

Чернівецький національний університет імені Юрія Федьковича

(повне найменування закладу вищої освіти)

Факультет математики та інформатики

(назва факультету/навчально-наукового інституту)

Кафедра математичного моделювання

(назва кафедри)

Декан

“ЗАТВЕРДЖУЮ”

Мартинюк О.В.

“ 25 ” 06 2024 року

РОБОЧА ПРОГРАМА

навчальної дисципліни

Технології машинного навчання

(назва навчальної дисципліни)

вибіркова

(вказати: обов'язкова / вибіркова)

Освітньо-професійна програма "Інформаційні технології та управління проектами"

(назва програми)

Спеціальність 122 – Комп'ютерні науки

(вказати: код, назва)

Галузь знань 12 – Інформаційні технології

(вказати: шифр, назва)

Рівень вищої освіти перший (бакалаврський)

(вказати: перший бакалаврський/другий магістерський)

Факультет математики та інформатики

(назва факультету/ навчально-наукового інституту, на якому здійснюється підготовка фахівців за вказаною освітньо-професійною програмою)

Мова навчання українська

(вказати: на якій мові читається дисципліна)

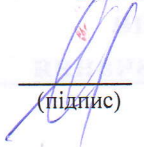
Чернівці 2024 рік

Робоча програма навчальної дисципліни Технології машинного навчання складена відповідно до освітньо-професійної програми "Інформаційні технології та управління проектами", затвердженої Вченою радою Чернівецького національного університету імені Юрія Федьковича (протокол № 7 від 29 квітня 2024 р.)

Розробники: канд. ф.-м.н. , доц. Дорошенко І.В.

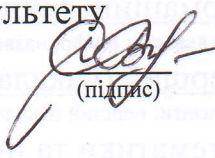
Погоджено з гарантом ОП і затверджено на засіданні кафедри математичного моделювання

Протокол № 18 від "25" червня 2024 року

Завідувач кафедри  Черевко І.М.
(підпис) (прізвище та ініціали)

Схвалено методичною радою факультету математики та інформатики

Протокол № 11 від "25" червня 2024 року

Голова методичної ради факультету
математики та інформатики  Сікора В.С.
(підпис) (прізвище та ініціали)

1. Мета навчальної дисципліни: формування у студентів сучасного наукового світогляду в області методів машинного навчання; наукової уяви про задачі, що розв'язуються з допомогою методів машинного навчання, вивчення методів класифікації і регресії з вчителем, а також методи кластерного аналізу (без вчителя); знайомство студентів з сучасними технологіями машинного навчання та тенденціями розробки і застосування; подальше становлення і вдосконалення інформаційної та програмної культури майбутніх фахівців.

2. Результати навчання:

Студент повинен мати навички (набути досвід): - розробки інструментальних засобів аналізу даних методами машинного навчання.

Формування компетентностей, а саме

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК4. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

ЗК8. Здатність генерувати нові ідеї (креативність).

ЗК11. Здатність приймати обґрунтовані рішення.

ФК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

ФК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

В результаті навчального курсу студенти повинні

знати: – основні задачі машинного навчання; основні типи даних та методи побудови матриць суміжності; основні методи машинного навчання; основні засоби мов R для розв'язання задач машинного навчання.

вміти:

- виконувати обчислення, пов'язані з навчанням і роботою моделей машинного навчання;
- вибирати набір ознак для класифікації і кластеризації та проводити попередню обробку даних; виконувати обчислення, пов'язані з навчанням і роботою моделей машинного навчання;
- вміти будувати різні типи алгоритмів машинного навчання;
- визначати оптимальний метод для кожної задачі;
- розв'язувати задачі машинного навчання засобами R.

Програмні результати навчання:

ПРН1. Застосовувати знання основних форм і законів абстрактнологічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

ПРН5. Проектувати, розробляти та аналізувати алгоритми розв'язання обчислювальних та логічних задач, оцінювати ефективність та складність алгоритмів на основі застосування формальних моделей алгоритмів та обчислюваних функцій.

ПРН8. Використовувати методологію системного аналізу об'єктів, процесів і систем для задач аналізу, прогнозування, управління та проектування динамічних процесів в макроекономічних, технічних, технологічних і фінансових об'єктах.

ПРН12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

3. Опис навчальної дисципліни

3.1. Загальна інформація

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин						Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	4	8	3	90	22	-	-	22	46	-	залік

3.2. Структура змісту навчальної дисципліни

Назви змістових модулів і тем	Кількість годин												
	денна форма							Заочна форма					
	усього	у тому числі					усього	у тому числі					с.р.
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13	
Змістовний модуль 1. Основні поняття і визначення машинного навчання. Класифікація													
ТЕМА1. Вступ до машинного навчання. Лінійні методи машинного навчання	4	2	-	-	-	2							
ТЕМА2. Задачі навчання за прецедентами.	8	2	-	-	-	6							
ТЕМА3. Особливі види регресії в машинному навчанні. Логістична регресія	4	2		-		2							
ТЕМА4. Лінійні методи класифікації	4	2	-	-		2							
ТЕМА5. Методи мультикласової класифікації	6	2	-	-		4							
ТЕМА6. Лабораторна робота 1. Основи роботи в R.	4	-	-	2		2							
ТЕМА7. Лабораторна робота 2. Регресійний аналіз в R.	6	-	-	4		2							
ТЕМА8. Лабораторна робота	6	-	-	4		2							

3. Моделі класифікації												
Разом за змістовим модулем 1	42	10	-	10	-	22						
Змістовий модуль 2. Навчання без учителя. Композиції алгоритмів												
ТЕМА1. Робота з ознаками. Деревя рішень	4	2	-	-		2						
ТЕМА2. Композиції алгоритмів	6	2	-	-		4						
ТЕМА3. Навчання без учителя	4	2	-	-		2						
ТЕМА4. Кластеризація	9	4		-		5						
ТЕМА5. Машинне навчання на асоціаціях	7	2		-		5						
ТЕМА6. Лабораторна робота 4. Моделі кластеризації: метод <i>k</i> -середніх та агломера-тивна ієрархічна кластеризація	6	-	-	4		2						
ТЕМА7. Лабораторна робота 5. Моделі кластеризації: EM-алгоритм	6	-	-	4		2						
ТЕМА8. Лабораторна робота 6. Моделі кластеризації: алгоритми DBSCAN та OPTICS	6	-	-	4		2						
Разом за змістовим модулем 2	48	12	-	12		24						
Усього годин	90	22	-	22		46						

3.3 Тематика лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Основи роботи в R.	2
2	Регресійний аналіз в R.	4
3	Моделі класифікації	4
4	Моделі кластеризації: метод k -середніх та агломеративна ієрархічна кластеризація	4
5	Моделі кластеризації: ЕМ-алгоритм	4
6	Моделі кластеризації: алгоритми DBSCAN та OPTICS	4

3.4. Самостійна робота студента

№ з/п	Назва теми	Кількість балів
1	Особливості роботи з реальними даними Пропуски в даних. Попередня обробка ознак. Чистка даних. Категорійні ознаки: кодування, хешування, лічильники. Робота з текстами. Розріджені ознаки: векторизація, хешування, TF-IDF. Косинусна метрика.	10
2	Машинне навчання в прикладних задачах. Етапи аналізу даних. Робота з числовими ознаками. Робота з категоріальними та текстовими ознаками. Підготовка даних. Оцінювання якості роботи алгоритму.	10
3	Підходи до отримання ознак для складних даних Робота з зображеннями (фільтри, отримання ознак за допомогою нейромереж), текстами (word embeddings).	10
4	Колаборативна фільтрація. Задачі колаборативної фільтрації і матриця суб'єкти-об'єкти. Латентні методи на основі бі-кластеризації. Алгоритм Брегмана. Латентні методи на основі матричних розкладань. для розріджених даних.	10
5	Багатошарові нейронні мережі. Біологічний нейрон. Функції активації. Проблема повноти. Повнота двошарових мереж в просторі булевих функцій. Теореми Колмогорова, Стоуна, Горбаня (без доведення). Алгоритм зворотного поширення помилок. Метод пошарового налаштування мережі. Підбір структури мережі: методи поступового ускладнення мережі, оптимальне проріджування нейронних мереж.	10
6	Рекомендаційні системи Постановки задачі. Метрики якості. Методи, базовані на колаборативній фільтрації. Методи, базовані на матричних розкладах.	10

4. Освітні технології, методи навчання і викладання навчальної дисципліни

У викладання курсу використовуються такі освітні технології:

- Лекції та їх презентації.
- Онлайн-лекції.
- Лабораторні заняття.
- Групова робота, коли студенти розв'язують практичні завдання.
- Онлайн-тести та опитування: Використання системи MOODLE

Електронні підручники і ресурси репозитарію ЧНУ

Методи навчання:

МН 1 - лекція-візуалізація;

МН 8 – робота з тестами;

МН 9 – робота в групах;

МН 12 – дистанційне навчання з використанням відповідних онлайн-платформ

5. Критерії оцінювання результатів навчання з навчальної дисципліни

5.1. Критерієм підсумкового оцінювання

Змістовий модуль 1

- лабораторні роботи (25 балів)
- теоретичне опитування (5 балів)

Змістовий модуль 2

- лабораторні роботи (25 балів)
- теоретичне опитування (5 балів)

Залік (40 балів)

Критерієм успішного проходження здобувачем освіти підсумкового оцінювання може бути досягнення ним мінімальних порогових рівнів оцінок за кожним запланованим результатом навчання навчальної дисципліни.

Оцінювання знань студента на заліку, під час лабораторних занять та виконання індивідуальних завдань проводиться за такими критеріями:

- розуміння, ступінь засвоєння теорії та методології навчальної дисципліни для розв'язання проблем, що розглядаються;
- рівень знань з теорії дисципліни та понятійно-категоріального апарату, термінології, поняття і принципи предметної області навчальної дисципліни.
- повнота розкриття питання; уміння чітко формулювати визначення понять/термінів й пояснювати їх; здатність аргументувати відповідь;
- аналітичні міркування, порівняння, формулювання висновків; логічна послідовність, культура мови; емоційність та вміння переконувати.
- ступінь засвоєння фактичного матеріалу навчальної дисципліни;
- ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- вміння поєднувати теорію з практикою при розгляді виробничих ситуацій, проведенні аналізу, розв'язанні задач, проведенні розрахунків у процесі виконання індивідуальних завдань;
- застосування аналітичних підходів;
- здатність проводити критичну та незалежну оцінку певних проблемних питань, бачити слабкі й сильні сторони організації, обґрунтовувати можливості і загрози, що існують у зовнішньому середовищі організації;
- вміння пояснювати альтернативні погляди та наявність власної точки зору, позиції на певне проблемне питання;

- якість і чіткість викладення міркувань;
- обґрунтованість висновків щодо розробки стратегії розвитку досліджуваного підприємства (організації).

Дедлайни та перескладання. Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку. Перескладання модулів відбувається з дозволу деканату за наявності поважних причин (наприклад, лікарняний, участь у конференції, студентській олімпіаді).

Академічна доброчесність. Здобувачі вищої освіти самостійно виконують навчальні завдання, завдання поточного та підсумкового контролю результатів навчання. Обов'язковим є посилання на джерела інформації в разі використання ідей, розробок, тверджень.

Відвідування занять. Відвідування занять є обов'язковою умовою виконання навчального плану дисципліни. Форми навчання визначені затвердженим графіком освітнього процесу Чернівецького національного університету імені Юрія Федьковича.

Критерії оцінювання.

Оцінка знань здобувачів включає поточний та підсумковий контроль. Поточний контроль здійснюється впродовж семестру під час проведення лекційних та лабораторних занять. Підсумковий контроль має за мету – перевірку теоретичних знань здобувачів, виявлення навичок застосування перших при вирішенні практичних завдань, а також навиків самостійної роботи з навчальною і науковою літературою.

Загальна кількість балів, яку здобувач може отримати у процесі вивчення дисципліни становить 100 балів, з яких 60 балів (по 30 балів за перший та другий модуль) здобувач може одержати як суму результатів поточного контролю (контрольні, самостійні роботи та тестування) і 40 балів – на підсумковому модулі (заліку).

Білет містить чотири питання, з яких одне теоретичне, три практичних. Повна відповідь на кожне питання оцінюється 10 балами. За кожну помилку, яка допущена у відповіді, знімається певна кількість балів, а саме:

а) при відповіді на теоретичне питання у випадку неістотної помилки знімається 1-3 бали, а у випадку істотної 4-7 балів, якщо ж здобувач не опанував теоретичний матеріал дисципліни, плутається в означеннях, наводить логічно неправильні твердження, то знімається до 9 балів;

б) при оцінці практичного завдання за помилку, допущену при обчисленнях, знімається 1-2 бали, за істотну помилку, знімається 3-5 балів, якщо ж розв'язання задачі логічно неправильне, то знімається до 8 балів.

Підсумкова оцінка виставляється за результатами суми балів набраних за кожне питання білета з додаванням сумарної кількості балів за перший та другий модуль.

Розподіл балів, які отримують студенти

Поточне оцінювання (аудиторна та самостійна робота)						Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль №1			Змістовий модуль № 2				
ТЕМА 6.	ТЕМА 7.	ТЕМА 8.	ТЕМА 6.	ТЕМА 7.	ТЕМА 8.	40	100
10	10	10	10	10	10		

5.2. Шкала оцінювання: національна та ЄКТС

Оцінка за національною шкалою	Оцінка за шкалою ECTS	
	Оцінка (бали)	Пояснення за розширеною шкалою
Зараховано	A (90-100)	відмінно
	B (80-89)	дуже добре
	C (70-79)	добре
	D (60-69)	задовільно
	E (50-59)	достатньо
Не зараховано	FX (35-49)	(незадовільно) з можливістю повторного складання
	F (1-34)	(незадовільно) з обов'язковим повторним курсом

5.3. Засоби оцінювання

- перевірка та захист студентами письмових звітів про виконання кожної лабораторної роботи,
- експрес-опитування,
- тестові завдання.

6. Форми поточного та підсумкового контролю

1. Поточний (захист лабораторних робіт)
2. Модульний (тестування, виконання завдань)
3. Підсумковий (залік)

7. Рекомендована література

7.1. Основна

1. Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.
2. Системи машинного навчання: Методичні вказівки до лабораторних робіт, укл. І.В. Дорошенко. – Чернівці: ЧНУ, 2024. 112 с.
<https://archer.chnu.edu.ua/xmlui/handle/123456789/10073>
3. David Barber. Bayesian Reasoning and Machine Learning. – Cambridge University Press, 2012. – 697 p.

4. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Data Mining, Inference and Prediction. – Springer, 2018. – 745 p.
5. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. An Introduction to Statistical Learning (with applications in R). – Springer, 2018. – 426 p.
6. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014. — 739 p.
7. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
8. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012.
9. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
10. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.

7.2. Допоміжна

1. Doroshenko I.V., Knihnitska T.V., Deretorska T.I. Comparison of machine learning algorithms for predicting mortality from Covid-19 virus // Sworld Jornal Issue No11, Part 2 January 2022 – P. 72-77
2. Doroshenko I.V., Knihnitska T.V., Kreshtanovych M.A. Comparison of data clustering algorithms // Sworld Jornal Issue No23, Part 1 January 2024 – P. 116-127
3. Doroshenko, I., Knopov, O. & Vovk, L. Mathematical Models of Extreme Modes in Ecological Systems // Cybernetics and Systems Analysis.– 2022.– Vol.58, N5.– P.764–779.
4. Doroshenko I.V., Knihnitska T.V. Application of statistical analysis for medical data // International Scientific Periodical Journal "Modern engineering and innovative technologies" No31 Part 1 February 2024 – P. 86-91

8. Інформаційні ресурси

1. <http://cran.r-project.org/bin/windows/base/>