

# Моделирование вычислительных, телекоммуникационных, управляющих и социально-экономических систем

## Simulations of Computer, Telecommunications, Control and Social Systems

Научная статья

DOI: <https://doi.org/10.18721/JCSTCS.15204>

УДК 004.912, 004.85, 004.41



### РАЗРАБОТКА И ИССЛЕДОВАНИЕ МОДЕЛЕЙ МНОГОКЛАССОВЫХ КЛАССИФИКАТОРОВ ДЛЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ПОДГОТОВКИ ЗАЯВОК НА ПОРТАЛЕ ЕДИНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ В СФЕРЕ ЗАКУПОК

*Я.А. Селиверстов<sup>1</sup> ✉, А.А. Комиссаров<sup>2</sup>, А.А. Лесоводская<sup>3</sup>,  
П.Г. Бобыкин<sup>4</sup>, А.В. Подтихов<sup>5</sup>, С.С. Торсионов<sup>6</sup>,  
Д.А. Цирков<sup>7</sup>, С.А. Орлов<sup>8</sup>*

<sup>1</sup> Институт проблем транспорта имени Н.С. Соломенко РАН,  
Санкт-Петербург, Российская Федерация;

<sup>1,2,3,4,5,6,7</sup> Университет национальной технологической инициативы 2035,  
Москва, Российская Федерация;

<sup>8</sup> Томский государственный университет,  
г. Томск, Российская Федерация

✉ [Y.Seliverstov@2035.university](mailto:Y.Seliverstov@2035.university)

**Аннотация.** Обоснована актуальность разработки сервисов, способствующих подготовке тендерной документации, в части определения кода ОКПД 2 к формируемой заявке. Для решения задачи автоматической классификации заявок в соответствии с ОКПД 2 разработан алгоритм системы сравнительного анализа моделей классификаторов, осуществлена предобработка и запись в базу данных собранной информации в формате json. Разметка и подготовка данных для обучения моделей классификаторов осуществлена в среде PolyAnalyst. В качестве моделей многоклассовых классификаторов из библиотеки Scikit-Learn выбраны наивный байесовский классификатор, SVM-классификатор и классификатор на основе случайного леса. В качестве векторизаторов выбрана модель TFIDF и word-hashing. В качестве четвертого классификатора выбрана нейросетевая модель ruBert-base. Проведено обучение классификаторов и оценено качество их работы. По результатам валидации и тестирования лучшими оказались две модели: ruBert-base и модель наивного байесовского классификатора с векторизатором word-hashing. На основе результатов произведена тестовая классификация заявок.

**Ключевые слова:** рекомендательные системы, многоклассовая классификация, SVM, naive Bayes, ruBert-base, векторизаторы

**Для цитирования:** Селиверстов Я.А., Комиссаров А.А., Лесоводская А.А. и др. Разработка и исследование моделей многоклассовых классификаторов для рекомендательной системы подготовки заявок на портале единой информационной системы в сфере закупок // Computing, Telecommunications and Control. 2022. Т. 15, № 2. С. 43–62. DOI: 10.18721/JCSTCS.15204

Research article

DOI: <https://doi.org/10.18721/JCSTCS.15204>

UDC 004.912, 004.85, 004.41



## DEVELOPMENT AND RESEARCH OF MODELS OF MULTI-CLASS CLASSIFIERS FOR A RECOMMENDED SYSTEM FOR PREPARING APPLICATIONS ON THE E-PROCUREMENT

*Y.A. Seliverstov<sup>1</sup> ✉, A.A. Komissarov<sup>2</sup>, A.A. Lesovodskay<sup>3</sup>,  
P.G. Bovykin<sup>4</sup>, A.V. Podtikhov<sup>5</sup>, S.S. Torsionov<sup>6</sup>,  
D.A. Tsyrvkov<sup>7</sup>, S.A. Orlov<sup>8</sup>*

<sup>1</sup> IPT RAS, St. Petersburg, Russian Federation;  
<sup>1,2,3,4,5,6,7</sup> University National Technology Initiative 2035,  
Moscow, Russian Federation;

<sup>8</sup> National Research Tomsk State University,  
Tomsk, Russian Federation

✉ Y.Seliverstov@2035.university

**Abstract.** As a result of the analysis, the relevance of developing services that contribute to the preparation of tender documentation, in terms of determining the OKPD 2 code for the generated application, is indicated. To solve the problem of automatic classification of applications in accordance with OKPD 2, an algorithm for the system of comparative analysis of classifier models was developed. Further, preprocessing was carried out, and the collected information was written to the database in json format. Labeling and preparation of data for training classifier models was carried out in the PolyAnalyst environment. As a result of the analysis, a naive Bayes classifier, an SVM classifier, and a random forest classifier were selected as models of multiclass classifiers from the Scikit-Learn library. The TFIDF and word-hashing models were chosen as vectorizers. The ruBert-base neural network model was chosen as the fourth classifier. Classifiers were trained and the quality of their work was assessed. According to the results of validation and testing, two models turned out to be the best: ruBert-base and a model of a naive Bayes classifier with a word-hashing vectorizer. Based on the results, a test classification of applications was made.

**Keywords:** recommendation systems, multiclass classification, SVM, naive Bayes, ruBert-base, vectorizer

**Citation:** Seliverstov Y.A., Komissarov A.A., Lesovodskay A.A., et al. Development and research of models of multi-class classifiers for a recommended system for preparing applications on the e-procurement. *Computing, Telecommunications and Control*, 2022, Vol. 15, No. 2, Pp. 43–62. DOI: 10.18721/JCSTCS.15204

### Введение

Развитие современных информационных систем и технологий способствуют быстрому росту цифровизации во всех сферах государственного управления, в том числе и государственных закупок. Основную роль в этом процессе играют цифровые платформы, поскольку именно они формируют среду электронного взаимодействия потенциальных заказчиков и поставщиков, являясь провайдером сферы электронных закупок.

С 2016 года в России введена в эксплуатацию Единая информационная система (ЕИС) в сфере закупок [1]. ЕИС создана в целях информационного обеспечения контрактной системы в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд, закупок товаров, работ, услуг отдельными видами юридических лиц.

В настоящее время процедуры проведения тендеров занимают ключевое место в работе отдела закупок компании. Большинство современных рекомендательных сервисов и систем [2–4] функ-

ционируют на веб-площадках. К их числу относятся рекомендательные сервисы проведения тендеров [5], выбора тендеров с вероятностью оценки выигрыша участников тендера [6], проверки корректности тендерной документации [7] и др.

Одним из важных этапов формирования электронной тендерной документации является процедура формирования заявки и выбора соответствующего кодификатора из перечня ОКПД 2 – Общероссийского классификатора продукции по видам экономической деятельности.

Стремительный рост различных секторов производства и расширение многообразия создаваемых товаров, продуктов и услуг, а также репрофилирование предпринимательской деятельности самостоятельного субъекта усложняют задачу точной идентификации продуктовых границ отраслевого рынка в рамках Общероссийских классификаторов видов предпринимательской деятельности (ОКПД 2) [8].

По данным [9], ОКПД 2 содержит около 38 тысяч позиций. Столь большая величина затрудняет поиск и увеличивает временные затраты на его точное определение.

Указание неправильного кода ОКПД 2, его незнание или несоответствие реальной экономической деятельности коду ОКПД 2, указанному субъектом при регистрации, влечет множество нежелательных последствий, таких как определение ставки взносов при возникновении несчастных случаев; оформление патентов; потеря налоговых льгот, предоставляемых государством по некоторым видам экономической деятельности; блокирование платежей при внутренней банковской проверке; наложение высоких штрафных санкций при внесении ошибочного кода ОКПД 2 в Единую информационную систему в сфере закупок [10, 11].

Последнее подчеркивает актуальность разработки и внедрения рекомендательных сервисов проверки корректности заявки [12] в части выбора ОКПД 2 в системы электронного сопровождения корпоративного тендерного документооборота [13].

**Анализ предметной области.** В настоящее время рекомендательным системам посвящено множество работ российских и зарубежных ученых.

Теоретико-методологические особенности рекомендательных систем, а также различные характеристики и возможности методов прогнозирования в современных рекомендательных системах рассмотрены в [14, 15].

Подход совместной фильтрации, основанный на наивном байесовском классификаторе, изучен в [16]. Предложенная в [16] байесовская модель не только предоставляет рекомендации, но и дает объяснения этим прогнозам.

В [17] описана система, использующая кластеризацию и случайный лес в качестве многоуровневых стратегий для прогнозирования рекомендаций на основе оценок пользователей, ориентируясь при этом на мышление пользователей и текущие тенденции.

Решая предсказательную задачу моделирования динамических предпочтений пользователей на основе их исторического поведения, в [18] представлена глубокая двунаправленная последовательная модель под названием BERT4Rec для последовательной рекомендации. При обучении модели применяется задача Cloze, которая предсказывает элементы, используя как левый, так и правый контекст.

Влияние систем поиска и рекомендаций на продажи в электронной торговле изучено в [3]. В статье показано, как разные классы инструментов поиска и рекомендаций влияют на распределение продаж по продуктам, общий объем продаж и излишек потребителей.

В [19] авторы рассматривают тематическую диверсификацию, новый метод, разработанный для баланса и диверсификации персонализированных списков рекомендаций, чтобы отразить полный спектр интересов пользователя. Согласно [19] алгоритмы диверсификации использует Amazon для улучшения своих рекомендаций.

В [20] решается задача многоклассовой классификации слабotoксичного образовательного контента онлайн школ с использованием нейросетевой модели RoBERT.

В [21], опираясь на информационно-методическое обеспечение [22] на основе нейронной сети трансформера, разрабатывается метод многоклассовой классификации угроз.

В [23] и [24] рассмотрены системы на основе моделей классификации транспортных данных в части определения состояния качества дорог. В качестве моделей взяты классификаторы SVM и наивного байеса.

В [25] изучены методы машины опорных векторов и латентно-семантический анализ с различными векторизаторами для построения модели многоклассовой классификации слабоструктурированных текстовых документов.

В [26] выполнен подробный обзор подходов интеллектуального анализа текстов на русском языке.

Проведенный анализ свидетельствует о том, что модели классификаторов на основе наивного байесовского алгоритма, машины опорных векторов, случайного леса и нейросетевые алгоритмы, такие как BERT, широко используются в различных информационных системах, связанных с фильтрацией и классификацией веб-контента.

**Постановка задачи.** Цель настоящей работы – в исследовании современных программных методов классификации для определения наиболее точной модели классификации применительно к задаче классификации заявки тендерной документации к соответствующему классификатору из ОКПД 2. Данную модель классификации в последующем предполагается использовать в разработке рекомендательного сервиса, помогающего специалисту в сфере закупок по информации, содержащейся в подготавливаемой к тендеру заявке, быстро и точно определить соответствующий ей код первого уровня ОКПД 2.

Сервис планируется использовать совместно с Единой информационной системой в сфере закупок (портал <https://zakupki.gov.ru/>).

В рамках исследования планируется разработать краулер для сбора данных, осуществить сбор данных с портала ЕИС, произвести подготовку и разметку данных с использованием программных методов, построить и осуществить сравнительный анализ четырех моделей мультиклассовой классификации: наивного байесовского классификатора (Naive Bayes), классификатора на основе алгоритма случайного леса (Random Forest), классификатора на основе машины опорных векторов (SVM) и нейросетевой модели ruBERT (Bidirectional Encoder Representations from Transformers) [27]. Общее количество классов – 88. В качестве векторизаторов планируется рассмотреть два метода – TFIDF и word hashing [28].

### Основная часть

Построение плана исследования целесообразно начать с разработки обобщенного алгоритма системы сравнительного анализа моделей классификаторов. Алгоритм в общем виде состоит из процедур, представленных на рис. 1.

На первом этапе осуществляется сбора данных заявок. Для этих целей разработан краулер на языке Python3 с использованием библиотеки Requests. С помощью разработанного краулера осуществлен сбор заявок закупок с портала <https://zakupki.gov.ru> (рис. 2).

Информация о закупках доступна по гостевому доступу (рис. 3) и хранится на FTP-сервере (<ftp://ftp.zakupki.gov.ru>) в виде дампов в формате xml в директории fcs\_regions (рис. 4).

Сбор данных осуществляется по следующему алгоритму: 1) краулер заходит на ftp-сервер под гостевым доступом в директорию fcs\_regions; 2) проходит по регионам и загружает RETR-запросом по очереди каждый xml-файл закупок в память; 3) «на лету» из памяти осуществляет парсинг метаданных, содержащихся в xml-файле, преобразуя их в json в соответствии с заданной структурой; 4) записывает json в базу данных; 5) удаляет xml-файл из памяти; 6) переходит к следующему по очереди xml-файлу закупок.

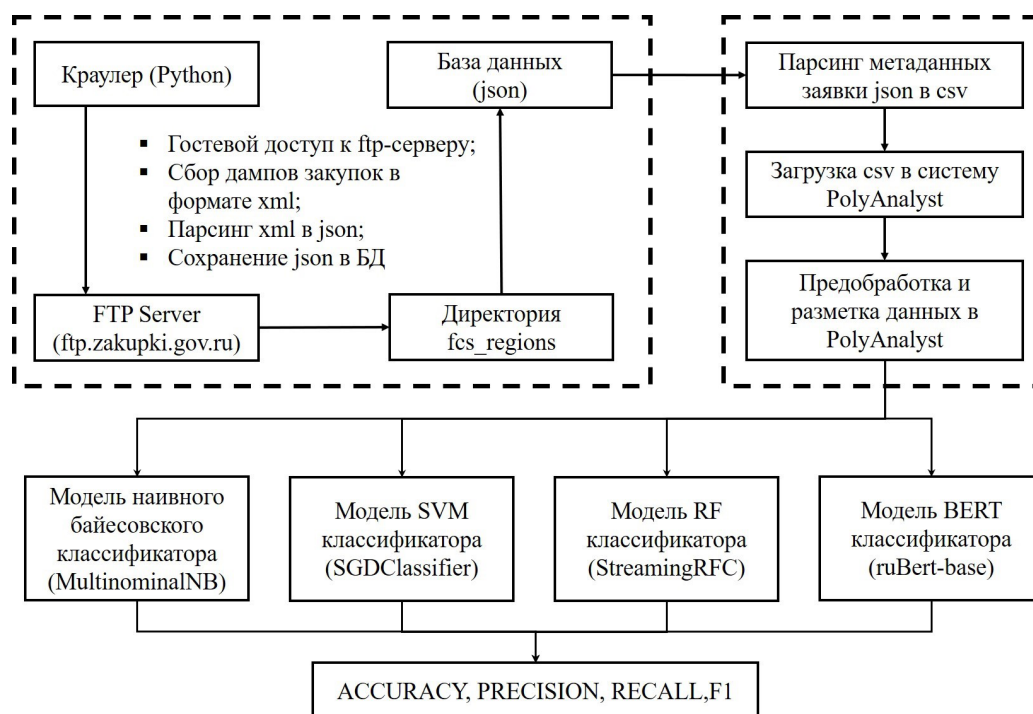


Рис. 1. Схема обобщенного алгоритма системы сравнительного анализа моделей классификаторов

Fig. 1. Scheme of the algorithm of the system for comparing classifier models

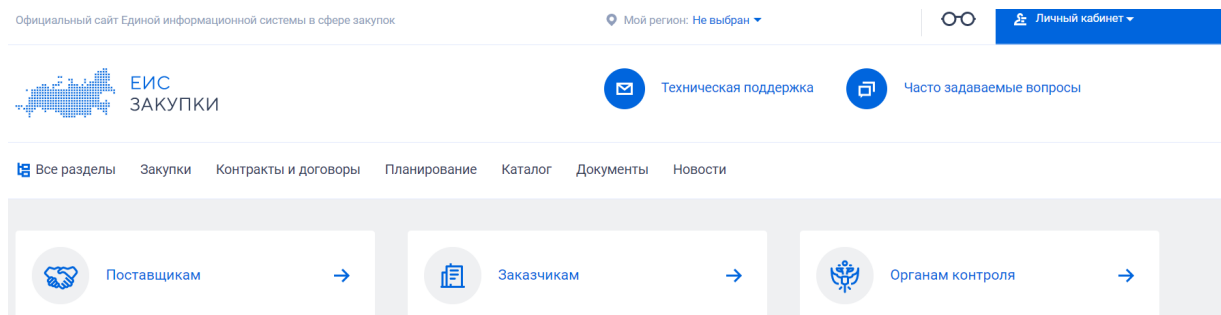


Рис. 2. Портал zakupki.gov.ru

Fig. 2. Portal zakupki.gov.ru

Цикл повторяется до тех пор, пока все метаданные заявок закупок не будут загружены в базу данных.

Структура json-файла закупок в базе данных представлена на рис. 5.

На втором этапе осуществляется парсинг лотов заявок в базе данных и перевод json в csv-формат с использованием скрипта, написанного на Python3, и загрузка csv в систему PolyAnalyst. В результате парсинга json формируется следующая структура csv файла метаданных закупок: id – номер заявки на ftp-сервере портала zakupki.gov.ru; number – реестровый номер извещения; reg – место подведения итогов; publish data – дата публикации заявки на портале ЕИС; purchaseResponsible – наименование организации заказчика; maxPrise – максимальная цена контракта закупки; purchaseObjectInfo – информация о закупках; name\_purchaseObject – наимено-



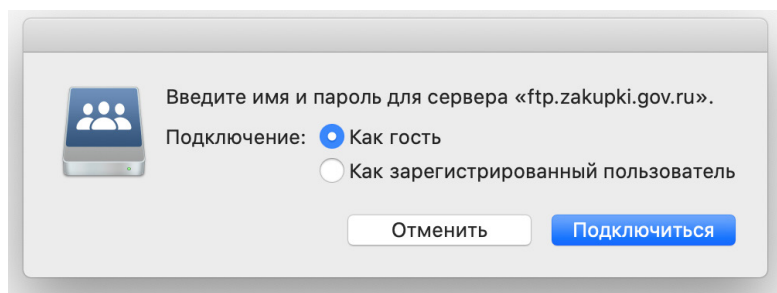


Рис. 3. Окно гостевого доступа к серверу ftp://ftp.zakupki.gov.ru  
Fig. 3. Guest access window to the ftp://ftp.zakupki.gov.ru server

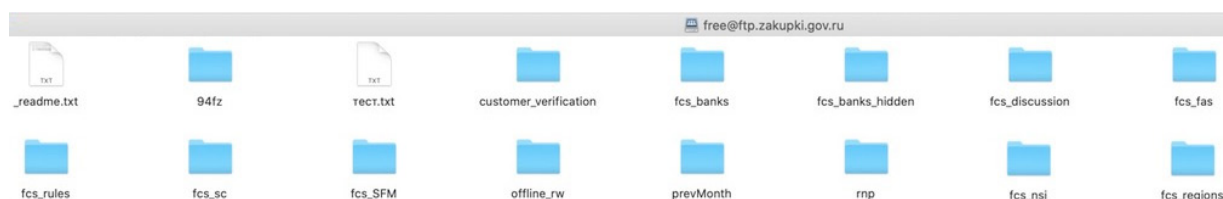


Рис. 4. Директории ЕИС на сервере ftp://ftp.zakupki.gov.ru  
Fig. 4. Data directories on the ftp server ftp://ftp.zakupki.gov.ru

вание объекта закупок; ОКПД2\_codes – код ОКПД2; ОКПД2\_names – наименование ОКПД2; ОКПД\_codes – код ОКПД; ОКПД\_names – наименование ОКПД; KRTU\_codes – код КРТУ; KRTU\_names – наименование КРТУ.

Структура csv файла закупок zakupki\_all.csv для загрузки в систему PolyAnalys для последующей обработки и разметки (рис. 6).

На третьем этапе в среде Polyanalyst [29] выполняется обработка и разметка данных. Графическое представление вычислительного скрипта представлено на рис. 7.

Вычислительный скрипт содержит следующие узлы:

1) узел производные колонки sum text – в данном узле формируется новое поле sum text путем объединения информации из смежных пяти полей:

```
sum text = [purchaseObjectInfo] + "\n" + [name_purchaseObject] + "\n" + [ОКПД2_names] + "\n" + [ОКПД_names] + "\n" + [KTRU_names];
```

2) узел модификации колонок – в данном узле идет переименование поля sum text в поле description;

3) узел CRC – в данном узле каждой уникальной строке присваивается уникальный номер;

4) узел обработки уникальных описаний – в данном узле по номерам CRC удаляются дубликаты записей, с помощью фильтра русского языка данные отфильтровываются от любых других иностранных слов, удаляются орфографические ошибки и «ё» заменяется на «е»;

5) узел объединения – в данном узле формируется исправленный текст с присвоением ему CRC;

6) узел фильтрации строк – в данном узле удаляются пустые поля description, т. е. те поля, в которых нет описания заявки;

7) узел zakupki\_clean.csv – исправленный и размеченный классификаторами ОКПД и ОКПД2 текст сохраняется в csv.

На четвертом этапе осуществляется разработка и сравнение моделей классификаторов на языке Python 3. Количество классов ОКПД 2 – 88, т. к. классификация осуществляется только

```

{
  "id": 123,
  "lot": {
    "currency": {
      "code": "RUB",
      "name": "Российский рубль"
    },
    "maxPrice": 245099.1,
    "restrictInfo": "не установлено",
    "financeSource": "средства ТФ ОМС Курганской области",
    "purchaseObjects": {
      "totalSum": 245099.1,
      "purchaseObject": {
        "sum": 245099.1,
        "OKEI": {
          "code": 236,
          "nationalCode": "КАЛ/Ч"
        },
        "OKPD": {
          "code": "40.30.10.111",
          "name": "Тепловая энергия, отпущенная тепловыми электростанциями"
        },
        "name": "Теплоснабжение",
        "price": 1287.15,
        "quantity": {
          "value": 190.42
        }
      }
    },
    "quantityUndefined": false,
    "customerRequirements": {
      "customerRequirement": {
        "customer": {
          "regNum": 243100024
        },
        "maxPrice": 245099.1,
        "deliveryTerm": "с 01.01.2014г. по 31.12.2014г.",
        "deliveryPlace": "Российская Федерация, 640018, Курганская область, Курган г, Советская, 81, - "
      }
    },
    "href": "http://zakupki.gov.ru//epz/order/notice/ep44/view/common-info.html?noticeId=123",
    "printForm": {
      "url": "http://zakupki.gov.ru//epz/order/notice/printForm/view.html?noticeId=123"
    },
    "placingWay": {
      "code": "EP44",
      "name": "Закупка у единственного поставщика (подрядчика, исполнителя)"
    },
    "attachments": {
      "attachment": {
        "url": "http://zakupki.gov.ru//44fz/filestore/public/1.0/download/priz/file.html?uid=EF3F84CF641E00A0E043AC110725D56E",
        "fileName": "обоснование.odt",
        "docDescription": "обоснование"
      }
    },
    "docPublishDate": "2014-01-09T09:48:28.348+04:00",
    "purchaseNumber": 24310002414000420,
    "purchaseObjectInfo": "Теплоснабжение здания ТФ ОМС Курганской области советская, 81",
    "purchaseResponsible": {
      "responsibleOrg": {
        "regNum": 243100024,
        "fullName": "Территориальный фонд обязательного медицинского страхования Курганской области",
        "factAddress": "Российская Федерация, 640018, Курганская обл, Курган г, Советская, д.81, -",
        "postAddress": "Российская Федерация, 640018, Курганская обл, Курган г, Советская, д.81, -"
      },
      "responsibleInfo": {
        "contactFax": "7-3522-463142",
        "contactEmail": "ur@ktfoms.orbitel.ru",
        "contactPhone": "7-3522-413637",
        "contactPerson": {
          "lastName": "Белов ",
          "firstName": "Николай",
          "middleName": "Григорьевич"
        },
        "orgFactAddress": "Российская Федерация, 640018, Курганская обл, Курган г, Советская, д.81, -",
        "orgPostAddress": "Российская Федерация, 640018, Курганская обл, Курган г, Советская, д.81, -"
      },
      "responsibleRole": "CU"
    }
  }
}

```

Рис. 5. Структура метаданных заявки в базе данных

Fig. 5. Json file of purchases in the database

по первому уровню. Предобработка табличных размеченных данных `zakupki_clean.csv` осуществляется с использованием библиотеки `Pandas`, лемматизация текста – с использованием библиотеки `rumystem3`, сериализация и десериализация объектов (запись, чтение и загрузка дампов) – с использованием модуля `pickle`, фильтрация стоп-слов с помощью библиотеки `nlTK`, очистка текста от лишних пробелов и знаков препинания осуществляется с помощью регулярных выражений. Модели многоклассовых классификаторов `MultinomialNB` (наивный байесовский классификатор), `SGDClassifier` (машина опорных векторов), `StreamingRFC` (случайный лес), а также `TFIDF` векторизатор импортируются из библиотеки `sklearn`. Векторизатор `word hashing` создается на основе триграмм с использованием функции `permutations` из библиотеки `itertools`. Модель нейросетевого классификатора на основе `ruBert-base` от Сбербанка [29] с токенизато-

#	id	number	reg	publish_date	purchaseResponsible	maxPrice	purchaseObjectInfo	name_purchaseObject	OKPD2...	OKPD2_names	OKPD_cod...	OKPD_nam...	KTRU_cod...	KTRU_names	
4481	25,912,894	00,000,000,000	Республика Карелия	3/30/2021	УПРАВЛЕНИЕ ФЕДЕРАЛЬНОЙ	104,552.20	Поставка картриджа для	Картридж Q661A для принт			28.23.25.000	28.23.25.000	Части и принадлежности	32.50.13.120-00327	Линза интравитреальной
4481	25,912,902	00,000,000,000	Республика Карелия	3/12/2021	ГОСУДАРСТВЕННОЕ БЮДЖЕ	941,078.00	Поставка линз интравитре	Линза интравитреальная для з			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	25,912,957	00,000,000,000	Республика Карелия	3/12/2021	МУНИЦИПАЛЬНОЕ КАЗЕНН	1,335,081.60	Приобретение благоустро	однокомнатная благоустро			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	25,913,087	00,000,000,000	Республика Карелия	3/12/2021	ГОСУДАРСТВЕННОЕ КАЗЕНН	34,077.00	Поставка хозяйственных т	Универсальные салфетки в г			13.92.29.190	13.92.29.190	И изделия текстильн		
4481	25,913,095	00,000,000,000	Республика Карелия	3/12/2021	ГОСУДАРСТВЕННОЕ КАЗЕНН	612,040.00	поставка автомобильног	Бензин автомобильный АИ-9			68.10.11.000	68.10.11.000	Услуги по покупке и	19.20.21.125-00001	Бензин автомобиль
4481	25,913,166	00,000,000,000	Республика Карелия	3/12/2021	МУНИЦИПАЛЬНОЕ КАЗЕНН	1,339,502.40	Приобретение благоустро	однокомнатная благоустро			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	25,913,213	00,000,000,000	Республика Карелия	3/12/2021	ГОСУДАРСТВЕННОЕ КАЗЕНН	1,880,000.00	поставка светового и зву	Петличный микрофон Студия			26.40.41.000	26.40.41.000	Микрофоны и подст	26.40.31.190-00000	Цифровой микрофон
4481	25,913,244	00,000,000,000	Республика Карелия	3/12/2021	ГОСУДАРСТВЕННОЕ БЮДЖЕ	1,511,973.60	Поставка экспресс-тестов	Набор реагентов для имму			21.20.23.111	21.20.23.111	Препараты диагнос		
4481	25,913,252	00,000,000,000	Республика Карелия	3/12/2021	МУНИЦИПАЛЬНОЕ КАЗЕНН	1,719,691.20	Приобретение благоустро	двухкомнатная благоустро			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	25,913,350	00,000,000,000	Республика Карелия	3/12/2021	МУНИЦИПАЛЬНОЕ КАЗЕНН	1,976,097.60	Приобретение благоустро	двухкомнатная благоустро			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	25,913,425	00,000,000,000	Республика Карелия	3/12/2021	МУНИЦИПАЛЬНОЕ КАЗЕНН	1,989,360.00	Приобретение благоустро	двухкомнатная благоустро			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	25,913,491	00,000,000,000	Республика Карелия	3/12/2021	МУНИЦИПАЛЬНОЕ КАЗЕНН	2,528,697.60	Приобретение благоустро	треккомнатная благоустро			68.10.11.000	68.10.11.000	Услуги по покупке и		
4481	1,407,288	00,000,000,000	Орловская област	7/1/2014	Администрация Орловского	1,133,543.48	Приобретение жилого пом	Благоустроенное жилое пом	70.12.11.000	Услуги по покупке и п					

Рис. 6. Структура csv файла закупок (zakupki\_all.csv) в PolyAnalyst  
Fig. 6. csv-file (zakupki all.csv) in PolyAnalyst

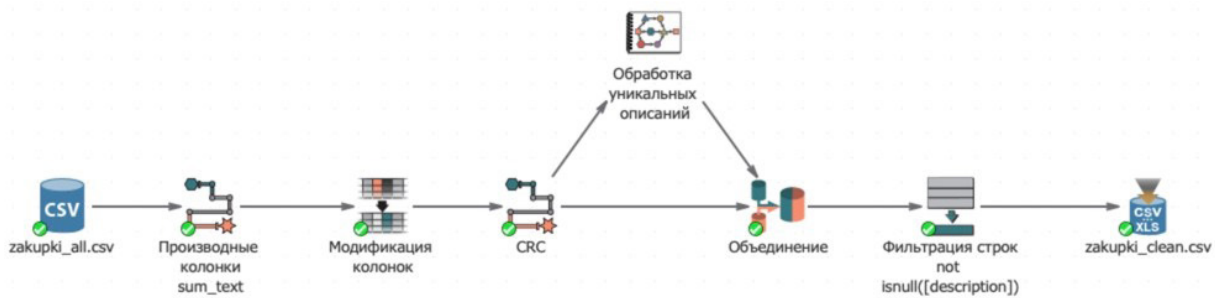


Рис. 7. Вычислительный скрипт обработки данных в PolyAnalyst  
Fig. 7. Data processing script in PolyAnalyst

classifier nb tfidf	
Train [1 / 5]	Accuracy = 54.99%, Precision = 46.71%, Recall = 41.61%, F1 = 39.64%: 100% ██████████ 459/459 [00:30<00:00, 15.00it/s]
Val [1 / 5]	Accuracy = 42.21%, Precision = 40.54%, Recall = 35.26%, F1 = 33.24%: 100% ██████████ 51/51 [00:01<00:00, 38.40it/s]
Train [2 / 5]	Accuracy = 54.34%, Precision = 47.06%, Recall = 48.03%, F1 = 42.33%: 100% ██████████ 459/459 [00:30<00:00, 15.22it/s]
Val [2 / 5]	Accuracy = 40.58%, Precision = 40.53%, Recall = 35.63%, F1 = 32.76%: 100% ██████████ 51/51 [00:01<00:00, 39.04it/s]
Train [3 / 5]	Accuracy = 53.68%, Precision = 46.12%, Recall = 49.75%, F1 = 42.08%: 100% ██████████ 459/459 [00:30<00:00, 15.26it/s]
Val [3 / 5]	Accuracy = 39.94%, Precision = 40.07%, Recall = 35.77%, F1 = 32.42%: 100% ██████████ 51/51 [00:01<00:00, 39.16it/s]
Train [4 / 5]	Accuracy = 53.28%, Precision = 45.68%, Recall = 50.49%, F1 = 41.75%: 100% ██████████ 459/459 [00:30<00:00, 15.28it/s]
Val [4 / 5]	Accuracy = 39.63%, Precision = 40.11%, Recall = 35.95%, F1 = 32.25%: 100% ██████████ 51/51 [00:01<00:00, 38.87it/s]
Train [5 / 5]	Accuracy = 53.03%, Precision = 45.45%, Recall = 50.88%, F1 = 41.55%: 100% ██████████ 459/459 [00:29<00:00, 15.34it/s]
Val [5 / 5]	Accuracy = 39.53%, Precision = 39.80%, Recall = 35.99%, F1 = 32.00%: 100% ██████████ 51/51 [00:01<00:00, 38.83it/s]

Рис. 8. Характеристики модели MultinomialNB с TFIDF  
Fig. 8. Characteristics of the MultinomialNB model with TFIDF





classifier svm hashing	
Train [1 / 5]	Accuracy = 34.95%, Precision = 23.10%, Recall = 14.24%, F1 = 13.95%: 100% ██████████ 459/459 [01:55<00:00, 3.97it/s]
Val [1 / 5]	Accuracy = 31.21%, Precision = 22.82%, Recall = 12.13%, F1 = 12.02%: 100% ██████████ 51/51 [00:03<00:00, 15.82it/s]
<b>Train [2 / 5]</b>	<b>Accuracy = 40.38%, Precision = 26.75%, Recall = 14.49%, F1 = 14.98%: 100% ██████████ 459/459 [01:46&lt;00:00, 4.30it/s]</b>
Val [2 / 5]	Accuracy = 31.37%, Precision = 22.19%, Recall = 11.33%, F1 = 10.92%: 100% ██████████ 51/51 [00:02<00:00, 20.84it/s]
Train [3 / 5]	Accuracy = 39.87%, Precision = 27.21%, Recall = 13.46%, F1 = 14.08%: 100% ██████████ 459/459 [01:49<00:00, 4.21it/s]
<b>Val [3 / 5]</b>	<b>Accuracy = 34.79%, Precision = 22.79%, Recall = 11.75%, F1 = 11.36%: 100% ██████████ 51/51 [00:02&lt;00:00, 20.95it/s]</b>
Train [4 / 5]	Accuracy = 39.02%, Precision = 27.41%, Recall = 12.74%, F1 = 13.44%: 100% ██████████ 459/459 [01:51<00:00, 4.13it/s]
Val [4 / 5]	Accuracy = 30.07%, Precision = 23.37%, Recall = 10.48%, F1 = 10.55%: 100% ██████████ 51/51 [00:02<00:00, 20.85it/s]
Train [5 / 5]	Accuracy = 38.12%, Precision = 27.31%, Recall = 12.17%, F1 = 12.91%: 100% ██████████ 459/459 [01:50<00:00, 4.15it/s]
Val [5 / 5]	Accuracy = 30.44%, Precision = 22.98%, Recall = 10.32%, F1 = 10.31%: 100% ██████████ 51/51 [00:02<00:00, 21.12it/s]

Рис. 11. Характеристики модели SGDClassifier с word hashing  
Fig. 11. Characteristics of the SGDClassifier model with word hashing

classifier rf tfidf	
Train [1 / 5]	Accuracy = 1.09%, Precision = 0.15%, Recall = 1.49%, F1 = 0.19%: 100% ██████████ 459/459 [2:56:29<00:00, 23.07s/it]
Val [1 / 5]	Accuracy = 1.22%, Precision = 0.23%, Recall = 1.48%, F1 = 0.24%: 100% ██████████ 51/51 [38:14<00:00, 44.99s/it]
Train [2 / 5]	Accuracy = 1.08%, Precision = 0.15%, Recall = 1.49%, F1 = 0.19%: 100% ██████████ 459/459 [8:34:57<00:00, 67.31s/it]
Val [2 / 5]	Accuracy = 1.10%, Precision = 0.21%, Recall = 1.45%, F1 = 0.22%: 100% ██████████ 51/51 [1:17:48<00:00, 91.54s/it]
Train [3 / 5]	Accuracy = 1.10%, Precision = 0.18%, Recall = 1.50%, F1 = 0.21%: 100% ██████████ 459/459 [10:34:42<00:00, 43.07s/it]
Val [3 / 5]	Accuracy = 1.23%, Precision = 0.25%, Recall = 1.50%, F1 = 0.26%: 100% ██████████ 51/51 [2:58:33<00:00, 34.99s/it]
Train 4 / 5]	Accuracy = 1.09%, Precision = 0.15%, Recall = 1.49%, F1 = 0.19%: 100% ██████████ 459/459 [14:21:37<00:00, 27.31s/it]
Val [4 / 5]	Accuracy = 1.11%, Precision = 0.24%, Recall = 1.48%, F1 = 0.24%: 100% ██████████ 51/51 [4:32:11<00:00, 61.54s/it]
<b>Train 5 / 5]</b>	<b>Accuracy = 1.11%, Precision = 0.21%, Recall = 1.54%, F1 = 0.23%: 100% ██████████ 459/459 [18:55:12&lt;00:00, 47.31s/it]</b>
<b>Val [5 / 5]</b>	<b>Accuracy = 1.13%, Precision = 0.28%, Recall = 1.51%, F1 = 0.27%: 100% ██████████ 51/51 [6:32:11&lt;00:00, 44.32s/it]</b>

Рис. 12. Характеристики модели StreamingRFC с TFIDF  
Fig. 12. Characteristics of the StreamingRFC model with TFIDF

classifier rf hashing	
Train [1 / 5]	Accuracy = 1.11%, Precision = 0.12%, Recall = 1.47%, F1 = 0.17%: 100% ██████████ 459/459 [4:01:31<00:00, 31.57s/it]
Val [1 / 5]	Accuracy = 1.15%, Precision = 0.19%, Recall = 1.53%, F1 = 0.23%: 100% ██████████ 51/51 [53:37<00:00, 63.09s/it]
Train [2 / 5]	Accuracy = 1.12%, Precision = 0.12%, Recall = 1.48%, F1 = 0.18%: 100% ██████████ 459/459 [12:02:32<00:00, 94.45s/it]
Val [2 / 5]	Accuracy = 1.12%, Precision = 0.18%, Recall = 1.50%, F1 = 0.22%: 100% ██████████ 51/51 [1:48:05<00:00, 127.18s/it]
Train [3 / 5]	Accuracy = 1.12%, Precision = 0.14%, Recall = 1.50%, F1 = 0.21%: 100% ██████████ 459/459 [23:32:26<00:00, 27.44s/it]
Val [3 / 5]	Accuracy = 1.17%, Precision = 0.22%, Recall = 1.64%, F1 = 0.33%: 100% ██████████ 51/51 [2:35:13<00:00, 73.22s/it]
Train [4 / 5]	Accuracy = 1.12%, Precision = 0.12%, Recall = 1.48%, F1 = 0.18%: 100% ██████████ 459/459 [1:16:29:32<00:00, 65.48s/it]
Val [4 / 5]	Accuracy = 1.16%, Precision = 0.24%, Recall = 1.58%, F1 = 0.24%: 100% ██████████ 51/51 [1:48:05<00:00, 127.18s/it]
Train [5 / 5]	<b>Accuracy = 1.14%, Precision = 0.18%, Recall = 1.61%, F1 = 0.31%:</b> 100% ██████████ 459/459 [2:11:55:18<00:00, 72.48s/it]
Val [5 / 5]	<b>Accuracy = 1.18%, Precision = 0.32%, Recall = 1.88%, F1 = 0.30%:</b> 100% ██████████ 51/51 [3:21:43<00:00, 87.28s/it]

Рис. 13. Характеристики модели StreamingRFC с word hashing  
 Fig. 13. Characteristics of the StreamingRFC model with word hashing

ром BertTokenizer импортируется из библиотеки transformers. В качестве метрик качества классификаторов используются Accuracy, Recall, Precision и мера F1. Количество эпох обучения для каждого классификатора – 10.

Результаты обучения моделей классификаторов MultinomialNB, SGDClassifier, StreamingRFC, с векторизаторами TFIDF и word hashing на пяти эпохах с указанными метриками качества представлены на рис. 8–13. Маркером и шрифтом выделены лучшие значения на трейне и валидации.

Лучшие метрики обучения нейросетевой модели классификатора на основе ruBert-base соответствуют седьмой эпохе, они представлены на рис. 14–17.

Лучшей среди четырех представленных выше моделей классификаторов оказалась модель классификатора ruBert-base. В сравнении с моделью классификатора на основе наивного байеса с векторизатором word hashing на валидационной выборке (см. табл. 1), она показала результаты в среднем на 15 % выше.

Таблица 1

**Сравнение MultinomialNB + word hashing с ruBert-base**

Table 1

**Comparison of MultinomialNB + word hashing with ruBert-base**

		Accuracy	Precision	Recall	F1
MultinomialNB с word hashing	Train	0.5588	0.4553	0.4508	0.4244
	Val	0.4662	0.4117	0.4032	0.3817
ruBert-base	Train	1	1	1	1
	Val	0.7141	0.5802	0.5815	0.5760

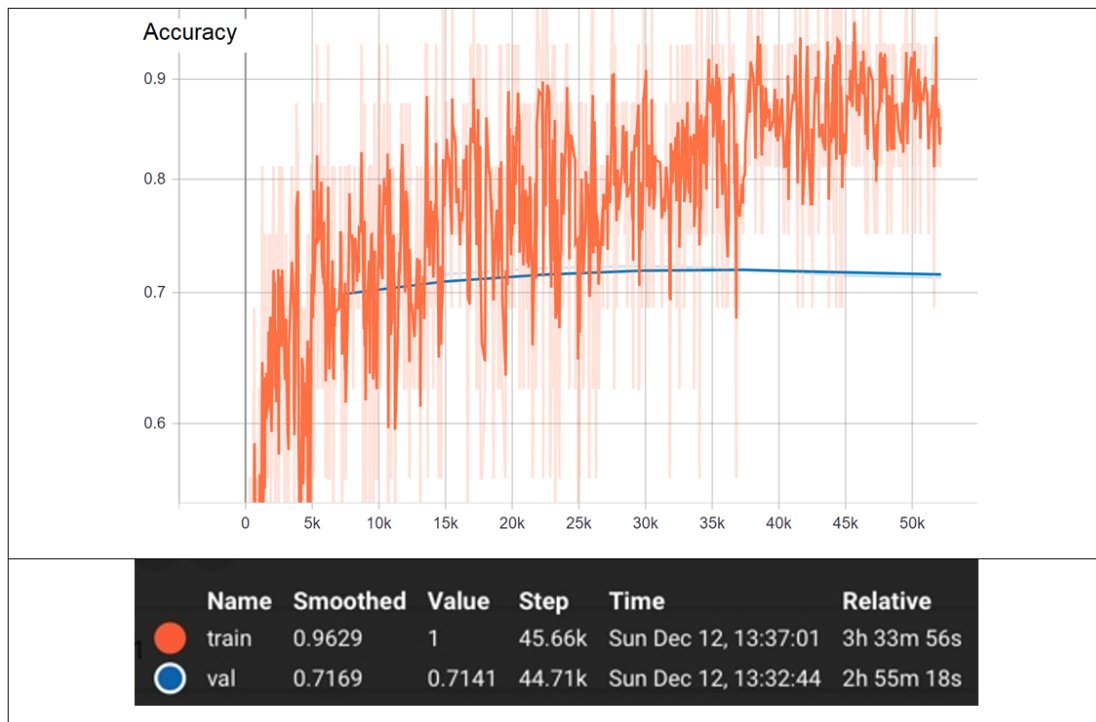


Рис. 14. Метрика Accuracy модели классификатора ruBert-base  
Fig. 14. Accuracy metric of ruBert-base classifier model

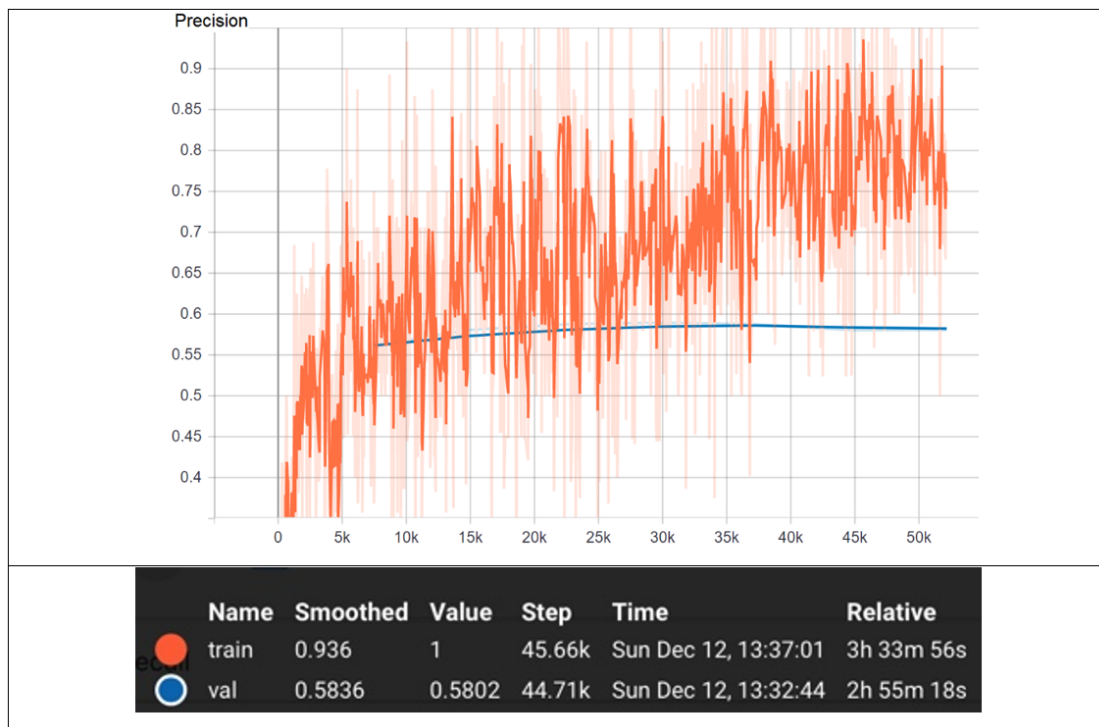


Рис. 15. Метрика Precision модели классификатора ruBert-base  
Fig. 15. The Precision metric of the ruBert-base classifier model

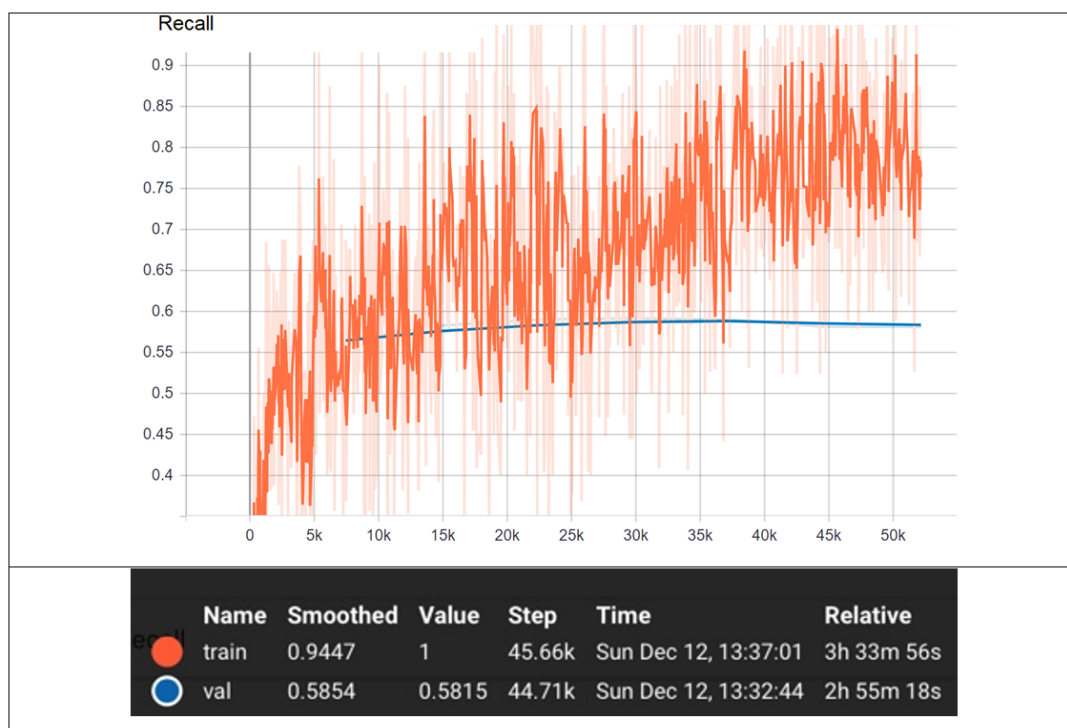


Рис. 16. Метрика Recall модели классификатора ruBert-base  
 Fig. 16. Recall metric of ruBert-base classifier model

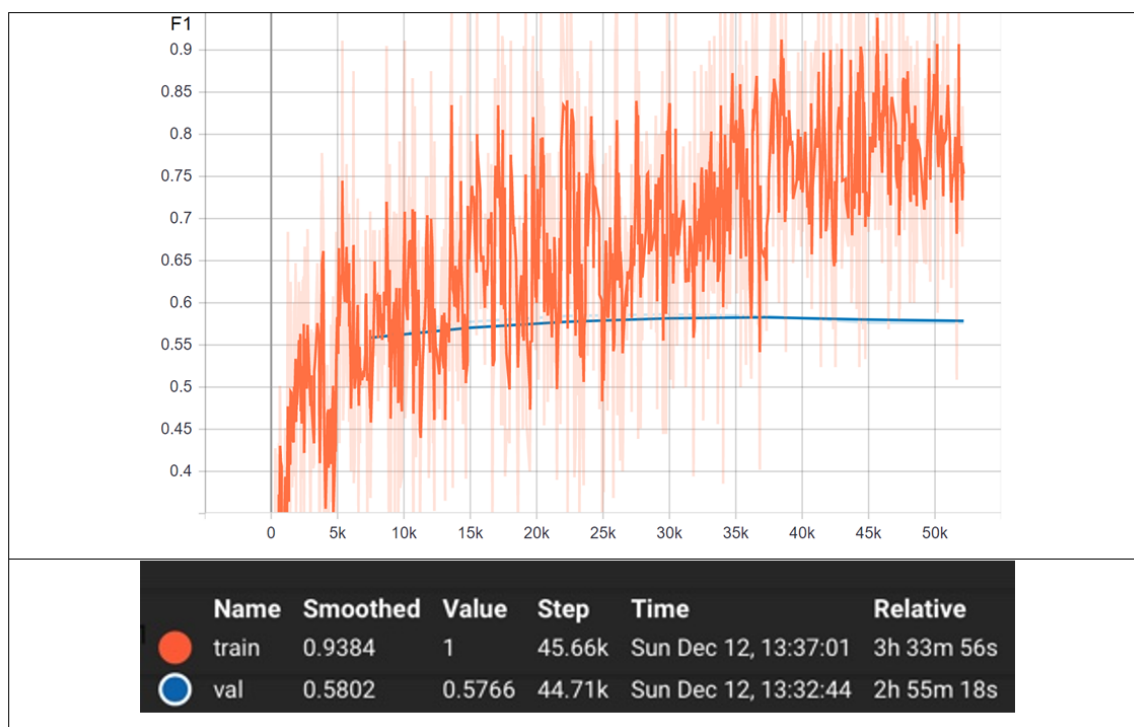


Рис. 17. Метрика F1 модели классификатора ruBert-base  
 Fig. 17. F1 metric of ruBert-base classifier model



Худшие результаты классификации показал алгоритм случайного леса, основные метрики качества не превысили 1 %. Это можно объяснить тем, что при решении задач многоклассовой классификации с большим количеством классов и обучающей выборкой случайный лес использует множество деревьев: каждому дереву в лесу передаются одни и те же входные данные, на основании которых оно должно вернуть свое предсказание. После чего также происходит голосование на полученных прогнозах. Весь этот процесс занимает много времени. Процесс разделения классов повторяется с использованием «жадной», рекурсивной процедуры, пока дерево не достигнет максимальной глубины, или в каждом узле не останутся только образцы одного класса. Выращивание деревьев такой размерности затратная процедура как по времени, так и по вычислительной мощности. Что подтвердилось в ходе обучения: уже на второй эпохе алгоритм случайного леса с векторизатором word hashing затратил на обучение более 12 часов.

Тестирование классификаторов в режиме предсказания класса заявки в соответствии с первым уровнем классификатора ОКПД2 (см. табл. 2) показало, что лучшие результаты продемонстрировали наивный байесовский классификатор и ruBert-base.

Таблица 2

**Результаты тестирования классификаторов в режиме предсказания кода заявки из ОКПД2**

Table 2

**Results of testing classifiers in the order code prediction mode from ОКПД2**

Текст	SVM		MultinomialNB		StreamingRFC		ruBert-base	ОКПД 2
	TFIDF	W-hash	TFIDF	W-hash	TFIDF	W-hash		
Текущий ремонт помещений	27	45	41	43	49	49	43	43
Поставка песка и гравия природного	28	14	8	8	49	62	8	8
Закупка персональных компьютеров	27	14	26	26	49	49	26	26
Закупка медицинских масок	32	13	14	32	49	80	14	32,21, (14)
На оказание услуг автомойки для автомобилей Федерального бюджетного учреждения здравоохранения «Центр гигиены и эпидемиологии в Камчатском крае»	74	71	20	45	62	80	45	45

Таким образом, на основе сравнительного анализа определено, что для рекомендательной системы по предсказанию кодов классификатора ОКПД 2 по тексту заявки целесообразнее использовать модели многоклассовой классификации на основе наивного байесовского алгоритма с векторизатором word hashing и нейросетевую модель ruBert-base.

**Выводы**

Осуществлена разработка и проведено сравнение моделей многоклассовых классификаторов для построения рекомендательной системы, используемой при подготовке заявок на портале ЕИС в сфере закупок <https://zakupki.gov.ru/>. Модели классификаторов обучались на данных, собранных с портала ЕИС.

В ходе сравнительного анализа хорошие результаты продемонстрировали модель наивного байесовского классификатора с векторизатором word hashing и модель ruBert-base. Метрики ка-

чества на модели наивного байесовского классификатора на валидационной выборке составили: accuracy = 0.4662, precision = 0.4117, recall = 0.4032, F1 = 0.3817. Метрики на модели ruBert-base составили: accuracy = 0.7141, precision = 0.5802, recall = 0.5817, F1 = 0.5760. В результате тестирования классификаторов в режиме предсказания кода заявки из ОКПД 2 на наборе из пяти примеров классификатор на основе наивного байеса с векторизатором word hashing отработал без ошибок, классификатор ruBert-base на четвертом примере определил класс неточно. Исходя из результатов сравнительного анализа, данные две модели классификаторов предлагается использовать для построения рекомендательной системы.

В дальнейшем планируется рассмотреть широкий класс нейросетевых моделей трансформерной архитектуры, такие как gpt [30], gpt-2 [31], gpt-3 [32], roBert [33], megatronBert [34] и др. Также планируется реализовать все уровни классификатора ОКПД2.

Подобные подходы позволят расширить функции существующих электронных торговых площадок в части развития широкого класса рекомендательных сервисов [35, 36] для интеллектуального анализа тендерной документации в сфере закупок.

## СПИСОК ЛИТЕРАТУРЫ

1. Мамедова Н.А., Александрова К.С., Позднякова Н.А. Анализ информационного обеспечения контрактной системы на примере официального сайта ЕИС в сфере закупок // Internat. J. of Professional Science. 2018. No. 4. Pp. 24–34.
2. Rashid A.M., Albert I., Cosley D., Lam S.K., McNee S.M., Konstan J.A., et al. Getting to know you: Learning new user preferences in recommender systems // Proc. of the Internat. Conf. on Intelligent User Interfaces. 2002. Pp. 127–34.
3. Hinz O., Eckert J. The impact of search and recommendation systems on sales in electronic commerce // Bus Inf Syst Eng. 2010. no. 2. Pp. 67–77. DOI: 10.1007/s12599-010-0092-x
4. Garcia Rodriguez M.J., Rodriguez Montequin V., Ortega Fernandez F., Villanueva Balsera J.M. Bidders recommender for public procurement auctions using machine learning: Data analysis, algorithm, and case study with tenders from Spain // Complexity. 2020. DOI: 10.1155/2020/8858258
5. Sunmola Funlade, Shehu Yusuf. A case study on performance features of electronic tendering systems // Procedia Manufacturing. 2020. Vol. 51. Pp. 1586–1591. DOI: 10.1016/j.promfg.2020.10.221
6. Mehrbod A., Grilo A. Advanced engineering informatics tender calls search using a procurement product named entity recognizer // Advanced Engineering Informatics. 2018. Vol. 36.
7. Garcia Rodriguez M.J., Rodriguez Montequin V., Ortega Fernandez F., Villanueva Balsera J.M. Public procurement announcements in Spain: Regulations, data analysis, and award price estimator using machine learning // Complexity. 2019.
8. Мигунов В.В. Проблема идентификации отраслевого рынка в ОКВЭД: методический подход // Конкурентоспособность территорий. Матер. XXIV Всерос. экономического форума молодых ученых и студентов. В 4-х ч. Екатеринбург, 2021. С. 64–66.
9. Молчанов А.Н., Реут Д.В., Смирнов С.О. Анализ проблемного поля основных общероссийских классификаторов технико-экономической информации (ОКВЭД 2, ОКПД 2, ОКОФ) // Управление развитием крупномасштабных систем MLSD'2017. Матер. X междунар. конф. в 2-х т. Институт проблем управления им. В.А. Трапезникова; РАН, 2017. С. 358–361.
10. Игнатова Г.В., Иноземцева В.В. Информационное обеспечение закупочной деятельности в цифровой экономике // Вестник Саратовского государственного социально-экономического университета. 2019. № 3 (77). С. 15–19.

11. **Горюнова Н.Д., Громов Д.Н., Кузьмин А.Е.** Организация внутреннего финансового контроля государственными заказчиками при планировании закупок товаров, работ и услуг // Управленческое консультирование. 2019. № 5 (125). С. 42–52. DOI: 10.22394/1726-1139-2019-5-42-52
12. **Schoenherr T., Tummala V.** Electronic procurement: A structured literature review and directions for future research // *Internat. J. of Procurement Management*. 2007. no. 1. Pp. 8–37. DOI: 10.1504/IJPM.20-07.015353
13. **Masudin I., Aprilia G.D., Nugraha A., Restuputri D.P.** Impact of e-procurement adoption on company performance: Evidence from Indonesian manufacturing industry // *Logistics*. 2021. no. 5. P. 16. <https://doi.org/10.3390/logistics5010016>
14. **Milano S., Taddeo M., Floridi L.** Recommender systems and their ethical challenges // *AI & Soc.* 2020. no. 35. Pp. 957–967. <https://doi.org/10.1007/s00146-020-00950-y>
15. **Isinkaye F.O., Folajimi Y.O., Ojokoh B.A.** Recommendation systems: Principles, methods and evaluation // *Egyptian Informatics J.* 2015. Vol. 16(3). Pp. 261–273. DOI: 10.1016/j.eij.2015.06.005
16. **Valdiviezo-Diaz P., Ortega F., Cobos E., Lara-Cabrera R.** A collaborative filtering approach based on naive Bayes classifier // *IEEE Access*. 2019. Vol. 7. Pp. 108581–108592.
17. **Ajesh A., Nair J., Jijin P.S.** A random forest approach for rating-based recommender system // 2016 *Internat. Conf. on Advances in Computing, Communications and Informatics (ICACCI)*. DOI: 10.1109/icac-ci.2016.7732225
18. **Sun F., Liu J., Wu J., Pei C., Lin X., Ou W., Jiang P.** BERT4Rec // *Proc. of the 28<sup>th</sup> ACM Internat. Conf. on Information and Knowledge Management – CIKM’19*. DOI: 10.1145/3357384.3357895
19. **Ziegler C.N., McNeel S.M., Konstan J.A., Lausen G.** Improving recommendation lists through topic diversification // *Proc. of the 14<sup>th</sup> Internat. Conf. on World Wide Web*. 2005. Pp. 22–32.
20. **Seliverstov Y.A., Komissarov A.A., Poslovskaia E.D., Lesovodskaya A.A., Podtikhov A.V.** Detection of low-toxic texts in similar sets using a modified XLM-RoBERTa neural network and toxicity confidence parameters // *XXIV Internat. Conf. on Soft Computing and Measurements (SCM)*. 2021. Pp. 161–164. DOI: 10.1109/SCM52931.2021.9507117
21. **Zueva N., Kabirova M., Kalaidin P.** Reducing unintended identity bias in Russian hate speech detection // *Proc. of the 4<sup>th</sup> Workshop on Online Abuse and Harms*. 2010. Pp. 65–69. DOI: 10.18653/v1/P17, arXiv preprint arXiv:2010.11666
22. **Smetanin S.** Toxic comments detection in Russian // *Proc. of the Internat. Conf. on Computational Linguistics and Intellectual Technologies. Dialogue 2020*. DOI: 10.28995/NNNN-NNNN/2020-19-1-11
23. **Seliverstov Y.A., Seliverstov S.A., Malygin I.G., Korolev O.** Traffic safety evaluation in Northwestern Federal District using sentiment analysis of Internet users’ reviews // *Transportation Research Procedia*. 2020. no. 50. Pp. 626–635. DOI: 10.1016/j.trpro.2020.10.074
24. **Селиверстов Я.А., Чигур В.И., Сазанов А.М., Селиверстов С.А., Свистунова А.С.** Разработка системы для тонового анализа отзывов пользователей портала “AUTOSTRADA.INFO/RU” // *Труды СПИИРАН*. 2019.18:2. С. 354–389.
25. **Wang Qing, Peng Rongqun, Wang Jiaqiang, Xie Yushu, Zhou Yangfan.** Research on text classification method of LDA- SVM based on PSO optimization. 2019. Pp. 1974–1978. DOI: 10.1109/CAC48633.2019.8996952
26. **Двойникова А.А., Карпов А.А.** Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных // *Информационно-управляющие системы*. 2020. № 4. С. 20–30. DOI: 10.31799/1684-8853-2020-4-20-30
27. **Kurатов Y., Arkhipov M.** Adaptation of deep bidirectional multilingual transformers for Russian language // *Computational Linguistics and Intellectual Technologies: Papers from the Annual Internat. Conf. Dialogue. RSUH, Moscow*, 2019. Pp. 333–340.
28. **Huang Po-Sen, He Xiaodong, Gao Jianfeng, Deng Li, Acero Alex, Heck Larry.** Learning deep structured semantic models for web search using clickthrough data. 2013. Pp. 2333–2338. DOI: 10.1145/2505515.2505665

29. **Usama Khalid, Mirza Omer Beg, Muhammad Umair Arshad.** RUBERT: A bilingual Roman Urdu BERT using cross lingual transfer learning // arXiv:2102.11278v1 [cs.CL] 22 Feb 2021.
30. **Radford A., Narasimhan K., Salimans T., Sutskever I.** Improving language understanding by generative pre-training // URL: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf), 2018.
31. **Vig J., Belinkov Y.** Analyzing the structure of attention in a transformer language model. 2019. Pp. 63–76. DOI: 10.18653/v1/W19-4808
32. **Dehouche Nassim.** Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3): The best time to act was yesterday. The next best time is now // Ethics in Science and Environmental Politics. 2021. P. 21. DOI: 10.3354/ese00195
33. **Liu Yinhan, Ott Myle, Goyal Naman, Du Jingfei, Joshi Mandar, Chen Danqi, Levy Omer, Lewis Mike, Zettlemoyer Luke, Stoyanov Veselin.** RoBERTa: A robustly optimized BERT pretraining approach. 2019.
34. **Shoeybi Mohammad, Patwary Md. Mostofa Ali, Puri Raul, Legresley Patrick, Casper Jared, Catanzaro Bryan.** Megatron-LM: Training multi-billion parameter language models using model parallelism. 2019 // URL: <https://deepai.org/publication/megatron-lm-training-multi-billion-parameter-language-models-using-model-parallelism>
35. **Allal-Chérif Oihab, Virginia Simón-Moya, Ballester Antonio C.** Intelligent purchasing: How artificial intelligence can redefine the purchasing function // J. of Business Research. 2021. no. 124. Pp. 69–76. DOI: 10.1016/j.jbusres.2020.11.050
36. **Milana Carlo, Ashta Arvind.** Artificial intelligence techniques in finance and financial markets: A survey of the literature // Strategic Change. 2021. no. 30. Pp. 189-209. DOI: 10.1002/jsc.2403

## REFERENCES

1. **Mamedova N.A., Aleksandrova K.S., Pozdnyakova N.A.** Analiz informatsionnogo obespecheniya kontraktnoy sistemy na primere Ofitsialnogo sayta YeIS v sfere zakupok. *International Journal of Professional Science*, 2018, no. 4, Pp. 24–34. (rus)
2. **Rashid A.M., Albert I., Cosley D., Lam S.K., McNee S.M., Konstan J.A., et al.** Getting to know you: Learning new user preferences in recommender systems. *Proceedings of the International Conference on Intelligent User Interfaces*, 2002, Pp. 127–34.
3. **Hinz O., Eckert J.** The impact of search and recommendation systems on sales in electronic commerce. *Bus Inf Syst Eng*, 2010, no. 2, Pp. 67–77. DOI: 10.1007/s12599-010-0092-x
4. **Garcia Rodriguez M.J., Rodriguez Montequín V., Ortega Fernandez F., Villanueva Balsera J.M.** Bidders recommender for public procurement auctions using machine learning: Data analysis, algorithm, and case study with tenders from Spain. *Complexity*, 2020. DOI: 10.1155/2020/8858258
5. **Sunmola Funlade, Shehu, Yusuf.** A case study on performance features of electronic tendering systems. *Procedia Manufacturing*, 2020, Vol. 51, Pp. 1586–1591. DOI: 10.1016/j.promfg.2020.10.221
6. **Mehrbod A., Grilo A.** Advanced engineering informatics tender calls search using a procurement product named entity recognizer. *Advanced Engineering Informatics*, 2018, Vol. 36.
7. **Garcia Rodriguez M.J., Rodriguez Montequín V., Ortega Fernandez F., Villanueva Balsera J.M.** Public procurement announcements in Spain: Regulations, data analysis, and award price estimator using machine learning. *Complexity*, 2019.
8. **Migunov V.V.** Problema identifikatsii otraslevogo rynka v OKVED: metodicheskiy podkhod. *Konkurentosposobnost territoriy. Materialy XXIV Vserossiyskogo ekonomicheskogo foruma molodykh uchenykh i studentov.* Yekaterinburg, 2021. Pp. 64–66. (rus)
9. **Molchanov A.N., Reut D.V., Smirnov S.O.** Analiz problemnogo polya osnovnykh obshcherossiyskikh klassifikatorov tekhniko-ekonomicheskoy inorfmatzii (OKVED 2, OKPD 2, OKOF). *Upravleniye razvitiyem*

*krupnomasshtabnykh sistem MLS D'2017. Materialy X mezhdunarodnoy konferentsii*. Institut problem upravleniya im. V.A.Trapeznikova; RAS, 2017. Pp. 358–361. (rus)

10. **Ignatova G.V., Inozemtseva V.V.** Informatsionnoye obespecheniye zakupochnoy deyatelnosti v tsifrovoy ekonomike. *Vestnik Saratovskogo Gosudarstvennogo Sotsialno-Ekonomicheskogo Universiteta*, 2019, no. 3 (77), Pp. 15–19. (rus)

11. **Goryunova N.D., Gromov D.N., Kuzmin A.Ye.** Organizatsiya vnutrennego finansovogo kontrolya gosudarstvennymi zakazchikami pri planirovanii zakupok tovarov, rabot i uslug. *Upravlencheskoye Konsultirovaniye*, 2019, no. 5 (125), Pp. 42–52. (rus). DOI: 10.22394/1726-1139-2019-5-42-52

12. **Schoenherr T., Tummala V.** Electronic procurement: A structured literature review and directions for future research. *International Journal of Procurement Management*, 2007, no. 1, Pp. 8–37. DOI: 10.1504/IJPM.2007.015353

13. **Masudin I., Aprilia G.D., Nugraha A., Restuputri D.P.** Impact of e-procurement adoption on company performance: Evidence from Indonesian manufacturing industry. *Logistics*, 2021, no. 5, P. 16. <https://doi.org/10.3390/logistics5010016>

14. **Milano S., Taddeo M., Floridi L.** Recommender systems and their ethical challenges. *AI & Soc.*, 2020, no. 35, Pp. 957–967. <https://doi.org/10.1007/s00146-020-00950-y>

15. **Isinkaye F.O., Folajimi Y.O., Ojokoh B.A.** Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 2015, Vol. 16(3), Pp. 261–273. DOI: 10.1016/j.eij.2015.06.005

16. **Valdiviezo-Diaz P., Ortega F., Cobos E., Lara-Cabrera R.** A collaborative filtering approach based on naive Bayes classifier. *IEEE Access*, 2019, Vol. 7, Pp. 108581–108592.

17. **Ajesh A., Nair J., Jijin P.S.** A random forest approach for rating-based recommender system. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. DOI: 10.1109/icaccci.2016.7732225

18. **Sun F., Liu J., Wu J., Pei C., Lin X., Ou W., Jiang P.** BERT4Rec. *Proceedings of the 28<sup>th</sup> ACM International Conference on Information and Knowledge Management – CIKM'19*. DOI: 10.1145/3357384.3357895

19. **Ziegler C.N., McNeel S.M., Konstan J.A., Lausen G.** Improving recommendation lists through topic diversification. *Proceedings of the 14<sup>th</sup> International Conference on World Wide Web*, 2005, Pp. 22–32.

20. **Seliverstov Y.A., Komissarov A.A., Poslovskaia E.D., Lesovodskaya A.A., Podtikhov A.V.** Detection of low-toxic texts in similar sets using a modified XLM-RoBERTa neural network and toxicity confidence parameters. *XXIV International Conference on Soft Computing and Measurements (SCM)*, 2021, Pp. 161–164. DOI: 10.1109/SCM52931.2021.9507117

21. **Zueva N., Kabirova M., Kalaidin P.** Reducing unintended identity bias in Russian hate speech detection. *Proceedings of the 4<sup>th</sup> Workshop on Online Abuse and Harms*, 2010, Pp. 65–69. DOI: 10.18653/v1/P17, arXiv preprint arXiv:2010.11666

22. **Smetanin S.** Toxic comments detection in Russian. *Proceedings of the International Conference on Computational Linguistics and Intellectual Technologies*, Dialogue 2020. DOI: 10.28995/NNNNNNNN/2020-19-1-11

23. **Seliverstov Y.A., Seliverstov S.A., Malygin I.G., Korolev O.** Traffic safety evaluation in Northwestern Federal District using sentiment analysis of Internet users' reviews. *Transportation Research Procedia*, 2020, no. 50, Pp. 626–635. DOI: 10.1016/j.trpro.2020.10.074

24. **Seliverstov Y.A., Chigur V.I., Sazanov A.M., Seliverstov S.A., Svistunova A.S.** Razrabotka sistemy dlya tonovogo analiza otzyvov polzovateley portala "AUTOSTRADA.INFO/RU". *Tr. SPIIRAN*, 2019, 18:2, Pp. 354–389. (rus)

25. **Wang Qing, Peng Rongqun, Wang Jiaqiang, Xie Yushu, Zhou Yangfan.** *Research on text classification method of LDA- SVM based on PSO optimization*, 2019, Pp. 1974–1978. DOI: 10.1109/CAC48633.2019.8996952

26. **Dvoynikova A.A., Karpov A.A.** Analiticheskiy obzor podkhodov k raspoznavaniyu tonalnosti russkoyazychnykh tekstovykh dannykh. *Informatsionno-Upravlyayushchiye Sistemy*, 2020, no. 4, Pp. 20–30. (rus). DOI: 10.31799/1684-8853-2020-4-20-30



27. **Kuratov Y., Arkhipov M.** Adaptation of deep bidirectional multilingual transformers for Russian language. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue"*. RSUH, Moscow, 2019, Pp. 333–340.
28. **Huang Po-Sen, He Xiaodong, Gao Jianfeng, Deng Li, Acero Alex, Heck Larry.** *Learning deep structured semantic models for web search using clickthrough data*, 2013, Pp. 2333–2338. DOI: 10.1145/2505515.2505665
29. **Usama Khalid, Mirza Omer Beg, Muhammad Umair Arshad.** RUBERT: A bilingual Roman Urdu BERT using cross lingual transfer learning. *arXiv:2102.11278v1 [cs.CL] 22 Feb 2021*.
30. **Radford A., Narasimhan K., Salimans T., Sutskever I.** Improving language understanding by generative pre-training. Available: [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf), 2018
31. **Vig J., Belinkov Y.** *Analyzing the structure of attention in a transformer language model*, 2019, Pp. 63–76. DOI: 10.18653/v1/W19-4808
32. **Dehouche Nassim.** Plagiarism in the age of massive generative pre-trained transformers (GPT-3): The best time to act was yesterday. The next best time is now. *Ethics in Science and Environmental Politics*, 2021, P. 21. DOI: 10.3354/ese00195
33. **Liu Yinhan, Ott Myle, Goyal Naman, Du Jingfei, Joshi Mandar, Chen Danqi, Levy Omer, Lewis Mike, Zettlemoyer Luke, Stoyanov Veselin.** *RoBERTa: A robustly optimized BERT pretraining approach*, 2019.
34. **Shoeybi Mohammad, Patwary Md. Mostofa Ali, Puri Raul, Legresley Patrick, Casper Jared, Catanzaro Bryan.** Megatron-LM: Training multi-billion parameter language models using model parallelism. 2019. Available: <https://deepai.org/publication/megatron-lm-training-multi-billion-parameter-language-models-using-model-parallelism>
35. **Allal-Chérif Oihab, Virginia Simón-Moya, Ballester Antonio C.** Intelligent purchasing: How artificial intelligence can redefine the purchasing function. *Journal of Business Research*, 2021, no. 124, Pp. 69–76. DOI: 10.1016/j.jbusres.2020.11.050
36. **Milana Carlo, Ashta Arvind.** Artificial intelligence techniques in finance and financial markets: A survey of the literature. *Strategic Change*, 2021, no. 30, Pp. 189–209. DOI: 10.1002/jsc.2403

#### INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

**Селиверстов Ярослав Александрович**  
**Yaroslav A. Seliverstov**  
 E-mail: Y.Seliverstov@2035.university

**Комиссаров Андрей Александрович**  
**Andrey A. Komissarov**  
 E-mail: Andrew.Komissarov@gmail.com

**Лесоводская Алина Алексеевна**  
**Alina A. Lesovodskaya**  
 E-mail: a.lesovodskaya@2035.university

**Бовыкин Павел Геннадьевич**  
**Pavel G. Bovykin**  
 E-mail: p.bovykin@2035.university

**Подтихов Артур Владимирович**  
**Artur V. Podtikhov**  
 E-mail: a.podtikhov@2035.university

**Торсионов Станислав Сергеевич**  
**Stanislav S. Torsionov**  
E-mail: s.torsionov@2035.university

**Цирков Дмитрий Алексеевич**  
**Dmitriy A. Tsykov**  
E-mail: tsykov@2035.university

**Орлов Сергей Александрович**  
**Sergei A. Orlov**  
E-mail: orlov@ff.tsu.ru

*Поступила: 08.04.2022; Одобрена: 23.08.2022; Принята: 25.08.2022.*  
*Submitted: 08.04.2022; Approved: 23.08.2022; Accepted: 25.08.2022.*