

# Data Science Semester Projects

## Overview

This semester project will be completed in groups of two (2) students (no exceptions).

- Each group must choose their project topic as soon as possible.
- Only 11 groups maximum may select the same topic.
- Groups may be cross-section (students from different sections can collaborate).
- Detailed iteration deliverables and submission dates will be announced soon.
- You are required to select one project topic from the five listed options.
- Topics will be allocated on a first-come, first-served basis, so make your choice early.
- The proposed datasets and approaches are provided only as guidelines to help you get started.
- You may explore alternative methods or techniques as long as they align with the project's main objectives.
- **If you wish to use a different but relevant dataset, you must get it approved in advance before starting your implementation.**
- For any queries regarding the project, email me at [l227527@lhr.nu.edu.pk](mailto:l227527@lhr.nu.edu.pk)
- Explore kaggle if gpu required else u can also work locally on vs code or google colab

Each project is designed to span three structured iterations focusing on core data science skills from data wrangling and preprocessing to modeling and evaluation.

Students may select a topic of their choice based on their interest in research or development.

## Project Topics

### 1. Student Behaviour Detection (Video Analytics)

This project focuses on developing an AI-based video analytics system to detect and analyze different types of student behavior in classroom environments. The goal is to identify actions such as attentiveness, distraction, use of mobile phones, or other forms of inattentive behavior from classroom video footage. Using the [Classroom Behavior Dataset](#), you can train and evaluate your model to recognize and categorize these behaviors. Since the dataset is quite large, **you may use a smaller subset of the data** for practical experimentation and testing.

### 2. Eye Contact and Engagement Analysis

The Eye Contact and Engagement Analysis project focuses on developing a system that can automatically analyze a person's level of attention and engagement during online classes or meetings using computer vision. The primary goal is to evaluate whether a participant is attentive, distracted, or disengaged by analyzing their eye gaze direction, head pose, and facial orientation from webcam footage.

For this project, you can use the Columbia Gaze Dataset ([https://www.cs.columbia.edu/CAVE/databases/columbia\\_gaze/](https://www.cs.columbia.edu/CAVE/databases/columbia_gaze/)), which provides over 5,800 high-quality images of 56 subjects with annotated gaze directions and head poses. Each image includes labels for gaze angle and head orientation, enabling you to train or fine-tune deep learning models for gaze estimation and attention analysis.

Downloading link:

<https://drive.google.com/drive/folders/1ySrlePZd0AucIM3tnEysfaKPmq5bP-TU?usp=sharing>

### 3. Smart Price Recommender (Development Project)

You can work on a Smart Price Recommender system that focuses on grocery price comparison and recommendation by aggregating data from multiple local and online stores such as Jalal Sons, Al-Fatah, Imtiaz, Carrefour and other local stores. This project involves developing a complete data-driven pipeline from web scraping to building an interactive dashboard. You will start by designing automated web scraping scripts to collect and update product prices from various sources, followed by data wrangling, cleaning, and normalization to handle inconsistencies in currency formats and product naming. The cleaned data will then be used to train ensemble regression models or recommendation algorithms that identify the best deals and price trends across stores. Finally, you will build a basic frontend dashboard that visualizes price comparisons, analytics etc. **scrape atleast 5 stores**

### 4. AI Security Surveillance (Anomaly Detection in Video)

You can work on an AI Video Surveillance System designed to enhance security by automatically detecting suspicious activities and unauthorized entries using live video feeds. This project involves applying computer vision and deep learning techniques to analyze motion patterns, recognize and classify objects such as people, animals, and vehicles, and generate real-time alerts for potential threats. You will preprocess and clean publicly available datasets like the UCF Crime Dataset then train and evaluate models for object and activity detection using architectures such as YOLOv8, ResNet, or Vision Transformers. The project includes stages like dataset preparation, model training, and performance evaluation using metrics such as precision, recall, and F1-score. The final outcome will be a working prototype that automatically detects suspicious behavior from video footage, supported by performance analysis and result visualization. **You can reduce number of classes to 8-10 from 14 to reduce dataset size**

Dataset link: video: <https://www.kaggle.com/datasets/alirakhmaev/ucf-crime-full>

Frames: <https://www.kaggle.com/datasets/rumaisaarizwansaya/ucf-crime-frames-colab>

## 5. Image Segmentation for Medical Imaging

You can work on an Image Segmentation project focused on medical imaging, particularly targeting organ or lesion segmentation in MRI or CT scans such as brain tumor detection. This project involves using deep learning techniques to automatically identify and segment abnormal regions from medical images, enabling more accurate and efficient diagnostic support. You will use datasets like the Brain MRI Segmentation Dataset (BraTS 2021), which provides annotated medical scans for supervised training. The implementation will include developing and training segmentation models such as UNet, ResUNet, or Vision Transformer-based architectures, followed by evaluating their performance using metrics like Dice coefficient and Intersection over Union (IoU). The project allows you to gain hands-on experience with medical image preprocessing, model design, and performance evaluation, contributing to real-world healthcare AI applications in medical diagnostics and analysis.

Dataset link: <https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1>