МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Фізичний факультет
Кафедра фізики металів

Затверджено
На засіданні кафедри фізики металів
фізичного факультету
Львівського національного університету
імені Івана Франка
(протокол № 9 від 29.08.2022 р.)

Завідувач кафедри проф. Мудрий С. І.

Силабус
з навчальної дисципліни «Обробка і аналіз даних», що викладається в межах ОПП «Нанофізика та наноматеріали» першого (бакалаврського) рівня вищої освіти для здобувачів з спеціальності 105 Прикладна фізика та наноматеріали

Львів 2022

Назва дисципліни	Обробка і аналіз даних
Адреса викладання дисципліни	вул. Кирила і Мефодія, 8, 79005, м. Львів
Факультет та кафедра, за якою закріплена дисципліна	фізичний факультет, кафедра фізики металів
Галузь знань, шифр та назва спеціальності	Галузь знань – 10 Природничі науки Спеціальність – 105 Прикладна фізика та наноматеріали
Викладач дисципліни	Лектор: Никируй Юлія Семенівна, доцент кафедри фізики металів, к.ф.-м.н.; лабораторні заняття проводить: Штаблавий Ігор Іванович, доцент кафедри фізики металів, д.ф.-м.н.
Контактна інформація викладача	Yuliya.nykyruy@lnu.edu.ua https://physics.lnu.edu.ua/employee/nykyruj-yu-s ihor.shtablavyi@lnu.edu.ua https://physics.lnu.edu.ua/employee/shtablavyj-i-i
Консультації з курсу відбуваються	Консультації в день проведення лекцій та лабораторних занять (за попередньою домовленістю). Також можливі консультації через електронну пошту або он-лайн засобами Microsoft Teams, Telegram.
Сторінка курсу	https://physics.lnu.edu.ua/course/obrobka-i-analiz-danykh-105-prykladna-fizyka-ta-nanomaterialy-op-nanofizyka-ta-nanomaterialy
Інформація про дисципліну	Дисципліна «Обробка і аналіз даних» є вибірковою дисципліною для підготовки бакалавра за спеціальністю 105 Прикладна фізика та наноматеріали, яка викладається в 8 семестрі в обсязі 4,0 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Дисципліну «Обробка і аналіз даних» розроблено таким чином, щоб надати відповідні теоретичні знання, уміння, навички, загальні та фахові компетентності для розв'язання комплексних проблем у галузі науки про дані. Дисципліна розглядає найбільш поширені методи та алгоритми аналізу даних.
Мета та цілі дисципліни	Метою дисципліни «Обробка і аналіз даних» є навчання студентів основних питань теорії та практики використання методів обробки інформації для підвищення ефективності наукових досліджень та формування наукового уявлення про сучасні методи аналізу даних Цілі вивчення дисципліни: набуття знань про принципи та алгоритми, що лежать в основі сучасних систем аналізу даних; оволодіння технологіями та методиками збору, попередньої підготовки та аналізу експериментальних даних; набуття практичних навичок роботи з конкретними програмними та мовами програмування - засобами аналізу даних.
Література для вивчення дисципліни	**Базова:** 1. Статистична обробка експериментальних даних: Навчальний посібник / О.П. Мельниченко, І.Л. Якименко, Р.Л. Шевченко − Біла Церква, 2006.− с. 2. Методи аналізу даних: навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с **Допоміжна:** 1. Копей В. Б. Мова програмування Python для інженерів і науковців: Навчальний посібник / В. Б. Копей - Івано-Франківськ: ІФНТУНГ, 2019. - 275 с. 2. Маттес Е. Пришвидшений курс Python. – Львів : ВСЛ, 2021. 3. Селіверстов Р., Мельничин А. Основи програмування мовою Python: навч. посібник. – Львів : ЛНУ імені Івана Франка, 2020.

	Додаткові матеріали також буде запропоновано для кожної теми окремо.
Тривалість курсу	один семестр
Обсяг курсу	**120** годин, з яких 48 годин аудиторних занять (з них 16 годин лекцій, 32 годин лабораторних занять та 72 годин самостійної роботи).
Очікувані результати навчання	В результаті вивчення даного курсу студенти повинні **знати:** Елементи теорії імовірності, основи статистичного аналізу, кореляційного, регресійного та кластерного аналізу даних; основні алгоритми обробки даних, формати збереження даних. **вміти:** Вміти застосовувати базові математичні знання, з аналітичної геометрії, лінійної алгебри, математичного аналізу, теорії ймовірностей та математичної статистики, математичного моделювання. Вміти обробляти та зберігати дані, володіти основними методами аналізу даних. Мати навички роботи із сучасною обчислювальною технікою, вміти використовувати стандартні пакети прикладних програм і програмувати на рівні, достатньому для реалізації чисельних методів розв'язування фізичних задач, комп'ютерного моделювання фізичних та астрономічних явищ і процесів, виконання обчислювальних експериментів. Володіти навиками програмувати використовуючи мови високого рівня
Ключові слова	Обробка даних, статистичний аналіз даних, випадкові величини, кластеризація, регресія, кореляція
Формат курсу	Очний
	проведення лекцій, лабораторних занять і консультації для кращого розуміння тем
Теми	Наведено у таблиці 1, таблиці 2
Підсумковий контроль, форма	Підсумковий контроль: залік в кінці семестру.
Пререквізити	Для вивчення курсу студенти повинні знати основні закони та поняття з курсів фізики та математики, вміти застосовувати набуті раніше знання з курсів математичного аналізу, диференційних рівнянь та комп'ютерних технологій для розв'язку практичних завдань; володіти навиками пошуку та опрацювання спеціалізованої літератури, розв'язку алгебраїчних і диференційних рівнянь, побудови та аналізу графічних залежностей.
Навчальні методи та техніки, які буде використано під час викладання курсу	Презентації, лекції, робота за комп'ютером.
Необхідне обладнання	персональний комп'ютер, загальновживані комп'ютерні програми й операційні системи, проєктор
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням: • лабораторні заняття: 80% семестрової оцінки; максимальна кількість балів 80 • контрольна робота 20% семестрової оцінки. Максимальна кількість балів 20. Підсумкова максимальна кількість балів 100. **Академічна доброчесність** здобувачами вищої освіти передбачає самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання. Списування, втручання в

	роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій чи усній роботі студента є підставою для її незарахування викладачем, незалежно від масштабів обману. **Відвідання занять** є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і практичні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися усіх строків, визначених для виконання усіх видів робіт, передбачених курсом. **Література**. Усю література, яку студенти не зможуть знайти самостійно, буде надано викладачами виключно в освітніх цілях без права її передавання третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих. **Політика виставлення балів**. Враховуються бали, набрані на практичних та поточному тестуванні. При цьому обов'язково враховуються присутність на заняттях та активність студента під час заняття; недопустимість пропусків та запізнень на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях, не пов'язаних із навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін. Жодні форми академічної недоброчесності не толеруються.
Питання на контрольні роботи	1. Елементи теорії імовірності. Імовірність події, густина імовірності, теорема додавання ймовірностей. Які події називають сумісними/несумісними, неможливими, рівноможливими? Розподіл Гауса (запишіть формулу, зобразіть графічно). 2. Статистичний аналіз. Нульова і альтернативна гіпотези. Що таке генеральна сукупність та вибірка, репрезентативна/не репрезентативна вибірка. Що таке надійна імовірність, яких значень найчастіше задають надійній імовірності. Які є міри центральної тенденції? Що таке нормальний розподіл, зобразіть графічно, які його характеристики. Одномодальний, бімодальний розподіл (дайте визначення, зобразіть графічно) 3. Кореляційний аналіз Що таке кореляція? В чому полягає кореляційний аналіз? Як знайти коефіцієнт кореляції Пірсона (формула)? Якими інструментами (програмними засобами) можна знайти коефіцієнт кореляції Пірсона? Якими ще коефіцієнтами користуються в кореляційному аналізі? 4. Регресійний аналіз Що таке регресія, які види регресії розрізняють? Чим відрізняється регресійний аналіз від кореляційного аналізу? Як знайти параметри регресії (формула)? Якими інструментами (програмними засобами) можна знайти параметри регресії? Коли регресію можна використовувати для прогнозування? 5. Кластерний аналіз. У чому полягає кластерний аналіз? З яких етапів складається процес кластеризації? Які є труднощі (проблеми) кластеризації? До якого методу машинного навчання відноситься кластеризація, для яких прикладних задач може бути використана? Які методи кластеризації ви знаєте? Опишіть один або кілька методів (алгоритмів) кластеризації. Якими інструментами (програмними засобами) можна виконати кластеризацію?
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

Таблиця 1

Схема курсу «Обробка і аналіз даних»

Тижні	Назва теми (перелік питань)	Форма діяльності та обсяг годин	Термін виконання
1	Вступ. Предмет науки про дані (Data Science). Теорія ймовірності. Випадкові величини.	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
2	Описова статистика. Розподіли випадкових величин.	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
3	Кореляційний та регресійний аналізи	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
4	Структури даних. Алгоритми обробки даних. Збереження даних. Бази даних.	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
5	Середовище Python для обробки і аналізу даних	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
6	Багатовимірні масиви і матриці, математичні функції і операції з цими масивами. Бібліотека NumPy, Pandas.	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
7	Візуалізація даних як елемент аналізу. Візуалізація даних в Python.	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень
8	Класифікація і кластеризація даних. Базові алгоритми кластеризації	Лекції – 2 год. самостійна робота – 9 год.	1 тиждень

Таблиця 2

Теми лабораторних занять

Тижні	Назва теми	Форма діяльності та обсяг годин	Термін виконання
1	Вступне заняття	Лабораторні – 2 год.	1 тиждень
2,3	Використання MS Excel для первинної обробки даних	Лабораторні – 4 год.	2 тижні
4, 5, 6	Основи програмування мовою PYTHON. Теорія ймовірності і комбінаторика.	Лабораторні – 6 год.	3 тижні
7, 8, 9	Статистична обробка даних.	Лабораторні – 6 год.	3 тижні
10	Проміжне підсумкове заняття	Лабораторні – 2 год.	1 тиждень
11, 12	Регресійний аналіз даних	Лабораторні – 4 год.	2 тижні
13,14,15	Кластерний аналіз даних	Лабораторні – 6 год.	3 тижні
16	Підсумкове заняття	Лабораторні – 2 год.	1 тиждень