

Department of Computer Science

Module Learning Outcomes

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.

I. Assessment Background/Scenario

You are working as an AI engineer at a reputable organisation. You have a client from the telecommunication sector who is facing a customer churn problem due to fierce competition in the market. Customer churn, also known as 'customer attrition', is defined as the rate at which customers stop using an organisation's services within a certain timeframe. Your client is interested in applying AI techniques to explore and model its customer data. This modelling can help the company to forecast their customers' behaviour and potentially help reduce the churn rate. You have carried out initial data exploration and found:

- The data provided by the company is imbalanced as there are far more retained customers than 'churn' customers.
- The data contains numeric and nominal attributes.
- There are a few missing values in the data.

You've had a meeting with your client and have agreed to model the data using artificial intelligence techniques - namely, supervised learning and feature selection optimisation. Feature selection is important in removing irrelevant attributes and helps reduce computation cost. You are expected to present a report to your client by constructing two robust models which must follow the guidelines presented below:

- Design and build a supervised learning model on the full data.

- Use optimisation techniques (learned in this module) to find a subset of relevant features.
- Design and build a supervised learning model on the derived subset of features.
- Critically evaluate the two learning models (with and without feature selection).
- Evaluate the robustness of the generated models by applying appropriate validation techniques (and identifying a suitable subset of data for validation).

While setting the parameters of the optimisation methods, pay special attention to selecting an appropriate fitness function (evaluation criteria). The fitness function plays an important role in evolving the potential solutions and setting the direction of the search.

The dataset will be provided at the point of assessment.

II. Assessment Tasks

Your task is to design and build solutions, and to write a report based on the discussed scenario and the data provided. You should clearly draw on the 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. Further formatting details are given below.

Overall Academic Quality (10% weighting) & 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 the results demonstrated.
4. What should be reconsidered in future.

Task 1 (10% weighting) - Introduction

(suggested - 700 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 - 1000 words)

Given the scenario above, research and identify the main areas of investigation which 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 efficiently address those business problems?
3. How are these techniques being evaluated in the context of the 'problem'?
4. Critically evaluate various approaches/solutions presented in the literature.

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 learning algorithm, and one 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 optimisation techniques you would apply for feature selection and why (techniques covered in this module are: hill climbing, simulated annealing, tabu search and genetic algorithms).
4. Which supervised learning techniques you would apply and why (techniques covered in this module are: artificial neural networks, decision tree, 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 are adequately stated and discussed.

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

(suggested - 1000 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 algorithm and supervised learning technique.
3. Discuss how these results can help the business solve the customer churn problem.
4. Your arguments should also be supported by the relevant literature.

Task 5 (10% weighting) - Conclusion

(suggested - 500 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. Any recommendation(s) resulting from your analysis.

III. Deliverables

You should submit the following to the Canvas submission point:

- A zipped file containing your solution source files (not executables), in .zip format ONLY
- A separate file containing your report, in either .docx or .pdf 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.

NOTE: Failure to submit all the required elements will result in a grade of **ZERO**. Work NOT submitted in the requested format will **NOT** be considered.

Report Format

Your report should be no more than **4,000 words** in total. For each task contributing to your report, we have provided *guidelines* above on the suggested word counts for each section; however, it is your choice how to use the word count limit across the whole of this report.

You should also adhere to the following formatting instructions.

- You should submit a single word-processed file as **.docx** or **.pdf**. Other formats are NOT acceptable and are not accessible by your marker.
- Page counts that exceed the overall limits will not be graded. A line will be drawn at the limit as indicated above.

- You must state on the front page of your report the number of words/pages used and this will be checked.
- The main text should be written using a consistent **sans serif font** and font size.
- All images and diagrams must be clear and viewable on the page without scaling. They should 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.
- Your report should be fully justified OR left justified, but NOT centrally justified
- You should not exceed more than 3 levels of section headings, i.e. main heading, sub-heading 1, sub-heading 2. Your title is not classed as a heading.
- All source material that is used, whether by direct quotation or not, must be acknowledged, following the IEEE referencing style. See the University of York Academic Integrity site.
- Appendices may be used but should **not exceed 3 additional pages** and all content must be referred to and discussed in the main body of the text. Any content that is not cross-referenced within the report will NOT be considered.
- Appendices should ONLY be used for supportive information, such as over-large figures or tables of data. They are NOT a device to incorporate material that would otherwise cause you to exceed the page limit. These are not included in the page count.
- Your reference list should come after any appendices and is not included in the page count. It should be formatted using the IEEE guidelines.

DO NOT include your examination number, student ID or your name anywhere within your report.

General presentation guidance

- Coloured text and highlighting should not be used for accessibility reasons, and for ease of reading.
- Bullet points should only be used where the content is a definitive list and should be used sparingly.
- Tables should only be used for tabulated data and not as a means of structuring passages of text.

- Paragraphs must not be excessively long. “Wall of text” answers which do not contain breaks at logical points are not acceptable.
- Paragraphs should not consist of a single sentence.
- You are advised to use direct quotes minimally. They should only be used to support your points, and do not attract marks directly.
- You are expected to undertake further research to answer assessment questions. It is not sufficient to use only the module notes/content.

NOTE: Any work submitted beyond the word-count or stated limits will not be considered when grading the work.

IV. Marking Criteria

MLO	Criteria	Available marks
Overall Academic Quality		
1, 2, 3, 4	Clear and coherent across all tasks with appropriate, relevant and effective referencing and citation.	10
Executive Summary		
1, 2, 3, 4	Full and clear overview of report.	10
Task 1 – Introduction		
1	Clear description of the problem and its significance to the sector. Discussion of relevant background/underpinning theory in AI. Alignment of the business problem to the field of AI. Description of the solution(s) presented.	10
Task 2 – Literature Review		
4	Examines current, relevant problem areas. Considers and evaluates proposed solutions. Includes a critical evaluation of any, and all relevant approaches/solutions under review.	20
Task 3 – Research Design		

1, 2, 3, 4	To cover pre-processing techniques, selection of features and optimisation techniques, evaluation of selected techniques and discussion of algorithmic parameters, accompanied by a clear and justified rationale in each case.	20
Task 4 - Experimental Result and Analysis		
2, 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
Task 5 - Conclusion		
1, 2, 3, 4	A clear and concise presentation of findings with consideration of limitations and further development.	10
	TOTAL:	100