## Комплексне контрольне завдання

### **Задача 1: Класифікація фізичної активності на основі даних з акселерометра**

* **Дані:** [Data for activity recognition](https://www.kaggle.com/datasets/kosovanolexandr/data-for-activity-recognition/data)
* **Завдання**: Побудувати модель, яка розпізнає тип фізичної активності (наприклад, ходьба чи біг) за даними з акселерометра.

### **Задача 2: Класифікація фейкових новин**

* **Дані:** [Fake News Classification](https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification)
* **Завдання**: Побудувати модель, яка класифікує новини як **правдиві** або **фейкові**.

### **Задача 3: Прогноз виживання пасажирів Титаніка**

* **Дані:**[Titanic Survival Prediction Dataset](https://www.kaggle.com/datasets/yasserh/titanic-dataset)
* **Завдання**: Побудувати модель, яка прогнозує, чи вижив пасажир, на основі характеристик (вік, клас, стать, ціна квитка, кількість родичів на борту тощо).

### 

### **Структура роботи (обов’язкові етапи):**

1. **Огляд та попередній аналіз даних:**
   * Завантаження даних.
   * Перевірка наявності пропусків, базова статистика.
2. **Візуалізація:**
   * Для акселерометра: побудувати графіки зміни значень по осях, розподіли активностей.
   * Для фейкових новин: аналіз довжини текстів, частоти слів, wordcloud.
   * Для Titanic Dataset: візуалізувати залежності між характеристиками та виживанням (наприклад, розподіли за віком і статтю, heatmap кореляцій, bar chart виживання по класах, статі тощо).
3. **Попередня обробка:**
   * Нормалізація/скейлінг (для числових).
   * Токенізація/TF-IDF/ембедінги (для тексту).
4. **Моделювання:**
   * Побудова базової моделі (Logistic Regression / Random Forest / SVM / інше).
   * Порівняння з іншими моделями.
5. **Оцінка моделі:**
   * Розділення на train/test.
   * Метрики: accuracy, precision, recall, F1-score, confusion matrix.
6. **Висновки:**
   * Що вдалося? Що можна покращити?

### 

### **📊 Критерії оцінювання (30 балів):**

| **Критерій** | **Макс. балів** | **Опис** |
| --- | --- | --- |
| **1. Візуалізація даних** | 5 | Наявність змістовних графіків, що допомагають зрозуміти дані |
| **2. Якість попередньої обробки** | 5 | Коректна обробка пропусків, масштабування/токенізація |
| **3. Побудова моделей** | 5 | Чітке пояснення вибору моделі, реалізація без помилок |
| **4. Оцінка результатів** | 5 | Використання відповідних метрик, інтерпретація результатів |
| **5. Якість коду та оформлення** | 5 | Чистий, зрозумілий код, коментарі, логічна структура ноутбуку |
| **6. Висновки та аналітика** | 5 | Глибина розуміння результатів, ідеї для покращення |