

Міністерство освіти і науки України
Харківський національний університет імені В.Н. Каразіна
Кафедра теоретичної та прикладної системотехніки

“ЗАТВЕРДЖУЮ”

В.о. декана факультету комп'ютерних наук


Світлана КОЛОВАНОВА
“ 30 ” червня 2023 р.



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Комп'ютерні інформаційні технології Data Stream Mining

рівень вищої освіти другий (магістерський)

спеціальність 123 Комп'ютерна інженерія

освітня програма Комп'ютерна інженерія

вид дисципліни за вибором

факультет комп'ютерних наук

2023 / 2024 навчальний рік

Програму рекомендовано до затвердження вченою радою факультету комп'ютерних наук

«29» червня 2023 року, протокол № 14

РОЗРОБНИКИ ПРОГРАМИ:

кандидат технічних наук, доцент кафедри теоретичної та прикладної системотехніки

Бакуменко Ніна Станіславівна;

кандидат технічних наук, доцент кафедри теоретичної та прикладної системотехніки

Стрілець Вікторія Євгенівна

Програму схвалено на засіданні кафедри теоретичної та прикладної системотехніки

«08» червня 2023 року, протокол № 13

Завідувач кафедри теоретичної та прикладної системотехніки


Сергій ШМАТКОВ.

Програму погоджено з гарантом освітньої програми «Комп'ютерна інженерія»

Гарант освітньої програми «Комп'ютерна інженерія»


Олен ТОЛСТОЛУЗЬКА

Програму погоджено методичною комісією факультету комп'ютерних наук

«21» червня 2023 року, протокол № 12

Голова методичної комісії факультету комп'ютерних наук


Лариса ВАСИЛЬСВА

ВСТУП

Програма навчальної дисципліни «Комп'ютерні інформаційні технології Data Stream Mining» розроблена відповідно до освітньо-професійної програми підготовки другого (магістерського) рівня спеціальності 123 «Комп'ютерна інженерія» освітньої програми «Комп'ютерна інженерія».

1. Опис навчальної дисципліни

1.1. **Метою** викладання навчальної дисципліни є:

засвоєння студентами основ наукових і математичних положень, що лежать в основі функціонування програмних засобів, систем для обробки великих даних, методології машинного навчання для дослідження складних ієрархічних багаторівневих систем (СІБС) та процесів, управління складними комп'ютерними системами; вироблення навичок з адаптації стандартних алгоритмів до нових – чисельних рішень складних прикладних задач, а також придбання знань про пакети прикладних програм спеціального призначення.

Об'єктом вивчення дисципліни «Інтелектуальні комп'ютерно-інтегровані технології управління виробничими процесами» є методи наукових досліджень, організації та планування експерименту, збирання даних та моделювання в комп'ютерних системах, а саме – методи аналізу стохастичних процесів з використанням сучасних моделей; методи планування і виконання комп'ютерних експериментів та управління ними, сучасна методологія машинного навчання дослідження СІБС та процесів, управління складними комп'ютерними системами, у якій розробляються методи й алгоритми, які регуляризують, в задачах аналізу, інтелектуальної обробки даних і синтезу СІБС та процесів, управління складними комп'ютерними системами; а також шляхи використання для цієї мети сучасних комп'ютерних систем, спеціалізованих пакетів прикладних програм.

Предметом вивчення є методи машинного навчання в задачах аналізу, інтелектуальної обробки даних і синтезу складних ієрархічних багаторівневих систем (СІБС) та процесів, управління складними комп'ютерними системами, оцінки їх ефективності та ін., для рішення яких розробляється математичне забезпечення комп'ютерних систем, а також використовуються спеціалізовані пакети прикладних програм.

1.2. **Основними завданнями вивчення навчальної дисципліни** є вивчення і набуття навичок застосування:

- математичних моделей і методів кластеризації даних (стратифікація прецедентів на підгрупи (класи));
- системних математичних моделей розпізнавання станів систем і процесів на основі даних моніторингу;
- прямої і зворотної задачі розрахунку розмірних конструкторських ланцюгів і методи їх вирішення;
- математичних алгоритмів для вирішення задач робастного оцінювання;
- основних характеристики та принципів роботи з спеціалізованими пакетами прикладних програм (ППП) для побудови моделей процесів управління нелінійними динамічними системами на основі статистичних даних про змінний стану;
- основних характеристик та принципів роботи з спеціалізованими ППП для вирішення завдань РОП&ІД систем і процесів.

В ході вивчення дисципліни у студента повинні формуватися такі компетентності.

Загальні компетентності (ЗК)

ЗК01. Здатність генерувати нові ідеї (креативність).

ЗК03. Здатність до абстрактного мислення, аналізу та синтезу.

Спеціальні (фахові, предметні) компетентності (ФК)

ФК01. Здатність здійснювати автоматизацію складних технологічних об'єктів та комплексів, створювати кіберфізичні системи на основі інтелектуальних методів управління та цифрових технологій з використанням баз даних, баз знань, методів штучного інтелекту, робототехнічних та інтелектуальних мехатронних пристроїв.

ФК04. Здатність аналізувати виробничо-технологічні системи і комплекси як об'єкти автоматизації, визначати способи та стратегії їх автоматизації та цифрової трансформації.

ФК07. Здатність застосовувати спеціалізоване програмне забезпечення та цифрові технології для розв'язання складних задач і проблем автоматизації та комп'ютерно-інтегрованих технологій.

ФК08. Здатність розробляти функціональну, технічну та інформаційну структуру комп'ютерно-інтегрованих систем управління організаційно-технологічними комплексами із застосуванням мережевих та інформаційних технологій, програмно-технічних керуючих комплексів, промислових контролерів, мехатронних компонентів, робототехнічних пристроїв та засобів людино-машинного інтерфейсу.

1.3. Кількість кредитів – 6

1.4. Загальна кількість годин – 180

1.5. Характеристика навчальної дисципліни	
За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
2-й	2-й
Лекції	
24 год.	год.
Практичні, семінарські заняття	
	год.
Лабораторні заняття	
24 год.	год.
Самостійна робота	
112 год.	год.
Індивідуальні завдання	
20 год.	

1.6. Відповідно до вимог освітньо-кваліфікаційного рівня підготовки за результатами вивчення дисципліни студенти повинні –

знати:

- наукові і математичні положення, що лежать в основі функціонування програмних засобів, систем для оброблення великих даних;
- методи наукових досліджень, організації та планування експерименту, збирання даних та моделювання в комп'ютерних системах, а саме – методи аналізу стохастичних процесів з використанням сучасних моделей; методи планування і виконання комп'ютерних експериментів та управління ними та ін.
- методи машинного навчання для розв'язання задач аналізу, інтелектуальної обробки даних і синтезу складних ієрархічних багаторівневих систем СІБС, управлінню складними комп'ютерними системами;

уміти:

- здійснювати вибір методів машинного навчання для розв'язання задач аналізу, інтелектуальної обробки даних і синтезу СІБС, управління складними комп'ютерними системами;
- проводити верифікацію математичних методів, оцінку якості математичних методів на основі існуючих критеріїв;
- вирішувати задачі чисельного характеру аналізу, інтелектуальної обробки даних і синтезу СІБС, управління складними комп'ютерними системами з застосуванням спеціалізованих пакетів;
- створювати дослідницькі групи для проведення аналізу та обробки великих масивів даних;
- створювати та використовувати нове програмне забезпечення для аналізу та обробки великих масивів даних;
- застосовувати сучасні підходи і методи машинного навчання та створення ефективних систем автоматизації складними технологічними та організаційно-технічними об'єктами;
- пояснювати, кількісно та якісно оцінювати, корегувати отримані результати.

придбати навички:

- планування та виконання експериментальних і теоретичних досліджень та випробувань, вибирати для цього придатних методи та інструменти, здійснювати статистичну обробку даних, оцінювати адекватності отриманих результатів;
- застосування методів машинного навчання для розв'язання задач аналізу, інтелектуальної обробки даних і синтезу СІБС, управління складними комп'ютерними системами;
- проведення верифікації математичних методів, оцінки якості математичних методів на основі існуючих критеріїв;
- вирішення задач чисельного характеру з застосуванням спеціалізованих пакетів;

мати уявлення:

- про роль методів машинного навчання у створенні сучасних складних технічних систем; перспективах розвитку методів машинного навчання; про основні проблеми розробки сучасного програмного забезпечення для розв'язання задач аналізу, інтелектуальної обробки даних і синтезу СІБС, управління складними комп'ютерними системами та ін.

В результаті вивчення дисципліни у студента повинні формуватися такі *програмні результати навчання (ПРН)*.

ПРН01. Створювати системи автоматизації, кіберфізичні виробництва на основі використання інтелектуальних методів управління, баз даних та баз знань, цифрових та мережевих технологій, робототехнічних та інтелектуальних мехатронних пристроїв.

ПРН04. Застосовувати сучасні підходи і методи моделювання та оптимізації для дослідження та створення ефективних систем автоматизації складними технологічними та організаційно-технічними об'єктами.

ПРН05. Розробляти комп'ютерно-інтегровані системи управління складними технологічними та організаційно-технічними об'єктами, застосовуючи системний підхід із врахуванням нетехнічних складових оцінки об'єктів автоматизації.

ПРН07. Аналізувати виробничо-технічні системи у певній галузі діяльності як об'єкти автоматизації і визначати стратегію їх автоматизації та цифрової трансформації.

ПРН08. Застосовувати сучасні математичні методи, методи теорії автоматичного керування, теорії надійності та системного аналізу для дослідження та створення систем автоматизації складними технологічними та організаційно-технічними об'єктами, кіберфізичних виробництв.

ПРН10. Розробляти і використовувати спеціалізоване програмне забезпечення та цифрові технології для створення систем автоматизації складними організаційно-технічними об'єктами, професійно володіти спеціальними програмними засобами.

ПРН11. Дотримуватись норм академічної доброчесності, знати основні правові норми щодо захисту інтелектуальної власності, комерціалізації результатів науково-дослідної, винахідницької та проектної діяльності.

ПРН12. Збирати необхідну інформацію, використовуючи науково-технічну літературу, бази даних та інші джерела, аналізувати і оцінювати її.

ПРН13. Знати методи управління процесами різної природи, які побудовані на основі сучасних та перспективних методів математики, системного аналізу, штучного інтелекту.

2. Тематичний план навчальної дисципліни

Розділ 1. Методи оцінювання інформативності змінних стану при невизначеності даних.

Тема 1. Методи оцінюванні диференціальної інформативності з врахування точності вимірювання змінних стану и наявність парної кореляції між ними: кореляційного аналізу, дисперсійного аналізу і методи розпізнавання образів. Методи розпізнавання образів: детерміністські (дискримінантного аналізу, багатовимірною шкалювання і логічні), ймовірно-статистичні (методи Байеса, послідовного аналізу і оцінювання на основі теорії інформації). Стохастичний аналіз інформативності: індекси Соболя. Taguchi S / N Ratio.

Тема 2. Оцінювання інформативності на основі методів структурно-параметричного аналізу і синтезу регресійних моделей: факторного аналізу (головних компонент (МГК), нелінійні МГК, Грамма-Шмідта, аналізу компонентів на основі теорії інформації) і спрямованого перебору (ітеративні – на основі різних типів аппроксиматорів, в тому числі ШНМ, що навчаються), послідовного аналізу варіантів, вагові з адаптацією, локально-стохастичні на основі самоорганізації.

Розділ 2. Системні математичні моделі розпізнавання станів систем і процесів на основі методів глибокого навчання.

Тема 3. Згорткові нейронні мережі. Загальна архітектура, операція згортки, доповнення нулями (zero-padding) та операція зміщення (strides), операція об'єднання (pooling), операція Dropout, мета та алгоритм застосування, функціонали якості роботи нейронної мережі (НМ). Навчання НМ: нормалізація, пакетна нормалізація (Batch нормалізація), регуляризація, стохастичний градієнтний спуск, mini batch – стохастичний градієнтний пуск, метод імпульсів стохастичного градієнтного спуску, прискорений градієнт Нестерова, оптимізатор Adagrad, оптимізатор RMSProp, оптимізатор Adadelta.

Тема 4. Рекурентні нейронні мережі. Загальна структура. Типи рекурентних нейронних мереж. Навчання рекурентної нейронної мережі. Алгоритм Back Propagation Through Time. Мережі LSTM. Загальний опис архітектури. Forget gate, Output gate, Block of output.

Розділ 3. Прогнозування багатовимірних часових рядів на основі концепції тренд-аналізу. Математичні моделі контролю стану (ММКС) динамічних систем і процесів.

Тема 5. Загальні відомості про часові ряди. Визначення часового ряду (ЧР). Стаціонарність ЧР. Автоковаріаційна, автокореляційна, часткова автокореляційна функції ЧР.

Тема 6. Моделювання ЧР. Згладжене ковзне середнє. Експоненційне ковзне середнє. Регресивні моделі: згладжування часового ряду (змінного середнього і експоненційного згладжування), авторегресійні трендові моделі: лінійні (ARIMA, GARCH і SSM) та нелінійні (ШНМ з тимчасовими затримками, рекурентні ШНМ).

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин					
	Денна форма					
	Всього	у тому числі:				
		Л	ПЗ	Лаб. роб.	Інд.	СР
1	2	3	4	5	6	7

Розділ 1. Методи оцінювання інформативності змінних стану при невизначеності даних						
Тема 1. Методи оцінюванні диференціальної інформативності з врахуванням точності вимірювання змінних стану и наявність парної кореляції між ними.	22	2		4		16
Тема 2 Оцінювання інформативності на основі методів структурно-параметричного аналізу і синтезу регресійних моделей.	34	4		4	6	20
Всього за розділом 1	56	6		8	6	36
Розділ 2. Системні математичні моделі розпізнавання станів систем і процесів на основі методів глибокого навчання						
Тема 3. Згорткові нейронні мережі	40	6		6	8	20
Тема 4. Рекурентні нейронні мережі.	28	4		4		20
Всього за розділом 2	68	10		10	8	40
Розділ 3. Прогнозування багатовимірних часових рядів на основі концепції тренд-аналізу. Математичні моделі контролю стану (ММКС) динамічних систем і процесів						
Тема 5. Загальні відомості про часові ряди.	30	4		2	6	18
Тема 6. Моделювання ЧР.	26	4		4		18
Всього за розділом 3	56	8		6	6	36
Усього годин	180	24	0	24	20	112

4. Теми лабораторних робіт

№ п/п	Назва теми	Кількість годин
1	Зниження розмірності набору ознак за допомогою головних компонент	4
2	Навчання НМ за допомогою алгоритму зворотного поширення похибки	4
3	Порівняння якості класифікації за допомогою повнозв'язної мережі та згорткової НМ	4
4	Обробка тексту за допомогою рекурентної НМ	4
5	Прогнозування ЧР за допомогою моделей ARIMA(p,q,k)	4
6	Прогнозування ЧР за допомогою LSTM мережі	4
	Разом	24

5. Завдання для самостійної роботи

№ п/п	Зміст	Кількість годин
1	Дослідити математичні моделі і методи кластеризації даних (стратифікація прецедентів на підгрупи (класи)).	16
2	Дослідити системні математичні моделі розпізнавання станів систем і процесів на основі даних моніторингу.	16
3	Дослідити методи розв'язання прямої та зворотної задач розрахунку розмірних конструкторських ланцюгів і методи їх вирішення.	18
4	Дослідити приклади застосування меметичних алгоритмів для вирішення задач робастного оцінювання.	16

5	Ознайомлення з основними характеристиками та принципами роботи з спеціалізованими пакетами прикладних програм (ППП) для побудови моделей процесів управління нелінійними динамічними системами на основі статистичних даних про змінний стану.	18
6	Ознайомлення з основними характеристиками та принципами роботи з спеціалізованими пакетами прикладних програм (ППП) для вирішення завдань РОП&ІД систем і процесів.	18
7	Підготовка до курсового проекту (роботи).	20
8	Підготовка до підсумкової контрольної роботи	10
	Разом	132

6. Індивідуальні завдання

Індивідуальне завдання пов'язане із застосуванням методів машинного навчання в конкретному завданні, розробкою програми для його реалізації і обґрунтуванням корисності і ефективності прийнятого рішення.

Індивідуальне завдання виконується у вигляді 1 контрольної роботи і курсового проекту (роботи).

7. Методи контролю

Контроль роботи студентів при вивченні дисципліни і засвоєння ними навчального матеріалу здійснюється на практичному занятті шляхом проведення контрольних опитувань і захисту звітів з практичних завдань. Підсумковий контроль здійснюється при виконанні 1 контрольної роботи, захисті курсової роботи і на екзамені.

Студенти, що не виконали впродовж семестру 1 контрольну роботу, а також не представили і не захистили звіти з практичних завдань і курсову роботу, до екзамену не допускаються.

Екзаменаційний білет містить два теоретичних і одне практичне питання. Максимальна кількість балів за відповіді на кожне теоретичне питання складає по 12 балів, на практичне питання – 16 балів.

При дистанційному навчанні видача практичних завдань та контроль їх виконання здійснюється за допомогою сервісу дистанційного навчання Google Classroom. Лекційні заняття проводяться із використанням сервісу відео-конференцій Google Meet., Google Classroom.

Підсумковий контроль у вигляді екзамену (заліку) проводиться шляхом відповіді на білет та он-лайн опитування (сервіси відео-конференцій Google Meet, Google Classroom).

8. Схема нарахування балів

Бали за поточний контроль знань впродовж семестру (по темах)						Контрольна робота, передбачена навчальним планом	Курсова робота	Разом	Іспит	Сума
Розділ 1		Розділ 2		Розділ 3						
T1	T2	T3	T4	T5	T6	1				
5	5	5	5	5	5	10	20	60	40	100

T1, T2 ... – теми розділів.

За темою T1 студент отримує 5 балів за виконання практичної роботи 1.

За темою T2 студент отримує 5 балів за виконання практичної роботи 2.

За темою T3 студент отримує 5 балів за виконання практичної роботи 3.

За темою T4 студент отримує 5 балів за виконання практичної роботи 4.

За темою T5 студент отримує 5 балів за виконання практичної роботи 5.

За темою Т6 студент отримує 5 балів за виконання практичної роботи 6.

Критерії оцінювання знань студентів за практичні роботи

Вимоги	Кількість балів
<ul style="list-style-type: none"> ▪ Завдання відзначається повнотою виконання без допомоги викладача. ▪ Визначає рівень поінформованості, потрібний для прийняття рішень. Вибирає інформаційні джерела. ▪ Робить висновки і приймає рішення у ситуації невизначеності. Володіє уміннями творчо-пошукової діяльності. 	8
<ul style="list-style-type: none"> ▪ Завдання – повні, з деякими огріхами, виконані без допомоги викладача. ▪ Планує інформаційний пошук; володіє способами систематизації інформації; ▪ Робить висновки і приймає рішення у ситуації невизначеності. Володіє уміннями творчо-пошукової діяльності. 	6-7
<ul style="list-style-type: none"> ▪ Завдання відзначається неповнотою виконання без допомоги викладача. ▪ Студент може зіставити, узагальнити, систематизувати інформацію під керівництвом викладача; вільно застосовує вивчений матеріал у стандартних ситуаціях. 	4-5
<ul style="list-style-type: none"> ▪ Завдання відзначається неповнотою виконання за консультацією викладача. ▪ Застосовує запропонований вчителем спосіб отримання інформації, має фрагментарні навички в роботі з підручником, науковими джерелами; ▪ Вибирає відомі способи дій для виконання фахових методичних завдань. 	3
Завдання відзначається фрагментарністю виконання за консультацією викладача або під його керівництвом.	1-2

Критерії оцінювання знань студентів за контрольну роботу

Вимоги	Кількість балів
Повнота виконання завдання повна, студент здатен формулювати закони та закономірності, структурувати судження, умовиводи, доводи, описи.	9-12
Повнота виконання завдання повна, студент здатен формулювати операції, правила, алгоритми, правила визначення понять.	5-8
Повнота виконання завдання елементарна, студент здатен вибирати відомі способи дій для виконання фахових завдань.	3-5
Повнота виконання завдання фрагментарна.	1-2

Критерії оцінювання екзаменаційних робіт студентів

Вимоги	Кількість балів
Показані всебічні систематичні знання та розуміння навчального матеріалу; безпомилково виконані завдання.	35-40
Показані повні знання навчального матеріалу; помилки, якщо вони є, не носять принципового характеру.	30-35
Показано повне знання необхідного навчального матеріалу, але допущені помилки.	20-30
Показано повне знання необхідного навчального матеріалу, але допущені суттєві помилки	10-20

Показано недосконале знання навчального матеріалу, допущені суттєві помилки.	5-10
Показано недосконале знання навчального матеріалу, допущені суттєві помилки, які носять принциповий характер; обсяг знань не дозволяє засвоїти предмет.	1-5

Критерії оцінювання виконання курсової роботи

Параметри оцінювання	Діапазон оцінки, балів	Критерії оцінювання за бальною шкалою
Оцінювання якості курсової роботи	0 - 20	
Вміння чітко та стисло викласти основні результати дослідження	0-5	0 – студент неспроможний чітко та стисло викласти основні результати дослідження
		5 – студент невпорядковано викладає основні результати дослідження
		10 – студент спроможний чітко та стисло викласти основні результати дослідження
Використання роздаткового ілюстративного матеріалу	0-10	0 – роздатковий ілюстративний матеріал не використано
		5 – доповідь супроводжена роздатковим ілюстративним матеріалом
Повнота, глибина, обґрунтованість відповідей на питання	0-5	0 – студент неспроможний надати відповіді на поставлені питання
		5 – студент надає неповні, поверхові, необґрунтовані відповіді на поставлені питання
		10 – студент надає повні, глибокі, обґрунтовані відповіді на поставлені питання

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90 – 100	відмінно	зараховано
70-89	добре	
50-69	задовільно	
1-49	незадовільно	не зараховано

9. Рекомендована література

Основна література

1. Elgendy M. Deep Learning for Vision Systems / Mohamed Elgendy., 2020. – 480 с. – (1st edition).
2. Buduma N. Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms / Nikhil Buduma., 2017. – 283 с.
3. О.М. Пупена, Д.В. Мацебула. Промислові інформаційні мережі та інтеграційні технології. Конспект лекцій. – К. НУХТ. – 89 с.
4. MacKay David J.C. Information Theory, Inference and Learning Algorithms. – Cambridge University Press, 2003. – 628 p.
5. Sutton Richard S., Barto Andrew G. Reinforcement Learning: An Introduction. – Cambridge, Massachusetts: MIT Press, 2018. – 426 p.
6. Hastie Trevor, Tibshirani Robert, Friedman Jerome. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. – New York: Springer, 2009. – 745 p.

Допоміжна.

1. Meniaailov Ievgen, Mathematical Models and Methods of Effective Estimation in Multi-Objective Optimization Problems under Uncertainties/ Ievgen Meniaailov, Olexandr Khustochka, Kateryna Ugryumova, Sergey Chernysh, Sergiy Yepifanov, Mykhaylo Ugryumov // Advances in Structural and Multidisciplinary Optimization: Proceedings of the 12th World Congress of Structural and Multidisciplinary Optimization (WCSMO12) By Axel Schumacher (05th - 09th, June 2017, Braunschweig, Germany).– SpringerLink, 2018.– 2115 p. (ISBN: 978-331-967-987-7) (Paper No. 0011, P.411-427)
2. Rasmussen C. E., Williams C. K. I. Gaussian Processes for Machine Learning. –Cambridge, Massachusetts: MIT Press, 2006. – 248 p.
3. Bishop Christopher M. Pattern Recognition and Machine Learning. – New York: Springer, 2006. – 738 p.
4. Deep Learning Tutorial. – LISA lab, University of Montreal, 2015. – 167 p.
5. Hal Daume III. A course in machine learning. – Edited by John Mark Ockerbloom (onlinebooks@pobox.upenn.edu), 2015. – 193 p.

10. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. <https://www.datacamp.com/courses/deep-learning-in-python>
2. <https://www.kaggle.com/learn/machine-learning>
3. Datasets. – <https://www.kaggle.com>