DAT7016 - Applied Artificial Intelligence

Assessment 2 - Portfolio

Author: Shivang Shukla - ss21@bolton.ac.uk

Due Date: 29th April 2024 (No later than 23:59)

Assignment Instructions: Please read carefully

1. Weighting: **60% of module marks.**
2. Assessment number - 2.
3. Assessment Type - Coursework.
4. Assessment Name - Portfolio.

Presenting Your Work

1. For writing **report**: use LaTeX, MS Word, LibreOffice Writer or any similar word processing tool.
2. While it is encouraged to compose this assignment using LaTeX, it is important to note that no additional marks will be granted for doing so.
3. Please ensure that the written assessment is processed in **Arial** or **Calibri Light** font size **12**.
4. There should be double-spacing and each page should be numbered.
5. The body of the **submitted work** must be in your own words.
6. Make sure to assign a number and clear label to every table and figure.
7. Position the figure's captions **underneath** to the corresponding figures and place the table's captions **above** all the respective table.
8. Do not forget to cross-reference all figures and tables.
9. There is a word limit of **3000 words**.
10. You should include a word count at the end of the assessment (excluding references, figures, tables and appendices).
    - Upon exceeding the specified word limit, one of the following penalties will apply:
      (a) Up to 10% over the specified word length = no penalty
      (b) 10 – 20% over the specified indicative word length = 5 marks subtracted (but if the assessment would normally gain a pass mark, then the final mark to be

no lower than the pass mark for the assessment).

(c) More than 20% over the indicative word length = if the assessment would normally gain a pass mark or more, then the final mark will capped at the pass mark for the assessment.

11. Writing and constructing your submission by copying and pasting chunks of text from various external sources (websites/books) is **not acceptable**. Please check the module guide for further information.
12. It is important to indicate clearly in your own work where you have included the work of others. In Computer Science, this could include the reuse of designs and code as well as copying or quoting text.
13. Please avoid the usage of first-person pronouns such as "I", "we", "me", "my" etc. in formal report writing style.
14. The implementation of your work should follow **Python 3.7 or higher** versions.
15. For any comparisons, please use tabular format.
16. It is mandatory to **present your work** to your module tutor, before uploading to on Turnitin otherwise **"No marks will be awarded for the submitted work"**.
17. The presentations will be on **6th May 2024** i.e. Monday.
18. Each presentation will last for 10 minutes including 2 minutes for Q and A.

References

1. It is expected that the Reference List will contain between **fifteen to twenty sources**. As a MINIMUM the Reference List should include **four refereed academic journals and five academic books**.
2. All written work should be referenced using the standard University of Bolton referencing style– see: `https://libguides.bolton.ac.uk/resources/referencing/`

Submitting Your Work

1. The file must be uploaded to Turnitin using the link provided on **AIN7301 Moodle page**.
2. **For this assignment, you are only permitted to submit your work only once.**
3. Ideally, the report should be submitted in a **PDF format**, followed by all the instructions provided in the assessment brief.
   - The **Report** must be in **PDF** format and should be uploaded to Turnitin with the filename "xxxxxxAIN7301A2.pdf" where xxxxxx is your student number.
   - **Associated codes** must be in the **PY** format and should be uploaded to Turnitin with the filename "xxxxxxAIN7301A2.py" where xxxxxx is your student number.
   - The **presentation** file must be in the **PPTX** format and should be uploaded to Turnitin with the filename "xxxxxxAIN7301A2.pptx" where xxxxxx is your student number.
4. The submissions will be marked anonymously so you **must not include your name** in your submissions.

5. While submitting your work via Turnitin, you are implicitly declaring the work to be your own.
6. **Email submissions are not acceptable** and no marks will be awarded for any of the work submitted via email.
7. If Turnitin detects a similarity value exceeding 20%, you will be required to justify any content that may appear plagiarised. In this context, justification entails demonstrating that your work is not simply a verbatim reproduction sourced from external materials, but rather showcases a genuine understanding and possession of similar knowledge.
8. It's important to note that ***"no marks will be awarded for any work that remains unjustified"*** and should the conditions outlined in point 8 of the "Submitting Your Work" section be met.

Deliverable

1. Report
2. Presentation and Live Demo.
3. Associated Codes in a **single-click executable** file.

Specific Assessment Criteria

1. This assessment will be graded against the University of Bolton's "General Assessment Criteria for Written Assessments Level HE7".This is provided in the module guide.
2. This assignment is designed to assess your ability to apply AI algorithms effectively in real-world scenarios, critically evaluate their output and communicate your findings professionally. Therefore, emphasise both technical competence as well as analytical skills in your submissions as the below-mentioned criteria will also be taken into consideration while grading the submitted work:
   - Problem Selection and Justification: **5%**
   - Algorithm Selection and Implementation: **30%**
   - Evaluation and Analysis: **35%**
   - Originality, Creativity and Independence: **15%**
   - Report Writing and Presentation: **15%**
3. All the deliverables are integrated and should be submitted as per the instructions provided. ***No marks will be awarded if any of the listed 'deliverables' are not submitted within the stipulated time frame.***

Learning Outcomes and Assessments

1. This assessment element contributes to the learning outcomes for the overall coursework assessment as follows:

Learning Outcome 2	Assessment No. and Type
LO2: Apply appropriate AI algorithms in real-world contexts, demonstrating originality, creativity and independence.	• Assessment 2 - Portfolio
Learning Outcome 3	
LO3: Critically evaluate the output of the innovative AI algorithms used to address complex problems.	• Assessment 2 - Portfolio

Late Submission

- Late work will be subject to the following penalties:
  1. Up to 7 calendar days late = 10 marks subtracted but if the assignment would normally gain a pass mark, then the final mark should be no lower than the pass mark for the assignment.
  2. More than 7 calendar days late = This will be counted as non-submission and no marks will be recorded.
  3. Late submission of assessments on refer and those which are graded Pass/Fail only, is not permitted unless an **extension is approved**.

Academic Misconduct

1. It is expected that students engage critically with the topics, providing evidence-based arguments and insights supported by academic literature and real-world examples.
2. Plagiarism will not be tolerated and may result in severe penalties.
3. Please follow "**Academic Misconduct**" section provided in the module guide for further details.

Assignment Objective:

This assignment aims to assess your ability to apply appropriate AI algorithms in real-world contexts, demonstrating originality, creativity and independence. Additionally, you will critically evaluate the output of these innovative AI algorithms used to address complex problems.

Instructions:

Problem Selection (5%):

1. Identify and select a complex real-world problem(s) that can be addressed using AI algorithms.
2. Justify your choice of problem(s) and discuss the significance and relevance in the context of AI research and applications.

Please note that the list provided below may offer guidance on potential real-world problem-based projects. Alongside the listed topics, you'll also have the option to explore the URLs provided below as well. The resource list and URLs may also assist you in discovering the latest project ideas, recent advancements, ongoing problems, system requirements, most recent literature and associated publicly available source code and may be helpful to achieve the final design solution.

List of a Few Example Projects

1. **Text**
    - Autonomous Tagging of StackOverflow Questions
        - Make a multi-label classification system that automatically assigns tags for questions posted on a forum such as StackOverflow or Quora.
        - Dataset: StackLite or 10% sample
    - Keyword/Concept identification
        - Identify keywords from millions of questions
        - Dataset: StackOverflow question samples by Facebook
    - Topic identification
        - Multi-label classification of printed media articles to topics
        - Dataset: Greek Media monitoring multi-label classification
2. **Natural Language Processing**
    - Sentence to Sentence semantic similarity
        - Can you identify question pairs that have the same intent or meaning?
        - Dataset: Quora question pairs with similar questions marked
    - Fight online abuse
        - Can you confidently and accurately tell whether a particular comment is abusive?
        - Dataset: Toxic comments on Kaggle
    - Open Domain question answering
        - Can you build a bot which answers questions according to the student's age or curriculum?
        - Facebook's FAIR is built in a similar way to Wikipedia.
        - Dataset: NCERT books for K-12/school students in India, NarrativeQA by Google DeepMind and SQuAD by Stanford
    - Automatic text summarization
        - Can you create a summary with the major points of the original document?
        - Abstractive (write your own summary) and Extractive (select pieces of text from the original) are two popular approaches
        - Dataset: CNN and DailyMail News Pieces by Google DeepMind
    - Copy-cat Bot
        - Generate plausible new text which looks like some other text
        - Obama Speeches? For instance, you can create a bot which writes some new

speeches in Obama's style
    - Trump Bot? Or a Twitter bot which mimics @realDonaldTrump
    - Narendra Modi bot saying "doston"? Start by scraping off his Hindi speeches from his personal website
    - Example Dataset: English Transcript of Modi speeches
  - Sentiment Analysis
    - Do Twitter Sentiment Analysis on tweets sorted by geography and timestamp.
    - Dataset: Tweets sentiment tagged by humans
3. **Forecasting**
  - Univariate Time Series Forecasting
    - How much will it rain this year?
    - Dataset: 45 years of rainfall data
  - Multi-variate Time Series Forecasting
    - How polluted will your town's air be? Pollution Level Forecasting
    - Dataset: Air Quality dataset
  - Demand/load forecasting
    - Find a short-term forecast on the electricity consumption of a single home
    - Dataset: Electricity consumption of a household
  - Predict Blood Donation
    - We're interested in predicting if a blood donor will donate within a given time window.
    - More on the problem statement at Driven Data.
    - Dataset: UCI ML Datasets Repo
4. **Recommendation System**
  - Movie Recommender
    - Can you predict the rating a user will give on a movie?
    - Do this using the movies that the user has rated in the past, as well as the ratings similar users have given similar movies.
    - Dataset: Netflix Prize and MovieLens Datasets
  - Search + Recommendation System
    - Predict which Xbox game a visitor will be most interested in based on their search query
    - Dataset: BestBuy
  - Can you predict Influencers in the Social Network?
    - How can you predict social influencers?
    - Dataset: PeerIndex
5. **Vision**
  - Image classification
    - Datasets:
      * CIFAR-10
      * ImageNet

        * MS COCO is the modern replacement to the ImageNet challenge
        * MNIST Handwritten Digit Classification Challenge is the classic entry point
        * Character recognition (digits) is the good old Optical Character Recognition problem
        * Bird Species Identification from an Image using the Caltech-UCSD Birds dataset dataset
    - Diagnosing and Segmenting Brain Tumors and Phenotypes using MRI Scans
        * Dataset: MICCAI Machine Learning Challenge aka MLC 2014
    - Identify endangered right whales in aerial photographs
        * Dataset: MOAA Right Whale
    - Can computer vision spot distracted drivers?
        * Dataset: State Farm Distracted Driver Detection on Kaggle
- Bone X-Ray competition
    - Can you identify if a hand is broken from an X-ray radiograph automatically with better than human performance?
    - Stanford's Bone XRay Deep Learning Competition with MURA Dataset
- Image Captioning
    - Can you caption/explain the photo a way humans would?
    - Dataset: MS COCO
- Image Segmentation/Object Detection
    - Can you extract an object of interest from an image?
    - Dataset: MS COCO, Carvana Image Masking Challenge on Kaggle
- Large-Scale Video Understanding
    - Can you produce the best video tag predictions?
    - Dataset: YouTube 8M
- Video Summarisation
    - Can you select the semantically relevant/important parts from the video?
    - Example: Fast-Forward Video Based on Semantic Extraction
    - Dataset: Unaware of any standard dataset or agreed-upon metrics? I think YouTube 8M might be a good starting point.
- Style Transfer
    - Can you recompose images in the style of other images?
    - Dataset: fzliu on GitHub shared target and source images with results
- Chest XRay
    - Can you detect if someone is sick from their chest XRay? Or guess their radiology report?
    - Dataset: MIMIC-CXR at Physionet
- Clinical Diagnostics: Image Identification, classification & segmentation
    - Can you help build open-source software for lung cancer detection to help radiologists?
    - Link: Concept to Clinic challenge on DrivenData

- Satellite Imagery Processing for Socioeconomic Analysis
  - Can you estimate the standard of living or energy consumption of a place from night time satellite imagery?
  - Reference for Project details: Stanford Poverty Estimation Project
- Satellite Imagery Processing for Automated Tagging
  - Can you automatically tag satellite images with human features such as buildings, roads, waterways and so on?
  - Help free the manual effort in tagging satellite imagery: Kaggle Dataset by DSTL, UK

6. **Covid 19**
   - Bing Coronavirus
     - Classify Bing Queries as either specific (e.g. about a specific location) or generic. You might have to figure out a more exact definition of specific or generic though
     - Dataset: BingCoronavirusQuerySet
   - Covid Clinical Data
     - Rank and sort high-risk patients using clinical data. Pick an interpretable approach if you can.
     - Dataset: CovidClinicalData
     - **If you haven't already, checkout Kaggle's Covid19 Section as well. It has datasets and ideas both.**
7. **Music and Audio**
   - Music/Audio Recommendation Systems
     - Can you tell if two songs are similar using their sound or lyrics?
     - Dataset: Million Songs Dataset and its 1
     - Example: Anusha et al
   - Music Genre recognition using neural networks
     - Can you identify the musical genre using their spectrograms or other sound information?
     - Datasets: FMA or GTZAN on Keras
     - Get started with Librosa for feature extraction

Please follow the example URLs to explore more about the **recent advancements, most recent literature** and associated **publicly available source code**.

1. `https://paperswithcode.com/`
2. `https://github.com/topics`
3. `https://scholar.google.co.uk/`

Algorithm Selection and Implementation (30%):

1. Choose appropriate AI algorithm(s) based on the problem(s) identified.
2. Implement the selected algorithm(s) using suitable programming languages or AI frameworks.

3. Provide a detailed explanation of the implementation process, including any modifications or enhancements made.

Evaluation and Analysis (35%):

1. Evaluate the performance and output of the AI algorithms based on predefined metrics or criteria.
2. Critically analyse the strengths, weaknesses, opportunities and threats associated with the algorithm(s) used.
3. Discuss any ethical considerations or implications arising from the application of these algorithms.

Originality, Creativity and Independence (15%):

1. Demonstrate originality, creativity and independence in your approach to solving the problem(s) using AI algorithms.
2. Highlight any novel ideas, methodologies or solutions implemented during the project.

Report Writing and Presentation (15%):

1. Prepare a well-structured and professionally written **3000 words** report documenting your findings and insights. The structuring of the report must follow the sequence provided below:
    - Front Page
        - University Logo
        - Department
        - Module Code
        - Assessment Name and Number
        - Student Number
    - Abstract
    - Acknowledgement
    - Table of Contents
    - List of Figures
    - List of Tables
    - Introduction
    - Literature Review
    - Methodology
    - Implementation
    - Results
    - Conclusion and Future Direction
    - References
2. Design a visually appealing presentation that effectively communicates key aspects of your project for 08 minutes followed by another 2 minutes for questions and answers.
3. Please ensure clarity, coherence and accuracy in both the report and presentation.

Recommendations:

Before moving forward, consider reading the suggestions provided below that may help you to save some of your valuable time.

1. **Start Reading on the Background before the start of your project:** Ideally, by the time you start working full-time on the project, you should know what to do, how it relates to existing approaches and have some idea of how to do it. This might also be a good time to get familiar with your machine learning framework of choice (for example PyTorch!).
    - Before selecting and finalising any topics/papers/articles/manuscripts, please make sure that the source code is publicly available and can be re-implemented.
    - Before implementation, please double-check the hardware and software requirements that are generally provided under "specification or Requirements".
    - Use of GPU could be the most basic requirement in such tasks. So, please make sure that you have access to a machine compatible with the specified requirement.
    - If you are using Google Colaboratory `https://colab.research.google.com/`, please ensure that the GPU is activated otherwise, you may have to spend most of your time resolving some strange and unknown errors.
    - A *light implementation* of the task is also acceptable. Here, light means: reducing the size of the dataset to save some processing time.
2. **Use Version Control:** You really should use version control for your assessment. There is nothing worse than losing all your hard work days before the deadline. If you do not have one already, open a GitHub account `https://github.com/`. As a student, you get free private repositories. If you do not know about version control, learn it now and thank yourself later.
3. **Evaluate using Random Repeats:** In academia, it is unlikely that you have access to more than a handful of GPUs during your project. However, particularly in deep reinforcement learning, it is important to not draw premature conclusions from a single or few experiments. Ideally, you want to repeat experiments multiple times and as mentioned, get a sense of the robustness of different starting conditions and hyper-parameters.
4. **Start Writing Early and Consistently throughout this assessment:** If you are doing this mini-project, your work will be assessed based on your written report, presentation and implementation. **The work will not just be based on the outstanding work that you did in your implementation but did not get enough time to write it, clearly.** So, start writing early and do not underestimate the effort of disseminating your research. State your aims, hypothesis and contributions clearly and allow the reader to follow your thought process. Explain your design choices and discuss your findings clearly. Ideally, you should write your report consistently during the course of this assignment. That way, you force yourself to think your next steps through and it is less likely that you forget about any important information when the deadline gets close.

5. **Remember: "TIMELINE IS IMPORTANT".**