УДК 004.056.52 doi: 10.15827/0236-235X.148.547-553 2024. Т. 37. № 4. С. 547–553

Совершенствование метода оценки стойкости пароля аутентификации пользователя компьютерных систем на основе использования известных уязвимостей

И.Г. Сидоркина [1], С.В. Михалищев [1]✉

[1] Поволжский государственный технологический университет, г. Йошкар-Ола, 424000, Россия

Ссылка для цитирования
Сидоркина И.Г., Михалищев С.В. Совершенствование метода оценки стойкости пароля аутентификации пользователя компьютерных систем на основе использования известных уязвимостей // Программные продукты и системы. 2024. Т. 37. № 4. С. 547–553. doi: 10.15827/0236-235X.148.547-553

Информация о статье
Группа специальностей ВАК: 2.3.6
Поступила в редакцию: 26.07.2024 После доработки: 25.08.2024 Принята к публикации: 27.08.2024

Аннотация. Актуальность исследования обусловлена растущими угрозами безопасности пользовательских данных в современной цифровой среде, где выбор надежных паролей играет критическую роль в защите информации. Усовершенствованный метод основан на применении модели машинного обучения CatBoost, он улучшает существующие подходы к анализу структуры паролей и автоматическому формулированию требований к их сложности. При этом учитываются известные уязвимости, связанные с выбором пользователями слабых или скомпрометированных паролей. Особенностью усовершенствованного метода является глубокий анализ обучающего набора данных и использование словарей запрещенных и скомпрометированных паролей. Это позволяет повысить точность обученной модели и учесть динамически изменяющиеся требования безопасности. Основные результаты работы демонстрируют значительное повышение точности определения стойкости паролей аутентификации пользователей компьютерных систем. При анализе обучающего набора данных, используемого при обучении модели машинного обучения, были выявлены: неактуальное отнесение паролей к определенному классу надежности новым стандартом безопасности; наличие скомпрометированных паролей; отсутствие классифицированных паролей, у которых расстояние от левого символа кодовой таблицы unicode до правого минимально. В отличие от известных решений, проверка пароля по словарям проводилась на начальном этапе, до обучения модели. Это не создает дополнительной нагрузки на модель и не позволяет пользователям использовать запрещенные и скомпрометированные пароли. Практическая значимость работы заключается в интеграции предложенного усовершенствованного метода в системы аутентификации пользователей компьютерных систем. Это позволит исключить использование слабых и скомпрометированных паролей, повысит их эффективность и уровень защиты пользовательских данных, а также снизит риск успешных атак злоумышленников. Кроме того, предложенный усовершенствованный метод может быть адаптирован для различных систем безопасности при интеграции в существующие механизмы проверки паролей. Данное исследование вносит вклад в развитие методов цифровой безопасности и может быть полезно для специалистов в области информационной безопасности и при разработке программного обеспечения.

Ключевые слова: метод определения стойкости пароля, машинное обучение, аутентификация, безопасность информации, модель прогнозирования

Введение. Сохранность пользовательских данных является ключевым аспектом в современной цифровой среде, где злоумышленники постоянно совершают новые атаки и пытаются проникнуть в системы для получения конфиденциальной информации. В этом контексте одной из наиболее критических уязвимостей является выбор ненадежных паролей, которые могут быть легко взломаны с помощью различных методов атаки, таких как перебор, словарные атаки и использование радужных таблиц. Несмотря на широкое использование различных схем аутентификации, проблема слабых паролей остается актуальной и требует надежных методов защиты.

Ранее проведенные исследования в этой области посвящены различным аспектам проверки паролей и методам улучшения стойкости паролей аутентификации. Некоторые из них фокусировались на анализе сложности паролей с использованием статистических методов, в то время как другие исследования предлагали алгоритмы проверки паролей на основе сложных математических моделей.

Однако большинство разработок из-за недостаточной точности или сложности внедрения в реальные системы сталкиваются с определенными ограничениями. Кроме того, эти методы часто не учитывают возможное использование словарей паролей и динамически изменяющиеся требования безопасности.

Данная работа стремится заполнить этот пробел в исследованиях, предлагая усовершенствованный метод проверки стойкости паролей

аутентификации. Формируя базу запрещенных и скомпрометированных паролей, можно повысить точность обученной модели и исключить фактор использования злоумышленниками словарей паролей. Этот подход позволяет учитывать динамические изменения в требованиях к безопасности и обеспечивать более высокий уровень защиты данных.

Цель работы заключается в исключении возможности использования слабых и скомпрометированных паролей для обеспечения защиты пользовательских данных и в усовершенствовании метода определения стойкости пароля аутентификации.

Анализ существующих методов

Согласно отчету [1], в 2020 году 89 % всех типов взломов связано со злоупотреблением учетными данными, включая атаки грубой силы и повторное использование учетных данных. Это обстоятельство подчеркивает важность использования машинного обучения для определения стойкости пароля, что может уменьшить риск большинства возможных атак. В таком контексте значительное количество современных исследований фокусируется на разработке эффективных методов определения стойкости пароля с помощью машинного обучения [2, 3].

Основные уязвимости стойкости паролей аутентификации пользователей компьютерных систем выявила компания Nord Security Inc. (разработчик приложений для хранения и генерации паролей). Ее ежегодные исследования показали, что большая часть пользователей компьютеров и Интернета избегают использования надежных паролей, выбирая простые, легко запоминающиеся комбинации цифр и слов, которые уже давно известны даже рядовым пользователям [4]. Наиболее популярные пароли, применяемые пользователями во всем мире, представлены в таблице 1. Как следствие, такие системы взламываются за секунды, нанося урон персональным данным и конфиденциальной информации.

Для обеспечения безопасности пользовательских данных под запрет должны попадать пароли, включенные в словари запрещенных паролей, содержащих популярные имена, города, дни месяца, части логинов пользователей, номера компьютеров и т.д., и в словари скомпрометированных паролей.

Использование методов машинного обучения [5] для построения верификаторов позволяет алгоритмам формулировать требования к надежности паролей наиболее гибко. Алгоритмы машинного обучения, такие как классификация, регрессия и нейронные сети, анализируют структуру и характеристики вводимых паролей, определяя, какие из них являются потенциально слабыми или уязвимыми. Обучаясь на данных о различных типах паролей и их стойкости, модели машинного обучения могут более точно оценивать стойкость, выявляя скрытые паттерны и слабые места, неочевидные на первый взгляд.

Существующие методы к созданию верификаторов надежности паролей (https://www.kaggle.com/datasets/morph1max/password-security-sber-dataset/data, [6]), как правило, используют популярные модели машинного обучения, такие

Таблица 1

Наиболее распространенные (скомпрометированные) пароли 2021–2023 гг.

Table 1

Most common (compromised) passwords in 2021–2023

2021		2022		2023	
Пароль	**Количество пользователей**	**Пароль**	**Количество пользователей**	**Пароль**	**Количество пользователей**
123456	103 170 552	password	4 929 113	123456	4 524 867
123456789	46 027 530	123456	1 523 537	admin	4 008 850
12345	32 955 431	123456789	413 056	12345678	1 371 152
qwerty	22 317 280	guest	376 417	123456789	1 213 047
password	20 958 297	qwerty	309 679	1234	969 811
12345678	14 745 771	12345678	284 946	12345	728 414
111111	13 354 149	111111	229 047	password	710 321
123123	10 244 398	12345	188 602	123	528 086
1234567890	9 646 621	col123456	140 505	Aa123456	319 725
1234567	9 396 813	123123	127 762	1234567890	302 709

как SGDClassifier, KNeighborsClassifier, CatBoost [7– 9].

Сравнительный результат моделей машинного обучения показал сопоставимые результаты всех проанализированных методов. В данной работе использовалась модель градиентного бустинга CatBoost» [10], создающая решающую модель прогнозирования в виде ансамбля слабых моделей прогнозирования, обычно деревьев решений.

Сохраним предложенную в существующих методах классификацию паролей по уникальным признакам:

- длина пароля (len);
- количество заглавных букв (count_upper_sym);
- количество прописных букв (count_lower_sym);
- количество спецсимволов (count_special_sym);
- количество цифр (count_digit);
- расстояние от левого символа до правого на основе кодовой таблицы unicode (distance_unicode);
- количество уникальных символов в пароле (count_unique_sym).

Итоги экспериментов подтвердили, что наибольшее влияние на стойкость оказывает длина пароля, за ней идут регистр, спецсимволы и наличие цифр (см. рисунок).

Рассмотрим результаты проведенных исследований.

Обучающий набор данных, используемый в существующих методах, содержал в себе 100 000 уникальных классифицированных паролей. Данные в наборе не сбалансированы: 74 % паролей с классом «1», 14 % слабых паролей с классом «0», 12 % сложных паролей с классом «2».

Поскольку набор паролей с классом сложности «1» считается несбалансированным, сравнивать его можно только по метрике $F1_{macro}$, которая учитывает производительность модели для каждого класса независимо от его размера в наборе [11]:

$$F1_{macro} = \frac{1}{N}\sum_{N}^{i=1} F1_i,$$

где N – количество классов; $F1_i$ – F1-мера для каждого отдельного класса.

При работе с обучающим набором данных выполнялись случайное перемешивание и разбивка на обучающий (80 %) и тестовый (20 %) наборы. Для тестирования моделей использовалась перекрестная проверка с пятью делениями набора данных.

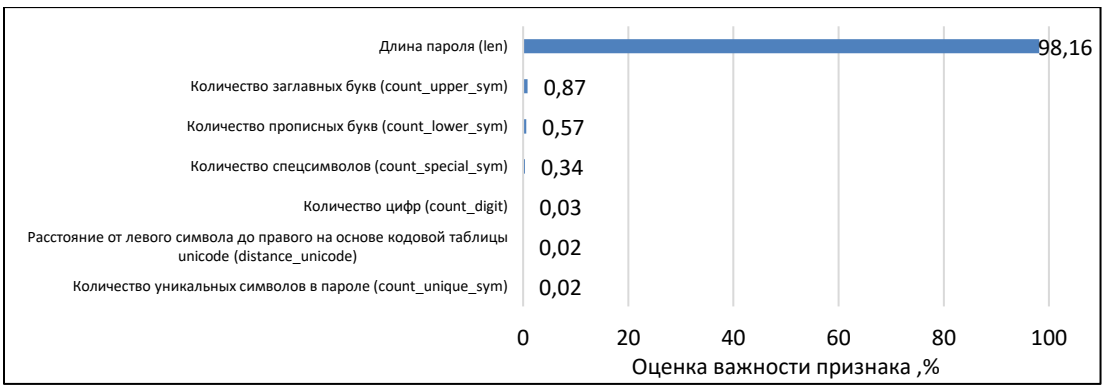
Предложенный метод показал, что применение модели машинного обучения на обучающем наборе методом кросс-валидации по метрике Precision дает почти 100-процентный результат. Однако тестирование модели на валидной выборке, содержащей 70 уникальных классифицированных паролей, показало невысокий результат: $F1_{macro} \approx 0{,}7(7)$.

Выявленные недостатки данного метода могут быть обусловлены

а) отсутствием фильтрации обучающего набора данных перед началом обучения на предмет использования скомпрометированных и запрещенных паролей;

б) использованием обучающего набора данных на основе кодовой таблицы unicode, в котором не представлены пароли с минимальным расстоянием от левого до правого символа;

в) доверительным отношением к классификации паролей в обучающем наборе данных, которая не подвергалась сомнению и не проходила дополнительных проверок.



Зависимость надежности пароля от признака

Dependence of password strength on attribute

Данные недостатки и достоинства учтены при разработке усовершенствованного метода оценки стойкости пароля аутентификации пользователя компьютерных систем.

Определение стойкости пароля аутентификации

В соответствии с документом ФСТЭК (https://fstec.ru/dokumenty/vse-dokumenty/prikazy/prikaz-fstek-rossii-ot-18-fevralya-2013-g-n-21) для достижения требований четвертого уровня защищенности персональных данных пароль должен содержать не менее 8 символов при алфавите пароля не менее 70 символов. Максимальное
количество неуспешных попыток аутентификации (ввода неправильного пароля) до блокировки – 4.

Проанализируем валидную выборку, используемую для тестирования рассмотренных моделей.

Наибольший процент ошибок пришелся на пароли, у которых расстояние от левого символа до правого на основе кодовой таблицы unicode минимально. Представим примеры паролей, стойкость которых отнесена обученной моделью к классу надежности «2»: «bbaaccdaa aabdcbada», «dadaaabaacbbadbadc», «31222a23 2331131312», «32a331132312121232» и аналогичные. В действительности их стойкость по валидному набору данных равна единице.

Для устранения ошибки, связанной с отсутствием в обучающем наборе данных паролей, характеризующихся минимальной дисперсией кодов unicode для крайних символов, в обучающий набор данных будет добавлено 20 паролей, подобных «deeddfdffeddddfdee», «4456554 6d455644664» и аналогичных, со значением класса надежности, равным единице.

Перед началом обучения модели применим базу скомпрометированных паролей (https://www.kaggle.com/datasets/joebeachcapital/top-10-million-passwords?resource=download), содержащую 999 997 записей, и исключим эти записи из обучающего набора данных. После применения обучающий набор данных, содержащий 100 020 уникальных значений, классифицированных по уровням сложности, сократился до 96 776 значений.

Протестируем усовершенствованную модель. Для обучения будем использовать модель градиентного бустинга CatBoost. Классификацию паролей выполним по признакам, приведенным на рисунке.

Основную результирующую оценку модели будем выполнять по метрике $F1_{macro}$, используя формулу, представленную выше.

Результат обучения и оценки модели по метрике $F1_{macro}$ равен 1.

Повторное тестирование модели на валидной выборке, используемой в рассмотренных аналогах, показало: $F1_{macro} \approx 1$.

Сгенерируем валидный набор данных из 100 паролей. Для генерации использовался сервис «Рандомус» (https://randomus.ru/password). Для проверки надежности каждого пароля и отнесения его к определенному классу стойкости использовался сервис от лаборатории Касперского (https://password.kaspersky.com/ru). В результате получили валидный набор, содержащий несбалансированные данные: 60 паролей с классом «1», 15 слабых паролей с классом «0», 25 сложных паролей с классом «2».

Проверим обученную модель на валидной выборке по метрике $F1_{macro}$, используя следующий алгоритм.

1. Преобразование массивов:

– преобразуем массив предсказанных значений pred_valid в одномерный;

– преобразуем столбец strength DataFrame df_valid_features в одномерный массив.

2. Вычисление F1-меры:

– используем функцию f1_score для вычисления с макроусреднением на основе одномерных массивов pred_valid и df_valid_features. strength.

3. Вывод значения F1-меры с макроусреднением.

4. Создание и обработка массива ошибок:

– создаем массив ошибок со значением типа булево, сравнивая значения массивов pred_valid и df_valid_features.strength;

– выбираем номера строк из DataFrame df_valid_features, где обнаружены ошибки.

5. Вывод номеров строк с ошибками.

В результате выполнения алгоритма установлено, что значение F1-меры с макроусреднением составляет 0,8154121863799283. Ошибки обнаружены в валидном наборе данных с номерами строк: 5, 6, 7, 8, 9, 10, 11, 12, 13 и 14.

Анализ допущенных ошибок показал, что обученная модель отнесла 10 слабых паролей к классу надежности «1» и ни разу не ошиблась при классификации сильных паролей. Пароли, стойкость которых отнесена обученной моделью к классу надежности «1»: «e_4wKT~1», «_3>hN84O», «_S$8/I/x», «-JNkz{s6», «-3@Vy Kl.», «e_4wKT~1», «_3>hN84O», «_S$8/I/x», «-JNkz{s6» и «-3@VyKl». В действительности

их стойкость по валидному набору данных равна нулю. Все представленные пароли, на которых обученная модель показала низкую точность при классификации, имеют одинаковую длину, равную 8 знакам.

Обучающий набор данных содержит 16 816 паролей длиной 8 знаков, все из них были отнесены к классу надежности «1». Однако валидный набор данных включал 10 паролей той же длины, которые были отнесены к классу надежности «0». Это расхождение в классификации обучающего и валидного наборов данных и привело к значительным ошибкам.

Для улучшения точности модели было принято решение о пересмотре и корректировке обучающего набора данных. Все 16 816 паролей длиной 8 знаков переклассифицированы из класса надежности «1» в класс надежности «0». Это изменение отражает актуальные требования к надежности паролей и необходимость их пересмотра в соответствии с новыми стандартами безопасности.

После внесения этих корректировок точность обученной модели значительно улучшилась. Результаты тестирования на валидной выборке показали точность по метрике $F1_{macro}$, равную 1, что свидетельствует о правильной классификации всех паролей в валидной выборке (табл. 2). Таким образом, реализация усовершенствований метода привела к увеличению точности обученной модели на 23,2 %.

Для получения результата, близкого к стопроцентной точности, были предприняты следующие шаги.

1. Исключение из обучающего набора паролей по словарю скомпрометированных и запрещенных паролей, что позволило снизить вероятность переобучения модели на неинформативных данных для обобщения результатов на новых, ранее не рассматриваемых паролях.

2. Добавление в обучающую выборку паролей, у которых расстояние от левого символа до правого на основе кодовой таблицы unicode минимально. Этот шаг был направлен на повышение энтропии генерируемых паролей, что затрудняет их взлом методами перебора и криптоанализа.

3. Переклассификация паролей длиной 8 знаков. Все 16 816 паролей длиной 8 знаков были переклассифицированы из класса надежности «1» в класс надежности «0». Это изменение отражает актуальные требования к надежности паролей и необходимости их пересмотра в соответствии с новыми стандартами безопасности.

Такую модель следует применять только после проверки пользовательского пароля на наличие его в словарях скомпрометированных и запрещенных паролей. Исключение слабых паролей из анализа не только повышает эффективность работы модели, но и снижает объем вычислений, необходимых для обработки данных. Такой подход обеспечивает более быструю и точную оценку стойкости паролей, что в конечном итоге способствует надежной защите пользовательских данных.

Заключение

Таким образом, показано, что применение модели машинного обучения «градиентный бустинг с CatBoost», использование словаря скомпрометированных и запрещенных паролей при ее обучении, а также применение анализа и корректировка обучающего набора данных повышают эффективность метода точности определения стойкости пароля аутентификации до значений, близких к 100 %, что является существенным усилением защиты пользовательских данных от атак злоумышленников.

Таблица 2

Результат тестирования обученной модели на валидной выборке

Table 2

Result of testing the trained model on a valid sample

Выборка	Точность (precision)	Полнота (recall)	F1-мера (fl-score)	Поддержка (support)
Класс надежности 0	1,00	1,00	1,00	15
Класс надежности 1	1,00	1,00	1,00	60
Класс надежности 2	1,00	1,00	1,00	25
Точность (accuracy)			1,00	100
Среднее (macro avg)	1,00	1,00	1,00	100
Средневзвешенное (weighted avg)	1,00	1,00	1,00	100

При создании модели любой набор данных перед использованием должен быть проверен и приведен в соответствие действующим требованиям к защищенности персональных данных. Это осуществляется с учетом словаря скомпрометированных и запрещенных паролей и технических возможностей злоумышленников. Модель машинного обучения (обученная на специально подготовленном наборе данных с вероятностью, близкой к 100 %) позволит классифицировать надежность пароля аутентификации и проинформирует пользователя об ограничениях на использование слабых паролей.

В процессе разработки усовершенствованного метода оценки стойкости пароля были учтены достоинства существующих решений, включая использование алгоритмов машинного обучения и статистического анализа. Проведенный анализ также выявил их недостатки, такие как отсутствие комплексной проверки данных и неэффективное использование известных уязвимостей. Внедрение БД скомпрометированных и запрещенных паролей позволило устранить указанные недостатки, что способствовало повышению надежности оценки стойкости паролей.

Применение усовершенствованного метода определения стойкости пароля аутентификации повысит эффективность защиты информации и обеспечит повышение безопасности в системах управления доступом, в мобильных/экранных приложениях, в веб-приложениях, таких как онлайн-банкинг и электронная почта.

Список литературы

1. Jartelius M. The 2020 data breach investigations report – a CSO's perspective. Network Security, 2020, vol. 2020, no. 7, pp. 9–12. doi: 10.1016/S1353-4858(20)30079-9.

2. Sarkar S., Nandan M. Password strength analysis and its classification by applying machine learning based techniques. Proc. ICCSEA, 2022, pp. 1–5. doi: 10.1109/ICCSEA54677.2022.9936117.

3. Sakya S.S., Mauparna M.N. Building a multi-class password strength generator and classifier model by augmenting supervised machine learning techniques. Research Square, 2022. doi: 10.21203/rs.3.rs-1820885/v1.

4. Zaidi T., Garai S., Biradar T.V. Exploring the landscape of password managers for individual users through innovative solution. In: Information Technology Security. STEEE, 2024, pp. 69–99. doi: 10.1007/978-981-97-0407-1_4.

5. Назарова А.Д. Анализ надежности паролей для защиты данных // Умная цифровая экономика. 2022. Т. 2. № 4. С. 41–46.

6. Беликов В.В., Прокуронов И.А. Построение верификатора стойкости пароля с использованием классических методов машинного обучения и рекуррентной LSTM нейронной сети // Russ. Tech. J. 2023. Т. 11. № 4. С. 7–15. doi: 10.32362/2500-316X-2023-11-4-7-15.

7. Li X., Orabona F. On the convergence of stochastic gradient descent with adaptive stepsizes. Proc. Int. Conf. on Artificial Intelligence and Statistics. PMLR, 2019, vol. 89, pp. 983–992.

8. Lubis A.R., Lubis M., Khowarizmi A. Optimization of distance formula in K-Nearest Neighbor method. BEEI, 2020, vol. 9, no. 1, pp. 326–338. doi: 10.11591/eei.v9i1.1464.

9. Coronado-Blázquez J. Classification of Fermi-LAT unidentified gamma-ray sources using CatBoost gradient boosting decision trees. MNRAS, 2022, vol. 515, no. 2, pp. 1807–1814. doi: 10.1093/MNRAS/STAC1950.

10. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: Unbiased boosting with categorical features. Proc. NeurIPS, 2018, vol. 31, pp. 6638–6648.

11. Lever J., Krzywinski M., Altman N. Classification evaluation. Nat. Methods, 2016, vol. 13, pp. 603–604. doi: 10.1038/nmeth.3945.

Software & Systems doi: 10.15827/0236-235X.148.547-553 2024, 37(4), pp. 547–553

Assessing the strength of a computer system user authentication password: Improving the method based on using known vulnerabilities

Irina G. Sidorkina [1], Stanislav V. Mikhalishchev [1]✉

[1] Volga State University of Technology, Yoshkar-Ola, 424000, Russian Federation

For citation

Sidorkina, I.G., Mikhalishchev, S.V. (2024) 'Assessing the strength of a computer system user authentication password: Improving the method based on using known vulnerabilities', *Software & Systems*, 37(4), pp. 547–553 (in Russ.). doi: 10.15827/0236-235X.148.547-553

Article info
Received: 26.07.2024 After revision: 25.08.2024 Accepted: 27.08.2024

Abstract. The relevance of the research is due to the growing threats to user data security in the modern digital environment; the choice of strong passwords plays a critical role in protecting information. The improved method is based on applying the CatBoost machine-learning model. It improves existing approaches to analyzing password structure and automatic formulating requirements for password complexity. It takes into account well-known vulnerabilities associated with users choosing weak or compromised passwords. A special feature of the improved method is a deep analysis of the training data set and using dictionaries of forbidden and compromised passwords. This allows improving the accuracy of the trained model and take into account dynamically changing security requirements. The main results of the work demonstrate a significant improvement in the accuracy of password strength determination for authentication of computer system users. When analyzing the training dataset used to train the machine-learning model, the authors identified some issues. These were irrelevant assignment of passwords to a certain security class by a new security standard; compromised passwords; the absence of classified passwords that have a minimum distance from the left character of the Unicode code table to the right one. In contrast to known solutions, password verification by dictionaries took place at the initial stage, before training the model. It does not create additional load on the model and does not allow users to use forbidden and compromised passwords. The practical significance of this work is in implementing the improved method in user authentication systems. This will eliminate weak and compromised passwords, increase their efficiency and the level of user data protection, reduce the risk of successful attacks. Furthermore, the authors can adapt the proposed method for various security systems when integrated into existing password verification mechanisms. This research contributes to the development of digital security techniques. It can be useful for information security and software engineering professionals.
Keywords: password strength determination method, machine learning, authentication, information security, prediction model

References

1. Jartelius, M. (2020) ‘The 2020 data breach investigations report – a CSO's perspective’, *Network Security*, 2020(7), pp. 9–12. doi: 10.1016/S1353-4858(20)30079-9.
2. Sarkar, S., Nandan, M. (2022) ‘Password strength analysis and its classification by applying machine learning based techniques’, *Proc. ICCSEA*, pp. 1–5. doi: 10.1109/ICCSEA54677.2022.9936117.
3. Sakya, S.S., Mauparna, M.N. (2022) ‘Building a multi-class password strength generator and classifier model by augmenting supervised machine learning techniques’, *Research Square*. doi: 10.21203/rs.3.rs-1820885/v1.
4. Zaidi, T., Garai, S., Biradar, T.V. (2024) ‘Exploring the landscape of password managers for individual users through innovative solution’, in *Information Technology Security. STEEE*, pp. 69–99. doi: 10.1007/978-981-97-0407-1_4.
5. Nazarova, A.D. (2022) ‘Password strength analysis for data protection’, *Smart Digital Economy*, 2(4), pp. 41–46 (in Russ.).
6. Belikov, V.V., Prokuronov, I.A. (2023) ‘Password strength verification based on machine learning algorithms and LSTM recurrent neural networks’, *Russ. Tech. J.*, 11(4), pp. 7–15 (in Russ.). doi: 10.32362/2500-316X-2023-11-4-7-15.
7. Li, X., Orabona, F. (2019) ‘On the convergence of stochastic gradient descent with adaptive stepsizes’, *Proc. Int. Conf. on Artificial Intelligence and Statistics. PMLR*, 89, pp. 983–992.
8. Lubis, A.R., Lubis, M., Khowarizmi, A. (2020) ‘Optimization of distance formula in K-Nearest Neighbor method’, *BEEI*, 9(1), pp. 326–338. doi: 10.11591/eei.v9i1.1464.
9. Coronado-Blázquez, J. (2022) ‘Classification of Fermi-LAT unidentified gamma-ray sources using CatBoost gradient boosting decision trees’, *MNRAS*, 515(2), pp. 1807–1814. doi: 10.1093/MNRAS/STAC1950.
10. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A. (2018) ‘CatBoost: Unbiased boosting with categorical features’, *Proc. NeurIPS*, 31, pp. 6638–6648.
11. Lever, J., Krzywinski, M., Altman, N. (2016) ‘Classification evaluation’, *Nat. Methods*, 13, pp. 603–604. doi: 10.1038/nmeth.3945.

Авторы
Сидоркина Ирина Геннадьевна [1], д.т.н.,
профессор, заведующий кафедрой,
igs592000@mail.ru
Михалищев Станислав Вячеславович [1],
магистр, соискатель,
mihalischevstas@gmail.com

[1] Поволжский государственный
технологический университет,
г. Йошкар-Ола, 424000, Россия

Authors
Irina G. Sidorkina [1], Dr.Sci. (Engineering),
Professor, Head of Chair,
igs592000@mail.ru
Stanislav V. Mikhalishchev [1],
Master of Science, Candidate,
mihalischevstas@gmail.com

[1] Volga State University of Technology,
Yoshkar-Ola, 424000,
Russian Federation