

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДНУ «ІНСТИТУТ МОДЕРНІЗАЦІЇ ЗМІСТУ ОСВІТИ»
WARSAW UNIVERSITY OF TECHNOLOGY (М. ВАРШАВА, ПОЛЬЩА)
ІНСТИТУТ ПРОБЛЕМ МАТЕМАТИЧНИХ МАШИН І СИСТЕМ
НАЦІОНАЛЬНОЇ АКАДЕМІЇ НАУК УКРАЇНИ
ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
НАЦІОНАЛЬНИЙ ТРАНСПОРТНИЙ УНІВЕРСИТЕТ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ



Перша міжнародна
науково-практична конференція

«Штучний інтелект та інформаційні технології»

3–4 червня 2024 р.

Київ НУХТ 2024

УДК 004

Наукові праці Першої міжнар. наук.-практ. конф. «Штучний інтелект та інформаційні технології» (АІТ-2024), 3–4 червня 2024 р. (Київ, Україна). К. : НУХТ, 2024. 422 с.

У працях конференції наведено доповіді за напрямками:

- тенденції та досягнення в розробленні й застосуванні методів і практичних інструментів штучного інтелекту;
- інтелектуальні системи управління та аналізу даних;
- використання інформаційних технологій та штучного інтелекту в освіті;
- кіберзахист критичної інформаційної інфраструктури;
- використання інтернет-речей у науці й виробництві;
- математичне моделювання складних об'єктів.

Праці конференції будуть корисні науковим та інженерно-технічним працівникам, студентам ЗВО та всім, хто цікавиться сучасними інформаційними системами та телекомунікаційними технологіями.

Подано в авторській редакції.

Автори матеріалів несуть повну відповідальність за достовірність наведеної інформації та відповідність матеріалів нормам законодавства, моралі й етики.

ISBN 978966-612-323-0

© НУХТ, 2024

UDC 004

Proceedings of the 1st international scientific and practical conference «Artificial Intelligence & Information Technology» (AIIT-2024), June 3–4, 2024 (Kyiv, Ukraine). Kyiv : NUFT, 2024. 422 p.

The proceedings contain papers on the following topics:

- trends and achievements in the development and application of methods and practical tools of artificial intelligence;
- intelligent systems for data management and analysis;
- using information technology and artificial intelligence in education;
- cybersecurity of critical information infrastructure;
- using the Internet of Things (IoT) in science and production;
- mathematical modeling of complex objects.

The collection will be useful to scientists, researchers, students, and everyone interested in modern information technology and artificial intelligence.

Submitted in the authors' edition.

The authors are fully responsible for the accuracy of provided information, as well as for the papers' compliance with the laws, morals and ethics.

ISBN 978966-612-323-0

© NUFT, 2024

38. Юхта М. А. Сучасні інтелектуальні технології в Україні і світі.....	93
39. Яненко С. Л. Дипфейки: загрози та методи боротьби.....	95
40. Яценко Т. Ю. Тенденції та досягнення в розробленні й застосуванні методів і практичних інструментів штучного інтелекту.....	97
2 СЕКЦІЯ: ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ.....	99
41. Білодід Д. В. Інтелектуальний аналіз даних мовою програмування JavaScript.....	100
42. Бойко І. О. Роль штучного інтелекту в аналізі великих даних.....	102
43. Вороньков Д. В., Харкянен О. В. Використання мови Python для кластеризації методом k-середніх.....	104
44. Вороньков Є. В., Харкянен О. В., Костіков М. П. огляд бібліотек мови Python для розв'язання задач аналізу даних.....	107
45. Грама М. П., Серветник М. В. Застосування інтелектуальних систем обробки інформації як засіб розв'язання проблем управління сучасним підприємством.....	110
46. Гундар О. А., Костіков М. П. Використання бібліотек Python для аналітики даних.....	112
47. Дзюбан І. І., Харкянен О. В. Аналіз великих даних у реальному часі: технології та алгоритми.....	114
48. Залецький І. М. Дослідження сучасних технологій для аналізу текстової інформації на онлайн-ресурсах у регулярній роботі менеджерів.....	116
49. Зубрецька Н. А., Ложніков В. А. Використання інструментів бізнес-аналітики при підготовці фахівців із інформаційних технологій.....	118

ОГЛЯД БІБЛІОТЕК МОВИ PYTHON ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ АНАЛІЗУ ДАНИХ

Вороньков Є. В., Харкянен О. В., Костіков М. П.

Національний університет харчових технологій, Київ, Україна

E-mail: zhenia.v.voronkov@gmail.com

Overview of Python Libraries for Data Analysis

Data analysis is a crucial component of modern research and business analytics. Python has become one of the most popular programming languages for data analysis due to its simplicity and extensive library support. This review highlights the importance of selecting the right Python libraries for various data analysis tasks, including data manipulation, visualization, machine learning, statistical analysis, and natural language processing. Understanding the capabilities and applications of these libraries is essential for efficient and effective data analysis.

Аналіз даних є ключовим компонентом сучасних досліджень і бізнес-аналітики. Python став однією з найбільш популярних мов програмування для аналізу даних завдяки своїй простоті та широкому набору бібліотек. Огляд охоплює основні бібліотеки Python для різних методів аналізу даних — від опрацювання і візуалізації до машинного навчання та обробки природної мови.

Обробка та маніпулювання даними. Pandas:

- призначення: обробка та аналіз структурованих даних;
- методи: фільтрація, групування, злиття, агрегація даних, робота з датами та часом;
- застосування: аналіз фінансових даних, підготовка даних для машинного навчання.

NumPy:

- призначення: обробка числових даних та виконання математичних операцій;
- методи: робота з багатовимірними масивами, математичні операції, лінійна алгебра, генерація випадкових чисел;
- застосування: наукові обчислення, створення алгоритмів машинного навчання.

Візуалізація даних. Matplotlib:

- призначення: створення графіків та візуалізацій;
- методи: лінійні графіки, гістограми, стовпчикові діаграми, scatter plots;
- застосування: візуалізація результатів експериментів, презентація даних.

Seaborn:

- призначення: статистична візуалізація даних;
- методи: кореляційні графіки, box plots, violin plots, heatmaps;

- застосування: візуалізація розподілів даних, дослідження зв'язків між змінними.

Машинне навчання. Scikit-learn:

- призначення: алгоритми машинного навчання та попередня обробка даних;
- методи: класифікація, регресія, кластеризація, зниження вимірності;
- застосування: прогнозування продажів, класифікація текстів.

TensorFlow та Keras:

- призначення: глибоке навчання та нейронні мережі;
- методи: створення, тренування та оцінка нейронних мереж;
- застосування: розпізнавання образів, обробка природної мови.

PyTorch:

- призначення: глибоке навчання та нейронні мережі;
- методи: динамічне обчислення графів, створення та тренування моделей;
- застосування: розпізнавання мовлення, генеративні моделі, статистичний аналіз.

Statsmodels:

- призначення: статистичне моделювання та тестування гіпотез;
- методи: регресійний аналіз, часові ряди, ANOVA;
- застосування: економічне моделювання, медична статистика.

SciPy:

- призначення: наукові та технічні обчислення;
- методи: інтеграція, оптимізація, інтерполяція, спеціальні функції;
- застосування: інженерні обчислення, фізичні моделі.

Обробка тексту та NLP (Natural Language Processing). NLTK (Natural

Language Toolkit):

- призначення: обробка природної мови;
- методи: токенізація, стемінг, лемматизація, парсинг;
- застосування: аналіз тексту, навчальні завдання.

SpaCy:

- призначення: обробка природної мови;
- методи: токенізація, POS-тегування, named entity recognition, синтаксичний аналіз;
- застосування: автоматизація обробки текстів, чат-боти.

Gensim:

- призначення: моделювання та семантичний аналіз;
- методи: Word2Vec, Doc2Vec, LDA (Latent Dirichlet Allocation);
- застосування: моделювання, аналіз великих текстових даних.

Розглянемо використання бібліотек мови Python для розв'язання задач аналізу даних в різних галузях людської діяльності.

1. Фінансовий сектор.

Pandas та NumPy: використовуються для аналізу фінансових даних, прогнозування ринкових тенденцій та оцінки ризиків. Pandas дозволяє ефективно маніпулювати великими обсягами даних, тоді як NumPy забезпечує швидке виконання математичних обчислень.

Scikit-learn: використовується для створення моделей машинного навчання для прогнозування цін акцій, кредитного скорингу та виявлення шахрайства.

2. Медицина та охорона здоров'я.

TensorFlow і Keras: використовуються для розробки моделей глибокого навчання, які допомагають в діагностиці захворювань, обробці медичних зображень та персоналізованій медицині.

Statsmodels: використовується для статистичного аналізу клінічних даних, моделювання результатів лікування та аналізу виживання пацієнтів.

3. Маркетинг та реклама.

Gensim: використовується для аналізу великих обсягів текстових даних, таких як відгуки клієнтів та дослідження даних з соціальних мереж. Допомагає в розумінні настроїв споживачів та виявленні трендів.

NLTK та SpaCy: використовуються для обробки текстових даних, автоматизації процесів аналізу відгуків та створення чат-ботів для взаємодії з клієнтами.

Python-бібліотеки для аналізу даних забезпечують широкий спектр інструментів для вирішення різноманітних задач. Вибір конкретної бібліотеки залежить від специфіки задачі та вимог до продуктивності, гнучкості та масштабованості. Pandas та NumPy є основою для багатьох інших бібліотек, тоді як TensorFlow, PyTorch та Scikit-learn пропонують потужні інструменти для машинного та глибокого навчання. Бібліотеки для обробки тексту, такі як NLTK та SpaCy, забезпечують ефективну роботу з текстовими даними.

Література

1. McKinney W. (2017) Python for Data Analysis. O'Reilly Media.
2. Harris C. R., Millman K. J., van der Walt S. J. et al. (2020) Array programming with NumPy. Nature.
3. Hunter J. D. (2007) Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering.
4. Pedregosa F., Varoquaux G., Gramfort A. et al. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.
5. Abadi M., Barham P., Chen J. et al. (2016) TensorFlow: A System for Large-Scale Machine Learning. OSDI.
6. Reimers N., Gurevych I. (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. EMNLP.