Module title:	**ACS341**	**Assignment Name:**	**Coursework 1**
Person responsible and contact details:	**Dr John Oyekan**	**Assignment weighting:**	**35%**
Assignment released:	**15th of March 2022**	**Assignment Due:**	**28th of April 2022**

Assignment due date: Please hand in by **11pm** on the **28th of April 2022**; this course work makes up 35% of your total module mark. Submit your report as a pdf file on Blackboard. Also include your orange file (.ows), Python and MATLAB codes as part of your submission.

Extenuating Circumstances: If you have any extenuating circumstances (medical or special circumstances) that might have affected your performance on the assignment, please complete an extenuating circumstances form. Late submission rules apply with a reduction in 5% for every additional late day and a score of zero after 5 days.

Unfair means: All work must be **completed as individuals**. References should be used to support your domain analysis research. Suspected unfair means will be investigated and will lead to penalties. For more information on the university unfair means' guidance, please check: http://www.shef.ac.uk/ssid/exams/plagiarism.

The challenge: A company wishes to understand the reasons why their machines fail and under what conditions. This will enable them to conduct predictive maintenance. You have been provided a data set of machine performance which includes features such as the machine's wear, torque, process temperature and failure. The dataset is made up of a total of 8 features. The description for the features is provided in the Appendix below. Please write a report that addresses the various tasks below.

Tools to use: Majority of the MATLAB code you need to complete the assignment are available from various lab sessions. If you are comfortable using Python, you are free to use it. You are also free to use Orange for various aspects of the coursework as required.

Tasks and Mark Scheme: The aim of this coursework is to design, implement and evaluate an effective machine learning pipeline for predicting machine failure. The specific tasks and corresponding mark scheme are given in the table below. It is up to you how you approach this problem, design a solution and write-up your results. For each task, the mark within the grade boundary will be based on your description in your report, results and code.

Task/Assessment Description	Mark Range	Level of achievement
Task 1: Conduct and write a domain analysis that discusses the important features to take into consideration when working in the predictive maintenance for machine domain**.** Discuss how what you have found from your domain analysis will support and be carried over to other parts of your pipeline. To give you a head start, take a look at this article and its references:**https://www.sciencedirect.com/science/article/abs/pii/S0965997810001663**	0-10%	1
Task 2: Achieve level 1 as well as conduct data cleaning, pre-processing and feature engineering. Discuss how you used your understanding of the domain from level 1 to	10-20%	2

support this task. This could involve deciding which features to drop and which relevant features to keep.		
Task 3. Achieve all the previous levels as well as build a regression model (decide which hypothesis function $\boldsymbol{h_\theta}$ that is best to use e.g polynomial or linear etc) to predict the values of Torque from Rotational speed. Discuss the suitability of your chosen hypothesis function. Show plots of the predictive capability of your model in relation to the original data points.	20-30%	3
Task 4. Achieve all the previous levels as well as use learning curves to discuss how effective your regression model machine learning pipeline is at preventing overfitting and underfitting.	30-45%	4
Task 5. Achieve all the previous levels plus discuss which cross validation technique you applied in Task 4 above and why.	45-50%	5
Task 6. Using the features you consider most important to this challenge, apply Logistic Regression to build a model that predicts when a machine would break. Use the classification metrics of confusion matrix, accuracy, precision and recall to explain your results.	50-65%	6
Task 7. Use Decision Trees to provide transparent explanation of your machine learning pipeline. Also compare the results of the Decision Tree with Logistic Regression using the classification metrics of confusion matrix, accuracy, precision and recall to explain your results.	65-80%	7
Task 8. Achieve all the previous levels and the below: • Using the dataset given, compare the results of the machine learning algorithms above with the results of • Two other algorithms that we have not covered in class. • Discuss the mathematical peculiarities of the algorithms you have chosen (strengths and weaknesses) and how they impact the results you obtain. • Apply the appropriate metrics to compare the algorithms you have chosen with the ones we have used in class.	80-100%	8

Technical Report and code

Write your results in no more than a 15 page technical report. Make sure your report has a table of content, sections, discussion and conclusion.

Support your report with an orange pipeline design and MATLAB and/or Python code(s). Make sure you provide comments in your MATLAB (or/and Python) code as well as instructions on how to run it. Hand in your report (.pdf), software (Orange and MATLAB (or/and Python)) via Blackboard by 11pm on the **28th of April 2022**.

This course work makes up 35% of your total module mark.

Appendix

Product ID: This is the machine ID.
Type: This feature relates to the type of machine.
Air Temperature [K]: This is the air temperature around the machine measured in Kelvin (K).
Process temperature [K]: This is the temperature, measured in Kelvin (K), that is generated by the machine during operation.
Rotational speed [rpm]: Measured in Revolutions per minute, this is the speed the machine's spindle runs at.
Torque [Nm]: Is the measure of force used to generate angular momentum during rotations by the machine.
Tool wear [min]: This measures the gradual failure of parts of the machine due to regular operation
Machine failure: This indicates if the machine has failed or not.