Northumbria University NEWCASTLE
Newcastle • London • Amsterdam

COURSEWORK ASSESSMENT SPECIFICATION

Module Title:	Machine Learning on Cloud
Module Number:	LD7187
Module Tutor Name(s):	Dr Bilal Hassan/Dr Husnain Sherazi/Dr Karim Mohammed Rezaul/Alan/ Alan Vallavaraj/Sharib Anis
Academic Year:	2022-23
% Weighting (to overall module):	100%
Coursework Title:	Report

Dates and Mechanisms for Assessment Submission and Feedback

Date of Handout to Students: Week commencing TBC via Blackboard
Mechanism for Handout to Students: Via Blackboard; briefing via online blackboard collaborate and face 2 face session.
Date and Time of Submission by Student: Submitted on **18th May 2023** (no later than 23:59)
Mechanism for Submission of Work by Student: The report must be submitted Turnitin on Module Blackboard site.
Date by which Work, Feedback and Marks will be returned to Students: Within 20 working days after the submission date.
Mechanism for return of assignment work, feedback and marks to students: Formal feedback will be made available via Blackboard following completion of all reviews and internal moderation of results.

Learning Outcomes tested in this assessment

This assignment will assess the following learning outcomes:

1. Demonstrate critical knowledge and understanding of the methods, theories and techniques of machine learning.
2. Critical appraisal of cloud computing architectures and tools used to deploy machine learning techniques.
3. Develop practical skills to critically analyse, design and implement machine learning solution for a given scenario.
4. Be able to demonstrate critical awareness of the social, ethical and sustainability issues in the context of AI and cloud.

General Information

This is an INDIVIDUAL piece of work contributing towards the module assessment. Deliverables should include access guide to model deployment and a report document, which includes (critical discussion and analysis, research, and snapshot evidences of tasks carried out and justification of techniques used). Submission will be in the form of a MS word report (4000 words).

This assignment constitutes 100% towards the final mark for this module. Any queries relating to this assignment should be directed to;

bilal Hassan (b.hassan@northumbria.ac.uk)

Assessment brief: FIFA World Cup 2022

(The original case study and dataset can be found from Kaggle). The FIFA World Cup is the most prestigious football tournament in the world. The championship has been awarded every four years since the start of the tournament in 1930. In the tournament, 32 teams, including the host nation, compete for the title at different stadiums in the host country. This year, Qatar will host the 2022 tournament, for which the first match will be played in November 2022. Before the start of the tournament, to predict the winner of the tournament is an exciting discussion always and if this is done through machine learning, it is most likely that it is going to be true. To accomplish this task, the dataset provided includes a complete overview of all international soccer matches played since 90s. On top of that, the strength of each team is provided by incorporating actual FIFA rankings including several other attributes. The original data can be downloaded from the following URL (Please signin/signup using person or organizational email to download).

The original dataset contains several extra attributes and missing values. As ML developer, this is your task to analyze and preprocess the data using appropriate machine learning techniques. Later, to propose and develop a machine learning model (an application deployed on cloud) **to predict the winner of the tournament. You prediction may be a**

probability of winning for the five teams. In order to achieve this, you are required to accomplish following task on the AWS.

A typical workflow of machine learning on the cloud involves, data preprocessing, selection and training of model, model evaluation and deployment for prediction. During coursework, instructor will guide you how to perform different machine learning tasks on the AWS.

Note: This case study is based upon a challenging dataset provided on *Kaggle.*

Assessment Criteria/Marking Scheme:

The work will be marked out of 100 in line with the University's marking grades and according to the following assessment criteria:

Description	Marks
Task 1: Cloud feasibility study – 10% **(suggested word limit for this section is 500 words)**	
In your report, explore and analyse different cloud platform for commercial machine learning solutions deployment in terms of feasibility to the given scenario. The discussion should be based on state-of-the-art literature.	10
Task 2: Data analysis and opportunity identification – 10% **(suggested word limit for this section is 500 words)**	
In your report, explore and present facts about given data and a comprehensive discussion on similar problems and techniques used to solve those.	10
Task 3: Data pre-processing – 20% **(suggested word limit for this section is 1000 words)**	
Generate data for the model training after executing appropriate pre-processing steps like wrangling, ingestion and transformation etc. using AWS Data Wrangler UI or Python. Only perform whatever steps are essential and provide reasoning for that in the report.	20

Task 4: Model selection and training - 30% **(suggested word limit for this section is 1000 words)**	
Select an appropriate model and train it, Using the SageMaker Python SDK or Use Amazon SageMaker Built-in Algorithms. The SDK includes lot of different machine learning models, choose whatever is more appropriate to the problem and outcome of analysis performed in Task 1. Discuss the whole process in detail.	30
Task 5: Model evaluation and visualization - 15% **(suggested word limit for this section is 500 words)**	
Evaluate the model, once you are done with training, evaluate it for the overfitting etc. You can see what measure or visualization would be appropriate to showcase the performance of chosen model using following link. Discuss results.	15
Task 6: Model deployment - 15% **(suggested word limit for this section is 500 words)**	
Deploy your model using SageMaker Hosting Services (desired) or discuss step by step procedure for deployment without using SageMaker. Discuss in detail what best professional practices can be employed to comply with ethical and privacy concerns.	15
Total	**100**

Grading Guidance

Distinction (70 and above):

A distinction submission will include an end-to-end implementation e.g. fully automated from pre-processing to deployment. It will also provide a comprehensive report on each step of implementation along with reasoning of every particular technique used. It will also showcase the final form of application by providing access mechanism to it and a quick access guide.

Commendation (60-69):

A commendation submission will include source code snippets performing various tasks e.g. pre-processing, training, evaluation or deployment etc. in attempt to accomplish tasks. It will also provide a report having discussion on these tasks provided with little or no reasoning. It will provide access to application with a quick guide.

Pass (50-59):

A pass submission will include random code snippets developed for few task and their output. The report will contain discussion on these tasks accomplished. It will be missing a critical analysis of different techniques used with little or no access mechanism.

Fail (less than 50):

A fail submission will provide little or no implementation of tasks assigned and their incomplete output. The report will lack clarity and reasoning for each task performed. It will also miss the access mechanism and quick access guide to application.

Academic Integrity Statement: You must adhere to the university regulations on academic conduct. Formal inquiry proceedings will be instigated if there is any suspicion of plagiarism or any other form of misconduct in your work. Refer to the University's Assessment Regulations for Northumbria Awards if you are unclear as to the meaning of these terms. The latest copy is available on the University website. (Accessed on 21.08.2022)

Formative Feedback

There will be an opportunity for formative feedback during the semester. You are advised to start working on this assignment as early as possible so that you can seek clarification from the module tutor regarding any questions you might have during the semester. Note that tutors will not predict your grade, and you should not take the lack of comment on any aspect of your work as indicating that it is correct. You should make every effort to take advantage of formative feedback as tutors will not comment on draft work at other times. Remember that you will get more useful feedback from us by asking specific questions than just presenting us with your documentation and asking, 'Is this right?'

Penalties for Exceeding Word Limits:

The following penalties will be applied after any reductions in mark due to late submission have been made, Penalties will be applied as defined in the University Policy on Word Limits Policy. (Accessed 21.08.2022)

The actual word count is to be declared on the front of the assessment submission.

Under the word limit	No Penalty: In not making use of the full word count, students may have self-penalised their work. If students have been able to achieve the requirements of the assessment using fewer words than allocated, they will not be penalised.
Up to 10% over word limit	No Penalty: Situation flagged by tutor in feedback but over-run is tolerated and no deduction is made from the final mark.
More than 10% over the word limit	The marker will stop reading when they judge that the word count exceeds the recommended word count by more than 10% i.e. for a 3000 word essay, the marker will read only the first 3300 words and will indicate on the text where they stop reading. The content following this point will not be read and a mark will be awarded only for the content up to this point.

Late Submission Policy:

For coursework submitted up to 1 working day (24 hours) after the published hand-in deadline without approval, 10% of the total marks available for the assessment (i.e.100%) shall be deducted from the assessment mark. Penalties will be applied as defined in the University Policy on the Late submission work. (Accessed 21.08.2022)

For clarity: a late piece of work that would have scored 65%, 55% or 45% had it been handed in on time will be awarded 55%, 45% or 35% respectively as 10% of the total available marks will have been deducted.

Failure to submit: The University requires all students to submit assessed coursework by the deadline stated in the assessment brief. Where coursework is submitted without approval after the published hand-in deadline, penalties will be applied as defined in the University Policy on the Late submission of work. (Accessed 21.08.2022)