

**ASSESSMENT BRIEF**

**ACADEMIC YEAR 2023/24**  **TRIMESTER 1-2**

**LEVEL: 5**

**MODULE TITLE:** Machine Learning

**MODULE CODE:** MOD008371

**COURSE(S): BSc(Hons) Data and Analytical Science**

**YOUR TASKS**

**TASK ONE: Element 010**– Learning Outcome 1, 2, 3 & 4

Coursework 3000-word equivalent - worth 100% of your module grade.

The deadline by which this assignment must be uploaded and presented is: **2pm on 22nd April 2024**

**TASK ONE: Element 010 – Learning Outcome** 1, 2, 3

**ASSESSMENT INTRODUCTION:**

This assignment gives you an opportunity to choose from the Machine Learning techniques you have researched and studied and apply your choice to a practical decision-making problem. You are also expected to assess the performance of your solution.

**Task 1: Model Development (Outcome 4)**

**Deliverable**: Trained and tested machine learning model, or models, for a given scenario.

Design and implement a Machine Learning model to decide whether a car driver is safe to overtake on a single-carriageway road. Ignore all traffic except your vehicle, the vehicle you are overtaking and an oncoming vehicle.

You are provided with a black-box code routine, which you can call as often as you like. It returns the following data:

1. speed of your vehicle in m/s **(OvertakingSpeedMPS)**

2. speed of the oncoming vehicle in m/s (**OncomingSpeedMPS)**

3. distance to the oncoming vehicle in m (**InitialSeparationM)**

4. the expected outcome – true or false (**Success)**

The data items 1-3 vary randomly with each call.

You will be provided with 3 versions of the black-box routine:

1. A basic version that returns reliable (but unrealistic) data, that is based on a simple algorithm to determine the success.
2. An intermediate version that returns data similar to the basic version but with a hidden variable that factors in the speed of the vehicle you are overtaking.
3. An advanced version building on the intermediate version and incorporates a more realistic algorithm that factors in further variables affecting the phsysics of the situation, such as wind-speed, but in turn results in a less predictable result.

You should use this data to build training and testing sets and then develop a solution using a machine-learning technique, such as a decision tree, linear or logistic regression, an artificial neural network, a genetic algorithm, etc.

The key is to iterate over the training of your selected model or models, improving your data preprocessing, model selection, and hyperparameter tuning based on model evaluation results. Since you are dealing with hidden variables and uncertainties, consider models that can handle noise well and explore techniques to quantify prediction uncertainties.

You may include code from the course material or published material provided that you reference your source and explain in detail how you used and adapted the material.

You should analyse the performance of your solution by comparing the outcome it predicts with the outcomes from the code routines. You should experiment and adapt your implementation, recording your attempts as you go, to fine-tune the algorithm to produce a reliable prediction.

**Further investigation.**

It should be possible to develop a separate model, utilising the advanced black-box routine data to predict the overtaking speed required to successfully overtake in a given situation (**OvertakingSpeedMPS)**, when provided with the **InitialSeparation**, **OncomingSpeed**, and **Success** data. This task would likely require a different type of model to the initial problem presented. Additional marks are available if you can test and recommend a model to solve this problem.

You must upload your code (.ipynb files) to GitHub and submit your slides for Task 2 to Canvas by 2pm on Monday 22nd April, 2024. Following this you will present your solution at a time scheduled with your tutor for the afternoon of 22nd April.

You will present your solution and decisions made during the implementation in task 2.

**Task 2: Presentation (Outcome 1, 2 & 3)**

**Deliverable**: A presentation on the implementation of machine learning models for a given scenario with a demonstration of the solution.

You will need to produce a 30-minute presentation demonstrating your understanding of machine learning concepts and how these are implemented to provide predictable results to the problems presented in task 1.

A suggested presentation structure follows:

1. Introduction to the scenario and data preprocessing.
2. Demonstration of the functional model.
3. Discussion of technique/s chosen. How does it work? Why did you choose it? Why use this technique over another? What is the economic, social, legal, and ethical context of data used in machine learning applications such as used in this project compared to other scenarios.
4. Code discussion
   1. Commentary on the key features of your programmed code; you will demonstrate the code live during the presentation.
   2. Explain how your program makes use of any published material that it employs.
5. Present an analysis of the quality of your program's results and report any steps you have tried to improve that quality. (If your program does not deliver a practical solution, discuss how you would investigate the quality.). You may include figures of data to help illustrate your points, or, for higher marks you might provide some statistical analysis of the results, including graphs for illustration.
6. Further Investigation, if carried out.
7. A conclusion to the presentation.
8. A link to a GitHub with your full code listings, timestamped to demonstrate that no further changes have been made following the presentation.

You must upload your slides for the presentation by 2pm Monday, 22nd April, 2024, and then attend at a time schedule with your tutor for the afternoon, where you will have 30 minutes to demonstrate your solution and present your findings.

**YOUR FEEDBACK**

**FORMATIVE FEEDBACK**

Throughout the module you are advised to take advantage of formative feedback opportunities. These are:

Week 17

Week 19

Week 21

Week 24

**SUMMATIVE FEEDBACK**

You will receive comments on your work within 20 working days of the deadline detailed above. Your feedback may include your un-moderated fine grade, but please note that this will remain subject to change until it has been ratified by the Module Approval Panel. Your final, approved grade will appear on your student record after each Assessment Panel.

Please refer to key dates on e:Vision.

Your work will be assessed against learning outcomes defined in the module definition; in undertaking the tasks outlined above, you should ensure that you evidence these. Once your assessor has confirmed that you have met the requirements of each learning outcome, the assessment criteria and marking standards (below) will be used to decide your fine grade.

|  |  |
| --- | --- |
| **Learning Outcomes (threshold standards):** | |
|  | On successful completion of this module the student will be expected to be able to: |
| **Knowledge and understanding** | 1. Demonstrate an understanding of the key concepts underpinning machine learning. 2. Demonstrate knowledge of most common machine learning algorithms. |
| **Intellectual, practical, affective and transferable skills** | 1. Identify the most appropriate algorithms for the problem to be solved and exercise a practical understanding of the economic, social, legal and ethical context of data use in machine learning applications. 2. Design appropriate machine learning solutions, train and validate a range of models for supervised learning, evaluate their performance and their technical and non-technical impact, recognising their limitations. |

**Full Module Definition Form available separately via Canvas**

Assessment Criteria:

1. **Data Preprocessing and Understanding** (10%)
   * Demonstrates a thorough understanding of the data provided by the black-box routines.
   * Applies appropriate preprocessing techniques to prepare the data for model training.
2. **Model Implementation** (10%)
   * Justifies the choice of model and associated machine learning technique(s) for the given scenario compared to other potential models, providing a rationale for their selection over others and considering the task's complexities and data characteristics
3. **Understanding and Application of Machine Learning Concepts** (10%)
   * Demonstrates a comprehensive understanding of key machine learning concepts and their application to the given problem.
   * Clearly explains the chosen technique(s), providing a rationale for their selection over others.
   * Discusses the economic, social, legal, and ethical context of data use in machine learning applications, showing awareness of the broader implications of the work.
4. **Technical Discussion** (10%)
   * Provides a detailed discussion of the programmed code, including the use of any published material.
   * Demonstrates the effective implementation of the chosen model(s), including code efficiency, code quality (such as consistency) and the effective use of relevant libraries or frameworks.
5. **Training and Hyperparameter Tuning** (10%)
   * Shows a systematic approach to training the model, including the selection and tuning of hyperparameters.
   * Iterates over model training to improve performance, demonstrating an understanding of overfitting, underfitting, and model validation techniques.
6. **Model Evaluation and Comparison** (10%)
   * Evaluates the model's performance using appropriate metrics, comparing predicted outcomes with those from the code routine.
   * Discusses the reliability of the model and suitability for use in a real-world scenario, including its handling of uncertainties and hidden variables, and compares the performance across different versions of the black-box routine.
7. **Statistical Analysis and Data Presentation** (10%)
   * Includes statistical analysis of the results, where appropriate, with graphs or other figures to illustrate key points.
   * Demonstrates an ability to critically evaluate data and draw meaningful conclusions.
8. **Further Investigation and Innovation** (20%)
   * Explores the development of an additional model to predict the overtaking speed required for a successful manoeuvre, demonstrating creativity and problem-solving skills.
   * Tests and evaluates the secondary model, providing recommendations and justifications for its use.
9. **Structure, Clarity, and Cohesion of Presentation** (10%)
   * Presents information in a well-structured format, with appropriate citations.
   * Explanations are clear, concise, and well-organised, with appropriate use of technical language.