American University of Beirut
Department of Computer Science
CMPS 261 - Spring 2023-2024
Course Project

1 Deadlines and Grading

Task	Grade	Deadline
Grouping and Topic Selection	0%	06 March 2024
Report	60%	06 May 2024
Code	30%	06 May 2024
Model Testing	10%	06 May 2024

2 Tasks

2.1 Grouping and Topic Selection

Each group should be composed of three to four people. If you cannot find a group or if your group consists of less than three people, we will assign you to groups or merge groups to reach the desired group size. You can form groups from different sections. You can choose one of the below three topics to work on in your project:

1. Food Item Recognition from Images: This dataset consists of images of various food items that belong to nine different classes, and your role is to build a machine learning model to classify each food item into its respective class based on its image. Food item recognition has many applications in various domains industries such as hospitality management, recycling, waste management and nutrition.

2. Medical Text Classification: This dataset consists of short medical snippets that describe 5 different types of medical conditions. Your job is to classify a snippet into one of those five classes. This can have many applications such as assisting with systematic reviews, medical information retrieval, and medical chatbots.

3. Breast Cancer Classification: This dataset contains clinical and genomic data for patients with breast cancer. The dataset includes four types of breast cancer. Your task is to predict the type of cancer, and to identify potential biomarkers. Breast cancer is the most frequent cancer among women, impacting 2.1 million women each year. Breast cancer causes the greatest number of cancer-related deaths among women.

2.2 Report

At the end of the project, you should create a report describing how you tackled your chosen problem. The report should consist of the below sections. One report per group should be submitted, and you should use latex for typing.

1. Abstract: provide a summary describing the problem you chose, how you tackled it, and a summary of your results.

2. Introduction: describe your problem in detail and summarize your approach and findings

3. Dataset: describe the dataset you used to train your models, and any pre-processing steps you performed.

4. Models: describe the models you used to tackle your problem, and how you trained, tuned and validated them. Please be specific and concise. Also, provide the rationale behind your choices.

5. Results: describe the experiments you ran to train and evaluate your models. Also, provide all results of your experiments and discuss them.

6. Conclusion: provide a concise conclusion of your project, highlighting the main results you achieved, and what future work could be.

7. References: a list of references that you used throughout your project.

Your report will be graded based on the following rubric:

Excellent	The report is well written and contains all the required information (100 - 85 points)
Satisfactory	The report is missing some information or the writing could be enhanced (80 - 60 points)
Unsatisfactory	The report is poorly written, and lacks many of the required information (55 - 0 points)

2.3 Code

You should submit a notebook that includes all the code you used to tackle your problem including data cleaning and data processing, training, diagnosis, hyper-parameter tuning and model selection. You notebook should be clear, well-structured, and well commented. In addition to submitting the notebook through Moodle, you are required to maintain a Github repository with a properly documented code and include the link of that repository in your report. The Github repository should have your final trained model available for download. Your code will be graded based on the following rubric:

Excellent	The code is clear, well documented, comprehensive and runs smoothly (100 - 85 points)
Satisfactory	The code is not very clear, incomplete, there are issues running it, or it does not fully match the report (80 - 60 points)
Unsatisfactory	The code is poorly written, cumbersome, buggy and incomplete or does not match the report in the most part (55 - 0 points)

2.4 Model Testing

Your trained model will be downloaded from the Github repository and will be tested on a holdout test. You should submit a separate notebook for your project that can be used to upload your trained model and test it. This notebook should perform all preprocessing steps needed to apply the model on raw test data. Your model testing will be graded based on the following rubric:

Excellent	The model has a very good performance compared to other submitted models (100 - 85 points)
Satisfactory	The model has an acceptable performance that falls within the average of other submitted models (80 - 60 points)
Unsatisfactory	The model has a very low performance compared to other submitted models (55 - 0 points)