**Groupwork Assignment Brief for:   
COM747 Data Science and Machine Learning  
(worth 60% of module assessment)**

**Module Coordinator:** Raymond Bond (rb.bond@ulster.ac.uk)

**Teaching staff/contact**: Iftikhar Afridi

**A reminder of the learning outcomes for the module:**

1. Appraise appropriate methods to characterise various types of data.
2. Develop and critically evaluate analytical models to automate statistical inference and

analysis of data using machine learning algorithms.

1. Design experiments using appropriate processes to take into consideration of the Legal,

Social, Ethical and Professional issues in the development and use of sustainable,

inclusive and trustworthy systems.

1. Exercise initiative and independent responsibility regarding data integrity and effective

communication, working within a team and participate effectively in the peer review

process.

**Key dates for this assignment:**

* This assignment will be **released in week 1.**
* This **assignment is** **due to be submitted in week 11 (DATE to be added by QAHE at 12pm noon UK time).**
* *“You will receive feedback as per University Guidance which is currently set at 20 working days after submission.”*

**This assignment includes 2 components:**

* **Component 1:** Groupwork submission (worth 50% of the marks).
* **Component 2:** Individual submission (worth 50% of the marks).

**All materials for this assignment should be submitted electronically on Blackboard Learn.** No paper materials are required. Do remember to back up your work, code and computer files etc. When submitting your assignments do keep a copy of the receipt and take a screenshot of the submission screen.

**Remember, plagiarism is a serious offense** **– please ensure that you use your own words** even when referencing a source. Do remember to reference all sources as well as tutorials or code that you used/adapted.

* You can read more about plagiarism on the Ulster website:  
   <https://www.ulster.ac.uk/student/exams/cheating-and-plagiarism>.
* You can also access guidance on the use of AI on the Ulster website: <https://www.ulster.ac.uk/learningenhancement/cqe/strategies/ai>

**By undertaking and submitting your assignments, you agree with the following statement:**

“*I declare that this is all my own work. Any material I have referred to has been accurately referenced and any contribution of Artificial Intelligence technology has been fully acknowledged. I have read the University’s policy on academic misconduct and understand the different forms of academic misconduct. If it is shown that material has been falsified, plagiarised, or I have otherwise attempted to obtain an unfair advantage for myself or others, I understand that I may face sanctions in accordance with the policies and procedures of the University. A mark of zero may be awarded and the reason for that mark will be recorded on my file.*”

**Component 1 – Groupwork**

The module coordinator will allocate students to groups. In your allocated group, all group members are to work together to undertake a data science project and to co-author and submit a **research paper that should be around 4 pages in length (3000 words)** using the IEEE file format/template which can be found online (<https://www.ieee.org/conferences/publishing/templates.html>). The IEEE file format/template (Microsoft Word version) can also be found in the module assessment area on blackboard learn. You can either use the Microsoft Word IEEE template or the IEEE overleaf/LaTeX template. The body text in the IEEE template should be Times New Roman (size 10). Remember effective communication is a key part of writing research papers.

In your group, please decide who will be the team leader. The team leader is responsible for submitting the IEEE research paper on behalf of the whole team/group. The team leader should email **Iftikhar Afridi** **on iftikhar.afridi@qa.com** and inform them of who the team leader is for the group, preferably before the end of week 3 (approximately).

As a group, you might also want to think about using onedrive so that the group can share, write and add content using the same ‘living’ shared document. You should also consider using any other collaborative tools that may help support the communication and interaction between team members, for example, this can include tools such as Microsoft Teams, Blackboard Learn, Git and other tools. You might also consider having a shared data science lab book which could be a shared document that allows all team members to include data analysis and results etc.

In your allocated group, you should collectively undertake a data science project. As a group, you should first select an open dataset that is of interest to the group. The data science project can be related to your experience and is ideally related to a real-world context/challenge. Open datasets can be found on websites such as the UCI repository (<https://archive.ics.uci.edu/datasets>) and Kaggle (<https://www.kaggle.com/datasets>). You may also consider open datasets from PURE (<https://pure.ulster.ac.uk/en/datasets/>) where datasets from the school’s research centres or indeed from other research centres in Ulster may also be available.

As a group, all members should meet regularly and discuss the project. The group should setup a weekly team meeting and also meet during the lab sessions when possible. The group should also keep agendas and minutes of meetings and record those who were and were not in attendance.Your group should also use the scheduled lab time to get continued support in progressing this assignment.

Remember, this groupwork assessment will also include **self and peer assessment**. Self and peer assessment is when each group member evaluates/scores their own contribution to the project and also evaluates/scores the contribution made by the other group members. **This self and peer assessment may be used to adjust your group submission mark in instances where your contribution to the group submission is poor (more details are in this brief).** Each individual student should submit a self and peer assessment before the submission deadline. The self and peer assessment form will either be completed multiple times by you (once for each peer assessment and once for your own self-assessment) using an online Microsoft form, or the self and peer assessment form will be submitted once using a word document submitted to blackboard. The module coordinator will let you know whether you are to complete the self and peer assessment using the online Microsoft form or using a word document (in some circumstances you may be asked to complete the self and peer assessment by submitting an excel spreadsheet to blackboard).

This group project should include the use of a programming language such as R and/or Python as well as the implementation of data science and machine learning techniques to analyse your chosen open dataset. The project should also have a clear aim and objectives. As a group, you should go through a data science/machine learning lifecycle, for example, using CRISP-DM. Your group project should include data importing/cleaning/wrangling, statistical analysis, data visualisation, feature engineering and modelling (machine learning). This group-based data science project should also consider, discuss and/or be guided by the Legal, Social, Ethical and Professional issues and implications across the stages in your project and implement any appropriate processes. You should also consider any factors related to equality, diversity, and inclusion as well as sustainability when implementing your data science project, where appropriate.

**The IEEE research paper should include the following headings/sections:**

1. Title
2. Abstract
3. Introduction
   1. Related works
   2. Aim and objectives
4. Methods
5. Results
6. Discussion
   1. Limitations
7. Conclusion
8. References (in an IEEE referencing style or a referencing style similar to IEEE)

The research paper should detail the data access, the data science techniques, methods and packages that were used as well as any data analysis, results, visualisations and modelling (machine learning). The research paper should also present any evaluation and validation of the data analysis/machine learning etc. Remember to use scholarly references/citations (textbooks, journal papers and conference papers). Also be sure to use an IEEE referencing style or a referencing style similar to IEEE.

**Word limit of the paper**   
The word limit of the paper is 3000 words. We will allow up to 10% over this limit (i.e. 3300 words). However, we may not mark any content beyond 10% over the limit, so please ensure to adhere to the specified agreed limit of 3000 words. There are no other additional penalties applied beyond not marking/assessing content beyond 10% over the limit. The marker should note when this is the case and note in their own records – the approximate point in the paper when they stopped marking (likely at 3300 words). Word limit does not include references.

**You are required to include the number of words in your paper at the end of the document. (e.g. Word limit excluding references = 2965 words).**

**Self and peer assessment**  
Each student will score themselves and each member in their group out of 10 (where 5/10 is a kind of pass mark and a score of 10 is excellent). Students will use the following guide:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Self/peer assessment rubric for students to use. | | | | |
|  | **Score = 7 to 10** | **Score = 6 – 6.9** | **Score = 5 – 5.9** | **Score = 0 to 4.9** |
| **Self/Peer Assessment** | Excellent proactive engagement and contribution to this Groupwork assignment. | Good engagement and made a good contribution to this Groupwork assignment. | Satisfactory engagement but often reactive than proactive. Made a satisfactory contribution to Groupwork assignment. | Score = 0 to 4.9: Very limited engagement with peers and a very limited contribution to this Groupwork assignment. A score of 0 would mean no contribution at all and no engagement at all (e.g. did not attend any meetings). |

The rounded average self/peer assessment scores for each student will be calculated. If a student’s rounded average self/peer assessment score is 5 or above (given 5/10 is deemed a pass and deemed satisfactory) then the student is awarded all of the group paper marks. However, if a student’s rounded average self/peer review score is below 5, their mark for the group paper will be reduced using the following formula:

group paper mark \* (raw-average-self-peer-assessment-score / 5)

Worked example:

For example, if a student’s raw average self/peer assessment score and their rounded average self/peer assessment score was 3 out of 10 and the group paper had a mark of 30 out of 50 – then this individual student would be awarded 18 out of 50 for the group paper submission.

i.e. 30 \* (3 / 5) = 18

**Component 2 - Individual Component (50%)**

**As an individual, you are to submit an individual video presentation (video presentation should be 4 minutes in duration]) to blackboard learn as well as the group code document in PDF**.

Whilst the code document is co-authored by the group and is not part of the marking criteria, you should each submit this individually to evidence your involvement in the group work**.**

In this individual video presentation, you should first very briefly summarise the data science project. In your own words, you should then present the technical code implementation of the project and then describe the technical aspects and code that you mainly contributed to, which should be the main focus of the video. To be clear, you should mainly highlight your individual technical code contributions and discuss your approach to ensuring that your code contributions are readable, updatable, maintainable, and reusable. You can use powerpoint slides and/or demonstrate code in your code editor to describe and showcase the code and technical implementation details. It is appreciated that the code will likely have been co-developed collectively in the group but there will be aspects to the coding/implementation that you have predominantly led on. Also as a team member, you should be able to showcase your technical understanding on all or most other aspects of the code. **Use the assessment criteria in this document to guide/structure your video presentation.**

To create and record the video presentation, you will need to use screen casting software (e.g. Panopto). Please submit your video in .mp4 video file format and please do not use video formats such as .wmv.

**To be clear, you should work on this video presentation individually and not as part of a team, i.e. you should not be sharing your individual video presentation with others or working with others on this individual presentation/submission.**

**Duration limit of the individual video presentation**   
The duration limit of the video presentation is 4 minutes. We will allow up to 10% over this limit (i.e. 4mins and 24 seconds). However, we may not mark any content beyond 10% over the limit, so please ensure to adhere to the specified agreed limit of 4 minutes. There are no other additional penalties applied beyond not marking/assessing content beyond 10% over the limit. The marker should note when this is the case and note in their own records – the approximate point in the video when they stopped marking (likely around 4mins and 24 seconds).

You may wish to record a video and then edit this down to the required duration.

In summary, **the following files should be submitted before the deadline:**

* **Group submission:**
  1. The group/team leader should submit a PDF of the IEEE research paper (3000 words).
* **Individual submission:**

Each student should submit the following:

* 1. Video presentation (4 minutes in length). Preferably submitted in mp4 video file format.
  2. Any slides you may have used in the video (in PDF).
  3. Code implementation from the group project (in PDF). Ensure the PDF includes the code as text, i.e. do NOT include screenshots of code.
  4. Self and peer review form/s. The module coordinator/teaching team will provide a self-peer assessment word document and/or an online Microsoft form to complete self and peer assessments.

**General advice/notes:**

1. Coursework submitted late may not be marked and you may receive a mark of 0.
2. Ensure that the video presentation is uploaded correctly (check the video is playing in the browser within blackboard) – otherwise it may not be marked.
3. You may need to record the video multiple times so do allow yourself enough time to record your presentation.
4. Be sure to read the assessment criteria/rubrics to help guide your submissions.
5. Ensure code is structured and well commented.
6. **Do not plagiarise.** Use quotes when quoting a source. Include references and citations. Make sure you understand what plagiarism is and the penalties for plagiarism offences (<https://www.ulster.ac.uk/student/exams/cheating-and-plagiarism>).
7. All submitted assignment files should have the file name: "SurnameFirstNameBNumber"

e.g. BrownJohnB00001234\_Assignment.

**Assessment criteria/ rubrics**

Note: Rubrics are indicative.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Groupwork Submission Mark (50% of CW2) – Research Paper | | | | |
| Criteria | Distinction (70-100%) | Commendation  (60-69%) | Pass (50-59%) | Fail (0-49%) |
| **Overview of domain and existing work (10%)** | 7 – 10 marks  Excellent articulation of the domain, aim and objectives with exceptionally well written abstract and introduction. Excellent literature review with in-depth critically review of the research area which support the identified opportunity. | 6 – 6.9 marks  Good articulation of the domain, aim and objectives with a well written abstract and introduction. Evidence of good and extensive search of literature in the research area. | 5 – 5.9 marks  Adequate articulation of the domain, aim and objectives with a adequately well written abstract and introduction. Evidence of searching the literature. | 0 – 4.9 marks  Limited articulation of the domain, aim and objectives with a poorly written abstract and introduction to the domain. Little evidence of searching the relevant literature. |
| **Methodology and Results (20%)** | 14 – 20 marks  Excellent justification of the selected statistical and machine learning methods. Excellent use of statistics to characterise the dataset. Design of the experiment is comprehensive and innovative. Thorough results and critical evaluation of the proposed model with consideration to many aspects of the performance, perhaps including benchmarking. Excellent use of graphs and/or tables that were very well designed. | 12 – 13.8 marks  Good justification of the selected statistical and machine learning methods. Good use of statistics to characterise the dataset. Design of the  experiment is clear and appropriate. Well presented results and a good critical evaluation of the model. Good use of graphs and/or tables that were well designed. | 10 – 11.8 marks  Adequate justification of the selected statistical and machine learning methods. Adequate use of statistics to characterise the dataset – more statistical analysis and/or exploratory data analysis needed. Design of the experiment is adequate. Satisfactory results and evaluation of the model performance. A more critical evaluation would be useful. Could have made better use of well-designed graphs and/or tables. | 0 – 9.8 marks  Insufficient justification of the selected statistical and machine learning methods. Poor and/or very limited use of statistics to characterise the dataset – much more statistical analysis and/or exploratory data analysis needed. Poor presentation of results and evaluation of the model and/or a lack of detailed results and analysis. Really lacking in a detailed critical evaluation. Could have made better use of well-designed graphs and/or tables. |
| **Effective Communication, Discussion, Conclusions and References**  **(10%)** | 7 – 10 marks  Excellent awareness of the insights gained from the results, the limitations of the project and methods employed. Outstanding use of referencing from highly reputable sources. Conclusions are clear and reasonable. Excellent effective communication, layout and clarity. | 6 – 6.9 marks  Good awareness of the insights gained from the results, the limitations of project and methods employed. Comprehensive use of referencing though scope for more/higher quality articles to be included. Conclusions are clear and reasonable. Good effective communication, layout, and clarity. | 5 – 5.9 marks  Adequate awareness of the insights gained from the results, the limitations of project and methods of employed. References adequate, generally of good quality and used appropriately but with room for improvement. Adequate effective communication and layout etc. More clarity needed. | 0 – 4.9 marks  Limited awareness of the insights gained from the results, the limitations of project and methods of employed. Small number of references used, however these are of poor quality and/or not relevant. Poor effective communication and/or layout. More clarity needed. |
| **Description of professional, ethical, social and sustainability issues within project (10%)** | 7 – 10 marks  Comprehensive discussion of issues and adoption/consideration of appropriate Legal, Social, Ethical and/or Professional practices. The project was clearly guided by considering these factors. Also, excellent consideration of sustainable, inclusive, and trustworthy systems as well as equality and diversity. This is explained in an excellent level of detail. | 6 – 6.9 marks  Good discussion of issues and adoption/consideration of appropriate Legal, Social, Ethical and Professional practices. Also, good consideration of sustainable, inclusive, and trustworthy systems as well as equality and diversity.  Mostly comprehensive but perhaps lacking some detail. | 5 – 5.9 marks  Adequate discussion of issues and adoption/consideration of appropriate Legal, Social, Ethical and Professional practices. Adequate consideration of sustainable, inclusive, and trustworthy systems as well as equality and diversity.  However, some aspects were not explained clearly or complete. | 0 – 4.9 marks  Poor or no serious discussion of issues and adoption/consideration of appropriate Legal, Social, Ethical and Professional practices. Limited or no consideration of sustainable, inclusive and trustworthy systems. More consideration for equality and diversity is also needed. |

Total \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ out of 50

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Individual Submission (Mark 50% of CW2) – Video presentation  Only the video presentation will be marked against the following criteria. | | | | |
| Criteria | Distinction (70-100%) | Commendation  (60-69%) | Pass (50-59%) | Fail (0-49%) |
| **Short overview of the project (10%)** | 7 – 10 marks  Video included an excellent short overview of the project. Overview was extremely clear and very effective in summarising the main stages. | 6 – 6.9 marks  Video included a good short overview of the project. Could be slightly more effective and/or clearer in parts. | 5 – 5.9 marks  Video included an adequate short overview of the project. Summary could have been more effective and/or could have been clearer. | 0 - 4.9 marks  Overview of the project was not provided or was lacking structure, insight, effectiveness and/or clarity. |
| **Individual contribution and understanding of the technical implementation of the solution (i.e. understanding of code etc.) (30%)** | 21 – 30 marks  Video included excellent insight and clear description and justification of the code and technology used for the implementation. Individual understanding and individual contribution are clearly evident given the level of understanding provided. Demonstration of the code was excellent. There is evidence of innovative practice to ensure that the code is readable, updatable, maintainable, and reusable. | 18 – 20.7 marks  Video included good insight and description and justification of the code and technology used for the implementation. Individual understanding and individual contribution is good especially given the level of understanding provided. Demonstration of code was good. There is practice to ensure that the code is readable, updatable, maintainable, and reusable. | 15 – 17.7 marks  Video included adequate high level description and justification of the code and technology used for the implementation. Individual understanding and individual contribution is adequate especially given the level of understanding provided. Demonstration of code was adequate. There was no real evidence and/or presentation of how the code was created to be readable, updatable, maintainable, and reusable. | 0 – 14.7 marks  From the video, there is a lack of understanding of the technical implementation and the code. Poor description and justification of the code and technology used for the implementation. Individual understanding and individual contribution are poor. More understanding of the code is needed. Demonstration of code was poor. There was no evidence and/or presentation of how the code was created to be readable, updatable, maintainable, and reusable. |
| **Effective communication, and overall quality of the presentation (10%)** | 7 – 10 marks  Excellent presentation/demonstration that was very well organized and was very professional with excellent content and clarity. Excellent effective communication. | 6 – 6.9 marks  Good presentation/demonstration that was well organized with good content. Perhaps some more detail and/or clarity was needed. Good effective communication. | 5 – 5.9 marks  Adequate presentation/communication/ demonstration but could have been better. More detail and/or clarity was needed. | 0 - 4.9 marks  Poor presentation/communication/ demonstration. There was a real lack of detail and/or clarity throughout. More structure was needed. |

Total \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ out of 50