



Кам'янець-Подільський національний
університет імені Івана Огієнка
Фізико-математичний факультет
Кафедра комп'ютерних наук

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ АНАЛІЗ ВЕЛИКИХ НАБОРІВ ДАНИХ

1. Загальна інформація про курс

Назва курсу, мова викладання	АНАЛІЗ ВЕЛИКИХ НАБОРІВ ДАНИХ, мова викладання – українська
Викладач	Іванюк Віталій Анатолійович, доктор технічних наук, доцент, завідувач кафедри комп'ютерних наук
Профайл викладача	https://cs.kpnu.edu.ua/2019/11/04/ivaniuk-vitalij-anatolijovych/
E-mail:	wivanyuk@kpnu.edu.ua
Викладач	Кушнір Оксана Климівна, кандидат економічних наук, доцент, доцент кафедри економіки підприємства
Профайл викладача	https://kaf-ep.kpnu.edu.ua/kushnir-oksanaklymivna/
E-mail:	oksana.kushnir@kpnu.edu.ua
Сторінка курсу в MOODLE	https://moodle.kpnu.edu.ua/course/view.php?id=25456
Консультації	Розклад проведення консультацій: щочетверга з 16-00 до 18-00 в ауд. №44 корпусу №4; формат консультацій – групові та індивідуальні у вигляді співбесіди

2. Анотація до курсу

Навчальна дисципліна "Аналіз великих наборів даних" присвячена вивченню сучасних технологій збору, зберігання, обробки та аналізу великих обсягів даних. Програма курсу охоплює широкий спектр тем, від основних понять Big Data до застосування передових методів машинного навчання та хмарних технологій.

3. Мета і завдання курсу

Метою навчальної дисципліни "Аналіз великих наборів даних" є ознайомити студентів з фундаментальними концепціями Big Data, навчити використовувати інструменти та платформи для роботи з великими даними, розвинути навички аналізу даних, включаючи статистичний аналіз, машинне навчання та візуалізацію, підготувати фахівців, здатних вирішувати складні завдання з аналізу даних в різних сферах діяльності.

Завдання – набуття компетентностей, визначених освітньою програмою:

Здатність розв'язувати задачі дослідницького та/або інноваційного характеру у сфері комп'ютерних наук.

ЗК 01 Здатність до абстрактного мислення, аналізу та синтезу.

ЗК 02 Здатність застосовувати знання у практичних ситуаціях.

ЗК 05 Здатність вчитися й оволодівати сучасними знаннями.

ЗК 06 Здатність бути критичним і самокритичним.

ЗК 07 Здатність генерувати нові ідеї (креативність).

СК01 Усвідомлення теоретичних засад комп'ютерних наук.

СК 03 Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК 04 Здатність збирати і аналізувати дані (включно з великими) для забезпечення якості прийняття проєктних рішень.

СК 05 Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

СК 06 Здатність застосовувати чинні і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

СК 07 Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

СК 09 Здатність розробляти та адмініструвати бази даних та знань.

4. Результати навчання

Програмні результати навчання, визначені освітньою програмою:

ПРН 07 Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

ПРН 08 Розробляти математичні моделі та методи аналізу даних (включно з великим).

ПРН 09 Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

ПРН 11 Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

ПРН 12 Проєктувати та супроводжувати бази даних та знань.

ПРН 18 Збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до інформаційної або комп'ютерної системи, що розробляється, експлуатується чи супроводжується.

ПРН 20 Досліджувати та використовувати методи та засоби комп'ютерного моделювання для аналізу складних динамічних систем, в тому числі вбудованих.

5. Формат курсу

Стандартний курс (очний).

6. Обсяг і ознаки курсу

Найменування показників	Характеристика навчального курсу
Освітня програма, спеціальність	Освітньо-професійна програма: <i>Комп'ютерні науки та інформаційні технології</i> спеціальність: 122 <i>Комп'ютерні науки</i>
Рік навчання	1
Семестр навчання	1
Кількість кредитів ЄКТС	5
Загальний обсяг годин	150
Кількість годин навчальних занять	50
Лекційні заняття	24

Практичні заняття	0
Лабораторні заняття	26
Самостійна робота	100
Форма підсумкового контролю	Екзамен

7. Пререквізити і кореквізити курсу

Освітній компонент вивчається в першому семестрі.

8. Технічне й програмне забезпечення, обладнання

Для проведення лекцій необхідно мультимедійне забезпечення. При проведенні лабораторних робіт лабораторії із необхідним програмним забезпеченням.

Програмне забезпечення: платформа для хмарних обчислень Google Cloud Platform, мова програмування Python, бібліотеки NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow, інструмент для візуалізації Power BI, IDE PyChar, VisualCode.

Передбачається застосування об'єктно-модульного динамічного середовища навчання MOODLE.

9. Політика курсу (правила та вимоги)

Увесь навчальний контент розміщено в модульному середовищі навчання К-ПНУ імені Івана Огієнка – moodle. Підготовка та виконання завдань лабораторних робіт є обов'язковим для кожного студента. Проект виконується в групах обсягом від 3 до 5 осіб і є обов'язком для кожного студента.

Методи навчання: лекція, розповідь, пояснення, бесіда, інструктаж, демонстрація, ілюстрація, спостереження, лабораторний метод, практична робота, аналітичний, індуктивний, дедуктивний, проблемний, дослідницький.

Академічна доброчесність. Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел списування, втручання в роботу інших студентів становлять, але не обмежують приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі студента є підставою для її не зарахування викладачем, незалежно від масштабів плагиату чи обману. URL: <https://integrity.kpnu.edu.ua/>.

Відвідання занять. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Для того, щоб опрацювати питання пропущеної лекції, студент повинен підготувати питання пропущеної лекції та захистити їх на одному з лабораторних занять або під час консультації (усно). Виконання усіх лабораторних робіт є обов'язковим. Студенти зобов'язані дотримуватися термінів виконання усіх видів робіт, передбачених курсом.

Комунікування. Спілкування з викладачами здійснюється через електронну пошту та месенджери (Telegram), під час лекційних та лабораторних занять (участь у бесідах, дискусіях тощо).

Форми контролю.

Поточний контроль – перевірка виконання завдань лабораторних робіт, захист проектів. Підсумковий контроль – екзамен.

Неформальна та/або інформальна освіта.

У випадку, якщо здобувач освіти отримав знання у неформальній та інформальній освіті, зарахування результатів навчання здійснюється згідно Порядку визнання результатів навчання здобувачів вищої освіти, отриманих шляхом здобуття неформальної/інформальної освіти в Кам'янець-Подільському національному університеті імені Івана Огієнка, зокрема, якщо їх тематика відповідає змісту навчальної дисципліни (окремій темі або змістовому модулю).

Для неформальної освіти рекомендується Google Cloud Data Analytics Certificate. URL: <https://www.cloudskillsboost.google/paths/420>. Для зарахування отриманих знань необхідна наявність значків модулів або сертифікату курсу.

10. Схема курсу

№	Назви змістових модулів і тем	Кількість годин				
		усього	у тому числі			
			л	п	лаб	с.р.
	Загалом	150	24	0	26	100
	Модуль 1. Вступ та основи. Інструменти та платформи	24	6	0	4	14
1	Вступ. Поняття великих даних. Добування даних	6	2	0		4
2	Архітектурні моделі інженерії Big Data. MapReduce і новий програмний стек	6	2	0		4
3	Інструменти аналізу даних	12	2	0	4	6
	Модуль 2. Методи аналізу	26	6	0	8	12
4	Задачі і алгоритми (Пошук подібних об'єктів. Аналіз потоків даних. Аналіз посилань. Часткові предметні набори. Кластеризація. Реклама в Інтернеті)	10	2	0	4	4
5	Моделі машинного навчання.	10	2	0	4	4
6	Моніторинг та аналіз теплового режиму ІТ-обладнання (Аналіз даних теплового режиму комутаційного обладнання комп'ютерних мереж на основі відновлення сигналів температурних датчиків)	6	2			4
	Модуль 3. Бізнес-аналітика на основі Big Data	24	6	0	4	14
7	Діджиталізація економіки як елемент сталого розвитку	6	2	0		4
8	Аналіз великих даних та бізнес аналітика - інструменту для прийняття управлінських рішень	6	2	0		4
9	Візуалізація даних. Інструменти, створення ефективних візуалізацій.	12	2		4	6
	Модуль 4. Хмарні технології та Big Data	24	6	0	6	12
10	Data Analytics in Google Cloud. Управління та зберігання даних у хмарі.	10	2	0	4	4
11	Трансформація даних у хмарі. Візуалізація даних в хмарі	8	2	0	2	4
12	Підготовка до роботи аналітика хмарних даних	6	2			4
	Проект	52			4	48

11. Система оцінювання та вимоги

Розподіл балів між поточним і підсумковим контролем подано у таблиці.

Назви змістових модулів і тем	Кількість балів
Поточний контроль	30
Проект	30
Екзамен	40
Загалом	100

Поточний контроль (30 балів)

Максимальний бал оцінки поточної успішності здобувачів вищої освіти на кожному лабораторному занятті – 12.

Рівень досягнення	Оцінка	Критерії оцінювання
Початковий	1	Не розуміє основних понять, не може виконати прості завдання, не дотримується академічної доброчесності.
	2	Обмежене розуміння термінології, не може самостійно розв'язати навіть прості задачі.
	3	Може визначити деякі ключові поняття, але не може їх застосувати на практиці.
Середній	4	Може виконати прості завдання за зразком, але має труднощі з більш складними завданнями.
	5	Розуміє основні алгоритми, але не може їх пояснити або застосувати в нестандартних ситуаціях.
	6	Може виконати більшість завдань, але допускає помилки при аналізі даних та інтерпретації результатів.
Достатній	7	Демонструє розуміння основних концепцій, вміє використовувати різноманітні інструменти для аналізу даних, може пояснити отримані результати.
	8	Самостійно виконує складні завдання, вміє вибирати відповідні методи аналізу для різних типів даних.
	9	Виявляє творчий підхід до вирішення задач, може запропонувати альтернативні рішення.
Високий	10	Глибоко розуміє теоретичні основи, вміє розробляти власні алгоритми, проводить поглиблений аналіз даних.
	11	Демонструє високий рівень аналітичних навичок, може проводити наукові дослідження в галузі аналізу даних.
	12	Має вичерпні знання з дисципліни, може самостійно вирішувати складні наукові проблеми, є лідером у команді.

Здобувачу вищої освіти, який не виконав поставлених завдань в журнал обліку роботи академічної групи ставиться 0 балів.

Результати перевірки самостійної роботи також входять до поточного контролю.

Пропущені заняття здобувач вищої освіти має обов'язково відпрацювати. За відпрацьовані лекційні заняття (усна відповідь) оцінки не ставляться, за лабораторні заняття (звіт лабораторної роботи) нараховуються бали середнього (4, 5, 6), достатнього (7, 8, 9) та високого рівня (10, 11, 12).

Якщо здобувач вищої освіти не відпрацював пропущені навчальні заняття, не виправив оцінки 0,1,2,3, отримані на навчальних заняттях, він вважається таким, що має академічну заборгованість за результатами поточного контролю. Рейтингова оцінка у балах знань, умінь і навичок здобувача на навчальних заняттях з навчального (змістового) модуля обчислюється після проведення цих занять та ліквідації ним поточної заборгованості, пов'язаної з пропусками занять, невідповідністю або недостатньою підготовленістю до них відповідно до «Положення про рейтингову систему оцінювання навчальних досягнень здобувачів вищої освіти Кам'янець-Подільського національного університету імені Івана Огієнка» (нова редакція)» (https://drive.google.com/file/d/1aD_jeL-jGRbDWAgkQ58tdMxxbqQKufF/view).

Проект(30 балів)

Оцінювання проекту здійснюється під час його захисту.

Критерії оцінювання проекту:

Критерій	Максимальна оцінка за проєкт	Пояснення
Командна робота (40%)	12 балів	Ефективність командної роботи. Оцінюється індивідуально кожен член команди
Взаємодія в команді	4 бали	Здатність до співпраці, комунікація, вирішення конфліктів
Розподіл ролей	4 бали	Ефективність розподілу завдань, відповідальність за виконання
Досягнення результатів	4 бали	Внесок у загальний результат проєкту
Проект (60%)	18 балів	Якість виконання завдань. Оцінюється команда загалом
Підготовка даних	3 балів	Якість очищення даних, вибір ознак, візуалізація, підготовка до моделювання
Постановка задачі та вибір моделей	3 бали	Чіткість формулювання, обґрунтування вибору моделей, налаштування гіперпараметрів, порівняння моделей
Технічні аспекти	3 бали	Якість коду, використання бібліотек, документація коду
Оцінка моделей	3 бали	Вибір метрик, крос-валідація, аналіз результатів
Інтерпретація результатів та висновки	3 бали	Зрозумілість висновків, практична значимість, обмеження моделі, пропозиції щодо покращення
Презентація	3 бали	Структура презентації, якість візуалізацій, відповіді на запитання
Загальна оцінка за проєкт	30 балів	

Оцінка за кожним критерієм виставляється в балах від 0 до максимального значення для цього критерію. Загальна оцінка проєкту розраховується як сума балів за всіма критеріями.

Здобувачі вищої освіти, які за виконання завдань отримали рейтинговий бал менший 60% від максимальної кількості балів, виділених на цей вид роботи, а також ті, що не з'явилися на захист проєкту, вважаються такими, що мають академічну заборгованість за результатами поточного контролю, ліквідація якої є обов'язковою.

Екзамен (40 балів)

До екзамену допускаються здобувачі ВО, що успішно виконали усі завдання поточного контролю та успішно захистили проєкт (сумарна оцінка становить не менше 36 балів).

Екзаменаційні білети складаються із двох запитань теоретичного характеру. Відповіді на запитання оцінюються за 12-бальною шкалою за критеріями аналогічними до поточного контролю. Рейтингова оцінка у балах знань, умінь і навичок здобувача обчислюється відповідно до «Положення про рейтингову систему оцінювання навчальних досягнень здобувачів вищої освіти Кам'янець-Подільського національного університету імені Івана Огієнка (нова редакція)» (https://drive.google.com/file/d/1aD_jEL-jGRbDWAgkQ58tdMxxbqQKufF/view).

Здобувачам вищої освіти, які за іспит отримали незадовільну оцінку, дозволяється ліквідувати академічну заборгованість після належної підготовки. Ліквідація академічної

заборгованості за результатами семестрового контролю дозволяється до початку наступного семестру в час, визначений графіком ліквідації академічної заборгованості.

Оцінювання навчальних досягнень здобувачів вищої освіти з навчальної дисципліни здійснюється відповідно до таблиці 1 Положення про організацію освітнього процесу в Кам'янець-Подільському національному університеті імені Івана Огієнка (нова редакція) (<https://drive.google.com/file/d/1ZbMN35h-7ZSJBBOVvL2bTCaLtRbcQA86/view>)

Таблиця 1

Таблиця відповідності шкал оцінювання навчальних досягнень здобувачів вищої освіти

Рейтингова оцінка з кредитного модуля	Оцінка за шкалою ECTS	Оцінка за національною шкалою
90-100 і більше	A (відмінно)	відмінно
82-89	B (дуже добре)	добре
75-81	C (добре)	
67-74	D (задовільно)	задовільно
60-66	E (достатньо)	
35-59	FX (незадовільно з можливістю повторного складання)	незадовільно
34 і менше	F (незадовільно з обов'язковим проведенням додаткової роботи щодо вивчення навчального матеріалу кредитного модуля)	

12. Рекомендована література

Основна

1. Banachewicz K., Massaron L. The Kaggle book: Data analysis and machine learning for competitive data science. Birmingham, UK: Packt Publishing Ltd., 2022. 531 p.
2. Google Cloud Data Analytics Certificate. URL: <https://www.cloudskillsboost.google/paths/420>
3. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of massive data sets. Cambridge university press.
4. Ланде Д.В., Субач І.Ю., Гладун А.Я. Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. Київ 2021. – 168 с.
5. Федорчук В. А., Іванюк В. А., Тихоход В. А. Аналіз даних теплового режиму комутаційного обладнання комп'ютерних мереж на основі відновлен-ня сигналів температурних датчиків. Математичне та комп'ютерне моделю-вання. Серія: Технічні науки: зб. наук. праць / Інститут кібернетики імені В. М. Глушкова Національної академії наук України, Кам'янець-Подільський національний університет імені Івана Огієнка; [редкол.: О.М.Хіміч(відп. ред.) та ін.]. Кам'янець-Подільський: Кам'янець-Подільський національний університет імені Івана Огієнка, 2024. Вип. 25. С.114-128.
6. Чаплінський В.Р., Кушнір О.К. Свідер О.П. Аналіз великих даних та їх візуалізація для потреб бізнесу. Ефективна економіка. № 6 (2021). URL : <http://www.economy.nayka.com.ua/?op=1&z=8979>.

Додаткова

7. Bierer, Doug. Learn MongoDB 4. x: A guide to understanding MongoDB development and administration for NoSQL developers. Packt Publishing Ltd, 2020.
8. Davis, Adam L., and Adam L. Davis. "Spring data." Spring Quick Reference Guide: A Pocket Handbook for Spring Framework, Spring Boot, and More (2020): 43-59.

9. Digitalization of economics: inter-disciplinary and inter-branch approach : manual. Zhytomyr : Zhytomyr Polytechnic State University, 2023. 540 p. URL : https://dig2eco.ztu.edu.ua/wp-content/uploads/2023/06/maket_pidruchnik_digeco_31.05.23-u-vidavnicztvo.pdf.
10. Інтелектуальний аналіз даних: Практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 124 «Системний аналіз», освітніх програм «Системний аналіз і управління», «Системний аналіз фінансового ринку» / Н. І. Недашківська; КПІ ім. Ігоря Сікорського. Київ : КПІ ім. Ігоря Сікорського, 2021.– 105 с.
11. Jeyaraj, Rathinaraja, Ganeshkumar Pugalendhi, and Anand Paul. Big Data with Hadoop MapReduce: A Classroom Approach. Apple Academic Press, 2020.
12. Kumar S. Introduction to Parallel Programming / Subodh Kumar // Cambridge University Press, 2022. – 277 p.
13. Minelli M. Big Data. Big Analytics / M.Minelli, M.Chambers, A. Dhiraj //John Wiley & Sons, Inc., 2020. – 205 p.
14. Pollack, Mark, Oliver Gierke, Thomas Risberg, Jon Brisbin, and Michael Hunger. Spring Data: modern data access for enterprise Java. "O'Reilly Media, Inc.", 2021.
15. Raschka S. et al. Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python. – Packt Publishing Ltd, 2022.
16. Raschka S., Mirjalili V. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. – Packt Publishing Ltd, 2019.
17. Simon J. D. Prince. Understanding deep learning // The MIT Press, Cambridge, Massachusetts, 2023. 527 p.
18. The Kaggle Book. Data analysis and machine learning for competitive data science. Code Repository for The Kaggle Book, Published by Packt Publishingю URL: <https://github.com/PacktPublishing/The-Kaggle-Book>
19. Аналітика великих даних: принципи, напрямки і задачі (огляд) / О.С. Балабанов // Проблеми програмування. — 2019. — № 2. — С. 47-68.
20. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.
21. Економічна аналітика в бізнесі : навч. посібник / [О.С. Гринькевич, С.О. Матковський, А.В. Сидорова та ін.] ; за ред. О.С. Гринькевич, С.О. Матковського, А.В. Сидорової, Н.С. Струк. Львів : ЛНУ ім. Івана Франка, 2022. 480 с. URL : https://econom.lnu.edu.ua/wp-content/uploads/2016/04/Navchalnyy-posibnyk_2022.pdf.
22. Іванчук Я.В., Месюра В.І., Яровий А.А., Манжілевський О.Д. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних. Вінниця: ВНТУ, 2021. 69 с.
23. Інтелектуальний аналіз даних : навчальний посібник / А.О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 271 с.
24. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с.
25. Кушнір О.К., Макогончук Н.В., Янчик І.М. Utilization of big data analytics for logistic processes enhancing. *Інвестиції: практика та досвід*. № 10 (2024). С. 129-134. URL : <https://nayka.com.ua/index.php/investplan/article/view/3716/3751>.
26. Кушнір О.К., Чаплінський В.Р. Статистичні методи аналізу великих даних. *Modern Economics*. № 39 (2023). С. 75-81. URL : <https://modecon.mnau.edu.ua/statistical-methods-for-big-data/>.
27. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань: навчальний посібник. Львів: Магнолія, 2021. 276 с.

28. Олещенко Л.М. Технології оброблення великих даних. Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с.
29. Основи цифрової економіки : навчальний посібник / за ред. Н.А. Мазур. Кам'янець-Подільськ. ун-т ім. І. Огієнка. Чернівці : Чернівець. нац. ун-т ім. Ю. Федьковича, 2024. 304 с.
30. Сидорова А.В., Біленко Д.В., Буркіна Н.В. Бізнес-аналітика : навчально-методичний посібник. Вінниця: ДонНУ імені Василя Стуса. 2019. 104 с. URL : https://r.donnu.edu.ua/bitstream/123456789/105/1/79_%D0%9D%D0%9C%D0%9F_%D0%91%D1%96%D0%B7%D0%BD%D0%B5%D1%81-%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D1%82%D0%B8%D0%BA%D0%B0.pdf.
31. Сидорова А.В., Біленко Д.В., Буркіна Н.В. Бізнес-аналітика: навчально-методичний посібник. Вінниця: ДонНУ імені Василя Стуса. 2019. 104 с.
32. Тарн В.І., Гордієнко Ю. Г., Стіренко С.Г. Технології BIG DATA Практикум // Київ: КПІ ім. Ігоря Сікорського, 2022. – 56 с.

Інформаційні джерела

33. Google Developers Codelabs. URL: <https://codelabs.developers.google.com/>
34. Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. URL: <http://www.mmids.org/>
35. Kaggle. Learn. URL: <https://www.kaggle.com/learn>
36. TensorFlow. An end-to-end platform for machine learning. URL: <https://www.tensorflow.org/>