Department of Computer Science

Summative Coursework Set Front Page

Module Title:	Artificial Intelligence and Machine Learning
Module Code:	CSMAI21, CSMAI19
Lecturer responsible:	Dr Yevgeniya Kovalchuk
Type of Assignment:	Individual
Weighting of the Assignment:	50%, 100%
Page limit/Word count:	3,000 words excluding figures and tables
Expected hours spent for this assignment:	20 hours
Items to be submitted:	A single zip archive containing: 1) report (PDF or Word file) 2) dataset(s) (CSV file(s)) 3) Python script(s) (PY file(s) [not IPYNB])
Work to be submitted on-line via Blackboard Learn by:	**Friday 11th March 2022 at noon**
Work will be marked and returned by:	15 working days after the above deadline

NOTES

By submitting this work, you are certifying that it is all your own and has not been taken from any other person's work except where explicitly the works of others have been acknowledged, quoted, and referenced. You understand that failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly. The University's Statement of Academic Misconduct is available on the University web pages.

If your work is submitted after the deadline, *10%* of the maximum possible mark will be deducted for *each* working day (or part of) it is late. A mark of zero will be awarded if your work is submitted more than 5 working days late. You are strongly recommended to hand work in by the deadline as a late submission on one piece of work can impact on other work.

If you believe that you have a valid reason for failing to meet a deadline then you should complete an Extenuating Circumstances form and submit it to the Student Support Centre *before* the deadline, or as soon as is practicable afterwards, explaining why.

Assignment Description

You are required to find a dataset, 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 three different machine learning models that would address the problem, and draw conclusions and recommendations based on your findings. One of the three models must be based on a deep learning architecture implemented using the Keras Python library. 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://archive.ics.uci.edu/ml/datasets.php

https://www.kaggle.com/datasets

https://data.gov.uk/

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 *(summarising your work and results)*
3. Background and problem to be addressed *(justified and supported with references to literature)*
4. Dataset(s) description
5. Machine learning model N *(iterate for each of the three models)*
   - 5.1. Summary of the approach *(justified and supported with references to literature)*
   - 5.2. Data visualisation, preprocessing, feature selection *(supported with figures & Python snippets)*
   - 5.3. Model training and evaluation *(supported with Python snippets)*
   - 5.4. Results and discussion *(supported with tables/figures & Python snippets)*
6. Results comparison across the models built *(supported with tables, figures & Python snippets)*
7. Conclusion, recommendations and future work
8. References.

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 errors or missing details.	Most deliverables are implemented, but there are some significant errors, s/w principles are not followed, and/or lack of comments.	All required sections are covered, but structure is not well planned or 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 - 90%	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.