

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВОДНОГО ГОСПОДАРСТВА ТА ПРИРОДОКОРИСТУВАННЯ

Навчально-науковий інститут кібернетики, інформаційних технологій та інженерії

04-01-128S

СИЛАБУС SYLLABUS	Актуальні проблеми "Data Mining"	
	Actual problems of "Data Mining"	
Шифр за ОП Code in Degree Programme	ОК 9	
Освітній рівень Level of Education	Магістерський (другий)	
	Master's (second)	
Галузь знань Field of Knowledge	12	Інформаційні технології Information technologies
Спеціальність Field of Study	122	Комп'ютерні науки Computer Science
Освітня програма Degree Programme	Прикладна інформатика	
	Applied Informatics	

РІВНЕ – 2024

Силабус навчальної дисципліни «Актуальні проблеми "Data Mining"» для здобувачів вищої освіти ступеня «магістр», які навчаються за освітньо-професійною програмою «Прикладна інформатика» спеціальності 122 «Комп'ютерні науки» денної форми навчання. Рівне. НУВГП. 2024. 10 стор.

ОП на сайті університету: <https://ep3.nuwm.edu.ua/28870/>

Розробник силабусу:

Демчук Олена Станіславівна, к.т.н., доцент кафедри комп'ютерних наук та прикладної математики;

Силабус схвалений на засіданні кафедри комп'ютерних наук та прикладної математики

Протокол № 6 від "12" грудня 2023 року

Завідувач кафедри: *Турбал Ю.В., д.т.н., професор.*


Керівник (гарант) ОП: *Мічута О.Р., к.т.н., доцент*

Схвалено науково-методичною радою з якості ННІКІТІ
Протокол №5 від "26" лютого 2024 року

Голова науково-методичної ради з якості ННІКІТІ:

_____ Мартинюк П. М., д.т.н., професор

© НУВГП, 2024

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ	
«Актуальні проблеми "Data Mining"»	
ЗАГАЛЬНА ІНФОРМАЦІЯ	
Ступінь вищої освіти	<i>магістр</i>
Освітня програма	Прикладна інформатика
Спеціальність	122 «Комп'ютерні науки»
Рік навчання, семестр	<i>1 рік, 2 семестр</i>
Кількість кредитів	4
Лекції:	<i>20 годин</i>
Лабораторні заняття:	<i>20 годин</i>
Самостійна робота:	<i>80 годин</i>
Курсова робота:	<i>Не передбачено</i>
Форма навчання	<i>денна</i>
Форма підсумкового контролю	<i>залік</i>
Мова викладання	<i>українська</i>
ІНФОРМАЦІЯ ПРО ВИКЛАДАЧА	
Лектор	 <p><i>Демчук Олена Станіславівна, кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук та прикладної математики</i></p>
Вікіситет	http://wiki.nuwm.edu.ua/index.php/Демчук_Олена_Станіславівна
ORCID	http://orcid.org/0000-0002-8318-5009
Як комунікувати	o.s.demchuk@nuwm.edu.ua Актуальні оголошення на сторінці дисципліни в системі MOODLE https://exam.nuwm.edu.ua/course/view.php?id=2801
ІНФОРМАЦІЯ ПРО ДИСЦИПЛІНУ	
Мета і завдання	

Мета: вивчення основних понять та актуальних проблем Data Mining; оволодіння методами, моделями, алгоритмами та програмними засобами Data Mining.

Завдання: сформувані у студентів теоретичні знання та практичні вміння у сфері видобутку та інтелектуального аналізу даних, що виникають при дослідженні різноманітних інформаційних, технічних, економічних, екологічних, соціальних та інших систем.

У результаті вивчення навчальної дисципліни студенти будуть **знати:**

- основні поняття та актуальні проблеми Data Mining;
- сучасні методи і моделі Data Mining;
- програмні засоби Data Mining.

вміти:

- вибирати методи і моделі Data Mining при вирішенні конкретних практичних задач;
- використовувати сучасні програмні засоби Data Mining;
- аналізувати та інтерпретувати отримані результати при вирішенні прикладних задач.

Посилання на розміщення освітнього компонента на навчальній платформі Moodle, на платформі освітніх програм та їхніх освітніх компонентів

<https://exam.nuwm.edu.ua/course/view.php?id=2801>

<https://nuwm.edu.ua/dystsypliny>

Передумови вивчення навчальної дисципліни

Дисципліни, що передують вивченню дисципліни «Актуальні проблеми "Data Mining"»: «Теорія комп'ютерних систем та методологія їх проєктування», «Системи керування базами даних та знань», «Сучасні технології об'єктно-орієнтованого програмування», «Розробка проєктів засобами C#.NET», «Сучасні методи розпізнавання образів».

Результати вивчення дисципліни стануть у нагоді при підготовці кваліфікаційної магістерської роботи.

Компетентності

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ФК1. Усвідомлення теоретичних засад комп'ютерних наук.

ФК2. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

ФК4. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття рішень.

Програмні результати навчання (ПРН)

ПРН1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.

ПРН20. Здійснювати математичне та комп'ютерне моделювання об'єктів, процесів і явищ, розробляти алгоритми чисельного розв'язання задач відповідних задач.

Структура та зміст навчальної дисципліни

ЗМІСТОВИЙ МОДУЛЬ 1. Основні поняття Data Mining. Задачі Data Mining. Методи і підходи до їх розв'язання. Факторний аналіз. Дерева рішень. Штучні нейронні мережі

Тема 1. Основні поняття, сфери застосування, методи і стадії Data Mining. Процес Data Mining
Поняття Data Mining. Предмет, історія виникнення Data Mining. Особливості технології Data Mining та її відмінності від інших методів аналізу даних. Застосування Data Mining у бізнесі, торгівлі, банківській справі, страхуванні, медицині, науці тощо. Задачі Data Mining. Статистичні та кібернетичні методи Data Mining. Стадії Data Mining. Процес Data Mining. Підготовчі етапи процесу Data Mining. Дублювання даних. Очищення даних. Етапи очищення даних.

Тема 2. Статистичні методи Data Mining. Факторний аналіз. Метод головних компонент
Класичні статистичні методи Data Mining. Ідея факторного аналізу. Модель компонентного аналізу. Метод головних компонент та сфери його застосування.

Тема 3. Задачі класифікації та кластеризації, прогнозування та візуалізації
Задачі та види класифікації. Методи, що застосовуються для розв'язання задач класифікації. Задача кластеризації. Методи кластерного аналізу. Ієрархічний кластерний аналіз. Алгоритми неієрархічної кластеризації. Задача пошуку асоціативних правил. Характеристика і методи пошуку асоціативних правил. Прогнозування і часові ряди. Тренд, сезонність і цикл. Види помилок та прогнозів. Візуалізація інструментів Data Mining. Методи візуалізації. Принципи компонування візуальних засобів.

Тема 4. Дерева рішень для розв'язання задач класифікації та прогнозування
Метод дерев рішень та його переваги. Алгоритми дерев рішень. Метод опорних векторів. Лінійний SVM. Метод «найближчого сусіда». Байєсівська класифікація.

Тема 5. Застосування штучних нейронних мереж для розв'язання задач Data Mining
Поняття штучної нейронної мережі. Модель нейрону. Архітектура НМ. Основні парадигми навчання НМ. Нейронні мережі із зворотним розповсюдженням помилок. Закон навчання Хебба. Конкурентне навчання. Карти Кохонена. Нейронні мережі Хопфільда та Хемінга.

ЗМІСТОВИЙ МОДУЛЬ 2. Нечіткі експертні системи. Генетичні та інші еволюційні алгоритми

Тема 6. Методи прийняття рішень на основі нечітких множин та нечіткої логіки. Нечіткі експертні системи

Основи теорії нечітких множин. Методи нечіткої логіки. Нечіткі експертні системи. Етапи побудови нечіткого логічного виводу. Нечіткі алгоритми Мамдані, Цукамото, Такагі-Сугено та інші. Нейро-нечіткі технології.

Тема 7. Генетичні алгоритми та їх застосування. Еволюційні алгоритми
Основні поняття генетичних алгоритмів. Класичний генетичний алгоритм та його модифікації. Види еволюційних алгоритмів: еволюційні стратегії, еволюційне програмування, генетичні алгоритми, генетичне програмування. Гібридні системи.

Розподіл годин за темами змістових модулів

Лекції	Год	Лабораторні роботи	Год	Сам. робота (год.)	Всього (год.)	Навчальні матеріали
ЗМІСТОВИЙ МОДУЛЬ 1. Основні поняття Data Mining. Задачі Data Mining. Методи і підходи до їх розв'язання. Факторний аналіз. Дерева рішень. Штучні нейронні мережі						

Тема 1. Основні поняття, сфери застосування, методи і стадії Data Mining. Процес Data Mining	3	-	0	8	11	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12,13]
Тема 2. Статистичні методи Data Mining. Факторний аналіз. Метод головних компонент	1	ЛР-1. Факторний аналіз. Метод головних компонент	2	8	11	[1, 4, 5, 7,13]
Тема 3. Задачі класифікації та кластеризації, прогнозування та візуалізації	1	-	0	8	9	[1, 2, 3, 4, 5, 6, 7, 8, 13]
Тема 4. Дерева рішень для розв'язання задач класифікації та прогнозування	1	ЛР-2. Дерева рішень для розв'язання задач класифікації	2	8	11	[1, 3, 4, 7, 10, 13]
Тема 5. Застосування штучних нейронних мереж (ШНМ) для розв'язання задач Data Mining	6	ЛР-3. Штучні нейронні мережі для розв'язання задач класифікації	1	20	32	[1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13]
		ЛР-4. Штучні нейронні мережі для розв'язання задач регресії	1			
		ЛР-5. Моделювання і прогнозування часових рядів за допомогою ШНМ	1			
		ЛР-6. Класифікація для даних часового ряду за допомогою ШНМ	1			
		ЛР-7. Штучні нейронні мережі для розв'язання задач кластеризації	2			
МК-1	-		2		2	
За змістовим модулем 1	12		12	52	76	
ЗМІСТОВИЙ МОДУЛЬ 2 Нечіткі експертні системи. Генетичні та інші еволюційні алгоритми						
Тема 6. Методи прийняття рішень на основі нечітких множин та нечіткої логіки. Нечіткі експертні системи	5	ЛР-8. Нечіткі експертні системи. Алгоритм Мамдані	2	18	27	[1, 2, 3, 4, 7, 10, 13]
		ЛР-9. Нечіткі експертні системи. Алгоритм Сугено	2			
Тема 7. Генетичні алгоритми та їх застосування. Еволюційні алгоритми	3	ЛР-10. Розв'язання задач оптимізації за допомогою генетичних алгоритмів	2	10	15	[1, 2, 4, 6, 7, 10, 12, 13]
МК-2	0		2		2	
За змістовим модулем 2	8		8	28	44	
Разом	20		20	80	120	

Відповідність програмних результатів навчання навчальним матеріалам

Теми	ПР 01	ПР 20
Тема 1		
Тема 2		
Тема 3		
Тема 4		
Тема 5		
Тема 6		
Тема 7		

Форми та методи навчання

Методи навчання: інформаційно-ілюстративний, презентації, тренінги, обговорення, ситуаційні дослідження, командна робота.

Технології навчання: навчання у співробітництві, мозковий штурм, дослідницьке навчання.

Інструменти, обладнання, програмне забезпечення

-технічні засоби навчання: мультимедійне обладнання, ноутбук;

-програмне забезпечення: MS Windows, доступ до Інтернет;

-програмне забезпечення: система дистанційного навчання Moodle.

Порядок оцінювання програмних результатів навчання

Поточний контроль здійснюється за виконанням завдань лабораторних робіт; за підсумками роботи під час лекційних занять. Підсумковий контроль відбувається у вигляді проходження двох модульних контролів у формі тестування на університетській платформі MOODLE.

У тесті передбачено 27 запитань різної складності:

- рівень 1 – 20 запитання по 0,6 бали (12 балів),
- рівень 2 – 5 запитань по 1 балу (5 балів),
- рівень 3 – 2 запитання по 1,5 бали (3 бали).

Усього – 20 балів.

Усі форми контролю включено до 100-бальної шкали оцінювання. За конкретні пропозиції з удосконалення змісту навчальної дисципліни студентам також можуть бути зараховані додаткові бали (до 3 балів).

Шкала оцінювання навчальних досягнень студентів

Вид заняття	Бали
1. Поточна складова оцінювання	
1.1. Лабораторна робота 1. Факторний аналіз. Метод головних компонент	5
1.2. Лабораторна робота 2. Дерева рішень для розв'язання задач класифікації	5
1.3. Лабораторна робота 3. Штучні нейронні мережі для розв'язання задач класифікації	4
1.4. Лабораторна робота 4. Штучні нейронні мережі для розв'язання задач регресії	4
1.5. Лабораторна робота 5. Моделювання і прогнозування часових рядів за допомогою ШНМ	4
1.6. Лабораторна робота 6. Класифікація для даних часового ряду за допомогою ШНМ	4
1.7. Лабораторна робота 7. Штучні нейронні мережі для розв'язання задач кластеризації	4
1.8. Лабораторна робота 8. Нечіткі експертні системи. Алгоритм Мамдані	10
1.9. Лабораторна робота 9. Нечіткі експертні системи. Алгоритм Сугено	10
1.10. Лабораторна робота 10. Розв'язання задач оптимізації за допомогою генетичних алгоритмів	10
Всього поточна складова оцінювання:	60
2. Модульна складова оцінювання	
2.1. Модульний контроль №1	20
2.2. Модульний контроль №2	20
Всього підсумкова складова оцінювання:	40
Разом:	100

Рекомендована література

Основна

1. Марченко О. О., Россада Т. В. Актуальні проблеми Data Mining : навч. посіб. для студентів факультету комп'ютерних наук та кібернетики. Київ : КНУ ім. Т. Шевченка, 2017. 150 с.
2. Кононова К. Ю. Машинне навчання : методи та моделі : підручник. Харків : ХНУ імені В. Н. Каразіна, 2020. 301 с.
3. Дранишников Л.В. Інтелектуальні методи в управлінні: навчальний посібник – Кам'янське: ДДТУ, 2018. – 416 с.
4. Черняк О. І. Інтелектуальний аналіз даних : підручник. Київ : Знання, 2014. 599 с.
5. Грицюк П. М., Остапчук О.П. Аналіз даних: навчальний посібник – Рівне: НУВГП, 2008. – 218 с.
6. Зайченко Ю.П. Основи проектування інтелектуальних систем. Навчальний посібник. – К.: Видавничий Дім "Слово", 2004. – 352 с.
7. Бахрушин В. Є. Методи аналізу даних : навч. посіб. для студентів. Запоріжжя : КПУ, 2011. 268 с.
8. Бідюк П. І, Романенко В. Д., Тимощук О. Л. Аналіз часових рядів : підручник. Київ : ВПК «Політехніка», 2013. 599 с.

Допоміжна

9. Тимощук П.В. Штучні нейронні мережі: навч. посібник. – Львів: Видавництво Львівської політехніки, 2011. – 444 с.
10. Бідюк П. І., Гожий О. П. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень. Київ : ВПК «Політехніка», 2010, 335 с.
11. Шаховська Н. Б. Програмне та алгоритмічне забезпечення сховищ та просторів даних : монографія. Львів : Видавництво Львівської політехніки, 2010. 194 с.
12. Кононюк А.Ю. Нейронні мережі і генетичні алгоритми / А.Ю. Кононюк. – К.: ПП Корнійчук. – 2008. – 160 с.
13. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посіб. – К.: КНЕУ, 2007. – 376 с.

Інформаційні ресурси в Інтернет

14. COCOMO II Model [Електрон. ресурс]. – Режим доступу: http://sunset.usc.edu/csse/research/COCOMOII/cocomo_main.html
15. Softstar Systems. COCOMO and COSYSMO Estimation Tools [Електрон. ресурс]. – Режим доступу: <http://www.softstarsystems.com>

Поєднання навчання та досліджень

Здобувачі мають можливість додатково отримати бали за виконання індивідуальних завдань дослідницького характеру (до 3 балів), можуть бути долучені до написання та опублікування наукових статей з тематики навчальної дисципліни (до 5 балів), участі в науково-практичних конференціях інформаційно-педагогічного спрямування (до 5 балів).

ПОЛІТИКИ ВИКЛАДАННЯ ТА НАВЧАННЯ

Перелік соціальних, «м'яких» навичок (soft skills)

Комунікативність; вміння чітко формулювати свою думку; навички колективної роботи, навички ефективного мислення; вміння сприймати конструктивну критику; здатність до саморозвитку; стресостійкість та інші.

Дедлайни та перескладання

Поточні терміни захисту лабораторних робіт становлять два тижні після проведення заняття. Крайні терміни захисту лабораторних робіт регламентується останнім тижнем перед початком екзаменаційної сесії. У разі невиконання студентом вимог щодо поточного оцінювання протягом семестру (невчасне виконання) завдання) оцінку може бути знижено в межах 10%. Ліквідація академічної заборгованості здійснюється згідно з «Порядком ліквідації академічних заборгованостей у НУВГП», <https://ep3.nuwm.edu.ua/25072/>. За цим документом реалізується право студента на повторне проходження навчальної практики. Оголошення стосовно дедлайнів здачі та перездачі оприлюднюються на сторінці MOODLE <https://exam.nuwm.edu.ua/course/view.php?id=3623>

Неформальна та інформальна освіта

Студенти мають право на перезарахування результатів навчання, набутих у неформальній та інформальній освіті (<https://ep3.nuwm.edu.ua/28363/>). Студенти можуть самостійно на платформах Prometheus, Coursera, edEx, edEra, Future Learn опановувати матеріал для перезарахування результатів навчання. При цьому важливо, щоб знання та навички, що формуються під час проходження певного онлайн-курсу чи його частин, мали зв'язок з очікуваними програмними результатами навчальної дисципліни та перевірялись в підсумковому оцінюванні.
Важливо! Перед початком проходження обраних курсів необхідно отримати згоду викладача.

Правила академічної доброчесності

У разі виявлення копіювання результатів виконання завдань студенту завдання не зараховується. Студент повторно отримує завдання і виконує його самостійно. Документи стосовно академічної доброчесності (про плагіат, порядок здачі звіту, кодекс честі студентів, документи Національного агентства стосовно доброчесності) наведені на сторінці НУВГП <http://nuwm.edu.ua/sp/akademichna-dobrochesnisti>

Вимоги до відвідування

- Заняття відбуваються згідно розкладу <https://desk.nuwm.edu.ua/cgi-bin/timetable.cgi> офлайн або онлайн за допомогою Google Meet за лінком: <https://meet.google.com/>
- Консультації проводяться за потреби в режимі онлайн за допомогою Google Meet у домовлений час зі студентами.
- Здобувачі можуть на заняттях використовувати мобільні телефони та ноутбуки, але виключно в навчальних цілях.
- Студенту не дозволяється пропускати заняття без поважних причин.
- За наявності об'єктивних причин пропуску занять, студенти можуть самостійно ознайомитися з теоретичним матеріалом на платформі MOODLE <https://exam.nuwm.edu.ua/course/view.php?id=3623>.

Автор

О.С. Демчук

к.т.н., доцент кафедри комп'ютерних наук
та прикладної математики

Автор
Доцент

Олена ДЕМЧУК

Затверджено

Проректор з науково-педагогічної та
навчальної роботи

Валерій СОРОКА



документ підписаний КЕП
Номер документа СИЛ №627
Підписувач Сорока Валерій Степанович
Підписувач (дані КЕП):
Сертифікат 58E2D9E7F900307B04000000807E2D0054327D00