

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет прикладної математики та інформатики
Кафедра математичного моделювання соціально-економічних процесів

Затверджено

На засіданні
кафедри математичного моделювання
соціально-економічних процесів
факультету прикладної математики та
інформатики
Львівського національного університету
імені Івана Франка
(протокол № 1 від 28.08. 2024 р.)



Завідувач кафедри Петро СЕНЬО

Силабус з навчальної дисципліни
“ Інтелектуальний аналіз даних”,
що викладається в межах ОПП Системний аналіз
другого (магістерського) рівня вищої освіти для здобувачів з
спеціальності 124 – Системний аналіз

Львів 2024 р.

Назва дисципліни	Інтелектуальний аналіз даних
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
Факультет та кафедра, за якою закріплена дисципліна	Факультет прикладної математики та інформатики Кафедра математичного моделювання соціально-економічних процесів
Галузь знань, шифр та назва спеціальності	12 – інформаційні технології 124 – системний аналіз
Викладачі дисципліни	Костенко Світлана Борисівна, доцент кафедри математичного моделювання соціально-економічних процесів
Контактна інформація викладачів	Svitlana.kostenko@lnu.edu.ua; https://ami.lnu.edu.ua/employee/kostenko ; Головний корпус ЛНУ ім. І. Франка, каб. 361. м. Львів, вул. Університетська, 1
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/ лабораторних занять (за попередньою домовленістю).
Сторінка курсу	https://ami.lnu.edu.ua/course/dv1-mmsep-intelektualnyy-analiz-danykh
Інформація про дисципліну	Дисципліна “Інтелектуальний аналіз даних” є вибірковою дисципліною з спеціальності 124 – системний аналіз для освітньої програми Системний аналіз, яка викладається в 1-му семестрі в обсязі 5-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Інтелектуальний аналіз даних – дисципліна, призначена для вивчення підходів, що дозволять здійснювати пошук прихованих закономірностей в великих масивах даних з використовуючи методів машинного навчання.
Мета та цілі дисципліни	Метою дисципліни є поглиблення знань з машинного навчання та, завдяки цьому, розширення можливостей інтелектуального аналізу даних. Цілями дисципліни є збільшення інструментарію розв'язання задач класифікації, кластеризації, пошуку асоціативних правил, застосування ансамблей методів, використання відходів та метрик оцінки якості алгоритмів.
Література для вивчення дисципліни	<ol style="list-style-type: none"> 1. Басюк Т. М., Литвин В. В., Захарія Л. М., Кунанець Н.Е.. Машинне навчання: Навчальний посібник. Львів: Видавництво «Новий Світ - 2000», 2019. - 335 с. 2. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с. 3. Провост Ф., Фоусетт Т. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані / пер. з англ. Анастасія Дудченко. — 2-ге вид. — К. : Наш формат, 2020. — 400 с. 4. Deisenroth M. P., Faisal A. A., Soon O. Ch. Mathematics for Machine Learning. - Published by Cambridge University Press, 2020. – 417 5. https://scikit-learn.org/stable/

	<p>Додаткова література та інтернет джерела</p> <p>6. https://www.kdnuggets.com</p> <p>7. https://www.datasciencecentral.com/</p> <p>8. https://www.datacamp.com/</p>
Обсяг курсу	Загальний обсяг: 150 годин. Аудиторних занять: 48 год., з них 32 год. лекцій та 16 години лабораторних робіт. Самостійної роботи: 102 год.
Очікувані результати навчання	<p>Після завершення цього курсу студент буде:</p> <p>знати:</p> <ul style="list-style-type: none"> – метрики та методи оцінювання якості алгоритмів аналізу даних. – особливості застосування методів класифікації, кластеризації об'єктів та пошуку асоціативних правил об'єктів – підходи до пониження розмірності даних та зменшення обсягів даних – засоби мови програмування Python для виконання інтелектуального аналізу даних <p>вміти:</p> <ul style="list-style-type: none"> – обчислювати метрики якості алгоритмів аналізу даних та застосовувати методи оцінки якості алгоритмів аналізу даних – здійснювати препроцесінг даних та пониження розмірностей – будувати асоціативні правила; – застосовувати методи класифікації, а також їх ансамблі; – застосовувати методи кластеризації об'єктів та підбирати оптимальну кількість кластерів; – застосовувати оператори пакету <code>skipy</code> мови програмування Python для виконання інтелектуального аналізу даних.
Ключові слова	Асоціативні правила, задача кластеризації, задача класифікації, ансамблеві методи,
Формат курсу	Очний
Теми	Подані нижче в Схемі курсу
Підсумковий контроль, форма	Залік
Пререквізити	<p>Для вивчення курсу студенти потребують базових знань з дисциплін</p> <ul style="list-style-type: none"> - Машинне навчання; - Лінійна алгебра; - Теорія ймовірностей; - Методи оптимізації - Дискретна математика - Програмування <p>достатніх для сприйняття категоріального апарату дисципліни.</p>
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентації, лекції Індивідуальні завдання
Необхідне обладнання	Комп'ютер із встановленими необхідними для роботи середовищами

Критерії оцінювання (окремо для кожного виду навчальної діяльності)

Оцінювання проводиться за 100-бальною шкалою.

Оцінка ЄКТС	Оцінка в балах	Оцінка за національною шкалою		
		Екзамен, диференційований залік		Залік
A	90-100	5	відмінно	зараховано
B	81-89	4	дуже добре	
C	71-80		добре	
D	61-70	3	задовільно	
E	51-60		достатньо	
FX	21-50	2	незадовільно	не зараховано
F	0-20	2	незадовільно (без права перездачі)	не зараховано (без права перездачі)

Поточне оцінювання: впродовж семестру студент може отримати максимум 100 балів. Впродовж навчання студенти виконують 8 індивідуальних завдань, за виконання, оформлення звіту і захист яких може отримати по 10 балів за кожне, та підсумковий тест на 20 балів.

Критерії оцінювання практичних завдань.

% балів	Критерії оцінювання
80-100 %	студент повністю виконав умови завдання та відповів на всі запитання, пов'язані з тематикою завдання, інтерпретує отримані результати та може запропонувати інші підходи до вирішення поставленого завдання
60-79 %	студент не повністю виконав умови завдання, або не на всі питання відповів в повній мірі, проте підходи до реалізації правильно
40-59 %	студент виконав завдання з помилками, але самостійно їх виправляє, якщо на них вкаже викладач, на деякі запитання, пов'язані з тематикою завдання, відповідає з неточностями, проводить аналіз отриманих результатів з неточностями
20-39 %	студент виконав завдання частково, завдання реалізовано з помилками, які самостійно не може виправити, переважно не відповідає на запитання
1-19 %	студент виконав завдання частково або з грубими помилками, які самостійно не може виправити, демонструє незнання матеріалу
0 балів	студент не виконав завдання.

Для кожного завдання встановлено терміни здачі. Роботи, які здаються із порушенням термінів без поважних причин, оцінюються меншою кількістю балів. Запізнення до 7 днів зменшує отриману кількість балів на 25%, від 8 днів на 50 %.

	<p>Письмові роботи: письмові звіти про виконання індивідуальних завдань.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх видів письмових робіт та індивідуальних завдань, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали набрані за виконання, оформлення та захист індивідуальних завдань та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання	<p>Метрики оцінки задачі класифікації Accuracy (Точність), Precision (Прецизія) Recall (Повнота), F1-оцінка, ROC крива, площа AUC Confusion Matrix (Матриця помилок). Метрики оцінки для задачі класифікації задачі регресії (MAE), Середнє значення квадратів помилок (MSE), Корінь квадратний із середнього значення квадратів помилок (RMSE), Коефіцієнт детермінації (R-squared): Відсоток варіації у залежній змінній, який пояснюється незалежними змінними моделі. Метрики оцінки задачі кластеризації: Silhouette Score, Davies-Bouldin Index, Adjusted Rand Index (ARI). Методи оцінки: Holdout Method, K-fold Cross-Validation, Leave-One-Out Cross-Validation (LOOCV), Bootstrapping. Методи кластеризації. DBSCAN та Affinity Propagation. Асоціативні правила. Алгоритм Apriori. Секвенційний аналіз.</p>
Опитування	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>

СХЕМА КУРСУ

№	Назви змістових модулів і тем	Кількість годин		
		Лекції (год)	Практ. (год)	Сам. р-та
Модуль 1. Місце машинного навчання в аналізі даних та підготовка даних				

1	Інтелектуальний аналіз даних. Оцінювання якості алгоритмів. [1, 2, 3]	4	2	12,5
2	Методи попередньої підготовки даних. Методи пониження розмірності. [2, 3, 4]	4	2	12,5
Модуль 2. Особливості застосування методів машинного навчання для інтелектуального аналізу даних.				
3	Задача класифікації. Метод опорних компонент. [1,2]	4	2	12,5
4	Побудова дерев рішень та регресій. Зупинка алгоритмів. [1,2]	4	2	12,5
5	Асамблеві методи. [1,2,3]	4	2	12,5
6	Тема 6. Алгоритми кластеризації: на основі щільності DBSCAN та поширення подібності з автоматичним виявленням кількості кластерів Affinity Propagation. [1,2,5]	4	2	12,5
Модуль 3. Пошук закономірностей у послідовностях даних				
7	Побудова асоціативних правил. Секвенційний аналіз. [1, 2, 5]	4	2	12,5
8	Засоби мови програмування Python для виконання інтелектуального аналізу даних. [5]	4	2	14,5