Department of Computer Science

Summative Coursework Set Front Page

Module Title	Artificial Intelligence & Machine Learning
Module Code	CS3AM
Lecturer responsible	Dr. Muhammad Shahzad Dr. Ferran Espuny Pujol
Type of Assignment (e.g., technical report, portfolio exercise, in-class test)	Coursework
Individual or Group Assignment	Individual
Weighting of the Assignment	50%
Word count/page limit	5-7 pages (7 maximum) • Excluding the front/title page of information, references and appendices. • Including figures, diagrams, graphs, and tables. • Times New Roman, 12pt., 1.15 line spacing. • Table of content, abstract/introduction are not required. • The report should be clearly structured with a separate section (with appropriate subsection) for each task and a final conclusion.
Expected hrs spent for the assignment (set by lecturer)	20 hours (beyond lab sessions and also provided you attend all the lab sessions to strengthen your AI concepts)
Items to be submitted	A single zip archive containing: 1) report (PDF or Word file) 2) dataset(s) 3) Python script(s) (PY or IPYNB files)
Work to be submitted on-line via Blackboard Learn by	**9th December 2024, noon (Week-11)**
Work will be marked and returned by	15 working days after the above deadline
Note	

By submitting this work, you are certifying that you have read the assessment guidelines, which are displayed at the top of the Assessment Folder on the Blackboard course for this module, and that you have conformed to the associated policies and practises, including those on:

- Submitting your own work, not that of other people or systems, and the associated penalties for Academic Misconduct
- Submitting by the specified deadline, and the penalties associated with late submission (if allowed)
- The exceptional circumstances system (for applying for extensions)
- The use of a green sticker for students with relevant needs

1. Assignment description

You are required to find a dataset other than the one used in lab sessions and provided in BlackBoard, formulate a problem you want to address with the dataset (e.g., predict whether a mushroom is poisonous or not based on its characteristics), build, evaluate and compare two different machine learning models that would address the problem, and draw conclusions and recommendations based on your findings. One of the two models must be based on a deep learning architecture implemented using the TensorFlow/PyTorch in Python. The submission should include your report, dataset(s) and Python scripts with comments, all included in one zip-file. Your work should be original and produced by you. Copying whole tutorials, scripts or images from other sources is not allowed. Any material you borrow from other sources to build upon should be clearly referenced (use comments to reference in Python scripts); otherwise, it will be treated as plagiarism, which may lead to investigation and subsequent action.

You can use any open data, e.g.:

https://ieee-dataport.org/topic-tags/artificial-intelligence

https://archive.ics.uci.edu/ml/datasets.php

https://www.kaggle.com/datasets

https://data.gov.uk/

Some examples:

Optical Image data:

1. Building Detection and Roof Type Classification
   https://ieee-dataport.org/competitions/2023-ieee-grss-data-fusion-contest-large-scale-fine-grained-building-classification

2. So2Sat LCZ42 Dataset for land cover classification
   https://mediatum.ub.tum.de/1483140

3. DOTA: A Large-Scale Benchmark and Challenges for Object Detection in Aerial Images

https://captain-whu.github.io/DOTA/dataset.html

Weather and Climate Data:

4. Daily 0900 GMT observations from the university weather stations (back to 1908; there was a site change in 1968):
   https://metdata.reading.ac.uk/cgi-bin/climate_extract.cgi

5. Five-minute/hourly data from our automatic weather station back to 1 Sept 2014 (has a few missing dates):
   https://metdata.reading.ac.uk/cgi-bin/MODE3.cgi
   http://www.met.reading.ac.uk/~sws09a/MODE3_help.html
   For some further inspiration (visualisation of current data) and information around the above two data sources, check these resources:
   https://www.met.reading.ac.uk/weatherdata/wall_display.html
   https://research.reading.ac.uk/meteorology/atmospheric-observatory/atmospheric-observatorydata/ https://www.ecmwf.int/en/forecasts/charts/catalogue/

6. Daily energy demand over India by state, and (many) meteorological variables of interest averaged over each state (hourly/daily; 2013–present):
   https://gws-access.jasmin.ac.uk/public/incompass/kieran/kovalchuk/energy-india/
7. Daily observed river discharge at five stations over the Indus and its tributaries, with catchmentaveraged meteorological and hydrological variables (Jan 2015 to Jan 2021):
   https://gws-access.jasmin.ac.uk/public/incompass/kieran/kovalchuk/indus-river/
   Some notes on the provenance and metadata for the above two data sources:
   - River data are from here: http://www.wapda.gov.pk/index.php/river-flow-data
   - Energy demand data are scraped from PDF publications on the POSOCO website, e.g.:
     https://posoco.in/download/17-05-21_nldc_psp/?wpdmdl=37035
   - Catchment- and state-averaged variables were computed using ERA5 data, for which descriptions are available here:
     https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

Additional information

It ***expected*** that you will work on the coursework both in and out of the scheduled lab sessions especially in last two weeks, i.e., Week-9 and Week-10 before the coursework submission deadline. However, you are not bound to follow this schedule and may work on the coursework as you prefer.

2. Assignment submission requirements

Items to be submitted on-line through Blackboard Learn include a single zip archive containing:

1) report (PDF or Word file)

2) dataset(s)

3) Python script(s) (PY or IPYNB files)

Front page of the student's submission

(*the following are compulsory)*

Module Code: CS3AM

Assignment report Title: Coursework

Date (when the work completed):

Actual hrs spent for the assignment:

We will use information about how long you spent on the assignment when we review and balance coursework between modules for later years. An exact answer is not necessary, but please try to give a reasonable approximation.

Recommended Report Structure

1. Cover page with the title of your project; module code, title, convenor name; your name and student number; date.
2. Abstract (summarise your work and results)
3. Background and problem to be addressed (justify and support with references to literature)
4. Exploratory data analysis (dataset description and visualisation, support with Python code snippets and figures)
5. Data pre-processing and feature selection (support with Python code snippets)
6. Machine learning model N (iterate for each of the two models)
   6.1. Summary of the approach (justify why this ML algorithm, support with references to literature)
   6.2. Model training and evaluation (support with Python code snippets)
   6.3. Results and discussion (support with tables/figures)
7. Results comparison across the models built (support with tables/figures)
8. Conclusion, recommendations, and future work
9. References.

3. Assessment Classification

Classification Range	Typically, the work should meet these requirements
First Class (≥=70%)	Outstanding/excellent work with correct results, a good presentation of the workflows, code and results, and a critical analysis of the results. An outstanding work will present fully automated solutions based on advanced techniques. - All parts of the assignment are completed correctly, - comprehensive discussions, - helpful & precise comments (in code), - deep & insightful analysis, - excellent & compelling presentation of the work in the form of report writing.
Second Class Division 1 (60-69%)	Good work with mostly correct results and good discussions: most work has been carried out correctly. The presentation of

	work is good, well structured, clear, and complete with respect to the work done.
Second Class Division 2 (50-59%)	Acceptable work with little discussions: some significant part of the assignment is missing and/or has partially correct results. The presentation is, in general, accurate and complete, though it may lack some clarity and quality.
Third Class (40-49%)	Achievement of the minimum requirements with no discussion: most of the significant part of the assignment is missing and/or has mostly incorrect results. The presentation is, in general, partially accurate and not complete or clear.
Fail (<40%)	Incomplete solutions to limited part of the assignment with very little or no discussions. Most tasks have not been carried out with sufficient accuracy. Results may not be correct or technically sound. The presentation is not accurate/complete and lacks clarity.

4. Marking Scheme

Data Selection & Preprocessing (20%)	• [10 marks] Choice of dataset and workflow performing data understanding and preprocessing. • [10 marks] Reporting each of the data understanding and preprocessing step, findings, and the corresponding discussions.
Modelling (40%)	• [10 marks] Workflow performing the classification, regression, forecasting, or other chosen task. • [10 marks] Descriptions of the adopted algorithms and parameters. • [10 marks] Discussions and justifications of your selection of algorithms and parameters. • [10 marks] Presentation, explanation, and analysis of results, discussions, ablation, and justifications.
Evaluation (10%)	• [10 marks] Descriptions of performance measures and evaluation methods/strategies, discussions, and justifications.
Code (10%)	• [10 marks] Code quality, commenting, structure and organization, i.e., implemented in efficient way, with clear and accurate comments, and no errors in execution.
Report Quality (20%)	• [20 marks] Report structure, conclusion, references, quality of figures and tables. It is expected that no redundant or low-quality figures, or tables will be included.