**62FIT4ATI Assignment**

**Fall 2025**

1. **Mini-projects**

|  |  |
| --- | --- |
|  |  |

1. **Groups**

62FIT4ATI Student groups and midterm mini-projects

1. **Schedule**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Class ATI-CLC 01 | | | Class ATI-CLC 02 | | | Class ATI-CLC 03 | | |
| *Date* | *Group* | *Mini-Project* | *Date* | *Group* | *Mini-Project* | *Date* | *Group* | *Mini-Project* |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

1. **Activities**

* Each group will have 20 minutes for their presentation/demonstration and a 10 minute Q&A session. You are welcome to include additional relevant content in your presentation. Strict adherence to the time limit is required.
* For each topic, students should explain and demonstrate (with source code) the complete deep learning workflow for the problem, including:

1. Formulate/Outline the problem

2. Identify inputs and outputs

3. Prepare data: inspecting data and visualizing data

4. Specify neural network model: define network architecture; compile and train the model

5. Measuare the performance: confusion matrix and discussion

6. Inference on new data

7. Conclusion

* You are encouraged to use Google Colab for project demonstrations.

1. **Submission**

* Presentation slides and souce code must be uploaded to the course website by the group leader three days prior to the presentation.
* If you are using Google Colab, you must download your project to .ipynb file and submit to the course website.
* Name your presentation slides and .ipynb file as: "62FIT4ATI\_Group #\_Topic #".

**List of topics**

**Mini-Project 1 - Structured data classification - Heart Disease dataset**

You are working with a dataset provided by the Cleveland Clinic Foundation for Heart Disease. It's a CSV file with 303 rows. Each row contains information about a patient, and each column describes an attribute of the patient.

You are asked to design and train a deep learning model to predict whether a patient has a heart disease (a binary classification task).

Link to download the [Heart Disease dataset](https://drive.google.com/file/d/1TvCRdSErX0ox-DTibYYeBXjwgOvviU3m/view?usp=sharing).

**Mini-Project 2 - Structured data classification (multi-classification task)**

You are working with a dataset collected from surveys conducted in Mexico, Peru, and Colombia, which includes individual characteristics such as age, gender, dietary habits, and physical activity. The dataset contains labeled obesity risk categories representing multiple classes related to obesity levels.

You are asked to design and train a deep learning model to classify individuals into one of seven obesity-related categories, making this a structured multi-class classification task.

Link to download the Obesity Levels dataset:  
<https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels>

**Mini-Project 3 - Structured data classification (regression task)**

This project uses insurance cost data gathered from individuals with different demographic and health-related features such as age, BMI, smoking habits, and region. The dataset is structured tabular data designed to predict the expected medical insurance cost.

You are asked to design and train a deep learning model to predict the continuous insurance cost for an individual, framing this as a regression task.

Link to download the Medical Insurance Costs dataset from Kaggle:

<https://www.kaggle.com/datasets/joebeachcapital/medical-insurance-costs>

**Mini-Project 4 - Collaborative filtering recommendation system with neural network**

You will work with the Book-Crossing dataset, which contains explicit ratings given by users on various books in an online book community. The dataset captures hundreds of thousands of users and nearly three hundred thousand books, including user-item interaction data.

You are asked to design and train a deep learning-based collaborative filtering recommendation system to predict user preferences for books based on past rating behavior.

Link to download the Book-Crossing dataset from the official source:  
<https://www.kaggle.com/datasets/somnambwl/bookcrossing-dataset>

**Mini-Project 5 - Content-based filtering recommendation system with neural network**

This project involves a dataset of cooking recipes, including detailed ingredient lists and preparation instructions. The dataset allows building content-based filtering models by leveraging recipe features to recommend dishes to users based on their ingredient preferences.

You are asked to design and train a neural network-based content filtering recommendation system that suggests recipes aligned with user-supplied ingredients or preferences.

Link to download the Recipe Recommendation dataset from the provided research resource:  
<https://archive.ics.uci.edu/dataset/911/recipe+reviews+and+user+feedback+dataset>

**Mini-Project 6 - Neural network embedding recommendation system**

The dataset provided contains anonymized user interaction data from an e-commerce platform, including actions such as product views, cart additions, and purchases collected over several months. Item properties and category structures are also available for modeling.

You are asked to design and train a neural network embedding model to learn latent user and item representations for personalized recommendation in an online retail environment.

Link to download the Retailrocket E-commerce dataset from Kaggle:  
<https://www.kaggle.com/retailrocket/ecommerce-dataset>

**Mini-Project 7 - Convolutional neural network**

This project uses an image dataset of hand gestures representing digits 0 through 9, collected from a large group of participants. Each image is 100x100 pixels in RGB color and contains samples of American Sign Language digits.

You are asked to design and train a convolutional neural network to classify images into one of ten digit classes based on sign language gestures.

Link to download the Sign Language Digits dataset from its GitHub repository:  
<https://github.com/ardamavi/Sign-Language-Digits-Dataset>

**Mini-Project 8 - Convolutional neural network**

The dataset consists of images of 12 species of plant seedlings showing different growth stages, collected from various agricultural settings. Each image is labeled with the species name and provides a moderate complexity classification challenge.

You are asked to design and train a convolutional neural network to accurately classify plant seedlings by species from photographic images.

Link to download the Plant Seedlings Classification dataset from Kaggle:  
<https://www.kaggle.com/vbookshelf/v2-plant-seedlings-dataset>

**Mini-Project 9 - Convolutional neural network**

You will work with a dataset of microscopic blood smear images, containing cell samples labeled as either parasitized (malaria-infected) or uninfected. The images were collected by the National Institutes of Health (NIH) and offer a practical challenge in biomedical image classification relevant to disease diagnosis.

You are asked to design and train a convolutional neural network to classify blood cell images as infected or uninfected by malaria parasites, focusing on binary image classification.

Link to download the Malaria Cell Images dataset from Kaggle:

<https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria>

**Mini-Project 10 - Convolutional neural network**

This project uses a dataset of chest X-ray images collected to assist in diagnosing pneumonia in patients. The dataset contains labeled images showing normal lungs and lungs affected by pneumonia, sourced from the National Institutes of Health and hosted on Kaggle. Pneumonia detection from X-ray images is a critical application of deep learning in medical imaging, aiding timely diagnosis and treatment.

You are asked to design and train a convolutional neural network to classify chest X-ray images as either pneumonia-positive or normal, framing it as a binary image classification problem.

Link to download the Chest X-ray Pneumonia dataset from Kaggle:  
<https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>