



СУЧАСНІ МЕТОДИ І ТЕХНОЛОГІЇ ОБРОБЛЕННЯ НАДВЕЛИКИХ МАСИВІВ ДАНИХ

Робоча програма навчальної дисципліни (силабус)

Рівень вищої освіти	<i>Третій (освітньо-науковий)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки</i>
Освітньо-наукова програма	<i>Комп'ютерні системи і технології спеціального зв'язку</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна (денна, вечірня)</i>
Рік підготовки, семестр	<i>II рік підготовки, весняний семестр</i>
Обсяг дисципліни	<i>3 кредити ECTS</i>
Семестровий контроль / контрольні заходи	<i>Залік / модульна контрольна робота / реферат</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лекції та семінарські заняття: кандидат технічних наук, доцент Василь ЦУРКАН</i>
Розміщення курсу	<i>Google Classroom</i>

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Силабус освітнього компонента “Сучасні методи і технології оброблення надвеликих масивів даних” складено відповідно до освітньо-наукової програми “Комп’ютерні системи і технології спеціального зв’язку” для підготовки здобувачів вищої освіти третього (освітньо-наукового) рівня вищої освіти за спеціальністю 122 Комп’ютерні науки.

Метою навчальної дисципліни є підсилення у аспірантів наступних компетентностей:

- ZK02 Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ZK04 Здатність розв’язувати комплексні проблеми комп’ютерних наук на основі системного наукового світогляду та загального культурного кругозору із дотриманням принципів професійної етики та академічної доброчесності.
- СК02 Здатність застосовувати сучасні методології, методи та інструменти експериментальних і теоретичних досліджень у сфері комп’ютерних наук, сучасні цифрові технології, бази даних та інші електронні ресурси у науковій та освітній діяльності.
- СК03 Здатність виявляти, ставити та вирішувати дослідницькі науково-прикладні задачі та/або проблеми в сфері комп’ютерних наук, оцінювати та забезпечувати якість виконуваних досліджень.
- СК06 Здатність аналізувати та оцінювати сучасний стан і тенденції розвитку комп’ютерних наук та інформаційних технологій.
- СК07 Здатність створювати нові ідеї та розв’язувати складні науково-прикладні проблеми в галузі комп’ютерних наук, пов’язані з розробкою та дослідженням комп’ютерних систем та технологій спеціального зв’язку для потреб Держспецзв’язку.
- СК08 Здатність збирати дані з розподілених джерел, які використовуються для аналізу результатів розвідки, оцінки та/або планування.

Предмет навчальної дисципліни – сучасні методи, технології оброблення надвеликих масивів даних.

Програмні результати навчання, на покращення яких спрямована дисципліна:

- РН03 Формулювати і перевіряти гіпотези; використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних досліджень і математичного та/або комп’ютерного моделювання, наявні літературні дані.
- РН04 Розробляти та досліджувати концептуальні, математичні і комп’ютерні моделі процесів і систем, ефективно використовувати їх для отримання нових знань та/або створення інноваційних продуктів у комп’ютерній науці та дотичних міждисциплінарних напрямках.
- РН05 Планувати і виконувати експериментальні та/або теоретичні дослідження з комп’ютерних наук та дотичних міждисциплінарних напрямів з використанням сучасних інструментів, критично аналізувати результати власних досліджень і результати інших дослідників у контексті усього комплексу сучасних знань щодо досліджуваної проблеми.
- РН06 Застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи.
- РН07 Розробляти та реалізовувати наукові та/або інноваційні інженерні проєкти, які дають можливість переосмислити наявне та створити нове цілісне знання та/або

- професійну практику і розв'язувати значущі наукові та технологічні проблеми комп'ютерної науки з дотриманням норм академічної етики і врахуванням соціальних, економічних, екологічних та правових аспектів.
- PH08 Визначати актуальні наукові та практичні проблеми у сфері комп'ютерних наук, глибоко розуміти загальні принципи та методи комп'ютерних наук, а також методологію наукових досліджень, застосувати їх у власних дослідженнях у сфері комп'ютерних наук та у викладацькій практиці.
- PH12 Створювати нові та ефективні ідеї та методи розв'язання складних науково-прикладних проблем в галузі комп'ютерних наук, пов'язаних з розробкою та дослідженням комп'ютерних систем та технологій спеціального зв'язку для Держспецзв'язку.
- PH13 Розробляти дієві рекомендації ключовим зацікавленим сторонам на основі збору, обробки даних та аналізу даних.

2. Пререквізити та постреквізити навчальної дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Успішне вирішення завдань навчальної дисципліни базується на засвоєні аспірантами знань та умінь, сформованих у них, в результаті навчання на попередніх рівнях вищої освіти.

Компетенції, знання та уміння, одержані в процесі вивчення освітнього компонента можуть використовуватися для виконання завдань дисертаційних досліджень у межах наукової складової.

3. Зміст навчальної дисципліни

Семестр 4.

Розділ 1. Теоретичні основи оброблення надвеликих масивів даних.

Тема 1. Методи оброблення надвеликих масивів даних.

Основні виклики надвеликих масивів даних. Життєвий цикл оброблення надвеликих масивів даних. Аналітика надвеликих масивів даних. Завдання оброблення надвеликих масивів даних. Підходи до оброблення надвеликих масивів даних. Методи попереднього оброблення надвеликих масивів даних. Методи обирання ознак. Методи класифікування надвеликих масивів даних. Методи машинного навчання. Методи забезпечення кібербезпеки надвеликих масивів даних.

Розділ 2. Практичні основи оброблення надвеликих масивів даних.

Тема 2. Технології оброблення надвеликих масивів даних.

Основні технології оброблення надвеликих масивів даних. Технологія збереження і оброблення надвеликих масивів даних. Екосистема Hadoop. Технологія паралельного оброблення надвеликих масивів даних. Фреймворк MapReduce. Технологія зберігання даних у нереляційних форматах (структурованих, напівструктурованих, неструктурованих, поліморфних). NoSQL. Технологія зберігання надвеликих обсягів неструктурованих даних. Озера даних. Технологія машинного навчання. Технологія глибокого навчання. Технологія блокчейн. Інженерія кібербезпеки надвеликих масивів даних.

4. Навчальні матеріали та ресурси

Основна література:

1. Sasikala S., Renuka Devi D. Research practitioner's handbook on Big Data analytics. New York : Apple Academic Press, 2023. 310 p.
2. Shi Y. Advances in Big Data Analytics. Cham : Springer, 2022. 728. DOI: <https://doi.org/10.1007/978-981-16-3607-3>.
3. Ланде Д.В., Субач І.Ю., Гладун А.Я. Оброблення надвеликих масивів даних (BIG DATA) : навчальний посібник. Київ : КПІ ім. Ігоря Сікорського, 2021. 168 с. URL: <https://ela.kpi.ua/handle/123456789/46129>.

4. Олещенко Л.М. Технології оброблення великих даних: конспект лекцій. Київ : КПІ ім. Ігоря Сікорського, 2021. 227 с. URL: <https://ela.kpi.ua/handle/123456789/42206>.

5. Mongeau S., Hajdasinski A. Cybersecurity Data Science: Best Practices in an Emerging Profession. Cham : Springer, 2021. 415 p. DOI: <https://doi.org/10.1007/978-3-030-74896-8>.

Додаткова література:

1. Kaliraj P., Devi T. Big Data Applications in Industry 4.0. London : CRC Press, 2022. 422 p.

2. Erl T., Khattak W., Buhler P. Big Data Fundamentals : Concepts, Drivers & Techniques. Boston : Pearson, 2016. 240 p.

3. Satchidananda Dehuri S., Chen Y.-W. Advances in Machine Learning for Big Data Analysis. Cham : Springer, 2022. 239. DOI: <https://doi.org/10.1007/978-981-16-8930-7>.

4. Schintler L.A., McNeely C.L. Encyclopedia of Big Data. Cham : Springer, 2022. 976 p. DOI: <https://doi.org/10.1007/978-3-319-32010-6>.

5. Savas O., Deng J. Big Data Analytics in Cybersecurity. London : CRC Press, 2017. 336 p.

Навчальний контент

5. Методика опанування навчальної дисципліни

Номери, назви розділів, тем і питання навчальних занять, посилання на літературу		Кількість годин				
		Всього	у тому числі			
			Лекції	Семінарські заняття	Практичні заняття	СР
Розділ 1. Теоретичні основи оброблення надвеликих масивів даних.						
Тема 1	Методи оброблення надвеликих масивів даних.	40	2	18		20
Заняття 1/1	Основні виклики надвеликих масивів даних.	2	2			
Заняття 1/2	Життєвий цикл оброблення надвеликих масивів даних.	4		2		2
Заняття 1/3	Аналітика надвеликих масивів даних.	4		2		2
Заняття 1/4	Завдання оброблення надвеликих масивів даних.	4		2		2
Заняття 1/5	Підходи до оброблення надвеликих масивів даних.	4		2		2
Заняття 1/6	Методи попереднього оброблення надвеликих масивів даних.	4		2		2
Заняття 1/7	Методи обирання ознак.	4		2		2
Заняття 1/8	Методи класифікування надвеликих масивів даних.	5		2		3
Заняття 1/9	Методи машинного навчання.	5		2		3
Заняття 1/10	Методи забезпечення кібербезпеки надвеликих масивів даних.	4		2		2
Разом за розділом 1		40	2	18		20
Розділ 2. Практичні основи оброблення надвеликих масивів даних.						
Тема 2	Технології оброблення надвеликих масивів даних	36	2	16		18
Заняття 2/1	Основні технології оброблення надвеликих масивів даних.	2	2			

Заняття 2/2	Технологія збереження і оброблення надвеликих масивів даних.	4		2		2
Заняття 2/3	Технологія паралельного оброблення надвеликих масивів даних.	4		2		2
Заняття 2/4	Технологія зберігання даних у нереляційних форматах.	4		2		2
Заняття 2/5	Технологія зберігання надвеликих обсягів неструктурованих даних.	4		2		2
Заняття 2/6	Технологія глибокого навчання.	4		2		2
Заняття 2/7	Технологія блокчейн.	5		2		2
Заняття 2/8	Інженерія кібербезпеки надвеликих масивів даних.	5		2		2
Заняття 2/9	Модульна контрольна робота.	6		2		4
Разом за розділом 2		36	2	16		18
Реферат		6				6
Залік		8		2		6
Всього годин		90	4	36		50

6. Самостійна робота аспіранта

Головними видами самостійної роботи аспірантів є: самостійна підготовка до аудиторних занять та самостійна підготовка до заліку.

Доцільно час самостійної підготовки для поглибленого вивчення та закріплення навчального матеріалу розподілити наступним чином:

№ з/п	Назва теми та перелік основних питань (перелік дидактичного забезпечення, посилання на літературу)	Кількість годин СР
1.	<p>Тема 1. Методи оброблення надвеликих масивів даних.</p> <ol style="list-style-type: none"> 1. Проблематика використання надвеликих масивів даних. 2. Фреймворки оброблення надвеликих масивів даних. 3. Порівняння етапів аналітики надвеликих масивів даних. 4. Інструментальні засоби аналітики надвеликих масивів даних. 5. Особливості завдання оброблення надвеликих масивів даних. 6. Проблематика попереднього оброблення надвеликих масивів даних. 7. Поняття ройового інтелекту. 8. Фреймворки розпаралелювання оброблення надвеликих масивів даних. 9. Проблематика класифікування надвеликих масивів даних. 10. Застосування надвеликих масивів даних в аспекті забезпечення кібербезпеки. <p>Основна література: [1-5]. Додаткова література: [1-5].</p>	20
2.	<p>Тема 2. Технології оброблення надвеликих масивів даних.</p> <ol style="list-style-type: none"> 1. Порівняння етапів оброблення надвеликих масивів даних. 2. Фреймворк Apache Spark. 3. Компоненти оброблення надвеликих масивів даних. 4. Моделі ACID і BASE. 5. Найкращі практики використання озер даних. 6. Прикладні застосування глибокого навчання. 7. Прикладні застосування блокчейну. 8. Надвеликі масиви даних забезпечення кібербезпеки. <p>Основна література: [1-5].</p>	18

	Додаткова література: [1-5].	
	Реферат	6
3	Підготовка до заліку.	6
Всього годин		50

Політика та контроль

7. Політика навчальної дисципліни

Система вимог, які викладач ставить перед аспірантом:

- правила відвідування занять: заборонено оцінювати присутність або відсутність здобувача на аудиторному занятті, в тому числі нараховувати заохочувальні або штрафні бали. Відповідно до РСО даної дисципліни бали нараховують за навчальну активність на семінарських заняттях і за модульну контрольну роботу;
- правила поведінки на заняттях: аспірант має можливість отримувати бали за навчальну активність на семінарських заняттях і за модульну контрольну роботу, передбачені РСО дисципліни. Використання засобів зв'язку для пошуку інформації в Інтернеті здійснюється за умови надання вказівок викладачем;
- політика дедлайнів та перескладань: якщо аспірант не проходив або не з'явився на семінарське заняття чи модульну контрольну роботу (без поважної причини), результат виконання ним завдання оцінюється у 0 балів. Перескладання семінарських занять та модульної контрольної роботи в даному випадку не передбачено.

Академічна доброчесність. Політика щодо академічної доброчесності: Кодекс честі Національного технічного університету України «Київський політехнічний інститут» (<https://kpi.ua/files/honorcode.pdf>) встановлює загальні моральні принципи, правила етичної поведінки осіб та передбачає політику академічної доброчесності для осіб, що працюють і навчаються в університеті, якими вони мають керуватись у своїй діяльності, в тому числі при вивченні та складанні контрольних заходів з дисципліни. Здобувач вищої освіти повинен знати, що викладання ґрунтується на засадах академічної доброчесності – сукупності етичних принципів та визначених законом правил, якими мають керуватися учасники освітнього процесу під час навчання, викладання та провадження наукової (творчої) діяльності з метою забезпечення **довіри** до результатів навчання та/або наукових (творчих) досягнень. **Порушеннями академічної доброчесності** вважаються: академічний **плагіат**, самоплагіат, фабрикація, фальсифікація, **списування**, **обман**, хабарництво, необ'єктивне оцінювання.

Кожний здобувач вищої освіти зобов'язаний дотримуватися принципів академічної доброчесності. Письмові завдання з використанням часткових або повнотекстових запозичень з інших робіт без зазначення авторства – це **плагіат**. Використання будь-якої інформації (текст, фото, ілюстрації тощо) мають бути правильно процитовані з посиланням на автора. Якщо ви не впевнені, що таке плагіат, фабрикація, фальсифікація, порадьтеся з викладачем. До здобувачів вищої освіти, у роботах яких буде виявлено списування, плагіат чи інші прояви недоброчесної поведінки можуть бути застосовані штрафні бали.

Зарахування результатів навчання, **набутих у неформальній/інформальній освіті** здійснюється відповідно до Положення про визнання в КПП ім. Ігоря Сікорського результатів навчання, набутих у неформальній/інформальній освіті, затвердженого та уведено в дію наказом від 09.05.2023 р. № НОН/157/2023 (<https://osvita.kpi.ua/node/179>).

У випадку запровадження обмежувальних заходів, що унеможливають організацію і здійснення освітнього процесу в навчальних приміщеннях у складі груп, проведення навчальних занять з даної навчальної дисципліни можна здійснювати віддалено з використанням технологій дистанційного навчання.

Навчальні матеріали та ресурси, зазначені у розділі 4 цієї робочої програми навчальної дисципліни (силабусі) є відкритими, не містять відомостей з обмеженим доступом і можуть бути оприлюднені з використанням технологій дистанційного навчання,

а сама програма не потребує коригування у випадку проведення навчальних занять у дистанційному режимі.

Правила призначення заохочувальних та штрафних балів. Заохочувальні бали виставляються за: активну участь на лекціях, науково-практичних конференціях, семінарах тощо. Кількість заохочуваних балів на більше 10. Штрафні бали виставляються за порушення політики навчальної дисципліни. Кількість штрафних балів на більше 10.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: доповіді на семінарських заняттях, модульна контрольна робота.
Семестровий контроль: залік.

Оцінювання результатів навчання аспірантів здійснюється відповідно до Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського затвердженого та уведеного в дію наказом № 1/273 від 14.09.2020 р., зі змінами, внесеними наказом № НОН/131/2022 від 03.05.2022 р.

Рейтингова оцінка трансформується до університетської системи оцінювання згідно з табл. 1.

Таблиця 1 – Переведення рейтингових балів до оцінок за університетською шкалою
Рейтингові бали, RDOцінка за університетською шкалою

Кількість балів	Оцінка
95-100	Відмінно
85-94	Дуже добре
75-84	Добре
65-74	Задовільно
60-64	Достатньо
Менше ніж 60	Незадовільно

Рейтингова система оцінювання результатів навчання.

Рейтинг аспіранта з навчальної дисципліни “Сучасні методи і технології оброблення надвеликих масивів даних” складається з балів, що він отримує за:

- 1) доповіді на семінарських заняттях;
- 2) модульну контрольну роботу;
- 3) написання реферату.

Система рейтингових (вагових) балів та критерії оцінювання.

1. Доповіді на семінарських заняттях.

Ваговий бал – 4. Максимальна кількість балів на всіх заняттях дорівнює 4 балів x 17 доп. = 68 балів.

Критерії оцінювання	Бали
Аспірант показав глибоке знання предмету, повно і чітко зробив доповідь, правильно і акуратно оформив презентацію доповіді, показав здатність вільно застосовувати свої теоретичні знання під час відповіді на запитання.	4
Оцінка зменшується, коли: – виконано всі перелічені вище вимоги, але є несуттєві недоліки; – доповідь загалом є правильною, але неповною; – відповіді на запитання вірні, але не повні; – була потрібна допомога викладача у вигляді правлень і додаткових питань;	3

– презентація доповіді оформлено неохайно.	
Бали зменшуються, коли при виконанні завдання є суттєві недоліки.	2
В інших випадках.	0

2. Модульна контрольна робота.

Ваговий бал – 16. Максимальна кількість балів за модульну контрольну роботу у семестрі дорівнює: 16 балів x 1 = 16 балів.

Критерії оцінювання для завдання	Бали
Завдання виконано повно та без зауважень.	14–16
Допущена неточність (неточності): – несуттєвого характеру; – при правильній, але не повній відповіді.	11–13
Протягом відповіді допущено помилки суттєвого характеру.	9–10
Відповідь відсутня	0

3. Реферат.

Ваговий бал – 16. Максимальна кількість балів за реферат у семестрі дорівнює: 16 балів x 1 = 16 балів.

Критерії оцінювання для завдання	Бали
Реферат підготовлено повно та оформлено без зауважень.	15-16
Допущена неточність несуттєвого характеру при підготовці та оформленні реферату.	12-14
Допущена неточність (неточності) при підготовці та оформленні реферату.	9-11
Під час підготовки та оформлення реферату допущено помилки принципового характеру.	8
Реферат відсутній	0

4. Залікова контрольна робота оцінюється 100 балами. Контрольне завдання цієї роботи складається з трьох запитань (два – теоретичних, одне – практичне).

Кожне теоретичне питання оцінюється 25 балами за такими критеріями:

- “відмінно” – повна відповідь (не менше 90% потрібної інформації), надані відповідні обґрунтування та особистий погляд – 25 ... 23 балів;
- “добре” – достатньо повна відповідь (не менше 75% потрібної інформації) з незначними неточностями – 22 ... 19 балів;
- “задовільно” – неповна відповідь (не менше 60% потрібної інформації) з деякими помилками – 18 ... 15 балів;
- “незадовільно” – незадовільна відповідь – 0 балів.

Практичне питання оцінюється 50 балами за такими критеріями:

- “відмінно” – повна відповідь (не менше 90% потрібної інформації), надані відповідні обґрунтування та особистий погляд – 50 ... 45 балів;
- “добре” – достатньо повна відповідь (не менше 75% потрібної інформації) з незначними неточностями – 44 ... 38 балів;
- “задовільно” – неповна відповідь (не менше 60% потрібної інформації) з деякими помилками – 37 ... 30 балів;
- “незадовільно” – незадовільна відповідь – 0 балів.

Максимальний рейтинг за семестр складає 100 балів.

Розрахунок шкали рейтингу:

Сума вагових балів контрольних заходів протягом семестру складає:

$$r_c = 68 + 16 + 16 = 100 \text{ балів.}$$

Рейтингова шкала з навчальної дисципліни складає:

$$RD = r_c = 100 \text{ балів.}$$

Аспірант, який набрав протягом семестру необхідну кількість балів ($RD \geq 60$), може отримати залікову оцінку (залік) так званим “автоматом” відповідно до набраного рейтингу. У такому разі до відомості семестрового контролю вносяться бали та відповідні оцінки.

Аспірант, який у семестрі отримав більше 60 балів, може взяти участь у заліковій контрольній роботі з метою підвищення оцінки. У цьому разі бали, отримані ним на заліковій контрольній роботі, є остаточними.

Якщо оцінка за залікову контрольну роботу більша ніж за рейтингом, аспірант отримує оцінку за результатами залікової контрольної роботи.

Якщо оцінка за залікову контрольну роботу менша, ніж за рейтингом, викладач застосовує жорстку рейтингову систему оцінювання. При цьому попередній рейтинг аспіранта з навчальної дисципліни (за винятком заохочувальних балів) скасовується і він отримує оцінку з урахуванням результатів залікової контрольної роботи.

Умовою допуску до залікової контрольної роботи є: написання модульної контрольної роботи, написання реферату та отримання рейтингу не менше 36 балів.

Знання аспірантів оцінюються за шкалою: “Відмінно”, “Дуже добре”, “Добре”, “Задовільно”, “Достатньо”, “Незадовільно” із наступним перерахуванням в бали РСО згідно з табл. 1.

9. Додаткова інформація з навчальної дисципліни

Перелік тем, які виносяться на залік.

1. Основні виклики надвеликих масивів даних.
2. Життєвий цикл оброблення надвеликих масивів даних.
3. Аналітика надвеликих масивів даних.
4. Завдання оброблення надвеликих масивів даних.
5. Підходи до оброблення надвеликих масивів даних.
6. Методи попереднього оброблення надвеликих масивів даних.
7. Методи обирання ознак.
8. Методи класифікування надвеликих масивів даних.
9. Методи машинного навчання.
10. Методи забезпечення кібербезпеки надвеликих масивів даних.
11. Основні технології оброблення надвеликих масивів даних.
12. Технологія збереження і оброблення надвеликих масивів даних.
13. Технологія паралельного оброблення надвеликих масивів даних.
14. Технологія зберігання даних у нереляційних форматах.
15. Технологія зберігання надвеликих обсягів неструктурованих даних.
16. Технологія глибокого навчання.
17. Технологія блокчейн.
18. Інженерія кібербезпеки надвеликих масивів даних.

Рекомендовано до ухвалення Методичною комісією ІСЗЗІ КПІ ім. Ігоря Сікорського на засіданні Спеціальної кафедри № 5 (протокол від 29.06.2023 № 11).