Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
(повна назва)

Кафедра програмної інженерії
(повна назва)

ЗАТВЕРДЖУЮ
Декан факультету
А.Л. Єрохін
(підпис, ініціали, прізвище)
«01» вересня 2021р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Технології обробки великих даних

(назва навчальної дисципліни)

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

спеціальність 121 Інженерія програмного забезпечення
(код і повна назва спеціальності)

освітньо-наукова програма (вибіркова)
(професійна або наукова)

Інженерія програмного забезпечення
(повна назва програми)

Харків – 2021 р.

Розробник(и): К.С. Смеляков, професор кафедри програмної інженерії, д.т.н., професор

(ініціали, прізвище, посада, науковий ступінь, вчене звання)

Робочу програму схвалено на засіданні кафедри програмної інженерії

Протокол від «31» серпня 2021 р. № 1

Завідувач кафедри _____ З.В. Дудар

(підпис) (ініціали, прізвище)

Схвалено методичною комісією факультету комп'ютерних наук

Протокол від «31» 08 2021 р. № 1

Голова методичної комісії _____ О.Ф. Лановий

(підпис) (ініціали, прізвище)

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<table>
<tr><td rowspan="2">Найменування показників</td><td rowspan="2">Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень</td><td colspan="4">Характеристика навчальної дисципліни</td></tr>
<tr><td colspan="2">денна форма навчання</td><td colspan="2">заочна форма навчання</td></tr>
<tr><td rowspan="2">Кількість кредитів – 8</td><td>Галузь знань
12 – інформаційні технології</td><td colspan="4" rowspan="2">Дисципліни професійної та практичної підготовки</td></tr>
<tr><td>Напрям підготовки:
121 – інженерія програмного забезпечення</td></tr>
<tr><td></td><td rowspan="6">Спеціальність:
121 – інженерія програмного забезпечення</td><td colspan="4">Рік підготовки:</td></tr>
<tr><td>Змістових модулів – 4</td><td colspan="2">1-й</td><td colspan="2"></td></tr>
<tr><td rowspan="3"></td><td colspan="4">Семестр</td></tr>
<tr><td colspan="2">2-й</td><td colspan="2"></td></tr>
<tr><td colspan="4">Кількість годин</td></tr>
<tr><td>Загальна кількість годин – 240</td><td colspan="2">240 год.</td><td colspan="2"></td></tr>
<tr><td rowspan="15">Тижневих годин для денної форми навчання:

аудиторних:
самостійної роботи студента:</td><td rowspan="15">Рівень вищої освіти:
третій (освітньо-науковий)</td><td colspan="4">Аудиторні: 1) лекції</td></tr>
<tr><td colspan="2">48</td><td colspan="2"></td></tr>
<tr><td colspan="4">2) практичні</td></tr>
<tr><td colspan="2">48 год.</td><td colspan="2"></td></tr>
<tr><td colspan="4">3) лабораторні</td></tr>
<tr><td colspan="2">-</td><td colspan="2"></td></tr>
<tr><td colspan="4">4) консультації</td></tr>
<tr><td colspan="2">16 год.</td><td colspan="2"></td></tr>
<tr><td colspan="4">Самостійна робота</td></tr>
<tr><td colspan="2">128</td><td colspan="2"></td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">2) курсова робота, год.</td></tr>
<tr><td colspan="2">-</td><td colspan="2"></td></tr>
<tr><td colspan="4">Вид контролю:</td></tr>
<tr><td>-</td><td>залік</td><td></td><td></td></tr>
</table>

Співвідношення кількості годин аудиторних занять до загальної кількості годин становить:

для денної форми навчання – 46,(6)%.

2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета: засвоєння аспірантами основних теоретичних відомостей щодо сучасних підходів до аналізу великих даних; набуття практичних навичок використання сучасних програмних засобів обробки великих даних.

За результатом вивчення навчальної дисципліни аспірант повинен:

– *знати*: області застосування Big Data, процес data science у зв’язку з Big Data, методи очищення, інтеграція та перетворення даних, моделі і методи машинного навчання, у тому числі процес моделювання, локальні принципи роботи з великими даними (на одному комп'ютері), аспекти вибору алгоритмів обробки даних і відповідних структур даних, технології розподілення і зберігання великих даних, технології SQL і NoSQL, графові бази даних, основні методі і технології аналізу тексту і, відповідно, способи і технології візуалізації даних.

– *уміти*: застосовувати інструменти та моделі аналізу великих даних (у зв’язку з Data Science технологіями), будувати і обирати моделі даних, виконувати попередню обробку даних (очищення, інтеграція, перетворення даних та інше), використовувати методи машинного навчання, уміти обирати ефективний алгоритм / технологію обробки даних, а також структури для роботи з великими даними (локально і глобально), застосовувати методологію розподілення і зберігання великих даних, уміти використовувати сучасні бази даних (перш за все NoSQL і графові бази даних), проводити поглиблений аналіз тексту, застосовувати класифікатори повідомлень, застосовувати методи аналізу даних високого рівня, а також технології візуалізації даних, виконувати дослідження реальних сховищ даних, презентувати результатів наукових досліджень; здійснювати дослідження з використанням цих ресурсів та технологій.

– *володіти* (компетенції): здатність застосовувати методологію та технології великих даних (у зв’язку з методологією Data Science), реалізовувати відповідні методи, алгоритми й сучасні технології для дослідження складних об'єктів, систем і явищ, перевіряти отримані результати та інтерпретувати їх.

Компетентності:

Здатність застосовувати методологію та технології інтелектуального аналізу даних, реалізовувати його методи й алгоритми для дослідження складних об'єктів і систем, перевіряти отримані результати та інтерпретувати їх.

Здатність використовувати, адаптувати та розробляти інформаційні технології вирішення задач у інженерії програмного забезпечення та дотичних до неї міждисциплінарних напрямах щодо управління, підтримки прийняття рішень, пошуку та аналізу даних.

Здатність виконувати інтерпретацію результатів досліджень з урахуванням їх наукового значення та результатів експериментальної перевірки.

Програмні результати навчання:

Знати та розуміти основні методи аналізу даних; вміти застосовувати інструменти та моделі аналізу даних (пакети прикладних програм, онлайн ресурси й відповідні технології) в дослідженні реальних систем та презентації результатів наукових досліджень у різних формах; здійснювати науково-педагогічну діяльність з використанням цих ресурсів.

Уміти застосовувати, удосконалювати та розробляти нові математичні моделі та методи проектування, підтримки й супроводу сучасних інформаційних технологій, а також виконувати їх експериментальну перевірку.

3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Базова методологія Big Data

Тема 1. Область застосування Data Science і Big Data

Тема 2. Процес Data Science

Тема 3. Етапи процесу Data Science. Частина 1

Тема 4. Етапи процесу Data Science. Частина 2

Тема 5. Очищення, інтеграція та перетворення даних

Тема 6. Машинне навчання

Тема 7. Машинне навчання. Процес моделювання

Тема 8. Типи машинного навчання

Змістовий модуль 2. Основні аспекти роботи з Big Data

Тема 1. Робота з великими даними на одному комп'ютері

Тема 2. Вибір ефективного алгоритму

Тема 3. Вибір структури даних

Тема 4. Рекомендації програмістам при роботі з великими наборами даних

Тема 5. Розподілення і зберігання великих даних

Змістовний модуль 3. Робота з базами даних

Тема 1. Робота з NoSQL. Частина 1

Тема 2. Робота з NoSQL. Частина 2

Тема 3. Графові бази даних

Тема 4. Графічна база даних Neo4j

Змістовний модуль 4. Аналіз текстових даних і візуалізація

Тема 1. Поглиблений аналіз тексту

Тема 2. Методи глибокого аналізу тексту

Тема 3. Класифікація повідомлень. Частина 1

Тема 4. Класифікація повідомлень. Частина 2

Тема 5. Дослідження і підготовка даних

Тема 6. Аналіз даних

Тема 7. Способи візуалізації даних

4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	Усього	у тому числі					Усього	у тому числі				
		л	п	лб	конс	с.р.		л	п	лб	конс	с.р.
Змістовий модуль 1. Базова методологія Big Data												
Тема 1. Область застосування Data Science і Big Data		2										
Тема 2. Процес Data Science		2			2							
Тема 3. Етапи процесу data science. Частина 1		2										
Тема 4. Етапи процесу data science. Частина 2		2	4		2							
Тема 5. Очищення, інтеграція та перетворення даних		2	4									
Тема 6. Машинне навчання		2			2							
Тема 7. Машинне навчання. Процес моделювання		2										
Тема 8. Типи машинного навчання		2	4		2							
Разом за зміст. мод. 1	68	16	12	-	8	32						
Змістовий модуль 2. Основні аспекти роботи з Big Data												
Тема 1. Робота з великими даними на одному комп'ютері		2	4									
Тема 2. Вибір ефективного алгоритму		2	4		2							
Тема 3. Вибір структури даних		2	4									
Тема 4. Рекомендації програмістам при роботі з великими наборами даних		2	4									
Тема 5. Розподілення і зберігання великих даних		2	4									
Разом за зміст. мод. 2	64	10	20	-	2	32						
Змістовний модуль 3. Робота з базами даних												
Тема 1. Робота з NoSQL. Частина 1		2										
Тема 2. Робота з NoSQL. Частина 2		2	4									
Тема 3. Графові бази даних		2	4		2							
Тема 4. Графічна база даних Neo4j		2	4									
Разом за зміст. мод. 3	54	8	12	-	2	32						
Змістовний модуль 4. Аналіз текстових даних і візуалізація												
Тема 1. Поглиблений ана-		2										

ліз тексту												
Тема 2. Методи глибокого аналізу тексту		2			2							
Тема 3. Класифікація повідомлень. Частина 1		2										
Тема 4. Класифікація повідомлень. Частина 2		2										
Тема 5. Дослідження і підготовка даних		2			2							
Тема 6. Аналіз даних		2										
Тема 7. Способи візуалізації даних		2	4									
Разом за зміст. мод. 4	54	14	4	-	4	32						
Усього годин	**240**	**48**	**48**	**-**	**16**	**128**						

5 ТЕМИ ПРАКТИЧНИХ ЗАНЯТЬ

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Вступний приклад використання Hadoop	4	
2	Дослідницький аналіз даних	4	
3	Побудова моделей і моделювання	4	
4	Побудова рекомендаційної системи всередині бази даних	4	
5	Індексація і інші допоміжні операції в базі даних	4	
6	Аналіз великих даних для оцінки риску кредитування. Частина 1	4	
7	Аналіз великих даних для оцінки риску кредитування. Частина 2	4	
8	Аналіз великих даних для оцінки риску кредитування. Частина 3	4	
9	Аналіз великих для діагностики захворювань. Частина 1	4	
10	Аналіз великих для діагностики захворювань. Частина 2	4	
11	Аналіз великих для діагностики захворювань. Частина 3	4	
12	Використання сучасних засобів візуалізації даних	4	
	Загальна кількість	**48**	

6 САМОСТІЙНА РОБОТА

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1	Вивчення теоретичного матеріалу з використанням конспектів і навчальної літератури	48	
2	Підготовка до практичних занять	24	
3	Підготовка до семестрового контролю	4	
5	Самостійне опрацювання матеріалу за літературними джерелами: Тема 1. Основні джерела Big Data. Тема 2. Big Data у фактах та цифрах. Тема 3. Засоби імплементації Big Data проектів. Тема 4. Класифікація сховищ даних. Тема 5. Засоби отримання статичних і потокових даних. Тема 6. Засоби візуалізації багатовимірних даних. Тема 7. Загальна характеристика Apache Software Foundation стосов-	52	

	но проектів Big Data. Тема 8. Особливості використання Hadoop MapReduce. Тема 9. Додатки Apache Software Foundation для роботи с базами даних. Тема 10. Класифікація і характеристика програмних додатків SAP. Тема 11. Інтеграція Big Data додатків на платформі SAP. Тема 12. Візуалізація рішень у додатках SAP. Тема 13. Класифікація методів прогнозування. Тема 14. Методи прогнозування на основі штучного інтелекту. Тема 15. Методи прогнозування у багатовимірних просторах.		
	Загальна кількість	**128**	

7 МЕТОДИ НАВЧАННЯ

Метод навчання – це упорядкована діяльність викладача і аспірантів, спрямована на досягнення заданої мети навчання. Основні методи навчання:

- пояснювально-ілюстративний (лекції);
- практичний (практичні заняття);
- робота з навчально-методичною літературою (конспектування, самостійне опрацювання заданих розділів, виконання завдань).
- перевірка знань та умінь (за результатами виконання практичних занять, індивідуальних завдань).

8 МЕТОДИ КОНТРОЛЮ ТА РЕЙТИНГОВА ОЦІНКА ЗА ДИСЦИПЛІНОЮ

8.1 Розподіл балів, які отримують аспіранти (Кількісні критерії оцінювання)

Для оцінювання роботи аспіранта протягом 1-го та 2-го семестрів підсумкова рейтингова оцінка $O_{\text{сем}}$ розраховується як сума оцінок, які аспірант набрав протягом цих семестрів, виконуючи всі види контролю, передбачені робочою програмою.

Вид заняття / контрольний захід	Оцінка $O_{\text{сем}}$
ПЗ №1	2...5
ПЗ №2	2...5
ПЗ №3	2...5
ПЗ №4	2...5
ПЗ №5	2...5
ПЗ №6	2...5
ПЗ №7	2...5
ПЗ №8	2...5
ПЗ №9	2...5
ПЗ №10	2...5
ПЗ №11	2...5
ПЗ №12	2...5
Залік, тестування	**36…40**
Всього	60…100

Як форма підсумкового контролю для дисципліни наприкінці 2-го семестру використовується залік. При цьому виді контролю підсумкова рейтингова оцінка $P_{\text{п}} = O_{\text{сем}}$.

Отримані бали переводяться за національною шкалою та шкалою ECTS.

8.2 Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки.

Знати: основні області застосування Big Data, процес data science у зв'язку з Big Data, методи очищення, інтеграція та перетворення даних, базові моделі і методи машинного навчання, у тому числі процес моделювання, локальні принципи роботи з великими даними, аспекти вибору алгоритмів обробки даних і відповідних структур даних, технології розподілення і зберігання великих даних, технології SQL і NoSQL, графові бази даних, основні методі і технології аналізу тексту і, відповідно, способи і технології візуалізації даних.

Необхідний обсяг умінь для одержання позитивної оцінки.

Уміти застосовувати інструменти та моделі аналізу великих даних (у зв'язку з Data Science технологіями), будувати і обирати моделі даних, виконувати попередню обробку даних (очищення, інтеграція, перетворення даних та інше), використовувати базові методи машинного навчання, уміти обирати ефективний алгоритм / технологію обробки даних, а також структури для роботи з великими даними (локально і глобально), застосовувати методологію розподілення і зберігання великих даних, уміти використовувати сучасні бази даних (перш за все NoSQL і графові бази даних), проводити аналіз тексту, застосовувати класифікатори повідомлень, застосовувати методи аналізу даних високого рівня, а також технології візуалізації даних, виконувати дослідження реальних сховищ даних, презентувати результатів наукових досліджень; здійснювати дослідження з використанням цих ресурсів та технологій.

Критерії оцінювання знань та вмінь аспіранта для отримання заліку.

Задовільно, D, E (60-74). Мати мінімум знань і умінь: знати основні поняття, означення та терміни аналізу великих даних, вміти розв'язувати найпростіші практичні завдання. Відпрацювати всі практичні заняття, тестові завдання.

Добре, C (75-89). Твердо знати мінімум. Знати основні поняття, означення та терміни основних методів і технологій аналізу даних, працювати з сучасними базами даних Big Data; вміти розв'язувати практичні завдання з використанням методів Data Science. Вміти вибирати та використовувати потрібні програмні продукти для вирішення практичних завдань.

Відмінно, A, B (90-100). Знати всі теми. Знати основні поняття, означення та терміни аналізу даних та основних методів і технологій, вміти працювати з сучасними базами даних Big Data; вміти розв'язувати практичні завдання з поясненням та обґрунтуванням. Розв'язувати задачі машинного навчання, задачі прогнозування.

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
96 – 100	**A**	відмінно	зараховано
90-95	**B**		
75-89	**C**	добре	
66-74	**D**	задовільно	
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

8 МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА

8.1 Базова література

1. Davy Cielen, Arno D. B. Meysman, and Mohamed Ali Introducing Data Science: Big data, machine learning, and more, using Python tools. – Manning, 2016. – 320 p.
2. Виктор Майер-Шенбергер, Кеннет Кукьер Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. – Издательство Манн, Иванов и Фербер, 2013. – 240с.
3. Андреас Вайгенд BIG DATA. Вся технология в одной книге. – Издательство Эксмо, 2018. – 414с.
4. Брендан Тирни, Джон Келлехер Наука о данных. – Издательство Альпина Диджитал, 2020. – 175с.
5. Конспект лекцій з дисципліни.

8.2 Допоміжна література

6. Алексей Благирев, Наталья Хапаева Big data простым языком. – Издательство АСТ, 2019. – 153с.
7. Тим Филлипс Управление на основе данных. – Издательство Манн, Иванов и Фербер, 2017. – 117с.
8. Коул Нафлик Данные: визуализируй, расскажи, используй. – Издательство Манн, Иванов и Фербер, 2020. – 179с.
9. https://open.sap.com (Driving Business Results with Big Data)
10. https://hadoop.apache.org/

9 ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

1. Мова програмування Python.