***CDS 525 Group Project***

# Assignment Description

You are suggested to pick up one of the following topics to complete your project **or bring your own idea and dataset related to deep learning.**

# Topic & Dataset.

1. **Fake News Detection**

fakenews.csv is a collection of 4729 unique values that have been labeled as either real or fake:<https://www.kaggle.com/datasets/iamrahulthorat/fakenews-csv/data>

# Sentiment Analysis for Financial News

dataset contains two columns, "Sentiment" and "News Headline", sentiment can be negative, neutral or positive:

<https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-for-financial-news/data>

# Social Media Sentiment Analysis

Sentiment dataset:<https://www.kaggle.com/datasets/mdismielhossenabir/sentiment-analysis>

# Harmful Brain Activity Classification

Classify seizures and other patterns of harmful brain activity in critically ill patients: <https://www.kaggle.com/competitions/hms-harmful-brain-activity-classification/data>

# Handwriting Recognition

KMNIST dataset: <https://github.com/rois-codh/kmnist> built-in dataset in Pytorch\*

# Image Classification

CIFAR 100 dataset: <https://www.cs.toronto.edu/~kriz/cifar.html> built-in dataset in Pytorch\*

\*Note: Pytorch built-in can be found here: <https://pytorch.org/vision/stable/datasets.html>

**Model Design**. For your project, you need to consider the following issues and write codes for selecting your model:

* 1. Model
  2. Loss Function
  3. Optimizer
  4. Other Hyperparameters (e.g., batch size, learning rate, number of epochs)
  5. Fine-tuning and pre-training (for large language/Transformer model)

**Performance Visualization**. Choose **One Model** for your project.

Training loss, training accuracy, and test accuracy with the change of number of epochs (you can decide the no. of epochs by using a validation set or other methods like setting a terminal condition) [**one figure**]

1. Training loss, training accuracy, and test accuracy with the change of number of epochs by using a different loss function from (1) [**one figure**].
2. Training loss, training accuracy, and test accuracy with the change of number of epochs by using different scales of learning rate (e.g., 0.1, 0.01, 0.001, 0.0001) and other settings are the same as (1) and (2) [**two figures**].
3. Training loss, training accuracy, and test accuracy with the change of number of epochs by using different batch sizes (e.g., 8, 16, 32, 64, 128) and other settings are the same as (1) and

(2) [**two figures**].

1. Visualize your predicted labels together with their corresponding inputs and actual labels of the first 100 results in the test set [**one figure or one table**].

When developing the project, you should consider making good use of the deep learning models you have learned on this course. Reasons for using the selected deep learning models, your pre-processing steps, how to select parameters and the performance et al. should be discussed in the written report.

# Group Allocation

The group formation will be uploaded to the MOODLE if you find a suitable member. Each group has about **4-6 MEMBERS.**

# Written Reports and Source Codes (20%)

*Written Report*

The written report is **not less than 2000 words**. The report should contain the following contents:

Introduction: briefly introduce the problem and the idea of this project and illustrate your

work: what is it used for? and what are the current popular techniques of your selected topic? (Mainly corresponding to **Topic & Dataset in Section 1**);

Design & Functions: illustrate the main functions of your system, illustrate the DL techniques

and models used in your system, and the reasons why you use these techniques and models (Mainly corresponding to **Model Design in Section 1**);

Demonstration & Performance: describe the datasets; show and visualize the performance of your proposed model (Mainly corresponding to **Performance Visualization in Section 1**); Conclusion: summarize your system as well as the limitations/future work.

References: cite the relevant sources

*Note: Please avoid pasting* ***a large amount of source codes*** *in your written report, and you only need to highlight some key source codes in the written report.*

*Marking criteria*

Problem Solving & Application of Knowledge (30%):

* whether the problem is well defined and formulated in the project;
* whether the problem is well addressed in the project;
* whether the basic requirements are fulfilled;
* whether knowledge is applied correctly;
* whether the solution is suitable for the problem.

Analysis & Critical Thinking (40%):

* whether the solution is effective;
* whether knowledge is applied wisely;
* whether the design of model is well-analyzed;
* whether the required performance visualization is included in the report.

Presentation & Illustration (20%):

* whether the problem is clearly defined in the written report;
* whether the knowledge applied is well-illustrated;
* whether the written report is organized elaborately;
* whether the critical points in the project is well-revealed;
* whether the report is written in a concise manner.

Implementation & Organization (10%):

* whether the program code is well-organized;
* whether the application is implemented elegantly and professionally.

# Presentation (10%)

Each group will prepare a presentation. The presentation should be about 20 minutes long. It should be **no more than 25 minutes**. The order of the group for the presentation will be randomly allocated by the instructor during the lecture.

The contents of the oral presentation should be consistent with that of the written report.

*Marking criteria*

Appropriate time allocation and pace (10%): Did the student allocate time appropriately, and

mange time effectively, with smooth progression? Did the student use appropriate pace? Did the presentation start punctually?

Clear, logically organized and relevant content (15%): Was information included always

relevant? Were presented points clearly stated and developed? Did the materials flow extremely well? Were the materials well organized? Were there any ambiguities are left unexplained?

Making effective use of presentation tools (20%): Was there balanced and proper use of

presentation tools with little or no distraction (e.g., appropriate animation/pictures, appropriate information on one slide, good color combination, clear titles, etc.)?

Using good body language, eye contact, and appropriate voice tone (20%): Did the student

show balanced posture, enthusiasm and confidence? Did the student make good eye contact with audience? Did the student use voice tone effectively?

Gaining/holding attention (20%): Did the student provide good motivation to engage

the audience’s interest? Did the student present the contents in a manner that captivates the audience’s attention?

Clarity of speech/Accuracy of grammar & pronunciation (15%): Was the voice consistently comprehensible? Were grammar and pronunciation accurate?

*Submission Details:*

1. You must ensure that all your project files used for this task and the report sit in a directory called “Group Assignment – Your Group Name”.
2. All files are required to be uploaded and a link to the “Group Assignment” directory submitted to Moodle.
3. Please make sure that unit Instructor and TA have access to the folder.
4. A **link** to the demo video of your app running **must be submitted**.
5. It would be great if you could submit your **GitHub** link.
6. This is a **Group** assignment, and you should submit it **by 8 pm, Thursday, Week 10**.