



ТЕОРЕТИЧЕСКИЙ МИНИМУМ ПО BIG DATA

ВСЁ, ЧТО НУЖНО ЗНАТЬ О БОЛЬШИХ ДАННЫХ



Annalyn Ng & Kenneth Soo

Numsense!
Data Science for the
Layman

No Math Added

Анналин Ын, Кеннет Су

ТЕОРЕТИЧЕСКИЙ МИНИМУМ ПО BIG DATA

ВСЁ, ЧТО НУЖНО ЗНАТЬ О БОЛЬШИХ ДАННЫХ



Санкт-Петербург · Москва · Екатеринбург · Воронеж
Нижний Новгород · Ростов-на-Дону
Самара · Минск

2019

ББК 32.973.233-018
УДК 004.62
Ы11

Ын Анналин, Су Кеннет

Ы11 Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. — СПб.: Питер, 2019. — 208 с.: ил. — (Серия «Библиотека программиста»).

ISBN 978-5-4461-1040-7

Сегодня Big Data — это большой бизнес.

Нашей жизнью управляет информация, и извлечение выгоды из нее становится центральным моментом в работе современных организаций. Неважно, кто вы — деловой человек, работающий с аналитикой, начинающий программист или разработчик, «Теоретический минимум по Big Data» позволит не утонуть в бушующем океане современных технологий и разобраться в основах новой и стремительно развивающейся отрасли обработки больших данных.

Хотите узнать о больших данных и механизмах работы с ними? Каждому алгоритму посвящена отдельная глава, в которой не только объясняются основные принципы работы, но и даются примеры использования в реальных задачах. Большое количество иллюстраций и простые комментарии позволят легко разобраться в самых сложных аспектах Big Data.

16+ (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ББК 32.973.233-018
УДК 004.62

Права на издание получены по соглашению с Annalyn Ng и Kenneth Soo. Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

ISBN 978-9811110689 англ.

Authorized translation from the English language edition, titled «Numsense! Data Science for the Layman: No Math Added» (ISBN 9789811110689) by Annalyn Ng & Kenneth Soo

© 2017 to present

ISBN 978-5-4461-1040-7

© Перевод на русский язык ООО Издательство «Питер», 2019

© Издание на русском языке, оформление ООО Издательство «Питер», 2019

© Серия «Библиотека программиста», 2019

© Перевод с английского языка Тимохин А. В., 2018

Краткое содержание

Предисловие.....	12
Введение	16
Почему Data Science?.....	18
Глава 1. Об основах без лишних слов	21
Глава 2. Кластеризация методом k-средних.....	39
Глава 3. Метод главных компонент	51
Глава 4. Ассоциативные правила	65
Глава 5. Анализ социальных сетей	77
Глава 6. Регрессионный анализ.....	93
Глава 7. Метод k-ближайших соседей и обнаружение аномалий.....	107
Глава 8. Метод опорных векторов	117
Глава 9. Дерево решений.....	127
Глава 10. Случайные леса.....	137
Глава 11. Нейронные сети	149
Глава 12. A/B-тестирование и многорукие бандиты.....	167
Приложения	179
Глоссарий.....	188
Литература и ссылки на источники.....	199
Об авторах	204