

Міністерство освіти і науки України
Національний технічний університет
“Дніпровська політехніка”

Кафедра інформаційних технологій та комп’ютерної інженерії



«ЗАТВЕРДЖЕНО»

завідувач кафедри

Гнатушенко В.В.

« 30 » _____ червня 2022 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
“Методи та техніки аналізу великих даних”

Галузь знань	12 Інформаційні технології
Спеціальність	всі спеціальності галузі
Рівень вищої освіти	другий (магістерський)
Статус.....	вибіркова
Загальний обсяг	4 кредити ЄКТС (120 годин)
Форма підсумкового контролю	диференційований залік
Термін викладання	2-й семестр
Мова викладання	українська

Викладач: доц. Кожевников А.В.

Пролонговано: на 20__/20__ н.р. _____ (_____) “__” 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) “__” 20__ р.
(підпис, ПІБ, дата)

Дніпро
НТУ «ДП»
2022

Робоча програма навчальної дисципліни “Методи та техніки аналізу великих даних” для магістрів галузі знань 12 Інформаційні технології / Нац. техн. ун-т. “Дніпровська політехніка”, каф. ІТКІ. – Д. : НТУ “ДП”, 2022. – 12 с.

Розробники: Кожевников А.В. – доцент, кандидат технічних наук, доцент кафедри інформаційних технологій та комп’ютерної інженерії.

Робоча програма регламентує:

- мету дисципліни;
- дисциплінарні результати навчання;
- базові дисципліни;
- обсяг і розподіл за формами організації освітнього процесу та видами навчальних занять;
- програму дисципліни (тематичний план за видами навчальних занять);
- алгоритм оцінювання рівня досягнення дисциплінарних результатів навчання (шкали, засоби, процедури та критерії оцінювання);
- інструменти, обладнання та програмне забезпечення;
- рекомендовані джерела інформації.

Робоча програма призначена для реалізації компетентнісного підходу під час планування освітнього процесу, викладання дисципліни, підготовки студентів до контрольних заходів, контролю провадження освітньої діяльності, внутрішнього та зовнішнього контролю забезпечення якості вищої освіти, акредитації освітніх програм у межах спеціальності.

ЗМІСТ

1 МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ.....	4
2. ОЧІКУВАНІ ДИСЦИПЛІНАРНІ РЕЗУЛЬТАТИ НАВЧАННЯ.	4
3 БАЗОВІ ДИСЦИПЛІНИ.....	4
4 ОБСЯГ І РОЗПОДІЛ ЗА ФОРМАМИ ОРГАНІЗАЦІЇ ОСВІТНЬОГО ПРОЦЕСУ ТА ВИДАМИ НАВЧАЛЬНИХ ЗАНЯТЬ.....	4
5 ПРОГРАМА ДИСЦИПЛІНИ ЗА ВИДАМИ НАВЧАЛЬНИХ ЗАНЯТЬ.....	5
6 ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ.....	7
6.1 Шкали.....	7
6.2 Засоби та процедури.....	7
6.3 Критерії.....	8
7 ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ...	12
8 РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ	12

1 МЕТА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета дисципліни – формування у здобувачів вищої освіти умінь та компетентностей щодо методів та інформаційних технологій обробки даних, розмір яких перевищує можливості звичайних програмних аналітичних платформ та баз даних по аналізу, зберіганню і управлінню великими даними (Big Data).

2 ОЧІКУВАНІ ДИСЦИПЛІНАРНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Дисциплінарні результати навчання (ДРН)	
шифр ДРН	зміст
ДРН – 01	Проводити аналіз об'єктів, які оперують з великими даними, та обґрунтовувати вибір алгоритмів обробки даних
ДРН – 02	Демонструвати знання інформаційних технологій обробки великих даних
ДРН – 03	Застосовувати методи регресійного аналізу для обробки великих даних
ДРН – 04	Застосовувати методи класифікації для обробки великих даних
ДРН – 05	Застосовувати методи штучного інтелекту для рішення задач кластеризації даних та прогнозування
ДРН – 06	Застосовувати методи факторного аналізу для обробки великих даних
ДРН – 07	Розробляти та використовувати спеціалізоване програмне забезпечення для обробки і візуалізації багатовимірних даних

3 БАЗОВІ ДИСЦИПЛІНИ

Дисципліна не потребує додаткових вимог до базових дисциплін. Міждисциплінарні зв'язки: курс ґрунтується на знаннях, отриманих з вивчених дисциплін за попереднім рівнем освіти.

4 ОБСЯГ І РОЗПОДІЛ ЗА ФОРМАМИ ОРГАНІЗАЦІЇ ОСВІТНЬОГО ПРОЦЕСУ ТА ВИДАМИ НАВЧАЛЬНИХ ЗАНЯТЬ

Вид навчальних занять	Обсяг, години	Розподіл за формами навчання, години			
		денна		заочна	
		аудиторні заняття	самостійна робота	аудиторні заняття	самостійна робота
лекційні	60	20	40	6	54
практичні	-	-	-	-	-
лабораторні	60	39	21	6	54
РАЗОМ	120	59	61	12	108

5 ПРОГРАМА ДИСЦИПЛІНИ ЗА ВИДАМИ НАВЧАЛЬНИХ ЗАНЯТЬ

Шифри ДРН	Види та тематика навчальних занять	Обсяг складових, години
	ЛЕКЦІЇ	60
ДРН-1, ДРН-2	Вступ Мета і завдання дисципліни “ Методи та техніки аналізу великих даних”. Сфери застосування Big Data	2
ДРН-1	Тема 1. Подання даних та їх попередня обробка Визначення. Шкали виміру ознак. Життєвий цикл даних. Попередня обробка даних. Алгоритм ZET заповнення пробілів у таблицях даних. Метадані. Життєвий цикл метаданих	4
ДРН-1, ДРН-3	Тема 2. Прогнозування стохастичних залежностей. Регресійний аналіз Завдання регресійного аналізу. Оцінка параметрів рівнянь парної регресії. Система нормальних рівнянь. Основні моделі парної регресії. Критерії оцінки якості моделей парної регресії. Багатовимірний лінійний регресійний аналіз. Відбір факторних змінних. Мультиколінеарність. Алгоритм Фаррара-Глобера. Оцінка параметрів рівняння множинної лінійної регресії. Стандартизоване рівняння множинної лінійної регресії. Покроковий відбір факторів множинної лінійної регресії Аналіз та прогнозування часових рядів. Автокорреляція часових рядів. Стаціонарність часових рядів. AR та MA процеси	10
ДРН-1, ДРН-4	Тема 3. Методи класифікації даних. Кластерний аналіз Огляд методів класифікації даних. Завдання кластерного аналізу. Міри відстані та збіжності. Основні метрики. Основні методи кластерного аналізу: ієрархічні агломеративні та дивізімні методи, метод К-середніх Мак-Куїна. Нечітка кластеризація, метод С-середніх	6
ДРН-1, ДРН-4	Тема 4. Методи класифікації даних. Деревя рішень Структура дерева рішень і цільова функція. Критерії обрання атрибута розбиття дерева. Критерії припинення розбиття. Відсікання гілок. Алгоритм ID3	6
ДРН-1, ДРН-5	Тема 5. Методи класифікації даних. Нейронні мережі Елементи нейронної мережі. Функції активації нейрона. Основні архітектури мереж. Помилка мережі. Навчання мереж. Правила Хебба. Правило Відроу-Хоффа. Реалізація бінарного класифікатора та предиктора лінійного часового тренду на основі нейронних мереж.	6
ДРН-1, ДРН-6	Тема 6. Факторний аналіз Завдання факторного аналізу. Факторне відображення. Визначення факторних навантажень. Метод головних компонент.	6
ДРН-2, ДРН-7	Тема 7. Програмні інструменти Big Data. Мова Python Статистичні функції мови Python. Аналіз даних засобами мови: регресійний, кластерний, факторний. Реалізація дерев рішень та нейронних мереж засобами мови.	10

Шифри ДРН	Види та тематика навчальних занять	Обсяг складових, години
ДРН-2, ДРН-7	Тема 8. Програмні інструменти Big Data. Платформа Hadoop Екосистема Hadoop. Складові платформи: Hadoop Common, YARN, MapReduce, файлова система HDFS. Пісочниця Apache Hadoop	10
	ЛАБОРАТОРНІ РОБОТИ	60
ДРН-2, ДРН-7	Лабораторна робота 1. Deductor Academic Ознайомлення з інтерфейсом користувача аналітичної платформи Deductor Academic. Експорт, імпорт і візуалізація даних	12
ДРН-2, ДРН-7	Лабораторна робота 2. Deductor Academic Сховище даних	12
ДРН-2, ДРН-7	Лабораторна робота 3. Deductor Academic Індивідуальна робота за вибором викладача: Очищення даних OLAP-технологія аналізу даних Автокореляційний аналіз часового ряду ABCD-аналіз даних XYZ-аналіз даних Прогнозування часового ряду за допомогою нейронної мережі Регресійне прогнозування часового ряду Класифікація даних за допомогою дерев рішень Кластеризація даних ієрархічними методами Кластеризація даних за допомогою самоорганізуючихся карт Кохонена Пошук асоціативних правил	12
ДРН-3, ДРН-7	Лабораторна робота 4. MathCAD Регресійний аналіз	12
ДРН-4, ДРН-7	Лабораторна робота 5. MathCAD Кластерний аналіз	12
	РАЗОМ	120

6 ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Сертифікація досягнень студентів здійснюється за допомогою прозорих процедур, що ґрунтуються на об'єктивних критеріях відповідно до Положення університету «Про оцінювання результатів навчання здобувачів вищої освіти».

Досягнутий рівень компетентностей відносно очікуваних, що ідентифікований під час контрольних заходів, відображає реальний результат навчання студента за дисципліною.

6.1 Шкали

Оцінювання навчальних досягнень студентів НТУ «ДП» здійснюється за рейтинговою (100-бальною) та інституційною шкалами. Остання необхідна (за офіційною відсутністю національної шкали) для конвертації (переведення) оцінок мобільних здобувачів.

Шкали оцінювання навчальних досягнень студентів НТУ «ДП»

Рейтингова	Інституційна
90...100	відмінно / Excellent
74...89	добре / Good
60...73	задовільно / Satisfactory
0...59	незадовільно / Fail

Кредити навчальної дисципліни зараховується, якщо студент отримав підсумкову оцінку не менше 60-ти балів. Нижча оцінка вважається академічною заборгованістю, що підлягає ліквідації відповідно до Положення про організацію освітнього процесу НТУ «ДП».

6.2 Засоби та процедури

Зміст засобів діагностики спрямовано на контроль рівня сформованості знань, умінь, комунікації, автономності та відповідальності студента за вимогами НРК до 7-го кваліфікаційного рівня під час демонстрації регламентованих робочою програмою результатів навчання.

Здобувач на контрольних заходах має виконувати завдання, орієнтовані виключно на демонстрацію дисциплінарних результатів навчання (розділ 2).

Засоби діагностики, що надаються здобувачам на контрольних заходах у вигляді завдань для поточного та підсумкового контролю, формуються шляхом конкретизації вихідних даних та способу демонстрації дисциплінарних результатів навчання.

Засоби діагностики (контрольні завдання) для поточного та підсумкового контролю дисципліни затверджуються кафедрою.

Види засобів діагностики та процедур оцінювання для поточного та підсумкового контролю дисципліни подано нижче.

Засоби діагностики та процедури оцінювання

ПОТОЧНИЙ КОНТРОЛЬ			ПІДСУМКОВИЙ КОНТРОЛЬ	
навчальне заняття	засоби діагностики	процедури	засоби діагностики	процедури
лекції	контрольні завдання за кожною темою	виконання завдання під час лекцій	комплексна контрольна робота (ККР)	визначення середньозваженого результату поточних контролів;
лабораторні	перевірка та захист	виконання лабораторних робіт		виконання ККР під час заліку за бажанням студента

Під час поточного контролю лекційні заняття оцінюються шляхом визначення якості виконання контрольних конкретизованих завдань. Лабораторні заняття оцінюються якістю виконання та захисту лабораторних робіт.

За наявності рівня результатів поточних контролів з усіх видів навчальних занять не менше 60 балів, підсумковий контроль здійснюється без участі студента шляхом визначення середньозваженого значення поточних оцінок.

Незалежно від результатів поточного контролю кожен студент під час заліку має право виконувати ККР, яка містить завдання, що охоплюють ключові дисциплінарні результати навчання.

Кількість конкретизованих завдань ККР повинна відповідати відведеному часу на виконання. Кількість варіантів ККР має забезпечити індивідуалізацію завдання.

Значення оцінки за виконання ККР визначається середньою оцінкою складових (конкретизованих завдань) і є остаточним.

Інтегральне значення оцінки виконання ККР може визначатися з урахуванням вагових коефіцієнтів, що встановлюється кафедрою для кожної складової опису кваліфікаційного рівня НРК.

6.3 Критерії

Реальні результати навчання студента ідентифікуються та вимірюються відносно очікуваних під час контрольних заходів за допомогою критеріїв, що описують дії студента для демонстрації досягнення результатів навчання.

Для оцінювання виконання контрольних завдань під час поточного контролю лекційних і лабораторних занять в якості критерію використовується коефіцієнт засвоєння, що автоматично адаптує показник оцінки до рейтингової шкали:

$$O_i = 100 a/m,$$

де a – число правильних відповідей або виконаних суттєвих операцій відповідно до еталону рішення; m – загальна кількість запитань або суттєвих операцій еталону.

Зміст критеріїв спирається на компетентнісні характеристики, визначені НРК для магістерського рівня вищої освіти (подано нижче).

**Загальні критерії досягнення результатів навчання
для 7-го кваліфікаційного рівня за НРК**

Опис кваліфікаційного рівня	Вимоги до знань, умінь/навичок, комунікації, відповідальності і автономії	Показник оцінки
Знання		
♦ спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері професійної діяльності або галузі знань і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у галузі та на межі галузей знань	Відповідь відмінна – правильна, обґрунтована, осмислена. Характеризує наявність: - спеціалізованих концептуальних знань на рівні новітніх досягнень; - критичне осмислення проблем у навчанні та/або професійній діяльності та на межі предметних галузей	95-100
	Відповідь містить не грубі помилки або описки	90-94
	Відповідь правильна, але має певні неточності	85-89
	Відповідь правильна, але має певні неточності й недостатньо обґрунтована	80-84
	Відповідь правильна, але має певні неточності, недостатньо обґрунтована та осмислена	74-79
	Відповідь фрагментарна	70-73
	Відповідь демонструє нечіткі уявлення студента про об'єкт вивчення	65-69
	Рівень знань мінімально задовільний	60-64
Рівень знань незадовільний	<60	
Уміння/навички		
♦ спеціалізовані уміння/навички розв'язання проблем, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур; ♦ здатність інтегрувати знання та розв'язувати складні задачі у широких або мультидисциплінарних контекстах; ♦ здатність розв'язувати проблеми у нових або незнайомих середовищах за наявності неповної або	Відповідь характеризує уміння: - виявляти проблеми; - формулювати гіпотези; - розв'язувати проблеми; - оновлювати знання; - інтегрувати знання; - провадити інноваційну діяльність; - провадити наукову діяльність	95-100
	Відповідь характеризує уміння/навички застосовувати знання в практичній діяльності з не грубими помилками	90-94
	Відповідь характеризує уміння/навички застосовувати знання в практичній діяльності, але має певні неточності при реалізації однієї вимоги	85-89
	Відповідь характеризує уміння/навички застосовувати знання в практичній діяльності, але має певні неточності при реалізації двох вимог	80-84
	Відповідь характеризує уміння/навички застосовувати знання в практичній діяльності, але має певні неточності при реалізації трьох вимог	74-79

Опис кваліфікаційного рівня	Вимоги до знань, умінь/навичок, комунікації, відповідальності і автономії	Показник оцінки
обмеженої інформації з урахуванням аспектів соціальної та етичної відповідальності	Відповідь характеризує уміння/навички застосовувати знання в практичній діяльності, але має певні неточності при реалізації чотирьох вимог	70-73
	Відповідь характеризує уміння/навички застосовувати знання в практичній діяльності при виконанні завдань за зразком	65-69
	Відповідь характеризує уміння/навички застосовувати знання при виконанні завдань за зразком, але з неточностями	60-64
	Рівень умінь/навичок незадовільний	<60
Комунікація		
♦ зрозуміле і недвозначне донесення власних знань, висновків та аргументації до фахівців і нефахівців, зокрема до осіб, які навчаються	Зрозумілість відповіді (доповіді). <i>Мова:</i> правильна; чиста; ясна; точна; логічна; виразна; лаконічна. <i>Комунікаційна стратегія:</i> – послідовний і несуперечливий розвиток думки; – наявність логічних власних суджень; – доречна аргументації та її відповідність відстоюваним положенням; – правильна структура відповіді (доповіді); – правильність відповідей на запитання; – доречна техніка відповідей на запитання; – здатність робити висновки та формулювати пропозиції; використання іноземних мов у професійній діяльності	95-100
	Достатня зрозумілість відповіді (доповіді) та доречна комунікаційна стратегія з незначними хибами	90-94
	Добра зрозумілість відповіді (доповіді) та доречна комунікаційна стратегія (сумарно не реалізовано три вимоги)	85-89
	Добра зрозумілість відповіді (доповіді) та доречна комунікаційна стратегія (сумарно не реалізовано чотири вимоги)	80-84
	Добра зрозумілість відповіді (доповіді) та доречна комунікаційна стратегія (сумарно не реалізовано п'ять вимог)	74-79
	Задовільна зрозумілість відповіді (доповіді) та доречна комунікаційна стратегія (сумарно не реалізовано сім вимог)	70-73
	Задовільна зрозумілість відповіді (доповіді) та комунікаційна стратегія з хибами (сумарно не реалізовано дев'ять вимог)	65-69
	Задовільна зрозумілість відповіді (доповіді) та комунікаційна стратегія з хибами (сумарно не реалізовано 10 вимог)	60-64
	Рівень комунікації незадовільний	<60
Відповідальність і автономія		
	Відмінне володіння компетенціями:	95-100

Опис кваліфікаційного рівня	Вимоги до знань, умінь/навичок, комунікації, відповідальності і автономії	Показник оцінки
<ul style="list-style-type: none"> ♦ управління робочими або навчальними процесами, які є складними, непередбачуваними та потребують нових стратегічних підходів; ♦ відповідальність за внесок до професійних знань і практики та/або оцінювання результатів діяльності команд та колективів; ♦ здатність продовжувати навчання з високим ступенем автономії 	<ul style="list-style-type: none"> - використання принципів та методів організації діяльності команди; - ефективний розподіл повноважень в структурі команди; - підтримка врівноважених стосунків з членами команди (відповідальність за взаємовідносини); - стресовитривалість; - саморегуляція; - трудова активність в екстремальних ситуаціях; - високий рівень особистого ставлення до справи; - володіння всіма видами навчальної діяльності; - належний рівень фундаментальних знань; - належний рівень сформованості загальнонавчальних умінь і навичок 	
	Упевнене володіння компетенціями автономії та відповідальності з незначними хибами	90-94
	Добре володіння компетенціями автономії та відповідальності (не реалізовано дві вимоги)	85-89
	Добре володіння компетенціями автономії та відповідальності (не реалізовано три вимоги)	80-84
	Добре володіння компетенціями автономії та відповідальності (не реалізовано чотири вимоги)	74-79
	Задовільне володіння компетенціями автономії та відповідальності (не реалізовано п'ять вимог)	70-73
	Задовільне володіння компетенціями автономії та відповідальності (не реалізовано шість вимог)	65-69
	Задовільне володіння компетенціями автономії та відповідальності (рівень фрагментарний)	60-64
	Рівень автономії та відповідальності незадовільний	<60

7 ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

Використовуються лабораторна й інструментальна база випускової кафедри інформаційних технологій та комп'ютерної інженерії, комп'ютерне та мультимедійне обладнання, дистанційна платформа Moodle. При проведенні лабораторних занять використовується вільно розповсюджувана аналітична платформа Deductor Academic.

8 РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Zgurovsky M.Z. Big Data: Conceptual Analysis and Applications. [Текст] / M.Z. Zgurovsky, Y.P. Zaychenko // Springer, 2021, 298 p.
2. Stanislaw Osowski. Sieci Neuronowe do Przetwarzania Informacji [Текст] / Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2000, 344 с.
3. Akerkar R. Models of Computation for Big Data [Текст] / R. Akerkar // Springer, 2018, 110 p.
4. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics [Текст] / P. Ghavami // CreateSpace Independent Publishing Platform, 2016, 204 p.
5. Feeney K. Engineering Agile Big-Data Systems [Текст] / K. Feeney, J. Davies, J. Welch, S. Hellmann, C. Dirschl, A. Koller, P. Francois, A. Marciniak // River Publishers, 2018, 436 p.
6. Мороз Б.І. Лабораторний практикум з курсу: “Аналіз даних та процесів”. [Електрон. ресурс]. Режим доступа: <https://do.nmu.org.ua/course/view.php?id=3214>.
7. Big Data Fundamentals courses [Електрон. ресурс]. Режим доступа: <https://cognitiveclass.ai/learn/big-data>.
8. Big Data Analytics [Електрон. ресурс]. Режим доступа: <https://cognitiveclass.ai/learn/analytics/>.
9. Форман Дж. Много цифр: Анализ больших данных при помощи Excel [Текст] / Дж. Форман // М.: Альпина Паблишер, 2016, 464 с.
10. Kozhevnikov A. V. Estimation of the population density spatial distribution using clutter model [Текст] / A. V. Kozhevnikov, Ye. A. Krivosheyev // Науковий вісник НГУ – Дніпропетровськ: НГУ, 2011, №2, с. 31 – 36.
11. UCI Machine Learning Repository [Електрон. ресурс]. Режим доступа: <http://archive.ics.uci.edu/ml/index.php/>.

Навчальне видання

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
“Методи та техніки аналізу великих даних”
для бакалаврів спеціальності 123 “Комп’ютерна інженерія ”

Розробник: доц. Кожевников А.В.

В редакції автора

Підготовлено до друку
у Національному технічному університеті
«Дніпровська політехніка».
Свідоцтво про внесення до Державного реєстру ДК № 1842 від 11.06.2004
49005, м. Дніпро, просп. Д. Яворницького, 19.