**Міністерство Освіти І НАУКИ України**

**Національний університет "Львівська політехніка"**

Інститут **КНІТ**

Кафедра **ПЗ**

### ЗВІТ

До лабораторної роботи № 1

**З дисципліни:** *“Технології ШІ в інженерії даних”*

**На тему:** *“Вирішення класичних задач аналітики передбачення”*

**Лектор:**

асист. каф. ПЗ

Капковський Ю.І.

**Виконав:**

ст. гр. ПЗ-45

Хруставчук М.Л.

**Прийняла:**

асист. каф. ПЗ

БаускО.Є.

« \_\_\_\_ » \_\_\_\_\_\_\_\_ 2025 р.

∑= \_\_\_\_\_ .

Львів – 2025

**Тема роботи:** Вирішення класичних задач аналітики передбачення.

**Мета роботи:** Отримати практичні навички роботи з класичним машинним навчанням використовуючи алгоритми регресії і класифікації.

**TЕОРЕТИЧНІ ВІДОМОСТІ**

.

**ЗАВДАННЯ**

Підзадача №1

1. Розробити ПЗ яке використовує вхідний датасет, вчить на ньому алгоритм поліноміальної регресії, потім демонструє результати порівняння роботи навченої моделі з test датасет.
2. Зробити аналогічне до пункту 1, але на базі алгоритму Elastic Net.
3. Зробити звіти порівняння отриманих результатів двома регресіями і сформулювати головні три висновки.

Підзадача №2

1. Розробити ПЗ яке використовує вхідний датасет, вчить на ньому алгоритм класифікації (логістична регресія), потім демонструє результати порівняння роботи навченої моделі з test датасет.

**ХІД ВИКОНАННЯ**

*Бізнес-задача та обґрунтування вибору датасету*

*Бізнес-задача*

У сучасній виноробній індустрії якість продукції має вирішальне значення для конкурентоспроможності виробника. Традиційно якість вина оцінюється експертами-дегустаторами, проте цей процес є дорогим, тривалим і суб’єктивним. Бізнес-цілі полягають у створенні програмного забезпечення, яке дозволяє на основі фізико-хімічних показників вина (кислотність, вміст цукру, рН, концентрація сульфатів, рівень алкоголю тощо) автоматично прогнозувати його якість.

У рамках завдання можна виділити дві підзадачі, які безпосередньо відповідають вимогам лабораторної роботи:

1. Регресійна аналітика
   * Використання поліноміальної регресії для моделювання нелінійних залежностей між показниками (наприклад, взаємодія алкоголю і кислотності).
   * Використання Elastic Net регресії для відбору найбільш значущих факторів серед великої кількості ознак і уникнення переобучення.
   * Бізнес-результат: можливість передбачити числовий рейтинг вина (від 0 до 10), що дозволить виробнику контролювати якість ще на етапі виробництва.
2. Класифікація (логістична регресія)
   * Перетворення задачі в класифікаційну: наприклад, прогнозування того, чи належить вино до «високоякісних» (7–10 балів), «середніх» (5–6) чи «низьких» (менше 5).
   * Бізнес-результат: автоматичне визначення категорії якості допоможе швидко сегментувати продукцію для різних ринкових сегментів і приймати рішення щодо ціноутворення чи маркетингових стратегій.

*Обґрунтування вибору датасету*

Для реалізації описаної бізнес-задачі обрано [Wine Quality Dataset](https://archive.ics.uci.edu/dataset/186/wine+quality). Він містить:

* + 4898 записів для білого вина та 1599 записів для червоного, що разом дає понад 6 тисяч прикладів – цього достатньо для статистично значущого навчання моделей.
  + 11 числових ознак, які описують хімічні характеристики вина. Це створює як можливості для пошуку нелінійних залежностей (поліноміальна регресія), так і потребу у відборі релевантних змінних (Elastic Net).
  + Цільова змінна – оцінка якості (0–10), яка підходить як для регресії, так і для класифікації після категоризації.

Отже**,** використання Wine Quality Dataset є доцільним, оскільки воно дає можливість продемонструвати роботу алгоритмів поліноміальної регресії, Elastic Net і логістичної регресії на практичній бізнес-проблемі, яка має високу прикладну цінність для харчової промисловості.

**ВИСНОВКИ**

На цій лабораторній роботі я ознайомився з джерелами та застосуванням випадкових чисел, алгоритмами генерування псевдовипадкових чисел для використання в системах захисту інформації.