**ЗАТВЕРДЖУЮ**

Проректор з науково-педагогічної роботи

\_\_\_\_\_\_\_\_\_\_ Андрій МОРОЗОВ

\_\_\_ \_\_\_\_\_\_\_\_\_ 2021 р.

**ТЕСТОВІ ЗАВДАННЯ**

**з навчальної дисципліни**

**«Аналіз великих даних у фінансах»**

для здобувачів вищої освіти освітнього ступеня «магістр»

спеціальності 051 «Економіка»

освітньо-професійна програма «Економіка»

факультет бізнесу та сфери обслуговування

(назва факультету)

кафедра цифрової економіки та міжнародних економічних відносин

(назва кафедри)

Схвалено на засіданні кафедри цифрової економіки та міжнародних економічних відносин

 (назва кафедри)

\_28\_ \_\_\_\_\_\_08\_\_\_\_ 2021 р., протокол № \_13\_\_

Завідувач кафедри

\_\_\_\_\_\_ Катерина ШИМАНСЬКА

**Розробник**

Бондарчук Віталій Вікторович - заступник завідувача кафедри, доцент кафедри міжнародних економічних відносин, кандидат економічних наук

The Big Data Analytics in finance syllabus is developed in the framework of ERASMUS+ CBHE project «Digitalization of economic as an element of sustainable development of Ukraine and Tajikistan» / DigEco618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

*This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which maybe made of the information contained therein.*

*Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі.*

Житомир

2021-2022 навчальний рік

Аналіз великих даних у фінансах

[Електронний ресурс] : тестові завдання з дисципліни «Аналіз великих даних у фінансах» для здобувачів вищої освіти «Магістр» за спеціальністю: 051 «Економіка». – Житомир : Державний університет «Житомирська політехніка», 2021. – Режим доступу:

Тестові завдання розроблено в рамках проєкту Erasmus+ “Діджиталізація економіки як елемент сталого розвитку України та Таджикистану (DigEco) 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP”/ The proramm is developed in the framework of ERASMUS+ CBHE project “Digitalization of economic as an element of sustainable development of Ukraine and Tajikistan” / DigEco 618270-EPP-1-2020-1-LT-EPPKA2-CBHE-JP

Цей проект фінансується за підтримки Європейської Комісії. Цей документ відображає лише погляди автора, і Комісія не несе відповідальності за будь-яке використання інформації, що міститься в документі/This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained there in.

Розробник: Бондарчук Віталій Вікторович - заступник завідувача кафедри, доцент кафедри міжнародних економічних відносин, кандидат економічних наук

© Державний університет «Житомирська політехніка», 2021

© В. В. Бондарчук, 2021

# ТЕСТОВІ ЗАВДАННЯ

1. **Поєднання структурованих, напівструктурованих та неструктурованих даних, які можуть бути видобуті для отримання інформації та використані в проектах машинного навчання, прогнозного моделювання та інших передових програм аналітики, це:**
   1. Великі дані
   2. Традиційні дані
   3. Логічні дані
   4. Стандартні дані
2. **Які з видів даних охоплює Big Data**
   1. Структуровані дані
   2. Неструктуровані дані
   3. Напівструктуровані дані
   4. Усі перераховані варіанти
3. **Яка з перелічених країн не є фаворитом на ринку великих даних**
   1. США
   2. Канада
   3. Китай
   4. Україна
4. **Життєвий цикл великих даних поділяють на:**
   1. 3 етапи
   2. 5 етапів
   3. 7 етапів
   4. 9 етапів
5. **Дані, які необхідні для глибокого аналізу складаються з**
   1. Даних якими володіє компанія
   2. Даних придбаних у фірм які спеціалізуються на зборі інформації
   3. Даних з відкритих джерел
   4. Усі відповіді вірні
6. **Етап на якому здійснюється визначення наборів даних, необхідних для аналітичних проектів і їх джерел, це**
   1. Етап ідентифікації даних
   2. Етап збору і фільтрації даних
   3. Етап перевірки та очищення даних
   4. Етап агрегування та представлення даних
7. **Виберіть чотири головні характеристики Big Data:**
   1. Virtualization, Volume, Variability, Vehicle
   2. Variety, Velocity, Volume, Value
   3. Verification, Volume, Velocity, Visualization
   4. Video, Value, Variety, Volume
8. **Виберіть важливі події, що вплинули на формування тренда великих даних:**
   1. Розробка мови Pyhton
   2. Розробка фреймворку Hadoop
   3. Винахід принципу MapReduce
   4. Розробка мови Java
   5. Розробка мови R
9. **Виберіть варіант, в яких дані структуровані**
   1. Дані щодо продажів підприємства, які наведені у вигляді помісячних звітів у форматіMS Word
   2. Таблиця з щоденними показами температури приміщення за рік в файлі формату csv
   3. Текст наукового посібника, що наведений у форматі PDF  
      4 Бібліотека фільмів, представлених у форматі mpeg4 на одному жорсткому диску
10. **Hadoop – це:**
    1. Розподілена СУБД, що дозволяє обробляти великі дані
    2. Набір утиліт, а також програмний каркас для виконання розподілених програм, щопрацюють на кластерах
    3. Мова виконання завдань відповідно до парадигми MapReduce
    4. Розподілена файлова система, що призначена для зберігання файлів великого обсягу
11. **Розставте послідовність етапів проекту аналітики відповідно до методології CRISPDM**

|  |  |
| --- | --- |
| 1) | а) Розуміння бізнес-цілей (Business understanding); |
| 2) | б) Впровадження (Deployment); |
| 3) | в) Моделювання (Modeling); |
| 4) | г) Початкове вивчення даних (Data Understanding); |
| 5) | д) Оцінка (Evaluation) |
| 6) | е) Підготовка даних (Data Preparation) |

1. **Для чого аналітику необхідна «аналітична пісочниця»?**
   1. Для високопродуктивної аналітики за рахунок використання оперативної пам’яті та inDBоперацій.
   2. Для зберігання всіх отриманих від замовника даних.
   3. Для побудови звітів про результати аналізу.
   4. Для зниження витрат, пов’язаних з реплікацією даних.
2. **Які з перелічених засобів доцільно використовувати для аналізу даних, що представлені єдиним csv-файлом, розміром більшим за 100 ГБ?**
   1. Data Warehouse.
   2. Hadoop.
   3. «Аналітична пісочниця».
   4. Python.
   5. PHP.
3. **На якому з етапів процесу CRISP-DM відбувається перевірка якості даних?**
   1. Розуміння бізнес-цілей (Business understanding).
   2. Початкове вивчення даних (Data Understanding).
   3. Підготовка даних (Data Preparation).
   4. Моделювання (Modeling).
   5. Оцінка (Evaluation).
   6. Впровадження (Deployment).
4. **До слабких сторін Hadoop можна віднести... Виберіть зайве.**
   1. Hadoop – це фреймворк, а не готове рішення.
   2. Hive та Pig мають безліч архітектурних обмежень порівняно з реляційною СУБД.
   3. Hadoop легко встановити і налаштувати, але складніше супроводжувати.
   4. Hadoop обробляє тільки дуже велику кількість даних, які зберігаються розподілено на безлічі вузлів.
   5. У деяких випадках Hadoop дуже повільний.
5. **Принцип MapReduce полягає в тому, щоб:**
   1. Здійснювати обчислення на вузлах, де інформація спочатку була збережена.
   2. Використовувати обчислювальні потужності систем зберігання.
   3. Використовувати функціональне програмування для вирішення задач масивно-паралельної обробки.
   4. Використовувати мову програмування Phyton.
6. **До методів машинного навчання не належить**
   1. Класифікація
   2. Кластеризація
   3. Виявлення викидів
   4. Фільтрація
   5. Уніфікація
7. **До семантичного аналізу не відноситься**
   1. Обробка природної мови
   2. Обробка тексту
   3. Аналіз емоційного забарвлення висловлювань
   4. Аналіз математичних формул
8. **Що не відноситься до візуального аналізу**
   1. Кольорові карти
   2. Часові ряди
   3. Сітьові графіки
   4. Зіставлення просторових даних
   5. Кореляція
9. **Метод, який фокусується на аналізі відносин між об'єктами в мережі, це**
   1. Сітьовий аналіз
   2. Кореляційний аналіз
   3. Факторний аналіз
   4. Системний аналіз
10. **Візуальним каналом є**
    1. Точка
    2. Лінія
    3. Частина поверхні
    4. Колір
11. **Вид візуалізації «Історичний потік»** 
    1. Допомагає стежити за еволюцією документа, над створенням якого працює одночасно велику кількість авторів
    2. Використовується для відстеження просторового зміни інформації
    3. Подання будь-якої корисної інформації в формі цікавої розповіді
    4. Передбачає створення великої кількості різних візуальних уявлень одних і тих же даних
12. **Клас алгоритмів машинного навчання, що використовуються для виявлення прихованих закономірностей в даних.**
    1. Навчання без вчителя
    2. Навчання признаків
    3. Навчання з підкріпленням
    4. Навчання з вчителем
13. **Клас алгоритмів машинного навчання, при яких прогноз будується на основі закономірностей в даних і, крім того, продовжує поліпшуватися по мірі надходження нових результатів**
    1. Навчання без вчителя
    2. Навчання признаків
    3. Навчання з підкріпленням
    4. Навчання з вчителем
14. **Сукупність упорядкованих, класифікованих даних про деяке масове явище або процес це...**
    1. Графічні дані
    2. Статистичні дані
    3. Математичні дані
    4. Систематичні дані
    5. Великі дані
15. **Ряди розподілу одиниць сукупності за ознаками, що мають кількісний вираз, називаються це...**
    1. Генеральна сукупність
    2. Вибірка
    3. Статистичний ряд розподілу
    4. Варіаційний ряд
16. **Забезпечення доступу до загальнодоступних даних усім користувачам, захист цих даних від спотворення та блокування зловмисниками - це...**
    1. Доступність
    2. Цілісність
    3. Конфіденційність
    4. Загальноприйнятність
17. **Що називають Інформаційною безпекою?**
    1. Сукупність заходів для захисту даних та інформаційної системи від випадкових або навмисних пошкоджень та несанкціонованого доступу
    2. Сукупність засобів і методів віддаленого зберігання й опрацювання дани
    3. Норми поведінки осіб у інформаційному просторі
    4. Сукупність антивірусних програм
18. **Чим пояснюється надійність сучасних засобів збереження інформації?**
    1. Ефективними способами передавання і копіювання даних
    2. Механічною міцністю й термічною витривалістю цих засобів
    3. Довговічністю цих засобів
    4. Цифровим кодуванням даних
19. **Портал відкритих даних Міністерства цифрової трансформації України**
    1. Data.gov
    2. <https://open-data.europa.eu/>
    3. Data.worldbank.org
    4. Data.gov.ua

# РЕКОМЕНДОВАНА ЛІТЕРАТУРА

**Основна**

1. Chapman, Arthur D. “Principles and methods of data cleaning.” 2005
2. D. Hand, H. Mannila, P. Smyth. 2001. Principles of Data Mining
3. Daniel Keim, JörnKohlhammer, Geoffrey Ellis und Florian Mansmann. „Visual Analytics“. 2010
4. Data Science from Scratch by Joel Grus (O‘REILLY) Elements of Statistical Learning: Data Mining, Inference, and Prediction, by Trevor Hastie, Robert Tibshirani and Jerome Friedman
5. Dimitri P. Bertsekas and John N. Tsitsiklis. Introduction to Probability. Charles Wheelan. Naked Statistics: Stripping the Dread from the Data. W. W. Norton and Company, 2013.
6. E. Friedmann, T. Dunning. 2015. Time Series Databases: New Ways to Store and Access Data
7. Guyon, Isabelle, Nada Matic and Vladimir Vapnik. "Discovering Informative Patterns and Data Cleaning." 1996
8. J. Han, M. Kamber. 2011. Data Mining. Concepts and Techniques
9. James Thomas und Kristin Cook. „Illuminating the Path“. 2005 vis.pnnl.gov/pdf/RD\_Agenda\_VisualAnalytics.pdf
10. Osborne, Jason W. “Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data.” 2013
11. Python for Data Analysis by Wes McKinney (O‘REILLY)
12. R. M. Müller, H.-J. Lenz. 2013. Business Intelligence
13. Steven Skiena. “The Data Science Design Manual” <http://www.data-manual.com/>
14. Гурвиц Дж., Ньюджент А., Халпер Ф., Кауфман М. Просто о больших данных: пер. с англ. М.:Эксмо, 2015. 400 с.
15. Кобзаренко Д.Н., Мустафаев А.Г. Учебное пособие дисциплины«Анализ больших данных» для направления подготовки 38.03.05«Бизнесинформатика», профиль «Электронный бизнес». – Махачкала: ДГУНХ, 2019 г.– 107 с.
16. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем имыслим. – М.: Манн, Иванов и Фербер, 2017. – 240 c
17. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. СПб:Университет ИТМО, 2018. 52 с.
18. Силен Дэвид, МайсманАрно, Али МохамедОсновы Data Science иBig Data. Python и наука о даных. СПб.: Питер, 2017. 336 с.
19. Томас Єрл, ВаджидХаттак, Пол БулерОснови Big Data: Концепції, алгоритми та технології/Пер.з англ. Анатолія Гладуна;За наук.ред. Олексія Найди. Дніпро: «Баланс Бізнес Букс», 2018. 320 с.
20. Фрэнк Билл. Революция в аналитике. Как в эпоху Big Data улучшить вашбизнес с помощью операционной аналитики М.: Альпина Паблишер, 2014. 430 с
21. Фрэнкс Билл. Укрощение больших данных М.: Манн, Иванов и Фербер, 2014. 352 с.
22. ЫнАнналин, Су Кеннет Теоритический минимум по Big Data, что нужно знать о больших даных. СПб.: Питер, 2019. 208 с.

**Допоміжна**

1. 36 кращих інструментів для візуалізації даних. URL: <https://toplead.com.ua/ua/blog/id/38-luchshih-instrumentov-dlja-vizualizacii-dannyh-160/>
2. Big data от А до Я. Часть 3: Приемы и стратегии разработки MapReduce-приложений URL: https://habrahabr.ru/company/dca/blog/270453/ 5
3. BigData от А до Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce URL: <https://habrahabr.ru/company/dca/blog/267361/>
4. Большие данные в финансовой отрасли: обзор и оценка перспектив развития мирового и российского рынков. URL: <file:///C:/Users/PC/Downloads/idc-26012016.pdf>
5. Великі дані в індустрії фінансових послуг – від даних до аналізу. URL: <https://www.finextra.com/blogposting/17847/big-data-in-the-financial-services-industry---from-data-to-insights>
6. Великі дані у фінансах – ваш посібник з аналізу фінансових даних. URL: <https://www.talend.com/resources/big-data-finance/>
7. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методами исследованию операций с использованием пакетов STATISTICA и EXCEL: учебноепособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013. – 464 с.
8. Гобарева Я.Л., Городецкая О.Ю., Золотарюк А.В. Бизнес-аналитикасредствами Excel М.: Вузовский учебник, ИНФРА-М, 2013. –336 с.
9. Захист персональних даних . URL: <https://www.atcominvestbank.com/uk/about/personal-data-protection> URL:
10. ІЕ безпека і захист інформації. URL: <https://infotel.ua/ua/IT-bezopasnost-i-zacshita-informatsii-1/>
11. Ілюк А. Ризики, пов'язані із захистом персональних даних в контексті BigData. *Юридична газета online*. URL: https://yur-gazeta.com/publications/practice/inshe/riziki-povyazani-iz-zahistom-personalnih-danih-v-konteksti-big-data.html
12. Квартальний Н. Великі дані у фінансах – роль аналізу фінансових даних. URL: https://inoxoft.com/blog/big-data-in-financial-services-role-of-financial-data-analysis-inoxoft/
13. Константинівна І.О. Сучасні криптосистеми. URL: <https://sites.google.com/site/sucasnikriptosistemik/home>
14. Круковець Д. Можливості Data Science в центральних банках: огляд. URL: https://journal.bank.gov.ua/uploads/articles/249\_2\_Krukovets\_Ukr.pdf
15. Мега Шах. Як використовувати аналітику великих даних для вдосконалення фінансової індустрії URL: https://www.techfunnel.com/fintech/how-to-use-big-data-analytics-to-improve-finance-industry/
16. Основные понятия безопасности для работы с Кластеры больших данных SQL Server URL: <https://docs.microsoft.com/ru-ru/sql/big-data-cluster/concept-security?view=sql-server-ver15>
17. Охорона банківської таємниці: правові засади. URL: <http://obt.inf.ua/page10.html#top>
18. Про Систему BankID Національного банку. URL: <https://bank.gov.ua/ua/bank-id-nbu>
19. Распределенные базы и хранилища данных : Электронный учебник / Н. Аносова, О. Бородин, Е. Гаврилов и др. – НОУ "ИНТУИТ" URL: <http://www.intuit.ru/studies/courses/1145/214/info>
20. Распределенные файловые системы. Технологии хранения и обработкибольших объемов данных / Computer Science Center URL: https://compscicenter.ru/ courses/ big-data/ 2015-spring/ classes/1117/.
21. Сем Палмер. Аналіз великих даних у фінансових послугах. URL: <https://www.devteam.space/blog/big-data-analytics-in-financial-services/>
22. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посіб.К.: КНЕУ, 2007. 376 с.
23. Черняк Л. Безопасность больших данных. URL:<https://www.osp.ru/os/2013/02/13034551>
24. Что нужно знать о Big Data в финансах: краткое руководство. URL: <https://aboutdata.ru/2017/06/06/big-data-and-finance/>
25. Шандрівська О. Є., Кириленко А. А.Особливості ідентифікації ризиків ринку big data. [URL: http://science.lpnu.ua/sites/default/files/journal-paper/2021/jun/23774/menedzhment121-84-97.pdf](URL:%20http://science.lpnu.ua/sites/default/files/journal-paper/2021/jun/23774/menedzhment121-84-97.pdf)
26. Шипунов А.Б., Балдин Е.М., Волкова П.А., Коробейников А.И., Назарова С.А., ПетровС.В., Суфиянов В.Г. Наглядная статистика. Используем R! - М.: ДМК Пресс, 2012. – 298с.
27. Шулин С. Будущее фынансового сектора за Big Data. URL: <https://bluescreen.kz/articles/budushhee-finansovogo-sektora-za-bigdata/>
28. Що таке великі дані у фінансах. URL: <https://corporatefinanceinstitute.com/resources/knowledge/other/big-data-in-finance/>
29. Як великі дані покращують банківські та фінансові системи. URL: <https://innovecs.com/blog/big-data-in-banking-and-financial-systems/>
30. Як і для чого використовують візуалізацію даних? URL: http://eidos.org.ua/novyny/yak-i-dlya-choho-vykorystovuvaty-vizualizatsiyu-danyh/

**Рекомендовані джерела інформації**

1. Законодавчі та нормативно-правові документи. URL: [http://zakon.rada.gov.ua/](http://zakon.rada.gov.ua/7)
2. <https://books.google.com/ngrams>
3. <https://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf>
4. [https://www.data.gov](https://www.data.gov/)
5. <http://data.europa.eu/euodp/de/about>
6. <https://www.govdata.de/>
7. <http://transparenz.hamburg.de/open-data/>
8. <https://dasl.datadescription.com/datafiles/>
9. <http://linkedscience.org/>
10. <https://www.opensciencedatacloud.org/>
11. <https://philogb.github.io/jit/>
12. <http://tagesnetzwerk.de/>
13. [https://tweetping.net](https://tweetping.net/)
14. <https://flowingdata.com/2017/08/18/catalog-of-visualization-types-to-find-the-one-that-fits-your-dataset/>
15. <https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/>
16. <https://callingbullshit.org/tools/tools_proportional_ink.html>
17. <http://www.visualisingdata.com/2014/04/the-fine-line-between-confusion-and-deception>
18. <http://tylervigen.com/spurious-correlations>
19. www.domo.com/learn/infographic-data-never-sleeps
20. http://www.originlab.de/
21. http://support.minitab.com
22. https://flowingdata.com/tag/treemap/
23. http://mars.wiwi.hu-berlin.de/mediawiki/teachwiki/index.php/Analysis\_of\_Mortality
24. http://jn.physiology.org/content/104/4/2103
25. http://www.sigchi.org/chi96/proceedings/papers/Tweedie/lt1txt.htm
26. http://rosuda.org/software/Gauguin/gauguin.html