Міністерство освіти і науки України
Вінницький національний технічний університет

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ ТА МАШИННЕ НАВЧАННЯ
ЧАСТИНА 1
БАЗОВІ МЕТОДИ ТА ЗАСОБИ АНАЛІЗУ ДАНИХ

Навчальний посібник

Вінниця
ВНТУ
2021

УДК 004.652
І86

Рекомендовано до друку Вченою радою Вінницького національного технічного університету Міністерства освіти і науки України (протокол № 15 від 31.05.2021 р.)

Автори**:**
Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський

Рецензенти:
Р. Н. Квєтний, доктор технічних наук, професор
Т. Б. Мартинюк, доктор технічних наук, професор
І. Г. Цмоць, доктор технічних наук, професор

І86 **Інтелектуальний** аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с.
ISBN 978-966-641-874-9

Посібник містить теоретичний матеріал і приклади розв’язування практичних задач з інтелектуального аналізу даних; цілі, практичні завдання, переліки контрольних питань та вимоги до знань студентів, потрібні для виконання лабораторних робіт, що стосуються першого модуля навчальної дисципліни «Інтелектуальний аналіз даних та машинне навчання2.

Розрахований на студентів комп’ютерних спеціальностей всіх форм навчання.

УДК 004.652

ISBN 978-966-641-874-9

ЗМІСТ

ВСТУП

Навчальний посібник призначений для студентів спеціальності 122 – «Комп'ютерні науки». Зміст навчального посібника відповідає плану і програмі дисципліни «Інтелектуальний аналіз даних та машинне навчання».

Аналіз даних – сфера математики та інформаційних технологій, яка займається побудовою і дослідженням математичних методів і обчислювальних алгоритмів отримання бази знань із експериментальних даних [1–3]. Дана сфера охоплює процеси збирання, структурування і моделювання даних, їх дослідження та фільтрації з метою отримання корисної інформації та прийняття рішень. Математичні методи побудови моделей на основі даних об'єднуються в науку, яка називається машинним навчанням [4]. Сукупність наук, спрямованих на методи і технології роботи з даними, називають Data Science [1–7].

У зв'язку зі стрімким розвитком інформаційних технологій за останні десятиліття дані природним чином стали накопичуватися в цифровому вигляді, і тому наукова галузь інтелектуального аналізу даних стала активно розвиватися. До теперішнього час накопичені дані досягли значних обсягів, за допомогою яких з'явилася можливість отримувати з них користь: знаходити в даних приховані залежності, а з їх допомогою прогнозувати нові результати і робити рекомендації, що дозволяють оптимізувати різні процеси і витрати ресурсів [8]. Методи аналізу даних і машинного навчання активно розвиваються і знаходять своє застосування не тільки в економіці, фізиці, але й у соціальних науках, журналістиці, лінгвістиці, юриспруденції, політології та гуманітарних науках тощо [9].

Тому актуальним є поява навчального посібника з дисципліни «Інтелектуальний аналіз даних», який дозволить формувати у студентів компетенції, пов'язані з напрямом науки Data Science. Отримані навички дозволять у перспективі швидко та ефективно інтегруватися в розв'язання професійних задач на стику предметних галузей та комп'ютерних технологій, які сьогодні є передовими, але вже в найближчій перспективі стануть звичною практикою.

Перша частина навчального посібника містить потрібний теоретичний матеріал з базових методів і засобів аналізу даних в програмі PYTHON. Також даний навчальний посібник містить велику кількість прикладів розв'язування практичних задач інтелектуального аналізу даних за допомогою вбудованих бібліотек програми PYTHON; цілі, практичні завдання, переліки контрольних питань та вимоги до знань студентів, потрібні для виконання лабораторних робіт, що віднесені, згідно з навчальною програмою, до першого модуля дисципліни «Інтелектуальний аналіз даних та машинне навчання».

1 МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА АНАЛІЗ ДАНИХ ПРОГРАМНИМИ ЗАСОБАМИ PYTHON

1.1 Засоби роботи з багатовимірними масивами даних бібліотеки NumPy

Дана бібліотека призначена для різного роду математичних обчислень і роботи з багатовимірними масивами [1]. Вона досить потужна, має великий математичний потенціал, гнучкий і зрозумілий інтерфейс. Основна структура NumPy – це певний багатовимірний масив, який створюється на основі отриманих даних. Для створення деякого масиву потрібно імпортувати дану бібліотеку, після чого викликати функцію array, яка приймає на вхід дані у вигляді деякої послідовності

import numpy as np

x = np.array([1,2,3,4]).

За допомогою команди *x.dtype* можна визначити, які дані зберігаються в масиві. Тип даних може визначатися автоматично на основі того, які є вхідні дані, а також можна його вказувати при створенні масиву як додатковий аргумент

x = np.array([1,2,3,4], *dtype=np.int*64).

У даному випадку було створено певний одновимірний масив і для визначення розмірності потрібно використати команду *x.shape*. Атрибут *shape* повертає набір розмірності по всіх осях, які у нас є в масиві. Тобто, при двовимірному масиві було б отримано пару чисел, а саме: кількість рядків і число стовпців, але в даному випадку для одновимірного масиву отримують тільки одне значення.

Для створення двовимірного масиву потрібно ввести

m = np.array([[2,3,4], [5,6,7]]).

За допомогою NumPy масиви можна створювати декількома способами. Ми розглянули метод створення масиву на основі *list*, також можна спробувати створити масиви на основі вбудованої функції. Припустимо, якщо потрібно отримати масив, заповнений тільки одиницями, то можна скористатися функцією *ones*, яка як аргумент приймає розмір нашого масиву, в результаті ми отримуємо вже масив, заповнений тільки одиницями,

m = np.ones(5).

Таким же способом можна створити масив, заповнений тільки нулями,

m = np.zeros(2).

Також можна створити деяку одиничну матрицю, де по діагоналі будуть розміщені одиниці, а інші значення будуть нульовими

m = np.eye(6).

Для створення масиву, заповненого випадковими значеннями, можна скористатися методом *random*

*m = np.random.randint(*10, *size* = (2, 3)).

ЛІТЕРАТУРА

1. Уэс М. Python и анализ данных / пер. с англ. Слинкин А. А. М. : ДМК Пресс, 2015. 482 с.

2. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. СПб. : Питер, 2018. 576 с.: ил.

3. Копец Д. Классические задачи Computer Science на языке Python. СПб. : Питер, 2020. 256 с.

4. Брюс П., Брюс Э. Практическая статистика для специалистов Data Science / пер. с англ. СПб. : БХВ-Петербург, 2018. 304 с.

5. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. СПб. : Питер, 2017. 336 с.

6. Грае Дж**.** Data Science. Наука о данных с нуля / пер. с англ. СПб. : БХВ-Петербург, 2017. 336 с.: ил.

7. Y. Ivanchuk, K. Koval, A. Halianovska. The algorithm for simulation of a nonlinear dynamic Lorence System. Proc. XII International Scientific-Practical Conference «Internet-Education-Science-2020» dedicated to the 25-th anniversary of the Computer Science Department, 2020 May 26–29, P. 123–124.

8. Моделювання та оптимізація систем : підручник / Дубовой В. М., Квєтний Р. Н., Михальов О. І., Усов А. В. Вінниця : ПП «ТД «Едельвейс», 2017. 804 с.

9. Месюра В. І., Ваховська Л. М. Основи проектування систем штучного інтелекту. Лабораторний практикум. Частина 1. Способи подання задач і пошуку розв’язків. Лабораторний практикум. Вінниця : ВНТУ, 2009. 108 с.

10. Polhul T., Yarovyi A. «Development of a method for fraud detection in heterogeneous data during installation of mobile applications». (2019) Eastern-European Journal of Enterprise Technologies, 1 (2–97), pp. 65–75. – DOI: https://doi.org/10.15587/1729-4061.2019.155060.

11. Польгуль Т. Д., Яровий А. А. Аналіз різнорідних даних в інтелектуальних системах виявлення шахрайства. Вісник Вінницького політехнічного інституту. 2019. № 2 (143). С. 78–90. – DOI: https://doi.org/10.31649/1997-9266-2019-143-2-78-90.

12. M. Granik, V. Mesyura and A. Yarovyi, «Determining Fake Statements Made by Public Figures by Means of Artificial Intelligence,» 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 2018, pp. 424–427. – DOI: https://doi.org/ 10.1109/STC-CSIT.2018.8526631.

13. Яровий А. А. Методи та засоби організації високопродуктивних паралельно-ієрархічних обчислювальних систем із рекурсивною архітектурою : монографія. Вінниця : ВНТУ, 2016. 363 с.

Навчальне видання

Іванчук Ярослав Володимирович
Месюра Володимир Іванович
Яровий Андрій Анатолійович
Манжілевський Олександр Дмитрович

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ ТА МАШИННЕ НАВЧАННЯ
ЧАСТИНА 1
БАЗОВІ МЕТОДИ ТА ЗАСОБИ АНАЛІЗУ ДАНИХ

Навчальний посібник

Рукопис оформив *Я. Іванчук*

Редактор *В. Дружиніна*

Оригінал-макет підготовила *Т. Криклива*

Підписано до друку 03.11.2021 р.
Формат 29,7×42¼. Папір офсетний.
Гарнітура Times New Roman.
Друк різографічний. Ум. друк. арк. 4,14.
Наклад 50 (1-й запуск 1–21) пр. Зам. № 2021-114.

Видавець та виготовлювач
Вінницький національний технічний університет,
інформаційний редакційно-видавничий центр.
ВНТУ, ГНК, к. 114.
Хмельницьке шосе, 95,
м. Вінниця, 21021.
Тел. (0432) 65-18-06.
press.vntu.edu.ua;
E-mail: kivc.vntu@gmail.com.
Свідоцтво суб'єкта видавничої справи
серія ДК № 3516 від 01. 07.2009 р