

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
«ПОЛТАВСЬКА ПОЛІТЕХНІКА ІМЕНІ ЮРІЯ КОНДРАТЮКА»**

**Навчально-науковий інститут фінансів, економіки, управління та права
Кафедра економіки, підприємництва та маркетингу**



ЗАТВЕРДЖУЮ

Проректор з науково-педагогічної
та навчальної роботи

А.М. Мартиненко

« 30 » 2024 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«АНАЛІТИКА ВЕЛИКИХ ДАНИХ»

підготовки бакалавра

спеціальності 076 Підприємництво та торгівля

**Полтава
2024 рік**

Робоча програма навчальної дисципліни «Аналітика великих даних» для студентів спеціальності **076 Підприємництво та торгівля**, першого (бакалаврського) рівня вищої освіти. Складена відповідно до освітньо-професійної програми «Економіка бізнесу» 2024 року.

Розробник: Щербініна С.А., доцент кафедри економіки, підприємництва та маркетингу, кандидат економічних наук.

Погоджено:

Гарант освітньо-професійної програми  О.В. Хадарцев

Робоча програма затверджена на засіданні кафедри економіки, підприємництва та маркетингу

Протокол від «27» серпня 2024 року № 20

Завідувач кафедри економіки, підприємництва та маркетингу  М.Б. Чижевська

«27» серпня 2024 року

Схвалено навчально-методичною комісією Навчально-наукового інституту фінансів, економіки, управління та права

Протокол від «29» серпня 2024 року № 1

Голова навчально-методичної комісії ННІФЕУП  Є.А. Карпенко

«29» серпня 2024 року

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни		
		форма навчання		
		денна	заочна	дистанційна
Кількість кредитів – 4	Галузь знань 07 Управління та адміністрування	вибіркова		
Загальна кількість годин – 120				
Модулів – 1	Спеціальність 076 «Підприємництво та торгівля»	Рік підготовки:		
Змістових модулів – 2		2-й	2-й	2-й
		Семестр		
		4-й	4-й	4-й
		Лекції		
Індивідуальне завдання не передбачено	Ступінь вищої освіти – бакалавр	20 год.	6 год.	–
		Практичні		
		22 год.	6 год.	–
		Лабораторні		
		–	–	–
		Самостійна робота		
		78 год.	108 год.	120 год.
		Індивідуальна робота:		
		0 год.	0 год.	0 год.
		Вид контролю: диференційований залік		

Примітка.

Співвідношення кількості годин аудиторних занять до самостійної та індивідуальної роботи становить:

- для денної форми навчання – 42/78;
- для заочної форми навчання – 12/108;
- для дистанційної форми навчання – 0/120.

2. МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Навчальна дисципліна «Аналітика великих даних» відноситься до складу вибіркового компоненту освітньо-професійної програми бакалавра «Економіка бізнесу» зі спеціальності 076 «Підприємництво та торгівля» та має на меті формування у студентів системних знань та практичних навичок для ефективного збору, зберігання, обробки та аналізу великих масивів даних для прийняття обґрунтованих управлінських рішень в різних сферах діяльності.

Вивчення навчальної дисципліни забезпечує **формування та розвиток у студентів компетентностей**, визначених освітньо-професійною програмою:

Загальні компетентності (ЗК):

- Здатність застосовувати отримані знання в практичних ситуаціях.
- Навички використання інформаційних і комунікаційних технологій.
- Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

Спеціальні (фахові) компетентності спеціальності (СК):

- Здатність обирати та використовувати відповідні методи, інструментарій для обґрунтування рішень щодо створення, функціонування підприємницьких і торговельних структур.
- Здатність аналізувати, оцінювати, обирати і застосовувати ефективні методи та інструменти ведення бізнесу, визначати перспективні напрямки і шляхи розвитку підприємництва та торгівлі, обґрунтовувати доцільність їх реалізації.

3. ПЕРЕДУМОВИ ДЛЯ ВИВЧЕННЯ ДИСЦИПЛІНИ

Передумовами для вивчення дисципліни «Аналітика великих даних» є навчальна дисципліна, яка має бути вивчена студентами раніше: «Математика для економіки та управління».

4. ОЧІКУВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ З ДИСЦИПЛІНИ

Формулювання результатів навчання базується на результатах навчання, визначених освітньо-професійною програмою «Економіка бізнесу» (програмних результатах навчання) спеціальності 076 «Підприємництво та торгівля».

Програмні результати, а саме:

Використовувати сучасні комп'ютерні і телекомунікаційні технології обміну та розповсюдження професійно спрямованої інформації у сфері підприємництва і торгівлі.

Організовувати пошук, самостійний відбір, якісну обробку інформації з різних джерел для формування банків даних у сфері підприємництва та торгівлі.

Володіти методами та інструментарієм для обґрунтування управлінських рішень щодо створення й функціонування підприємницьких та торговельних структур.

Визначати і застосовувати ефективні методи та інструменти ведення бізнесу, визначати та обґрунтовувати перспективні напрямки і шляхи розвитку підприємництва та торгівлі.

5. КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Критерієм успішного проходження здобувачем освіти підсумкового оцінювання може бути досягнення ним мінімальних порогових рівнів оцінок за кожним запланованим результатом вивчення навчальної дисципліни.

Мінімальний пороговий рівень оцінки варто визначати за допомогою якісних критеріїв і трансформувати в мінімальну позитивну оцінку числової (рейтингової) шкали.

Сума балів	Значення ЄКТС	Оцінка	Критерій оцінювання	Рівень компетентності
90-100	A	Відмінно	Здобувач демонструє повні й міцні знання навчального матеріалу в	Високий, що повністю

Сума балів	Значення ЄКТС	Оцінка	Критерій оцінювання	Рівень компетентності
			обсязі, що відповідає робочій програмі дисципліни, правильно й обґрунтовано приймає необхідні рішення в різних нестандартних ситуаціях. Власні пропозиції здобувача в оцінках і вирішенні практичних задач підвищує його вміння використовувати знання, які він отримав при вивченні інших дисциплін.	забезпечує вимоги до знань, умінь і навичок, що викладені в робочій програмі дисципліни.
82-89	В	Дуже добре	Здобувач демонструє гарні знання, добре володіє матеріалом, що відповідає робочій програмі дисципліни, робить на їх основі аналіз можливих ситуацій та вміє застосовувати теоретичні положення при вирішенні практичних задач, але допускає окремі неточності. Вміє самостійно виправляти допущені помилки, кількість яких є незначною.	Достатній, що забезпечує здобувачу самостійне вирішення основних практичних задач.
74-81	С	Добре	Відповідь правильна, але має певні неточності, недостатньо обґрунтована та осмислена. Характеризують вміння застосовувати знання в практичній діяльності, але має певні неточності при реалізації двох-чотирьох вимог. Добра зрозумілість відповіді (доповіді) та доречна комунікаційна стратегія. Добре володіння компетенціями автономності та відповідальності	Достатній, конкретний рівень, за вивченим матеріалом робочої програми навчальної дисципліни. Додаткові питання про можливість використання теоретичних положень для практичного використання викликають ускладнення.
64-73	Д	Задовільно	Здобувач загалом добре володіє матеріалом, знає основні положення матеріалу, що відповідають робочій програмі дисципліни, робить на їх основі аналіз можливих ситуацій та використовує для рішення характерних/типових практичних завдань на професійному рівні. Додаткові питання про можливість використання теоретичних положень для практичного використання викликають ускладнення.	Достатній, конкретний рівень, за вивченим матеріалом робочої програми дисципліни.
60-63	Е	Достатньо	Здобувач засвоїв основний теоретичний матеріал, передбачений робочою програмою дисципліни, та розуміє постановку стандартних практичних завдань, має пропозиції щодо напрямку їх вирішень. Розуміє основні положення, що є визначальними в курсі, може вирішувати подібні завдання тим, що	Середній, що забезпечує достатньо надійний рівень відтворення основних положень дисципліни.

Сума балів	Значення ЄКТС	Оцінка	Критерій оцінювання	Рівень компетентності
			розглядалися з викладачем, але допускає значну кількість неточностей і грубих помилок, які може усувати за допомогою викладача.	
35-59	FX	Незадовільно з можливістю повторного складання заліку	Здобувач засвоїв основний теоретичний матеріал, передбачений робочою програмою дисципліни, та розуміє постановку стандартних практичних завдань, має пропозиції щодо напрямку їх вирішень і володіє основними положеннями на рівні, який визначається як мінімально допустимий. Правила вирішення практичних завдань з використанням основних теоретичних положень пояснюються з труднощами. Виконання практичних завдань значно формалізовано: є відповідність алгоритму, але відсутнє глибоке розуміння роботи та взаємозв'язків з іншими дисциплінами.	Середній, що є мінімально допустимим у всіх складових навчальної дисципліни.
0-34	FX	Незадовільно з обов'язковим повторним вивченням дисципліни	Здобувач повністю не виконав вимог робочої програми навчальної дисципліни. Його знання на підсумкових етапах навчання є фрагментарними. Здобувач не допущений до здачі заліку.	Незадовільний, здобувач не підготовлений до самостійного вирішення задач, які окреслює мета та завдання дисципліни.

6. ЗАСОБИ ДІАГНОСТИКИ РЕЗУЛЬТАТІВ НАВЧАННЯ

Засобами оцінювання та методами демонстрування результатів навчання є:
 диференційований залік;
 опитування;
 тестування;
 виконання практичних завдань;
 виконання контрольної роботи № 1, 2 (для студентів дистанційної форми навчання);
 виконання завдань самостійної роботи.

7. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Основи Big Data та методи аналізу.

Тема 1. Big Data – сутність та особливості аналізу.

Визначення Big Data. Основні характеристики Big Data. Джерела Big Data. Відмінності аналізу Big Data від традиційного аналізу даних. Основні методи аналізу Big Data: машинне навчання, статистичний аналіз, інтелектуальний аналіз даних (Data Mining). Задачі, що вирішуються за допомогою аналізу Big Data. Приклади використання Big Data в різних галузях. Переваги та обмеження використання Big Data.

Практичне заняття № 1.

Тема 2. Методи і моделі інтелектуального аналізу Big Data в економіці.

Методи Data Mining: кластеризація (clustering), класифікація (classification), асоціація (associations), прогнозування (forecasting). Машинне навчання (machine learning): навчання з

учителем (supervised learning), навчання без учителя (unsupervised learning), навчання з підкріпленням (reinforcement learning). Статистичний аналіз: регресійний аналіз, кореляційний аналіз, аналіз часових рядів. Технології Text Mining. Моделі інтелектуального аналізу Big Data в економіці: моделі прогнозування попиту; моделі оцінки ризиків; моделі сегментації клієнтів; моделі виявлення шахрайства; моделі для аналізу фінансових ринків.

Практичне заняття № 2.

Тема 3. Передобробка великих даних.

Проблеми, пов'язані з «сирими» даними (шум, пропущені значення, невідповідності). Вплив якості даних на результати аналізу. Основні етапи передобробки. Очищення даних (data cleaning): видалення дублікатів; обробка пропущених значень (imputation); виявлення та виправлення аномалій (outlier detection); виправлення невідповідностей. Інтеграція даних (data integration): об'єднання даних з різних джерел; вирішення конфліктів у даних. Трансформація даних (data transformation): нормалізація та стандартизація; дискретизація; агрегація. Зменшення розмірності даних (data reduction): агрегація даних; видалення надлишкових атрибутів; використання методів головних компонент (PCA).

Практичне заняття № 3.

Тема 4. Технології інтелектуальної обробки даних: Text, Data, Web, Social Mining.

Визначення та основні завдання Text Mining, Data Mining, Web Mining, Social Mining. Методи та алгоритми Text Mining: обробка природної мови (NLP); аналіз тональності (sentiment analysis); тематичне моделювання (topic modeling); вилучення інформації (information extraction). Методи та алгоритми Data Mining: кластеризація (clustering); класифікація (classification); асоціативний аналіз (association analysis); прогнозування (forecasting). Методи та алгоритми Web Mining: аналіз структури веб-сайтів (web structure mining); аналіз контенту веб-сайтів (web content mining); аналіз поведінки користувачів (web usage mining). Методи та алгоритми Social Mining: аналіз соціальних мереж (social network analysis); виявлення спільнот (community detection); аналіз впливу (influence analysis); аналіз соціальних настроїв. Практичне застосування.

Практичне заняття № 4.

Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил.

Визначення кластеризації та її основні завдання. Типи алгоритмів кластеризації: ієрархічні алгоритми (агломеративні та дивізійні); алгоритми, засновані на центроїдах (k-середніх, k-медоїдів); алгоритми, засновані на щільності (DBSCAN); метрики відстані та їх вплив на результати кластеризації. Оцінка якості кластеризації (силуетний аналіз, індекс Девіса-Болдіна). Визначення асоціативного аналізу та його основні завдання. Алгоритми пошуку асоціативних правил: алгоритм Apriori, Алгоритм FP-growth. Практичне застосування.

Практичне заняття № 5-6.

Змістовий модуль 2. Інструменти та безпека Big Data.

Тема 6. Програмні інструменти Big Data.

Платформи для зберігання та обробки Big Data. Apache Hadoop. Apache Spark. NoSQL бази даних. Інструменти для аналізу даних. Python та його бібліотеки. R та його пакети. Інструменти візуалізації даних. Інструменти для обробки поточкових даних. Хмарні платформи для Big Data.

Практичне заняття № 7.

Тема 7. Аналіз даних засобами мови Python.

Основні бібліотеки Python для аналізу даних. NumPy: робота з масивами; математичні операції. Pandas: робота з табличними даними (DataFrame); очищення та трансформація даних; аналіз даних. Matplotlib та Seaborn: візуалізація даних; побудова графіків та діаграм. Практичні методи аналізу даних за допомогою Python. Імпорт та експорт даних (CSV, Excel, JSON). Очищення та підготовка даних (обробка пропущених значень, видалення дублікатів).

Статистичний аналіз даних (розрахунок середнього, медіани, стандартного відхилення). Групування та агрегація даних. Візуалізація даних для виявлення закономірностей. Приклади аналізу даних у Python. Аналіз даних продажів. Аналіз даних соціальних мереж. Аналіз фінансових даних.

Практичне заняття № 8.

Тема 8. Пакет R для аналізу Big Data.

Основи R для аналізу даних. Статистичний аналіз в R. Описова статистика. Візуалізація даних. Регресійний аналіз. Аналіз часових рядів. Кластерний аналіз. R для Big Data. Обробка великих даних за допомогою пакетів (data.table, bigmemory). Інтеграція R з Hadoop та Spark (rhdfs, sparklyr). Паралельні обчислення в R (parallel, foreach). Використання R в хмарних середовищах. Практичні приклади аналізу Big Data в R. Аналіз даних соціальних мереж. Аналіз фінансових даних. Аналіз веб-трафіку.

Практичне заняття № 9.

Тема 9. Візуалізація великих даних.

Основні принципи візуалізації великих даних. Вибір правильного типу візуалізації для різних типів даних. Принципи ефективного дизайну візуалізації (ясність, простота, інформативність). Інтерактивність та дослідження даних через візуалізацію. Інструменти візуалізації. Типи візуалізації для великих даних. Графіки та діаграми для числової інформації. Візуалізація категоріальних даних. Візуалізація геопросторових даних. Візуалізація мережевих даних. Візуалізація текстових даних. Інструменти та технології візуалізації. Python (Matplotlib, Seaborn, Plotly, Bokeh). R (ggplot2, Shiny). Хмарні платформи візуалізації. Практичні приклади візуалізації великих даних.

Практичне заняття № 10.

Тема 10. Інформаційна безпека у сфері Big Data. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data.

Особливості Big Data та їх вплив на інформаційну безпеку. Загрози та вразливості в сфері Big Data. Загрози конфіденційності даних: несанкціонований доступ; витік даних; ідентифікація особистості за анонімізованими даними. Загрози цілісності даних: маніпуляції з даними; помилки в даних; вплив шкідливого ПЗ. Загрози доступності даних: DoS-атаки; втрата даних через технічні збої; вплив людського фактору. Вразливості програмного забезпечення та інфраструктури. Підходи до забезпечення безпеки Big Data. Шифрування даних (на етапі зберігання та передачі). Контроль доступу та аутентифікація. Анонімізація та псевдонімізація даних. Моніторинг та аудит безпеки. Виявлення та запобігання вторгненням. Забезпечення фізичної безпеки інфраструктури. Розробка політик та процедур безпеки. Використання технологій машинного навчання для виявлення аномалій. Нормативно-правове регулювання.

Практичне заняття № 11.

8. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

а) для денної форми навчання

Назви змістових модулів і тем	Кількість годин					
	денна форма					
	усього	у тому числі				
лек		пр	лаб	інд	с.р.	
Змістовий модуль 1. Основи Big Data та методи аналізу.						
Тема 1. Big Data – сутність та особливості аналізу.	12	2	2	-	-	8
Тема 2. Методи і моделі інтелектуального аналізу Big Data в економіці.	12	2	2	-	-	8
Тема 3. Передобробка великих даних.	12	2	2	-	-	8
Тема 4. Технології інтелектуальної обробки даних: Text, Data, Web, Social Mining.	12	2	2	-	-	8
Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил.	14	2	4	-	-	8
Разом за змістовим модулем 1	62	10	12	-	-	40
Змістовий модуль 2. Основні елементи системи управління витратами						
Тема 6. Програмні інструменти Big Data.	12	2	2	-	-	8
Тема 7. Аналіз даних засобами мови Python.	12	2	2	-	-	8
Тема 8. Пакет R для аналізу Big Data.	12	2	2	-	-	8
Тема 9. Візуалізація великих даних.	10	2	2	-	-	6
Тема 10. Інформаційна безпека у сфері Big Data. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data.	12	2	2	-	-	8
Разом за змістовим модулем 2	58	10	10	-	-	38
Усього годин	120	20	22	-	-	78

б) для заочної форми навчання

Назви змістових модулів і тем	Кількість годин					
	заочна форма					
	усього	у тому числі				
лек		пр	лаб	інд	с.р.	
Змістовий модуль 1. Основи Big Data та методи аналізу.						
Тема 1. Big Data – сутність та особливості аналізу.	12	-	-	-	-	12
Тема 2. Методи і моделі інтелектуального аналізу Big Data в економіці.	12	2	2	-	-	8
Тема 3. Передобробка великих даних.	12	-	-	-	-	12
Тема 4. Технології інтелектуальної обробки даних: Text, Data, Web, Social Mining.	12	2	2	-	-	8
Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил.	14	2	2	-	-	10
Разом за змістовим модулем 1	62	6	6	-	-	50
Змістовий модуль 2. Основні елементи системи управління витратами						
Тема 6. Програмні інструменти Big Data.	12	-	-	-	-	12
Тема 7. Аналіз даних засобами мови Python.	12	-	-	-	-	12
Тема 8. Пакет R для аналізу Big Data.	12	-	-	-	-	12
Тема 9. Візуалізація великих даних.	10	-	-	-	-	10

Назви змістових модулів і тем	Кількість годин					
	заочна форма					
	усього	у тому числі				
		лек	пр	лаб	інд	с.р.
Тема 10. Інформаційна безпека у сфері Big Data. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data.	12	-	-	-	-	12
Разом за змістовим модулем 2	58	-	-	-	-	58
Усього годин	120	6	6	-	-	108

в) для дистанційної форми навчання

Назви змістових модулів і тем	Кількість годин					
	дистанційна форма					
	усього	у тому числі				
лек		пр	лаб	інд	с.р.	
Змістовий модуль 1. Основи Big Data та методи аналізу.						
Тема 1. Big Data – сутність та особливості аналізу.	12	-	-	-	-	12
Тема 2. Методи і моделі інтелектуального аналізу Big Data в економіці.	12	-	-	-	-	12
Тема 3. Передобробка великих даних.	12	-	-	-	-	12
Тема 4. Технології інтелектуальної обробки даних: Text, Data, Web, Social Mining.	12	-	-	-	-	12
Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил.	14	-	-	-	-	14
Разом за змістовим модулем 1	62	-	-	-	-	62
Змістовий модуль 2. Основні елементи системи управління витратами						
Тема 6. Програмні інструменти Big Data.	12	-	-	-	-	12
Тема 7. Аналіз даних засобами мови Python.	12	-	-	-	-	12
Тема 8. Пакет R для аналізу Big Data.	12	-	-	-	-	12
Тема 9. Візуалізація великих даних.	10	-	-	-	-	10
Тема 10. Інформаційна безпека у сфері Big Data. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data.	12	-	-	-	-	12
Разом за змістовим модулем 2	58	-	-	-	-	58
Усього годин	120	-	-	-	-	120

9. ПЕРЕЛІК ПИТАНЬ ДЛЯ СЕМІНАРСЬКИХ ЗАНЯТЬ

№ з/п	Назва питань (тем)	Кількість годин		
		форма навчання		
		денна	заочна	дистанційна
	Семінарські заняття не передбачені			

10. ПЕРЕЛІК ПИТАНЬ ДЛЯ ПРАКТИЧНИХ ЗАНЯТЬ

№ з/п	Назва теми	Кількість годин		
		Форма навчання		
		денна	заочна	дистанційна
1	Тема 1. Big Data – сутність та особливості аналізу. Визначення Big Data та аналіз її характеристик на прикладі реальних наборів даних.	2	—	—
2	Тема 2. Методи і моделі інтелектуального аналізу Big Data в економіці. Прогнозування економічних показників за допомогою методів машинного навчання.	2	2	—
3	Тема 3. Передобробка великих даних. Підготовка даних електронної комерції для аналізу.	2	—	—
4	Тема 4. Технології інтелектуальної обробки даних: Text, Data, Web, Social Mining. Видобування знань з даних за допомогою методів Data Mining.	2	2	—
5	Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил. Кластеризація даних за допомогою алгоритму k-середніх.	2	2	—
6	Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил. Застосування асоціативного аналізу для аналізу кошика покупок.	2	—	—
7	Тема 6. Програмні інструменти Big Data. Робота з Apache Hadoop для обробки великих наборів даних.	2	—	—
8	Тема 7. Аналіз даних засобами мови Python. Аналіз та візуалізація даних з використанням Pandas та Matplotlib.	2	—	—
9	Тема 8. Пакет R для аналізу Big Data. Основи роботи з R для аналізу великих даних.	2	—	—
10	Тема 9. Візуалізація великих даних. Візуалізація великих даних за допомогою R.	2	—	—
11	Тема 10. Інформаційна безпека у сфері Big Data. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data. Аналіз загроз та вразливостей Big Data систем.	2	—	—
	Разом	22	6	—

11. ПЕРЕЛІК ПИТАНЬ ДЛЯ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ з/п	Назва питань (тем)	Кількість годин		
		форма навчання		
		денна	заочна	дистанційна
	Лабораторні заняття не передбачені			

12. САМОСТІЙНА РОБОТА

Метою самостійної роботи студента є: навчитися користуватися бібліотечними фондами і каталогами, працювати з історичними та літературними джерелами, складати конспекти, аналізувати матеріал, порівнювати різні наукові концепції та робити висновки.

Види самостійної роботи студента:

- опрацювання лекційного матеріалу;

- підготовка до практичних занять;
- опрацювання тем курсу, які виносяться на самостійне вивчення, за списками літератури, рекомендованими в робочій програмі навчальної дисципліни;
- підготовка до поточного тестування;
- відвідування консультацій (згідно графіку консультацій кафедри);
- підготовка двох контрольних робіт для здобувачів дистанційної форми навчання;
- підготовка до складання підсумкового контролю за контрольними питаннями.

**Питання
для самостійного вивчення студентами**

№ з/п	Назва теми	Кількість годин		
		Форма навчання		
		денна	заочна	дистан- ційна
1	Тема 1. Big Data – сутність та особливості аналізу. Основні відмінності Big Data у порівнянні з традиційними базами даних. Концепція «5V» Big Data. Основні джерела Big Data. Основні методи аналізу Big Data. Етичні та юридичні проблеми пов'язані з використанням Big Data. Перспективи розвитку Big Data.	8	12	12
2	Тема 2. Методи і моделі інтелектуального аналізу Big Data в економіці. Дослідження алгоритмів класифікації (логістична регресія, дерева рішень, випадковий ліс, нейронні мережі) та їх застосування для прогнозування кредитних ризиків, виявлення шахрайства та класифікації економічних даних. Аналіз метрик оцінки якості класифікації (точність, повнота, F1-міра, AUC-ROC) та їх застосування в економічних дослідженнях.	8	8	12
3	Тема 3. Передобробка великих даних. Дослідження методів обробки пропущених значень (заміна середнім, медіаною, модою, прогнозування). Виявлення та виправлення аномалій (використання статистичних методів, візуалізація). Видалення дублікатів та невідповідностей у даних. Аналіз впливу якості даних на результати аналізу. Дослідження методів об'єднання даних з різних джерел (злиття, об'єднання, конкатенація). Вирішення конфліктів у даних (вибір правильних значень, узгодження форматів). Аналіз проблем, пов'язаних з інтеграцією даних. Дослідження методів нормалізації та стандартизації даних (min-max масштабування, Z-масштабування). Дискретизація та бінаризація даних. Агрегація та узагальнення даних. Аналіз впливу трансформації даних на результати аналізу. Дослідження методів видалення надлишкових атрибутів (фільтрація, обгортка, вбудовані методи). Використання методів головних компонент (PCA) для зменшення розмірності даних. Аналіз впливу зменшення розмірності на результати аналізу.	8	12	12
4	Тема 4. Технології інтелектуальної обробки даних: Text, Data, Web, Social Mining. Дослідження методів обробки природної мови (NLP) для аналізу текстових даних. Вивчення алгоритмів аналізу тональності та їх застосування. Огляд методів тематичного	8	8	12

№ з/п	Назва теми	Кількість годин		
		Форма навчання		
		денна	заочна	дистан- ційна
	моделювання та вилучення інформації. Дослідження основних методів Data Mining (кластеризація, класифікація, асоціативний аналіз, прогнозування). Огляд методів класифікації (логістична регресія, дерева рішень, нейронні мережі) та їх застосування. Аналіз застосування Data Mining в різних галузях (аналіз даних продажів, клієнтської бази). Дослідження методів аналізу структури веб-сайтів. Вивчення методів аналізу контенту веб-сайтів. Дослідження методів аналізу соціальних мереж. Аналіз застосування Social Mining в різних галузях (аналіз трендів, громадської думки).			
5	Тема 5. Алгоритми кластеризації. Асоціація, алгоритми пошуку асоціативних правил. Дослідження ієрархічних алгоритмів кластеризації (агломеративні та дивізійні). Вивчення алгоритмів, заснованих на центроїдах (k-середніх, k-медоїдів). Огляд алгоритмів, заснованих на щільності (DBSCAN). Аналіз метрик відстані та їх вплив на результати кластеризації. Вивчення методів оцінки якості кластеризації (силуетний аналіз, індекс Девіса-Болдіна). Дослідження алгоритму Apriori для пошуку асоціативних правил. Вивчення алгоритму FP-growth для пошуку асоціативних правил. Аналіз основних понять асоціативного аналізу (елементи, транзакції, підтримка, достовірність, ліфт). Порівняльний аналіз алгоритмів пошуку асоціативних правил.	8	10	14
6	Тема 6. Програмні інструменти Big Data. Дослідження архітектури та компонентів Hadoop. Вивчення принципів розподіленого зберігання та обробки даних у Hadoop. Аналіз застосування Hadoop для обробки великих наборів даних. Дослідження архітектури та компонентів Spark. Вивчення принципів обробки даних в пам'яті та паралельних обчислень у Spark. Аналіз застосування Spark для аналізу великих даних у реальному часі та машинного навчання. Дослідження різних типів NoSQL баз даних (ключ-значення, документо-орієнтовані, стовпцеві, графові). Вивчення принципів зберігання та запитів до даних у NoSQL базах даних. Аналіз застосування NoSQL баз даних для зберігання та обробки Big Data.	8	12	12
7	Тема 7. Аналіз даних засобами мови Python. Вивчення основних типів даних Python (списки, словники, кортежі) та їх використання для зберігання та обробки даних. Вивчення основних операторів та функцій Python для маніпулювання даними. Вивчення бібліотеки NumPy для роботи з масивами та виконання математичних операцій. Вивчення структури даних DataFrame та її використання для зберігання та обробки табличних даних. Вивчення методів Pandas для очищення, трансформації та агрегації даних. Вивчення методів Pandas для аналізу даних (статистичний аналіз, групування, зведені таблиці).	8	12	12

№ з/п	Назва теми	Кількість годин		
		Форма навчання		
		денна	заочна	дистанційна
8	Тема 8. Пакет R для аналізу Big Data. Використання пакету caret для навчання моделей машинного навчання на великих наборах даних. Оптимізація моделей машинного навчання для обробки великих даних. Використання пакету mlr для машинного навчання на великих даних.	8	12	12
9	Тема 9. Візуалізація великих даних. Дослідження принципів інтерактивної візуалізації та її переваги. Вивчення методів створення інтерактивних графіків та діаграм. Аналіз інструментів для створення інтерактивних дашбордів. Дослідження сервісів візуалізації даних в хмарних платформах AWS, Azure та GCP. Вивчення методів інтеграції інструментів візуалізації з хмарними сервісами. Аналіз переваг та недоліків візуалізації даних в хмарних середовищах.	6	10	10
10	Тема 10. Інформаційна безпека у сфері Big Data. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data. Дослідження методів шифрування даних (на етапі зберігання та передачі). Вивчення методів контролю доступу та аутентифікації. Аналіз методів анонімізації та псевдонімізації даних. Дослідження методів моніторингу та аудиту безпеки. Вивчення методів виявлення та запобігання вторгненням. Аналіз методів забезпечення фізичної безпеки інфраструктури. Дослідження методів розробки політик та процедур безпеки. Вивчення методів використання технологій машинного навчання для виявлення аномалій. Аналіз нормативно-правового регулювання у сфері безпеки Big Data.	8	12	12
	Всього	78	108	120

13. ІНДИВІДУАЛЬНІ ЗАВДАННЯ

Індивідуальне завдання не передбачено планом.

14. МЕТОДИ НАВЧАННЯ

При викладанні дисципліни застосовуються словесні, наочні та практичні методи навчання.

Словесні і наочні використовуються під час лекцій та інструктажів, практичні при проведенні практичних занять.

Під час проведення лекцій використовуються такі словесні методи як розповідь, пояснення та наочні методи: ілюстрація, демонстрація.

Перед проведенням практичних занять викладачами проводяться інструктажі: вступні, поточні, підсумкові.

15. МЕТОДИ КОНТРОЛЮ

Поточний контроль успішності засвоєння здобувачами навчального матеріалу може здійснюватися шляхом опитування й оцінювання знань студентів під час практичних занять, оцінювання виконання студентами самостійної роботи та індивідуальних завдань, проведення

і перевірки письмових контрольних робіт, тестування або в ході індивідуальних співбесід зі студентами під час консультацій. Вибір конкретних форм і методів поточного контролю знань студентів залежить від викладача і доводиться до їхнього відома на першому практичному занятті. Модульний контроль є частиною поточного контролю і має на меті перевірку засвоєння студентом певної сукупності знань та вмінь, що формують відповідний модуль. Він реалізується шляхом проведення спеціальних контрольних заходів (у формі тестування чи написання студентами контрольних робіт), проводиться наприкінці змістового модулю за рахунок аудиторних занять або самостійної роботи для дистанційної форми навчання, під час групових консультацій або ж за рахунок часу, відведеного на самостійну роботу студентів. На підставі результатів модульного контролю здійснюється міжсесійний контроль (атестація).

Підсумковий контроль здійснюється у формі семестрового диференційованого заліку.

16. РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ СТУДЕНТИ

**Нарахування балів з навчальної дисципліни «Аналітика великих даних»
за видами робіт
денна форма навчання**

Вид робіт (контролю)	Перелік тем									
	Тема 1	Тема 2	Тема 3	Тема 4	Тема 5	Тема 6	Тема 7	Тема 8	Тема 9	Тема 10
	Практичне заняття									
	1	2	3	4	5	6	7	8	9	10
Виконання практичних завдань	3	3	3	3	3	3	3	3	3	3
Тестування					5				5	5
Виконання завдань самостійної роботи	2	2	2	2	2	2	2	2	2	2
Всього за темами	5	5	5	5	15	5	5	5	10	10
Підсумковий контроль знань	30									
Всього за результатами вивчення навчальної дисципліни	100									

Шкала та критерії оцінювання виконання практичних завдань

Бали	Критерії оцінювання
3	Завдання виконано в повному обсязі, без помилок. Оформлення відповідає всім вимогам. Робота чітка, логічно послідовна, демонструє правильне застосування знань. Висновки повно і коректно розкривають суть завдання, є наочним прикладом для подальшого використання. Завдання виконано вчасно.
1,5	Завдання виконано частково або з незначними помилками. Оформлення відповідає основним вимогам, але є дрібні недоліки. Висновки неповно розкривають суть завдання або містять незначні неточності. Робота потребує деяких коригувань або доповнень.
0	Завдання не виконано або виконано з суттєвими помилками. Оформлення роботи неналежне або неповне. Висновки відсутні або некоректні, завдання не відповідає вимогам і не може бути оцінене через значні помилки або відсутність виконання.

Оцінювання тестування:

- кожна правильна відповідь оцінюється у фіксовану кількість балів ($0,25 \times 20 = 5$);
- правильність відповідей перевіряється відповідно до ключа тестів.

Шкала та критерії оцінювання виконання завдань самостійної роботи

Бали	Критерії оцінювання
2	Завдання виконано в повному обсязі, без помилок. Робота має високий рівень логічної послідовності, глибокий аналіз та чітке оформлення. Висновки розкривають сутність завдання, демонструючи здатність до критичного осмислення матеріалу та самостійного застосування отриманих знань у практичному контексті.
1	Завдання виконано частково, містить суттєві помилки або недоліки в оформленні. Висновки недостатньо розкривають суть завдання, важливі аспекти подано фрагментарно. Робота демонструє певний рівень розуміння матеріалу, але потребує значних виправлень або уточнень.
0	Завдання не виконано або виконано з критичними помилками, що унеможливають коректне розуміння роботи. Оформлення не відповідає вимогам, а висновки відсутні або некоректні. Робота не демонструє засвоєння матеріалу.

Нарахування балів з навчальної дисципліни «Аналітика великих даних» за видами робіт заочна форма навчання

Вид робіт (контролю)	Перелік тем									
	Тема 1	Тема 2	Тема 3	Тема 4	Тема 5	Тема 6	Тема 7	Тема 8	Тема 9	Тема 10
	Практичне заняття									
		1		2	3					
Виконання практичних завдань	-	4	-	4	4	-	-	-	-	-
Тестування					9					9
Виконання завдань самостійної роботи	4	4	4	4	4	4	4	4	4	4
Всього за темами	4	8	4	8	17	4	4	4	4	13
Підсумковий контроль знань	30									
Всього за результатами вивчення навчальної дисципліни	70									

Шкала та критерії оцінювання виконання практичних завдань

Бали	Критерії оцінювання
4	Завдання виконано в повному обсязі, без помилок. Оформлення відповідає всім вимогам. Робота чітка, логічно послідовна, демонструє правильне застосування знань. Висновки повно і коректно розкривають суть завдання, є наочним прикладом для подальшого використання. Завдання виконано вчасно.
2	Завдання виконано частково або з незначними помилками. Оформлення відповідає основним вимогам, але є дрібні недоліки. Висновки неповно розкривають суть завдання або містять незначні неточності. Робота потребує деяких коригувань або доповнень.
0	Завдання не виконано або виконано з суттєвими помилками. Оформлення роботи неналежне або неповне. Висновки відсутні або некоректні, завдання не відповідає вимогам і не може бути оцінене через значні помилки або відсутність виконання.

Оцінювання тестування:

- кожна правильна відповідь оцінюється у фіксовану кількість балів ($0,45 \times 20 = 9$);
- правильність відповідей перевіряється відповідно до ключа тестів.

Шкала та критерії оцінювання виконання завдань самостійної роботи

Бали	Критерії оцінювання
4	Завдання виконано в повному обсязі, без помилок. Робота має високий рівень логічної послідовності, глибокий аналіз та чітке оформлення. Висновки розкривають сутність завдання, демонструючи здатність до критичного осмислення матеріалу та самостійного застосування отриманих знань у практичному контексті.
2	Завдання виконано частково, містить суттєві помилки або недоліки в оформленні. Висновки недостатньо розкривають суть завдання, важливі аспекти подано фрагментарно. Робота демонструє певний рівень розуміння матеріалу, але потребує значних виправлень або уточнень.
0	Завдання не виконано або виконано з критичними помилками, що унеможливають коректне розуміння роботи. Оформлення не відповідає вимогам, а висновки відсутні або некоректні. Робота не демонструє засвоєння матеріалу.

Нарахування балів з навчальної дисципліни «Аналітика великих даних» за видами робіт дистанційна форма навчання

Вид робіт (контролю)	Перелік тем									
	Тема 1	Тема 2	Тема 3	Тема 4	Тема 5	Тема 6	Тема 7	Тема 8	Тема 9	Тема 10
Виконання контрольної роботи	35					35				
Всього за темами	35					35				
Підсумковий контроль знань	30									
Всього за результатами вивчення навчальної дисципліни	70									

Шкала та критерії оцінювання виконання контрольної роботи

Бали	Критерії оцінювання
35	Відповідь надана у письмовій формі, повна (не менше 95% потрібної інформації), без помилок або з незначними неточностями. Виклад матеріалу чіткий, логічний, із правильним використанням термінів та обґрунтованими висновками
30	Відповідь надана у письмовій формі, достатньо повна (не менше 85% потрібної інформації). Можливі незначні помилки або неточності, які не впливають на загальне розуміння матеріалу. Виклад логічний, є аргументовані висновки
25	Відповідь надана у письмовій формі, частково неповна (не менше 70% потрібної інформації). Можливі окремі помилки або неточності у формулюваннях. Аналіз та висновки потребують уточнення або доповнення
20	Відповідь надана у письмовій формі, містить лише ключові моменти (не менше 50% потрібної інформації), але є суттєві недоліки. Можливі логічні помилки, недостатнє обґрунтування висновків
15	Відповідь надана у письмовій формі, але коротка (менше 40% потрібної інформації). Присутні значні помилки або пропущені важливі аспекти. Виклад непослідовний, висновки слабо аргументовані
10	Відповідь містить лише основні тези (менше 30% потрібної інформації). Наявні серйозні логічні помилки, виклад неструктурований, висновки поверхневі або слабо аргументовані
5	Відповідь містить лише окремі спроби відтворення матеріалу, але вони не розкривають суті питання. Значні логічні помилки, виклад хаотичний або уривчастий
0	Відповідь не надана або містить виключно несуттєву інформацію, що не дозволяє оцінити знання з теми

Оцінювання знань студентів за рейтинговою оцінкою

Рейтингова оцінка	Оцінка за шкалою ECTS	Оцінка за національною шкалою
90 – 100	A – відмінно	5 – відмінно
82 – 89	B – дуже добре	4 – добре
74 – 81	C – добре	
64 – 73	D – задовільно	3 – задовільно
60 – 63	E – достатньо	
35 – 59	FX – незадовільно з можливістю повторної атестації	2 – незадовільно
0 - 34	F – незадовільно з обов’язковим повторним вивчення дисципліни	

Шкала оцінювання знань студентів за рейтинговою оцінкою: національна та ECTS

Загальна трудомісткість дисципліни – 100 балів. До 70 балів студент отримує протягом семестру в процесі поточного контролю знань та до 30 балів – в результаті підсумкового контролю знань.

1. Поточний контроль. Бали, отримані впродовж семестру, за видами навчальної діяльності розподіляються наступним чином (розподіл орієнтовний): робота на практичних заняттях, опитування, тестування, виконання практичних завдань, поточної контрольної роботи, завдань самостійної роботи – до 70 балів.

Виконання контрольних робіт для дистанційної форми навчання – до 70 балів.

Присутність на лекціях і практичних заняттях не оцінюється в балах. Пропуски занять підлягають обов’язковому відпрацюванню в індивідуальному порядку під час консультацій. Пропущене заняття має бути відпрацьоване впродовж двох наступних тижнів, при тривалій відсутності студента на заняттях з поважної причини встановлюється індивідуальний графік відпрацювання пропусків, але не пізніше початку екзаменаційної сесії.

Студент, який повністю виконав програму навчальної дисципліни і отримав достатню рейтингову оцінку (не менше 25 балів), допускається до підсумкового контролю з дисципліни.

2. Підсумковий контроль. Підсумковим контролем є диференційований залік. Він здійснюється відповідно до вимог «Положення про організацію освітнього процесу в Національному університеті «Полтавська політехніка імені Юрія Кондратюка»..

17. МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ

1. Щербініна С.А. Робоча програма навчальної дисципліни «Аналітика великих даних» для студентів першого (бакалаврського) рівня вищої освіти спеціальності 076 «Підприємництво та торгівля». Полтава: Національний університет імені Юрія Кондратюка, 2024. 19 с.

18. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. Харків: ХНУРЕ, 2021. 92 с.
2. Фісун М.Т., Кравець І.О., Казмірчук П.П., Ніколенко С.Г. Інт Олещенко Л.М. Технології оброблення великих даних. Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с.
3. Waller M. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*. 2022. № 74(2). P. 77-84.
4. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications: Springer, 2020. 298 p. Електронний ресурс. Режим доступу: <https://vdoc.pub/download/big-data-conceptual-analysis-and-applications-40crds1qhta0>

5. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark: Springer, 2019. 105 p.

Допоміжна

1. Joel Grus. Data Science from Scratch. First Principles with Python. O'Reilly Media. 2019. 403 p.
2. Чаплінський В. Р., Кушнір О. К., Свідер О. П. Аналіз великих даних та їх візуалізація для потреб бізнесу. *Ефективна економіка*. 2021. № 6. URL: <http://www.economy.nayka.com.ua/?op=1&z=8979>. DOI: [10.32702/2307-2105-2021.6.89](https://doi.org/10.32702/2307-2105-2021.6.89)
3. Балабанов О.С. Аналітика великих даних: принципи, напрямки і задачі (огляд). *Проблеми програмування*. 2019. № 2. С. 47-68.
4. Ланде Д.В., Гладун А.Я., Субач І.Ю. Аналітика Big Data для кібербезпеки: навч.-методич. посіб. Київ: Міжнародний науково-навчальний центр інформаційних технологій та систем НАН України та МОН України, 2021. 220 с.
5. Романюк О., Павлов С., Бобко О., Завальнюк Є., Решетнік О. Аналіз великих даних у комп'ютерній графіці, *Оптико-електронні інформаційно-енергетичні технології*, Том. 47, № 1, 2024. с. 50–57.

19. ІНФОРМАЦІЙНІ РЕСУРСИ

1. Дистанційний курс навчальної дисципліни «Аналітика великих даних». URL: <https://dist.nupp.edu.ua/course/view.php?id=6577>