Department of Computer Science

Summative Coursework Set Front Page

Module Title	Artificial Intelligence
Module Code	CS3AI18
Lecturer responsible	Dr. Muhammad Shahzad
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	4 pages (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)	10 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	**24th November 2023 (noon)**
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 programming assignments in the scheduled labs from **Week-7 and onwards**. However, you are not bound to follow this schedule and you may start working over the assignment as soon as you like outside the scheduled periods.

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: CS3AI18

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 & marking scheme

Assessment Criteria	**Dataset(s) & Question(s)**	**Modelling**	**Code**	**Report**
Weighting	**20%**	**40%**	**20%**	**20%**
0 – 29%	Inappropriate dataset or lack of its initial analysis and understanding; ill-formulated questions.	Missing or inappropriate data pre-processing, feature selection, modelling and/or results interpretation.	Missing or not compiling/ executing.	Not appropriately structured with main sections missing.
30 – 39%	Appropriate dataset, but its initial analysis is poor, and/or oversimplified questions.	Incomplete or significant errors in data pre-processing, modelling and/or results interpretation.	Compiling and executing but implementing only some deliverables.	Badly planned and/or some sections and/or referencing to code missing.
40 – 49%	Fair dataset and questions, but significant errors in initial dataset analysis or not fully justified questions.	Fair data pre-processing, feature selection, modelling and results interpretation, but with some significant	Most deliverables are implemented, but there are some significant errors, s/w	All required sections are covered, but structure is not well planned or

		errors or missing details.	principles are not followed, and/or lack of comments.	major details missing.
50 – 59%	Satisfactory dataset and justified questions, but some minor errors in initial analysis.	Good data pre-processing, feature selection, modelling and results interpretation, but with some minor errors or missing details.	All deliverables are implemented, but there are some minor errors, not all s/w principles are not followed, and/or insufficient/ inaccurate comments.	Well planned with all required sections present, but some details or code referencing missing or not clearly explained.
60 – 69%	Good choice of dataset and questions with fair impact and no errors in initial analysis.	Good data pre-processing, feature selection, modelling and results interpretation, with no errors.	All deliverables are implemented with no errors, but code is not optimised and/or with insufficient comments.	Well planned and clearly formulated with all required sections present, but with some minor details missing.
70 – 79%	Very good choice of dataset and questions with significant impact, no errors in initial analysis.	Very strong case of pre-processing, feature selection, modelling and results interpretation, with attention to details and no errors.	All deliverables are implemented in efficient way, following s/w principles, with clear and accurate comments, and no errors.	Very well planned and clearly presented, with appropriate and sufficient referencing to code and literature.
80 – 89%	Excellent choice of dataset and questions with major impact, no errors in initial analysis.	Excellent pre-processing, feature selection, modelling and results interpretation, error free with some advanced techniques employed and several settings tested.	All deliverables are implemented in efficient way, following s/w principles, employing some advanced methods, with clear and accurate comments, and no errors.	Excellent, complete, clearly presented professional work, with appropriate and sufficient referencing to code and literature.

90 – 100%	Outstanding choice of dataset and questions with significant impact, no errors in initial analysis.	Outstanding pre-processing, feature selection, modelling and results interpretation, error free with some novel techniques employed suitable for publication.	All deliverables are implemented in efficient way, following s/w principles, employing some advanced/novel methods, with clear and accurate comments, and no errors.	Outstanding, complete, clearly presented professional work, with appropriate referencing to code and literature, and suitable for publication.