

Опануй числа!

КУПИТИ

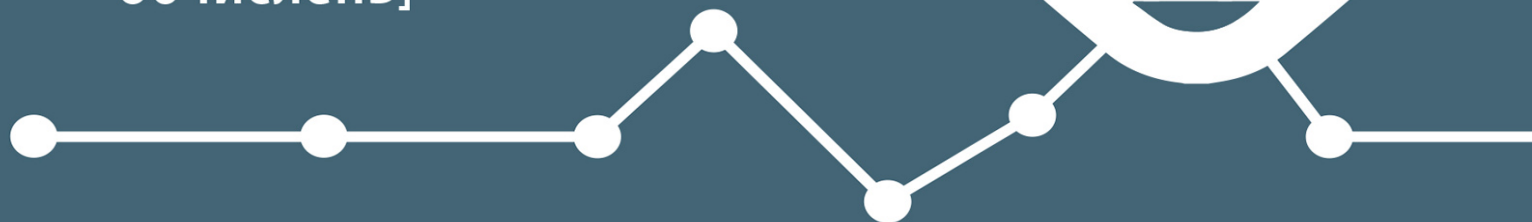
Сучасний світ перенасичений інформацією, зокрема даними. Величезною кількістю даних! Як серед них не загубитися? Як їх досягнути? Як використовувати в повсякденному житті та різноманітних галузях? Як аналізувати? Як інтерпретувати?

«Опануй числа!» - це зручний графічний опис ключових алгоритмів обробки даних, корисний як вступ для новачків у цій галузі, огляд для ділових людей, що працюють з аналітиками, чи стимул для тих, хто прагне знати, що відбувається з їхніми даними.

67294
ОПАЯЧУЙ
67294
ЧИСТА!

НАУКА ПРО ДАНІ ДЛЯ НЕФАХІВЦІВ

[без складних математичних
обчислень]



ВИДАВНИЧИЙ ДІМ
ФАБУЛА
#PRO

Анналін Нг та Кеннет Су

Вступне слово

Великі дані — це вже великий бізнес. Оскільки вони дедалі більше домінують у нашому житті, їхня монетизація зробилася основним завданням майже кожної організації. Методи розпізнавання закономірностей і прогнозування відкривають нові можливості для бізнесу. Прикладом слугують системи рекомендації товарів. Вони є безпрограшним варіантом як для покупців, так і для продавців, оскільки рекомендують першим товари, які можуть їх зацікавити, що, зі свого боку, приносить більший прибуток другим.

Але великі дані — це лише частина головоломки. Наука про дані, яка дає нам змогу аналізувати та використовувати їх, є багатогранною дисципліною, що охоплює машинне навчання, статистику та суміжні галузі математики. Зауважте, що машинне навчання посідає особливо важливе місце в цьому описі, оскільки є основним рушієм, який уможлиблює розпізнавання шаблонів і прогнозування. Алгоритми машинного навчання, що лежать в основі науки про дані, у поєднанні з даними здатні призвести до неоціненних відкриттів і нових способів використання інформації, яка вже є у нас під рукою.

Аби усвідомити, як наука про дані рухає сучасну інформаційну революцію, непосвяченим слід краще зрозуміти цю сферу. Проте побоювання щодо необхідних навичок змушують декого уникати цієї галузі — і це попри високий попит на грамотність у роботі з даними.

Саме тут на допомогу приходить «Опануй числа! Наука про дані для нефахівців». Після ознайомлення з книжкою Анналін Нг і Кеннета Су я переконався, що вона цілком відповідає своїй назві. Автори висвітлюють науку про дані саме для нефахівців, тому складну математику, яку подано на високому рівні, навмисно не описано детально. Але це не означає, що зміст книжки вихолощено. Насправді інформація, що міститься в цій праці, є ґрунтовною, і її перевага полягає саме в тому, що вона чітка й лаконічна.

«Чи є користь від такого підходу?» — запитаєте ви. Насправді її дуже багато! Я стверджую, що для неспеціаліста такий підхід є найкращим.

Уявіть нефахівця, який зацікавлений у вивченні роботи автомобіля. Загальний огляд складових частин автівки, імовірно, буде менш складним, ніж технічний текст із фізики горіння. Те саме стосується вступного курсу науки про дані: якщо ви зацікавлені у вивченні цієї галузі знань, легше почати із загальних понять, перш ніж заглиблюватися в математичні формули.

Вступ до книжки на кількох сторінках ознайомить непосвячених із фундаментальними поняттями. Це гарантує, що кожен почне читати, маючи загальне уявлення, що таке наука про дані. Важливі поняття, як-от вибір алгоритму, що їх часто не беруть до уваги у вступних матеріалах, також розглянуто одразу — це прищеплює читачам відчуття нагальної потреби сформулювати власне розуміння цієї галузі знань і надає всеосяжну основу для досягнення такої мети.

Існує безліч концепцій, які Анналін і Кеннет могли би вважати гідними висвітлення у своїй книжці, а також чимала кількість способів їх презентації. Автори зосередилися переважно на алгоритмах машинного навчання, що мають найбільше значення для науки про дані, а також додали кілька сценаріїв на основі конкретних завдань — і це було чудовим рішенням. Випробувані та перевірені алгоритми, як-от кластеризація методом k -середніх, дерево ухвалення рішень і метод k -найближчих сусідів, отримали належну оцінку. Також пояснено новітні алгоритми класифікації та ансамблеві алгоритми, наприклад, метод опорних векторів, який часто відлякує своєю складною математикою, і метод випадкових лісів. Також розглянуто нейронні мережі, які є рушійною силою нинішнього шаленого захоплення глибоким навчанням.

Поєднання алгоритмів з інтуїтивно зрозумілими прикладами використання — ще одна перевага цієї книжки. Під час пояснення методу випадкових лісів у контексті прогнозування злочинності або ж методу кластеризації в контексті профілювання кіноманів обрані приклади забезпечують чіткість і практичне розуміння. Водночас відсутність будь-яких згадок про вищу математику підтримує зацікавленість і мотивацію до того, що, як передбачається, є першим кроком читача у вивченні науки про дані.

Я дуже рекомендую книжку «Опануй числа! Наука про дані для нефахівців» новачкам, які шукають точку входу в науку про дані або алгоритми, що нею керують. Мені не спадає на думку інше подібне видання. З появою цієї книжки більше немає причин, аби математика завадила вам здобувати знання.

Метью Мейо, науковець із даних
і редактор сайту KDnuggets
[@mattmayo13](#)

Передмова

«Опануй числа! Наука про дані для нефахівців» написано двома ентузіастами науки про дані — Анналін Нг (Кембриджський університет) і Кеннетом Су (Стенфордський університет).

Ми помітили: хоча науку про дані дедалі частіше використовують для полегшення ухвалення рішень безпосередньо на робочому місці, багато хто й досі мало знає про цю галузь знань. Тому ми зібрали навчальні інструкції в одній книжці, щоби більше людей мали змогу навчатися — байдуже, чи то студент, підприємливий бізнесмен або ж будь-хто з допитливим розумом.

Кожна така інструкція пояснює важливі функції та передумови використання окремого методу науки про дані, не використовуючи складної математики й зайвої наукової термінології. Ми також ілюструємо ці методи даними та прикладами з реального життя.

Ми не змогли б написати цієї книжки самотужки.

Висловлюємо вдячність нашій редакторці й добрій подрузі Соні Чан за те, що вона вміло поєднала наші різні стилі написання та забезпечила плавність викладу матеріалу.

Ми дякуємо Дорі Тан, нашому талановитому графічному дизайнеру, за макет книжки та дизайн обкладинки оригінального видання.

Щира подяка нашим друзям Мішель По, Деннісу Чу та Маркові Хо за безцінні поради щодо того, як поліпшити зрозумілість контенту.

Ми також дякуємо професору Лонгу Нгуену (Мічиганський університет, місто Анн-Арбор), професору Персі Лянгу (Стенфордський університет) і доктору Михалу Косінському

(Стенфордський університет) за те терпіння, яке вони виявляли до нас, а ще за те, що ділилися з нами своїми експертними порадами.

Нарешті, ми хотіли би подякувати одне одному за те, що час від часу сперечалися, як це роблять хороші друзі, але завжди залишалися разом до кінця, аби завершити розпочату справу.

Чому наука про дані?

Уявіть себе молодим лікарем.

До вас звертається пацієнт зі скаргами на задишку, біль у грудях і періодичну печію. Ви його обстежуєте. Виявляється, його кров'яний тиск і частота серцевих скорочень у нормі, а ще в нього не було жодних попередніх захворювань.

Під час огляду ви виявляєте, що пацієнт повненький. Оскільки його симптоми характерні для людей з надмірною вагою, ви запевняєте його, що все під контролем, і рекомендуєте знайти час для фізичних вправ.

Подібна ситуація занадто часто призводить до встановлення хибного діагнозу, оскільки пацієнти із серцево-судинними захворюваннями мають симптоми, схожі на ознаки звичайного ожиріння. Лікарі часто не проводять подальшого обстеження, яке могло би виявити більш серйозне захворювання.

Судження кожної людини обмежені її суб'єктивним досвідом і неповнотою знань. Це погіршує процес ухвалення рішень і може, як у випадку з недосвідченим лікарем, завадити подальшому обстеженню, яке здатне допомогти дійти більш точних висновків.

Саме тут стане в пригоді наука про дані.

Замість покладатися на судження однієї людини, методи цієї науки дають нам змогу використовувати інформацію з більшої кількості джерел даних для ухвалення кращих рішень. Наприклад, ми в змозі перевірити історії хвороби пацієнтів зі схожими симптомами, аби на основі цієї статистики виявити можливі діагнози, на які раніше не звертали уваги.

Завдяки сучасним обчислювальним технологіям і передовим алгоритмам ми здатні:

- виявляти приховані тенденції у великих масивах даних;
- використовувати тенденції для прогнозування;
- обчислювати ймовірність кожного можливого результату;
- швидко отримувати точні результати.

Цю книжку написано простою мовою (без складної математики) як легкий вступ до науки про дані та її алгоритмів. Аби допомогти вам зрозуміти ключові поняття, ми використовуємо інтуїтивно зрозумілі пояснення та багато візуальних зображень.

Кожному алгоритму присвячено окремий розділ, у якому наведено приклад з реального життя, що пояснює, як працює алгоритм. Дані, використані для цих прикладів, доступні в інтернеті, а їхні джерела вказано наприкінці книжки.

Якщо вам потрібно повторити вивчений матеріал, зверніться до висновків після кожного розділу. Наприкінці книжки ви також знайдете зручні довідкові таблиці, у яких порівнюються переваги та недоліки кожного алгоритму. Також видання містить глосарій загальновживаних термінів.

За допомогою цієї книжки ми сподіваємося дати вам практичне розуміння науки про дані, аби ви могли скористатися її перевагами для ухвалення кращих рішень.

Тож почнімо.

1. Коротко про основи

Аби краще зрозуміти, як працюють алгоритми науки про дані, необхідно розпочати з основ. Саме тому це найдовший розділ у книжці. Він удвічі довший за інші, які безпосередньо присвячені алгоритмам. Проте завдяки цьому вступу ви отримаєте ґрунтовне уявлення про фундаментальні кроки, які присутні майже в усіх дослідженнях у галузі науки про дані. Ці базові процеси допоможуть вам оцінити контекст, а також обмеження, які виникають у процесі вибору відповідних алгоритмів для використання в дослідженнях.

Існують чотири ключові етапи дослідження в галузі науки про дані. Передусім дані мають бути оброблені та підготовлені до аналізу. Потім на основі вимог нашого дослідження складається короткий список відповідних алгоритмів. Після цього параметри алгоритмів мають бути налаштовані для оптимізації результатів. Нарешті, усе це завершується побудовою моделей, які згодом порівнюють для вибору найкращої.

1.1. Підготовка даних

Наука про дані пов'язана з даними. Якщо їхня якість низька, навіть найскладніший аналіз дасть лише непереконливі результати. У цьому розділі ми розглянемо основні формати даних, які зазвичай використовуються в аналізі, а також методи оброблення даних для поліпшення результатів.

Формат даних

Аби представити дані для аналізу, найчастіше використовують табличну форму (табл. 1). Кожен рядок є *елементом даних*, що описує окреме спостереження, а кожен стовпчик — змінною, що описує елемент даних. Змінні також відомі під назвами *атрибути*, *ознаки* або *розмірності*.

Змінні						
Елементи даних	Номер покупки	Біологічний вид покупця	Дата	Фрукти	Риба	Загальна сума покупок, \$
	1	пінгвін	1 січня	1	так	5,3
	2	ведмідь	1 січня	4	так	9,7
	3	кролик	1 січня	6	ні	6,5
	4	кінь	2 січня	6	ні	5,5
	5	пінгвін	2 січня	2	так	6
	6	жирафа	3 січня	5	ні	4,8
	7	кролик	3 січня	8	ні	7,6
	8	кіт	3 січня	?	так	7,4

Таблиця 1. Уявний набір даних про купівлю тваринами продуктів у супермаркеті. Кожен рядок — це купівля, а кожен стовпчик містить інформацію про неї

Кінець безкоштовного уривку. Щоби читати далі, придбайте, будь ласка, повну версію книги.

Про авторів

Анналін Нг — спеціалістка зі штучного інтелекту в Google Cloud. Її кар'єра в галузі науки про дані охоплює комерційний і державний сектори; вона працювала в *Amazon*, *Disney Research*, а також в уряді Сінгапура, зокрема міністерстві праці та міністерстві оборони. Анналін отримала ступінь бакалавра в Мічиганському університеті (Анн-Арбор), де вона також працювала на волонтерських засадах викладачем статистики для студентів. Згодом здобула ступінь магістра в Кембриджському університеті.

Кеннет Су має 7-річний досвід застосування науки про дані в державній політиці в уряді Сінгапура, а зараз очолює команду з управління двосторонніми відносинами з європейськими партнерами. Під час пандемії COVID-19 він керував загальнонаціональними ініціативами з діджиталізації для сінгапурської програми «Розумна нація» та Офісу цифрового уряду, зокрема впровадженням системи відстеження контактів. Кеннет Су отримав ступінь магістра статистики в Стенфордському університеті та був найкращим студентом протягом усіх трьох років навчання на бакалавраті з математики, операційних досліджень, статистики та економіки (MORSE) в Університеті Ворика.

Аби краще опанувати науку про дані, відвідайте блог:
www.algobbeans.com

Анналін Нг, Кеннет Су

ОПАНУЙ ЧИСЛА! Наука про дані для нефахівців



Видавництво «Фабула»
2024

Оригінальна назва твору:
Numsense! Data Science for the Layman

© 2017 Annalyn Ng and Kenneth Soo

© О. Буйвол, пер. з англ., 2023

© ВД «Фабула», 2024

ISBN 978-617-522-178-5 (epub)

Усі права збережено. Жодна частина цієї книжки не може бути відтворена або передана в будь-якій формі або будь-якими засобами, електронними чи механічними, включно з фотокопією, записом чи будь-якою системою зберігання та пошуку інформації, без письмового дозволу власників авторських прав.

Електронна версія створена за виданням:

Нг Анналін

Н11 Опануй числа! Наука про дані для нефахівців / Анналін Нг та Кеннет Су / пер. з англ. О. Буйвол. — Харків : ВД «Фабула», 2024. —

168 с.

ISBN 978-617-522-177-8

Сучасний світ перенасичений інформацією, зокрема даними. Величезною кількістю даних! Як серед них не загубитися? Як їх осягнути? Як використовувати в повсякденному житті та різноманітних галузях? Як аналізувати? Як інтерпретувати?

«Опануй числа!» — це зручний графічний опис ключових алгоритмів обробки даних, корисний як вступ для новачків у цій галузі, огляд для ділових людей, що працюють з аналітиками, чи стимул для тих, хто прагне знати, що відбувається з їхніми даними.

УДК 519.25

Шановний читачу!

Спасибі, що придбали цю книгу.

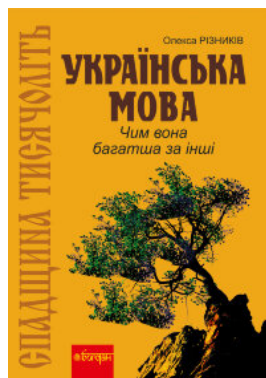
Нагадуємо, що вона є об'єктом Закону України «Про авторське і суміжні право», порушення якого карається за статтею 176 Кримінального кодексу України «Порушення авторського права і суміжних прав» штрафом від ста до чотирьохсот неоподатковуваних мінімумів доходів громадян або виправними роботами на строк до двох років, з конфіскацією та знищенням всіх примірників творів, матеріальних носіїв комп'ютерних програм, баз даних, виконань, фонограм, програм мовлення та обладнання і матеріалів, призначених для їх виготовлення і відтворення. Повторне порушення карається штрафом від тисячі до двох тисяч неоподатковуваних мінімумів доходів громадян або виправними роботами на строк до двох років, або позбавленням волі на той самий строк, з конфіскацією та знищенням всіх примірників, матеріальних носіїв комп'ютерних програм, баз даних, виконань, фонограм, програм мовлення, аудіо- і відеокaset, дискет, інших носіїв інформації, обладнання та матеріалів, призначених для їх виготовлення і відтворення. Кримінальне переслідування також відбувається згідно з відповідними законами країн, де зафіксовано незаконне відтворення (поширення) творів.

Щиро сподіваємося, що Ви з повагою поставитеся до інтелектуальної праці інших і ще раз Вам вдячні!

Рекомендована література



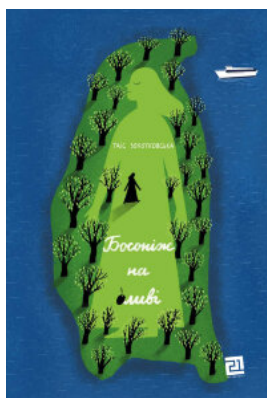
Порцелянова лялька.
Роман



Спадщина тисячоліть:
Українська мова. Чим
вона багатша за інші?



Три обіцянки. Роман



Босоніж на оливі



Київські канікули.
Роман



Країна призначення

Перейти до категорії
Математика



КУПИТИ