COM00166M

Department of Computer Science

Applied Artificial Intelligence

SUMMATIVE ASSESSMENT BRIEF

|  |  |
| --- | --- |
| **Author** | Joseph Akinyemi/Saul Cross |
| **Assessment type** | Summative assignment |
| **Weighting** | 100% |
| **Release** | Week 3 |
| **Deadline** | Monday following Week 8, 13:00 (UK time) \* |

\* If this date falls on a UK public holiday or a University of York closure day, the submission date will change. Please check the submission point in the ‘Assignments’ area of the module in Canvas for the exact submission deadline.

# Module Learning Outcomes

The module learning outcomes for this module are as follows:

1. Select and apply appropriate AI algorithms and methodologies, with consideration for optimisation and scale to meet business objectives and performance targets.
2. Critically evaluate AI methodologies through experimental design, exploratory modelling, and hypothesis testing.
3. Critically analyse techniques for the extraction of data from systems, ensuring standards of data quality and consistency for processing by AI systems
4. Identify and discuss appropriate application areas and problems for current AI techniques, such as: neural network, deep learning, genetic algorithms and local search approaches.

This assessment addresses **all** the module learning outcomes listed above.

# Assessment Background/Scenario

A non-governmental organisation (NGO) is applying for a grant to support them in combating child malnutrition across several countries worldwide. To increase their chances of securing funding, they have decided to analyse reputable databases to find relevant information that can support their campaign for improved child nutrition across the globe. They have found some datasets by the World Health Organisation (WHO) which seem to suit their needs. These datasets contain statistics on infant breastfeeding rates as well as child mortality rates (0 - 5 years) across different countries and year periods. Now, they have to present a strong argument for the correlation between these datasets and how this relates to their child nutrition campaign.

You have been consulted as an AI engineer to help with the analysis of these datasets and to reveal insights from them as follows:

* Your initial observations from your exploration of these datasets are as follows:
* The data contains some missing values
* The data mostly contains numeric values which may not be properly formatted
* The data contains relatively few features
* Both datasets seem to cover different ranges of years.

In your meeting with your client, you have agreed to use relevant Machine Learning (ML) techniques (supervised, unsupervised, etc.) and AI search or optimisation techniques. You are expected to present a report to your client by constructing a robust model which must follow the guidelines presented below:

1. Preprocess the datasets to create a single dataset which contains the needed information to predict mortality rates for different years for each country.
2. Use AI search or optimisation techniques (whichever is appropriate) to align the year periods for each country across all the datasets.
3. Based on the dataset you have created, build a supervised or unsupervised ML model to predict the mortality rates for each country for the different years possible.
4. Justify your design decisions for tasks 1, 2 & 3.
5. Critically evaluate the learning model you have built
6. Evaluate the robustness of your model by applying appropriate validation techniques (and identifying a suitable subset of data for validation).

While setting the parameters of the search or optimisation method, pay special attention to selecting appropriate metrics (evaluation criteria). The chosen metrics will play a critical role in the relative success or failure of the potential solution(s) and in setting the direction of the search or optimisation.

# Assessment Tasks

You are required to design and build solutions, and to write a report based on the discussed scenario and the data provided. You should clearly draw on current literature and use examples from your work throughout this module as supporting evidence for the approach. Your report should provide an initial executive summary and consist of five clear sections - one for each task outlined below. Further formatting details are also given below.

## Overall Academic Quality (10% weighting)

Covered by all tasks.

## Executive Summary (10% weighting)

(suggested - 300 words)

Overview/summary of the report which should at least contain:

* 1. What was achieved/undertaken.
  2. What processes were applied.
  3. What tools were used and why.
  4. What results were obtained and what they demonstrate.
  5. What should be reconsidered in future.

## Task 1 (10% weighting) - Introduction

(suggested - 500 words)

In the context of the scenario provided, your introduction should at least contain:

* 1. A brief description of the business problem and its significance to the relevant sector.
  2. Background information on the field of AI.
  3. A description of the link between the business problem and the field of AI.
  4. A brief description of the proposed solution.

## Task 2 (20% weighting) - Literature Review

(suggested - 700 words)

Given the scenario above, research and identify the main relevant areas of investigation that the research community is currently tackling. Consider the following questions:

* 1. What are the current ‘problem’ areas and how is AI helping to solve these problems?
  2. What techniques have been developed to effectively address those business problems?
  3. What tools have been used and what are the selection criteria for them?
  4. How are these techniques being evaluated in the context of the ‘problem’?
  5. Critically evaluate various approaches/solutions presented in the literature.
  6. Present a discussion around these questions and consider how current research could potentially change or improve your solution to the given scenario.

## Task 3 (20% weighting) - Research Design

(suggested - 500 words)

Given the scenario above, design and discuss the potential modelling solution(s). You are required to design the solution(s) of the presented scenario. Moreover, you need to strongly justify the techniques selected in the context of the ‘problem in hand’. You must select one supervised/unsupervised learning algorithm and one search or optimisation algorithm to complete this task.

Your report should clearly cover the following:

* 1. Any assumptions you are making about the given scenario.
  2. Any pre-processing you would undertake to make the data fit for purpose.
  3. Which search or optimisation techniques you would apply to align the year periods and why (the techniques covered in this module are: hill climbing, simulated annealing, tabu search and genetic algorithms).
  4. Which supervised/unsupervised learning techniques you would apply and why (the techniques covered in this module are: artificial neural networks, decision trees, k-means clustering, linear regression, naïve Bayes and support vector machines).
  5. An evaluation of the techniques applied in terms of the accuracy of their results (or any other suitable evaluation measure).
  6. Algorithmic parameters should be adequately stated and discussed.
  7. You will need to supply additional evidence in the form of a zip archive containing your cleaned/pre-processed datasets, the source code files in an appropriate format for both the cleaning/pre-processing and the implementation of the chosen solution/s.
  8. Data must be presented as either **.csv** or **.ipynb** with solutions presented as source code in **Java** or **Python** or as complete **.arff** files. Failure to provide the supporting files in the formats stated or in an unreadable form will result in the invalidation of this task.

## Task 4 (20% weighting) - Experimental Results and Analysis

(suggested - 700 words)

After carrying out the modelling of the data provided, both with and without feature selection, this section must at least cover the following points:

* 1. Present your findings in a clear and concise manner.
  2. Discuss your results in the context of the selected optimisation/search algorithm and supervised/unsupervised learning technique.
  3. Discuss how these results can help the NGO to present a sound justification of their grant application.
  4. Your arguments should also be supported by the relevant literature.

## Task 5 (10% weighting) - Conclusion

(suggested - 300 words)

Your conclusion must at least cover the following points:

* 1. A summary of the main points.
  2. A discussion of the significance of your results.
  3. Present the limitations (if any) of your approach in a clear and concise manner
  4. Any recommendation(s) resulting from your analysis.

# Deliverables

Your assignment should be laid out following the formatting guidelines that are specified in the ‘Submission Formatting’ page in Canvas. This includes restrictions on the length of the appendices, expectations on how your work should be presented and any penalties when these expectations are not met.

Your report should not exceed 3,000 words in total and consist of 6 clear sections – one for each task (plus an executive summary). Your response to one section/task will not contribute to grades in another. Further formatting details and essential points are given below.

You should submit the following to the Canvas submission point:

* A zipped file containing your solution source files (not executables) and pre-processed data in .zip format ONLY
* A separate file containing your report, in **.docx** format **ONLY.**

If you are submitting multiple files, you **must** upload all files simultaneously to ensure that they are marked as a single submission. If you want to resubmit one component of your work, you need to re-upload all other files at the same time: every submission must include **all** of the deliverables listed in the assessment brief.

Appendices are limited to 5 pages as per the ‘Submission Formatting’ guidelines.

## Referencing

You are required to use the [IEEE referencing style](https://subjectguides.york.ac.uk/referencing-style-guides/ieee) for citing books, articles, and all other sources (such as websites) used in your assignment.

Good referencing is essential in order to meet the standards of academic integrity set by the University. All of your sources must be acknowledged, regardless of whether you have included direct quotes or not. Visit your **Academic Integrity Tutorial** module in Canvas for additional guidance on effective referencing.

# Marking Criteria

|  |  |  |  |
| --- | --- | --- | --- |
| Learning Outcome | Section/Task | Criteria | Available marks |
|  | **Overall Academic Quality** | | |
| 1, 2, 3 & 4 | All Tasks | Clear and coherent across all tasks with appropriate, relevant and effective referencing and citation. | **10** |
|  | **Executive summary** | | |
| 1, 2, 3 & 4 | N/A | Full and clear overview of report. | **10** |
|  | Introduction | | |
| 1 | Task 1 | Clear description of the problem and its significance to the sector (the NGO, funding body and relevant stakeholders). Discussion of relevant background/underpinning theory in AI. Alignment of the problem to the field of AI. Description of the solution(s) presented. | **10** |
|  | **Literature Review** | | |
| 4 | Task 2 | Examines current, relevant problem areas. Considers and evaluates proposed solutions. Includes a critical evaluation of any, and all relevant approaches/solutions under review. | **20** |
|  | **Research Design** | | |
| 1, 2, 3 & 4 | Task 3 | To cover pre-processing techniques, features selection, supervised/unsupervised ML model and optimisation techniques, evaluation of selected techniques and discussion of algorithmic parameters, accompanied by a clear and justified rationale in each case.  **Non-submission of relevant supporting evidence (zip file) invalidates this task** | **20** |
|  | **Experimental Results and Analysis** | | |
| 2 & 4 | Task 4 | A clear and concise presentation of the results to include mapping of results to the problem; reference to relevant literature and discussion of the solutions in context. | **20** |
|  | **Conclusion** | | |
| 1, 2, 3 & 4 | Task 5 | A clear and concise presentation of findings with consideration of limitations and further development. | **10** |
|  | **TOTAL**: | | **100** |

**NOTE:** Failure to submit the report will result in an overall grade of **ZERO**. Failure to submit the required supporting zip file containing your solution source files will result in a grade of **ZERO** for task 3. Work NOT submitted in the requested formats (**.docx** **ONLY** for the report and source files in an appropriate and readable format: .arff, .csv, .ipynb .py, or .java) may **NOT** be considered.

# Marking Criteria: Grade breakdown

## Overall Academic Quality 10%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | The documentation is poorly structured and flows badly, with little or no evidence of an effective use of referencing and citation. |
| 40-49% | Marginal fail | The documentation is not well structured and does not flow well, with very limited/ineffective use of referencing and citation. |
| 50%-59% | Pass | The documentation has a clear structure and is both legible and coherent but could flow better and make much better use of referencing and citation. |
| 60%-69% | Merit | The documentation is well structured with a clear and coherent flow and uses referencing and citation effectively to support most of the discussion. |
| 70%-100% | Distinction | The documentation is very well structured with a clear and coherent flow and uses referencing and citation very effectively throughout the discussion. |

## Executive Summary 10%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | The executive summary is significantly incomplete. |
| 40-49% | Marginal fail | The executive summary lacks clarity, providing only limited coverage of the report. |
| 50%-59% | Pass | The executive summary is reasonably clear and provides basic coverage for most of the key areas. |
| 60%-69% | Merit | The executive summary is clear, giving a good coverage of the key areas of the report. |
| 70%-100% | Distinction | The executive summary is both clear and concise with full coverage of the report. |

## Task 1 - Introduction 10%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | The introduction is very limited with little or no theoretical underpinning, an undefined or poorly defined project and no clear alignment to the field of artificial intelligence. |
| 40-49% | Marginal fail | The introduction is limited with little theoretical underpinning, the project is not clearly defined, and there is limited relevance to the field of artificial intelligence. |
| 50%-59% | Pass | The introduction is basic with limited theoretical underpinning, the project definition and reference to relevant concepts in the field of artificial intelligence. |
| 60%-69% | Merit | The introduction gives a clear and concise coverage of the required points with most relevant theoretical underpinnings covered, a well-defined project, and a well-supported discussion of the relevant concepts in the field of artificial intelligence. |
| 70%-100% | Distinction | The introduction provides a very clear and concise coverage of the required points with most relevant theoretical underpinnings covered, a very well-defined project and a very well-supported discussion of relevant concepts in the field of artificial intelligence. |

## Task 2 - Literature Review 20%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | Coverage is very limited with little or no relevance in the sources selected, is not a good match for the subject matter under review and with very limited referencing and citation. |
| 40-49% | Marginal fail | There is limited coverage and the sources selected could be more relevant as they are not a good match to the subject matter under review; referencing and citation are present but limited. |
| 50%-59% | Pass | There is basic coverage from mostly relevant sources that do match the subject matter under review to an extent, but referencing and citation could be much more thoroughly and effectively used. |
| 60%-69% | Merit | There is some detailed and effective coverage based on relevant sources and this is a good match to the subject matter under discussion that makes effective use of referencing and citation. |
| 70%-100% | Distinction | A very detailed yet concise literature review that identifies a good range of relevant sources and aligns well with the subject under discussion, with very effective use of referencing and citation seen throughout. |

## Task 3 - Research Design 20%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | The required supporting evidence is either missing or the research design element is very limited, with little or no valid coverage of the dataset/subset selection, pre-processing, selection criteria for the methods/algorithms applied, feature selection and parameterisation. |
| 40-49% | Marginal fail | The research design element is limited, with limited coverage of the dataset/subset selection, pre-processing, selection criteria for the methods/algorithms applied, feature selection and parameterisation, and lacks rationale or justification for the decisions made. |
| 50%-59% | Pass | The research design element is basic, with some coverage of the dataset/subset selection, pre-processing, selection criteria for the methods/algorithms applied, feature selection and parameterisation; with limited rationale or justification for the decisions made. |
| 60%-69% | Merit | The research design element is effective, with reasonable coverage of the dataset/subset selection, pre-processing, selection criteria for the methods/algorithms applied, feature selection and parameterisation, and some rationale or justification for many of the decisions made. |
| 70%-100% | Distinction | The research design element is both clear and effective, with a good level of coverage of the dataset/subset selection, pre-processing, selection criteria for the methods/algorithms applied, feature selection and parameterisation, and clear rationale or justification for most of the decisions made. |

## Task 4 - Experimental Results and Analysis 20%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | Severely limited evidence of any analysis being carried out with very limited or no presentation of results or findings, very limited or no evaluation of the results with respect to the problem. |
| 40-49% | Marginal fail | Limited evidence of any analysis being carried out with limited presentation of results or findings, limited evaluation of the results with respect to the problem. |
| 50%-59% | Pass | Basic evidence of analysis and evaluation of the results with respect to the problem and there is only very limited rationale or justification given for the approach(es) taken. |
| 60%-69% | Merit | Some good evidence of analysis and evaluation of the results with respect to the problem and reasonable rationale or justification given for the approach(es) taken and with some support from wider reading. |
| 70%-100% | Distinction | Clear and detailed (in-depth) evidence of analysis and evaluation of the results with respect to the problem and a solid rationale or justification given for the approach(es) taken which is well supported by wider reading. |

## Task 5 - Conclusion 10%

|  |  |  |
| --- | --- | --- |
| 0-39% | Fail | The conclusion does not provide any meaningful insight into the findings or give any real consideration to limitations or potential future development. |
| 40-49% | Marginal fail | The conclusion does not give full coverage of the findings and has very limited discussion of limitations and/or potential future development. |
| 50%-59% | Pass | The conclusion is basic, summarising the key points in the findings only and with limited consideration of limitations and potential future development. |
| 60%-69% | Merit | The conclusion provides a rich coverage of most of the key points in the findings and considers limitations and potential future development effectively with some support from wider reading. |
| 70%-100% | Distinction | The conclusion provides full coverage of the key points in the findings, with a clear discussion of constraints/limitations and potential for improvement/further development that is well supported by wider reading. |

# Assessment Submission

You will submit your assessment in the ‘Assignments’ area of the module in Canvas. Please check your Canvas module for the specific submission date for this assignment.

This assessment requires you to anonymously upload your submission to Canvas. If you are submitting multiple files, you must upload all files simultaneously to ensure that they are marked as a single submission. If you want to resubmit one component of your work, you need to re-upload all other files at the same time: every submission must include **all** files required by the assessment brief.

We recommend that you allow at least 30 minutes before the deadline to upload your submission, as failure to upload your assessment file within the allotted time is not admissible as an exceptional circumstance.

The webpage [How do I submit an online assignment?](https://community.canvaslms.com/t5/Student-Guide/How-do-I-submit-an-online-assignment/ta-p/503) provides further technical information on how to upload an assessment. The advice given here comes directly from Canvas. We do not recommend uploading assignments by mobile. We recommend you view the submission immediately after uploading your work to ensure the correct file has been submitted and no technical errors have occurred.

If you face any technical difficulties whilst trying to submit this assessment, then contact Canvas support on [support@instructure.com](mailto:support@instructure.com) or +44 80 0060 8442 (available 24 hours) in advance of the deadline. You should also email [york-online-assessment@york.ac.uk](mailto:york-online-assessment@york.ac.uk) as a matter of urgency to report the issue and receive further instruction.

# Assessment Policies

This assessment is subject to the policies stated on the ‘Summative Assessment Policies’ page in Canvas. These policies include (but are not limited to):

* Academic Integrity and submission of student work to Turnitin
* Advice on anonymising your assessment
* Penalties for late submission
* Marking policy for multiple submissions
* The Fit to Sit / Submit policy
* Passing mark and module reassessment

Please ensure that you have read and understood these policies before starting the assessment.