

Чернівецький національний університет імені Юрія Федьковича

(повне найменування закладу вищої освіти)

Факультет математики та інформатики

(назва інституту/факультету)

Кафедра математичного моделювання

(назва кафедри)

СИЛАБУС навчальної дисципліни

Системи машинного навчання

обов'язкова

(вказати: обов'язкова)

Освітньо-професійна програма Системний аналіз

(назва програми)

Спеціальність 124 - Системний аналіз

(вказати: код, назва)

Галузь знань 12 - Інформаційні технології

(вказати: шифр, назва)

Рівень вищої освіти другий (магістерський)

(вказати: перший (бакалаврський)/другий (магістерський)/третій (освітньо-науковий))

математики та інформатики

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

Мова навчання українська

(вказати: на яких мовах читається дисципліна)

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

(вказати авторів (викладач (ів)), їхні посади, наукові ступені, вчені звання)

Профайл викладача (-ів) <http://matmod.fmi.org.ua/pro-kafedru/spivrobitnyky/doroshenko-irina-viktorivna/>

Контактний тел. 0504340655

E-mail: i.doroshenko@chnu.edu.ua

Сторінка курсу в Moodle <https://moodle.chnu.edu.ua/course/view.php?id=3473>

Консультації

1. Анотація дисципліни (призначення навчальної дисципліни).

Призначення дисципліни – вивчення методів, що застосовуються для побудови складних моделей та алгоритмів із метою вирішення завдань класифікації, кластеризації та прогнозування. Ці аналітичні моделі дозволяють дослідникам, науковцям із даних, інженерам та аналітикам «виробляти надійні, повторювані рішення і результати» та розкривати «приховані розуміння» шляхом навчання з історичних співвідношень та тенденцій у даних.

2. 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.

Основними завданнями: набуття практичних навичок і знань в області технологій машинного навчання. У результаті вивчення даної дисципліни студенти повинні освоїти основні методи навчання з вчителем (Байєсівський класифікатор, лінійний дискримінантний аналіз Фішера, логістична регресія, метод опорних векторів, дерева рішень, випадковий ліс) і без вчителя – кластеризація розбиттям, ієрархічна кластеризація та нечітка кластеризація. Також, в результаті освоєння матеріалу, студенти повинні вивчити основні практичні прийоми роботи з інформацією мовами R.

3. Пререквізити. Теорія ймовірностей та математична статистика, статистика, аналіз даних, програмування.

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

Метою вивчення дисципліни є формування компетентностей, а саме здатність:

- спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань);
- розробляти системи підтримки прийняття рішень та рекомендаційні системи;
- оцінювати ризики, розробляти алгоритми управління ризиками в складних системах різної природи;
- моделювати, прогнозувати та проектувати складні системи і процеси на основі методів та інструментальних засобів системного аналізу;
- застосовувати теорію і методи Data Science для здійснення інтелектуального аналізу даних з метою виявлення нових властивостей та генерації нових знань про складні системи;
- управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів;
- здійснювати захист прав інтелектуальної власності, комерціалізацію результатів досліджень та інновацій.

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:

- будувати та досліджувати моделі складних систем і процесів застосовуючи методи системного аналізу, математичного, комп'ютерного та інформаційного моделювання;
- застосовувати методи розкриття невизначеностей в задачах системного аналізу, розкривати ситуаційні невизначеності та невизначеності в задачах взаємодії, протидії та конфлікту стратегій, знаходити компроміс при розкритті концептуальної невизначеності;

Тема 3. Нейронні мережі	9	2	-	2		5						
Тема 4. Навчання без учителя	8	1	-	2		5						
Тема 5. Машинне навчання на асоціаціях	7	1		1		5						
Тема 6. Кластеризація	9	2		2		5						
Разом за змістовим модулем 2	48	9	-	9		30						
Усього годин	90	15	-	15		60						

5.3. Зміст завдань для самостійної роботи

№ з/п	Назва теми
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.
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.
3	. Навчання з учителем
4	Підходи до отримання ознак для складних даних Робота з зображеннями (фільтри, отримання ознак за допомогою нейромереж), текстами (word embeddings).
5	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.
6	Багатошарові нейронні мережі. Біологічний нейрон. Функції активації. Проблема повноти. Повнота двошарових мереж в просторі булевих функцій. Теорема Колмогорова, Стоуна, Горбаня (без доведення). Алгоритм зворотного поширення помилок. Метод пошарового налаштування мережі. Підбір структури мережі: методи поступового ускладнення мережі, оптимальне проріджування нейронних мереж.
7	Рекомендаційні системи Постановки задачі. Метрики якості. Методи, базовані на колаборативній фільтрації. Методи, базовані на матричних розкладах.

6. Система контролю та оцінювання

Види та форми контролю

1. Поточний (захист лабораторних робіт, опитування теоретичного матеріалу)
2. Модульний (тестування, виконання завдань)
3. Підсумковий (екзамен)

Засоби оцінювання

Засобами оцінювання та демонстрування результатів навчання можуть бути:

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

Критерії оцінювання результатів навчання з навчальної дисципліни

Шкала оцінювання: національна та ЄКТС

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

Розподіл балів, які отримують студенти

Поточне оцінювання (<i>аудиторна та самостійна робота</i>)									Кількість балів (екзамен)	Сумарна к-ть балів
Змістовий модуль №1					Змістовий модуль № 2					
T1	T2	T3	T4	T5	T6	T7	T8	T9	40	100
-	10	10	5	5	10	10	5	5		

T1, T2 ... T9 – теми змістових модулів.

7. Рекомендована література -основна

1. Hastie T., Tibshirani R, Friedman J. The Elements of Statistical Learning (2nd edition). Springer, 2009.
2. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
3. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012.
4. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
5. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.
6. Николенко С.И. Курс лекцій по машинному обучению — слайды. Електронний ресурс. Режим доступу: <http://logic.pdmi.ras.ru/sergei/index.php?page=mlaptu09>
7. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.
8. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.
9. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985.
10. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989.
11. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.
12. HastieT., TibshiraniR., FriedmanJ. The Elementsof Statistical Learning. Springer, 2014. — 739 p.
13. Bishop C.M. Pattern Recognitionand Machine Learning. — Springer, 2006. — 738 p