



**ДЕРЖАВНА СЛУЖБА СТАТИСТИКИ УКРАЇНИ
НАЦІОНАЛЬНА АКАДЕМІЯ СТАТИСТИКИ ОБЛІКУ
ТА АУДИТУ**

**КАФЕДРА СТАТИСТИКИ, ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА
МАТЕМАТИЧНИХ МЕТОДІВ В ЕКОНОМІЦІ**



ЗАТВЕРДЖУЮ
Перший проректор
наукowo-педагогічної роботи
Л.Є. Момотюк
«01» вересня 2022 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ОК 32. ВСТУП ДО BIG DATA

Ступінь вищої освіти **Бакалавр**

Галузь знань **05 Соціальні та поведінкові науки**

Спеціальність **051 «Економіка»**

Освітньо-професійна програма **«Прикладна статистика та бізнес-аналітика»**

Київ
2022 рік

Робоча програма з дисципліни «Вступ до Big Data» для студентів спеціальності 051 «Економіка», 31 серпня 2022 року. 19 с.

Розробник: **Горобець О. О.**, кандидат економічних наук, доцент кафедри статистики, інформаційних технологій та математичних методів в економіці.

Робоча програма затверджена на засіданні кафедри статистики інформаційних технологій та математичних методів в економіці.

Протокол від «30» серпня 2022 року, № 1

Схвалено Вченою радою обліково-статистичного факультету НАСOA

Протокол від «31» серпня 2022 року, № 1

1. Опис навчальної дисципліни

Найменування Показників	Спеціальність, освітня програма, рівень освіти	Характеристика навчальної дисципліни
Кількість кредитів – 4	Спеціальність 051 “Економіка”	Основна
Модулів – 1	Освітньо-професійна програма: Прикладна статистика та бізнес- аналітика	Рік підготовки:
Змістових модулів – 2		4-й
Загальна кількість годин – 120		Семестр
		8-й
	Лекції	
Тижневих годин для денної форми навчання: аудиторних – 5 самостійної роботи студента – 4	Рівень освіти: перший (бакалаврський)	36 год.
		Практичні 34 год.
		Самостійна робота
		50 год.
		Вид контролю: екзамен

2. Мета та завдання навчальної дисципліни

2.1. Метою викладання навчальної дисципліни «Вступ до Big Data» є ознайомлення студентів із сферою Big Data, демонстрація переваг та недоліків імплементації Big Data в діяльність органів офіційної статистики, представлення наявного інструментарію, що використовується для збирання, обробки, групування та аналізу Big Data та подальшого представлення статистичної інформації.

2.2. Основними завданнями вивчення навчальної дисципліни «Вступ до Big Data» є: оволодіння основними поняттями у сфері Big Data, усвідомлення ролі та значення Big Data для сучасних аналітиків та для органів статистики, ознайомлення з принципами появи та розповсюдження Big Data, організації сховищ великих даних, із базовими алгоритмами збереження, пошуку та аналізу Big Data, формування навичок роботи з програмним забезпеченням Big Data та використання статистичних методів для аналізу Big Data.

2.3. У результаті вивчення навчальної дисципліни студент повинен **знати:**

- ✓ сутність поняття Big Data та взаємопов'язаних термінів;
- ✓ способи та методи імплементації Big Data у різні сфери життєдіяльності;
- ✓ властивості, класифікацію, структурування та методи зберігання Big Data;
- ✓ властивості, особливості та джерела Big Data;
- ✓ характеристики та види архітектур сховищ Big Data;
- ✓ основні алгоритми роботи зі сховищем Big Data (MapReduce);
- ✓ екосистеми, фреймворки для обробки та зберігання Big Data;
- ✓ основні методи статистичного аналізу та прогнозування Big Data.

2.4. Найменування компетентностей (код та назва за ОПП), формування яких забезпечує вивчення навчальної дисципліни: Інтегральна компетенція - Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми в економічній сфері, які характеризуються комплексністю та невизначеністю умов, що передбачає застосування теорій та методів економічної науки.

ЗК7. Навички використання інформаційних і комунікаційних технологій.

ЗК8. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

СК7. Здатність застосовувати комп'ютерні технології та програмне забезпечення з обробки даних для вирішення економічних завдань, аналізу інформації та підготовки аналітичних звітів.

СК10. Здатність використовувати сучасні джерела економічної, соціальної, управлінської, облікової інформації для складання службових документів та аналітичних звітів.

2.5. Передумови для вивчення навчальної дисципліни – наявність сформованих ґрунтовних знань із «Статистики», «Економічної статистики», «Статистичного моделювання», «Інформатики».

2.6. Результати навчання (код та назва за ОПП):

ПРН 9. Використовувати інформаційні та комунікаційні технології для вирішення соціально-економічних завдань, підготовки та представлення аналітичних звітів.

ПРН 13. Ідентифікувати джерела та розуміти методологію визначення і методи отримання соціально-економічних даних, збирати та аналізувати необхідну інформацію, розраховувати економічні та соціальні показники.

ПРН 16. Вміти використовувати дані, надавати аргументацію, критично оцінювати логіку та формувати висновки з наукових та аналітичних текстів з економіки.

3. Програма навчальної дисципліни Змістовний модуль 1. Теоретико-методичні засади Big Data

Тема 1. Історія появи та поширення терміну Big Data. Предмет та завдання Big Data

Поняття, предмет та загальна концепція Big Data. Джерела та причини виникнення великих об'ємів даних. Переваги та недоліки Big Data. Завдання Big Data. Практичний досвід імплементації Big Data.

Рекомендована література:

Базова: [6, 7]

Допоміжна: [1, 2, 6, 19]

Тема 2. Глобальні групи даних. Категорії даних

Поняття про глобальні групи даних – Shallow data, Deep Data, Micro-data, Nano-data, їх структурування та класифікацію. Ідентифікація машинних даних, потокових даних, озер даних. Категорії структурованих, неструктурованих та напівструктурованих даних. Інструменти для роботи з Big Data.

Рекомендована література:

Базова: [2, 6, 7]

Допоміжна: [1]

Тема 3. Роль та значення Big Data для офіційної статистики. Забезпечення якості та конфіденційності

Big Data як альтернативне джерело статистичних даних для органів офіційної статистики. Нормативно-правове забезпечення якості та конфіденційності даних. Поняття «якість» та «конфіденційність» в контексті Big Data. Параметри якості даних. Вимоги до якості даних. Виміри якості даних. Поняття «життєвий цикл даних». Формування вартості даних. Типи конфіденційних та персональних даних в структурі Big Data.

Рекомендована література:

Базова: [6]

Допоміжна: [1, 5]

Електронні ресурси: [6]

Змістовний модуль 2. Методи та інструменти обробки Big Data

Тема 4. Екосистема Apache Hadoop: архітектура, реплікація, читання і запис даних

Характеристика екосистеми Apache Hadoop. Класифікація алгоритмів для опрацювання Big Data. Найпоширеніші методи аналізу та прогнозування Big Data в екосистемі Apache Hadoop: метод k -середніх; метод опорних векторів регресійний аналіз, метод головних компонент, метод асоціативних правил, аналіз соціальних медіа (SNA). Математичний аналіз SNA.

Рекомендована література:

Базова: [1, 3, 5, 8]

Допоміжна: [4, 7, 9]

Електронні ресурси: [2]

Тема 5. Програмний каркас Hadoop MapReduce

Характеристика інтерфейсу MapReduce. Організація функціонування MapReduce. Утиліта Hadoop Streaming. Технологія опрацювання даних «Відображення» – «Згортання».

Рекомендована література:

Базова: [3]

Допоміжна: [4]

Електронні ресурси: [2]

Тема 6. Алгоритмізація даних. Обчислення даних за допомогою нейромереж

Поняття «алгоритм». Особливості алгоритмізації даних. Поняття «нейромережа». Роль та значення нейромереж в процесі обробки Big Data.

Рекомендована література:

Базова: [4, 9]

Допоміжна: [11]

Тема 7. Фреймворки Apache Spark та Hive для обробки та зберігання Big Data

Характеристика фреймворку Apache Spark. Особливості проекту Hive. Роль та значення Apache Spark Hive для обробки, аналізу та зберігання Big Data.

Рекомендована література:

Базова: [1, 7]

Допоміжна: [4, 8]

Електронні ресурси: [3, 4]

Тема 8. Платформи хмарних обчислень Big Data

Платформи Amazon Web Services та Google Cloud Platform: особливості, переваги та недоліки використання. Сервіс Microsoft Azure: загальна характеристика статистичних інструментів для обробки Big Data.

Рекомендована література:

Базова: [2]

Допоміжна: [4, 8]

Електронні ресурси: [1]

Тема 9. Платформа Power BI як інструмент візуалізації Big Data

Функціонал платформи Power BI. Основні компоненти набору Power BI. Переваги та недоліки Power BI. Статистичне моделювання великих обсягів даних за допомогою інструментів Power BI.

Рекомендована література:

Базова: [7]

Допоміжна: [2, 3]

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин				
	усього	у тому числі			
		л	п	лаб	с.р.
Змістовий модуль 1. Теоретико-методичні засади Big Data					
Тема 1. Історія появи та поширення терміну Big Data. Предмет та завдання Big Data	13	4	4	-	5
Тема 2. Глобальні групи даних. Категорії даних	13	4	4	-	5
Тема 3. Роль та значення Big Data для офіційної статистики. Забезпечення якості та конфіденційності	13	4	4	-	5
Модульна контрольна робота	2	-	2	-	
Разом за змістовим модулем 1	41	12	14	-	15
Змістовий модуль 2. Методи та інструменти обробки Big Data					
Тема 4. Екосистема Apache Hadoop: архітектура, реплікація, читання і запис даних	13	4	4	-	5
Тема 5. Програмний каркас Hadoop MapReduce	12	4	2	-	6
Тема 6. Алгоритмізація даних. Обчислення даних за допомогою нейромереж	14	4	4	-	6
Тема 7. Фреймворки Apache Spark та Hive для обробки та зберігання Big Data	12	4	2	-	6
Тема 8. Платформи хмарних обчислень Big Data	12	4	2	-	6
Тема 9. Платформа Power BI як інструмент візуалізації Big Data	14	4	4	-	6
Модульна контрольна робота	2	-	2	-	-
Разом за змістовим модулем 2	79	24	20	-	35
Усього годин	120	36	34	-	50

5. Теми лекційних занять

№ з/п	Назва теми	Кількість годин
1	2	3
1.	<p style="text-align: center;">Тема 1. Історія появи та поширення терміна Big Data. Предмет та завдання Big Data План</p> <p>1. Поняття, предмет, завдання та загальна концепція Big Data. 2. Джерела та причини виникнення великих об'ємів даних. 3. Переваги та недоліки Big Data. 4. Практичний досвід імплементації Big Data.</p>	4
2.	<p style="text-align: center;">Тема 2. Глобальні групи даних. Категорії даних План</p> <p>1. Поняття про глобальні групи даних – Shallow data, Deep Data, Micro-data, Nano-data, їх структурування та класифікацію 2. Ідентифікація машинних даних, потокових даних, озер даних. 3. Категорії структурованих, неструктурованих та напівструктурованих даних. 4. Інструменти для роботи з Big Data.</p>	4
3.	<p style="text-align: center;">Тема 3. Роль та значення Big Data для офіційної статистики. Забезпечення якості та конфіденційності План</p> <p>1. Big Data як альтернативне джерело статистичних даних для органів офіційної статистики 2. Нормативно-правове забезпечення якості та конфіденційності даних. 3. Поняття «якість» та «конфіденційність» в контексті Big Data. Параметри якості даних. Вимоги до якості даних. Виміри якості даних. 4. Поняття «життєвий цикл даних». Формування вартості даних. 5. Типи конфіденційних та персональних даних в структурі Big Data.</p>	4
4.	<p style="text-align: center;">Тема 4. Екосистема Apache Hadoop: архітектура, реплікація, читання і запис даних План</p> <p>1. Характеристика екосистеми Apache Hadoop. Класифікація алгоритмів для опрацювання Big Data. 2. Найпоширеніші методи аналізу та прогнозування Big Data в екосистемі Apache Hadoop: метод <i>k</i>-середніх; метод опорних векторів регресійний аналіз, метод головних компонент, аналіз соціальних медіа (SNA).</p>	4

1	2	3
5.	<p style="text-align: center;">Тема 5. Програмний каркас Hadoop MapReduce</p> <p style="text-align: center;">План</p> <p>1.Характеристика інтерфейсу MapReduce. 2.Організація функціонування MapReduce. Утиліта Hadoop Streaming. 3.Технологія опрацювання даних «Відображення» – «Згортання». 4.Розподіл та обробка даних.</p>	4
6.	<p style="text-align: center;">Тема 6. Алгоритмізація даних. Обчислення даних за допомогою нейромереж</p> <p style="text-align: center;">План</p> <p>1.Особливості алгоритмізації даних. 2.Поняття про нейромережі. Роль та значення нейромереж в процесі обробки Big Data.</p>	4
7.	<p style="text-align: center;">Тема 7. Фреймворки Apache Spark та Hive для обробки та зберігання Big Data</p> <p style="text-align: center;">План</p> <p>1.Характеристика фреймворку Apache Spark. 2.Особливості проєкту Hive. 3.Роль та значення Apache Spark Hive для обробки, аналізу та зберігання Big Data.</p>	4
8.	<p style="text-align: center;">Тема 8. Платформи хмарних обчислень Big Data</p> <p style="text-align: center;">План</p> <p>1.Платформи Amazon Web Services та Google Cloud Platform: особливості, переваги та недоліки використання. 2.Сервіс Microsoft Azure: загальна характеристика статистичних інструментів для обробки Big Data.</p>	4
9.	<p style="text-align: center;">Тема 9. Платформа Power BI як інструмент візуалізації Big Data</p> <p style="text-align: center;">План</p> <p>1.Функціонал платформи Power BI. 2.Основні компоненти набору Power BI. 3.Переваги та недоліки Power BI. 4.Статистичне моделювання великих обсягів даних за допомогою інструментів Power BI.</p>	4
Разом		36

6. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	2	3
1	<p style="text-align: center;">Тема 1. Історія появи та поширення терміна Big Data. Предмет та завдання Big Data План</p> <p>1.Інтерпретування поняття Big Data. 2.Особливості імплементації Big Data в практичну діяльність.</p>	4
2	<p style="text-align: center;">Тема 2. Глобальні групи даних. Категорії даних План</p> <p>1. Особливості збирання та обробки глобальних груп Big Data – Shallow data, Deep Data, Micro-data, Nano-data. 2. Джерела машинних даних, потокових даних та озер даних. 3. Особливості збирання та обробки структурованих, неструктурованих та напівструктурованих даних.</p>	4
3	<p style="text-align: center;">Тема 3. Роль та значення Big Data для офіційної статистики. Забезпечення якості та конфіденційності План</p> <p>1.Аналіз нормативно-правового забезпечення якості та конфіденційності даних. 2.SWOT-аналіз Big Data. 3. Вимірювання якості глобальних даних: прикладний аспект.</p>	4
4	<p style="text-align: center;">Тема 4. Екосистема Apache Hadoop: архітектура, реплікація, читання і запис даних План</p> <p>1. Встановлення та налаштування Apache Hadoop. 2. Інтерфейс Apache Hadoop, можливості для статистиків.</p>	4
5	<p style="text-align: center;">Тема 5. Програмний каркас Hadoop MapReduce План</p> <p>1.Інтерфейсу MapReduce: можливості для статистиків. 2.Функціонування MapReduce: прикладний аспект опрацювання даних.</p>	4

1	2	3
6	<p>Тема 6. Алгоритмізація даних. Обчислення даних за допомогою неймереж</p> <p>План</p> <p>1. Побудова алгоритмів та виявлення причинно-наслідкових зв'язків</p> <p>2. Функціонування неймереж: досвід використання в Data Science</p>	4
7	<p>Тема 7. Фреймворки Apache Spark та HIVE для обробки та зберігання Big Data</p> <p>План</p> <p>1. Роль та значення фреймворку Apache Spark та проєкту HIVE для бізнес-аналітики .</p> <p>2. Інтерфейс та особливості практичного застосування Apache Spark та HIVE статистиками.</p>	2
8	<p>Тема 8. Платформи хмарних обчислень Big Data</p> <p>План</p> <p>1. Можливості, переваги та недоліки практичного використання платформ Amazon Web Services та Google Cloud Platform.</p> <p>2. Інтерфейс Amazon Web Services та Google Cloud Platform</p> <p>3. Можливості сервісу Microsoft Azure для статистиків.</p>	4
9	<p>Тема 9. Платформа Power BI як інструмент візуалізації Big Data</p> <p>План</p> <p>1. Інтерфейс платформи Power BI.</p> <p>2. Особливості обробки великих масивів даних у Power BI.</p> <p>3. Представлення Big Data у Power BI: побудова таблиць, графіків, діаграм.</p>	4
Разом		36

7. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	2	3
1.	<p>Тема 1. Історія появи та поширення терміна Big Data. Предмет та завдання Big Data План для самостійного вивчення</p> <p>1.Методи та інформаційні технології обробки Big Data. 2.Сфери застосування Big Data. 3. Сучасні тенденції в розвитку систем обробки великих обсягів даних.</p>	5
2.	<p>Тема 2. Глобальні групи даних. Категорії даних План для самостійного вивчення</p> <p>1.Життєвий цикл даних. Попередня обробка даних. 2.Алгоритм ZET заповнення пробілів у таблицях даних. 3.Метадані. Життєвий цикл метаданих.</p>	5
3.	<p>Тема 3. Роль та значення Big Data для офіційної статистики. Забезпечення якості та конфіденційності План для самостійного вивчення</p> <p>1. Принципи офіційної статистики в контексті забезпечення конфіденційності та якості даних. 2. Забезпечення конфіденційності даних в мережі Інтернет</p>	5
4.	<p>Тема 4. Екосистема Apache Hadoop: архітектура, реплікація, читання і запис даних План для самостійного вивчення</p> <p>1.Архітектура та компоненти, кластери, їх вузли, файлова система HDFS. 2. Доступ до сервісів Hadoop засобами мов програмування.</p>	5
5.	<p>Тема 5. Програмний каркас Hadoop MapReduce План для самостійного вивчення</p> <p>1. Програмні оболонки, вибір інтерфейсу та мови програмування при створенні програмних продуктів. 2. MapReduce: використання бібліотек та стандартних (ринкових) програмних продуктів, потоки даних.</p>	6

1	2	3
6.	<p>Тема 6. Алгоритмізація даних. Обчислення даних за допомогою нейромереж План для самостійного вивчення</p> <p>1. Особливості побудови алгоритмів. 2. Властивості штучних нейронних мереж. Біологічний прототип 3. Найпоширеніші помилки побудови алгоритмів.</p>	6
7.	<p>Тема 7. Фреймворки Apache Spark та Hive для обробки та зберігання Big Data План для самостійного вивчення</p> <p>1. Структуризація Apache Spark. Практичний досвід використання інструментарію фреймворку в практичну діяльність. 2. Особливості фреймворку Hive. Практичний досвід використання інструментарію фреймворку в практичну діяльність.</p>	6
8.	<p>Тема 8. Платформи хмарних обчислень Big Data План для самостійного вивчення</p> <p>1. Особливості обробки та зберігання даних на платформах Amazon Web Services та Google Cloud Platform 2. Особливості обробки та зберігання даних за допомогою сервісу Microsoft Azure.</p>	6
9.	<p>Тема 9. Платформа Power BI як інструмент візуалізації Big Data План для самостійного вивчення</p> <p>1. Інструментарій платформи Power BI. 2. Особливості візуалізації Big Data у Power BI</p>	6
	Разом	50

8. Методи навчання

Викладання курсу «Вступ до Big Data» передбачає використання лекційних та практичних занять, майстер-класів та самостійної роботи.

Для досягнення мети та завдань курсу, впродовж лекційних та семінарських занять передбачено застосування наступних методів навчання:

- за джерелами знань: словесні – розповідь, пояснення, лекція, консультація; наочні – демонстрація, ілюстрація; практичні – практична робота, ситуаційні завдання;

- за характером логіки пізнання: аналітичний, синтетичний, аналітико-синтетичний, індуктивний, дедуктивний;

- за рівнем самостійної розумової діяльності використовуються методи: проблемний, частково-пошуковий, дослідницький;

- за рівнем інноваційності: ситуаційного аналізу, ігрові імітаційні технології, дискусія, робота в групах.

Відповідно до Положення про організацію навчання студентів НАСООА за індивідуальним графіком, як виняток, студенти 3 курсу першого (бакалаврського) рівня у разі неможливості відвідувати заняття з поважних причин, що підтверджено відповідними документами, мають змогу отримати індивідуальний графік навчання, що дає їм право, визначеним у п.2.2 Положення, бути відсутніми на лекційних, семінарських та практичних заняттях та передбачає самостійне опрацювання студентом навчального матеріалу.

Студент, який має право на індивідуальний графік навчання, протягом перших двох тижнів нового семестру складає графік за визначеною формою (додаток 2 до Положення), погоджує його з кожним викладачем дисциплін, передбачених навчальним планом відповідного семестру.

Студенти, які навчаються за індивідуальним графіком повинні своєчасно та в повному обсязі виконувати завдання поточного і модульного контролів на платформі Moodle або представляти виконані завдання особисто викладачеві під час консультацій. Виконання індивідуального графіку навчання є підставою для допущення студента до екзаменаційної сесії на загальних підставах.

Консультавання студентів, як мають індивідуальний графік навчання, здійснюється за семестровим графіком консультацій викладача. Окремий час консультавання цих студентів не передбачено.

За виконання індивідуального графіку навчання відповідає студент.

9. Методи контролю

Для визначення успішності навчання використовуються контрольні заходи. Контрольні заходи включають поточний та підсумковий контроль.

Поточний контроль має на меті перевірку рівня підготовленості студента до виконання конкретного завдання.

Підсумковий контроль проводиться з метою оцінки результатів навчання після закінчення вивчення дисципліни (семестровий контроль) або відокремлених за робочим навчальним планом модулів.

Під час вивчення даної дисципліни використовуються такі форми поточного контролю:

- ✓ опитування під час виконання практичних робіт;
- ✓ відповіді на контрольні питання до практичних робіт;
- ✓ тестування.

Відповідно до Положення про внутрішню систему забезпечення якості освіти в НАСООА та Положення про академічну доброчесність у НАСООА у разі виявлення викладачем порушення студентом норм та правил академічної доброчесності, викладач зобов'язаний застосувати норми п.11.8.3 Положення про внутрішню систему забезпечення якості освіти в НАСООА, зокрема: «повторне проходження оцінювання (тест, контрольна робота, залік, іспит тощо); повторне проходження відповідного освітнього компонента (модуля) освітньої програми».

10. Розподіл балів, які отримують студенти

Поточний контроль										Підсумковий контроль (екзамен)	Загальна оцінка
Змістовий модуль 1				Змістовий модуль 2							
T1	T2	T3	Модульна КР	T4	T5	T6	T7	T8	Модульна КР		
5	5	5	15	5	5	5	5	5	15	30	100

10.Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового роботи, практики	для заліку
90 – 100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D		
60-63	E	задовільно	
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

11. Методичне забезпечення

1. Опорний конспект лекцій з дисципліни «Big Data» для здобувачів вищої освіти першого (бакалаврського) рівня / Укладач Горобець О.О. Київ: НАСООА, 2022 (в електронній формі, розміщений на платформі Moodle).

2. Методичні рекомендації для підготовки до семінарських занять з дисципліни «Big Data» для здобувачів вищої освіти першого (бакалаврського) рівня галузі знань 05 «Соціальні та поведінкові науки» спеціальності 051 «Економіка» / Укладач О.О.Горобець. Київ: НАСОА, 2022.

3. Методичні рекомендації для самостійної роботи з дисципліни «Big Data» для здобувачів вищої освіти першого (бакалаврського) рівня галузі знань 05 «Соціальні та поведінкові науки» спеціальності 051 «Економіка» / Укладач О.О.Горобець. Київ: НАСОА, 2022.

4. Тести для самоперевірки, завдання для поточного (модульного) та підсумкового контролю знань здобувачів (в електронній формі, розміщені на платформі Moodle).

12. Рекомендована література

Базова

1. Bruce P., Bruce A., Gedeck P. (2020). *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R Python*. O'Reilly Media, 368.
2. Wengrow J. *A Common-Sense Guide to Data Structures and Algorithms*. 2nd Ed. *Pragmatic Bookshelf*, 456.
3. White T. (2015). *Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale*. O'Reilly Media. 4th Edition, 756.
4. Бородкіна І. Л. *Теорія алгоритмів: навч. посіб.* Київ: Центр навчальної літератури, 2020. 184 с.
5. Мармоза А. *Теорія статистики*. Київ: Центр учбової літератури, 2021. 592 с.
6. Осауленко О. Г. *Офіційна статистика у системі національної інформаційної безпеки: моногр.* Київ: ТОВ «Август Трейд», 2017. 367 с.
7. Провост Ф. Фоусетт Т. *Data Science для бізнесу. Як збирати, аналізувати і використовувати дані*. Київ, Наш формат, 2019. 400 с.
8. Руденко В. *Математична статистика*. Київ: Центр навчальної літератури, 2019. 304 с.
9. Субботін С. О. *Нейронні мережі: теорія та практика: навч. посіб.* Житомир: Видавець О. О. Євенок, 2020. 184 с.

Допоміжна

1. Horobets O. (2021). Research Data as a Result of Research Activities: the Role and Significance for the Official Statistics, *Journal of the Knowledge Economy, Springer*, vol. 12(3), pages 1424-1436.
2. Indurkha N. (1997). *Predictive Data Mining: A Practical Guide (The Morgan Kaufmann Series in Data Management Systems)*. 1st Edition, 228.
3. Kamiński, B., Jakubczyk, M. & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Cent. Eur. J. Oper. Res.* 26, 135–159. <https://doi.org/10.1007/s10100-017-0479-6>

4. Koitzsch, K. (2017). Overview: Building Data Analytic Systems with Hadoop. In: Pro Hadoop Data Analytics. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-1910-2_1
5. Radermacher W. (2017). Official Statistics 4.0. Swiss Statistical Days 2017. doi: <https://doi.org/10.13140/RG.2.2.16604.9024>
6. Song D., Golin E. (1993). Fine-grain visualization algorithms in dataflow environments. *VIS '93: Proceedings of the 4th conference on Visualization '93*, pp.126–133.
7. Taleb, I., Serhani, M.A., Bouhaddioui, C. et al. (2021). Big Data quality framework: a holistic approach to continuous quality management. *J Big Data*, 8, 76. <https://doi.org/10.1186/s40537-021-00468-0>
8. Wu C., Buuya R., Ramamohanarao K. (2016). Chapter 1. Big Data Analytics = Machine Learning + Cloud Computing. *Big Data Principles and Paradigms*. Ed. R. Buuya R. N. Calheiros A. V. Dastjerdi. Elsevier. 3-38. Retrieved from: <https://www.sciencedirect.com/science/article/pii/B9780128053942000015>
9. Горобець О. О. Великі дані – джерело статистичної інформації: на прикладі книговидавничої галузі. *Науковий вісник Національної академії статистики, обліку та аудиту: зб. наук. пр.*. 2019. №1-2. С. 7-13.
10. Кейт О'Нілл. Big Data: зброя математичного знищення. Як великі дані збільшують нерівність і загрожують демократії. Київ: Bookchef, 2020. 336 с.
11. Нейронні мережі : теорія та практика: навч. посіб. Житомир : Вид. О. О. Євенок, 2020. 84 с.

Електронні ресурси

1. Amazon Web Services. URL: <https://aws.amazon.com>
2. Apache Hadoop. URL: <https://hadoop.apache.org/>
3. Apache Hive. URL: <https://hive.apache.org/>
4. Apache Spark. URL: <https://spark.apache.org/>
5. Azure. URL: <https://azure.microsoft.com>
6. Про офіційну статистику: Закон України від 16.08.2022 р. № 2524-IX. URL: <https://zakon.rada.gov.ua/laws/show/2524-20#Text>