LSBU EST 1892

Coursework Specification

CW_Specification_CSI_6_ARI_23-24

Read this coursework specification carefully, it tells you how you are going to be assessed, how to submit your assessment on-time and how (and when) you'll receive your marks and feedback.

Module Code	CSI_6_ARI
Module Title	Artificial Intelligence
Lecturer	Dr Bugra Alkan
% of Module Mark	60%
Distributed	**06/03/2024**
Submission Method	Submit online via this Module's Moodle site
Submission Deadline	**19/04/2024 at 18:00***
Release of Feedback & Marks	Feedback and provisional marks will be available in the Gradebook on Moodle from 03/05/2024

* Module leader will be available for any troubleshooting related to coursework submission up to the deadline time however please submit well in advance in case you run into technical issues.

Note:
The pass mark for this Coursework is **30%** (30 out of 100). The module has an overall pass mark of **40%** (40 out of 100) (The final grade for this module will be based on a total of 40% Exam and 60% CW).

Coursework Aim:

This coursework requires the student to work on a real-world engineering problem where they need to tackle various tasks such as data preparation, evaluation, and discussion of the performance of the proposed Artificial Intelligence (AI) solution.

The aim of this coursework is to demonstrate the ability of the student to set up a Machine Learning (ML) processing pipeline, starting from data preparation, and ending with critical appraisal of the obtained results, and to present their work thoroughly and professionally.

The work should therefore present all the steps required to prepare and process the data, with a reasonable description of all the relevant choices performed when alternative methods were available.

Coursework Details:

Type:	Technical Report (**60%** Overall Grade)
Overview	The primary goal of this individual CW assignment is to assess your understanding of fundamental theory, concepts, and various algorithms in artificial intelligence and machine learning, as well as your ability to apply appropriate tools/platforms, such as Python, MATLAB, and R, to carry out a data analytics project.
	Project Background:
	In this individual assignment, you will be acting as the AI expert working for Arasaka Electric Railway (AER), which is a leading Shinkansen bullet train operator based in Hokkaido, Japan. In recent months, the company has lost an important portion of its customer base due to travel delays related to strike actions and poor performance. To identify the underlying causes of concerns, the company conducted fieldwork over the past few weeks, gathering survey opinions from a representative sample of bullet train passengers. The survey results and travel data from customers have been gathered and stored in a database for a more in-depth analysis of cause-and-effect relationships. The company now wants to hire an expert who can further explore the collected data and potentially employ artificial intelligence methods to delve deeper into the underlying root causes of the issue.
	In this coursework, you are required to follow a proper methodology and apply various techniques covered in the CSI-6-ARI Artificial Intelligence lectures and tutorials: i) to analyse the dataset assigned to you to address the business concerns and problems that have been raised, ii) to build ML models that can accurately classify customer satisfaction based on the given travel and survey features, iii) to verify and compare your models using various key performance indicators, and iv) to discuss your results and offer guidance to the company.
	You will be given a '.csv' file including the on-time performance of the trains along with the passenger's information and passengers' feedback on various parameters related to the travel along with their overall experience. In the dataset, each row represents a sample that contains a survey of a passenger who was explicitly asked whether they were delighted with their overall travel experience, and that is captured in the data of the survey report under the variable labelled 'Overall_Experience' (binary response, 0: not satisfied and 1: satisfied). The objective of this coursework is to understand which parameters play an important role in swaying passenger feedback towards a positive scale.
	Dataset
	Download your dataset from the module's VLE site via this **link**. The dataset (unique to each student) consists of 30000 samples twenty-three attributes (i.e., features), a response (i.e., binary overall experience) and a participant ID. You should download and use the dataset corresponding to your student ID number. For example, if your ID number is '3420062', then you need to use 'CSI_6_ARI_23_24_01'.
Tasks	You are required to undertake the following tasks:

	1. Problem Identification
	  • Learn about the basic characteristics of the dataset, such as the specific business context associated with the data, the total number of attributes (dimensions, variables), the data type of each attribute, the value range/mode, skewness, and kurtosis of each attribute, the total number of instances, and simple data exploration with essential plotting, among other things.
	  • Provide a small literature review on the assigned machine learning algorithms.
	
	Five randomly selected classification algorithms are assigned to each student. In this CW, you must only use these assigned algorithms to carry out your ML project. Please see this **link** for further info.
	
	2. Data Preparation
	  • Identify which variables to be used in which analysis using exploratory data analysis.
	  • Choose appropriate methods for data pre-processing, including handling missing data, encoding categorical variables, data standardisation (e.g., data scaling, data normalisation, etc.,), and train/validation dataset splitting (Please see the related module tutorials on the data pre-processing).
	
	3. Model Development
	  • Using the pre-processed dataset undertake the machine learning tasks you have identified. You are required to apply **FIVE** ML algorithms to this problem. Algorithms are assigned to each student randomly. Please see your algorithms using **this link**.
	  • Select the appropriate Key Performance Indicators (KPIs) for evaluating the models' efficacy and provide an in-depth explanation of how they will be used.
	  • To build the optimal models, relevant hyperparameters could be considered for **EACH** of the assigned algorithms. You can use grid search, elbow methodology, random search, Bayesian search, or any other algorithms.
	
	4. Model Evaluation and Comparison
	  • Compare the performances of the developed models on both test and train datasets (you can create test, validation and train sets from the allocated dataset using cross-validation techniques) considering the selected KPI metrics.
	  • Discuss the meaningfulness and usefulness of the models built, and how these models can be used to address the original business concerns.
	  • Suggest next steps for the company to take in developing this into a viable business option.

Word Count:	As a guide, **aim** for **3,500 words**, excluding Title page, Table of Contents, tables and figures, footnotes, and bibliography. The **maximum word limit** is **5,000 words**. References and appendix will not count towards the word

	total. Add a word count at the title page. (See this **Link** for inserting the word count in your document.)
Structure of your report:	Your deliverables for this machine learning project are: **i)** a formal written **technical report, ii) a short (5 minutes) video presentation** explaining your technical implementation**, and iii)** your **Python/MATLAB/R code**. Specific details of technical report are given below.
	Technical Report:
	Note: Your report should be **narrative in style**, with human explanation and commentary. A "report" that is *merely* a collection of screenshots, code and data dumps will be **graded very poorly.**
	The report must contain the following sections:
	Title Page
	Your student ID number and the title of the report **must be provided**. On the title page, **do not** include any other personal information including your name.
	Table of Content
	List of Figures *(Optional but recommended)*
	List of Tables *(Optional but recommended)*
	Abstract Abstract provides a summary of your work.
	Section 1. Introduction This section provides a general presentation of the problem with relevant background.
	Section 2. Dataset This section should describe the data and its content (i.e., features and labels). Also, it must provide an exploratory data analysis (e.g., pair-wise correlation plot, etc.) on the dataset.
	Section 3. Data Pre-processing This section contains data pre-processing steps i.e., data standardisation (e.g., normalisation), train/validation split, and dimensionality reduction (if applicable), etc.
	Section 4. Algorithms Provide a **short summary** of the assigned algorithms with their technical details (possibly with equations and diagrams).
	Section 5. Implementation All the processing steps of the assigned ML methods are described in this section (including hyperparameter tuning steps). Selected ML methods, hyperparameter tuning approach, and KPI metrics (e.g., AUC, ROC, Acc, etc.,) must be discussed.

Section 6. Results
Results are presented, along with an evaluation of the effectiveness of the ML methods used, and a critique of the implemented solutions is provided in this section. Do not forget to include any necessary diagrams and visuals.

Section 7. Discussion
This section provides a discussion on the results found and outlines suggestions on the next steps for the company to take in developing this into a viable business solution.

Section 8. Conclusion
This section concludes the report by summarising the findings.

References

Note: Sub-sections can be added if more clarity is required.

How are marks awarded:	The **technical report** will be **marked** using the marking criteria provided in this coursework specification document. All the students are advised to fully understand the marking criteria before starting the coursework.
Presentation:	In total you must submit three files: *i)* your technical report work (.pdf), *ii)* the source code (Python notebook, MATLAB, or R file), and *iii)* a short video recording.

The report:
There is no specific report template assigned to this CW. You can use any format or template of your choice. Recommendations are as follows: A common 10- or 12-point font (Calibri is good). Margins: 1.5 inches on the left, 1 inch each for top, bottom and right. Paragraphs for sections: 1 or 1.15 spaced, first line indented 0.5 inch, full justified. Make sure that figures and tables are captioned.

If you used any source (please **avoid online sources without peer-review**), these sources must be acknowledged and referenced, and a bibliography should be provided.

Do not include your student's name in the report. Work must be submitted as a PDF, and must be named as follows:
CSI_6_ARI_2324_CW_<Student_ID>.pdf
(Please replace <Student_ID> with your student number)

The source code:
The source code must be submitted via VLE.
Name your source code file as follows:
CSI_6_ARI_2324_CWCode_<Student_ID>.*<extension>*
(Please replace <Student_ID> with your student number)

The video recording:
You are required to provide a video recording demonstrating your technical implementation. You can choose to either create a PowerPoint presentation or discuss over your source code depending on your preference. The video recording must **not exceed five minutes** and

	should solely focus on your machine learning implementation. You can use Panopto, Open Broadcaster Software, or PowerPoint to create your video recordings. The recommended format is .mp4. Video recordings **must not exceed 50 megabytes**. The **presenter must be clearly visible** in the video recordings. Please note that all student data will be stored **securely**, available only to the markers on the module, and retained alongside the module content for quality purposes. Name your video presentation as follows: **CSI_6_ARI_2324_Video_<Student_ID>.**_<extension>_ (Please replace <Student_ID> with your student number) Do not submit any archive file format, such as zip or rar. Do not merge your documents into a single pdf or an archive file. If your code or video file exceeds 50 MB in size, you can send it to your module leader via email (alkanb@lsbu.ac.uk).
Referencing:	The students should use either IEEE Style referencing or Harvard style referencing. Please see LSBU Harvard referencing guide for more information. **Note**: Inappropriate reference formats or mixed referencing will be penalised as stated in the assessment rubric's academic integrity criteria!
Regulations:	Make sure you understand the University Regulations on expected academic practice and academic misconduct. Please refer to LSBU's late submission and extenuating circumstances notification procedures via this link. **Note in particular:** Your work must be your own. Markers will be attentive to both the plausibility of the sources provided as well as the consistency and approach to the writing of the work. Simply, if you do the research and reading, and then write it up on your own, giving the reference to sources, you will approach the work in the appropriate way and will cause not give markers reason to question the authenticity of the work. All quotations must be credited and properly referenced. Paraphrasing is still regarded as plagiarism if you fail to acknowledge the source for the ideas being expressed. **TURNITIN**: When you upload your work to the Moodle site it will be checked by anti-plagiarism software. Your similarity index for the report must not be more than **20%**. Any report with more than a **20%** similarity index will be subject to Academic Misconduct Investigation.

Learning Outcomes

This assessment (CW in the table below) will fully or partially assess the following learning outcomes for this module.

Learning outcome	Assessed by
	Coursework Report
A. Knowledge and understanding	
• Appraise a range of techniques that have been employed to develop intelligent systems of various kinds.	Fully
• Consistently producing and reviewing research informed work which applies and is at the forefront of the developments in the domain	Partially
B. Intellectual skills	
• Evaluate AI problems for current and future feasibility and suggest approaches that might be applied.	Fully
C. Practical skills	
• Develop applications that exhibit intelligence in a specific context using established techniques.	Fully
D. Transferable skills	
• Evaluate the possibilities and limitations of intelligent systems being implemented now and, in the future, and assess their suitability for diverse applications.	Partially
• Self-manage your study time and work effectively to meet deadlines, select and evaluate appropriate knowledge, skills, etc.; also select and evaluate supporting resources/tools for a particular purpose, as well as being able to make effective contributions as team member/leader when required.	Partially

Assessment Criteria and Weighting

LSBU marking criteria have been developed to help tutors give you clear and helpful feedback on your work. They will be applied to your work to help you understand what you have accomplished, how any mark given was arrived at, and how you can improve your work in future.

Marking Criteria

Criteria	Weighting	100-80%	79-70%	69-60%	59-50%	49-40%	39-30%	29-0%
Subject Knowledge Understanding and application of subject knowledge. Contribution to subject debate.	10%	Shows sustained breadth, accuracy and detail in understanding key aspects of subject. Contributes to subject debate. Awareness of ambiguities and limitations of knowledge.	Shows breadth, accuracy and detail in understanding key aspects of subject. Contributes to subject debate. Some awareness of ambiguities and limitations of knowledge.	Accurate and extensive understanding of key aspects of subject. Evidence of coherent knowledge.	Accurate understanding of key aspects of subject. Evidence of coherent knowledge.	Understanding of key aspects of subject. Some evidence of coherent knowledge	Some evidence of superficial understanding of subject. Inaccuracies.	Little or no evidence of understanding of subject. Inaccuracies.
Testing and Problem-Solving Skills Design, implementation, testing and analysis of product/process/system/idea/solution(s) to practical or theoretical questions or problems.	20%	Autonomous creation and novel implementation. Adapts to unforeseen practical and theoretical challenges to achieve identified goals.	Almost entirely autonomous creation and implementation. Adapts to unforeseen practical and theoretical challenges to achieve identified goals.	Mainly autonomous creation and implementation. Adapts to unforeseen practical and theoretical challenges to achieve identified goals.	Some autonomy to create and implement. Some adaption made to unforeseen practical and theoretical challenges to achieve identified goals.	Exploration of possible solution(s). Use of established approaches to resolve practical and theoretical problems.	Limited exploration of possible solution(s) using established approaches to resolve practical and theoretical problems.	Little or no exploration of solution(s). Question or problem unresolved.
Experimentation Problem-solving and exploration of the assessment including risk taking and testing of ideas.	30%	Evidence of rich, in-depth exploration and highly innovative problem-solving. Responds to assessment in a way that pushes boundaries and/or makes exciting new connections possible within discipline. Finds a balance between experimentation and resolution to create a highly coherent and innovative piece of work.	Evidence of in-depth exploration and innovative problem-solving. Responds to assessment in a way that pushes boundaries and/or makes new connections possible within discipline. Finds a balance between experimentation and resolution to create a coherent and innovative piece of work.	Clear evidence of exploration and experimentation. The work may not find the balance between experimentation and the attempt to create a coherent piece of work.	There is some problem-solving evident but lacks creative exploration and experimentation	Some evidence of limited problem solving. Appropriate response to assessment but there is little development and exploration evident.	Some evidence of engagement with assessment but it is insufficiently developed. Unable to engage with the challenges of assessment.	Little or no evidence of engagement with the assessment, it shows little development.
Practical Competence Skills to apply theory to practice or to test theory	30%	Consistently applies comprehensive practical and/or technical skills which reflect the correct application of theory to context.	Applies comprehensive practical and/or technical skills which reflect the correct application of theory to context.	Applies practical and/or technical skills very well and with careful application of theory to context.	Applies practical and/or technical skills correctly and with some application of theory to context with only minimal errors.	Applies a sufficient level of practical and/or technical skills but with some errors. Limited application of theory to context.	Applies some aspects of the practical and/or technical skills but these are incomplete or contain important errors. Very limited application of theory to context.	Limited application of some aspects of the practical and/or technical skills. No application of theory to context.

Academic Integrity[1] Acknowledges and gives credit to the work of others follows the conventions and practices of the discipline including appropriate use of referencing standards for discipline.	10%	Consistent, error free application of relevant referencing conventions with great attention to detail.	Consistent, error free application of relevant referencing conventions.	Consistent application of relevant referencing conventions with few errors.	Application of relevant referencing conventions, with some errors and / or inconsistencies.	Generally correct application of relevant referencing conventions, with some errors and / or inconsistencies.	Limited application of referencing conventions and / or errors.	Very limited or no application of referencing conventions, and/or multiple errors.

[1]The application of this criterion is independent of the process outlined in the Student Academic Misconduct Procedure.

How to get help

Student would be required to attend a coursework specification discussion session on Wednesday 6ᵗʰ March 2024 during class session. However, if you have any related questions, please feel free to contact your Module Leader via MS Teams or email on alkanb@lsbu.ac.uk as soon as possible.

Clinic Sessions will be announced via VLE.

Note that clinic sessions will be based "on a first come, first served basis".

Resources & Additional guidance

All the module's lectures, tutorial handouts, and the references recommended in the module guide.

Please use Google Scholar to search for research papers on your chosen topic to understand the academic writing style. The electronic journal publication resources, those are not open access may require you to use LSBU credentials for institutional access.

Please visit:
https://libguides.lsbu.ac.uk/LSBU-Library-and-Learning-Resources/Home
to learn more on how to access research E-resources and publications.

Quality assurance of coursework specifications

Coursework specifications within CSI division go through internal (for new modules with 100% coursework also through external) moderation. This is to ensure high quality, consistency, and appropriateness of the coursework as well as to share best practice within the CSI division.