School of Computing, Napier University

Assessment Brief

1. Module number	**SET11121 002**
2. Module title	**Data Wrangling**
3. Module leader	*Yanchao Yu*
4. Tutor with responsibility for this Assessment	Yanchao Yu *(Y.Yu@napier.ac.uk)* *Jeff Mitchell (J.Mitchell@napier.ac.uk)*
5. Assessment	Coursework
6. Weighting	70% of module assessment
7. Size and/or time limits for assessment	Up to 1000 words. You can additionally include an unlimited number of tables, figures and a reference list.
8. Deadline of submission Your attention is drawn to the penalties for late submission	15/04/22 - 3 pm UK time
9. Arrangements for submission	Your Coursework must be submitted via Moodle. **Further submission instructions are included in the attached specification and on Moodle**
10. Assessment Regulations	All assessments are subject to the University Regulations**.**
11. The requirements for the assessment	See Attached
12. Special instructions	See Attached
13. Return of work	Feedback and marks will be provided **within four weeks** of submission.
14. Assessment criteria	Your coursework will be marked using the marking sheet attached as Appendix A. This specifies the criteria that will be used to mark your work. Further discussion of criteria is also included in the coursework specification attached.

Assessment Brief

The assignment aims to cover the learning outcomes specified for the module:

LO1: Analyse the concepts and process of data analysis, including pre-processing and preparation of data.
LO2: Analyse and evaluate modelling methods and techniques in data analysis.
LO3: Integrate data analysis algorithms to conduct data analysis and visualisation.
LO4: Critically interpret and evaluate results generated by analysis techniques.
LO5: Integrate specialised techniques for dealing with heterogeneous data sets.

The goal of this assignment is to use machine learning approaches to solve **one** of the two tasks described below.

Task 1: Computer Vision

Using the **provided** dataset (see below), develop a deep neural network (i.e. a Convolutional Neural Network) for the image classification problem. Your proposed model must be a Convolutional Neural Network with an architecture proposed by you. You should compare the performance (i.e. classification accuracy) against a pre-trained model already provided within Keras. Specifically:

1. Create a convolutional neural network (CNN). You must *motivate* the choice of CNN architecture.
2. Provide evidence of parameters tuning, such as learning rate, filter sizes, fully-connected layers.
3. Provide evidence of a comparison of your model, using appropriate evaluation metrics of your choice, against one of the predefined models within Keras. You MUST choose one from this list.
4. Write a report outlining your solution, including a description of the model architecture (including a graphical representation of the architecture), evidence of parameters tuning etc, as well as the evaluation results (including relevant tables and graphs) and critical evaluation/reflection.

Data
For this assignment you **should** use the CIFAR-10 dataset which can be obtained by the Keras library as follows:

```
from keras.datasets import cifar10
```

The goal of this exercise is not to produce a state-of-the-art model. If your model performs poorly with respect to the predefined model you have chosen, do not worry—this is not what we are testing. It should however be appropriately motivated and evaluated (you will be tested on those aspects).

Task 2: Natural Language Processing

Using the **provided** dataset on Moodle, develop a machine learning algorithm of your choice for text classification. You can use any combination of text representation and machine learning approaches. Specifically:

1. Choose and implement a text representation technique (such as a bag of words, word embeddings etc.). Justify your choice.
2. Using the text representation from (1), build a machine learning model for text classification (e.g. decision tree, neural network etc.). Justify your choice.
3. Evaluate your model using the appropriate metrics of your choice.
4. Write a report outlining your solution, including a description of the machine learning setup (including a description and motivation of the text representation technique, an appropriate graphical representation if relevant), evidence of parameters tuning (if relevant), as well as the evaluation results (including relevant tables and graphs) and critical evaluation/reflection.

Data
For this assignment, you **should** use the dataset provided on Moodle.

Again, the goal of this task is not to produce a state-of-the-art model. If your chosen model performs poorly by your selected metrics, do not worry—this is not what we are testing. It should however be appropriately motivated and evaluated (you will be tested on those aspects).

Tips and Clarifications
- If you are struggling to make something work with the volume of data present, you can subsample (for instance, randomly pick a proportion of the dataset).
- You **must** use Python and its libraries to tackle this task. You are strongly encouraged to make use of existing libraries for model building and evaluation, rather than writing your own unless you specifically need to do something with no library support (e.g. if you want to do feature engineering).

Your report can include images, such as your neural network architecture design (for Task 1 or for Task 2 if you choose a deep learning approach), and plots. For architecture design, many programs can be used. You can draw your neural network architecture with PowerPoint.

Deadline: Friday 15 April at 3 pm (UK time).

<u>You will submit:</u>
1. One .pdf file of up to 1000 words excluding tables, references etc as outlined earlier.
2. The code of your solution as a notebook. If you do any pre-processing to the data, please also include the script you use to do this (or a list of the commands run).

Marking:
Both tasks will be marked as follows:

25% Method/Model;

10% Evaluation

For the report [35%]:

5% Structure;

15% Content;
15% Criticality/Discussion.

See Appendix A for more explanations.

Feedback:
Apart from the markings, you will also receive text feedback from Moodle 3 - 5 weeks after the submission deadline. The feedback will contain a further explanation about what you have done good and what needs to improve, corresponding to your marks.

Late submission policy
Coursework submitted after the agreed deadline will be marked at a maximum of 40%. Coursework submitted over five working days after the agreed deadline will be given 0%.

Extensions
If you require an extension, please contact the module leader **before** the deadline. Extensions are only provided for exceptional circumstances and evidence may be required. See the Fit to Sit regulations for more details.

Plagiarism
Plagiarised work will be dealt with according to the university's guidelines (Please read - especially if this is the first time in a UK university): http://www2.napier.ac.uk/ed/plagiarism/

Appendix A: Marking Scheme

	No Submission	Very poor	Inadequate	Adequate	Good	Very good	Excellent	Outstanding
B1 Method / Model **25%**	No work submitted	Code with bugs and algorithm /model not well described	Code with bugs but algorithm /model well described	Code with a minor bug but algorithm /model not well described and justified	Code with a minor bug but algorithm /model well described and justified	Code without bugs but algorithm /model not described or justified	Code without bugs but algorithm /model not described and justified in great detail	Code without bugs and algorithm /model described and justified in detail
B2 Evaluation **10%**	No work submitted	Not appropriate evaluation metric chosen	Neither the evaluation setup nor the results are described appropriately	Evaluation setup is not justified but almost correctly executed and results are mentioned	Evaluation setup is not justified but correctly executed and results are mentioned	Evaluation setup is somewhat justified and results are somewhat mentioned and discussed	Evaluation setup is somewhat justified but results fully described and discussed	Evaluation setup is justified and results fully described and discussed

B3 Report / Reflection **35%**	No work submitted	Poor report that misses essential parts, and results/method not critically described	Poor report that misses essential parts, and results/method and not adequate reflection provided	Results adequately described and analysed, although the report would benefit from some further critical analysis.	Good report and reflection. However, more explanations would be needed.	Good report, reflection and explanation of results.	Excellent report, reflection and explanation of results, providing detailed analysis of results.	Excellent report, reflection and excellent insights of results (including outlining limitations and strengths).