

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій
Кафедра комп'ютерної інженерії та інформаційних систем



ЗАТВЕРДЖУЮ

Декан ФІТ

Тетяна ГОВОРУЩЕНКО.

05 вересня 2024 р.

СИЛАБУС

Навчальна дисципліна Аналітика великих даних
Рівень вищої освіти другий (магістерський)

Загальна інформація

Позиція	Зміст інформації
Викладач(і)	Кисіль Тетяна Миколаївна
Профайл викладача	http://kiis.khmnmu.edu.ua/personnel/kysil-tetyana-mykolayivna/
E-mail викладача(ів)	kysil_tanya@ukr.net
Контактний телефон	заповнюється за домовленістю
Сторінка дисципліни в ІСУ	https://msn.khmnmu.edu.ua/course/view.php?id=8241
Навчальний рік	2024-2025
Консультації	Очні: середа, 5-а пара, 1-113; онлайн: за необхідністю та попередньою домовленістю

Характеристика дисципліни

Форма здобуття освіти	Курс	Семестр	Загальне навантаження		Кількість годин						Форма семестрового контролю			
			Кредити ЄКТС	Години	Аудиторні заняття				Індивідуальна робота студента	Самостійна робота, в т.ч. ІРС	Курсовий проект	Курсова робота	Залік	Іспит
					Разом	Лекції	Лабораторні роботи	Практичні заняття						
ОД		непарний	8	240	85	34	34	17		155			+	
Разом ДФН			8	240	85	34	34	17		155			1	

Анотація дисципліни

Дисципліна "Аналітика великих даних" є однією із вибірових дисциплін підготовки магістрів.
Дисципліна викладається для здобувачів другого (магістерського) рівня вищої освіти денної форми навчання. При викладанні дисципліни використовуються активні і творчі форми проведення занять, зокрема, методи проблемного навчання.

Мета і завдання дисципліни

Метою і завданнями дисципліни є формування у студентів системи знань та практичних навичок в галузі використання методів та засобів обробки великих масивів даних; розвиток професійних умінь у сфері інженерії даних та знань щодо розроблення і підтримки комп'ютерних та інформаційних систем, які базуються на інтелектуальному аналізі великих даних.

Очікувані результати навчання.

Студент, який успішно завершив вивчення дисципліни, повинен: вміти розв'язувати задачі аналізу та синтезу засобів, характерних для спеціальності; провадити дослідницьку та/або інноваційну діяльність, здійснювати критичний аналіз, оцінювати та синтезувати нові складні ідеї під час аналізу великих даних в предметних областях, притаманних спеціальності; використовувати різні види та форми рухової активності для активного відпочинку та здорового способу життя.

Тематичний і календарний план вивчення дисципліни

№ тижня	Тема лекції*	Тема практичного заняття*	Тема лабораторної роботи*	Самостійна робота студентів		
				Зміст	Год.	Література
1	Big Data і Data Science: сутність та інструментарій	Застосування методів очищення даних	Застосування методів очищення даних.	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №1 та до практичної роботи №1.	17	[1-5]
2	Класифікація методів аналізу великих даних					
3	Метадані великих даних	Python: імпорт і попередня обробка даних	Python: імпорт і попередня обробка даних	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №2 та до практичної роботи №2.	17	[1-5]
4	Методи очищення даних					
5	OLAP-аналіз	Python: OLAP-аналіз	Python: OLAP-аналіз Класифікація (Дерево рішень)	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №3 та до практичної роботи №3.	17	[1-5]
6	Методи візуалізації великих даних					
7	Методи групування даних	Класифікація	Класифікація (Дерево рішень)	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №4 та до практичної роботи №4.	17	[1-5]
8	Методи групування даних: класифікація даних					
9	Методи групування даних: кластеризація даних	Кластерний аналіз	Кластерний аналіз	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №5 та до практичної роботи №5. Робота над курсовою роботою	17	[1-5]
10	Методи аналізу та моделювання залежностей					
11	Секвенційний аналіз	Асоціація	Асоціація .	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №6 та до практичної роботи №6.	17	[1-5]
12	Регресійний аналіз					
13	Методи аналізу динаміки	Застосування методів аналізу динаміки	Застосування методів аналізу динаміки	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №7 та до практичної	17	[1-5]
14	Ансамблі моделей					

				роботи №7.		
15	Нейронні мережі та їх застосування при аналізі великих даних	Штучний інтелект та нейромережі	Побудова найпростішої нейромережі	Опрацювання лекційного матеріалу. Підготовка до лабораторної роботи №8 та до практичної роботи №8.	17	[1-5]
16	Технології Text Mining					
17	Технології Map Reduce			Підготовка до контрольної роботи	19	[1-5]

Примітка: * Лекції і практичні заняття проводяться по дві години; послідовність проведення занять визначається розкладом (може не відповідати нумерованим тижням)

Політика дисципліни.

Організація освітнього процесу з дисципліни відповідає вимогам положень про організаційне і навчально-методичне забезпечення освітнього процесу, освітньої програми та навчальному плану. Студент зобов'язаний відвідувати лекції, практичні заняття згідно з розкладом, не запізнюватися на заняття, завдання виконувати відповідно до графіка. Пропущене практичне заняття студент зобов'язаний опрацювати самостійно у повному обсязі і відвітати перед викладачем не пізніше, ніж за тиждень до чергової атестації. До практичних занять студент має підготуватися за відповідною темою і проявляти активність.

Критерії оцінювання результатів навчання.

Поточний контроль здійснюється під час лекційних та практичних занять, а також у дні проведення контрольних заходів, встановлених робочим планом дисципліни. Семестровий контроль проводиться у формі заліку. При цьому при виведенні остаточної оцінки враховуються результати поточного контролю.

Перед вивченням дисципліни, як правило, проводиться вхідний контроль знань з дисциплін, що їй передують і забезпечують.

Кожний вид роботи з дисципліни оцінюється за чотирибальною шкалою. Семестрова підсумкова оцінка визначається як середньозважена з усіх видів навчальної роботи, виконаних і зданих позитивно з врахуванням коефіцієнта вагомості. Вагові коефіцієнти змінюються залежно від структури дисципліни і важливості окремих її видів робіт. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав залік, вважається невстигаючим.

При оцінюванні знань студентів використовуються різні засоби контролю, зокрема: усне опитування перед допуском до виконання практичної та лабораторної роботи – здійснюється на її початку; якість виконання, набуття теоретичних знань і практичних навичок перевіряється шляхом захисту кожної практичної та лабораторної роботи згідно з робочою програмою дисципліни і робочим навчальним планом.

Оцінка, яка виставляється за лабораторне заняття, складається з таких елементів: усне опитування студентів перед допуском до виконання роботи; знання теоретичного матеріалу з теми; якість оформлення звіту; вміння студента обґрунтувати прийняті конструктивні рішення; своєчасний захист роботи.

При оцінюванні знань студентів викладач керується такими критеріями.

Оцінку „зараховано”, за шкалою ECTS – А, отримує студент за глибоке і повне опанування змісту навчального матеріалу, в якому він легко орієнтується, понятійного апарату, за уміння зв'язувати теорію з практикою, вирішувати практичні завдання, висловлювати і обґрунтовувати свої судження. Така оцінка передбачає грамотний, логічний виклад відповіді (як в усній, так і в письмовій формі), якісне зовнішнє оформлення. Студент повинен набути практичних навичок із оцінювання критичних програмних систем.

Оцінка „зараховано”, за шкалою ECTS – А, виставляється студенту, який глибоко засвоїв основні принципи оцінювання критичних програмних систем та вмів раціонально застосовувати їх при проектуванні програмних систем. Студент не повинен вагатися при видозміні запитання, повинен робити детальні та узагальнюючі висновки.

Оцінку „зараховано”, за шкалою ECTS – В, отримує студент за повне засвоєння навчального матеріалу, володіння понятійним апаратом, орієнтування у вивченому матеріалі, свідоме використання знань для вирішення практичних завдань, грамотний виклад відповіді, але у змісті і формі відповіді мали місце окремі неточності (похибки), нечіткі формулювання закономірностей тощо. Відповідь студента повинна будуватись на основі самостійного мислення.

Оцінку „зараховано”, за шкалою ECTS – С, отримує студент за правильну відповідь з однією-двома суттєвими помилками.

Оцінки „зараховано”, за шкалою ECTS – D, заслуговує студент, який виявив знання основного навчально-програмного матеріалу в обсязі, необхідному для подальшого навчання та практичної діяльності за професією, що

справляється з виконанням практичних завдань, передбачених програмою. Як правило, відповідь студента будується на рівні репродуктивного мислення, студент слабо знає структуру курсу, допускає помилки у відповіді, засвоїв і набув практичних навичок з оцінювання критичних програмних систем, але припустився неточностей. Вагається при відповіді на видозмінене запитання, разом з тим студент володіє знаннями, що дозволяють йому під керівництвом викладача усунути неточності у відповіді.

Оцінки „зараховано”, за шкалою ECTS – E, заслуговує студент за неповне опанування програмного матеріалу, але отримані знання і набуті практичні навички із оцінювання якості та експертизи ПЗ.

Оцінка „незараховано”, за шкалою ECTS – FX, виставляється, коли студент має розрізнені, безсистемні знання, не вміє виділяти головне і другорядне, допускається помилок у визначенні понять, перекручує їх зміст, хаотично і невпевнено викладає матеріал, не може використовувати знання при вирішенні практичних завдань. Як правило, оцінка "незараховано", за шкалою ECTS – F, виставляється студенту, який не може продовжити навчання без додаткових знань з курсу.

На основі результатів поточного контролю виставляється підсумкова семестрова оцінка. На основі аналізу контролю знань викладач удосконалює курс лекцій, звертаючи особливу увагу на ті розділи, чи теми, з яких було найбільше неточних відповідей, що свідчить про методичні чи інші недоліки при висвітленні вказаних тем або розділів.

Аналогічно вносяться корективи в методичні посібники для практичних та лабораторних робіт, детальніше розглядаються принципові питання при виконанні практичних та лабораторних робіт та їх захисті.

Структурування дисципліни за видами робіт і оцінювання результатів навчання студентів денної форми навчання у семестрі за ваговими коефіцієнтами

Аудиторна робота						Самостійна, індивідуальна робота	Залік
II семестр							
Лабораторні роботи №:			Практичні роботи №			Контроль	
1	2	3	1	2	3	Контрольна робота	
ВК:						0,8	
						0,2	

Умовні позначення: Т – тема дисципліни; ВК – ваговий коефіцієнт.

Співвідношення вітчизняної шкали оцінювання і шкали оцінювання ECTS

Оцінка ECTS	Інституційна шкала балів	Інституційна оцінка	Критерії оцінювання	
A	4,75–5,00	5	Зараховано	Відмінно – глибоке і повне опанування навчального матеріалу і виявлення відповідних умінь та навичок
B	4,25–4,74	4		Добре – повне знання навчального матеріалу з кількома незначними помилками
C	3,75–4,24			Добре – в загальному правильна відповідь з двома-трьома суттєвими помилками
D	3,25–3,74	3		Задовільно – неповне опанування програмного матеріалу, але достатнє для практичної діяльності за професією
E	3,00–3,24	3		Задовільно – неповне опанування програмного матеріалу, що задовольняє мінімальні критерії оцінювання
FX	2,00–2,99	2	Незараховано	Незадовільно – безсистемність одержаних знань і неможливість продовжити навчання без додаткових знань з дисципліни
F	0,00–1,99	2		Незадовільно – необхідна серйозна подальша робота і повторне вивчення дисципліни

Питання для підсумкового контролю з дисципліни

1. Що таке Pandas? Призначення та застосування.
2. Назвіть основні класи методів обробки великих даних. Дайте стисло характеристику.
3. Суть Text Mining та його завдання.
4. Суть Map Reduce та його завдання.
5. Дайте визначення інтелектуального аналізу даних.

6. Що таке розвідувальний аналіз.
7. Що таке якість даних?
8. Які цілі підготовки даних до аналізу?
9. Який атрибут називається цільовим?
10. Що таке значимий і незначний атрибут?
11. Що таке відбір атрибутів?
12. У чому полягає завдання класифікації? Наведіть практичний приклад.
13. Що таке навчання з учителем і без учителя?
14. Завдання класифікації є описовим або прогнозуючим і чому?
15. Навіщо потрібна навчальна і тестова вибірки для вирішення завдання класифікації?
16. Які існують підходи для поділу вихідної вибірки на навчальну і тестову?
17. Метод класифікації 0R.
18. Метод класифікації за одним правилом 1R.
19. Метод класифікації PRISM.
20. Наївна Баєсова класифікація NaiveBayes.
21. Метод побудови дерев рішень CART
22. Метод побудови дерев рішень Id3 та C4.5.
23. Метод опорних векторів SMO.
24. Метод k найближчих сусідів kNN.
25. Як оцінити якість побудованої моделі класифікації?
26. Що таке матриця помилок? Як її інтерпретувати?
27. Що означають параметри чутливість, специфічність, точність? Як їх розрахувати?
28. Що таке параметр Каппа? Що він показує?
29. Що таке аналіз витрати-вигоди?
30. Як порівняти роботу двох класифікаторів?
31. Що таке ансамблі класифікаторів і адаптивний бустінг? Для чого вони застосовуються?
32. У чому полягає завдання регресії? Наведіть приклад.
33. Чим завдання регресії схоже і чим відрізняється від завдання класифікації?
34. Завдання регресії є описовим або прогнозуючим і чому?
35. Метод лінійної регресії.
36. Метод опорних векторів для вирішення завдання регресії.
37. Метод найближчих сусідів для вирішення завдання регресії.
38. Як оцінити якість побудованої моделі для завдання регресії?
39. Прогнозування часових рядів.
40. У чому полягає завдання кластеризації? Наведіть практичний приклад.
41. Що таке навчання з учителем і без учителя? До якого типу відноситься завдання кластеризації?
42. Завдання кластеризації є описовим або прогнозуючим і чому?
43. Чим визначається «схожість» об'єктів при вирішенні задачі кластеризації?
44. Що таке однорівнева і ієрархічна кластеризація?
45. Що таке чітка і нечітка кластеризація?
46. Які є підходи до розрахунку відстані між кластерами?
47. Алгомеративная і дівізімная ієрархічна кластеризація?
48. Метод кластеризації k-середніх.
49. Метод нечіткої кластеризації fuzzy c-means.
50. Метод кластеризації k-медоїд.
51. Ієрархічні методи кластеризації.
52. Імовірнісний метод кластеризації
53. Метод кластеризації COBWEB.
54. Метод кластеризації заснований на щільності розташування об'єктів DBSCAN.
55. Як оцінити якість побудованої моделі для завдання кластеризації?
56. У чому полягає завдання пошуку асоціативних правил? Наведіть практичний приклад.
57. Що таке частий набір?
58. Що таке сильне асоціативне правило?
59. З яких двох кроків складається пошук асоціативних правил?
60. У чому полягає принцип Apriori?
61. Як формуються правила зі знайдених частих наборів?
62. Опишіть алгоритм Apriori

74. Які переваги характерні для використання байесовських мереж?
75. Які задачі вирішуються за допомогою методу опорних векторів?
76. Що являє собою група синапсів нейрону?
77. Визначите головну функцію штучного нейрона.
78. У чому полягає навчання мереж, що самоорганізуються?
79. Які існують традиційні методи візуалізації?
80. Назвіть характеристики однієї з основних тенденцій в області візуалізації.
81. Що означає інтегрованість сховища даних?
82. Назвіть основні концепції сховища даних.
83. Якщо набір даних упорядкований та у ньому присутня сезонна або циклічна компонента, то яку мінімальну кількість даних необхідно мати для можливості аналізу?
84. Що повинна робити якісна програма очищення даних?
85. Які характеристики повинна мати якісна програма очищення даних?
86. На якому етапі перетинається робота фахівця предметної області й фахівця з видобутку даних?
87. Назвіть характеристики, властиві SAS Enterprise Miner.
88. Рішення яких задач передбачають алгоритми аналізу даних в PolyAnalyst?
89. Чи існує необхідність тимчасового або постійного копіювання даних для аналізу в системі KXEN?
90. Який компонент KXEN дозволяє виявити природні групи (кластери) у наборі даних?
91. Назвіть слабкі сторони використання готового програмного забезпечення Data Mining.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Черняк О. І. Інтелектуальний аналіз даних : підручник / О. І. Черняк, П. В. Захарченко. – К.: Знання, 2014. – 599 с.
2. Hans Weber. Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data and Machine Learning. Режим доступу: <https://www.amazon.com/dp/B08FP9YZXN?tag=uid10-20>.
3. Jay M. Patel. Getting Structured Data from the Internet: Running Web Crawlers/Scrapers on a Big Data Production Scale 1st ed. Edition. Режим доступу: <https://www.amazon.com/dp/1484265750?tag=uid10-20>.
4. Dursun Delen. Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd Edition (FT Press Analytics) 2nd Edition. Режим доступу: <https://www.amazon.com/dp/0136738516?tag=uid10-20>
5. Dinesh Sachdev. Thinking Big: Developers Guide for Big Data Engineering & Analytics. Режим доступу: <https://www.amazon.com/dp/B08TQD9GWQ?tag=uid10-20>

ІНФОРМАЦІЙНІ РЕСУРСИ

1. Модульне середовище для навчання MOODLE. Доступ до ресурсу: <https://msn.khnu.km.ua>.
2. Електронна бібліотека університету. Доступ до ресурсу: http://lib.khnu.km.ua/asp/php_f/page_lib.php.

Розробник:



к. ф. - м. н., доц. Тетяна КИСІЛЬ

Погоджено:
Зав. каф. КІС:



к. т. н., доц. Ірина ЗАСОРНОВА