

COM00166M

Department of Computer Science

Applied Artificial Intelligence

SUMMATIVE ASSESSMENT BRIEF

Author	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.

I. Module Learning Outcomes

The module learning outcomes (MLO's) for this module are as follows:

- MLO 1.** Select and apply appropriate AI algorithms and methodologies, with consideration for optimisation and scale to meet business objectives and performance targets.
- MLO 2.** Critically evaluate AI-methodologies through experimental design, exploratory modelling, and hypothesis testing.
- MLO 3.** Critically analyse techniques for the extraction of data from systems, ensuring standards of data quality and consistency for processing by AI-systems.
- MLO 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.

II. Assessment Background/Scenario

In recent years, there has been a steady decline in the population and variety of pollinating insects throughout the United Kingdom; most notably, bee populations have fallen due to a number of factors, including the use of specific pesticides, the introduction of invasive species and an increase in colony collapse disorder.

The decline in pollinating insect populations is a significant cause for concern with specific impacts on farming and agriculture in addition to other, longer-term environmental concerns. To address this, the UK government, through DEFRA (Department for the Environment, Food and Rural Affairs), have been engaging with landowners, farmers and a number of environmental organisations to find potential solutions to slow the decline of pollinating insect populations and encourage increased biodiversity in order to slow down or even reverse these trends.

Bee Positive is a non-government organisation working with both Government departments and farmers to inform and evaluate the development of schemes specifically focused on supporting bee populations and schemes to encourage biodiversity, which specifically encourage growth in bee populations.

Datasets have been collected covering the UK Biodiversity indicators in the following areas of interest:

- Agriculture & Environment Schemes;
- Habitat Connectivity;
- Insects of the Wider Countryside;
- Plants of the Wider Countryside;
- Pollinating Insects.

It is believed that AI can be used to develop models and applications for planning, forecasting and evaluation, and you have been tasked with the preliminary investigation, which should include developing an example proof of concept application of how this data can be used and which methods/approaches (utilising AI) should be applied. Your work should include the following:

- Your initial observations from your exploration of the provided datasets;
- Cleaning and pre-processing actions to address missing or incompatible values across the datasets, data alignment and the potential use of simulated data;
- Clear classification of the attributes within the datasets, including consideration of datatypes and formatting;

- Selection of appropriate features for analysis and as a basis to construct models;
- Identification and testing of suitable algorithms for investigating/evaluating the data.

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 proof of concept model, which must follow the guidelines presented below:

1. Preprocess the datasets to create a single dataset which contains the needed information to derive meaningful interpretations in the context of the proposed application (this should include the use of simulated data if applicable);
2. Use of AI search or optimisation techniques (whichever is appropriate) in the pre-processing and cleaning of the data to ensure that the maximum amount of viable data is available for modelling;
3. Based on the dataset you have created, build a supervised or unsupervised ML model to answer questions in the context of the application related to actions to increase/support biodiversity and pollinating insect populations.
4. Give clear and concise justifications of your design decisions for tasks 1, 2 & 3 above.
5. Provide a critical evaluation of your machine learning model you have developed.
6. Provide an evaluation of the accuracy of your machine learning 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.

At this level of study, there may be multiple “right answers” to any given problem. You may have to make assumptions about ambiguous or incomplete information, interpret the information provided, or make justifiable decisions about approaches and techniques to apply. These should flow from the information provided, your understanding of the course materials, and your further research around the subject area under assessment.

To support your work, meet the requirements of a passing grade for each assessment criterion, and potentially achieve higher grades, you must state your

assumptions, interpretations and decisions, with justifications, clearly in your answer. This assists the marker in understanding your work and giving you appropriate credit for it.

Equally, you must incorporate or refer to other authors' research to support your assumptions, interpretations and decisions, with justifications. Academic writing relies on evidence from diverse sources to construct a robust argument. Instead of summarising individual sources, integrate key evidence into a cohesive discussion and ensure that citations and references are used appropriately. This approach demonstrates that your work is based on more than just personal opinion.

III. 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)

1. Overview/summary of the report, which should at least contain:
2. What was achieved/undertaken.
3. What processes were applied.
4. What tools were used and why.
5. What results were obtained and what they demonstrate.
6. 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 problem and its significance to the relevant sector.

2. Background information on the field of AI.
3. A description of the link between the 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 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) to 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 (including any use of simulated data).
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 - plus recognised variants).

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, any additional simulated data, and all source code files in appropriate formats 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 (other formats for both data and source code may be permissible, but must be clearly documented and must be readable). ***Please note that failure to provide the supporting files in the formats stated or in a readable form may result in the invalidation of this task, as while the supporting files do not directly contribute to the marks awarded, they are required as evidence of the work being carried out as stated.***

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 justify further research into the use of AI in this area.
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:

- A. A summary of the main points.
- B. A discussion of the significance of your results.
- C. Present the limitations (if any) of your approach in a clear and concise manner
- D. Any recommendation(s) resulting from your analysis.

IV. Deliverables

Your submission must contain the following files:

1. One word-processed document in **.doc**, **.docx** or **.pdf** format containing your report.
2. One zipped file containing your solution source files (not executables) and pre-processed data in **.zip** 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.

Your submission must include a cover sheet and a table of contents. The cover sheet must include:

- The module name and module code
- The title of your submission
- The **total word count** of your report.

Please note: to ensure your work is marked anonymously, **do not** put any identifying information, such as name, exam number or student ID, on your submission (including the cover sheet). Please refer to the ‘Sitting assignments in Canvas’ page in your module in Canvas for guidance on how to fully anonymise your submission.

Referencing

You are required to use the [IEEE referencing style](#) 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 your sources must be acknowledged, regardless of whether you’ve included direct quotes or not. Visit your **Academic Integrity Tutorial module** in Canvas for additional guidance on effective referencing.

To learn more about academic integrity at the University of York, visit the [Academic integrity webpage](#).

Word count

Your report must not exceed **3,000** words in total and consist of **5** clear sections to address the main task. Your response to one section will not contribute to grades in another. Please include your word count on the cover sheet of your submission.

If you exceed the word count, the markers will mark your work up to the word count maximum and then will stop marking. They will base your mark on what they have read so far.

What is in the word count?

The word count does include all words between your title (excluded) and the last word of your summary/conclusion. Any captions, tables of data, code or text in diagrams will be included in your word count.

The word count does not include:

- Cover sheet
- Abstract/executive summary
- Contents page
- Acknowledgements
- Reference lists
- Appendices

Appendices

- Appendices must only be used for supportive information, such as over-large figures or tables of data.
- They are not a device to incorporate material which would otherwise cause you to exceed the word limit.
 - Where this is the case, the appendix item, as a whole, will not be considered.
- Appendices must not exceed **5** additional pages, and all content must be referred to and discussed in the main body of the text.
- Appendices must be clearly labelled, and may contain textual information such as a key or short note that informs the reader of the nature of the data, model or other figure being presented.

Layout and formatting guidelines

- You may choose to include a table of contents, figures or tables; however this will not affect your mark.
- The main text must be written using a consistent sans serif font, in size 12. Apply 1.5 or 2.0 line spacing.
- Pages must be oriented in portrait mode and have a consistent margin/border that is at least 2cm (20mm) on all sides (including top and bottom).
- The main text must be left-aligned, not right-aligned, centrally-aligned or justified. The main text must be presented in meaningful paragraphs, rather than a single wall of text.
- Coloured text and highlighting must not be used for accessibility reasons (sufficient contrast), and ease of reading.
- Bullet points must only be used where the content is a definitive list, and must be used sparingly. They must not be used to structure large bodies of text.
- Footnotes must not be used, and content provided within them will not be considered.
- You must not exceed more than 3 levels of section headings. i.e. main heading, sub-heading 1, sub-heading 2. Your title and/or subtitle are not classed as a heading. Numbering of sub-sections to one level of decimals (e.g. 2.1) is desirable; numbering to two levels (e.g. 2.1.2) is acceptable. Numbering to further levels (e.g. 2.1.2.3) is not normally desirable.
- Conventions for headings and indentation at various levels must be applied consistently.

Images, Tables, Code and Equations

- All images and diagrams must be clear and viewable on the page without scaling.
- Tables must only be used for tabulated data and not as a means of structuring passages of text.
- All tables of data must be written into the document, not embedded as images.
- Equations may be embedded as screenshots/images in the main body of your work, provided they are viewable on the page without scaling. Equations that are added as part of the main text will be included in the word count.

- Code samples exceeding a few lines must be included as screenshots or images, and must be readable without scaling. Where they are added as main text, they will be included in the word count.
- There may be instances where it is necessary to embed a few lines of code or refer to a construct name within your writing, as text. These instances will be included in your word count. Such ‘hand-typed’ code must be made distinct, to enable it to be identified within your writing. For example, by using the Courier font. You must consider the reading experience and flow of your discussion, when making this judgement.
- All images, tables, and screenshots must be accompanied by appropriate captions and be referred to and discussed in the main body of the text. Those that are not will not be considered.

V. Acceptable use of AI

The following information provides specific guidance for this assessment about what level of AI use is appropriate for this assessment. Please refer to the page called ‘AI in Assessment Framework’ in your module in Canvas for further guidance about how to understand and apply this framework to your assessment.

Important: Remember that in all cases you must submit work that is your own, acknowledging any part of it that has been provided by another source.

Level of GenAI - integration	Description	Level applies (Y/N)?
1. No GenAI*	GenAI must not be used	N
	<p>Your assignment should be produced using information sourced by you from your learning materials and academic sources and cited appropriately.</p> <p>You are still permitted to use GenAI tools embedded within software for structure suggestions, checking spelling, grammar and referencing, but GenAI cannot be used to translate work.</p>	
2. GenAI*-assisted idea development	Students are not allowed to create content in submitted assessments using GenAI	N
	<p>GenAI can only be used for pre-assessment tasks such as for brainstorming, idea generation, and structuring.</p> <p>You are still permitted to use GenAI tools embedded within software for structure suggestions, checking spelling, grammar and referencing, but GenAI cannot be used to translate work.</p>	
3. GenAI*-assisted editing	Students can use GenAI to edit and enhance the clarity and quality of their work	Y
	<p>Students will write the full first draft of an assessment and GenAI is used to edit this draft.</p> <p>You are still permitted to use GenAI tools embedded within software for structure suggestions, checking spelling, grammar and referencing, but GenAI cannot be used to translate work.</p>	
4. GenAI* Task Completion, Human Evaluation	Students are permitted to use GenAI for some or most elements of the task and demonstrate effective use of GenAI	N
	<p>You are still permitted to use GenAI tools embedded within software for structure suggestions, checking spelling, grammar and referencing, but GenAI cannot be used to translate work.</p>	
5. Full GenAI*	Students are expected to use GenAI creatively and collaboratively to complete the task and demonstrate effective use of GenAI including the use of GenAI to translate work	N

*GenAI = Generative AI

VI. 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 may result in a reduced grade for task 3. Work NOT submitted in the requested formats (.doc, .docx or .pdf ONLY for the report and source files in an appropriately documented and readable format: e.g. .arff, .csv, .ipynb .py, or .java) may NOT be considered.

VII. Submitting work to Canvas

You will submit your assessment in the [Assignments](#) area of the module in Canvas. Before starting your assignment:

- check your Canvas module for the specific submission deadline for this assignment;
- read the page called ‘Sitting assignments in Canvas’, which you can find in the ‘Assessment Policies and Procedures’ unit of your module in Canvas.

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.

You should 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.

We strongly recommend that you [check the submission 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 or +44 80 0060 8442 (available 24 hours) in advance of the deadline. You should also report the issue and email a copy of your work to york-online-assessment@york.ac.uk before your deadline, and check your emails regularly for further instruction.

VIII. 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
- Penalties for late submission
- Marking policy for multiple submissions
- The York Online Disruptions affecting Assessment policy
- Passing mark and module reassessment

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