}\|} \phi_j, \quad (2)$$

то есть считается средний прирост эффективности от добавления  $j$ -го признака в коалицию. Отличием SHAP является то, что он может работать с обученной моделью и определять информативность признаков на основе оценки тестовых экземпляров, в то время как другие методы определяют информативность с учетом процесса обучения.

### Результаты отбора признаков

При проведении отбора признаков рассматривались 50, 100, 200, 400 и 500 признаков, также, как и в ранее проведенном эксперименте с ГА [2]. Результаты представлены в Таблице 2. На Рисунке 1 изображены пять наиболее информативных признаков с отражением влияния признака на классификацию по методам ППО, ОПО, на основе регуляризации и ГА на основе признаков с наибольшим значением SHAP.

Таблица 2 – Результаты отбора признаков  
Table 2 – Results of feature selection

Число авторов	Метод	Число признаков				
		50	100	200	400	500
		Точность, %				
2	ГА	95,9±3	98,3±4	96,2±2	98,6±3	96,1±2
	ППО	90,0±5	92,0±4	93,2±4	89,3±3	86,4±3
	ОПО	89,9±3	92,4±5	91,2±3	88,6±3	91,1±3
	РО	96,1±3	95,9±4	96,6±4	98,9±2	96,0±2
	SHAP	94,5±2	95,2±5	93,6±4	92,3±2	93,9±3
5	ГА	94,8±2	92,4±2	95,0±4	95,5±3	94,7±3
	ППО	80,1±4	82,8±2	91,0±4	94,1±3	89,7±2
	ОПО	82,4±4	82,4±5	90,2±2	90,2±4	92,8±3
	РО	92,7±3	93,3±3	96,3±3	96,6±4	95,1±3
	SHAP	93,0±3	93,0±5	90,4±2	89,7±4	89,7±4
10	ГА	80,3±3	82,9±4	82,1±3	81,0±3	83,7±4
	ППО	76,7±4	77,4±5	75,4±4	74,7±4	71,5±3
	ОПО	65,2±3	69,3±4	71,0±3	75,7±4	72,6±5
	РО	83,4±4	84,2±3	86,2±3	88,5±4	82,7±4
	SHAP	74,5±4	78,2±4	75,1±5	73,1±3	75,7±4

Таблица 2 (продолжение)  
Table 2 (extended)

Число авторов	Метод	Число признаков				
		50	100	200	400	500
		Точность, %				
20	ГА	60,7±3	63,2±3	72,9±3	73,7±3	69,1±3
	ППО	59,2±5	56,1±5	64,2±4	65,9±4	64,2±2
	ОПО	55,4±3	60,3±5	66,5±4	67,1±2	66,4±5
	РО	64,6±5	66,6±3	74,1±4	78,9±4	72,1±3
	SHAP	55,2±4	62,9±4	65,6±5	60,9±4	62,5±2
50	ГА	33,8±2	38,1±4	42,4±4	44,4±3	40,2±2
	ППО	12,2±6	20,8±4	30,2±5	36,3±5	37,4±2
	ОПО	19,2±4	22,9±5	24,1±4	34,9±5	36,2±5
	РО	35,1±4	36,4±4	43,2±5	47,3±6	45,2±2
	SHAP	21,3±5	22,1±4	25,1±4	29,1±5	33,8±4

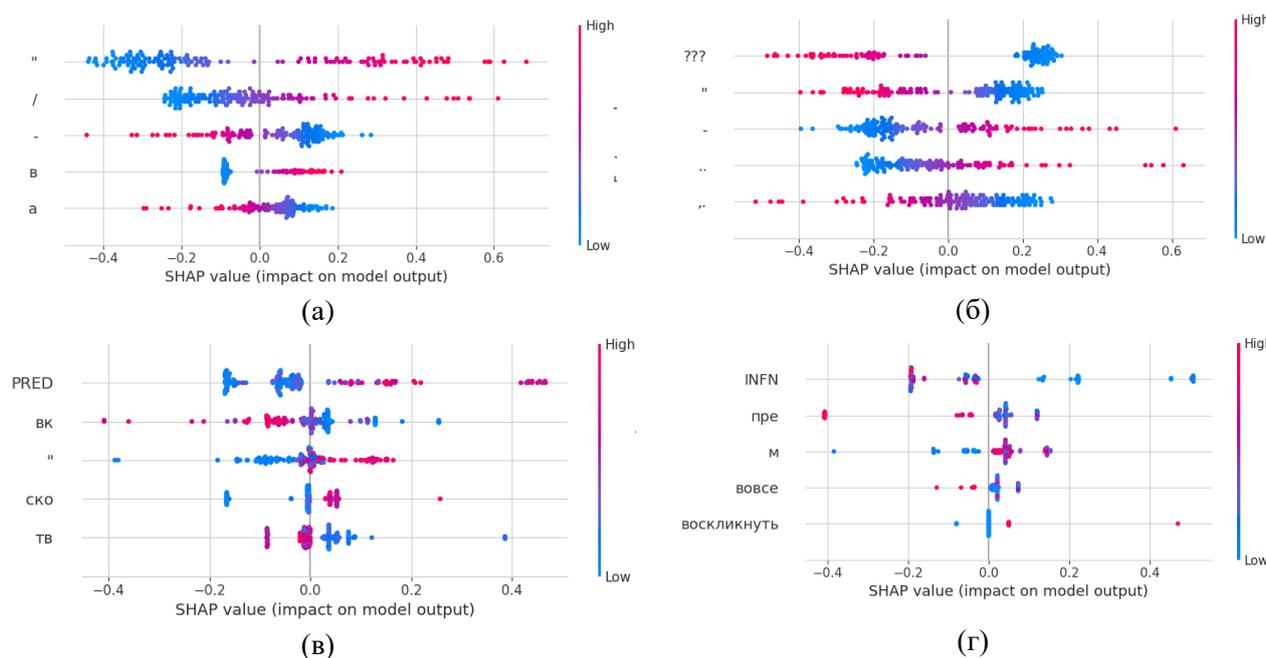


Рисунок 1 – Наиболее информативные признаки (а) ГА, (б) РО, (в) ППО, (г) ОПО  
Figure 1 – The most informative features (a) GA, (b) RS, (c) FS, (d) BS

Рисунки демонстрируют информативность признаков, в данном ключе рассматриваемую как прирост точности при их использовании. По оси ординат указаны наименования признаков, по оси абсцисс – влияние на результат модели. Каждый классификационный объект (текст) представлен одной точкой для каждого признака. Скопление точек отражает плотность, то есть множество объектов, имеющих близкие значения SHAP. Красный цвет точки указывает на высокую информативность, а синий – на низкую. Например, в случае метода ОПО отрицательные значения SHAP признака «вовсе» информативны, а положительные, напротив, избыточны. Вертикально удлинненные точки на графике представляют собой скопления точек данных, которые распределены вдоль оси X, чтобы избежать перекрытия. Такой метод визуализации используется для того, чтобы сделать каждую точку данных отдельной и видимой.

Результаты позволяют установить неэффективность методов прямого и обратного отбора признаков, поскольку точность при таких методах не превосходит исходные

результаты ГА, причем в случае 20 и 50 классов ППО и ОПО уступают ГА на 10 % и более. В качестве причин низкой точности можно выделить полный перебор, лежащий в основе методов. Признаки, выделенные как информативные методом SHAP, при совместном использовании с SVM демонстрируют конкурентные результаты только при малом количестве признаков – 50, 100 и классов – 2, 5, 10. Особенность метода – продемонстрировать хорошие результаты при низкой размерности признакового пространства обоснована тем, что в случае расширения множества признаков, с каждым новым признаком информативность нового элемента ниже, чем прошлого. Основанный на регуляризации метод РО позволил достигнуть поставленных целей и добиться улучшения точности для 20 и 50 авторов до 5 %. В качестве обоснования можно отметить ограничение данного метода – работу с линейными моделями, поэтому SVM использовался только с линейным ядром. Также ограничение состоит во влиянии параметра регуляризации, как следствие, можно использовать только ограниченный список моделей классификации. Ограничением ППО и ОПО является фактор случайности при выборе первого признака, на основе которого формируется дальнейшее множество. При неудачном выборе первого элемента последующие могут быть отобраны на основе настройки под конкретные тестовые данные, используемые при выполнении отбора. К тому же такие признаки могут быть мало информативными на последующих итерациях работы классификатора.

Исходя из представленных результатов, уменьшение количества признаков более, чем вдвое (400) не только не снижает точность классификации, но и позволяет улучшить результат определения автора текста на обоих наборах данных. Набор 400 признаков, на котором была достигнута максимальная точность, содержит частотные распределения шести знаков пунктуации, восьми частей речи, 20 униграмм символов, 107 биграмм символов, 98 триграмм символов и 165 слов из частотного словаря [18].

### Результаты экспериментов по определению авторства художественных и искусственно-сгенерированных текстов

В Таблице 3 приведены результаты эксперимента, посвященного классификации художественных и сгенерированных моделями текстов и сравнении fastText и SVM с отбором признаков в решении задачи. Эксперимент был проведен в трех вариациях: стандартный процесс классификации, усложнение задачи путем внедрения искусственных текстов, сгенерированных RuGPT-2 и RuGPT-3 на основе оригинального датасета.

Таблица 3 – Использование метода отбора на основе регуляризации и fastText для сгенерированных текстов

Table 3 – Results of feature selection using RS and fastText for generated texts

Число авторов	Метод	Исходные тексты	Исходные тексты + RuGPT-2	Исходные тексты + RuGPT-3
		Точность, %		
2	SVM с отбором признаков	98,9±2	96,5±4	95,7±4
	fastText	98,2±5	95,4±5	95,0±4

Таблица 3 (продолжение)  
Table 3 (extended)

Число авторов	Метод	Исходные тексты	Исходные тексты + RuGPT-2	Исходные тексты + RuGPT-3
5	SVM с отбором признаков	98,6±4	94,4±3	90,2±4
	fastText	95,0±4	90,3±5	90,4±5
10	SVM с отбором признаков	92,5±4	89,2±4	86,7±4
	fastText	92,2±6	87,2±4	85,3±4
20	SVM с отбором признаков	78,9±4	71,8±4	70,0±3
	fastText	69,9±4	63,2±4	62,5±4
50	SVM с отбором признаков	47,3±6	35,6±5	35,1±4
	fastText	56,8±6	48,4±5	40,9±5

Использование искусственных текстов усложняет процесс определения автора текста. Благодаря тщательному формированию признакового пространства и установлению метода отбора признаков, SVM оказался более устойчив к запутыванию модели. Максимальная потеря точности для fastText составила 16 %, а для SVM – 12 %. Наибольшая потеря точности зафиксирована при проведении экспериментов для 20 и 50 классов, в то время как для 2 и 5 классов, несмотря на осложнение задачи, потери не превышают 3 % от результата, полученного без использования сгенерированных образцов. Средняя точность классификации оригинальных образцов выше в случае fastText, однако SVM с отбором признаков превосходит fastText в случае атаки на метод посредством внедрения сгенерированных образцов, что свидетельствует об эффективности подобранного признакового пространства.

### Заключение

При выборе метода идентификации автора текста исследователи должны учитывать объем текстов. Это важно для точного определения авторства, что может иметь значение при расследовании случаев киберпреступлений или утечек информации. Обобщая результаты, были сформулированы следующие рекомендации:

1. В обучающий набор данных должны входить только тексты, точно принадлежащие исследуемым авторам.
2. Для художественных текстов необходимо использовать не менее трех текстов длиной 15000. Увеличение объема текстов повышает точность идентификации, что критично для выявления источников дезинформации или мошенничества.
3. В условиях ограниченных ресурсов следует применять классические методы машинного обучения с отбором информативных признаков. В случаях, когда есть риск преднамеренного искажения текста (например, в кибератаках), предпочтительнее использовать НС из-за их способности автоматического выявления признаков.
4. И последнее: чем меньше число авторов-кандидатов, тем выше точность методов идентификации. Различающая способность составляет 85 % и более для 2, 5 и 10 авторов в случае наборов литературных и сгенерированных текстов.

В будущих исследованиях планируется использование доверительных метрик и калибровочных кривых, что позволит снизить требования к объему данных и использовать даже небольшие наборы данных без ущерба для точности.

## СПИСОК ИСТОЧНИКОВ

1. Romanov A., Kurtukova A., Shelupanov A., Fedotova A., Goncharov V. Authorship identification of a Russian-language text using support vector machine and deep neural networks. *Future Internet*. 2020;13(1):3. DOI: 10.3390/fi13010003.
2. Fedotova A., Romanov A., Kurtukova A., Shelupanov A. Authorship attribution of social media and literary Russian-language texts using machine learning methods and feature selection. *Future Internet*. 2021;14(1):4. DOI: 10.3390/fi14010004.
3. Wu H., Zhang Z., Wu Q. Exploring syntactic and semantic features for authorship attribution. *Applied Soft Computing*. 2021;111:107815–107822. DOI: 10.1016/j.asoc.2021.107815.
4. Khomytska I., Bazylevych I., Teslyuk V. The statistical parameters of Ivan Franko's authorial style determined by the chi-square test. *2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), 2022*. p. 73–76. DOI: 10.1109/CSIT56902.2022.10000491.
5. Chekhovich Y. V., Khazov A. V. Analysis of duplicated publications in Russian journals. *Journal of informetrics*. 2022;16(1):101246. DOI: 10.1016/j.joi.2021.101246.
6. Исаченко В.В., Апанович З.В. Система анализа и визуализации для кросс-языковой идентификации авторов научных публикаций. *Вестник Новосибирского государственного университета. Серия: Информационные технологии*. 2018;16(2):49–61. DOI: 10.25205/1818-7900-2018-16-2-49-61.
7. Agun H.V., Yilmazel O. Incorporating topic information in a global feature selection schema for authorship attribution. *IEEE Access*. 2019;7:98522–98529 DOI: 10.1109/ACCESS.2019.2930536.
8. Kou G., Yang P., Peng Y., Xiao F., Chen Y., Alsaadi F.E. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*. 2020;86:105836. DOI: 10.1016/j.asoc.2019.105836.
9. Bardamova M., Hodashinsky I. Hybrid algorithm for tuning feature weights in a fuzzy classifier. *2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT), Yekaterinburg, Russia, 2021*. p. 0354–0357. DOI: 10.1109/USBREIT51232.2021.9455030.
10. Yaseen A., Laftah W., Kadhum I., Hamad A. Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system. *Pattern Recognition*. 2022;108912. DOI: 10.1016/j.patcog.2022.108912.
11. Uchendu A., Le T., Lee D. Attribution and obfuscation of neural text authorship: A data mining perspective. *ACM SIGKDD Explorations Newsletter*. 2023;25(1):1–18. DOI: 10.48550/arXiv.2210.10488.
12. Shamardina T. et al. Findings of the the ruatd shared task 2022 on artificial text detection in Russian. *arXiv preprint arXiv*. 2206;2022;01583. DOI: 10.48550/arXiv.2206.01583.
13. Xu W., Yuan, K., Li, W., Ding, W. An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution. *IEEE Transactions on Emerging Topics in Computational Intelligence*. 2022;7(1):76–88. DOI: 10.1109/TETCI.2022.3171784.
14. Yao G., Xiaojian H., Guanxiong W. A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain. *Expert Systems with Applications*. 2022;117002. DOI: 10.1016/j.eswa.2022.117002.
15. Abu Khurma R., Aljarah I., Sharieh A., Abd Elaziz M., Damaševičius R., Krilavičius T. A review of the modification strategies of the nature inspired algorithms for feature selection problem. *Mathematics* 2022;10(464). DOI: 10.3390/math10030464.

16. Borboudakis G., Tsamardinos I. Forward-backward selection with early dropping. *The Journal of Machine Learning Research*; 2019;20(1):276–314. DOI: 10.5555/3322706.3322714.
17. Le N.Q.K., Ho Q.T., Nguyen V.N., Chang J.S. BERT-Promoter: An improved sequence-based predictor of DNA promoter using BERT pre-trained model and SHAP feature selection. *Computational Biology and Chemistry*. 2022;99:107732. DOI: 10.1016/j.compbiolchem.2022.107732.
18. Новый частотный словарь русской лексики. URL: <http://dict.ruslang.ru/freq.php> (дата обращения 04.12.2023).

## REFERENCES

1. Romanov A., Kurtukova A., Shelupanov A., Fedotova A., Goncharov V. Authorship identification of a Russian-language text using support vector machine and deep neural networks. *Future Internet*. 2020;13(1):3. DOI: 10.3390/fi13010003.
2. Fedotova A., Romanov A., Kurtukova A., Shelupanov A. Authorship attribution of social media and literary Russian-language texts using machine learning methods and feature selection. *Future Internet*. 2021;14(1):4. DOI: 10.3390/fi14010004.
3. Wu H., Zhang Z., Wu Q. Exploring syntactic and semantic features for authorship attribution. *Applied Soft Computing*. 2021;111:107815–107822. DOI: 10.1016/j.asoc.2021.107815.
4. Khomytska I., Bazylevych I., Teslyuk V. The statistical parameters of Ivan Franko’s authorial style determined by the chi-square test. *2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), 2022*. p. 73–76. DOI: 10.1109/CSIT56902.2022.10000491.
5. Chekhovich Y. V., Khazov A. V. Analysis of duplicated publications in Russian journals. *Journal of informetrics*. 2022;16(1):101246. DOI: 10.1016/j.joi.2021.101246.
6. Isachenko V. V., Apanovich Z. Analysis and visualisation system for cross-lingual identification of authors of scientific publications. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Serija: Informacionnye tehnologii = Vestnik NSU. Series: Information Technologies*. 2018; 16(2):49-61. DOI: 10.25205/1818-7900-2018-16-2-49-61 (In Russ.).
7. Agun H.V., Yilmazel O. Incorporating topic information in a global feature selection schema for authorship attribution. *IEEE Access*. 2019;7:98522–98529 DOI: 10.1109/ACCESS.2019.2930536.
8. Kou G., Yang P., Peng Y., Xiao F., Chen Y., Alsaadi F.E. Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing*. 2020;86:105836. DOI: 10.1016/j.asoc.2019.105836.
9. Bardamova M., Hodashinsky I. Hybrid algorithm for tuning feature weights in a fuzzy classifier. *2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT), Yekaterinburg, Russia, 2021*. p. 0354–0357. DOI: 10.1109/USBREIT51232.2021.9455030.
10. Yaseen A., Laftah W., Kadhum I., Hamad A. Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system. *Pattern Recognition*. 2022;108912. DOI: 10.1016/j.patcog.2022.108912.
11. Uchendu A., Le T., Lee D. Attribution and obfuscation of neural text authorship: A data mining perspective. *ACM SIGKDD Explorations Newsletter*. 2023;25(1):1–18. DOI: 10.48550/arXiv.2210.10488.
12. Shamardina T. et al. Findings of the the ruatd shared task 2022 on artificial text detection in Russian. *arXiv preprint arXiv*. 2206;2022;01583. DOI: 10.48550/arXiv.2206.01583.

13. Xu W., Yuan, K., Li, W., Ding, W. An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution. *IEEE Transactions on Emerging Topics in Computational Intelligence*. 2022;7(1):76–88. DOI: 10.1109/TETCI.2022.3171784.
14. Yao G., Xiaojian H., Guanxiong W. A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain. *Expert Systems with Applications*. 2022;117002. DOI: 10.1016/j.eswa.2022.117002.
15. Abu Khurma R., Aljarah I., Sharieh A., Abd Elaziz M., Damaševičius R., Krilavičius T. A review of the modification strategies of the nature inspired algorithms for feature selection problem. *Mathematics*. 2022;10(464). DOI: 10.3390/math10030464.
16. Borboudakis G., Tsamardinos I. Forward-backward selection with early dropping. *The Journal of Machine Learning Research*. 2019;20(1):276–314. DOI: 10.5555/3322706.3322714.
17. Le N.Q.K., Ho Q.T., Nguyen V.N., Chang J.S. BERT-Promoter: An improved sequence-based predictor of DNA promoter using BERT pre-trained model and SHAP feature selection. *Computational Biology and Chemistry*. 2022;99:107732. DOI: 10.1016/j.compbiolchem.2022.107732.
18. New frequency dictionary of Russian vocabulary. URL: <http://dict.ruslang.ru/freq.php> (accessed on 04.12.2023). (In Russ.).

#### ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

**Романов Александр Сергеевич**, кандидат технических наук, доцент, доцент кафедры комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР), Томск, Российская Федерация.

*e-mail:* [alex.romanov@gmail.com](mailto:alex.romanov@gmail.com)

ORCID: [0000-0002-2587-2222](https://orcid.org/0000-0002-2587-2222)

**Aleksandr S. Romanov**, Candidate of Engineering Sciences, Associate Professor, Associate Professor at the Department of Complex Information Security of Electronic Computer Systems (KIBEVS), Tomsk State University of Control Systems and Radioelectronics (TUSUR), Tomsk, the Russian Federation.

*Статья поступила в редакцию 06.12.2023; одобрена после рецензирования 20.12.2023; принята к публикации 16.01.2024.*

*The article was submitted 06.12.2023; approved after reviewing 20.12.2023; accepted for publication 16.01.2024.*