

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет прикладної математики та інформатики
Кафедра інформаційних систем

Затверджено

На засіданні
кафедри інформаційних систем
факультету прикладної математики та
інформатики
Львівського національного університету
імені Івана Франка
(протокол № 1 від 28.08.2023 р.)

Завідувач кафедри Шинкаренко Г.А.



Силабус навчальної дисципліни
“Інтелектуальний аналіз даних”,
що викладається в межах ОПП Інформатика
першого (бакалаврського) рівня вищої освіти для здобувачів
за спеціальністю 122 – Комп’ютерні науки

Львів 2023 р.

Назва дисципліни	Інтелектуальний аналіз даних
Адреса викладання дисципліни	Львівський національний університет імені Івана Франка, вул. Університетська 1, м. Львів, Україна, 79000
Факультет та кафедра, за якою закріплена дисципліна	Факультет прикладної математики та інформатики Кафедра інформаційних систем
Галузь знань, шифр та назва спеціальності	Галузь знань: 12 Інформаційні технології Спеціальність: 122 Комп'ютерні науки
Викладачі дисципліни	Козій Ірина Ярославівна, доцент кафедри інформаційних систем; Соколовський Ярослав Іванович, професор кафедри інформаційних систем
Контактна інформація викладачів	iryna.koziy@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/kozij-i-ya Yaroslav.Sokolovskyy@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/sokolovskyy-ya-i Головний корпус ЛНУ ім. І. Франка, каб. 260. вул. Університетська 1, м. Львів, Україна, 79000
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю).
Сторінка курсу	https://ami.lnu.edu.ua/course/intellectual-analysis-of-data
Інформація про дисципліну	“Інтелектуальний аналіз даних” є вибірковою навчальною дисципліною зі спеціальності 122 Комп'ютерні науки для освітньо-професійної програми “Інформатика”, яку викладають у 8-му семестрі в обсязі 4-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Дисципліна “Інтелектуальний аналіз даних” спрямована на дослідження процесів виявлення знань, оволодіння методами та алгоритмами Data Mining. Розглядаються методи та алгоритми вирішення основних завдань аналізу: класифікація, регресія, пошук асоціативних правил, кластеризація. При розгляді цих тем основна увага зосереджується на формулюванні задач та програмуванні відповідних алгоритмів.
Мета та цілі дисципліни	Метою вивчення дисципліни “Інтелектуальний аналіз даних” є освоєння технологій оперативного та інтелектуального аналізу даних з візуалізацією результатів аналізу в процесі розв'язування прикладних задач у галузі комп'ютерних наук, пошуку у необроблених масивах даних раніше невідомих, практично корисних знань та закономірностей, необхідних для прийняття рішень.
Література для вивчення дисципліни	1. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с. 2. Обчислювальні аспекти аналізу даних на основі карт Кохонена. / Годич О. В та ін. Національний університет “Львівська політехніка”, 2011. 63-72 с. 3. Пасічник В.В., Шаховська Н.Б. Сховища даних: Навчальний посібник. – Львів: «Магнолія 2006», 2008. – 496 с. 4. Ситник В. Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. – К: КНЕУ, 2007. – 376 с. 5. Ситник В.Ф. Системи підтримки прийняття рішень: Навч.посібник. –

	<p>К.:КНЕУ, 2009. – 614 с.</p> <p>6. Черняк О.І.. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко ; Київ. нац. ун-т ім. Т. Шевченка. — К. : Знання, 2014. — 599 с.</p> <p>7. Inmon W.H. Building the Data Warehouse, 4th Edition. – Hoboken, NJ:Wiley, 2005. – 576 p.</p> <p>8. Kretowski, M. Evolutionary Decision Trees in Large-Scale Data Mining /M. Kretowski , - Cham: Springer International Publishing, 2019, 180 p.</p> <p>9. McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. O'Reilly Media. September 20, 2022. - 621 p.</p> <p>10. Ranga Suri N.N.R., Murty N., Athithan M.G. Outlier Detection: Techniques and Applications. A Data Mining Perspective / N.N.R. Ranga Suri, N. Murty, M.G. Athithan., - Cham: Springer International Publishing, 2019, 214 p.</p> <p>11. Witten I. Data Mining: Practical Machine Learning Tools and Techniques / Ian Witten, Eibe Frank, Mark A. Hall. – Morgan Kaufmann, 2016. – 654p.</p> <p>12. https://www.coursera.org/learn/machine-learning</p> <p>13. https://www.coursera.org/learn/machine-learning-with-python</p> <p>14. https://www.coursera.org/learn/data-analysis-with-python</p> <p>15. https://prometheus.org.ua/course/course-v1:IRF+ML101+2016_T3</p>
Обсяг курсу	Загальний обсяг: 120 годин. Аудиторних занять: 64 год., з них 32 год. лекцій та 32 год. лабораторних робіт. Самостійної роботи: 56 год.
Очікувані результати навчання	<p>Після завершення цього курсу студент повинен:</p> <p><i>Знати:</i> основні поняття та визначення інтелектуального аналізу даних; поширені методи та алгоритми вирішення основних завдань аналізу; критерії порівняння моделей і методів інтелектуального аналізу даних.</p> <p><i>Вміти:</i> обґрунтовувати вибір засобів і методів для розв'язку конкретних задач аналізу даних; створювати і застосовувати типові математичні моделі для дослідження явищ і процесів за результатами спостережень при розв'язку задач: класифікації, регресії, кластеризації, пошуку асоціативних правил; представляти результати аналізу в зручному для сприйняття вигляді, інтерпретувати їх у відповідності з поставленою задачею.</p>
Компетентності	<p><i>Загальні компетентності (ЗК):</i></p> <p>ЗК2. Здатність застосовувати знання у практичних ситуаціях.</p> <p>ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.</p> <p>ЗК6. Здатність вчитися й оволодівати сучасними знаннями</p> <p>ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</p> <p>ЗК11. Здатність приймати обґрунтовані рішення.</p> <p><i>Спеціальні компетентності (СК):</i></p> <p>СК2. 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.</p> <p>СК11. 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.</p>
Програмні результати навчання	<p>ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.</p>

	<p>ПР12. 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.</p>
Ключові слова	<p>Інтелектуальний аналіз даних Прогноз невідомих значень Аналіз виключень Багатовимірна модель даних Сховище даних, вітрина даних Класифікація Кластеризація Пошук асоціативних правил</p>
Формат курсу	<p>Очний, дистанційний Проведення лекцій, лабораторних робіт і консультацій. Ознайомлення з Internet курсами за даною тематикою Prometeus courses: https://prometeus.org.ua/course/course-v1:IRF+ML101+2016_T3 eDX courses: https://www.edx.org/course/analyzing-data-with-python https://www.edx.org/course/introduction-to-predictive-analytics-using-python https://www.edx.org/course/python-basics-for-data-science https://www.edx.org/course/python-for-data-science-2 або COURSERA courses: https://www.coursera.org/learn/machine-learning https://www.coursera.org/learn/machine-learning-with-python https://www.coursera.org/learn/data-analysis-with-python</p>
Теми	<p>Лекційні заняття:</p> <ol style="list-style-type: none"> 1. Введення до курсу. Поняття інтелектуального аналізу даних. Поняття «дані», «інформація» і «знання». Data Mining як частина ринку інформаційних технологій. 2. Системи підтримки прийняття рішень. Задачі систем підтримки прийняття рішень. Бази даних – основа систем підтримки прийняття рішень. Нефективність використання OLTP систем для аналізу даних. 3. Сховища даних. Концепція сховища даних. Організація СД. Структура систем підтримки прийняття рішень з фізичним СД. Структура систем підтримки прийняття рішень з віртуальним СД. Вітрини даних. ETL процес. Очистка даних. Сховища даних та аналіз. 4. OLAP системи. Багатовимірна модель даних. Визначення OLAP систем. Особливості OLAP систем. Тест FASMI. Архітектура OLAP систем (MOLAP, ROLAP, HOLAP). 5. Data Mining. Класифікація задач Data Mining. Задачі класифікації та регресії. Задача пошуку асоціативних правил. Задача кластеризації. Практичне застосування DataMining. Моделі Data Mining. Передбачувані (predictive), описові (descriptive) моделі. Методи Data Mining. Процес виявлення нових знань. Підготовка вхідних даних. 6. Класифікація та регресія. Постановка задачі. Представлення результатів. Класифікаційні правила. Деревя рішень. Математичні функції. 7. Методи побудови класифікаційних правил. 1R алгоритм. Метод Naive Bayes.

	<p>8. Дерева рішень. Методика “розділай та володарюй” для побудови дерев рішень. Алгоритм максимального покриття.</p> <p>9. Методи побудови математичних функцій. Лінійні методи. Метод найменших квадратів. Нелінійні методи. SVM метод.</p> <p>10. Карта Кохонена.</p> <p>11. Аналіз часових рядів.</p> <p>12. Пошук асоціативних правил. Постановка задачі. Секвенціальний аналіз. Представлення результатів. Алгоритм Apriori та його різновиди.</p> <p>13. Кластеризація. Постановка задачі кластеризації. Представлення результатів. Базові алгоритми кластеризації.</p> <p>14. Ієрархічні алгоритми кластеризації. Неієрархічні алгоритми кластеризації. Чітка та нечітка кластеризація. Алгоритм k-середніх (k-means). Алгоритм Fuzzy c-means.</p> <p>15. Кластеризація за Густавом Кесселем.</p> <p>16. Підсумкова лекція з обговоренням курсу. Підготовка до залікового тесту за матеріалами лекцій.</p> <p>Лабораторні заняття:</p> <p>1. 1R-метод. Представлення результатів.</p> <p>2. Алгоритм Naive Bayes. Застосування класифікатора Баєса для визначення спаму.</p> <p>3. Модифікований деревовидний алгоритм Баєса для вирішення завдань класифікації.</p> <p>4. Алгоритми: ID3, C4.5, CHAID, SIPINA, CN2, CART (Classification and Regression Tree) для побудови дерев рішень.</p> <p>5. Алгоритм максимального покриття для побудови дерев рішень.</p> <p>6. Метод найменших квадратів.</p> <p>7. Метод опорних векторів (SVM).</p> <p>8. Одношаровий перцептрон Розенблатта.</p> <p>9. Методи прогнозування часових рядів.</p> <p>10. Побудова асоціативних правил. Алгоритм Apriori.</p> <p>11. Карта Кохонена.</p> <p>12. Кластеризація даних. Представлення результатів. Алгоритм k-means, fuzzy c-means.</p> <p>13. Ієрархічні алгоритми кластеризації.</p> <p>14. Неієрархічні алгоритми кластеризації.</p> <p>15. Кластеризація за Густавом Кесселем.</p> <p>16. Кластеризація даних за допомогою нечітких відношень.</p>
Підсумковий контроль, форма	Залік у кінці семестру.
Пререквізити	<p>Для вивчення курсу студенти потребують базових знань з</p> <ul style="list-style-type: none"> - Алгебри та геометрії; - Дискретної математики; - Математичного аналізу; - Чисельні методи; - Теорії ймовірності та математичної статистики; - Програмування; - Баз даних та інформаційних систем.
Навчальні методи та техніки, які будуть використовуватися під час	Презентації, лекції. Індивідуальні завдання та проекти.

викладання курсу	
Необхідне обладнання	Комп'ютер із програмним забезпеченням Visual Studio 2019/2022, GNU Octave, Internet доступ до обчислювального кластера.
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> • індивідуальні завдання : 50% семестрової оцінки; максимальна кількість балів 50 • ознайомлення з безкоштовними курсами за даною тематикою в Internetі з виконанням індивідуальних завдань: 30% семестрової оцінки; максимальна кількість балів 30 • підсумковий тест: 20% семестрової оцінки; максимальна кількість балів 20 <p>Підсумкова максимальна кількість балів 100.</p> <p>Письмові роботи: Очікується, що студенти виконають одну письмову роботу (тест з теоретичних завдань) і звіт про виконання проекту.</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 практичні 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 при виконанні індивідуальних проектів, самостійній роботі 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; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання до заліку чи екзамену.	<p>Сховища даних. OLAP системи. Задачі класифікації, регресії, пошуку асоціативних правил та кластеризації. 1R-метод. Алгоритм Naive Bayes. Модифікований деревовидний алгоритм Басса для вирішення завдань класифікації. Алгоритми ID3, C4.5, CHAID, SIPINA, CN2, CART для побудови дерев рішень. Алгоритм покриття для побудови дерев рішень. Метод найменших квадратів.</p>

	<p>Метод опорних векторів. Одношаровий перцептрон Розенблатта. Методи прогнозування часових рядів. Алгоритм Аргіої. Карта Кохонена. Алгоритми k-means, fuzzy c-means. Ієрархічні алгоритми кластеризації. Неієрархічні алгоритми кластеризації.</p>
Опитування	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>