OXFORD BROOKES UNIVERSITY

MSc Advanced Computer Science

Postgraduate Diploma in Advanced Computer Science

COMP7032 Data Science and Machine Learning

Module Guide Semester 2, 2023- 24

School of Engineering, Computing and Mathematics

Faculty of Technology, Design and Environment

Module Leader: Inna Skarga-Bandurova

Office: E2.09, E Block, Wheatley

Phone: 01865 484584

Contents

Module introduction

Module leader contact details

Name: Inna Skarga-Bandurova

Room: E2.09 Wheatley Campus

Telephone: 01865 484584

Email: iskarga-bandurova@brookes.ac.uk

Office hours: Please see Google Calendar

Seminar leader contact details

Name: Bríd Brosnan

Email: bbrosnan@brookes.ac.uk

Office hours: Please see Google Calendar

Academic Liaison Librarian

Name: Penny Robertson

Email: probertson@brookes.ac.uk

Changes made to the module (if any) in response to student and other feedback

The lecture and seminar topics were updated to give students some additional time and opportunity to follow the coursework tasks more coherent.

Detailed marking schemes for Coursework 1 and Coursework 2 were added.

The submission date for CW1 is shifted by one week to allow students time to polish their reports after the presentation sessions taking into account suggestions received during the discussion.

Module Study Plan

Unit / Week	Lecture Topic	Topic for practical classes	Seminar/ Tutorial / Discussion
1	Introduction to the module; foundations of machine learning	Introduction to R. Exploring the real-world data.	CW set up. Data science problem proposal discussion
2	Regression: Linear regression.	Linear regression	Data preparation methods. **Data science problem proposal** – **due** (not marked)
3	Classification: Logistic Regression and k-Nearest Neighbours (kNN)	Logistic regression and visualisation. kNN and k-folding	Data exploration, transformation, visualisation and pre-processing techniques
4	Tree based methods: decision trees and random forests	Decision tree learning	Data exploration and preparation
5	Consolidation	Q&A session	**Propose own ML solutions** and **discuss with tutors** – due
6	Unsupervised Learning: Clustering, Principal Component analysis	Clustering, PCA	Design, implementation, modelling and evaluation
7	Unsupervised Neural Networks. SOM	SOM	Design, implementation, modelling and evaluation
8	Basics of Neural Networks. Perceptron.	Neural Networks	Report (Paper) preparation: Organising academic research papers
9	Support vector machines	SVM	Report (Paper) preparation: Ethical, legal, social and security issues in ML
10	Basics of Deep learning. Convolutional Neural Networks (CNN)	CNN	Presentation preparation:
11	CW2 Oral Presentations	CW2 Oral Presentations	**CW2: Presentation** submission deadline – **due**
12	Consolidation		**CW1: Report (Paper)** submission deadline – **due**

Module syllabus

Module aims

To provide students with the knowledge and skills necessary to design, develop and evaluate a complete data science project in a modern organisational context. This includes an introduction to Machine Learning, informed by the School's research in this area..

Learning outcomes

On successful completion of this module, students will be able to:		Brookes Attribute developed	Other Attributes developed, if applicable
1	Critically analyse and discuss key concepts in data science, including tools, approaches, and applications.	Research Literacy	Academic Literacy
2	Utilise effectively appropriate methods and tools for exploring, modelling, visualising and analysing data	Academic Literacy	Digital and Information Literacy, Research Literacy
3	Report data science results to a wider audience by tailoring them at different levels of detail.	Critical Self-Awareness & Personal Literacy	Academic Literacy
4	Design, implement and critically evaluate a data science product that addresses a given real world data problem.	Academic Literacy	Digital and Information Literacy, Research Literacy
5	Work in a context informed by relevant ethical, legal, social and security issues.	Active Citizenship	Critical Self-Awareness & Personal Literacy, Academic Literacy

Outline Syllabus

The module is designed to provide the students with an introduction to the most important and fundamental topics in machine learning, in particular:

- Introduction to the module; foundations of machine learning.
- Linear Regression
- Classification: Linear Discriminant Analysis and Logistic Regression
- K-nearest neighbour and model evaluation. Variable subset selection
- Tree based methods: decision trees and random forests
- Unsupervised Learning: Clustering, Principal Component analysis
- Support vector machines
- Neural Networks and Self Organising Maps
- Deep learning
- Convolutional Neural Networks
- Generative Adversarial Networks

Recommended reading list

Author(s)	Title	Publisher	Year
James G., Witten D., Hastie T. and Tibshirani R.	*An Introduction to Statistical Learning with Applications in R*	Springer, 2 ed.	2021
Zhao Y.	*R and Data Mining: Examples and Case Studies*	Elsevier	2015
Harrington P.	*Machine Learning in action*	Manning	2012
Wickham, H., Grolemund G.	*R for data science*	O'Reilly Media	2016

Assessment information

This module follows the principles of the University's Assessment and feedback policy developed in conjunction with the Student Union, to ensure good practice and transparency in assessment and feedback processes. The Assessment and feedback policy can be found in your Programme Handbook or on your Programme's Brookes Virtual site as well as on the University webpage of A-Z of Policies.

Please note: the Institutional University Handbook which will provide you with information that is central to your studies, including policies and regulations, student support and wellbeing and all the services available to you through Student Support. The core information is also available on Moodle via the drop down menu under 'Student Help'

The coursework is worth 100% of your final grade and consists of two parts:

- CW1 - the coursework report, executed as research paper (85%), and
- CW2 - a presentation (15%).

The overall assignment grade is determined as the weighted average of the report and the presentation.

Both items, the research paper and the presentation, should be submitted as two separate files via Turnitin using the links provided on Moodle. You are required to complete both CW1 and CW2.

Coursework 1

Your Assignment Brief can be found in Appendix 1 and Appendix 2 of this module guide. Here you will also find the assessment criteria outlining how marks will be awarded. You will be taken through each piece of coursework in a full briefing during your scheduled teaching hours. Details of how you should present your work and when the work is due are detailed below.

Submission date and instructions

Your CW1 must be uploaded to the dropbox in Moodle by 13:00 on Monday, week 12.

For electronic submission: Give the file(s) which you intend to upload a name which begins with your student number, for example, '12345678_CW1'.

Please ensure you submit your assignment no later than the deadline set above (these are fixed deadlines, but students may exceptionally secure an extension if last minute untoward circumstances affect your ability to submit on time). Please note the use of this extension is monitored and restrictions in place for overuse. You must make a request for any form of extension so please follow the link below to identify how.

https://www.brookes.ac.uk/students/your-studies/exceptional-circumstances/

The Blue Marking Card adjustment is only available to students who have an Inclusive Support Plan (ISP) specifying this adjustment. Eligible students who wish to use this adjustment must add a blue card:

https://www.brookes.ac.uk/students/inclusive-support-service/exams-and-assessments/blue-marking-cards

Recommendations for Reasonable Adjustments are made in accordance with the provisions of the Equality Act 2010. These are detailed in Inclusive Support Plans (ISPs) and need to be implemented unless there is a clear rationale for this not being possible, in which case we are accountable as an HEI for this decision. Alternatives should be considered, and further advice sought from the Inclusive Support Service to ensure we are compliant, consistent, and following best practice.

It may not be possible to give the full coursework extension period agreed in all Inclusive Support Plans for the coursework assignment. If your extension will take your deadline past the University final deadline for submission of work you will not be able to use the full extension. Consideration must be given, when determining a deadline for an extension period, to allowing enough time for submission of work and marking, so that the module and marks can be taken to the correct examination committee. If you have an Inclusive Support Plan you can check the full details of the adjustments, including whether you have coursework extensions in

https://generalssb-prod.ec.brookes.ac.uk/BannerExtensibility/customPage/page/student_ISP

If you have a coursework extension as part of an Inclusive Support Plan, contact your module leader if you are unclear about your deadlines.

Students who have an extension because of an ISP or a successful exceptional circumstances application (or both) can also use an extension if the need arises.

Contact the Inclusive Support Service if you would like to request a review of your Inclusive Support Plan (ISP) or to have your needs assessed for an ISP.

Coursework 2

Your Assignment Brief can be found in Appendix 1 of this module guide. Here you will also find the assessment criteria outlining how marks will be awarded. You will be taken through each piece of coursework in a full briefing during your scheduled teaching hours. Details of how you should present your work and when the work is due are detailed below.

Submission date and instructions

Your CW2 must be uploaded to the dropbox in Moodle by 13:00 on Monday, week 11.

The oral part of CW2 is delivered individually during scheduled time slots on Monday, week 11.

For electronic submission: Give the file(s) which you intend to upload a name which begins with your student number. For example '12345678_CW2'

Please ensure you submit your assignment no later than the deadline set above (these are fixed deadlines, but students may exceptionally secure an extension if last minute untoward circumstances affect your ability to submit on time). Please note the use of this extension is monitored and restrictions in place for overuse. You must make a request for any form of extension so please follow the link below to identify how.

https://www.brookes.ac.uk/students/your-studies/exceptional-circumstances/

The Blue Marking Card adjustment is only available to students who have an Inclusive Support Plan (ISP) specifying this adjustment. Eligible students who wish to use this adjustment must add a blue card:

https://www.brookes.ac.uk/students/inclusive-support-service/exams-and-assessments/blue-marking-cards

Recommendations for Reasonable Adjustments are made in accordance with the provisions of the Equality Act 2010. These are detailed in Inclusive Support Plans (ISPs) and need to be implemented unless there is a clear rationale for this not being possible, in which case we are accountable as an HEI for this decision. Alternatives should be considered, and further advice sought from the Inclusive Support Service to ensure we are compliant, consistent, and following best practice.

It may not be possible to give the full coursework extension period agreed in all Inclusive Support Plans for the coursework assignment. If your extension will take your deadline past the University final deadline for submission of work you will not be able to use the full extension. Consideration must be given, when determining a deadline for an extension period, to allowing enough time for submission of work and marking, so that the module and marks can be taken to the correct examination committee. If you have an Inclusive Support Plan you can check the full details of the adjustments, including whether you have coursework extensions in

https://generalssb-prod.ec.brookes.ac.uk/BannerExtensibility/customPage/page/student_ISP

If you have a coursework extension as part of an Inclusive Support Plan, contact your module leader if you are unclear about your deadlines.

Students who have an extension because of an ISP or a successful exceptional circumstances application (or both) can also use an extension if the need arises.

Contact the Inclusive Support Service if you would like to request a review of your Inclusive Support Plan (ISP) or to have your needs assessed for an ISP.

An oral part of assignment CW2 (5%) is an event assessment and therefore ISP coursework extensions cannot be given to this part. Exam adjustments (such as extra time and rest breaks) agreed through Inclusive Support Plans will be given. If this is a presentation, other adjustments may apply if specified in your ISP. If you have an Inclusive Support Plan you can check the full details of the adjustments, including whether you have an extra time allowance in

https://generalssb-prod.ec.brookes.ac.uk/BannerExtensibility/customPage/page/student_ISP

If you have a coursework extension as part of an Inclusive Support Plan, contact your module leader if you are unclear about your deadlines.

Students who have an extension because of an ISP or a successful exceptional circumstances application (or both) can also use an extension if the need arises.

Contact the Inclusive Support Service if you would like to request a review of your Inclusive Support Plan (ISP) or to have your needs assessed for an ISP.

Marking and moderation of your work

Following internal moderation, a sample of work is reviewed by the External Examiner for the programme to ensure that the standards applied are comparable to those at other institutions. To read how your work is moderated please go to your programme handbook for details.

Appendix 1: Coursework Assignment

COMP7032
Data Science and Machine Learning

Coursework Assignment (Semester 2, 2024)

Learning Outcomes

The coursework counts 100% of the total assessment of this module. It is designed to develop and assess your attainment of all learning outcomes, in particular

1. Critically analyse and discuss key concepts in data science, including tools, approaches, and applications.
2. Utilise effectively appropriate methods and tools for exploring, modelling, visualising and analysing data.
3. Report data science results to a wider audience by tailoring them at different levels of detail.
4. Design, implement and critically evaluate a data science product that addresses a given real world data problem.
5. Work in a context informed by relevant ethical, legal, social and security issues.

Task

The coursework includes two parts. You are required to complete both parts to pass this module.

Part	Task	Instruction
CW1: Written Assignment 1 (85% of the module mark)	Design, implement and critically evaluate a data science product that addresses a given real world problem, taking into account relevant ethical, legal, social and security issues. Produce a technical report in form of Research Paper.	Upload your Report as a single pdf file with a name that starts with your student number, of the form "*StudentNumber*_CW1"
CW2: Written Assignment 2 & Oral Presentation (15% of the module mark)	Presenting the results of the CW1, taking into the nature of the audience. Prepare a presentation and demonstrate the results.	1. Upload your Presentation as a single pptx file, of the form "*StudentNumber*_CW2". 2. Oral presentations will take place during the usual slots of Semester 2 (Weeks 11&12)

Presenting Coursework 1 and Coursework 2 for assessment

Assignment format

Coursework 1 must be presented as a single PDF file in the following format:

- It must be word-processed in 12 points Arial font and single-spaced, spacing after – 10 pt.
- All pages must be numbered.
- Margins must be as follows: Top: 1 inch, Bottom: 1 inch (2.5 cm), Left: 1.25 inches, Right: 1.25 inches (3.2 cm).

Coursework 2 must be presented as a single PPTX file in the following format:

- It should contain up to 10 slides
- All pages except the title page must be numbered.

Assignment length

The length of an assignment is limited by a set number of words to contribute towards the development of writing skills and to ensure all work is assessed equitably. We therefore require you to complete your assignments within the number of words specified in the assignment brief. The specified word count refers to the main body of the report and does not include front cover, title page, contents page, executive summary, reference list, bibliography or appendices. The word count does include headings, tables and in-text citations, but not equations or diagrams.

Appendices themselves will not be marked. However, inappropriate use of appendices will be taken into consideration when awarding the final mark.

Words that exceed the maximum stated words (+10%) will not be marked. If in doubt, you should discuss this with the Module Leader before submission.

Turnitin Assignments

Coursework 1 and Coursework 2 on this module will be submitted through Turnitin. Turnitin is a web-based tool that supports the development of good academic practice when preparing written work for assessment. This text-matching tool allows academic staff to check assignments for improper use of sources or potential plagiarism by comparing it against continuously up-dated databases (including web-pages and other student work).

The new Turnitin policy effective from September 2020 can be found via this link. https://www.brookes.ac.uk/regulations/

You are expected to be familiar with it. Guidance to help with interpretation and use of Turnitin is provided here: https://www.brookes.ac.uk/students/your-studies/turnitin/

Feedback

Feedback on your work will be provided in a range of ways at various times throughout this module, and different feedbacks will serve slightly different purposes.

Feedback is designed to support your learning and help you to improve subsequent work, so you need to engage and get the most out of the feedback provided. Please note that feedback is provided throughout the module not just on formal tasks. It will be provided on your work and contribution in class, on the formal assessment tasks and, in some circumstances, during staff

office hours. In particular, students will be provided with opportunities for formal feedback in the tutorial sessions. Students will be given a formative submission point for each assignment where verbal formative feedback is relevant. For the first assignment, students will be expected to summarise their formative feedback and say in the summative submission how they have responded to it. The tutorial sessions in weeks 2, 5 and 10 will be used to discuss your progress in the coursework and prepare initial submissions so it is important to follow the lectures especially in the first few weeks and think immediately about how to apply the material you have learned.

The submission date for CW1 is shifted by one week (from week 11 to week 12) to allow you time to polish your reports after the presentation sessions taking into account suggestions received during the discussion.

Please note that all marks are provisional until they are ratified by an Examination Committee. Following internal moderation, a sample of work is reviewed by the External Examiner for the programme to ensure that the standards applied are comparable to those at other institutions.

Submission Dates

Coursework 1: Monday of Week 12, 29th April 2024, 13:00.

Coursework 2: Monday of Week 11, 22nd April 2024, 13:00.

Submission Guidelines

Please ensure you submit your assignment no later than the deadline set above (these are fixed deadlines but students may exceptionally secure a 24 hour grace period if last minute untoward circumstances affect your ability to submit on time). Please see your Programme Handbook for more details.

Please note the use of this grace period is monitored and restrictions in place for overuse.

For student who are registered with the Dyslexia/SpLD Service, any submission through Moodle and Turnitin will trigger a notification of a Blue Card and there is no action required by a student.

Requests for extensions should be made via the Extenuating Circumstances procedure as usual.

Academic Conduct Regulations and Procedure:

The Coursework 1 and Coursework 2 are the individual assignments, therefore evidence of academic misconduct such as group work or collusion will be penalised.

For more information see: https://www.brookes.ac.uk/students/sirt/student-conduct/academic-misconduct/

Any use of the work of other people, including code and algorithms, must be clearly referenced.

See the regulations for details: https://www.brookes.ac.uk/regulations/

Resit coursework requirements

Do refer to the University Resit and Retake Policy https://www.brookes.ac.uk/students/your-studies/resits-and-retakes/

Coursework 1 – Technical Report (Research Paper)

Goal: You are required to create a report in the form of a research paper, based on a chosen data science problem related to machine learning.

Reward: 85% of the module mark
Total word count – 4,000 words
Set – Week 1
Due – Week 12

Problems in data science vary from one domain to another. In this coursework, you will select a dataset related to a real-world problem that can be solved by machine learning and best suits your area of interest. Your topic may be related to your graduate research or any other class project. However, you *must* make the relationship to your other work clear in your proposal and final report, and describe how you have extended this research for this module.

For this assignment, you will have to complete the following stages and write a report in the form of a research paper:

1. Data science problem proposal: This proposal should demonstrate a clear understanding of the chosen problem, data and methods.

2. Data exploration and preparation: The nature of the dataset may dictate some specific data exploration and preparation that can help inform the solutions. For example, higher dimensional datasets (those with too many attributes/columns) may require applying a data reduction method such as Principal Component Analysis (PCA).

3. Propose solution(s): In this step, you will choose **three** different machine learning algorithms to solve selected real-world problem.

4. Design, implementation, modelling and evaluation: design, model and implement the proposed algorithms in R/Python, and critically evaluate the solutions (analyse the performance according to a set of metrics).

All experiments must be thoroughly explained and accompanied with graphs and plots, where appropriate, and a set of R/Python scripts from which they can be reproduced.

Dataset

There is an abundance of websites that provide publicly available datasets. Example problem domains can be found here:

- A categorised list of datasets from GitHub https://github.com/caesar0301/awesome-public-datasets
- The UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php is another long-standing source of benchmark datasets for data mining and machine learning research.

- Kaggle https://www.kaggle.com/datasets also contains interesting real-world problems and datasets.

You can select a dataset from the above sources, or another one that is available online. Note that the dataset should be *publicly available*.
The chosen dataset should have a minimum of 1,000 instances (rows) and a minimum of 5 attributes / features (columns).

If you have difficulty finding an appropriate topic, feel free to talk with the module leader and co-teachers during office hours or post to the Question Form on Moodle.

Suggested Contents of the Report (Research Paper):

Title (draw readers' interest) – 8-15 words. Your paper title should be specific, concise, and descriptive.

Abstract (mini version of the paper) – **200 words**. A brief description of the content of the paper. Your abstract should also include the sentences summarising:

- The problem statement
  "This paper presents …."
- Indication of methodology
  "Using… (*something, you did not invent*)…. it was obtained …"
- Main findings and principal conclusion
  "Results show significant improvement over …"

Verb tense: Simple-past – refers to work done.

Introduction (evidence of importance, research questions) – 300 words. Introducing the problem and its significance, and outlining the contributions of your paper. Clearly describe the problem to be solved. Please reference any prior work that you plan to build upon.

- Claiming centrality: Make topic generalisations, review items of previous research.
  "In the past decade much research has focused on …"
- Finding a niche (counter-claiming): indicating a gap, question raising.
  "It remains unclear why/how…"
- Occupying the niche: outlining the purposes, announcing principal findings, indicating structure
  "The purpose of this study is to …"

Verb tense: Present – refers to established knowledge

State-of-the-art (review existing solutions, find a gap, look for debates) – 600 words. A short literature review of related work and benchmark solutions. As you prepare this part, consider the following directions:

- Literature on topic
- Literature on method
- Theoretical approach

This part should end with a paragraph describing benchmark (reference) model(s) that can be used to compare with yours.
Verb tense: Present – refers to established knowledge in literature

Methodology (study design, data sources, techniques) – 400 words. The section answers two main questions: (1) How were the data collected or generated? (2) How were they analysed?

- The sources and /or process of data collection
  "The data used for this study were collected by …"
- Techniques for data analysis
  - Data exploration and feature selection.
  - Approach - Describing your proposed solution(s).

The focus should be on how you *applied a method*, not on the mechanics of *doing a method.*
Verb tense: Simple past – refers to work done

Results (experiments, contribution) – 900 words

- Experiments - Presenting your results and a providing a critical discussion. The experiments should be described in sufficient detail that someone with a reasonable background in the area could reproduce your results.

Each subsection should include observations during experiments, observations about the results (compare/contrast different experiments) and appropriate performance metrics.
Verb tense: Simple past – refers to what was found, observed

Discussion (demonstrate your results, compare with benchmark) – 800 words

- Introduction: review findings, discuss outcomes, stake a claim
  "The findings of this study clearly show that …"
- Evaluation: analyse, offer explanations, reference the literature, state implications
  "One explanation for…"
- Conclusion: limitations, recommendations
  "The study is limited by…"

Verb tense: Present – emphasis on established knowledge, present results

Conclusion (summarise your principal findings) – 300 words. Your conclusion should refer back to your introduction, summarize three main points of your paper and wrap it all up with a final observation.

- Restate your research topic.
- Summarize the main points of your research.
- Connect the significance or results of the main points.

Verb tense: Present – emphasis on what should now be accepted as established knowledge

Legal, social, ethical, security and professional issues – 200 words.

Section should include a consideration of the wider legal, social, ethical, security and professional issues associated with the usage of machine learning for chosen real-world problem.

References

You are required to cite the work of others used in your research to avoid plagiarism and collusion. Remember each loan should have at least one citation (use the university recommended referencing style).
Referencing can follow any accepted style, as long as it is consistent. You can use Citing your references using the Harvard as default.

Appendices

Append your original source code (e.g., R code) or provide the link to an external repository where your code can be accessible to provide evidence of having run the experiments.

Note that **references** and **appendices** *are not counted* towards paper length. Therefore, if you need some extra space to explain the experiments or have some information you would like to include in your paper (e.g., big tables, images, methodology, etc.), you can include supplemental information as an appendix to your work.

These are generic section titles, which you may adapt appropriately to the application/problem that is investigated. You may include sections describing modifications of algorithms or developments that are novel and specific to your work. You may also include figures, tables, pseudo-code, and appendices with actual code.

Recommended software

We are **agnostic about your choice of code implementation** (although working in R would be a natural choice). Citations should be provided for any algorithms and software not developed in the scope of the project.

Report format

You can use the AAAI, AIAA, NIPS, or IEEE paper templates (the LaTeX version looks better, but Word will be accepted) – or you can use the template of another conference of your choice.

By default, please follow the formatting guidelines of IEEE Manuscript Template for Conference Proceedings (A4)

http://www.ieee.org/conferences_events/conferences/publishing/templates.html

Paper examples

Although there are plenty of research paper examples, for student projects we recommend starting from this library: https://web.stanford.edu/class/aa228/cgi-bin/wp/old-projects/

Additional information about organizing academic research paper can be found here.

Marking criteria

There is no ‘correct’ choice of topic or approach for your paper. In addition, your marks will not depend on who gets the best performance (since you might work on different tasks, it would be impossible to find out anyway), but on the clarity of the contribution, the depth of your analysis and the interest of your solution. Other parameters (in descending values) include:

Technical quality

- Significance of contribution
- Evidence and quality of the experiments
- Appropriateness of the approach
- Quality of discussion
- Quality of analysis of prior works
- Quality of conclusions

Analysis and discussion of the PLESS aspects

- Analysis of Professional, Legal, Ethical, Social, and Security issues.

Quality of the report

- Adequacy of the title
- Appropriateness of the abstract
- Structure of the paper
- Writing quality
- Adequacy of the references
- Clarity of the drawings, graphs, and tables

Marking Rubric for CW1 – Technical Report (Research Paper) 85%

Grade	ACADEMIC QUALITY OF THE PAPER
A - A+ 70%-100%	The title is specific, concise and descriptive, and indicates accurately the subject and scope of the study. Abstract is accurate (correctly reflect the contents of the paper), complete (covers the major parts of project) and non-evaluative. It is also concise (no excess wordiness or unnecessary information) and clear (readable, well-organised, not too jargon-laden). An introduction clearly identifies the subject area of interest. Succinct and clear problem description appropriate for machine learning. A clear statement of the rationale for your approach to the problem studied. Outstanding evidence of systematic review using multiple high quality academic sources. Logical, clear development of narrative. High quality references and citations. The closely related publications are summarized and cited providing adequate background to the research problem. The important features / results of similar studies are synthesized and relations to your own work is described. A gap in research is identified and description of own research problem is provided. Excellent explanation of study design and/or the overall methodological approach for investigating own research problem. An explanation of data sources and/or process of data collection is clear. A chosen approach for data analysis is justified. The description of study design, how you gathered the data, and the protocol for analysing the data are organized chronologically. Excellent technical quality (rigour of the experiments, data preparation, justification and correct application of the selected algorithms and suitability of the selection). Produced and demonstrated a comprehensive, high quality solution to the problem. There is some element of a novel approach to the problem or novel use of techniques. The approach described was quantitatively analysed according to a set of clearly described metrics, and the results of this analysis presented as figures/tables/graphs/etc. A paper, which could be, with minor modifications, suitable for a publication – or form the basis for a postgraduate project. Outstanding evaluation and discussion of the significance of the results (Why the results are important? How does the paper advance the state of the art? How would the results be useful to other researchers or practitioners? Is this a “real” problem or a small “toy” problem?). Excellent comparison of your results with the findings from other studies as well as description of lessons learned and potential limitations and weaknesses of your approach. The conclusions follow logically from the presented results, describe your main points, explain their significance and suggestions for further research. Legal, social, ethical, security and professional issues are fully considered.

	Appendices provide clear evidence of running the experiments with code that is excellently organized and commented. Sufficient information for the reader is provided to reproduce the results.
B - B+ 60% - 69%	Very good technical quality of the report. Produced and demonstrated very good quality solution to the problem. Very good evidence of systematic review using multiple high quality academic sources. Logical, clear development of narrative. Very good evaluation and discussion of the significance of the results. Appendices provide good evidence of running the experiments with code that is well organized and commented. Sufficient information for the reader is provided to reproduce the results. Legal, social, ethical, security and professional issues fully considered. Appropriate references and citations.
C – C+ 50% - 59%	Good technical quality. Produced and demonstrated good quality solution to the problem. Good evidence of reviewing multiple academic sources. Some references and citations. Good evaluation and discussion of the significance of the results. Legal, social, ethical, security and professional issues fully considered. Appendices provide an evidence of running the experiments with code that is organized and commented.
FPR 40% - 49%	Poor technical quality. Produced and demonstrated a solution to the problem. Appendices provide an evidence of running the experiments. Code lacks clarity. The solution either lacks depth or exhibits substantial, but not fatal weaknesses. Poor evidence of reviewing academic sources. Some references and citations. Some evaluation and discussion of the results. Legal, social, ethical, security and professional issues considered.
F 0-29%	Very poor adequate technical quality. Produced and demonstrated a solution to the problem, which is fatally flawed, despite some effort. Appendices do not include code. Poor evidence of reviewing academic sources. Little evaluation and discussion of the results. Little consideration of legal, social, ethical, security and professional issues. Narrative difficult to follow. Poor quality of references and citations.

Detailed Marking Scheme for CW 1 – Technical Report (Research Paper)

COMP7032 Coursework 1 - Marking and Feedback		
Student Name:		
Student Number:		
Total Mark:	0	*Out of 85*
Grade:	0	*Out of 100%*

Criteria/Quality of Work	Marks available	Marks awarded
Title	5 marks	
The title indicates accurately the subject and scope of the study (specific, concise and descriptive).	5	
Abstract	5 marks	
Abstract is accurate (correctly reflect the contents of the paper), complete (covers the major parts of project) and non-evaluative.	3	
It is concise (no excess wordiness or unnecessary information) and clear (readable, well-organised, not too jargon-laden).	2	
Introduction	10 marks	
An introduction clearly identifies the subject area of interest.	4	
Succinct and clear problem description appropriate for machine learning.	3	
A clear statement of the rationale for your approach to the problem studied.	3	
Literature Review (State-of-the-art)	10 marks	
The closely related publications are summarized and cited providing adequate background to the research problem.	4	
The important features / results of similar studies are synthesized and relations to your own work is described.	3	
Identification of a gap in research and description of your research problem.	3	
Methodology	10 marks	
Study design and/or the overall methodological approach for investigating your research problem (how you organised your study).	3	
An explanation of data sources and/or process of data collection.	3	
Justification of chosen approach for data analysis.	3	
The description of (1) study design, (2) how you gathered the data, and (3) the protocol for analysing the data are organized chronologically.	1	
Results	20 marks	

Technical quality of the solution (rigour of the experiments, data preparation, justification and correct application of the selected algorithms and suitability of model selection).	10	
The approach described was quantitatively analysed according to a set of clearly described metrics, and the results of this analysis presented as figures/tables/graphs/etc.	10	
Discussion	10 marks	
Explanations for the results and commenting on whether or not the results were expected	2	
Comparison of your results with the findings from other studies	4	
Description of lessons learned and potential limitations and weaknesses of your approach	4	
Conclusion	10 marks	
The conclusions follow logically from the presented results, describe your main points and explain their significance	5	
Suggestions for further research	5	
Legal, social, ethical, security and professional issues	5 marks	
Legal, social, ethical, security and professional issues fully considered.	5	
Writing quality of the report	-10 marks	
Structure of the paper	-2	
Proofreading the paper	-2	
Clarity of drawings, graphs and tables	-2	
Adequacy of references	-2	
Code implementation	-2	
	Total:	0

Coursework 2 – Presentation

Goal: The goal of this coursework is to effectively communicate the initial results of your research and development project, considering the audience's nature. You are required to prepare a concise and informative presentation, demonstrating the key aspects of your work.
The electronic version of the presentation must be uploaded to Moodle.

Reward: 15% of the module mark
Set – Week 1
Due – Week 11

Task: The key objective for the presentation is to demonstrate the basic elements of your report. You should provide the demonstration in such a way that allows the audience to understand the problem you dealt with, your project achievements, findings and future directions.

The time limit is 10 minutes per student.
.

Suggested content of the presentation.

The content of the presentation should be based on the results of the Technical Report (Research Paper). Therefore, ensure you completed the first comprehensive version your CW1 before week 11. This allows you to prepare presentation on time and improve CW1 following the suggestions obtained during discussion.

The basic design of the presentation should include no more than 10 slides as follows:

- Introduction (1 slide)
  Briefly introduce the context and purpose of your research.
- Problem Definition/Research Questions/Hypotheses (1 slide)
  Clearly outline the problem addressed, research questions, or hypotheses.
- Literature Review/Theory (1 slide)
  Summarize relevant literature or theoretical frameworks informing your research.
- Methods & Data (1 slide)
  Describe the methods employed and the nature of your data.
- Result Presentation/Findings (3-5 slides)
  Display and discuss key findings, supported by visuals where applicable.
- Conclusion (1 slide)
  Summarize your conclusions and briefly discuss the implications.

Presentation Submission.

Submit the presentation as a separate .pptx file. *Voiceover commentary* is optional.

Important Note: Ensure timely completion of the first version of CW1 before week 11 to facilitate adequate preparation and improvement of both the presentation and CW1 based on discussions and feedback.

Marking Rubric for CW2 – Presentation 15%

Grade (%)	**PRESENTATION – 15%**
A - A+ 70%-100%	Clear, well-paced excellent delivery. Appropriate introduction. Slides clearly laid out and legible. All aspects (problem statement, project achievement, future work) fully described and illustrated. Accommodating, but critical, attitude to alternative ideas. Clear and concise response. Deep understanding shown.
B - B+ 60% - 69%	Very good delivery and introduction. Layout and legibility mostly very good. Most aspects described and illustrated. Mostly clear and concise response. Good understanding and attitude to alternative ideas.
C – C+ 50% - 59%	Good delivery. Some introduction made. Layout and legibility good. Mostly good description of problem and project achievement. Mostly clear and concise response. Adequate understanding shown.
FPR 40% - 49%	Poor delivery. Layout and legibility adequate. Some description of problem and project achievement. Some understanding shown.
F 0%-29%	Failure to follow instructions or prepare in advance. Resulting talk clearly lacks overall the required aspects.
NS NON-SUBMISSION	N/A

Detailed Marking Scheme – CW2 Presentation

COMP7032 Coursework 2 Marking and Feedback		
Student Name:		
Student Number:		
Total Mark:	0	*Out of 15*
Grade:	0	*Out of 100%*

Criteria/Quality of Work	Marks available	Marks awarded
Written presentation	10 marks	
Organization of the slides (following the rubric for the report)	3	
All aspects (problem statement, project achievement, future work) fully described and illustrated.	7	
Oral presentation	5 marks	
Clear, well-paced excellent delivery.	2	
Clear and concise response. Deep understanding shown.	2	
Timing	1	
	Total:	0

Appendix 2: Assessment cover

OXFORD BROOKES UNIVERSITY

Assessment cover

STUDENTS, PLEASE COPY THIS PAGE AND USE AS THE COVER PAGE FOR YOUR SUBMISSION

Module No:	COMP7032	Module title:	**Data Science and Machine Learning**

Assessment title:	**Coursework 1: Technical Report (Research Paper)**

Due date and time**:**	**Monday, 29th April 2024, 13:00 (CW1)**

Estimated total time to be spent on assignment:	

LEARNING OUTCOMES

On successful completion of this module, students will be able to achieve the module following learning outcomes (LOs):
LO#1 Critically analyse and discuss key concepts in data science, including tools, approaches, and applications.
LO#2 Utilise effectively appropriate methods and tools for exploring, modelling, visualising and analysing data
LO#3 Report data science results to a wider audience by tailoring them at different levels of detail.
LO#4 Design, implement and critically evaluate a data science product that addresses a given real world data problem.
LO#5 Work in a context informed by relevant ethical, legal, social and security issues.

Engineering Council AHEP4 LOs assessed (from S2 2022-23)	
M1	Apply a comprehensive knowledge of mathematics, statistics, natural science and engineering principles to the solution of complex problems. Much of the knowledge will be at the forefront of the particular subject of study and informed by a critical awareness of new developments and the wider context of engineering
M2	Formulate and analyse complex problems to reach substantiated conclusions. This will involve evaluating available data using first principles of mathematics, statistics,

	natural science and engineering principles, and using engineering judgment to work with information that may be uncertain or incomplete, discussing the limitations of the techniques employed
M3	Select and apply appropriate computational and analytical techniques to model complex problems, discussing the limitations of the techniques employed
M4	Select and critically evaluate technical literature and other sources of information to solve complex problems
M5	Design solutions for complex problems that evidence some originality and meet a combination of societal, user, business and customer needs as appropriate. This will involve consideration of applicable health & safety, diversity, inclusion, cultural, societal, environmental and commercial matters, codes of practice and industry standards
M17	Communicate effectively on complex engineering matters with technical and non-technical audiences, evaluating the effectiveness of the methods used

STUDENT NAMES (ONLY IF GROUP ASSIGNMENT, OTHERWISE ANONYMOUS)

Student No:	Student Name:	Group Name and Number:
1.		
2.		
3.		
4.		
5.		

Statement of Compliance *(please tick to sign)*

☐ I declare that the work submitted is my own and that the work I submit is fully in accordance with the University regulations regarding assessments (*www.brookes.ac.uk/uniregulations/current*)

FORMATIVE FEEDBACK OPPORTUNITIES

SUMMATIVE FEEDBACK DELIVERABLES

Deliverable description and instructions	Weighting out of 100%

Individual reflection section: ● Propose further work that would offer improvements and enhancements. ● Evaluate personal learning and development in terms of technology/hardware/software/group work.	**10%**

Marking grid and peer marking form are attached at the end of this assignment.

ASSIGNMENT IN DETAIL