

Coursework Assignment Brief

Postgraduate

Academic Year 2019-20

Module Title:	Advanced Data Science	
Module Code:	CMP7161	
Assessment Title:	Coursework	
Assessment Identifier:	CWRK001	Weighting: 100%
School:	CDT	
Module Co-ordinator:	Yevgeniya Kovalcuk	
Hand in deadline date:	Coursework submission deadline is 12pm (midday) on 12 th May 2020.	
Return of Feedback date and format	Written feedback on 11 th June 2020.	
Re-assessment hand in deadline date:	12pm (midday) on 27 th July 2020.	
Support available for students required to submit a re-assessment:	Revisions sessions will be arranged for the period immediately preceding the hand in date.	
NOTE:	At the first assessment attempt, the full range of marks is available. At the re-assessment attempt the mark is capped and the maximum mark that can be achieved is 50%.	
Assessment Summary	<p>The coursework involves finding a dataset, formulating a research problem related to this dataset and developing a data analytics solution to address the problem. The assessment of the coursework will consist of evaluating the submitted presentation, dataset(s), project report and scripts. It will demonstrate the extent to which students have successfully fulfilled the learning outcomes and developed an analytical solution for their identified research problem. There are two deliverables:</p> <ol style="list-style-type: none"> 1) Presentation (20%); 2) Project report with code (80%). 	

IMPORTANT STATEMENTS

Standard Postgraduate Regulations

Your studies will be governed by the BCU Academic Regulations on Assessment, Progression and Awards. Copies of regulations can be found at <https://icity.bcu.ac.uk/Academic-Services/Information-for-Students/Academic-Regulations-2018-19>

For courses accredited by professional bodies such as the IET (Institution of Engineering and Technology) there are some exemptions from the standard regulations and these are detailed in your Programme Handbook

Cheating and Plagiarism

Both cheating and plagiarism are totally unacceptable and the University maintains a strict policy against them. It is YOUR responsibility to be aware of this policy and to act accordingly. Please refer to the Academic Registry Guidance at <https://icity.bcu.ac.uk/Academic-Registry/Information-for-Students/Assessment/Avoiding-Allegations-of-Cheating>

The basic principles are:

- Don't pass off anyone else's work as your own, including work from "essay banks". This is plagiarism and is viewed extremely seriously by the University.
- Don't submit a piece of work in whole or in part that has already been submitted for assessment elsewhere. This is called duplication and, like plagiarism, is viewed extremely seriously by the University.
- Always acknowledge all of the sources that you have used in your coursework assignment or project.
- If you are using the exact words of another person, always put them in quotation marks.
- Check that you know whether the coursework is to be produced individually or whether you can work with others.
- If you are doing group work, be sure about what you are supposed to do on your own.
- Never make up or falsify data to prove your point.
- Never allow others to copy your work.
- Never lend disks, memory sticks or copies of your coursework to any other student in the University; this may lead you being accused of collusion.

By submitting coursework, either physically or electronically, you are confirming that it is your own work (or, in the case of a group submission, that it is the result of joint work undertaken by members of the group that you represent) and that you have read and understand the University's guidance on plagiarism and cheating.

You should be aware that coursework may be submitted to an electronic detection system in order to help ascertain if any plagiarised material is present. You may check your own work prior to submission using Turnitin at the [Formative Moodle Site](#). If you have queries about what constitutes plagiarism, please speak to your module tutor or the Centre for Academic Success.

Electronic Submission of Work

It is your responsibility to ensure that work submitted in electronic format can be opened on a faculty computer and to check that any electronic submissions have been successfully uploaded. If it cannot be opened it will not be marked. Any required file formats will be specified in the assignment brief and failure to comply with these submission requirements will result in work not being marked. You must retain a copy of all electronic work you have submitted and re-submit if requested.

Learning Outcomes to be Assessed:

- 1) Understand different stages involved in building a data processing pipeline.
- 2) Critically evaluate and practice implementing a wide range of different algorithms and modern tools used to solve data science tasks.
- 3) Apply learned techniques to formulate and solve real-life data-based problems.
- 4) Communicate technical information in a range of formats appropriate to a specific audience.

Assessment Details:

Title: Building and evaluating data analysis pipelines

Type: Coursework

Style: Presentation, report and programming scripts

Rationale:

This coursework is most suited for assessing the learning outcomes of the module providing the practical nature of the Data Science field. The area is growing fast and the interest in machine learning solutions constantly increases. Learning to formulate and solving practical and research-oriented data-driven projects will ensure your continuing employability through development of analytical soft skills.

Description

For this coursework you are required to find a dataset (you can merge several), 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 and evaluate at least two different machine learning models that would address the problem, and draw conclusions and recommendations based on your findings. There are two deliverables:

- 1) Presentation (20%);
- 2) Project report with attached dataset(s) and Python code (80%).

Your presentation should be in a form of executive summary that one can read in 5 minutes. Recommended length of the report is 3,000 words excluding figures and tables.

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.:

<http://archive.ics.uci.edu/ml/datasets.php>

<https://www.kaggle.com/datasets>

<https://data.gov.uk/>

Additional information

Recommended Presentation Structure:

1. Title of your project, the module code and title, your name and student number
2. Abstract summarising your work
3. Dataset description and the problem to be addressed
4. Summary of data pre-processing and/or feature selection
5. Description of the models built, their evaluation and comparison
6. Summary of the results
7. Conclusion and recommendations

Recommended Report Structure:

1. Cover page with title of your project; module code, title, coordinator name; your name and student number; date.
2. Abstract
3. Introduction, background, aim and objectives
4. Dataset(s) description *(can be supported with figures and references to Python code)*
5. Problem to be addressed *(justified and supported with references to literature)*
6. Machine learning model N *(iterate for each model/algorithm)*
 - 6.1. Summary of the approach *(justified and supported with references)*
 - 6.2. Data pre-processing, visualisation, feature selection *(with references to Python code)*
 - 6.3. Model training, evaluation and testing *(with references to Python code)*
 - 6.4. Results and discussion *(supported with tables, figures and references to Python code)*
7. Results comparison across the models built *(supported with tables, figures and Python code)*
8. Conclusion, recommendations and future work
9. References

For advice on writing style, referencing and academic skills, please make use of the Centre for Academic Success: <https://icity.bcu.ac.uk/celt/centre-for-academic-success>

Workload: Recommended length of the report is 3,000 words excluding figures and tables. A typical student would be expected to spend a minimum of 40 hours working on the coursework to pass this assignment.

Marking Criteria

Assessment criteria and grading scale: Presentation (20%)

Assessment Criteria →	Visual Aid	Technical Content	Timing
Weighting	30%	50%	20%
Grading Criteria 0 – 29%	Incoherent, lacks basic blocks and/or content.	Incomplete and/or with major errors.	Significant lack of planning and inappropriate use of allowed time.
30 – 39%	Incoherent, lacks details on key project aspects.	Some important information is missing and/or there are significant errors.	Inappropriate distribution of time across sections and/or overrunning.
40 – 49%	Some errors or deficiencies in key points, lacks clarity.	Generally complete but with some errors and/or lack of understanding.	Within time limit, but rushed, not clear and lack of planning.
50 – 59%	Reasonable encapsulation of key points, but over/under-detailed and/or including minor errors.	Satisfactory level of proficiency, revealing understanding of basic concepts, but with minor errors.	Within time limit and clear, but the distribution of time across sections could be improved.
60 – 69%	Good encapsulation of key points, generally clear and coherent with good impact and error free.	Competent level of proficiency, with clear, insightful and error free artefacts.	Good planning and timekeeping.
70 – 79%	Very good encapsulation of key points, very clear, high impact and error free.	Very good level of proficiency, with clear, insightful and error free artefacts.	Very good planning and timekeeping.
80 – 100%	Excellent encapsulation of key points, extremely clear, high impact and error free.	Excellent level of proficiency, with clear, insightful and error free artefacts.	Excellent planning and timekeeping.

Assessment criteria and grading scale: Report and code (80%)

Assessment Criteria →	Dataset(s) & Question(s)	Modelling	Code	Report
Weighting	20%	40%	20%	20%
Grading Criteria 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%	Satisfactory dataset and questions, but significant errors in initial dataset analysis or not fully justified questions.	Satisfactory data pre-processing, feature selection, modelling and results interpretation, but with some major errors or missing details.	All deliverables are implemented, but there are some major 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 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 detail 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 – 100%	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.

Submission Details:

Format: The submission is by Moodle upload of one zip-file containing your presentation (in ppt/pdf format), report (in doc/pdf format), dataset(s) (in csv format) and Python code (as a collection of .py files).

Regulations:

If you submit an assessment late at the first attempt then you will be subject to one of the following penalties:

- if the submission is made **between 1 and 24 hours** after the published deadline the original mark awarded will be reduced by **5%**. For example, a mark of 60% will be reduced by 3% so that the mark that the student will receive is 57%. ;
- if the submission is made between **24 hours** and **one week (5 working days)** after the published deadline the original mark awarded will be reduced by 10%. For example, a mark of 60% will be reduced by 6% so that the mark the student will receive is 54%.
- **if the submission is made after 5 days following the deadline, your work will be deemed as a fail and returned to you unmarked.**

The reduction in the mark will not be applied in the following two cases:

- the mark is below the pass mark for the assessment. In this case the mark achieved by the student will stand
- where a deduction will reduce the mark from a pass to a fail. In this case the mark awarded will be the threshold (i.e.50%)

Please note:

- **If you submit a re-assessment late then it will be deemed as a fail and returned to you unmarked.**

Feedback:

There will be opportunities for formative verbal feedback on work during the lab sessions and presentation days. Written feedback will be provided on the summative assessment (submitted coursework package).

Marks and feedback on your work will normally be provided within 20 working days of its submission deadline.

Where to get help:

Individual support will be provided during normal lab hours. The tutor can be contacted by email to arrange to meet for further guidance and advice.

Students can get additional support from the library support for searching for information and finding academic sources. See their iCity page for more information:

<http://libanswers.bcu.ac.uk/>

The Centre for Academic Success offers 1:1 advice and feedback on academic writing, referencing, study skills and maths/statistics/computing. See their iCity page for more information: <https://icity.bcu.ac.uk/celt/centre-for-academic-success>

Link to My Assignment Planner tool: <http://library.bcu.ac.uk/MAP2/freecalc-mail/>

Fit to Submit:

Are you ready to submit your assignment? Review this assignment brief and consider whether you have met the criteria. Use this checklists to ensure that you have done everything needed:

- one .zip file that includes the following;
- one presentation in PPT or PDF format;
- one report in Word or PDF format with your name, student number and project title;
- dataset(s) in CSV format;
- optional: any additional CSV files representing data pre-processing stages;
- Python scripts (as many as required), each having a comment indicating the author.