

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

Затверджено

На засіданні кафедри КСП
факультету електроніки та комп'ютерних
технологій
Львівського національного університету
імені Івана Франка
(протокол № 1 від 30.08. 2022 р.)

Завідувач кафедри



Роман Шувар

Силабус з навчальної дисципліни
«Методи та технології інженерії даних»,
що викладається в межах ОПП
“Інженерія програмного забезпечення”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 121 – Інженерія програмного забезпечення

Назва дисципліни	Методи та технології інженерії даних
Адреса викладання дисципліни	Корпус факультету електроніки та комп'ютерних технологій, Львівський національний університет імені Івана Франка, вул. Драгоманова 50, м. Львів, 79005, вул. Ген. Тарнавського 107, м. Львів, 79011
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 121 – Інженерія програмного забезпечення
Викладачі дисципліни	Павлишенко Б.М., докт. техн. наук, професор Демків Л.С., канд. фіз.-мат. наук, доцент, доцент
Контактна інформація викладачів	bohdan.pavlyshenko@lnu.edu.ua lidiya.demkiv@lnu.edu.ua, https://electronics.lnu.edu.ua/employee/demkiv-l-s
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю). Для погодження часу онлайн консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	https://moodle.elct.lnu.edu.ua/course/view.php?id=176
Інформація про дисципліну	Дисципліна «Методи та технології інженерії даних» є дисципліною з спеціальності 121 – Інженерія програмного забезпечення для освітньої програми «Інформаційні системи та технології», яка викладається в 5-му семестрі в обсязі 4-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Мета та цілі дисципліни	Мета: реалізувати знайомство студента із способами створення цілісних та чистих даних, теоретичними принципами галузі інженерії даних та практичними технологіями, які найбільш затребувані в інженерії даних. Цілі: забезпечити знайомство студентів з особливостями обробки та зберігання різних типів даних, варіантами побудови сховищ даних; вивчити алгоритми попередньої обробки даних; опанувати теоретичний матеріал і практичне оволодіння сучасними графічно-інформаційними технологіями, комп'ютерними та програмними засобами створення цілісних даних, подання їх в графічній формі; вивчити методи і алгоритми обробки даних, визначити статистичні параметри даних; ознайомити 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 обробки даних.
Література для вивчення дисципліни	Основна література: 1. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем»)/ Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с 2. Литвин В. В. Методи та засоби інженерії даних та знань / В. В. Литвин // навчальний посібник з грифом МОНУ. — Львів : «Магнолія-2006», 2012. — 241 с.

	<p>3. Han, Jiawei. Data mining : concepts and techniques / Jiawei Han, Micheline Kamber, Jian Pei. – 3rd ed. ISBN 978-0-12-381479-1 Chapter 3. Data preprocessing 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. Extract, transform, and load (ETL) // Електронний ресурс. Режим доступу: https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl</p> <p>7. Edward L.Robinson Data Analysis for Scientists and Engineers // Pricenton University Press, 2016, - P.408, ISBN 9781400883066</p>
Обсяг курсу	64 години аудиторних занять. З них 32 години лекцій, 32 години лабораторних робіт та 56 година самостійної роботи
Очікувані результати навчання	<p>Після завершення цього курсу студент буде використовувати сучасні програмні засоби для попередньої обробки неідеальних реальних даних, запису даних у відповідні структури та сховища даних, моделювання та інтеграції даних; реалізовувати інтерактивні візуалізації даних; проводити необхідну попередню обробку даних для отримання чистих даних; визначати тип задачі аналізу та вирішувати її адекватно обраним методом з оптимально визначеними параметрами; оцінювати результати; робити змістовні висновки та інтерпретацію опрацювання даних.</p> <p>Після вивчення даного курсу здобувачі набудуть таких Загальних та Фахових компетентностей та Програмних результатів навчання:</p> <p>ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.</p> <p>ЗК2. Здатність застосовувати знання у практичних ситуаціях.</p> <p>ЗК6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</p> <p>ФК16. Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.</p> <p>ФК18. Здатність аналізувати, вибирати і застосовувати методи і засоби для забезпечення інформаційної безпеки (в тому числі кібербезпеки).</p> <p>ФК19. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.</p> <p>ФК20. Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв’язання завдань інженерії програмного забезпечення.</p> <p>ФК28. Володіння методами сучасних веб-технологій, хмарних технологій, великих даних та засобами розробки веб-застосунків.</p> <p>ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.</p> <p>ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.</p> <p>ПРН21. Знати, аналізувати, вибирати, кваліфіковано застосовувати засоби</p>

	забезпечення інформаційної безпеки (в тому числі кібербезпеки) і цілісності даних відповідно до розв'язуваних прикладних завдань та створюваних програмних систем. ПРН26. Знати та вміти застосовувати засоби інженерії програмного забезпечення для реалізації проєктів з використанням технологій науки про дані та штучного інтелекту.
Ключові слова	Структури даних, ETL, DATA WAREHOUSE, DATA LAKE
Формат курсу	Очний
	Проведення лекцій, лабораторних робіт та консультації для кращого розуміння тем
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Іспит в кінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань з дисциплін «Вища математика», «Дискретна математика», «Алгоритми та структури даних», «Бази даних».
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
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"> • лабораторні роботи: 50% семестрової оцінки; максимальна кількість балів 50. • Виконання теоретичних і практичних завдань на іспиті: 50% семестрової оцінки; максимальна кількість балів 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>Відвідування 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</p>

	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.
Питання до контрольних робіт	Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт розміщені на веб-сторінці.
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), год	Термін виконання
1	Поняття даних. Визначення кількості даних та проблеми швидкого збільшення кількості даних в сучасному світі. Професії, що працюють з даними. Відкриті дані — це цінний ресурс, який допомагає посилити цифрову та "реальну" економіку країни. Засоби вилучення, перетворення і завантаження даних. Поняття ETL, pipeline, datalake, data warehouse.	Лекція	1, 2, 5	Вступне заняття. Інструкція з техніки безпеки. Методи бібліотеки pandas для роботи з даними	2 тиж. семестру
2,	Формати зберігання даних. Джерела даних: таблиці, файли, бази даних, web-сервіси. Зчитування та запис файлових даних. Типи даних за шкалами вимірювання.	Лекція	1, 3, 5, 6	Вивчення методів бібліотеки pandas (python) для роботи з даними: сортування, групування, фільтрація, об'єднання даних. Основи ВІ аналітики.	6 тиж. семестру
3	Візуалізація даних як етап аналізу даних	Лекція	1, 5	Підготовка даних до візуалізації. Візуалізація даних.	6 тиж. семестру
4	Коваріація та кореляція даних. Кореляційний аналіз кількісних ознак. Побудова рекомендаційних систем.	Лекція	2, 3	Коваріація та кореляція даних.	7 тиж. семестру
5	Статистичний аналіз даних. Обчислення описової статистики даних. Статистичні моменти даних. Роботи з великою кількістю даних (оновлення статистики даних) Нормування та	Лекція	1, 2, 5, 6	Статистичний аналіз даних	7 тиж. семестру

	стандартизація даних				
6	Розподіли числових даних. Нормальний розподіл даних. Види розподілів. Перевірка гіпотез. Види гіпотез. P-value. Параметричні та непараметричні тести.		2, 6	Захисне заняття	7 тиж. семестру
7	Дослідницький аналіз даних. Етапи дослідницького аналізу даних.	Лекція	2, 5,	Дослідницький аналіз даних.	9 тиж. семестру
8	Створення даних для бізнес аналітики (BI)	Лекція	7	Побудова дашбордів за допомогою Tableau	10 тиж. семестру
9	Підготовка даних до класифікації. Редукція та класифікація даних. Метод головних компонент.	Лекція	2,4,7	Підготовка даних до класифікації та ML	11 тиж. семестру
10	Побудова дерев класифікації. Random Forest. Порівняння методів класифікації.	Лекція	1, 2, 4	Редукція та класифікація даних	12 тиж. семестру
11	Доступ до баз даних за допомогою модулів python. API Python для роботи з базами даних MySQL, SQLite, Принципи побудови DataLake.	Лекція	1, 2, 4	Доступ до баз даних за допомогою модулів python	14 тиж. семестру
12	Поняття ETL. Принципи побудови сховищ даних DataWareHouse.		6,7	Моделювання побудови сховищ даних DataWareHouse	14 тиж. семестру
13	Робота з нереляційними базами даних. JSON-формат та особливості його використання. Створення та парсинг формату. Типи даних json та масиви. Особливості зчитування json у python. Підготовка даних до кластеризації.	Лекція	2, 5,	JSON-формат та особливості його використання..	15 тиж. семестру
14	Графові бази даних. 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
15	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
16	Pipeline - трубопровід для опрацювання даних. Схеми побудови трубопроводів. Зберігання та оброблення даних в розподілених файлових системах. Особливості роботи з BigData.	Лекція	4, 5, 6, 7	Підсумкове заняття	16 тиж. семестру

