Посвящаю эту книгу
своему покойному дедушке
Владимиру Замашикову

Содержание

Об авторе

Никита Ткаченко возглавляет консалтинговое агентство Evalyn, специализирующееся на аудите обслуживания клиентов на основе искусственного интеллекта (ИИ) и технологических решениях по анализу данных. Он помогает организациям любого размера использовать ИИ и данные, чтобы оптимизировать процессы принятия решений, систематизировать операции и улучшать качество обслуживания покупателей. Обладая солидным опытом в области изысканий и аналитики, Никита также преподает курсы по инструментам исследования, руководит студентами и проводит академические исследования в Университете Сан-Франциско.

О техническом рецензенте

Сиджо Валаяккад Маникандан – опытный специалист в области искусственного интеллекта и науки о данных. Обладает обширным опытом управления крупномасштабными проектами в области науки о данных для корпораций из списка Fortune 500. Сиджо получил степень магистра наук в области деловой аналитики в знаменитом Техасском университете в Остине и степень бакалавра инженерных наук в Институте технологии и науки Бирлы в Пилани, Индия. Для достижения исключительных результатов он сочетает свои академические знания и практический опыт.

Сиджо внес значительный вклад в развитие науки о данных и научных исследований благодаря не только своим профессиональным и академическим достижениям, но и своей преданности сообществу. Он входит в состав нескольких авторитетных организаций, таких как Американская статистическая ассоциация, проводит независимые исследования, а также активно рецензирует научные и профессиональные монографии, был наставником молодых исследователей данных и руководил молодыми стартапами. Более того, его опыт позволил ему войти в состав жюри нескольких престижных премий, включая Webby Awards.

Будучи целеустремленным исследователем и идейным лидером, Сиджо продолжает оказывать глубокое влияние на сферу науки о данных, вдохновляя новое поколение исследователей данных. Его вклад продвинул междисциплинарную область науки о данных и научных исследований вперед и помог определить будущее этой быстро развивающейся отрасли, что сделало его бесценным активом для сообщества и провидцем в своей сфере.

Признательности

Посвящаю эту книгу своему покойному дедушке Владимиру Замашикову.

Выражаю свою глубочайшую благодарность Томасу Вайнанди, который убедил меня превратить черновой вариант в опубликованную книгу и направлял меня на протяжении всего процесса. Без него эта книга была бы невозможна.

Я глубоко признателен Парсу Рахими за его неизменную поддержку, тщательный анализ всех моих глав, правок и идей, а также за предоставленные художественные работы с элементами фантастики.

Также выражаю благодарность Алессандре Кассар и Майклу Джонасу за их поддержку на протяжении всего моего академического пути. Особая благодарность Питеру Лоренцену, Керу Гиббсу и Джесси Анттила-Хьюзу за их неоценимое наставничество и поддержку.

Я признателен Шивани Шукле, Брюсу Вайдику, Мише Цыгману, Марио Лиму, Стиву Треттелу, Мехмету Эмре, Эндрю Хоббсу, Арману Хачияну, Робизону Хубулашвили и Конраду Пошу за рецензирование глав и предоставленную поддержку. Кроме того, я благодарен преподавательскому составу, библиотекарям, сотрудникам писательского центра и Центра деловых исследований и инноваций (CBSI) Университета Сан-Франциско за их наставничество.

Гордон Гетти заслуживает благодарности за то, что вдохновил меня более уверенно писать о своих навыках.

В процессе публикации мне помогал коллектив издательства APress, в особенности Шауль Элсон, редактор-заказчик, Лаура Берендсон, редактор-разработчик, и Гриффин Винклер, ассистент редактора. Мой технический рецензент Сиджо Маникандан предоставил комментарии, которые значительно обогатили книгу дополнительными примерами и пояснениями. Выражаю особую благодарность всем сотрудникам издательства APress, благодаря чьим усилиям эта книга стала реальностью.

Выражаю глубокую признательность Джону Четвинду, Джейку Консгроуву, Тимофею Лопухову и Александру Рому за их поддержку, любопытство и доверие к моей работе.

Должен поблагодарить свою семью, включая мою маму Екатерину Ткаченко, чья вера в меня вдохновила меня на запуск этого проекта, моего отца Антона Ткаченко, приемную мать Дарью Ткаченко, а также бабушку и дедушку Валентину и Александра Ткаченко и Раису Замашикову.

Особая благодарность Анастасии Терновской за ее непоколебимую веру в меня.

Я очень признателен Марии Аксутенко за создание иллюстрации «кошка-исследователь».

Особая благодарность также коллективу Posit за их неоценимый вклад и поддержку сообщества.

Наконец, выражаю искреннюю признательность сообществу разработчиков открытого исходного кода и всем тем, кто щедро делился своими знаниями в интернете. Вы сделали это реальным, и я рад возможности принять участие в обсуждении. Надеюсь, что эта книга окажется полезной на вашем пути к овладению методами научных изысканий.

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Введение

Эта книга родилась из моих разочарований и опыта в сфере высшего образования и профессиональной деятельности. Она была написана на основе заметок и материалов весеннего курса 2023 года по разработке опросов под влиянием восторженных откликов и содержательных вопросов моих студентов.

Молодые специалисты в области исследования данных, как правило, на своих занятиях знакомятся с моделями, экспериментами и теориями и часто возвращаются к этим знаниям позже. Однако для проведения высококачественных изысканий и анализа требуется более глубокое понимание инструментов и того, «как» это делается, а не только того, «что» делается и «почему». Эти знания часто выходят за пределы того, чему учат в рамках стандартной учебной программы. В данной книге я стремлюсь преодолеть этот пробел, помогая вам перейти от знания того, что вы хотите делать, к пониманию того, как это делать. В свои главы я вложил сотни часов разочарования, так что вам не придется проходить этот путь самостоятельно.

Эта книга не является исчерпывающим руководством; если вы ищете именно его, то, возможно, вам лучше поискать в другом месте. Данная же книга может служить картой, очерчивающей необходимые инструменты и темы для вашего исследовательского путешествия. Ее цель состоит в том, чтобы развить вашу интуицию и подсказать места, где найти более подробную информацию. Главы намеренно сделаны краткими, а материал излагается по существу. Они призваны раскрывать и просвещать, а не утомлять. Вы узнаете об эффективном управлении данными, проведении воспроизводимых исследований, составлении обзора научных публикаций и методах написания текстов, а также об эффективной визуализации данных.

Эта книга, изначально написанная под впечатлением от моего обучения в аспирантуре по экономике, представляет ценность для всех дисциплин. В ней содержатся важные сведения для всех, кто работается с данными, от гуманитарных наук до анализа данных и естественных наук. Независимо от того, оттачиваете ли вы свой опыт в области анализа данных или являетесь в ней новичком, настоящая книга обещает предложить вам нечто ценное.

Приведенные в книге примеры в основном написаны на языке R, что предполагает базовое понимание языка, но это не обязательно. Некоторые главы, в особенности те, которые посвящены теории, вообще не требуют знаний

в области программирования. Материал оказался полезен широкому кругу читателей, включая веб-разработчиков, математиков, аналитиков данных и экономистов. Книга организована таким образом, чтобы быть всеохватывающей и давать информацию независимо от вашего уровня владения программированием или профессиональной подготовки.

Ее структура позволяет выбирать гибкие пути чтения; можно просматривать главы последовательно, чтобы усваивать материал систематически либо переходить непосредственно к наиболее актуальным для вас темам.

Установка языка R и среды RStudio

Добро пожаловать в захватывающий мир анализа данных на языке R, мастерски выполненного специально для статистического анализа и визуализации данных. Удобный синтаксис и воспроизводимость данного языка делают его идеальным выбором как для новичков, так и для профессионалов. Однако прежде чем приступить к разведывательному анализу данных и моделированию, важно провести различие между языком программирования R и интегрированной средой разработки (IDE) RStudio, которая расширяет функциональные средства языка R.

Скачивание и установка языка R

Язык R поддерживается и распространяется через всеобъемлющую сеть архивов R (Comprehensive R Archive Network, аббр. CRAN), которая обеспечивает доступ к самой свежей версии и ресурсам.

Для пользователей macOS:

1. Перейдите на веб-сайт CRAN[1].
2. Кликните по **Download R for macOS** (Скачать R для macOS).
3. Выберите подходящую версию:
   - для Apple Silicon (например, M1, M2) надо скачать версию, в названии которой указано «-arm64» (например, R-4.2.2-arm64.pkg);
   - для компьютеров Mac на базе Intel надо выбрать версию без «-arm64» (например, R-4.2.2.pkg).
4. Следуйте инструкциям мастера установки. Обычно достаточно стандартных настроек.

Для пользователей Windows:

1. Перейдите на веб-сайт CRAN.
2. Выберите **Download R for Windows** (Скачать R для Windows).
3. Выберите **base** (базовая), а затем первую ссылку в верхней части страницы (например, Download R-4.2.2 for Windows).
4. Установщик проведет вас по всему процессу. В целях более эффективной установки придерживайтесь стандартных настроек.

---

[1] См. https://cran.r-project.org/.