


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ**

Кафедра Програмної інженерії



ЗАТВЕРДЖУЮ
Декан факультету КН


А.Л.Єрохін
(підпис, ініціали, прізвище)
15.09.2020 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ТЕХНОЛОГІЇ DATA SCIENCE

рівень вищої освіти _____ магістерський _____
(бакалаврський, магістерський, освітньо-науковий)

спеціальність _____ 121 Інженерія програмного забезпечення _____

освітня-наукова програма - _____ Інженерія програмного забезпечення _____

освітня-професійна програма - _____ Інженерія програмного забезпечення _____

Харків, 2020

Розробники: К.С. Смеляков професор кафедри ПІ, д.т.н., професор
(ініціали, прізвище, посада, науковий ступінь, вчене звання)

Робочу програму схвалено на засіданні кафедри Програмної інженерії

Протокол № 1 від «1» 09 2020 року


Завідувач кафедри


(підпис)

Дудар З.В.

(ініціали, прізвище)

Керівник групи забезпечення якості


(підпис)

Лесна Н.С.

(ініціали, прізвище)

Схвалено методичною комісією факультету КН.

Протокол від «15» 09 2020 р. № 2

Голова методичної комісії


(підпис)

Лановий О.Ф.

(ініціали, прізвище)

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Характеристика навчальної дисципліни	
	денна форма навчання	заочна форма навчання
Кількість кредитів ЄКТС – <u>5</u>	Професійна та практична підготовка (обов'язкова, норм. дисц. – ПЗ; дисц. сам. вибору НЗ - ПЗС)	
Модулів – 5	Рік підготовки	
Змістових модулів – 5	2023/2024	
Індивідуальних завдання: РГЗ та КР _____	Семестр	
Загальна кількість годин – 150	2-й, 5 курс	
	Навчальні заняття:	
Тижневих годин для денної форми навчання: аудиторних – 2,4 самостійної роботи студента – 4,8 Тижневих годин для заочної форми навчання: аудиторних – 0,9 самостійної роботи студента – 6,3	1) лекції, год	
	26	
	2) практичні, год	
	3) лабораторні, год	
	16	
	4) консультації, год.	
	10	
	Самостійна робота, год	
	98	
в тому числі: 1) РГЗ та КР.		
Вид контролю: залік.		

2. МЕТА ДИСЦИПЛІНИ ТА ОЧІКУВАНІ РЕЗУЛЬТАТИ З ЇЇ ВИВЧЕННЯ

Метою дисципліни є визначення підходів і методології при розробці інноваційних додатків для високопродуктивних обчислень, які зорієнтовано на роботу із потоками даних, базами даних, вітринами даних, сховищами даних і озерами даних у реальному масштабі часу.

Визначення та аналіз основних моделей даних, методології обробки великих даних у реальному масштабі часу, а також інформаційних технологій обробки великих даних провідних світових виробників програмного забезпечення.

Виконання індивідуальних завдань практичної спрямованості для отримання відповідних навичок з дисципліни.

Наукова складова дисципліни – дослідження моделей і алгоритмів високопродуктивних обчислень.

Інформаційні технології: Apache HTTP Server, OpenMP, MPI, CUDA, Python.

Завдання дисципліни: за результатом вивчення дисципліни студенти мають володіти загальними / фаховими компетентностями та програмними результатами відповідно до змісту *освітньо-наукової програми Інженерія програмного забезпечення підготовки фахівців 2-го (магістерського) рівня вищої освіти спеціальності 121 Інженерія програмного забезпечення, та освітньо-професійної програми Програмне забезпечення систем підготовки фахівців 2-го (магістерського) рівня вищої освіти спеціальності 121 Інженерія програмного забезпечення, а саме:*

А) Загальними компетентностями:

1. Здатністю спілкуватися іноземною мовою для оцінювання джерел й отримання інформації з різноманітних джерел даних щодо створення профільного програмного забезпечення (ПЗ).
2. Здатністю проведення теоретичних та прикладних досліджень щодо створення інноваційного програмного забезпечення за напрямом високопродуктивних обчислень.
3. Здатністю удосконалювати свої навички на основі аналізу попереднього досвіду.
4. Здатністю генерувати нові ідеї відповідно до класифікації даних / джерел даних та форм їх використання щодо створення ПЗ для високопродуктивних обчислень.

Б) Фаховими компетентностями:

1. Здатністю аналізувати предметні області, формувати, аналізувати та моделювати вимоги до інноваційного ПЗ для високопродуктивних обчислень.
2. Здатністю розвивати і реалізовувати нові конкурентоспроможні інноваційні ідеї та інноваційні моделі для створення стартапів в ІТ та в інженерії ПЗ для високопродуктивних обчислень.
3. Здатністю оцінювати ступінь обґрунтованості застосування специфікацій, стандартів, правил і рекомендацій в професійній галузі та дотримуватися їх при реалізації процесів життєвого циклу цифрових продуктів та послуг для високопродуктивних обчислень.
4. Здатністю ефективно обирати інструменти та ресурси при плануванні / проектуванні сучасного програмного забезпечення у додатках з метою забезпечення ефективності аналізу не тільки статичних, але й поточкових даних.
5. Здатністю забезпечити виконання норм законодавства України відносно інтелектуальної власності та свідомо її застосовувати для захисту прав та економічних інтересів на інноваційне ПЗ для високопродуктивних обчислень.
6. Здатністю обґрунтовувати та вибирати критерії оцінки технічного стану основних й економічних показників при виробництві та експлуатації інноваційного ПЗ для високопродуктивних обчислень.

В) Програмні результати:

1. Знати і системно застосовувати методи аналізу та моделювання прикладної області, виявлення інформаційних потреб і збору вихідних даних для проектування інноваційного ПЗ для високопродуктивних обчислень.
2. Знати й вміти застосовувати сучасні технології розробки і тестування систем і сервісів для високопродуктивних обчислень.
3. Розробляти і оцінювати стратегії проектування інноваційного ПЗ для високопродуктивних обчислень; враховуючи обґрунтування вибору інструментів, сервісів та обладнання з точки зору якості кінцевого Data Science цифрового продукту.
4. Вміти приймати організаційно-управлінські рішення в умовах невизначеності для ефективності процесів та механізм розробки інноваційних проектів.
5. Набувати нові наукові і професійні знання, вдосконалювати навички, прогнозувати тенденції ринку ІТ стосовно потреб та інновацій у сфері високопродуктивних обчислень.
6. Формулювати, експериментально підтверджувати, обґрунтовувати і застосовувати на практиці в процесі розробки ПЗ конкурентоспроможні ідеї, методи, технології вирішення професійних, науково-технічних завдань в умовах невизначеності. А також в складних умовах функціонування Big Data сервісів.

3. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**Змістовий модуль 1. Основні поняття та визначення**

Тема 1. Вступ.

Тема 2. Можливості й інструменти Python.

Тема 3. Основні технології і рішення у предметній галузі.

Змістовий модуль 2. Робота з NumPy

Тема 4. Робота з масивами.

Тема 5. Робота з універсальними функціями.

Тема 6. Робота з структурами даних.

Змістовий модуль 3. Робота з Pandas

Тема 7. Робота з об'єктами.

Тема 8. Операції з даними.

Тема 9. Розширені можливості роботи з даними.

Змістовий модуль 4. Розвинені методи аналізу даних

Теми 10. Розвинені можливості Python і додатків.

Теми 11. Використання статистичних моделей.

Змістовий модуль 5. Методи штучного інтелекту

Тема 12. Використання інструментів штучного інтелекту.

Тема 13. Моделі і технології машинного навчання.

4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин						Кількість годин					
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
лк		пз	лаб.	конс	с. р.	конс		пз	лаб.	конс	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13
Змістовий модуль 1.												
Тема 1. Вступ.	8	2				6						
Тема 2. Можливості й інструменти Python.	10	2				6						
Тема 3. Основні технології і рішення у предметній галузі.	12	2			2	6						
Разом за змістовим модулем 1	30	6			2	18						
Змістовий модуль 2.												
Тема 4. Робота з масивами.	8	2				6						
Тема 5. Робота з універсальними функціями.	8	2				6						
Тема 6. Робота з структурами даних.	14	2		4	2	11						
Разом за змістовим модулем 2	30	6		4	2	23						
Змістовий модуль 3.												
Тема 7. Робота з об'єктами.	8	2				6						
Тема 8. Операції з даними.	8	2				6						
Тема 9. Розширені можливості роботи з даними.	14	2		4	2	11						
Разом за змістовим модулем 3	30	6		4	2	23						
Змістовий модуль 4.												
Тема 10. Розвинені можливості Python і додатків.	8	2				6						
Тема 11. Використання статистичних моделей.	8	2		4	2	11						
Разом за змістовим модулем 4	30	4		4	2	17						
Змістовий модуль 5.												
Тема 12. Моделі і технології машинного навчання.	14	2		4	2	11						
Тема 13. Використання інструментів штучного інтелекту.		2				6						
Разом за змістовим модулем 5	30	4		4	2	17						
Усього годин	150	26		16	10	98						

5. ТЕМИ ПРАКТИЧНИХ ЗАНЯТЬ

№ з/п	Теми занять	Обсяг, год.

6. ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ з/п	Теми занять	Обсяг, год.
1	Дослідження використання структур даних	4
2	Дослідження роботи з розвиненими структурами даних	4
3	Дослідження роботи з статистичними моделями даних	4
4	Дослідження використання моделей машинного навчання	4
	Загальна кількість	16

7. САМОСТІЙНА РОБОТА

№ з/п	Вид самостійної роботи	Форма навчання	
		Денна обсяг, год	Заочна обсяг, год
1	Тема 1. Основні поняття високопродуктивних обчислень.	6	
2	Тема 2. Високопродуктивні обчислення в ЄС.	6	
3	Тема 3. Засоби реалізації високопродуктивних обчислень.	6	
4	Тема 4. Сервіси високопродуктивних обчислень.	6	
5	Тема 5. Програмні засоби високопродуктивних обчислень.	6	
6	Тема 6. Засоби паралельної обробки даних.	11	
7	Тема 7. Класифікація об'єктів.	6	
8	Тема 8. Класифікація і операції з даними.	6	
9	Тема 9. Обробка нетривіальних моделей даних.	11	
10	Тема 10. Нетривіальні методи аналізу даних і сервіси.	6	
11	Тема 11. Генератори.	11	
12	Тема 12. НМ, дерева і візуалізація рішень.	6	
13	Тема 13. Методи прогнозування.	11	
	Загальна кількість	98	

8. ІНДИВІДУАЛЬНІ ЗАВДАННЯ

8.1 Розрахунково-графічні завдання (РГЗ) та контрольні роботи (КР) – не передбачені планом.

8.2 Курсова робота – не передбачена планом.

9. МЕТОДИ НАВЧАННЯ ТА ЗАСОБИ ОЦІНЮВАННЯ

Методи навчання є системними об'єктами, які об'єднують низку взаємопов'язаних дій викладача й студентів, спрямованих на виконання освітньої, розвивальної, виховної і контрольної функцій.

Методи навчання за джерелами знань:

- словесні (лекція, консультація)
- наочні (ілюстрації, демонстраційні приклади).

Вивчення дисципліни здійснюється традиційними методами із застосуванням новітніх інформаційних технологій, слайд лекцій.

Теоретичні знання, що викладаються під час лекцій, використовуються на практичних і лабораторних роботах, що проводяться у комп'ютерних аудиторіях, які обладнані сучасними комп'ютерними засобами.

1. Методи організації та здійснення навчально-пізнавальної діяльності за допомогою слайд-лекцій, пояснень на багатьох практичних прикладах.
2. Методи стимулювання й мотивації навчально-пізнавальної діяльності студентів у виконанні власних проектів з практичної реалізації завдань дисципліни.
3. Методи контролю (самоконтролю, корекції (самокорекції)), за ефективністю навчально-пізнавальної діяльності студента ці методи спрямовані на самостійну, творчу пізнавальну діяльність студентів, особливо при створенні власних проектів.
4. Універсальні методи поєднують самостійну роботу студентів під час практичних занять з інструктуванням, допомогою викладача, у результаті чого студенти набувають навичок самостійності та самостійною роботою студентів поза аудиторного навантаження. Крім цього студент має у своєму розпорядженні слайд-лекції, приклади розв'язання задач з роз'ясненнями – усе це поєднується в наочно-ілюстративно-практичний комплект матеріалів для навчання.

10. МЕТОДИ КОНТРОЛЮ ТА РЕЙТИНГОВА ОЦІНКА ЗА ДИСЦИПЛІНОЮ

Методи контролю – це способи діагностичної діяльності, які дозволяють здійснювати зворотний зв'язок у процесі навчання з метою отримання даних про успішність навчання, ефективність навчального процесу. Вони мають забезпечувати систематичне, повне, точне і оперативне отримання інформації про навчальний процес.

Виділяють наступні **методи контролю**: методи усного контролю, методи письмового контролю, методи практичного контролю, тести.

Методи усного контролю – виступ перед аудиторією за означеною проблематикою.

Усне опитування допомагає контролювати не лише знання, а й вербальні вміння, сприяє виправленню мовних помилок. Відтворення студентом раніше вивченого матеріалу сприяє кращому запам'ятовуванню, активному використанню наукових понять, що неможливо без достатнього застосування їх у мові.

Усне опитування може бути індивідуальним і фронтальним. За фронтального опитування студенти відповідають з місця, доповнюючи один одного. Частковим випадком фронтального опитування є групове опитування – 5-6 осіб одночасно. Індивідуальне опитування здійснюється у процесі проведення співбесіди під час практичних занять.

Запитання для усної перевірки знань поділяють на основні, додаткові і допоміжні. Основні запитання передбачають самостійну розгорнуту відповідь (наприклад, запитання щодо змісту лабораторного заняття). Додаткові - спрямовані на уточнення того, як студент розуміє певне питання, формулювання, формулу та ін. Допоміжні запитання мають за мету виправлення помилок та неточностей, якщо такі мали місце у відповіді студента. Усі запитання - логічні, чіткі, зрозумілі, а їх сукупність – послідовна і системна.

Письмовий контроль (аудиторна контрольна робота у вигляді тесту або відповіді на запитання) забезпечує глибоку і всебічну перевірку засвоєння, оскільки вимагає комплексу знань і умінь студента.

Виконання практичних завдань, домашньої контрольної роботи. У письмовій роботі студенту необхідно показати і теоретичні знання, і вміння застосовувати їх для розв'язування конкретних задач, проблем, крім того, виявляється ступінь оволодіння письмовою мовою, вміння логічно, адекватно проблемі скласти свій текст, давати оцінку, експерименту, проблемі.

Письмовий контроль можна здійснюється у вигляді відповідей на запитання, розв'язання задач під час виконання практичних робіт. Письмові роботи допомагають за короткий час з'ясувати рівень засвоєння матеріалу у великої кількості студентів. Результати письмових робіт можна проаналізувати і з'ясувати деталі і неточності у відповідях та діагностувати їх причини.

Практичний контроль – передбачає виявлення вмінь і навичок студентів, що набуті під час практичної діяльності (практичних занять, робота над власним проектом). Така перевірка дає змогу виявити, на якому рівні студент засвоїв теоретичні основи цих дій.

Програмований контроль здійснюється за допомогою бланкових тестів або в системі Open Test.

Поточний контроль дозволяє отримати достатньо повні дані про студента: і рівень його знань, умінь з предмету, і ставлення до навчання, ступінь його пізнавальної активності, свідомості, і вміння мислити, розв'язувати самостійно різноманітні завдання.

Іспит є найбільш активною перевіркою знань за визначений період навчання.

10.1 Розподіл балів, які отримують студенти (Кількісні критерії оцінювання)

Min/ max рейтингова оцінка	Вид заняття / контрольний захід									
	АКР 1 / 1 КТ	ЛР№ 1	АКР 2 / 2 КТ	ЛР№ 2	АКР 3 / 3 КТ	ЛР№ 3	АКР 4 / 4 КТ	ЛР№ 4	АКР 4 / 5 КТ	Разом за 1-у і 2-гу КТ
Відпрацювання та захист Min	12	3	9	3	9	3	9	3	9	60
Відпрацювання та захист Max	20	5	15	5	15	5	15	5	15	100

10.2 Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки.

- Знати основні поняття та визначення високопродуктивних обчислень.
- Знати особливості реалізації і імплементації проектів з високопродуктивних обчислень.
- Вміти працювати зі сховищами даних.
- Вміти застосовувати інструменти для отримання і інтеграції даних; потоків даних.
- Вміти використовувати сучасні засоби візуалізації даних.
- Вміти використовувати основні програмні додатки і бібліотеки для високопродуктивних обчислень.
- Вміти використовувати основні програмні додатки Пайтон.
- Знати основні методи й програмні додатки прогнозування.
- Володіти статистичними методами прогнозування.
- Володіти методами прогнозування на основі використання моделей штучного інтелекту (перш за все – нейроні мережі, дерева рішень).
- Володіти методами прогнозування у багатовимірних просторах. Володіти методами й технологіями підготовки даних (фільтрація, очищення, заповнення пропусків), методом головних компонент, а також методами розбиття багатовимірного простору.

Необхідний обсяг умінь для одержання позитивної оцінки.

- Вміти застосовувати основні додатки Пайтон.
- Вміти застосовувати основні додатки статистичного аналізу даних.
- Вміти використовувати основні моделі машинного навчання.

10.3 Критерії оцінювання роботи студента протягом семестру.

Як форма підсумкового контролю для дисципліни використовується письмовий (комбінований) залік. При цьому виді контролю підсумкова оцінка P_{Π} обчислюється за формулою : $P_{\Pi}=0.6 \cdot O_{\text{сем}}+0.4 \cdot O_{\text{ісп}}$, де $O_{\text{сем}}$ – оцінка за семестр у 100-бальній системі, $O_{\text{ісп}}$ – оцінка за залік у 100-бальній системі.

Задовільно, D, E (60-74). Відпрацювати та захистити усі лабораторні роботи, АКР. Знати основні поняття та визначення високопродуктивних обчислень; основні технології збору й обробки даних; обробку статичних і потокових даних; основні технології Пайтон.

Добре, C (75-89). Відпрацювати та захистити усі лабораторні роботи, АКР. Знати основні поняття та визначення високопродуктивних обчислень; основні технології збору й обробки даних; обробку статичних і потокових даних; основні технології Пайтон і сервісів. Крім того, знати й уміти застосовувати інструменти Predictive Analytics.

Відмінно, A, B (90-100). Відпрацювати та захистити усі лабораторні роботи, АКР. Знати основні поняття та визначення високопродуктивних обчислень; основні технології збору й обробки даних; обробку статичних і потокових даних; основні технології Пайтон і сервісів. Крім того, знати й уміти застосовувати інструменти Predictive Analytics. Досконально знати прогностичні моделі, оцінки ефективності їх застосування, оцінки похибки.

Критерії оцінювання знань та вмінь студента на заліку.

Задовільно, D, E (60-74). Висвітлити необхідний мінімум теоретичних знань. Знати методи розв'язання практичної задачі.

Добре, C (75-89). Показати добрі теоретичні знання матеріалу. Розв'язати практичне завдання.

Відмінно, A, B (90-100). Показати бездоганні знання теоретичного матеріалу та додаткового матеріалу. Розв'язати практичне завдання без помилок та вміти обґрунтувати своє рішення.

Білет для письмового заліку складається з двох теоретичних запитань та практичного завдання. Теоретичні запитання оцінюються за 100-бальною шкалою в 25 балів кожне, а завдання в 50 балів. Студент може не здавати залік, погодитися на оцінку, яку він заробив на протязі семестру за накопиченням.

Розподіл балів, які отримують студенти.

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для іспиту, курсового проекту (роботи), практики	для заліку
96-100	A	відмінно добре задовільно	зараховано
90-95	B		
75-89	C		
66-74	D		
60-65	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

11. МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА

11.1. Базова

1. Georg Hager, Gerhard Wellein Introduction to High Performance Computing for Scientists and Engineers (Chapman & Hall/CRC Computational Science). – CRC Press; 1st edition (July 2, 2010). – 356 pages.
2. Конспект лекцій з дисципліни.
3. Брендан Тірні, Джон Келлехер Наука про дані. - Видавництво Альпіна Діджітал, 2020. - 175с.
4. Луцків А.М., Лупенко С.А., Пасічник В.В. Паралельні та розподілені обчислення. – Львів Видавництво: Магнолія 2015, 566с.

5. Thomas Sterling, Maciej Brodowicz, Matthew Anderson High Performance Computing: Modern Systems and Practices. – Morgan Kaufmann; 1st edition (December 19, 2017). – 718 pages.
6. Андреас Вайгенд BIG DATA. Уся технологія в одній книзі. - Видавництво Ексмо, 2018. - 414с.
7. Robert Robey, Yuliana Zamora Parallel and High Performance Computing. – Manning (June 22, 2021). – 704 pages.

12. МАТЕРІАЛЬНО-ТЕХНІЧНЕ, ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

Інформаційні технології: Apache HTTP Server, OpenMP, MPI, CUDA, Python.