

## Особенности получения и использования открытых данных для проведения социально-экономических исследований

А.М. Патрусова<sup>1а</sup>, А.Н. Бильданов<sup>1b</sup>, А.В. Шадрина<sup>1с</sup>, Л.С.Вахрушева<sup>2d</sup>

<sup>1</sup> Братский государственный университет, ул. Макаренко, 40, Братск, Россия

<sup>2</sup> АО "Экзософт", Инновационный центр Сколково

<sup>а</sup> patrusova@mail.ru, <sup>б</sup> fossadoe@yandex.ru, <sup>с</sup> alinashadrina61@gmail.com, <sup>д</sup> lvahr@yandex.ru

Статья поступила 06.05.2024, принята 30.05.2024

*В статье представлен обзорный анализ особенностей и возможностей поиска, а также использования открытых данных для проведения исследований, актуальность которых обусловлена необходимостью изучения современных инструментов работы с данными для выявления различных факторов, характеризующих состояние отраслей экономики страны, региона и отдельной организации. Исследование направлено на изучение инструментария для получения данных из различных источников, на формирование алгоритма подготовки данных к проведению анализа, включающего получение данных и рассмотрение особенностей работы с файлами различных типов. Рассмотрены способы получения открытых данных: на примере наборов данных Федеральной налоговой службы – скачивание файлов в формате CSV; посредством сервиса DaData с применением протокола API – получение данных в формате JSON. Представлены инструменты для чтения файлов форматов CSV и JSON, полученных из открытых источников, это библиотека Pandas языка программирования Python и специализированная программа для работы с данными Orange. Предложен алгоритм реализации подготовки данных к анализу, включающий несколько этапов: поиск достоверного источника данных; выбор метода получения данных; получение данных в виде файлов; проверка полученных файлов на корректность; проведение разведочного анализа данных; построение и оценка модели. Раскрыта программная реализация решений открытия файлов в указанных форматах и изучения датасета с данными, что является задачей аналитика данных с применением методов, функций и их комбинаций спектра современных инструментов Data Science.*

**Ключевые слова:** данные; открытые данные; анализ данных; запрос.

## Features of obtaining and using open data for socio-economic research

A.M. Patrusova<sup>1 a</sup>, A.N. Bildanov<sup>1 b</sup>, A.V. Shadrina<sup>1 c</sup>, L.C. Vakhrusheva<sup>2 d</sup>

<sup>1</sup> Bratsk State University, 40 Makarenko str., Bratsk, Russia

<sup>а</sup> patrusova@mail.ru, <sup>б</sup> fossadoe@yandex.ru, <sup>с</sup> alinashadrina61@gmail.com <sup>д</sup> lvahr@yandex.ru

<sup>2</sup> Exosoft JSC, Skolkovo Innovation Center

Received 06.05.2024, accepted 30.05.2024

*The article presents an overview analysis of the features and capabilities of search, as well as the use of open data for conducting research, the relevance of which is due to the need to study modern tools for working with data to identify various factors characterizing the state of economic sectors of a country, region and individual organization. The research is aimed at studying the tools for obtaining data from various sources, at developing an algorithm for preparing data for analysis, including obtaining data and considering the features of working with files of various types. Methods for obtaining open data are considered: using the example of data sets of the Federal Tax Service – downloading files in CSV format; through the DaData service using the API protocol – obtaining data in JSON format. Tools for reading CSV and JSON files obtained from open sources are presented: the Pandas library of the Python programming language, a specialized program for working with Orange data. An algorithm for preparing data for analysis is proposed, which includes several stages: searching for a reliable data source; choosing a data acquisition method; receiving data in the form of files; checking received files for correctness; conducting exploratory data analysis; construction and evaluation of the model. The software implementation of solutions for opening files in specified formats and studying a dataset with data is disclosed, which is the task of a data analyst using methods, functions and their combinations from a range of modern Data Science tools.*

**Keywords:** data; open data; data analysis; request.

**Введение.** Не секрет, что 21-й век – это век данных. Количество накопленной информации растет очень быстро, объем данных стремительно увеличивается и на сегодняшний день измеряется в зеттабайтах (один зеттабайт равен 1 млрд терабайт). Поэтому специалистам сферы *Data Science* необходимо

решать задачи сбора, хранения, анализа, обеспечения безопасности данных и др. Интерес и потребность работы с данными приобретают особое значение в условиях развития цифровой экономики.

Анализ данных в условиях развития цифровой экономики является одним из важнейших инст-

рументов обеспечения конкурентоспо-собности отраслей экономики, отдельных организаций в силу возникающих возможностей по выявлению нетривиальных зависимостей различных факторов деятельности предприятий. Работа с данными позволяет генерировать и проверять различные гипотезы, строить предсказывающие и объясняющие модели, что может быть направлено на принятие своевременных управленческих решений как в отдельной сфере экономики, так и в конкретной организации.

Во всем разнообразии источников данных наиболее популярными являются сайты, содержащие открытые данные. Открытые данные — это концепция, отражающая идею о том, что определенные данные должны быть свободно доступны для машиночитаемого использования и дальнейшей публикации без ограничений авторского права, патентов и других механизмов контроля<sup>1</sup>. Интерес ученых к открытым данным постоянно растет и направлен на исследование концепции открытых данных, проблем реализации основных принципов открытых данных и методов их изучения [1–3].

Решение прикладных задач с применением анализа данных затрагивает различные отрасли экономики: данные о движении транспорта, покупках, социальной поддержке граждан и многое другое. Разнообразие предметных областей применения открытых данных подтверждает их востребованность для проведения научных исследований в различных направлениях: статистка, политика, журналистика, социальные сети, загрязнение воздуха и др. [4–9].

Известно, что размещение государственными органами и органами местного самоуправления информации о своей деятельности в сети Интернет в форме открытых данных осуществляется в рамках реализации подп. «г» п. 2 Указа Президента Российской Федерации от 07 мая 2012 г. № 601 «Об основных направлениях совершенствования системы государственного управления».

К примеру, Федеральная налоговая служба (ФНС), Федеральная служба государственной статистики (Росстат), Министерство финансов Российской Федерации (Минфин), Федеральная служба судебных приставов и другие ведомства в специальных разделах на официальных сайтах размещают открытые данные, которыми может воспользоваться любой исследователь для изучения, анализа и моделирования процессов на основе проведенных исследований открытых данных.

Кроме органов государственной власти, источниками открытых данных могут являться любые организации или сообщества. Наибольший объем данных публикуется в интернете государственными органами и научным сообществом. Независимо от

источника данных, открытые данные, в отличие от коммерческих, требуют проверки на корректность и актуальность.

Рассматривая открытые данные как механизм обеспечения прозрачности функционирования организаций, а также как объект научных интересов, современному исследователю требуется владеть инструментарием получения данных, включающим навыки работы с файлами различных типов [10–13].

*Целью статьи является* проведение обзорного анализа особенностей и возможностей использования открытых данных, представленных в файлах различных форматов.

**Получение открытых данных.** Для решения задач анализа данных их можно получать из различных источников, затем подготовить или обработать для использования.

Данные некоторых открытых источников доступны для использования путем простого скачивания.

Рассмотрим возможности сервиса «Прозрачный бизнес», который оперирует открытыми данными ФНС России<sup>2</sup>. Данный сервис содержит 90 наименований наборов данных (рис. 1). Здесь можно увидеть информацию о количестве скачиваний данных, что говорит о популярности наборов данных 5, 8, 27 соответственно: Адреса и платежные реквизиты инспекций — 172 055 скачиваний; Перечень вакантных должностей государственной гражданской службы в территориальных органах ФНС России — 101 354 скачиваний; Сведения о физических лицах, являющихся учредителями (участниками) нескольких юридических лиц — 2 194 174 скачивания. Количество наборов данных, представленных в формате CSV — 75, представленных в формате XML — 14.

№ п/п	наименование набора данных	формат набора данных	количество скачиваний
1	Государственный адресный реестр (Федеральная информационная адресная система)	XML	28894
2	Реестр операторов электронного документооборота	CSV	95810
3	Реестр лиц/наем на осуществление деятельности по производству и реализации лицензированной от гудделек полиграфической продукции	XML	3071
4	Налоговые ставки и льготы	XML	26154
5	Адреса и платежные реквизиты инспекций	CSV	172055
6	Перечень подведомственных организаций	CSV	83872
7	Реестр лиц/наем на осуществление деятельности по организации и осуществлению сделок ИЭП в бухгалтерском контроле или кассовых операциях	XML	12870
8	Перечень вакантных должностей государственной гражданской службы в территориальных органах ФНС России	CSV	101354
9	Налоговый календарь	XML	4267

Рис. 1. Фрагмент страницы сервиса «Прозрачный бизнес» ФНС

<sup>1</sup> Открытые данные [Электронный ресурс]. Режим доступа: [https://ru.wikipedia.org/wiki/Открытые\\_данные](https://ru.wikipedia.org/wiki/Открытые_данные) (дата обращения 01.05.2024)

<sup>2</sup> Федеральная налоговая служба [Электронный ресурс]. Режим доступа: <https://www.nalog.gov.ru/opendata/> (дата обращения 10.05.2024)

Перейдя на любую из 90 представленных страниц с данными, пользователь получает возможность изучить 18 характеристик набора данных, включая гиперссылку на набор данных (п. 8) и описание структуры набора данных (п. 10) (рис. 2). Используя указанные гиперссылки, путем скачивания можно получить наборы данных в файлах *data-20181205-structure-20181205.csv*, *structure-20170614.csv* и приступить к изучению данных. Рассмотренный способ получения данных является самым простым.

№	Имя набора данных
1	Идентификационный номер: 7707329152-job
2	Наименование набора данных: Перечень вакантных должностей государственной гражданской службы в территориальных органах ФНС России
3	Описание набора данных: Перечень вакантных должностей государственной гражданской службы в территориальных органах ФНС России
4	Эвлацелл набора данных: ФНС России
5	Отвественное лице: Кузнецов Вячеслав Германович (Советник государственной гражданской службы Российской Федерации 2 класса)
6	Телефон отвественного лица: +7 (495) 913-00-00 (доб. 18-63)
7	Адрес электронной почты отвественного лица: orendika@tax.gov.ru
8	Гиперссылка (URL) на набор: <a href="https://data.fedtax.ru/orendata/7707329152-job/data-20181205-structure-20181205.csv">https://data.fedtax.ru/orendata/7707329152-job/data-20181205-structure-20181205.csv</a>
9	Формат данных: csv
10	Описание структуры набора данных: <a href="https://data.fedtax.ru/orendata/7707329152-job/structure-20170614.csv">https://data.fedtax.ru/orendata/7707329152-job/structure-20170614.csv</a>

Рис. 2. Фрагмент характеристик набора данных «Перечень вакантных должностей государственной гражданской службы в территориальных органах ФНС России»

Другие открытые источники позволяют получать данные только с помощью интерфейса API. Иногда требуется предварительная регистрация на сайте организации с открытыми данными с целью контроля потоков и объемов скачивания.

Интерфейс прикладного программирования (API) — набор готовых классов, функций, структур и констант, предоставляемых приложением (библиотекой, сервисом) для использования во внешних программных продуктах<sup>1</sup>. Следовательно,

<sup>1</sup> Методические рекомендации по публикации открытых данных государственными органами и органами местного самоуправления, а также технические требования к публикации открытых данных. Версия 3.0. — Утверждены протоколом заседания Правительственной комиссии по координации деятельности Открытого правительства от 29 мая 2014 г. № 4.

но, с помощью программного интерфейса приложения API можно получить доступ со своего компьютера к данным на другом компьютере.

Протокол передачи данных при клиент-серверном взаимодействии (HTTP — *HyperText Transfer Protocol*) для обращения к ресурсу предполагает несколько способов, называемых HTTP-методами, среди которых самые используемые — GET, POST. Метод GET отвечает за получение данных с сервера, и с помощью этого метода пользователь имеет возможность запросить необходимую информацию. Метод POST отвечает за сохранение информации, создание новых записей, отправку данных на сервер. С помощью метода POST происходит передача данных на сервер для их сохранения, а также для того, чтобы запустить цепочку вычислений, в результате которых вернется ответ с данными.

Рассмотрим возможности сервиса *DaData*<sup>2</sup>, который посредством протокола API позволяет осуществлять проверку различных контактов — адресов, реквизитов индивидуальных предпринимателей (ИП) и юридических лиц (ЮЛ), ФИО, емейлов, телефонов.

Некоторые функциональные возможности раздела «API работы с ЕГРЮЛ и реестрами налоговой» сервиса *DaData*, заявленные на сайте компании, можно проверить на примере известной организации — ФГБОУ ВО «БрГУ».

**Сервис «Организация по ИНН».** Осуществление поиска организации или ИП по ИНН, КПП, ОГРН и получение сведений о бизнесе из государственных источников. При введении в тело POST-запроса ИНН ФГБОУ ВО «БрГУ» — 3805100148 — сервис *DaData* возвращает сведения об организации на нескольких страницах в *json*-формате, включая наименование организации, КПП, ОГРН, ФИО руководителя, виды деятельности и др. (рис. 3).

**Сервис «Компания по e-mail».** Получение сведений об организации, которой принадлежит адрес электронной почты. При введении в тело POST-запроса *e-mail* ФГБОУ ВО «БрГУ» — *rector@brstu.ru* — сервис *DaData* возвращает сведения об организации в *json*-формате, включая наименование организации, домен сайта, ИНН, ОГРН, ОКВЭД, город и др. (рис. 4).

<sup>2</sup> *DaData* [Электронный ресурс]. Режим доступа: <https://dadata.ru> (дата обращения 10.05.2024)

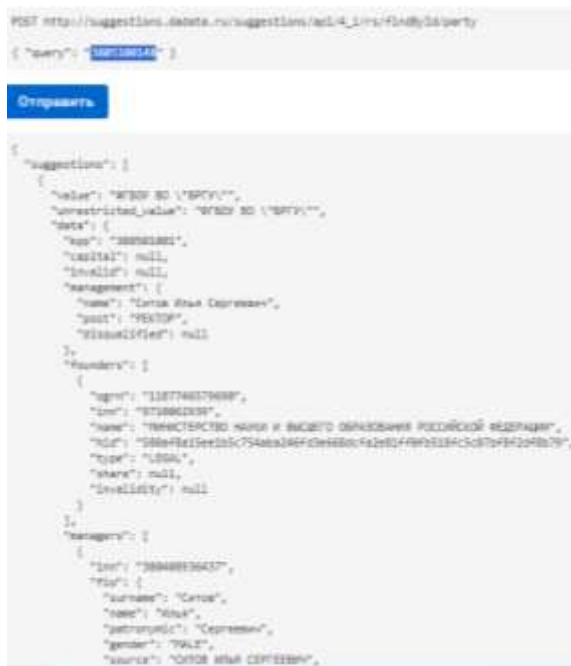


Рис. 3. Результат POST-запроса сервиса *DaData* по ИНН 3805100148 в *json*-формате

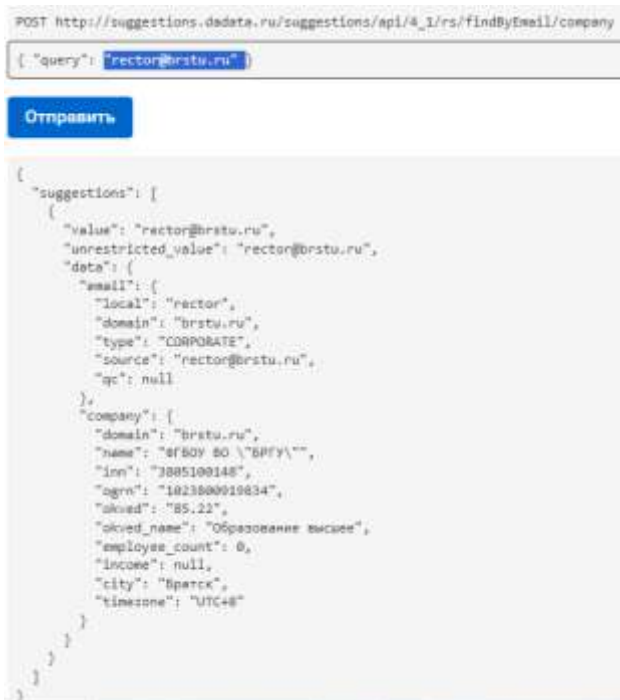


Рис. 4. Результат POST-запроса сервиса *DaData* по e-mail `rector@brstu.ru` в *json*-формате

Другим примером сервиса *DaData* является раздел «API работы с паспортными данными. Кем выдан паспорт» и заполнение наименования организации, выдавшей паспорт, по коду подразделения – 380-010 – сервис *DaData* выдает результат в *json*-формате, включая наименование организации в нужном падеже (ОТДЕЛОМ УФМС РОССИИ ПО ИРКУТСКОЙ ОБЛ. В Г. БРАТСКЕ И

БРАТСКОМ РАЙОНЕ), код региона и др. (рис. 5).



Рис. 5. Результат POST-запроса сервиса *DaData* по коду подразделения 380-010

Кроме рассмотренных разделов, сервис *DaData* содержит следующие возможности получения данных:

- API работы с почтовыми адресами и геокоординатами;
- API работы со справочниками Банка России;
- API работы с марками и моделями автомобилей;
- Прочие справочники (налоговые инспекции, мировые суды, классификаторы и др.).

Рассмотренные способы получения открытых данных – путем скачивания файлов и с применением протокола API – не являются исчерпывающими, но достаточны для проведения дальнейшей работы с данными.

Другим сервисом, позволяющим реализовать GET, POST-запросы, а также PUT, DEL-запросы, является сервис *Postman* (рис. 6).

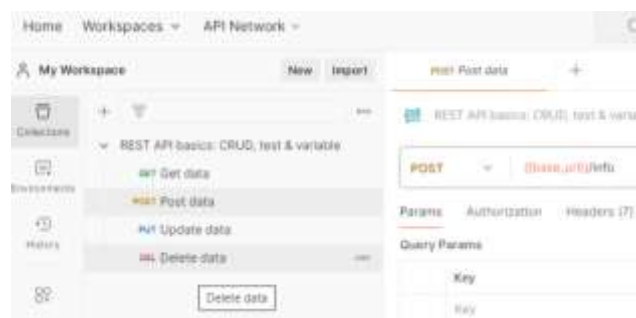


Рис. 6. Интерфейс сервиса *Postman*

**Особенности чтения файлов в различных форматах.** Как правило, данные хранятся в табличных формах, что является удобным как для изучения данных, так и для их обработки.

Самыми распространенными форматами файлов являются *xlsx*, *csv* и *json* [14]. При этом файл

формата *xlsx* — обычный файл, созданный в табличном процессоре *Excel*; *csv* — текстовый формат, предназначенный для хранения табличных данных; *json* — текстовый формат обмена данными, основанный на *JavaScript*.

Файлы форматов *xlsx* можно открыть и изучить в табличном процессоре *Excel*.

Чтобы посмотреть файл формата *csv*, можно использовать и *Excel*, и обычный Блокнот. Каждая строка файла формата *csv* — это строка соответствующей таблицы, а значения колонок в строке разделены специальным символом-разделителем.

Открыть *json*-файл можно любым текстовым редактором. Для удобства просмотра содержимого *json*-файла разработчики часто используют текстовый редактор с поддержкой синтаксиса разных языков программирования, в том числе с использованием языка программирования *Python* и библиотек *Json*, *Pandas*. Используя *Python* и его библиотеки, можно открыть файлы форматов *xlsx*, *csv*, *json* и реализовать этапы изучения данных и их подготовки к построению, например, математической модели.

Воспользуемся файлом о вакансиях ФНС России, полученным путем скачивания. Используя библиотеку *Pandas* языка программирования *Python*, можно получить датасет с описанием структуры набора данных «Перечень вакантных должностей государственной гражданской службы в территориальных органах ФНС России» (далее — Перечень вакансий ФНС России), представленной в файле *structure-20170614.csv* (рис. 7). Датасет с описанием структуры набора данных содержит 29 строк, 12 столбцов, некоторые ячейки не заполнены (значение *NaN*), где G1-G27 — показатели, характеризующие представленные вакансии (дата создания, наименование компании, должность, город, область и пр.).

Существует возможность получить список всех 29 показателей, например, дата создания вакансии, должность, город, уровень образования, семейное положение, владение английским языком и др. (рис. 8).

Далее, открыв файл *data-20181205-structure-20181205.csv*, получим датасет с набором данных о вакансиях ФНС России (рис. 9). Здесь представлены 314 строк (313 строк вакансий) в разрезе 29 столбцов (показателей). Функциональные возможности библиотеки *Pandas* открывают возможность изучить набор данных и провести необходимые исследования.

```
df2 = pd.read_csv('structure-20170614.csv', sep = ';')
df2
```

	Наименование таблицы	Unnamed: 1	Field name	English/Inname	description
0	NaN	GA	Номер вакансии	Vacancy number	Номер вакансии
1	Перечень вакантных должностей государственной ...	GB	Дата создания	date of creation	Дата создания
2	NaN	G1	Число прямых откликов	Number of direct responses	Число прямых откликов
3	NaN	G2	Название компании	Company name	Название компании
4	NaN	G3	Должность	Position	Должность
5	NaN	G4	Город	City	Город
6	NaN	G5	Область	Region	Область
7	NaN	G6	Регион	Region	Регион
8	NaN	G7	Уровень дохода, от	The level of income, from	Уровень дохода, от
9	NaN	G8	Уровень дохода, до	Income level, up to	Уровень дохода, до
10	NaN	G9	Возраст, от	Age, from	Возраст, от
11	NaN	G10	Возраст, до	Age, up to	Возраст, до
12	NaN	G11	Уровень образования	The level of education	Уровень образования
13	NaN	G12	Семейное положение	Family status	Семейное положение

Рис. 7. Датасет с описанием структуры набора данных из файла *structure-20170614.csv* с перечнем вакансий ФНС России

```
df2.iloc[:, 2]
```

0	Номер вакансии
1	Дата создания
2	Число прямых откликов
3	Название компании
4	Должность
5	Город
6	Область
7	Регион
8	Уровень дохода, от
9	Уровень дохода, до
10	Возраст, от
11	Возраст, до
12	Уровень образования
13	Семейное положение
14	Пол
15	Наличие детей
16	Соискатели из других городов
17	Владение английским языком
18	Источник вакансии
19	Тип работы
20	Права категории А
21	Права категории В
22	Права категории С
23	Права категории D
24	Права категории Е
25	Численность сотрудников в компании
26	Отрасль клиента
27	Опыт работы
28	Требования к квалификации

Name: Field name, dtype: object

Рис. 8. Список показателей перечня вакансий ФНС России

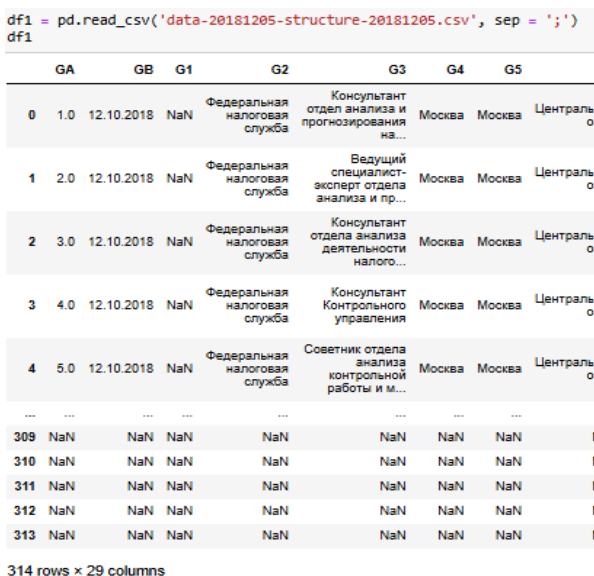


Рис. 9. Датасет с набором данных из файла *data-20181205-structure-20181205.csv* с перечнем вакансий ФНС России

Рассмотрим второй пример. Используя функциональные возможности библиотеки *Pandas*, можно получить датасет с описанием структуры набора данных «О задолженности по налогам и сборам, пеням и налоговым санкциям в бюджетную систему Российской Федерации по основным видам экономической деятельности» (далее – Задолженность в бюджетную систему России), представленной в файле *structure-20210511.csv* (рис. 10). Датасет с описанием структуры набора данных содержит 3 807 строк, 11 столбцов, некоторые ячейки не заполнены (значение NaN).

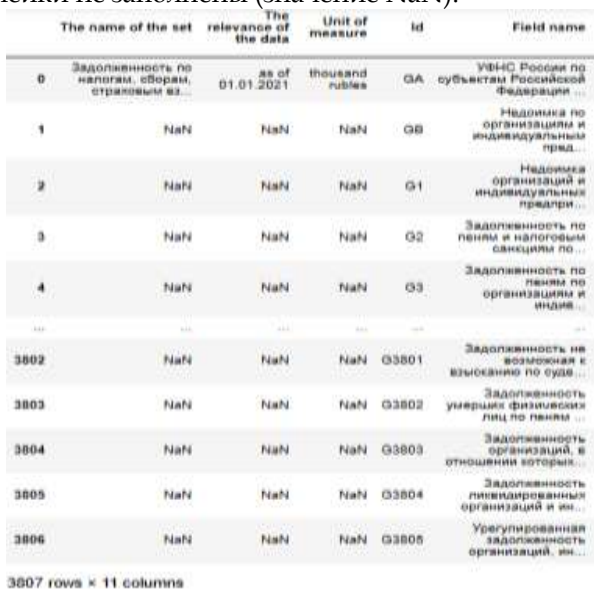


Рис. 10. Датасет с описанием структуры набора данных из файла *structure-20210511.csv* о задолженности в бюджетную систему России

Далее, открыв файл *data-20210511-structure-20210511.csv*, получим датасет с набором данных о задолженности в бюджетную систему России

(рис. 11). Здесь представлены 85 субъектов Российской Федерации в разрезе 3 807 показателей. Функциональные возможности библиотеки *Pandas* открывают возможность изучить набор данных и провести необходимые исследования.

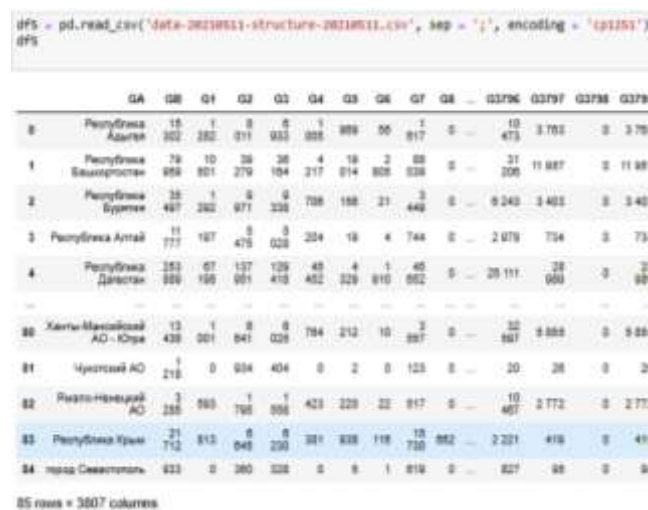


Рис. 11. Датасет с набором данных из файла *data-20210511-structure-20210511.csv* о задолженности в бюджетную систему России

Существует множество программ, позволяющих загружать файлы различных форматов и изучать наборы данных. Бесплатная программа для работы с данными *Orange* также позволяет ознакомиться с набором данных (рис. 12). *Orange* содержит множество средств для анализа данных и допускает отсутствие опыта программирования у исследователя.



Рис. 12. Результат загрузки файла *data-20210511-structure-20210511.csv* в программе *Orange* о задолженности в бюджетную систему России

Таким образом, получив файлы из источника открытых данных ФНС и используя современный инструментарий открытия файлов формата CSV и JSON, исследователю становятся доступны многие функции по изучению наборов данных. Одной из задач на данном этапе является подготовка данных и формирование итогового датасета, для решения которой проводится разведочный анализ данных, по результатам которого можно перейти к построению модели машинного обучения.

Автоматизация сбора данных из открытых баз данных является одной из актуальных задач в условиях стремительного роста объемов данных, включая открытые данные [15–18]. Программная реализация подобных решений является задачей аналитика данных с применением современных инструментов *Data Science*, что раскрывает возможность изучения поведения различных социально-экономических факторов, их связей, отражающих различные аспекты как отдельных сфер экономики, так и конкретной организации. Проверка всевозможных гипотез позволяет строить модели машинного обучения различных типов.

**Таблица 1.** Особенности работы с файлами формата CSV

Средства открытия файлов	Необходимость написания программного кода	Удобство визуализации содержимого файла	Возможность обработки данных
Excel	-	-	-
Orange	-	+	-
Python	+	+	+

**Таблица 2.** Особенности работы с файлами формата JSON

Средства открытия файлов	Необходимость написания программного кода	Удобство визуализации содержимого файла	Возможность обработки данных
Веб-ресурс <i>Json-formatter</i>	-	+	-
Python	+	+	+

Сравнительная оценка особенностей работы с файлами форматов CSV, JSON представлена в отношении ресурсов, рассмотренных в данной статье. Они бесплатны и общедоступны.

Рассматривая решение любой задачи анализа данных как проект [19; 20], предлагается алгоритм реализации подготовки данных к анализу, включающий следующие этапы:

1. Поиск достоверного источника данных.
2. Выбор метода получения данных.
3. Получение данных в виде файлов.
4. Проверка полученных файлов на корректность при открытии с помощью библиотек (*Pandas, Json*) языка программирования *Python*.
5. Проведение разведочного анализа данных.

Как правило, в подобных исследованиях необходимо определить целевую переменную, отражающую функциональную зависимость между различными факторами, при этом предметные области могут быть самыми разнообразными.

Например, с целью расширения предоставления услуг в сфере банковской деятельности организация может составить портрет клиента (набор и значения характеристик), который может взять кредит или оформить депозит. С целью построения правильной маркетинговой стратегии в сфере реализации товаров построение моделей классификации позволяет определить категорию цены определенных товаров в зависимости от их характеристик.

**Заключение.** Подводя итог проведенному исследованию, можно определить особенности работы с данными, полученными из открытых источников в различных форматах. В табл. 1 представлены особенности работы с файлами формата CSV, в табл. 2 — особенности работы с файлами формата JSON.

#### 6. Построение и оценка модели.

Реализация данного алгоритма позволит реализовать научный подход к анализу и обработке данных, что дает возможность прогнозировать поведение различных факторов на основе построения моделей машинного обучения с применением методов математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов и др.

Данные, их обработка и анализ, построение моделей машинного обучения направлены на своевременную выработку и принятие управленческих решений, поиск новых источников дохода и, в конечном итоге, повышение конкурентоспособности организации.

*Литература*

1. Стефогло С.Н., Томилова Н.И., Амиров А.Ж. Концепция открытых данных, основные принципы открытых данных // *Инновационная наука*. 2015. Т. 2, № 4. С. 53-55.
2. Парфенчик А.А. Открытые данные в контексте открытого Правительства и больших данных // *Информационные технологии и право: сб. материалов VII Междунар. науч.-практической конф.* (28 окт. 2021 г.). Минск, 2021. С. 292-297.
3. Боброва Е.В. Открытые данные и методы их изучения // *Информационный бюллетень ассоциации «История и компьютер»*. 2022. № 49. С. 199-201.
4. Лысова Н.В. Публикация официальной статистической информации в формате открытых данных // *Роль статистики в современном обществе и эффективном управлении: материалы Всерос. науч.-практической конф., посвящ. 100-летию образования гос. статистики в Коми крае (18-28 янв. 2019 г.)*. Сыктывкар, 2019. С. 375-381.
5. Карягин М.Е. Открытые государственные данные в поле публичной политики (по итогам общероссийского саммита «Открытые данные - 2015») // *Публичная политика*. 2017. № 1. С. 123-127.
6. Панюкова С.А. Роль открытых данных в развитии журналистики данных // *Знак: проблемное поле медиа образования*. 2015. № 1 (15). С. 25-33.
7. Оболенский Д.М., Шевченко В.И., Ченгарь О.В., Машенко Е.Н., Соина А.С. Сбор и предварительная обработка данных пользователей социальных сетей с помощью портала «Открытые данные» // *Modern Science*. 2020. № 7-2. Р. 369-378.
8. Ревенко В.В., Перькова М.А. Использование открытых данных для анализа загрязнения воздуха в городе // *Науч. мысль XXI века: результаты фундаментальных и прикладных исследований: сб. ст. междунар. науч.-практической конф.* (30 мая 2018 г.). Самара, 2018. С. 125-129.
9. Локтев А.П. Исследовательские данные как ключевой элемент открытой науки // *Наука, технологии и информация в библиотеках (Libway-2019): сб. тезисов докл. Междунар. науч.-практической конф.* (17-19 сент. 2019 г.). Иркутск, 2019. С. 129-130.
10. Кузовлев В.И., Орлов А.О. Повышение качества данных с использованием методики поиска аномалий на примере портала открытых данных правительства Москвы // *Инженерный вестн.* 2014. № 8. С. 7.
11. Miachin D.A., Pugach V.P., Avdeyuk S.S., Zunin V.V., Romanov A.Y. The open system for storing and processing of a dataset of combinational circuits // *Proceedings of the Institute for System Programming of the RAS*. 2023. V. 35, № 5. P. 81-90.
12. Nesterov S.A., Smolina E.M. The assessment of the results of a massive open online course using data mining methods // *Computing, Telecommunications and Control*. 2020. V. 13, № 1. P. 65-78.
13. Khayyat M., Aboulola O., Khayyat M. How open is open? A study of two Irish open government data websites // *Public Administration Issues*. 2022. № 5. P. 30-44.
14. Бородин А.Ф. Сравнительный анализ систем, предоставляющих открытые данные // *Дневник науки*. 2019. № 5 (29). С. 32.
15. Иванова Н.С., Савкина В.А., Карпулевич Е.А., Турдаков Д.Ю. Инструмент для сбора информации о научных публикациях из открытых баз данных: свид. о регистрации программы для ЭВМ RU 2023688841; заявл. 11.12.2023; опубл. 25.12.2023.
16. Резников Д.Ю. Система сбора и анализа открытых данных: свид. о регистрации программы для ЭВМ RU 2020662024; заявл. 29.09.2020; опубл. 06.10.2020.
17. Грабовой А.В., Каприелова М.С., Кильдяков А.С., Никитов А.В., Чехович Ю.В. База данных изображений из научных публикаций, находящихся в открытом доступе: свид. о регистрации базы данных RU 2022623549; заявл. № 2022623578 от 09.12.2022; опубл. 19.12.2022.
18. Лапушкин А.С., Шмойлов Д.В., Ладиков А.В., Ефремов А.А. Способ передачи данных на сервер с использованием открытого ключа: пат. на изобретение RU 2749182 С1; заявл. № 2020120438 от 19.06.2020; опубл. 07.06.2021.
19. Khaliev M.S.-U., Patrusova A.M., Natalevich L.F. Project management using it applications // *Journal of Physics: Conference Series. Ser. «International Conference on IT in Business and Industry, IT-BI 2021»*, 2021. P. 012127.
20. Patrusova A.M., Vahrusheva M.Y. Process management: some implementation aspects // *IOP Conference Series: Materials Science and Engineering. International Science and Technology Conference «FarEastCon 2019»*, 2020. P. 082028.