

НАУЧНАЯ СТАТЬЯ

УДК 330.45; 330.46; 330.47; 338



## Data Science в экономических процессах

**Алиса Романовна ДУДОВА**  ФГБОУ ВО «Саратовский национальный исследовательский  
государственный университет им. Н.Г. Чернышевского»  
410000, Российская Федерация, г. Саратов, ул. Вольская, 10А [alisiadudova@mail.ru](mailto:alisiadudova@mail.ru)

**Аннотация.** Исследовано экономическое применение Data Science (DS). DS – «наука о данных», зародилась на стыке статистического и интеллектуального анализа численных данных. В работе описаны основные наиболее перспективные направления и области эффективного использования различных инструментов DS. На примере функционирования российских рынков проиллюстрировано использование методов «науки о данных» в экономике. Показан конструктивный потенциал DS, который может использоваться в процессе целеполагания в условиях цифровизации экономики для обеспечения экономической безопасности Российской Федерации и построения эффективных маркетинговых стратегий.

**Ключевые слова:** Data Science, статистика, цифровизация, оптимизация, экономическая безопасность, прогнозирование, риски

**Для цитирования:** Дудова А.Р. Data Science в экономических процессах // Державинский форум. 2024. Т. 8. № 1. С. 119-127.

ORIGINAL ARTICLE

UDC 330.45; 330.46; 330.47; 338

## Data Science in economic processes

**Alisa R. DUDOVA**  Saratov State University  
10A Volskaya St., Saratov, 410000, Russian Federation [alisiadudova@mail.ru](mailto:alisiadudova@mail.ru)

**Abstract.** The economic application of Data Science (DS) has been investigated. DS – “data science”, originated at the junction of statistical and numerical data mining. The paper describes the main most promising areas and areas of effective use of various DS tools. Using the example of the Russian functioning markets, the use of “data science” methods in economics is illustrated. The constructive potential of DS is shown, which can be used in the process of goal-setting in the digitalization context of the economy to ensure the economic security of the Russian Federation and build effective marketing strategies.

**Keywords:** Data Science, statistics, digitalization, optimization, economic security, forecasting, risks

**For citation:** Dudova, A.R. (2024). Data Science in economic processes. *Derzhavinskii forum = Derzhavin Forum*, vol. 8, no. 1, pp. 119-127. (In Russ., abstract in Eng.)

---

## ПОСТАНОВКА ПРОБЛЕМЫ

В настоящее время актуальна проблема обработки, эффективного использования и хранения данных. Поскольку цифровизация обретает все большие масштабы, охватывая многие сферы жизнедеятельности общества (рис. 1 и 2), постепенно увеличиваются и объемы информации, на основе которой осуществляется аналитическая и прогнозная деятельность [1]<sup>1</sup>.

От эффективности обработки данных специалистами зависят определение векторов и расчет перспектив дальнейшего развития малых и крупных организаций. При этом очень важны форма представления и подачи данных, которые напрямую влияют на степень восприятия информации руководством, принимающим соответствующие решения. Поэтому целью данной работы будет исследование использования новейших методов Data Science (DS) для анализа социально-экономических процессов в современной экономике России.

## ОСНОВНЫЕ ПОНЯТИЯ DATA SCIENCE

Data Science – это наука о данных, а также профессиональная деятельность, смысл которой состоит в сборе, хранении и обработке больших объемов данных. «Дата-сайентисты» – это специалисты, разрабатывающие специальные модели и

---

<sup>1</sup> К примеру, из датчиков ежедневно производятся около трех миллиардов байт данных, мобильных устройств, онлайн-транзакций и социальных сетей. Интересно и то, что около 90 % данных в мире было создано всего лишь за последние 3 года [1].

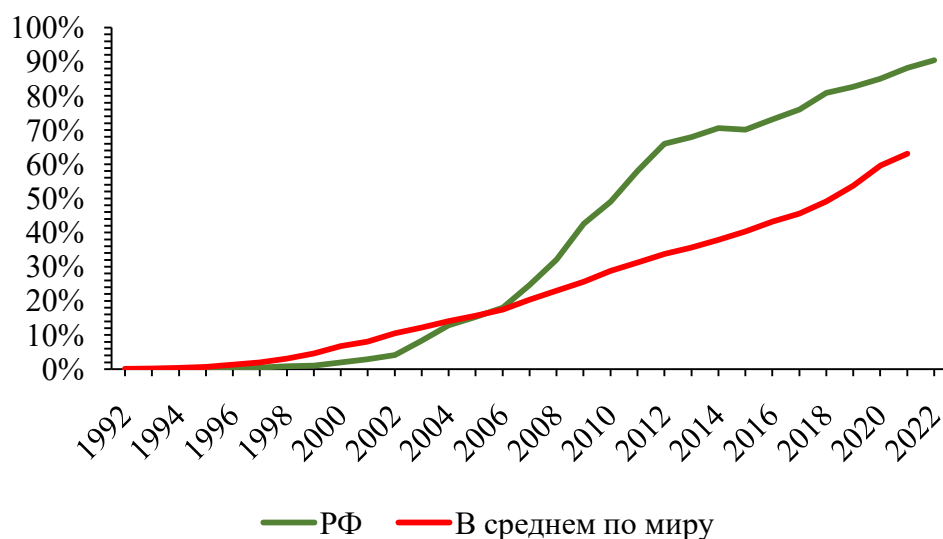
алгоритмы, которые позволяют упростить процедуру проверки сложного анализа больших объемов данных. Зачастую «дата-сайентисты» для облегчения создания моделей применяют автоматизированные инструменты машинного обучения. В свою очередь, экономисты, используя данные, высчитывают коэффициенты устойчивости/неустойчивости компаний, позволяющие наметить тенденции к развитию той или иной организации [2].

## ОСНОВНЫЕ ИНСТРУМЕНТЫ DATA SCIENCE

В качестве основного языка программирования специалистами по Data Science используется многофункциональный Python, который применяется для анализа данных, разработки моделей и написания рабочего кода. Помимо Python, в эконометрике также принято использовать языки R и SQL для написания запросов к источникам данных. В качестве интерактивной оболочки для различных экспериментов и разведочного анализа данных «дата-сайентисты» используют Jupyter Notebook. Наиболее распространенной и удобной, а поэтому популярной, средой разработки дата-сайентистов выступает PyCharm IDE. Помимо этого, в работе дата-сайентистов можно встретить и инструменты визуализации, среди которых выделяют: Tableau (для построения различных бордов, отчетов) и SAS (готовое решение по анализу данных).

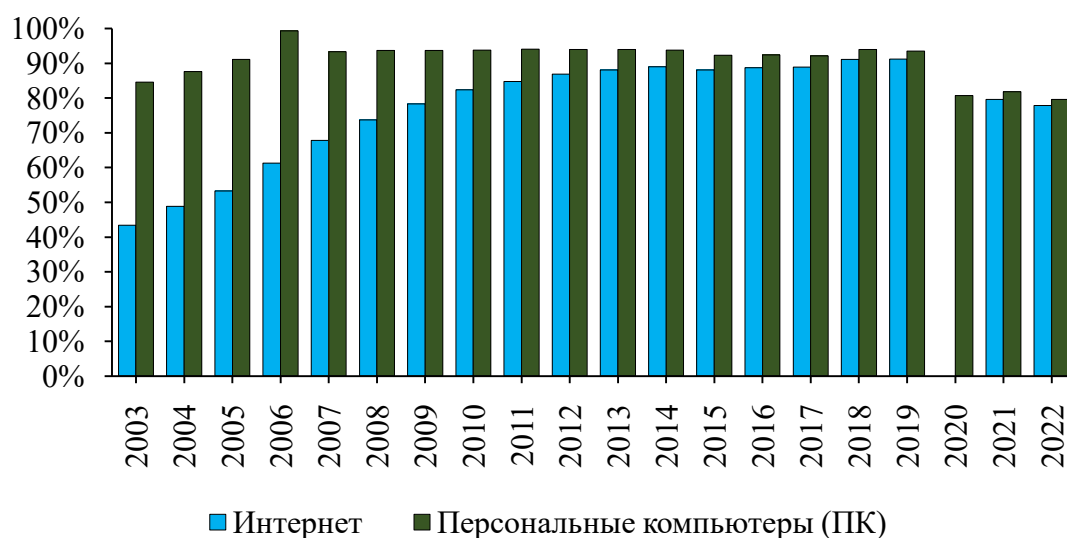
Если говорить о разведочном анализе (Exploratory Data Analysis), в Jupyter Notebook для построения графиков, схем и диаграмм применяют Matplotlib, Plotly, Seaborn, которых также хватает, чтобы исследовать различные зависимости в экономических данных. Инструменты

библиотеки NumPy позволяют оперировать матрицами, векторами, выполнять различные преобразования данных в области линейной алгебры и т. д.



**Рис. 1.** Динамика индивидуальных пользователей Интернета в РФ и в мире в 1992–2022 гг., % от общей численности населения (<https://data.worldbank.org/>)

**Fig. 1.** Dynamics of individual Internet users in the Russian Federation and on average around the world in 1992–2022, % of the total population (<https://data.worldbank.org/>)



**Рис. 2.** Удельный вес организаций, использовавших цифровые технологии: Интернет и ПК, – в России в 2003–2022 гг., % от общего числа организаций (<https://rosstat.gov.ru/statistics/>)<sup>2</sup>

**Fig. 2.** Share of organizations that used digital technologies: Internet and PC, – in Russia in 2003–2022, % of the total number of organizations (<https://rosstat.gov.ru/statistics/>)

<sup>2</sup> В связи с переходом на новые программные средства обработки данных, информация по показателю за 2020 г. не разрабатывалась (<https://rosstat.gov.ru/>)

Как уже было сказано, важнейшая задача дата-сайентиста – это разработка моделей, для которых используются простые конструкции, представленные в `scikit-learn`. Для создания более сложных конструкций используются библиотеки с бустинговыми алгоритмами: `XGBoost`, `CatBoost`, `LightGBM`, которые применяются для регрессии и классификации. С помощью различных фреймворков: `TensorFlow`, `Pytorch`, `Keras`, можно создавать нейронные сети для задач сегментации, распознавания, детекции и т. д. Для работы с большими массивами данных, с которыми нужно выполнять различные распределенные вычисления, дата-сайентисты часто используют экосистему `Hadoop`.

Специалисты по `Data Science` также активно применяют инструменты для автоматизации процессов, среди которых можно выделить: `Apache Airflow` и `Luigi`. С помощью этих универсальных инструментов можно следить за всеми задачами, производить вычисления, сбор данных, создание витрин, обучение и вариацию моделей, разработку и сохранение прогнозов и рекомендаций.

## DATA SCIENCE В РОССИЙСКОЙ ЭКОНОМИКЕ

На сегодняшний день существует множество экономических задач, в решении которых наука о данных позволит улучшить функционирование компаний, тем самым повысив их доходы. По данным Росстата, в 2022 г. более  $\frac{2}{3}$  из российских организаций использовали в своей деятельности специальные программные средства (рис. 3). Стоит отметить, что резкое падение числа фирм-пользователей специальных программных средств в 2020 г. – это статистический феномен, то есть результат изменения методологии. Наиболее распространенными «кейсами» для предприятий являются: оптимизация ценообразования товаров и

услуг<sup>3</sup>, прогнозирование наплыва/оттока клиентов и определение персонального профиля рекомендаций товаров. При этом любое конкретное решение будет определяться спецификой проекта и организации.

В финансовой сфере наука о данных применяется в следующих областях:

1) кредитный скоринг (автоматизированная система оценки платежеспособности заемщика): анализ случаев отказа или выдачи кредитов клиенту банка в зависимости от кредитной истории и других характеристик;

2) оценка вероятности наступления страховых случаев;

3) верификация пользователей и предотвращение мошеннических транзакций;

4) анализ показателей портфелей заемщиков;

5) прогноз спроса на наличные в банкоматах (такая модель, например, используется у банка «Райффайзенбанк»);

6) рекомендация инвестиционных инструментов (акций, облигаций) для создания сбалансированного портфеля (соответствующие алгоритмы используются в «Тинькофф. Инвестиции»)<sup>4</sup>.

Кроме того, `DS` находит применение и в сфере создания различных информационных систем<sup>5</sup>. Так, на базе `Data Science` моделей даже был создан онлайн-университет `GeekBrains`, в который входят под-

<sup>3</sup> Согласно данным Росстата, в 2022 г. специальные программные средства для управления закупками и продажами ТРУ использовали соответственно 33,7 и 26,6 % из обследованных российских организаций (<https://rosstat.gov.ru/statistics/>).

<sup>4</sup> Как компании превращают данные в деньги: обзор российских практик `Data Science`. URL: <https://proglib.io/p/kak-kompanii-prevrashchayut-dannye-v-dengi-obzor-rossiyskih-praktik-data-science-2020-07-18> (дата обращения: 27.09.2023).

<sup>5</sup> Согласно данным Росстата, в 2022 г. электронно-справочными системами пользовались, глобальными информационными сетями и обучающими программами пользовались соответственно 47,5, 28,4 и 26,1 % из обследованных российских организаций (<https://rosstat.gov.ru/statistics/>).

разделения по анализу данных: факультеты Искусственного интеллекта (ИИ), Аналитики Big Data и Data Engineering. Каждый из этих департаментов соответствует различным аспектам работы с данными и разным запросам рынка, что указывает на постепенную специализацию профессии «дата-сайентистов»<sup>6</sup>.

Еще одной важнейшей областью, где можно применять технологии Data Science, является сфера промышленности<sup>7</sup>. Важнейшими задачами при этом выступают:

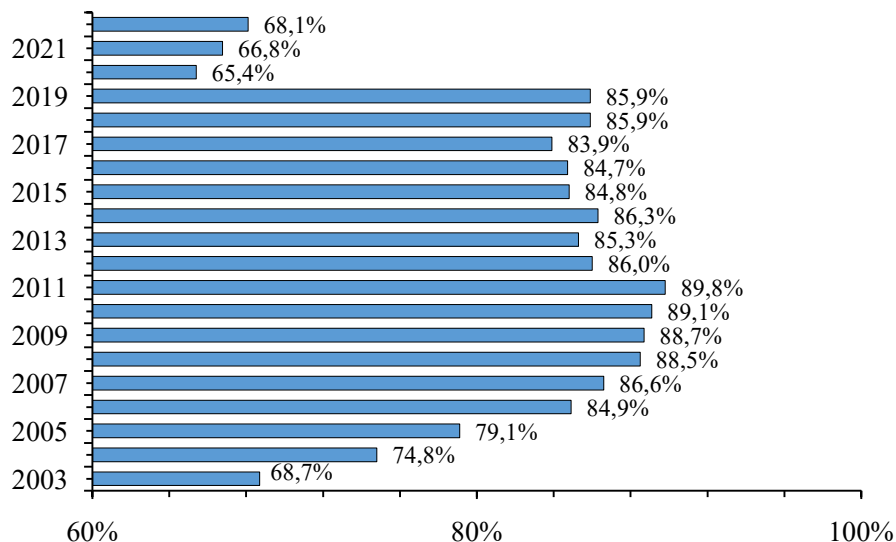
- 1) предотвращение хищений: контроль периметра промышленных территорий;
- 2) предотвращение несчастных случаев: контроль соблюдения зон ответственности, детекция ношения масок, определение аномального поведения;
- 3) задачи учета: считывание штрихкодов и буквенно-цифровых кодов для анализа перемещения грузов и других объектов;
- 4) контроль качества: своевременное обнаружение дефектов, контроль физических параметров;
- 5) робототехнические решения: локализация мобильных роботов, манипуляционные роботы, планирование траекторий.

## DATA SCIENCE И ЭКОНОМИЧЕСКАЯ БЕЗОПАСНОСТЬ

Еще одной задачей Data Science является обеспечение экономической безопасности. Всем известно, что предотвращение несанкционированного доступа, использования, раскрытия, искажения, изменения,

<sup>6</sup> Как компании превращают данные в деньги: обзор российских практик Data Science. URL: <https://proglib.io/p/kak-kompanii-prevrashchayut-dannye-v-dengi-obzor-rossijskih-praktik-data-science-2020-07-18> (дата обращения: 27.09.2023).

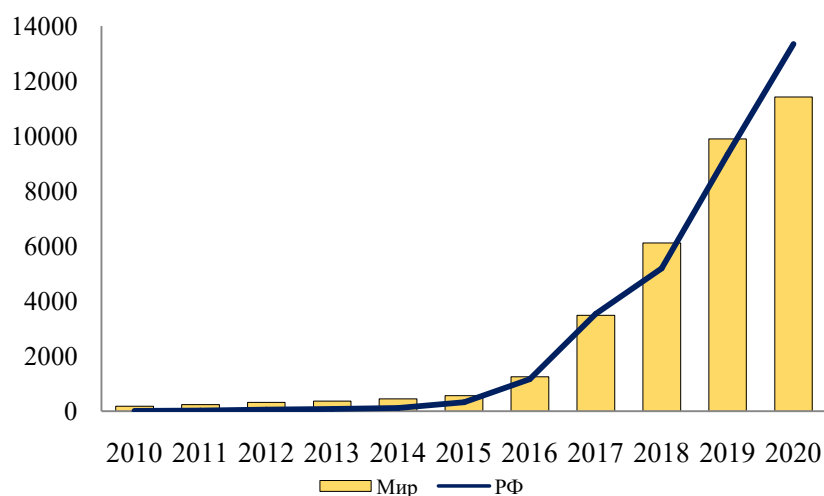
<sup>7</sup> Согласно данным Росстата, в 2022 г. специальные программные средства для управления автоматизированным производством и/или отдельными техническими средствами и технологическими процессами использовали 15,7 % из обследованных российских организаций (<https://rosstat.gov.ru/statistics/>).



**Рис. 3.** Удельный вес организаций, использовавших специальные программные средства, в России в 2003–2022 гг., % от общего числа организаций (<https://rosstat.gov.ru/statistics/>)<sup>8</sup>

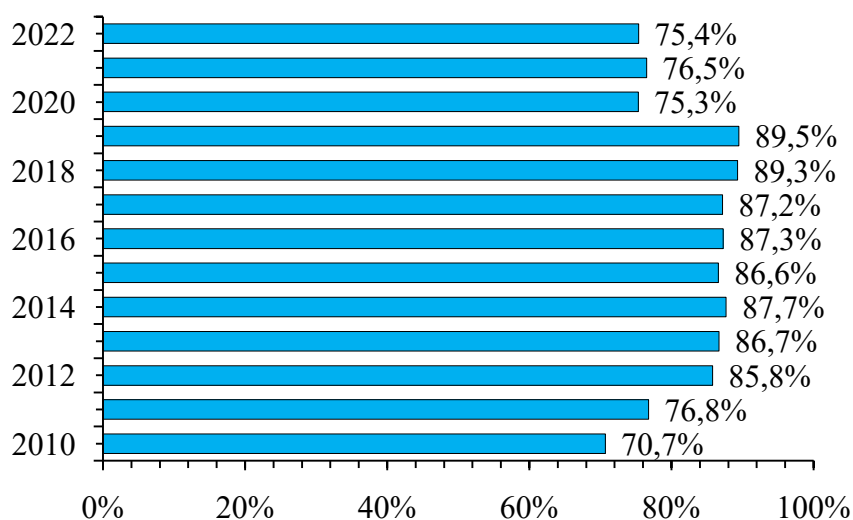
**Fig. 3.** Share of organizations that used special software in Russia in 2003–2022, % of the total number of organizations (<https://rosstat.gov.ru/statistics/>)

<sup>8</sup> По данным формы федерального статистического наблюдения № 3-информ «Сведения об использовании цифровых технологий и производстве связанных с ними товаров и услуг» (<https://rosstat.gov.ru/>).



**Рис. 4.** Использование безопасных веб-серверов в России и в среднем по миру в 2010–2020 гг., шт. на 1 млн человек (<https://data.worldbank.org/>)

**Рис. 4.** The use of secure web servers in Russia and on average around the world in 2010–2020, pcs. per 1 million people (<https://data.worldbank.org/>)



**Рис. 5.** Доля организаций, использовавших средства защиты информации, передаваемой по глобальным сетям, в России в 2010–2022 гг., % от общего числа организаций (<https://rosstat.gov.ru/statistics/>).

**Fig. 5.** The share of organizations that used means to protect information transmitted over global networks in Russia in 2010–2022, % of the total number of organizations (<https://rosstat.gov.ru/statistics/>).

исследования, записи или уничтожения информации становится важнейшей основополагающей задачей любой организации. Так, в Российской Федерации использование безопасных интернет-серверов в 2010–2020 гг. выросло в 781 раз, превысив даже среднемировое значение в

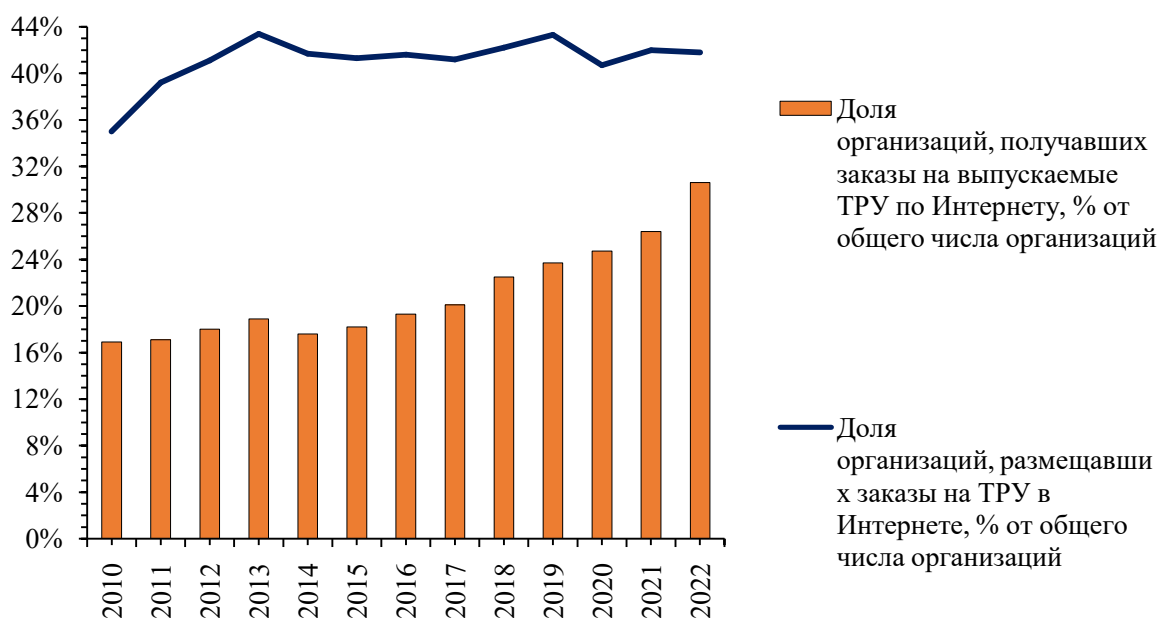
2020 г. (рис. 4). При этом в 2019 г. в среде российских организаций наблюдался практически 90 % охват по использованию средств защиты информации. Хотя после 2020 г. данный показатель снизился, оно все равно  $> \frac{3}{4}$  от всех организаций (рис. 5).

На практике обеспечение экономической безопасности: выявление и/или предсказание экономических угроз, – посредством Data Science осуществляется через анализ больших данных, алгоритмы машинного обучения и иные современные технологии обработки информации. Подобные процедуры включают в себя анализ тенденций развития рынка, прогнозирование экономических кризисов, обнаружение потенциальных финансовых мошенничеств. Используя алгоритмы машинного обучения, наука о данных может помочь в обнаружении необычных паттернов в финансовых данных, которые могут указывать на мошеннические действия, или, применяя анализ временных рядов и другие статистические методы, она может помочь прогнозировать возможные экономические кризисы, давая правительству, государственному и ком-

мерческому сектору экономики возможность заранее подготовиться и смягчить их последствия [1; 3; 4].

## DATA SCIENCE И МАРКЕТИНГ

Методы DS также активно используются в современном маркетинге. Хотя доли российских организаций, которые размещают или получают заказы на товары, работы и услуги, пока не превышали в 2022 г. соответственно 31 и 42 % от общего числа фирм, но с учетом масштаба охвата сети Интернет среди физических и юридических лиц в Российской Федерации существует значительный потенциал для роста в этом направлении (рис. 1, 2 и 6). Основные перспективные сферы применения Data Science в области маркетинга в Российской Федерации можно условно разбить на пять подгрупп:



**Рис. 6.** Доли организаций, получавших и размещавших заказы на ТРУ через сеть Интернет в России в 2010–2020 гг., % от общего числа организаций (<https://rosstat.gov.ru/statistics/>).

**Fig. 6.** The shares of organizations that received and placed orders for industrial and technical materials via the Internet in Russia in 2010-2020, % of the total number of organizations (<https://rosstat.gov.ru/statistics/>).

1) аналитика (благодаря разработкам в области Data Science маркетологи получают специфическую аналитику real-time, что позволяет им довольно быстро адаптироваться к какой-либо новой ситуации и принять меры по изменению стратегии. Таким образом, ускоряется планирование будущих компаний. Прогнозы, основанные на данных, позволяют выдвинуть гипотезу о том, как будет происходить ценообразование в дальнейшем и какую цену следует выбрать для сохранения или увеличения объемов продаж);

2) оптимизация бюджета (аналитика и некоторая предсказуемость, например, дальнейшего поведения потребителей, может существенно сэкономить денежные средства компании. Благодаря использованию различных методик Data Science маркетологи могут узнать о том, какой именно контент интересен конкретной аудитории, тем самым избегая лишних затрат на производство бесполезного и неактуального продукта);

3) персонализация и подробный анализ аудитории (современные методы науки о данных позволяют детально изучать целевую аудиторию, а именно: их интересы, поведение, наиболее предпочтительные каналы коммуникации). Это в некоторой степени позволяет компании усилить персонализацию предложений, так как она уже будет информирована о том, где клиент наиболее часто реагирует на рекламу, какие объявления его интересуют, как правильно организовать кросс-продажи и допродажи);

4) детальная сегментация аудитории по сотням различных параметров, чтобы не оставить без внимания какую-либо группу, которая в будущем может оказаться наиболее перспективной и мобильной);

5) оптимизация площадок (современные технологии Data Science позволяют выявить, на каких именно интернет-площадках, сайтах, приложениях и социальных сетях пользователи ведут себя

наиболее активно, а также определить, какие методы стоит использовать, чтобы улучшить уже существующий сервис. Data Science позволяет выявить, какой контент является наиболее удачным, а какой, наоборот, неэффективным и бесполезным, какие материалы потенциальные клиенты дочитывают до конца с интересом, а какие просто пролистывают).

## ЗАКЛЮЧЕНИЕ

Защита данных является ключевым аспектом, который обязательно следует учитывать при работе с экономической статистикой. Все данные должны храниться в безопасных и конфиденциальных базах во избежание утечек ценной информации. Изучение DS, а также разработка новых моделей Data Science полезны для современной экономики и бизнеса, поскольку позволяют:

1) автоматизировать и упростить множество процессов и вычислений, позволяющих уменьшить издержки и повысить эффективность бизнеса;

2) оценить текущие и будущие экономические и финансовые условия путем допуска сложных нелинейных отношений между прогнозами;

3) прогнозировать рентабельность и доходность фирм, правильно определять дальнейшие векторы развития организаций благодаря эффективной работе с колоссальными массивами данных;

4) предоставлять клиентам нестандартные решения, разработанные на базе искусственного интеллекта (ИИ), что в свою очередь может способствовать повышению конкурентоспособности фирмы, повышению качества товаров и услуг различных категорий;

5) интегрировать и дополнять официальные ключевые статистические показатели с использованием новых неструктурированных источников больших данных в режиме реального времени.



б) оптимизировать маркетинговые площадки, персонализацию предложений, проведение эффективной аналитической работы по клиентам и продуктам.

#### Список источников

1. *Barbaglia L., Consoli S., Manzan S., Reforgiato Recupero D., Saisana M., Tiozzo Pezzoli L.* Data Science technologies in economics and finance: a gentle walk-in // *Data Science for Economics and Finance*. 2021. pp. 1-17. [https://doi.org/10.1007/978-3-030-66891-4\\_1](https://doi.org/10.1007/978-3-030-66891-4_1)
2. *Гатман Ал.Дж.* Разберись в Data Science: как освоить науку о данных и научиться думать как эксперт. М.: Эксмо, 2023. 304 с.
3. *Коэльо Л.П., Ричарт В.* Построение систем машинного обучения на языке Python. М.: ДМК Пресс, 2016. 302 с.
4. *Вандер П.Дж.* Python для сложных задач: наука о данных и машинное обучение. СПб.: Питер, 2018. 576 с.
5. *Назаров А.Д., Плотников В.А.* Data Science и обеспечение экономической безопасности в эпоху цифровой экономики // *Известия Санкт-Петербургского государственного экономического университета*. 2023. № 4 (142). С. 151-155. <https://elibrary.ru/sgqzic>

#### References

1. *Barbaglia L., Consoli S., Manzan S., Reforgiato Recupero D., Saisana M., Tiozzo Pezzoli L.* (2021). Data Science technologies in economics and finance: a gentle walk-in. *Data Science for Economics and Finance*, pp. 1-17. [https://doi.org/10.1007/978-3-030-66891-4\\_1](https://doi.org/10.1007/978-3-030-66891-4_1)
2. *Gatman Al.Dzh.* (2023). *Razberis' v Data Science: kak osvoit' nauku o dannykh i nauchit'sya dumat' kak ehkspert* [Figure out Data Science: how to master the science of data and learn to think like an expert]. Moscow, Eksmo Publ., 304 p. (In Russ.)
3. *Koehl'o L.P., Richart V.* (2016). *Postroenie sistem mashinnogo obucheniya na yazyke Python* [Building Machine Learning Systems in Python]. Moscow, DMK Press, 302 p. (In Russ.)
4. *Vander P.Dzh.* (2018). *Python dlya slozhnykh zadach: nauka o dannykh i mashinnoe obuchenie* [Python for Complex Tasks: Data Science and Machine Learning]. St. Petersburg, Piter Publ., 576 p. (In Russ.)
5. *Nazarov A.D., Plotnikov V.A.* (2023). Data Science and economic security in the digital economy age. *Izvestiya Sankt-Peterburgskogo gosudarstvennogo ehkonomicheskogo universiteta = Bulletin of the St. Petersburg State University of Economics*, no. 4 (142), pp. 151-155. (In Russ.) <https://elibrary.ru/sgqzic>

---

#### Информация об авторе

**Дудова Алиса Романовна**, студентка экономического факультета, Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, г. Саратов, Российская Федерация, <https://orcid.org/0009-0006-2089-7080>, [alisiadudova@mail.ru](mailto:alisiadudova@mail.ru)

#### Information about the author

**Alisa R. Dudova**, Student of Economics Faculty, Saratov State University, Saratov, Russian Federation, <https://orcid.org/0009-0006-2089-7080>, [alisiadudova@mail.ru](mailto:alisiadudova@mail.ru)

---

Статья поступила в редакцию / The article was submitted 26.10.2023

Одобрена после рецензирования / Approved after reviewing 22.01.2023

Принята к публикации / Accepted for publication 01.03.2024