This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

Faculty of Engineering, Environment and Computing

7072CEM - MACHINE LEARNING

Assignment Brief (2021/2022)

Module Title **Machine Learning**	Individual or Group Project (2 people maximum), but individual paper submission	Cohort (Data Science MSc)	Module Code **7072CEM**
Coursework Title (e.g. CWK1) **Machine learning algorithms for solving real-world classification and clustering problems**			Hand out date: 18th May 2021
Lecturer **Dr. Xiaorui Jiang**			Due date and time: Date: 18th June 2021 Online: 18:00:00 (official time finalized by the course director, no extension)
Word Limit*: 5-6 A4 pages	Coursework type: Assignment		100% of Module Mark

Submission arrangement online via Aula: online via Aula or CU Moodle

File types and method of recording: Submit your report as a PDF or Word document using the 'Assignment' link in the 7072CEM Moodle page.

Mark and Feedback date (DD/MM/YY): 2 weeks after submission

Mark and Feedback method (e.g. in lecture, electronic via Aula): in Aula

Module Learning Outcomes Assessed:

1. Examine the fundamental concepts of machine learning, their implementation and application.
2. Prepare appropriate preparation of data sets and evaluate the performance of different learning algorithms on these data sets.
3. Appraise the different learning methodologies, their associated algorithms and their appropriateness to solve real-world problems.
4. Select and apply learning algorithms to various practical scenarios and evaluate their performance.
5. Critique trends in the current machine learning developments.

Task and Mark distribution:

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

1. See next pages

Notes:

1. You are expected to use the Coventry University APA style for referencing. For support and advice on this students can contact Centre for Academic Writing (CAW).
2. Please notify your registry course support team and module leader for disability support.
3. Any student requiring an extension or deferral should follow the university process as outlined here.
4. The University cannot take responsibility for any coursework lost or corrupted on disks, laptops or personal computer. Students should therefore regularly back-up any work and are advised to save it on the University system.
5. If there are technical or performance issues that prevent students submitting coursework through the online coursework submission system on the day of a coursework deadline, an appropriate extension to the coursework submission deadline will be agreed. This extension will normally be 24 hours or the next working day if the deadline falls on a Friday or over the weekend period. This will be communicated via your Module Leader.
6. You are encouraged to check the originality of your work by using the draft Turnitin links on Aula.
7. Collusion between students (where sections of your work are similar to the work submitted by other students in this or previous module cohorts) is taken extremely seriously and will be reported to the academic conduct panel. This applies to both courseworks and exam answers.
8. A marked difference between your writing style, knowledge and skill level demonstrated in class discussion, any test conditions and that demonstrated in a coursework assignment may result in you having to undertake a Viva Voce in order to prove the coursework assignment is entirely your own work.
9. If you make use of the services of a proof reader in your work you must keep your original version and make it available as a demonstration of your written efforts.
10. You must not submit work for assessment that you have already submitted (partially or in full), either for your current course or for another qualification of this university, with the exception of resits, where for the coursework, you may be asked to rework and improve a previous attempt. This requirement will be specifically detailed in your assignment brief or specific course or module information. Where earlier work by you is citable, i.e. it has already been published/submitted, you must reference it clearly. Identical pieces of work submitted concurrently may also be considered to be self-plagiarism.

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

7072CEM COURSEWORK:

MACHINE LEARNING ALGORITHMS FOR SOLVING REAL-WORLD CLASSIFICATION AND CLUSTERING PROBLEMS

Title: **Machine learning algorithms for solving real-world classification and clustering problems**

Individual Research Paper: 100% of the module mark

Context

During this module, you learned about different machine learning techniques, associated concepts and applications. We explored a number of classification algorithms, such as Generalized Logistic Regression, Linear Discriminant Analysis, Optimized K-nearest Neighbour, Bayesian and Statistical Methods, Support Vector Machines and Decision Trees. Also, we covered clustering algorithms, such as K-means, and feature selection and extraction methods, such as PCA. In this assignment, you will have to select an application related to a classification, clustering, or anomaly detection problem, and explore how best to apply machine learning algorithms to solve it. Basically, you are free to choose any of the following datasets (or another one agreed with your tutor in advance) and apply 3 classification or clustering techniques on it.

1. Bags of Words (Classification)
2. Daily and Sports Activities Dataset (Classification and Clustering)
3. Dresses Attribute Sales Dataset (Regression)
4. Or other (choose as you wish, but match techniques with the dataset)

You can combine and choose from the above mentioned algorithms or you can use or come up with a new classification or clustering algorithm!

The purpose of this coursework is to

- Examine the fundamental concepts of machine learning, their implementation and application.
- Perform appropriate preparation of a dataset and evaluate the performance of different learning algorithms on this dataset.
- Gain practical experience in selecting machine learning algorithms for solving a real-life classification or clustering problem.
- Demonstrate effectiveness in project teamwork and leadership.

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

You will be required to:

- Work in groups of 2, 1, or 3 (3 not preferred, and only when the project is very complex, and it should be approved by the lecturer);
- Actively participate in all activities;
- Write up individually on your own (using your own expression for describing the deliverable and emphasizing your contributions as well)
    - Contact the lecturer if you have doubt about this.
- Welcome to submit progress on your work regularly to get formative feedback and improve the final submission;
- Before your start, READ the four samples at the bottom of "Module Essentials >> Assessments" page. This is critically important for you to understand the elements and requirements of the CW.

Your final submission will include a "scientific paper" of up to 6 A4 pages (written individually, not as a group, if you worked in a group), based on the experience and results gained during the project work. You will have to acknowledge the contributions of all group members in your paper.
You are encouraged to target a certain conference or journal and submit the proposed paper to it. Submission guidelines can be found on the conference or journal web page you choose to submit to.
List of reputed conferences and journals:

1- IJCNN Conference
2- NeurIPS Conference
3- International Conference of Machine Learning
4- Machine Learning Journal
5- Neural Networks Journal
6- Others (please let us know)

The paper should broadly include the following sections:

- Abstract
- Introduction (where you introduce the problem along a short literature review of related work; if the literature review is longer, it is recommended to be a section on its own, which would be better)
- Problem and Data set(s) description (where you describe in detail the problem you want to solve and its significance)
- Methods (where you shortly describe the machine learning methods and/or other methods employed to solve the problem)

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

- Experimental setup (including data pre-processing, feature selection and extraction, classification/clustering parameters)
- Results
- Discussion and Conclusions
- References

These are **generic section titles**, which you may adapt appropriately to the application/problem that is being investigated. You may include sections describing modifications of algorithms or developments that are novel and specific to your work.

You may include figures, tables, pseudo-code, and appendices with the actual code that has been developed. You are free to use any programming language you are comfortable with (e.g., Matlab, Python, R, etc.)

More information of how to write a paper is available at the following link: "Crafting Papers on Machine Learning", by Pat Langley
(which can be found here if the previous link does not work http://www.machinelearning.ru/wiki/images/0/07/Langley00crafting.pdf).

You will need to follow the formatting guidelines of the IEEE Manuscript Template for Conference Proceedings (A4)

The group project general guidelines and milestones:
Please note, the following guidelines are good practice and should lead to better result, but you have the freedom to pick whatever is suitable for your style:

- Working in groups of 1, 2 (or 3, not recommended unless you have an extremely complex project which justifies collaboration with more people), you have to select a real-world classification/clustering problem and one or more appropriate dataset(s) as suggested above. You may also use the following links, which have numerous problems and datasets:
  - UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/;
  - ICML 2019 accepted papers: https://icml.cc/Conferences/2019/Schedule?type=Poster;
  - Kaggle competitions: http://www.kaggle.com/competitions;
  - Stanford machine learning projects: http://cs229.stanford.edu/projects2013.html, http://cs229.stanford.edu/projects2012.html, http://cs229.stanford.edu/projects2011.html, http://cs229.stanford.edu/projects2016.html .

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

- You will write a proposal (maximum of 1 A4 page), giving the title of the project, the names of all group members, the description of the problem and the plan of the work. You will need to submit this proposal to your tutor for formative feedback by 24 May 2021 (Monday).
- You will need to investigate and read related work in the next 2 days. You are welcomed to submit an individually written short literature review of your findings in order to get formative feedback. If you plan to do so, please do it before 28 May 2021 (Friday).
- In the following 2-3 weeks you have to select, implement and apply appropriate machine learning algorithms to the selected problem, performing data pre-processing, if needed, and record the results from the experiments.
- You are welcomed to report your progress to receive regular feedbacks from the module leader or tutor in the following 2 support weeks, via some scheduled meetings with the module leader.
- You have to write up your final paper, and submit it by the deadline specified on the first page.

Marking Criteria for the paper

Criterion	Mark
Technical quality	
1) Rigour and extent of the experiments.	10%
2) Correct application of the selected algorithms and suitability of the methods.	10%
3) Data preparation - technical quality.	10%
4) Extent of evidence of running the experiments provided in appendices.	10%

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

Evaluation 5) Evaluation and discussion of the results. Why the results are important? How would the results be useful to other researchers or practitioners? 6) Is this a "real" problem or a small "toy" problem? How does the paper advance the state of the art?	15% 5%
7) **Social, ethical, legal and professional considerations** related to the problem in question.	5%
Clarity of the writing: 8) Is there sufficient information for the reader to reproduce the results? Is the language used in the paper good? 9) References and general presentation; Are results clearly presented, with appropriate visualisations?	10% 5%
Originality: 11) Is there some original approach to the problem, original use of techniques? 12) Is there any (and how much) difference from previous contributions?	10% 10%

General marking guidelines (for all courseworks)

Mark allocation guidelines to students (to be edited by staff per assessment)

0-39	40-49	50-59	60-69	70+	80+
Work mainly incomplete and /or	Most elements completed; weaknesses	Most elements are strong, minor weaknesses	Strengths in all elements	Most work exceeds the standard expected	All work substantially exceeds the

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

weaknesses in most areas	outweigh strengths				standard expected

This document is for Coventry University students for their own use in completing their assessed work for this module and should not be passed to third parties or posted on any website. Any infringements of this rule should be reported to **facultyregistry.eec@coventry.ac.uk**.

Marking Rubric (To be edited by staff per each assessment)

GRADE	ANSWER RELEVANCE		ARGUMENT & COHERENCE		EVIDENCE		SUMMARY	
First **≥70**	Innovative response, answers the question fully, addressing the learning objectives of the assessment task. Evidence of critical analysis, synthesis and evaluation.		A clear, consistent in-depth critical and evaluative argument, displaying the ability to develop original ideas from a range of sources. Engagement with theoretical and conceptual analysis.		Wide range of appropriately supporting evidence provided, going beyond the recommended texts. Correctly referenced.		An outstanding, well-structured and appropriately referenced answer, demonstrating a high degree of understanding and critical analytic skills.	
Upper Second **60-69**	A very good attempt to address the objectives of the assessment task with an emphasis on those elements requiring critical review.		A generally clear line of critical and evaluative argument is presented. Relationships between statements and sections are easy to follow, and there is a sound, coherent structure.		A very good range of relevant sources is used in a largely consistent way as supporting evidence. There is use of some sources beyond recommended texts. Correctly referenced in the main.		The answer demonstrates a very good understanding of theories, concepts and issues, with evidence of reading beyond the recommended minimum. Well organised and clearly written.	
Lower Second **50-59**	Competently addresses objectives, but may contain errors or omissions and critical discussion of issues may be superficial or limited in places.		Some critical discussion, but the argument is not always convincing, and the work is descriptive in places, with over-reliance on the work of others.		A range of relevant sources is used, but the critical evaluation aspect is not fully presented. There is limited use of sources beyond the standard recommended materials. Referencing is not always correctly presented.		The answer demonstrates a good understanding of some relevant theories, concepts and issues, but there are some errors and irrelevant material included. The structure lacks clarity.	
Third **40-49**	Addresses most objectives of the assessment task, with some notable omissions. The structure is unclear in parts, and there is limited analysis.		The work is descriptive with minimal critical discussion and limited theoretical engagement.		A limited range of relevant sources used without appropriate presentation as supporting or conflicting evidence coupled with very limited critical analysis. Referencing has some errors.		Some understanding is demonstrated but is incomplete, and there is evidence of limited research on the topic. Poor structure and presentation, with few and/or poorly presented references.	
Fail **<40**	Some deviation from the objectives of the assessment task. May not consistently address the assignment brief. At the lower end fails to answer the question set or address the learning outcomes. There is minimal evidence of analysis or evaluation.		Descriptive with no evidence of theoretical engagement, critical discussion or theoretical engagement. At the lower end displays a minimal level of understanding.		Very limited use and application of relevant sources as supporting evidence. At the lower end demonstrates a lack of real understanding. Poor presentation of references.		Whilst some relevant material is present, the level of understanding is poor with limited evidence of wider reading. Poor structure and poor presentation, including referencing. At the lower end there is evidence of a lack of comprehension, resulting in an assignment that is well below the required standard.	
Late submission	0		0		0		0	