

виконання маніпуляцій із даними, їх пошуку і аналізу, коли на встановлення з'єднання з базою даних потрібно лише кілька рядків коду.

Такий зв'язок Python та MongoDB стає все популярнішим, оскільки обидва інструменти є вкрай потужними в області розробки та обробки даних. Python як мова програмування володіє зручним синтаксисом, що робить його обраною мовою для великої кількості розробників та аналітиків. MongoDB як БД типу NoSQL пропонує гнучку модель документів, що ідеально взаємодіє з динамічним та розширюваним характером Python. Це допомагає розробникам та аналітикам з легкістю працювати з неструктурованими даними, а також швидко й ефективно взаємодіяти з базою даних. Швидкість і гнучкість є ключовими факторами поєднання Python та MongoDB.

Підсумовуючи вкажемо, що інтеграція MongoDB у Data Science виходить за межі традиційного зберігання даних, слугуючи рушійною силою для інновацій. Окрім своєї фундаментальної ролі в управлінні даними, MongoDB надає переваги у вирішенні безпосередньо задач обробки та аналізу даних, поєднання великих наборів даних і машинного навчання.

Список використаних джерел

1. Why Use MongoDB: What It Is and What Are the Benefits. URL: <https://www.simplilearn.com/tutorials/mongodb-tutorial/what-is-mongodb>
2. Growing Significance of MongoDB in Data Science Field. URL: <https://www.edureka.co/blog/the-growing-significance-of-mongodb-in-data-science-field/>
3. Introduction to MongoDB for Data Science. URL: <https://www.knowledgehut.com/blog/data-science/mongodb-for-data-science>
4. Exploring MongoDB for Data Science and Python. URL: <https://www.linkedin.com/pulse/exploring-mongodb-data-science-python-aritra-pain-3rd-year/>

УДК 004.04

*Підруцький Д. А., здобувач 3 курсу спеціальності 122 Комп'ютерні науки,
Січко Т. В., канд. техн. наук, доцент, доцент кафедри інформаційних технологій*

BIG DATA В ОБРОБЦІ І АНАЛІЗІ ДАНИХ З РІЗНИХ СФЕР

Донецький національний університет імені Василя Стуса, м. Вінниця

У сучасному світі, охопленому хвилею технологічного розвитку, важко уявити галузь, яка залишається поза впливом великих обсягів даних, або так

званих Big Data. З допомогою високотехнологічних інструментів і методів обробки даних Big Data стає ключовим елементом у різних сферах – від бізнесу та медицини до науки і громадського управління.

Метою цього дослідження є ретельний аналіз та систематизація можливостей використання технологій Big Data в різних сферах. Зокрема, дослідження спрямоване на визначення самих технологій, того, як саме їх можливо застосовувати, їх методів пов'язаних із даними, а також можливих ризиків під час застосування цих технологій.

Для досягнення визначеної мети поставлено такі завдання:

- надати визначення технологій Big Data;
- визначити основні методи аналізу Big Data;
- визначити ключові аспекти використання технологій Big Data в різних сферах.

Big Data (укр. – Великі дані) – це великий масив структурованої та неструктурованої інформації, а також інструменти, підходи, методи обробки і зберігання даних. Важливість великих даних залежить не тільки від їх кількості, а й від того, як компанія їх інтерпретує та використовує. Через об'єм та різноманітність даних обробляти їх традиційним програмним забезпеченням неможливо [1].

Дані в галузі Big Data класифікуються згідно з «П'ятьма V», які визначають п'ять ключових характеристик [2] (рис. 1):



Рисунок 1. Основні риси Big Data

Перша характеристика «Об'єм (Volume)» – являє собою величезний обсяг накопиченої бази даних, який практично неможливо обробити та зберегти за допомогою традиційних методів. Рис. 1 демонструє, що цей обсяг інформації вимагає альтернативного підходу та використання спеціалізованих інструментів.

Друга характеристика «Швидкість (Velocity)» – визначає темпи накопичення та обробки даних, які постійно збільшуються. Сучасний попит на технології обробки даних у режимі реального часу позначається таким самим темпом зростання.

Третя характеристика «Різноманітність (Variety)» – позначає можливість одночасної обробки як структурованої, так і неструктурованої інформації. Структурована інформація, яка становить лише 5 % від загального обсягу, містить дані, які можна класифікувати, наприклад, інформацію з банківської бази даних. Неструктурована інформація, що становить решту 80 %, охоплює різноманітні дані, як-от фотографії, відео та інші, наприклад, із соціальних мереж.

Необхідність перевірки достовірності даних виражена у четвертій характеристиці – «Достовірність (Veracity)». Якість отриманої інформації може значно відрізнятися, що впливає на точність подальшого аналізу.

Остання, п'ята характеристика «Змінюваність (Variability)» – відображає невідповідність інформації, що може ускладнювати й іноді значно заважати процесам обробки та управління даними. Початок форми.

Тепер перейдемо до методів аналізу великих даних [3], оскільки великі обсяги інформації у самому своєму виді не мають значення для людини, необхідно провести їх аналіз для досягнення конкретної мети. Для обробки цих великих обсягів даних використовуються різноманітні інструменти, перелік яких постійно оновлюється. Серед них виділяють такі техніки та методики:

- класифікація (classification): використовується для передбачення поведінки споживачів у певному сегменті ринку;
- кластерний аналіз (cluster analysis): застосовується для класифікації об'єктів за групами, виявляючи їх спільні ознаки;
- масовий аналіз (crowdsourcing): використовується для збору інформації з різноманітних джерел;
- розробка даних (data mining): застосовується для виявлення раніше невідомих та корисних відомостей, які можуть бути використані для прийняття рішень у різних галузях;
- машинне навчання (machine learning): передбачає створення нейронних мереж, які самонавчаються та здатні якісно і швидко обробляти інформацію;
- обробка сигналів (signal processing): використовується для розпізнавання сигналів на тлі шуму та їх подальшого аналізу;
- навчання без нагляду (unsupervised learning): використовується для виявлення прихованих функціональних взаємозв'язків у даних;
- візуалізація (visualization): використовується для презентування результатів аналізу у вигляді діаграм і анімації.

Розглянемо сфери прямого застосування Big Data [4].

1. У медицині: шляхом аналізу обширної кількості медичних записів та обробки медичних знімків можна досягти точнішого й раніше поставленого діагнозу, глибшого розуміння природи різноманітних захворювань та розробки нових методів лікування і ліків.

2. У галузі сільського господарства відбувається справжня революція Big Data, що допомагає оптимізувати використання ресурсів для максимального збільшення врожаїв за мінімального впливу на екосистему та зниження вартості вирощених продуктів завдяки раціональному використанню обладнання і добрив.

3. У науці: завдяки аналізу великої кількості даних, отриманих від телескопів, НАСА може визначати хімічний склад атмосфер планет, віддалених на світлові роки, та робити припущення щодо їх придатності для життя.

4. Також є можливим прогнозування надзвичайних ситуацій: з використанням даних від численних сенсорів та супутників науковці можуть передбачати можливість землетрусів чи природних катаклізмів, а також моделювати поведінку людей під час надзвичайних ситуацій, що підвищує шанси на виживання.

5. Для запобігання злочинам: за допомогою нових технологій та аналізу обширних даних можливо автоматично виявляти фінансові махінації і відмивання грошей.

6. У торгівлі: торгові мережі використовують дані для випуску нових продуктів та впроваджують глобальні маркетингові кампанії для окремих сегментів покупців.

Хоча використання Big Data непереможно розширює можливості аналізу, це із собою це також несе певні ризики в конфіденційності користувачів, оскільки збір великої кількості особистих даних може порушити приватність і дати корпораціям можливість маніпулювати власниками цих даних. Недостатня захищеність даних може призвести до витоку інформації. Також висока залежність від технологій може призвести до труднощів у випадку системних збоїв і створює загрозу стійкості та доступності даних.

Отже, використання технологій Big Data визначається як важливий та перспективний напрям у різних галузях. Необхідність ретельного врахування етичних аспектів та забезпечення надійного захисту конфіденційності даних є невід'ємною умовою для досягнення максимального ефекту від застосування цих технологій. Одночасно це дослідження висвітлює необхідність усвідомлення ризиків, пов'язаних із використанням Big Data, зокрема порушення приватності, можливості маніпуляцій та недостатньої захищеності даних від кіберзагроз. Спрямоване на розширення розуміння можливостей та викликів, це дослідження стимулює подальший розвиток і ефективне впровадження технологій Big Data у різноманітних сферах, враховуючи всі аспекти їх використання.

Список використаних джерел

1. Що таке Big Data?: вебсайт. URL: <https://hub.kyivstar.ua/articles/shho-take-big-data> (дата звернення: 21.11.2023).

2. Що таке Big Data і як це працює: вебсайт. URL: <https://qagroup.com.ua/publications/shcho-take-big-data-i-yak-tse-pratciuiie/> (дата звернення: 21.11.2023).

3. Технології Big Data: ключові характеристики, особливості та переваги: вебсайт. URL: <https://cutt.ly/YwU8IX0m> (дата звернення: 21.11.2023).

4. Що таке Big Data: все що вам слід знати про великі дані?: вебсайт. URL: <https://cutt.ly/gwU8uYRS> (дата звернення: 21.11.2023).

5. Степанюк О. С., Січко Т. В. Особливості використання реляційних та нереляційних баз даних в Big Data. *Комп'ютерні технології обробки даних: матеріали всеукр. наук.-практ. конф., м. Вінниця, 2020. С. 103–106.*

6. Ткачук Н. О., Січко Т. В. Застосування Від Data у бізнесі. *Комп'ютерні технології обробки даних: матеріали всеукр. наук.-практ. конф., м. Вінниця, 2022. С. 224–226.*

УДК 004.67

Сніжинський М. В., здобувач 2 курсу спеціальності 122 Комп'ютерні науки СО «Магістр»,

Січко Т. В., канд. техн. наук, доцент, доцент кафедри інформаційних технологій

СИСТЕМА ОБРОБКИ ТА АНАЛІЗУ ДАНИХ У МЕДИЧНІЙ СФЕРІ НА ОСНОВІ ТЕХНОЛОГІЙ BIG DATA

Донецький національний університет імені Василя Стуса, м. Вінниця

Останніми роками відбулось значне збільшення обсягів даних у сфері охорони здоров'я завдяки впровадженню цифрових технологій, як-от: електронні медичні картки, медична візуалізація та геноміка. Ці дані включають інформацію про пацієнтів, їхні медичні історії, діагностичні результати та лікування. Донедавна обробка даних у сфері охорони здоров'я відбувалася з великими труднощами через їх обсяг та різноманітність. Використання технологій великих даних, як-от машинне навчання, штучний інтелект і розподілені обчислення, відкриває багатообіцяючі перспективи. Впровадження масштабованої та захищеної інфраструктури разом із передовими алгоритмами уможливує ефективну обробку даних, розпізнавання образів, прогностичне моделювання та аналіз результатів. До того ж включення інтегрованих систем і стандартизованих протоколів сприяє безперешкодному обміну даними і співпраці між медичними установами. Зміна парадигми охорони здоров'я на основі даних зумовлює необхідність створення комплексної системи обробки й аналізу великих обсягів медичних даних із використанням можливостей технологій великих даних.