

Научная статья

УДК 004.9+519.68

DOI 10.35266/1999-7604-2024-2-8



МОДУЛЬ ИНФОРМАЦИОННОЙ СИСТЕМЫ НА ОСНОВЕ НАИВНОГО БАЙЕСОВСКОГО КЛАССИФИКАТОРА ДЛЯ АВТОМАТИЗАЦИИ РАБОТЫ АПТЕКИ

Кирилл Валерьевич Святлов¹, Вадим Сергеевич Мошкин², Игорь Александрович Щукарев³✉

^{1, 2, 3}Ульяновский государственный технический университет, Ульяновск, Россия

¹k.svyatov@ulstu.ru, <https://orcid.org/0000-0001-8978-8567>

²v.moshkin@ulstu.ru, <https://orcid.org/0000-0002-9258-4909>

³blacxpress@gmail.com✉, <https://orcid.org/0000-0002-8860-8103>

Аннотация. Эффективным способом увеличения прибыли и укрепления рентабельности аптеки становится внедрение принципов автоматизации и управления деятельностью фармацевта с помощью информационных систем. Компании используют алгоритмы машинного обучения для корректировки своей стратегии, изучения отношения клиентов к своей организации посредством анализа отзывов и для повышения имиджа фирмы. Однако ручная обработка поступающих отзывов требует значительного времени и усилий фармацевта. Автоматизировать данный вид деятельности в статье предлагается за счет алгоритма наивного байесовского классификатора, реализованного средствами Python. Для обучения классификатора был создан собственный корпус размеченных текстов отзывов с двумя категориями, суммарное количество отзывов около 500. Для поиска отзывов использовался парсер, написанный на Python. В рамках предварительной обработки текста отзывов были выполнены: лемматизация, удаление знаков пунктуации, процедура приведения текста к нижнему регистру, токенизация и удаление стоп-слов, а в качестве способа векторизации текста был выбран метод «Bag of Words», или мешок слов. Согласно проведенным численным экспериментам, наивысшая точность классификатора достигалась при соотношении обучающей и тестовой выборки 80/20, без стоп-слов. При использовании классификатора на анализ 100 отзывов потребуется в восемь раз меньше времени по сравнению с их чтением человеком. Сам классификатор может быть представлен как отдельное приложение или как модуль информационной системы. Таким образом, растущее количество положительных отзывов у фирмы является показателем ее успешной работы и числа довольных клиентов, а рост имиджа позволит увеличить доверие покупателей к фирме и приведет к росту продаж.

Ключевые слова: классификация текстов, машинное обучение, корпус, токен, стоп-слово, лемматизация, обработка естественных языков, наивный байесовский классификатор, информационная система, Python

Для цитирования: Святлов К. В., Мошкин В. С., Щукарев И. А. Модуль информационной системы на основе наивного байесовского классификатора для автоматизации работы аптеки // Вестник кибернетики. 2024. Т. 23, № 2. С. 62–70. DOI 10.35266/1999-7604-2024-2-8.

INFORMATION SYSTEM MODULE BASED ON NAIVE BAYES CLASSIFIER FOR PHARMACY OPERATION AUTOMATION

Kirill V. Svyatov¹, Vadim S. Moshkin², Igor A. Shchukarev³✉

^{1, 2, 3}Ulyanovsk State Technical University, Ulyanovsk, Russia

¹k.svyatov@ulstu.ru, <https://orcid.org/0000-0001-8978-8567>

²v.moshkin@ulstu.ru, <https://orcid.org/0000-0002-9258-4909>

³blacxpress@gmail.com✉, <https://orcid.org/0000-0002-8860-8103>

Abstract. The introduction of the automation and management principles for pharmacists' activities using information systems is an effective way to increase profits and strengthen pharmacy profitability. Companies use machine learning algorithms to adjust their strategy, study customer reviews about their organization through feedback analysis, and improve the company's image. However, a significant amount of pharmacists' time and effort goes into the manual processing of incoming reviews. The article proposes automating the processing using the naive Bayes classifier algorithm, which is implemented in Python. To train the classifier, the authors created their own corpus of labeled review texts with two categories, with the total number of reviews being about 500. A parser written in Python was used to search for reviews. The following steps were taken during the preliminary processing of the reviews' text: lemmatization, elimination of punctuation, the text's letter conversion to lower case, tokenization, stop words removal, and text vectorization using the bag-of-words model. According to the numerical experiments carried out, the classifier's highest accuracy was achieved with an 80/20 ratio of training and test samples that did not include stop words. When using a classifier, analyzing 100 reviews takes 8 times less time than manual reading. The classifier itself can be presented as a separate application or as a module within an information system. Thus, the increasing number of positive reviews is an indicator of the company's efficient performance as well as the number of satisfied customers. The growth of its image will strengthen customers' loyalty to the company, resulting in greater sales.

Keywords: text classification, machine learning, corpus, token, stop word, lemmatization, natural language processing, naive Bayes classifier, information system, Python

For citation: Svyatov K. V., Moshkin V. S., Shchukarev I. A. Information system module based on naive Bayes classifier for pharmacy operation automation. *Proceedings in Cybernetics*. 2024;23(2):62–70. DOI 10.35266/1999-7604-2024-2-8.

ВВЕДЕНИЕ

Имидж является одним из важнейших средств достижения компанией своих целей. Зачастую при принятии покупателем решения о покупке товара или услуги репутация организации выходит на первый план. Парой даже негативные отзывы могут заставить клиента изменить свое решение. Потребители ощущают сомнения при совершении покупок при отсутствии какой-либо информации о компании. Поэтому растущее количество отзывов, желательно положительных, у фирмы является показателем ее успешной работы и числа довольных клиентов [1, 2]. Сегодня многие предприятия и организации используют в своей работе разнообразные информационные системы. Они могут быть связаны с различными областями деятельности предприятия, такими как автоматизация деятельности, бухгалтерия, управление персоналом и т.д. Под аптекой понимается специализированная организация системы здравоохранения, которая занимается производством, фасовкой и реализацией населению ассортимента лекарственных средств, биологически активных добавок и товаров медицинского назначения [3].

В ежедневной деятельности аптеки большое количество времени тратится, как правило, на работу с поставщиками, заказ товара, продажу лекарственных средств, консультирование покупателей, работу с ценниками, накладными и т.д., кроме того, необходимо позаботиться и о способах привлечения клиентов в аптеку, что позволит увеличить поток покупателей и приведет к росту прибыли.

Направление sentiment-анализа активно применяется на практике, а применение алгоритмов машинного обучения в коммерческих целях сегодня диктуется необходимостью классификации собранных статистических данных [4, 5]. Авторами работы [6] приводятся результаты численного эксперимента, в которых исследовались методы машинного обучения для классификации отзывов клиентов. В условиях современной экономики большинство предприятий делают ставку на долгосрочные отношения с клиентами и их позитивные эмоции в процессе потребления услуги. Для высокой конкурентоспособности предприятию необходимо эффективно управлять процессом оказания услуг, например с помощью анализа собираемой информации

от клиентов. Авторами проведено исследование отзывов клиентов с портала tophotels.ru. Показано, что машинное обучение позволяет классифицировать отзывы с точностью 85–88 %, а лемматизация повышает точность классификации отзывов на русском языке. В статье [7] анализируются отзывы клиентов относительно двух программных продуктов с целью сформулировать рекомендации по их улучшению для лиц, принимающих решение. Для решения поставленной задачи авторы использовали информацию, доступную на интернет-ресурсах. В качестве метода анализа был выбран наивный байесовский классификатор. Также авторами была разработана программа «OtClik». Методика анализа тональности текстовой информации включала этапы лемматизации и разбиение текста на униграммы. Программная реализация выполнена на языке Python в виде десктопного приложения на компьютер. Авторами работы [8] рассматривается проблема анализа отзывов об отелях в сфере туризма. Туристы рассказывают о своих впечатлениях от пребывания в отеле, оставляя отзывы. При наличии большого количества отзывов туристы не могут понять, содержат ли они положительные или отрицательные мнения. Чтобы быстро определить, являются ли отзывы положительными или отрицательными, необходимо провести соответствующий анализ. Авторами предлагается решение путем классификации положительных и отрицательных отзывов с использованием метода наивного байесовского классификатора.

В настоящей работе в рамках одного из направлений автоматизации аптеки предлагается использовать классификатор. В результате обратной связи собираются отзывы о работе аптечной организации, а с помощью обученного классификатора на основе размеченных данных фармацевт выбирает только позитивные отзывы и выставляет их на сайте компании с целью повышения привлекательности аптеки для покупателей, поэтому использование методов машинного обучения для обработки большого числа поступающих отзывов вместо их ручной

классификации более чем оправданно [9]. Однако перед анализом текста необходимо провести его предварительную обработку, например с помощью Python, который является высокоуровневым языком программирования с открытым исходным кодом. Для работы с Python необходима и интегрированная среда разработки или IDE. PyCharm – это кросс-платформенная IDE, которая предоставляет пользователю комплекс средств для написания кода с возможностью выявления в нем ошибок [10]. Чтобы оценить, насколько эффективно работает классификатор, можно использовать одну из стандартных метрик, например Accurasy. В качестве метода классификации данных был выбран наивный байесовский классификатор (Naive Bayes Classifier), среди преимуществ которого можно выделить высокую скорость работы и простоту программной реализации [11]. Байесовский классификатор можно использовать для систематизации документов с прямыми отношениями между признаками и соответствующими категориями, например для обнаружения рекламного контента, группировки отзывов об товарах, услугах или организациях.

МАТЕРИАЛЫ И МЕТОДЫ

Для решения поставленной задачи предлагается использовать алгоритм наивного байесовского классификатора, программная реализация которого была выполнена как отдельная функция информационной системы организации. Для решения задачи классификации предварительно был создан корпус текстов с двумя категориями – положительные и отрицательные отзывы, а чтобы компьютер мог обрабатывать информацию, она была представлена в машиночитаемой форме [12].

Предположим, что имеется отзыв о работе аптечного пункта $o_i \in O$, $i = 1, 2, \dots, N$, где $O = \{o_1, o_2, \dots, o_n\}$ – множество отзывов в корпусе, а N – размерность корпуса. Под K будем понимать набор соответствующих категорий (положительные и отрицательные отзывы) $K = \{k_1, k_2\}$. Используя обучающую и тестовую

вую выборки с помощью метода машинного обучения, была получена классифицирующая функция, которая задает отображение множества отзывов во множество категорий $f: O \rightarrow K$. В таком случае формула Байеса со строгими (наивными) предположениями о том, что значение любого заданного признака не зависит от значений других признаков, примет вид:

$$P(k|o) = \frac{P(o|k) \cdot P(k)}{P(o)},$$

где $P(k|o)$ – вероятность того, что отзыв o принадлежит категории k ;

$P(o|k)$ – вероятность встретить отзыв o среди всех отзывов категории k ;

$P(k)$ – вероятность встретить отзыв o категории $k \in K$ среди всех отзывов O ;

$P(o)$ – вероятность встретить отзыв o .

Шанс того, что отзыв o принадлежит категории k с учетом оценки апостериорного максимума (Maximum a posteriori estimation):

$$k_{\text{мапс}} = \arg \max_{k \in K} \frac{P(o|k) \cdot P(k)}{P(o)}.$$

Шанс встретить отзыв o всегда один и тот же для любого отзыва $o \in O$, т.е. $P(o)$ является константой и не может повлиять на ранжирование категорий:

$$k_{\text{мапс}} = \arg \max_{k \in K} (P(o|k) \cdot P(k)).$$

В естественном языке зачастую вероятность появления какого-либо слова зависит от контекста. Алгоритм байесовского классификатора представляет отзыв о работе аптечного пункта или аптеки как набор слов, вероятности которых условно не зависят друг от друга. Таким образом, условная вероятность отзыва аппроксимируется произведением условных вероятностей всех слов, входящих в отзыв:

$$P(o|k) \approx P(o_1|k) \cdot P(o_2|k) \cdots$$

$$\cdots P(o_n|k) = \prod_{i=1}^n P(o_i|k),$$

$$k_{\text{мапс}} = \arg \max_{k \in K} \left(P(k) \cdot \prod_{i=1}^n P(o_i|k) \right).$$

Чтобы оценить, насколько эффективно работает классификатор, можно использовать, например, метрику Ассигасу:

$$A = \frac{TP + TN}{TP + TN + FP + FN},$$

где TP – количество отзывов, которые принадлежат данной категории и которые были правильно определены классификатором;

TN – количество отзывов, которые не принадлежат данной категории и которые были правильно определены классификатором;

FP – количество отзывов, которые принадлежат данной категории и которые были неправильно определены классификатором;

FN – количество экземпляров, которые не принадлежат данной категории и которые были неправильно определены классификатором.

Наивный байесовский классификатор был реализован средствами Python с помощью библиотеки nltk, предназначенной для обработки естественного языка (рис. 1) [13]. Для обучения классификатора был создан собственный корпус размеченных текстов отзывов с двумя категориями. Для создания корпуса использовался парсер, написанный на Python (рис. 2) [14]. В рамках предварительной обработки текста отзывов были выполнены: лемматизация, удаление знаков пунктуации, процедура приведения текста к нижнему регистру, токенизация и удаление стоп-слов, а в качестве способа векторизации текста был выбран метод «Bag of Words», или мешок слов.

Пример положительного отзыва из корпуса: «Хорошая аптека с круглосуточным графиком работы. Профессиональные работники. Можно заказать лекарства, витамины и косметику через их сайт с быстрой доставкой в аптеку. Очень часто проходят акции на некоторые лекарства и средства, можно здорово сэкономить». Пример отрицательного отзыва из корпуса: «Цены на препараты не соответствуют. На сайте одна цена, а в аптеке другая, дороже в 2 раза. И дороже, чем по всей России. Ужасная аптека, разориться можно».


```
def slov(corpus):
    m = Mystem()
    lemma = m.lemmatize(corpus)
    lemmas = "".join(lemma).strip()
    corp_punct = re.sub( pattern: r'[\w\s]', repl: '', lemmas)
    corp_reg = corp_punct.lower()
    tokens = word_tokenize(corp_reg)
    corp_stop = [word for word in tokens if not word in stopwords.words('Russian')]
    corp_bigr = nltk.bigrams(corp_stop)
    return dict([(word, True) for word in corp_stop])
# print(slov(otz_1))

usage
def label(corpus):
    lab = collections.defaultdict(list)
    for i in corpus.categories():
        for j in corpus.fileids(categories=[i]):
            lab[i].append(slov(corpus.raw(fileids=[j])))
    return lab
# print(label(my_corpus))
# print(label(my_corpus).items())

n = 0.8
usage
def split(corpus):
    train = []
    test = []
    for i, j in corpus.items():
        k = int(len(j)*n)
        train.extend([(m, i) for m in j[:k]])
        test.extend([(m, i) for m in j[k:]])
    return train, test
train, test = split(label(my_corpus))
```

Рис. 1. Фрагмент кода классификатора на Python

Примечание: написано авторами для исследования.

```
2 import requests
3 html = requests.get('https://spb.blizko.ru/naiti/byd_zdorov/gde_kypit_lekarstva/apteki/company_reviews').text
4 soup = BeautifulSoup(html) # создаём экземпляр класса BeautifulSoup
5 links = soup.find_all('div', class_='text-') # получаем список ссылок и наименований
6 # print(html)
7 # print(links)
8
9 for i, link in enumerate(links):
10     name = link.text() # извлекаем наименование из блока со ссылкой
11     print(i)
12     print(f"{name}")
13     file = open(f"{i+1}.txt", "w", encoding="utf-8")
14     file.write(f"{name}")
15     file.close()
```

Рис. 2. Фрагмент кода парсера на Python

Примечание: написано авторами для исследования.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Для тестирования и оценки точности наивного байесовского классификатора были проведены серии экспериментов, результаты которых приведены в таблице. Точность числялась с помощью библиотеки nltk Python.

Как видно из таблицы, наивысшая точность классификатора достигается при соотношении обучающей и тестовой выборки 80/20 без стоп-слов. Для проверки корректности работы классификатора дополнительно были найдены два отзыва, не входящих

в исходный корпус, которые классификатор верно отнес к соответствующим категориям (рис. 3).

Согласно исследованиям, оптимальная скорость чтения составляет от 120 до 150 слов в минуту. Именно при такой скорости достигается наилучшее понимание смысла текста. При скорости чтения, большей чем 150 слов в минуту, например 180 (темп скороговороки), или меньшей 120 слов в минуту, человек с трудом воспринимает смысл прочитанного, что приводит к необходимости читать текст несколько раз [15]. Поэтому при больших объемах данных предпочтительнее использовать методы автоматической обработки информации, основанные на алгоритмах машинного обучения. Для целесообразности использования байесовского классификатора в рамках одного из направлений автоматизации аптеки

был проведен следующий эксперимент. Были собраны 100 отзывов и посчитано время, необходимое на чтение этих отзывов с целью их сортировки по категориям, и время, затраченное обученным байесовским классификатором. Результаты эксперимента приведены на рис. 4.

Как видно из рис. 4, на анализ 100 отзывов требуется в восемь раз меньше времени при использовании классификатора, реализованного средствами Python, по сравнению с чтением человеком. Это время может варьироваться в зависимости от конфигурации компьютера, на котором выполняется программа. Сам классификатор может быть представлен как отдельное приложение со своим графическим интерфейсом [16] (рис. 5a) или как модуль некоторой разработанной информационной системы (рис. 5b).

Таблица

Точность наивного байесовского классификатора в зависимости от соотношения обучающей и тестовой выборки

Обучение/тест, %	70/30	80/20	90/10
Со стоп-словами	0,86	0,85	0,86
Без стоп-слов	0,87	0,88	0,84

Примечание: составлено авторами на основании данных, полученных в исследовании.

```
D:\Programs\Python311\python.exe D:\Programs\OneDrive\Sample\PythonProject\b_cl\b_cl.py
Всего категорий: 2
Количество негативных отзывов: 250; Количество позитивных отзывов: 250
Train: 80.0 %; Обучаем на: 400; Тестируем на: 100; Точность: 0.88
Most Informative Features
    удобно = True          pos : neg = 15.7 : 1.0
    удобный = True         pos : neg = 13.0 : 1.0
    хороший = True         pos : neg = 12.5 : 1.0
    спасибо = True         pos : neg = 11.0 : 1.0
    отличный = True        pos : neg = 10.6 : 1.0
    вежливый = True         pos : neg = 10.4 : 1.0
    быстро = True          pos : neg = 8.6 : 1.0
    говорить = True        neg : pos = 7.7 : 1.0
    находить = True        pos : neg = 7.4 : 1.0
    вместо = True          neg : pos = 7.0 : 1.0
Тестирование на позитивном отзыве: pos
Тестирование на негативном отзыве: neg
```

Рис. 3. Результат обучения классификатора

Примечание: составлено авторами на основании данных, полученных в исследовании.

```
D:\Programs\Python311\python.exe D:\Programs\OneDrive\Sample\PythonProject\b_cl_test\t.py
Всего: 100 отзывов
['neg', 'pos', 'neg', 'neg', 'neg', 'pos', 'neg', 'pos', 'neg', 'pos', 'neg', 'pos', 'neg',
Потребуется для чтения 16.00 Минут
Время работы классификатора 2.00 минут

Process finished with exit code 0
```

Рис. 4. Время, необходимое на классификацию 100 отзывов

Примечание: составлено авторами на основании данных, полученных в исследовании.



Рис. 5. Классификатор (а) и пример интерфейса информационной системы с функцией классификатора отзывов (b)

Примечание: составлено авторами для исследования.

ЗАКЛЮЧЕНИЕ

Итак, в данной статье показано, что использование методов машинного обучения, а именно наивного байесовского классификатора, в рамках одного из направлений по автоматизации работы аптеки или аптечного пункта позволит существенно сэкономить время фармацевта. Предложенный байесовский классификатор, точность которого составляет порядка 88 %, позволит в восемь раз уменьшить время на обработку откликов в расчете на каждые 100 отзывов. Позитивные отзывы можно

публиковать в социальных сетях или на сайте компании. Таким образом, рост имиджа аптеки, формируемый, в частности, и на основе положительных отзывов, позволит увеличить доверие покупателей к фирме и приведет к росту продаж. В рамках дальнейшей деятельности прикладного характера планируется усовершенствование функции классификации отзывов в информационной системе, например путем автоматического удаления отрицательных отзывов после классификации, и сохранение оставшихся в указанный каталог.

Список источников

1. Белоконев С. Ю., Крохина В. О., Дронов А. И. Технологии имиджевого позиционирования компаний табачного и фармацевтического рынков в условиях рыночной конкуренции // Известия Тульского государственного университета. Гуманитарные науки. 2020. № 2. С. 93–101.
2. Гуськова О. В. Репутационный маркетинг как инструмент генерирования, мотивации, популяризации компании в интернет-среде // Инновационное развитие экономики. 2022. № 1–2. С. 138–143.

References

1. Belokonev S. Yu., Crohina V. O., Dronov A. I. Image positioning technologies of the tobacco and pharmaceutical companies in the highly competitive market. *Bulletin of Tula State University. Humanitarian Sciences*. 2020;(2):93–101. (In Russ.).
2. Guskova O. V. Reputation marketing as a tool for generation, motivation, popularization of a company in the internet environment. *Innovative Development of Economy*. 2022;(1–2):138–143. (In Russ.).

3. Скрипко А. А., Фёдорова Н. В., Клименкова А. А. Информационные технологии в фармации. В 4 ч. Ч. 4. Комплексная автоматизация деятельности аптечных организаций. Иркутск : ИГМУ, 2020. 84 с.
4. Сидикова Г. Р. Методы и инструменты сентимент-анализа // Современные проблемы лингвистики и методики преподавания русского языка в ВУЗе и школе. 2022. № 34. С. 974–985.
5. Большаков Н. И., Сидорова Е. В. Сравнительный анализ методов машинного обучения для задач классификации данных // Математические методы в технологиях и технике. 2023. № 8. С. 66–71. DOI 10.52348/2712-8873_MMTT_2023_8_66.
6. Богданова Д. Р. Оценка степени удовлетворенности клиентов сферы услуг на основе учета их эмоционально окрашенной информации // Системная инженерия и информационные технологии. 2021. Т. 3, № 3. С. 72–81.
7. Любченко М. А. Об одном опыте анализа данных и извлечения информации о программном продукте // Системная инженерия и информационные технологии. 2021. Т. 3, № 2. С. 75–80.
8. Farisi A. A., Sibaroni Y., Faraby S. A. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier // Journal of Physics: Conference Series. 2019. Vol. 1192. P. 012024.
9. Кугач В. В., Рылко Я. Н. Информационное наполнение и оформление интернет-сайтов аптечных организаций // Вестник фармации. 2022. № 1. С. 28–41.
10. Федяева И. А. Разработка метода отслеживания зависимостей для кэша вывода типов статического анализатора кода среды разработки Pycharm // XXIV Всерос. студенч. науч.-практич. конф., 05–06 апреля 2022 г., г. Нижневартовск. Ч. 3. Нижневартовск : Нижневартовский государственный университет, 2022. С. 209–214.
11. Хисамутдинов Д. С., Рыженко И. А., Павлова К. А. Автоматическая классификация документов // Научный альманах Центрального Черноземья. 2022. № 1–7. С. 143–157.
12. Томашевская В. С., Старичкова Ю. В., Яковлев Д. А. Использование машинного обучения для распознавания текстовых шаблонов литературных источников // Известия высших учебных заведений. Поволжский регион. Технические науки. 2022. № 3. С. 15–26. DOI 10.21685/2072-3059-2022-3-2.
13. Дрянкова Д. А. Искусственный интеллект в языке программирования Python // Modern Science. 2023. № 6–2. С. 15–19.
14. Болтовский Г. А. Создание парсера на языке Python с использованием библиотеки BeautifulSoup // Постулат. 2022. № 6. С. 1–6.
15. Максимова В. П., Черемных Е. О. Зависимость скорости чтения текста от формата носителя // Инновационное развитие регионов: потенциал науки и современного образования : материалы II Национал. науч.-практич. конф., 07 февраля 2019 г.,
3. Skripko A. A., Fyodorova N. V., Klimenkova A. A. Informationsnnye tekhnologii v farmatsii. In 4 parts. Part 4. Kompleksnaia avtomatizatsiia deiatelnosti aptechnykh organizatsii. Irkutsk: ISMU; 2020. 84 p. (In Russ.).
4. Sidikova G. R. Metody i instrumenty sentiment-analiza. *Sovremennye problemy lingvistiki i metodiki prepodavaniia russkogo iazyka v VUZe i shkole*. 2022;(34):974–985. (In Russ.).
5. Bolshakov N. I., Sidorova E. V. Comparative analysis of machine learning methods for problems of data classification. *Mathematical Methods in Technics and Technologies*. 2023;(8):66–71. DOI 10.52348/2712-8873_MMTT_2023_8_66. (In Russ.).
6. Bogdanova D. R. Assessment of the level of customer satisfaction in the service sphere based on their emotionally colored information. *System Engineering and Information Technologies*. 2021;3(3):72–81. (In Russ.).
7. Lyubchenko M. A. One experience in data analysis and extraction of information about the software product. *Systems Engineering and Information Technologies*. 2021;3(2):75–80. (In Russ.).
8. Farisi A. A., Sibaroni Y., Faraby S. A. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*. 2019;1192:012024.
9. Kugach V. V., Rylko Ya. N. Information content and design of Internet sites of pharmacy organizations. *Vestnik farmatsii*. 2022;(1):28–41. (In Russ.).
10. Fedyaeva I. A. Razrabotka metoda otslezhivaniia zavisimostei dlia kesha vyvoda tipov staticheskogo analizatora koda sredy razrabotki Pycharm. In: *Proceedings of the 24th All-Russian Student Research-to-Practice Conference*, April 5–6, 2022, Nizhnevartovsk. Pt. 3. Nizhnevartovsk: Nizhnevartovsk State University; 2022. p. 209–214. (In Russ.).
11. Khisamutdinov D. S., Ryzhenko I. A., Pavlova K. A. Automatic document. *Nauchnyi almanakh Tsentralno-go Chernozemia*. 2022;(1–7):143–157. (In Russ.).
12. Tomashevskaya V. S., Starichkova Yu. V., Yakovlev D. A. Using machine learning for recognition of text patterns of literary sources. *University Proceedings. Volga Region. Engineering Sciences*. 2022;(3):15–26. DOI 10.21685/2072-3059-2022-3-2. (In Russ.).
13. Dryankova D. A. Iskusstvennyi intellekt v iazyke programmirovaniia Python. *Modern Science*. 2023;(6–2):15–19. (In Russ.).
14. Boltovskiy G. A. Creating a Python parser using the BeautifulSoup library. *Postulat*. 2022;(6):1–6. (In Russ.).
15. Maksimova V. P., Cheremnykh E. O. Zavisimost skorosti chteniia teksta ot formata nositelia. In: *Proceedings of the 2nd National Research-to-Practice Conference “Innovatsionnoe razvitie regionov: potentsial nauki i sovremennogo obrazovaniia”*, February 7, 2019, As-

- г. Астрахань. Астрахань : Информационно-издательский центр, 2019. С. 176–178.
16. Щукарев И. А., Маркова Е. В. Разработка генератора паролей с использованием GUI MATLAB // Программные продукты и системы. 2022. № 3. С. 413–419.

Информация об авторах

К. В. Святлов – кандидат технических наук, доцент.

В. С. Мошкин – кандидат технических наук, доцент.

И. А. Щукарев – кандидат физико-математических наук.

trakhan. Astrakhan: Information Publishing Center; 2019. p. 176–178. (In Russ.).

16. Shchukarev I. A., Markova E. V. Developing a password generator using GUI MATLAB. *Software and Systems*. 2022;(3):413–419. (In Russ.).

Information about the authors

K. V. Svyatov – Candidate of Sciences (Engineering), Do-cent.

V. S. Moshkin – Candidate of Sciences (Engineering), Do-cent.

I. A. Shchukarev – Candidate of Sciences (Physics and Mathematics).