

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра системного проектування

Затверджено

На засіданні кафедри системного програмування
факультету електроніки та комп'ютерних
технологій
Львівського національного університету імені
Івана Франка
(протокол №1 від 28 серпня 2023 р.)

Завідувач кафедри  Роман ШУВАР

Силабус з навчальної дисципліни
«Технології аналізу даних / Technologies of Data Analysis»,
що викладається в межах ОП «Комп'ютерні науки»
другого (магістерського) рівня вищої освіти
для здобувачів зі спеціальності
122 – Комп'ютерні науки

Назва дисципліни	Технології аналізу даних / Technologies of Data Analysis
Адреса викладання дисципліни	м. Львів, вул. Драгоманова, 50
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра системного проектування
Галузь знань, шифр та назва спеціальності	12 Інформаційні технології, 122 Комп'ютерні науки
Викладачі дисципліни	Шувар Роман Ярославович, канд. фіз.-мат. наук, доцент, доцент
Контактна інформація викладачів	roman.shuvar@lnu.edu.ua https://electronics.lnu.edu.ua/employee/shuvar-r-ya-2
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю): кімн. 317, корпус факультету електроніки та комп'ютерних технологій, м. Львів, вул. Драгоманова, 50. Також можливі онлайн консультації через MS Teams або систему електронного навчання Moodle. Для погодження часу онлайн консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	https://electronics.lnu.edu.ua/course/tekhnohohii-analizu-danykh-technologies-of-data-analysis-122-kn http://194.44.208.156/moodle/course/view.php?id=59
Інформація про дисципліну	Дисципліна «Технології аналізу даних» є дисципліною вільного вибору студента зі спеціальності 122 Комп'ютерні науки для освітньої програми «Комп'ютерні науки», яка викладається в 2 семестрі в обсязі 3,0 кредити (за Європейською Кредитно-Трансферною Системою – ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено таким чином, щоб надати учасникам необхідні знання, обов'язкові для того, щоб використовувати сучасні технології та програмні засоби для дослідження та всебічного аналізу даних у різних галузях сферах людської діяльності. А також для попередньої обробки неідеальних реальних даних, для запису даних у відповідні структури даних, реалізовувати інтерактивні візуалізації даних, проводити необхідну попередню дослідницьку обробку даних, визначати тип задачі аналізу, використовувати нейромережі та машинне навчання для вирішення задач аналізу даних з оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію аналізу даних; пошук неочевидних закономірностей і здатність самостійно будувати гіпотези про взаємозв'язки досліджуваних даних.
Мета та цілі дисципліни	<i>Мета</i> викладання навчальної дисципліни “Технології аналізу даних” полягає у забезпеченні знайомства студентів з особливостями використання технологій аналізу даних для дослідження структурованих та неструктурованих даних. Її <i>цілі</i> – вивчити алгоритмів кластеризації, класифікації, регресії; опанувати теоретичний матеріал і практичне оволодіння сучасними графічно-інформаційними технологіями, комп'ютерними та програмними засобами створення цілісного аналізу даних, розробки моделей даних, подання їх в графічній формі, визначення статистичних параметрів даних. ознайомлення з базовими концепціями обробки даних, які дозволять правильно структурувати дані для подальшого їхнього опрацювання, візуалізації і моделювання, управління програмною інфраструктурою та інтерфейсом систем обробки даних, теорією і проектуванням систем аналізу даних.
Література для вивчення дисципліни	Основна література: 1. Литвин В. В. Методи та засоби інженерії даних та знань / В. В.

	<p>Литвин // навчальний посібник з грифом МОНУ. — Львів : «Магнолія-2006», 2012. — 241 с.</p> <p>2. Han, Jiawei. Data mining : concepts and techniques / Jiawei Han, Micheline Kamber, Jian Pei. – 3rd ed. – 2012. ISBN 978-0-12-381479-1 Chapter 3. Data preprocessing</p> <p>3. Michael R. Brzustowicz Data Science with Java Practical Method for scientists and engineers /Michael R. Brzustowicz. – O'REILLY, 2017. – 311p.</p> <p>4. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів /В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.</p> <p>5. Методичні вказівки до виконання лабораторних робіт з дисципліни “Інженерія прикладних інтелектуально-орієнтованих програмних продуктів” для студентів спеціальностей 121 “Інженерія програмного забезпечення” та 122 “Комп’ютерні науки та інформаційні технології” (всіх форм навчання) / В.М. Льовкін. – Запоріжжя : ЗНТУ, 2016. – 80 с.</p> <p>6. Інтелектуальний аналіз даних: Комп’ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп’ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 1,72 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 73 с.</p> <p>7. Литвин В.В. , Нікольський Ю.В., Пасічник В.В. Аналіз даних та знань. Навчальний посібник. – Видавництво Магнолія 2006, 2021р. – 276 с.</p> <p>8. Бідюк П. І. Байєсівський аналіз даних : монографія / П. І. Бідюк, І. О. Калініна, О. П. Гожий. – Херсон: Книжкове видавництво ФОП Вишемирський В.С., 2022. – 208 с.</p>
Обсяг курсу	Сумарно 90 год., з них 16 год. лекцій, 16 год. лабораторних робіт та 58 год. самостійної роботи.
Очікувані результати навчання	Після завершення цього курсу студент буде використовувати сучасні програмні засоби для створення моделей реальних даних, для запису даних у відповідні структури даних, реалізовувати інтерактивні візуалізації даних, проводити необхідну попередню обробку даних, визначати тип задачі аналізу, вирішувати її адекватно обраним методом з оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію;
Ключові слова	Типи даних, класифікація, кластеризація та регресія даних
Формат курсу	Очний
	Проведення лекцій, лабораторних робіт та консультації для кращого розуміння тем
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Залік в кінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань у галузі 12 – Інформаційні технології, зокрема з дисциплін «Вища математика», «Дискретна математика», «Алгоритми та структури даних», «Чисельні методи».
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентація, лекції, лабораторні роботи, обговорення, дискусія.

Необхідне обладнання	Мультимедіа, платформа Moodle, комп'ютерне програмне забезпечення: вільне програмне забезпечення Jupyter Notebook, середовище PyCharm, онлайн-сервіс Google Colab
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою. Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • 15 лабораторних робіт: 50% семестрової оцінки; максимальна кількість балів $15 \times 3 + 5 = 50$, де 5 додаткових балів надається за якість захистів. • 2 контрольні заміри (модулі): 50% семестрової оцінки; максимальна кількість балів $2 \times 25 = 50$. <p>Загалом упродовж семестру 100 балів.</p> <hr/> <p>Контрольні заміри проводяться у формі тестових завдань. Академічна доброчесність: Очікується, що лабораторні та контрольні 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 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 і лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Студенти зобов'язані дотримуватися усіх термінів визначених для виконання усіх видів робіт, 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 на поточному тестуванні, самостійній роботі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час лабораторного 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. Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання до контрольних робіт	Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт розміщені на веб-сторінці http://194.44.208.156/moodle/course/view.php?id=59
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання, лабораторна робота, самостійна робота, год.	Термін виконання
1	Пошук даних. Сучасні методи аналізу даних та їх практичне	Лекція	1, 7	Вступне заняття. Інструкція з техніки	2 тиж. семестру

	застосування. Формати зберігання даних. Класи, типи, об'єкти. Джерела даних: таблиці, файли, бази даних, web-сервіси. Типи і підходи до збору даних. Основні особливості і недоліки сучасних реальних даних.			безпеки. Академічна доброчесність. Методи бібліотеки pandas та scilearn для роботи з даними	
2,	Дослідницький аналіз даних. Етапи дослідницького аналізу даних. Методи для визначення основних характеристик даних.	Лекція	1, 2, 5,	Набори даних для машинного навчання та їх опис.	3 тиж. семестру
3	Поняття кластеризації даних та методи кластеризації даних; k-means, агломеративний алгоритм, DBSCAN.	Лекція	1, 2, 5,	Кластеризація даних. Порівняльний аналіз методів кластеризації	5 тиж. семестру
4	Класифікація даних. Алгоритми класифікації даних: дерева прийняття рішень, випадковий ліс, метод опорних векторів	Лекція	2, 5,	Створення моделі класифікація даних	7 тиж. семестру
5	Класифікація даних за допомогою нейромереж (Keras та PyTorch)	Лекція	2, 5,	Створення моделей класифікації за допомогою нейромереж	8 тиж. семестру
6	Попередній аналіз даних в задачах прогнозування попиту. Типи даних.	Лекція	1, 2,	Дослідницький аналіз даних	9 тиж. семестру
7	Візуалізація даних як етап аналізу даних	Лекція	1, 5,	Візуалізація даних як етап аналізу даних.	10 тиж. семестру
8	Коваріація та кореляція даних. Кореляційний аналіз кількісних ознак	Лекція	2, 31	Коваріація та кореляція даних.	12 тиж. семестру
9	Business intelligence – інструменти, технології і процеси, які використовуються для перетворення даних в інформацію.	Лекція	1, 2, 5	Business intelligence – інструменти, технології і процеси.	13 тиж. семестру
10	Прогнозування числових рядів за допомогою машинного навчання. Байєсівська регресія.	Лекція	1, 2, 4,8	Прогнозування числових рядів за допомогою машинного навчання.	14 тиж. семестру
11	Лінійна регресія на прикладі прогнозування продажів	Лекція	2, 5,	Прогнозування цін на нерухомість	15 тиж. семестру
12	Графи та структури даних на графах. Представлення та візуалізація даних за допомогою графів	Лекція	2, 3, 4, 5	Графи та структури даних на графах.	16 тиж. семестру
13	Особливості роботи з сирими гав даними. Чисельні системи представлення даних та частота дискретизації даних	Лекція	2. 4. 5	Особливості роботи з сирими гав даними.	16 тиж. семестру
14	Редукція та класифікація даних. Факторний аналіз. Метод головних компонент. Стекінг моделей. Крос-	Лекція	6,7	1. Редукція та класифікація даних	16 тиж. семестру

	валідація				
15	Аналітика текстових даних. Класифікація текстових даних.	Лекція	1, 2, 4	Класифікація текстових даних.	16 тиж. семестру
16	Класифікація зображень. Знаходження об'єктів на зображеннях.	Лекція	3, 7	Класифікація зображень.	16 тиж. семестру