

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

Технології машинного навчання

вibірко́ва

(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 перший (бакалаврський)

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

Факультет 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

I.V.

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

Профайл викладача (-ів) <https://mathmod.chnu.edu.ua/pro-nas/spivrobotnyky/doroshenko-iryna-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 є 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 мовою R.

3. Пререквізити. Теорія ймовірностей та математична статистика, Програмування, Аналіз даних.

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

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:

- вибирати набір ознак для класифікації і кластеризації та проводити попередню обробку даних; виконувати обчислення, пов'язані з навчанням і роботою моделей машинного навчання;

- розв'язувати задачі машинного навчання засобами R.

5. Опис навчальної дисципліни

5.1. Загальна інформація

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

5.2. Дидактична карта навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	усього	у тому числі					усього	у тому числі				
		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
Змістовий модуль 1.												
Н.Е1.1. Вступ до машинного навчання. Лінійні методи машинного навчання	4	2	-	-	-	2						
Н.Е1.2. Задачі навчання за прецедентами.	8	2	-	-	-	6						
Н.Е1.3. Особливі види регресії в машинному навчанні. Логістична регресія	4	2		-		2						
Н.Е1.4. Лінійні методи класифікації	4	2		-		6						
Н.Е1.5. Методи мультикласової класифікації	6	2		-		6						
Н.Е.1.6. Лабораторна робота 1. Основи роботи в R	4	-	-	2		4						
Н.Е.1.7. Лабораторна робота 2. Регресійний аналіз в R	6	-	-	4		2						
Н.Е. 1.8. Лабораторна робота 3. Моделі класифікації	6	-	-	4		2						
Разом за змістовим модулем 1	42	10	-	10	-	22						
Змістовий модуль 2.												
Н.Е. 2.1. Робота з ознаками. Дерева рішень	4	2	-	-		2						
Н.Е. 2.2. Композиції алгоритмів	6	2	-	-		4						
Н.Е. 2.3. Навчання без учителя	4	2	-	-		2						
Н.Е. 2.4. Кластеризація	9	4		-		5						
Н.Е. 2.5. Машинне навчання на асоціаціях	7	2		-		5						
Н.Е.2.6. Лабораторна робота 4. Моделі кластеризації: метод k -середніх та агломеративна ієрархічна кластеризація	6	-	-	4		2						

Н.Е.2.7. Лабораторна робота 5. Моделі кластеризації: EM-алгоритм	6	-	-	4	2						
Н.Е.2.8. 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						

5.3. Самостійна робота студента

Самостійна робота складає 46 годин. Розподіл самостійної роботи за видами навчальних робіт:

- підготовка до лекційних занять – 10 годин;
- підготовка до лабораторних занять – 20 годин;
- підготовка до підсумкового модуль-контролю – 16 годин.

Студенти можуть отримати до 10 балів в рахунок ІНДЗ

№ з/п	Назва теми ІНДЗ
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	Підходи до отримання ознак для складних даних Робота з зображеннями (фільтри, отримання ознак за допомогою нейромереж), текстами (word embeddings).
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.
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.
6	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 (залік)

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

Система оцінювання рівня навчальних досягнень ґрунтується на принципах ECTS та є накопичувальною. Протягом семестру студенти виконують 6 лабораторних робіт. Кожна лабораторна робота оцінюється кількістю балів за таблицею, наведеною нижче.

Виконуючи завдання лабораторної роботи, студент повинен оформити і завантажити для подальшої перевірки на сайт електронного навчання звіт разом із працездатними файлами програмної реалізації завдань ЛР (правила оформлення наведені на сторінці навчальної дисципліни на сайті).

50% балів, відведених на оцінювання ЛР, студент отримує за працюючий програмний продукт, в якому реалізовано всі завдання ЛР та оформлений звіт. Решта 50% балів виставляється після захисту студентом виконаного звіту. На захисті звіту з ЛР студент має відповісти на питання щодо постановки задачі та розробленого ним алгоритму реалізації кожного із завдань ЛР. При відповіді на теоретичні питання та питання щодо програмної реалізації алгоритму у випадку неістотної помилки знімається 10-20% балів, а у випадку істотної 20-40% балів, якщо ж студент не опанував теоретичний матеріал, плутається в означеннях, наводить логічно невірні твердження, то знімається до 50% балів від усієї суми балів за ЛР.

Максимальна кількість, яку можна набрати на підсумковому модулі (тестування) – 40 балів.

Підсумкова оцінка виставляється за результатами суми балів, набраних на змістових модулях під час семестру та підсумковому модулі згідно з нижче наведеною таблицею.

Розподіл балів, які отримують студенти

Поточне оцінювання (аудиторна та самостійна робота)						Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль №1			Змістовий модуль № 2				
Н.Е.1.6.	Н.Е.1.7.	Н.Е.1.8.	Н.Е.2.6.	Н.Е.2.7.	Н.Е.2.8.	40	100
10	10	10	10	10	10		

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
Незараховано	FX (35-49)	(незадовільно) з можливістю повторного складання

	F (1-34)	(незадовільно) з обов'язковим повторним курсом
--	----------	--

7. Політика освітнього процесу

Здобувач зобов'язаний своєчасно та якісно виконувати всі отримані завдання; за необхідністю з метою з'ясування всіх незрозумілих під час самостійної та індивідуальної роботи питань, відвідувати консультації викладача.

Студенти мають дотримуватись правил академічної доброчесності відповідно до "Кодексу академічної доброчесності ЧНУ". Політика дотримання академічної доброчесності (відповідно до Закону України "Про вищу освіту") – викладання навчальної дисципліни ґрунтується на засадах академічної доброчесності – сукупності етичних принципів та визначених законом правил, якими мають керуватися учасники освітнього процесу під час навчання, викладання та провадження наукової (творчої) діяльності з метою забезпечення довіри до результатів навчання та/або наукових (творчих) досягнень. Наявність академічного плагіату в студентських роботах є підставою для виставлення негативної оцінки. Списування студентів під час проведення модульної контрольної роботи є підставою для дострокового припинення її складання та виставлення негативної оцінки.

Складання / перескладання екзаменів відбувається за встановленим деканатом розкладом.

8. Рекомендована література

1. Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.
2. David Barber. Bayesian Reasoning and Machine Learning. – Cambridge University Press, 2012. – 697 p.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Data Mining, Inference and Predictio. – Springer, 2018. – 745 p.
4. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. An Introduction to Statistical Learning (with applications in R). – Springer, 2018. – 426 p.
5. HastieT., TibshiraniR., FriedmanJ. The Elementsof Statistical Learning. Springer, 2014. — 739 p.
6. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
7. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012.
8. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
9. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.

9. Інформаційні ресурси

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