EST 1892 LSBU

Coursework Specification

CW_Specification_CSI_6_ARI_24-25

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	**25/02/2025**
Submission Method	Submit online via this Module's Moodle site
Submission Deadline	**04/04/2024 at 18:00***
Release of Feedback & Marks	Feedback and provisional marks will be available in the Gradebook on Moodle from **02/05/2025**

* 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 Artificial Intelligence (AI) solutions.

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**, to carry out an industrial machine learning project. **Project Background:** The manufacturing industry has experienced a revolutionary shift with the rise of Industry 4.0, where advanced technologies like the Internet of Things (IoT), Artificial Intelligence (AI), and Cyber-Physical Systems converge to enable smart manufacturing systems. These systems facilitate real-time monitoring, predictive maintenance, and quality control, leading to higher efficiency and reduced operational costs. Among the various manufacturing processes, plastic injection moulding is particularly significant due to its versatility in producing high-quality, complex plastic components on a large scale. Ensuring consistent product quality in plastic injection moulding is challenging because numerous factors, such as melt temperature, injection pressure, cooling time, and mould temperature, significantly affect the final product. Traditional quality control methods, which rely on post-production inspections, are often time-consuming, costly, and prone to human error. By leveraging Machine Learning (ML) techniques, manufacturers can predict potential defects during the production process, enabling real-time process adjustments that prevent defective products from reaching the end of the production line. This coursework aims to predict the **probability of defects** in **plastic injection moulding** using advanced machine learning models. You will work with real-world manufacturing data, perform in-depth statistical analysis, develop robust predictive models, and build an interactive dashboard to present your findings. You will be provided with a CSV dataset containing information related to the plastic injection moulding process. Each row in the dataset represents a production cycle and includes process parameters such as mould temperature, injection pressure, cycle time, and volume. The target variable, **'quality'**, is a multiclass label categorising the final product into one of four classes: 1. **Waste**: Product fails to meet basic standards and must be scrapped. 2. **Acceptable**: Product meets minimum quality standards but is not ideal. 3. **Target**: Product meets the desired quality specifications. 4. **Inefficient**: Product is above acceptable but falls short of target quality due to process inefficiencies. The objective of this coursework is to **classify the quality class** of the products using advanced machine learning models. Identify which process parameters most significantly influence the quality classification. Develop models that assist in optimising manufacturing conditions to increase the production of target-quality parts while reducing waste and inefficiencies. Through this project, students will gain hands-on experience in data pre-processing,

	exploratory data analysis (EDA), feature engineering, statistical modelling, machine learning, and interactive dashboard development. **Dataset** Download your dataset from the module's VLE site via this **link**. The dataset (unique to each student) consists of 1000 samples. You should download and use the dataset corresponding to your student ID number. For example, if your ID number is '4022820, then you need to use CW_Dataset_4022820. **Dataset Reference: Polenta, A., Tomassini, S., Falcionelli, N., Contardo, P., Dragoni, A. F., & Sernani, P. (2022).** *A Comparison of Machine Learning Techniques for the Quality Classification of Molded Products*. Information, 13(6), 272. https://doi.org/10.3390/info13060272
Tasks	You are required to undertake the following tasks: **1. Problem Identification** • Learn about the basic characteristics of the dataset such as the total number of attributes, 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. **Important**: Five randomly selected classification algorithms (ANN and RF are compulsory for all students) are **assigned** to each student. You must **only use** the assigned algorithms to carry out your ML project. Train/test splits and k values for cross-validation are **also randomised**. Students **must use the parameters provided** to them. Please see this **link** for further info. **2. Data Preparation and Analyses** • Identify which variables to be used in which analysis using exploratory data analysis. • Choose appropriate methods for data pre-processing, including handling outliers, encoding categorical variables – if any, data standardisation and normalisation (e.g., data scaling, etc.,), and train/validation dataset splitting. • Apply statistical methods such as **Analysis of Variance (ANOVA) (**or an appropriate method**)** to validate relationships between process parameters and product quality classes. **3. Model Development** • Using the pre-processed dataset, develop **the machine learning models assigned to you** to predict the four quality classes. See your assigned algorithms using **this link**. • Implement **k-fold cross-validation** (k = 3, 5, or 10 as assigned) for robust model evaluation. • Select the appropriate Key Performance Indicators (KPIs) for evaluating the models' efficacy and provide an in-depth explanation of how they will be used. • **Perform hyperparameter tuning using GridSearchCV** to optimise model performance.

	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 be used to address the original business concerns. • Suggest **next steps for the company** to take in developing this into a viable business option. **5. Develop an interactive dashboard (using Streamlit, Dash, or an equivalent tool) to:** • Allow users to **input process parameters**. • Predict **product quality classes** using the trained models (use only the best model found). • Add displays such as overall production quality, product scrap rates, etc., • Visualise feature importance, confusion matrices, class-wise evaluation metrics, and ANOVA plots. 6. **Compile a final report** and **record a 5-minute video presentation** explaining the methodology, results, and dashboard functionality.
Word Count:	As a guide, **aim** for **4,000 words**, excluding Title page, 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 Jupyter Notebook 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. **Abstract** Abstract provides a summary of your work. **Section 1. Data Preprocessing & EDA (15%)**

- **Data Cleaning:** Identification and handling of missing values, outliers, and data inconsistencies.
- **Feature Engineering:** Creation of new features that improve model accuracy.
- **Visualisations:** High-quality, clearly labelled charts and graphs showing data distributions and feature relationships.
- **Insight Generation:** Interpretation of patterns and potential impacts on quality classes.

Section 2. Hypothesis Testing & ANOVA (15%)

- **Hypothesis Formulation:** Clear, testable statements linking process parameters to quality classes.
- **Statistical Test Application:** Correct use of hypothesis tests and ANOVA with appropriate assumptions considered.
- **Interpretation:** Meaningful insights derived from statistical analyses.

Section 3. Machine Learning Model Development (40%)

- **Model Selection:** Inclusion of five appropriate models with clear rationale.
- **Cross-Validation:** Correct implementation of k-fold validation with well-justified k values.
- **GridSearchCV Usage:** Systematic exploration of hyperparameters with performance improvements demonstrated.
- **Implementation:** Correct and well-documented code for multiclass classification.
- **Evaluation Metrics:** Use of appropriate metrics (Accuracy, Precision, Recall, F1-Score, ROC-AUC for each class).
- **Feature Importance:** Insightful analysis showing which features most affect each quality class.

Section 4. Interactive Dashboard Development (20%)

- **Functionality:** Fully operational dashboard allowing for user interaction and real-time predictions.
- **User Experience:** Clear navigation, responsive design, and intuitive interface.
- **Visualisation Integration:** Well-presented plots and tables showing model results and feature importance.

References
AI Statement

AI Statement Requirements
- Every submission must include an AI statement covering:
- Whether AI tools were used during any part of the coursework.
- How these tools were utilised (e.g., for code suggestions, data analysis, or report generation).
- A declaration confirming the student's understanding and responsibility for any AI-assisted content.

<table>
<tr><td></td><td>

- AI Statement can be found in this **link.**

Notes for Final Report:

- **Structure & Clarity:** Logical organisation with smooth flow between sections.
- **Methodology Explanation:** Detailed and well-explained approaches with justifications.
- **Result Interpretation:** Thorough discussion of model outcomes and recommendations.
- **References:** Proper citation of all sources in an appropriate format.

</td></tr>
<tr><td>How are marks awarded:</td><td>

Only 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.

Mark Distribution

	Weight (%)
Data Pre-processing & EDA	15%
Hypothesis Testing & ANOVA	15%
Machine Learning Model Development	30%
K-Fold Cross-Validation & Hyperparameter Tuning	10%
Interactive Dashboard Development	20%
Academic Integrity	10%

</td></tr>
<tr><td>Presentation:</td><td>

In total you must submit three files: *i*) your technical report work (.pdf), *ii*) the source code (Jupyter notebook file), and *iii*) a 5-mins 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_2425_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_2425_CWCode_<Student_ID>.*<extension>*
(Please replace <Student_ID> with your student number)

The video recording:

</td></tr>
</table>

	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/dashboard 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 show your ID at the start of your video). 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_2425_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

📊 Mark Distribution

	Weight (%)
Data Pre-processing & EDA	15%
Hypothesis Testing & ANOVA	15%
Machine Learning Model Development	30%
K-Fold Cross-Validation & Hyperparameter Tuning	10%
Interactive Dashboard Development	20%
Academic Integrity	10%

📝 Rubric Key Levels

Each section will be evaluated on a **five-level scale**, with corresponding marks allocated proportionally based on the section weight:

Level	Description	Percentage Range	Mark Allocation (out of 100)
5 – Excellent	Outstanding work exceeding all expectations.	85% - 100%	Full marks for the section
4 – Very Good	Strong work with minor improvements needed.	70% - 84%	80% - 84% of section marks
3 – Good	Solid work meeting key requirements.	55% - 69%	65% - 69% of section marks
2 – Satisfactory	Adequate but with notable shortcomings.	40% - 54%	50% - 54% of section marks
1 – Needs Improvement	Insufficient work with significant issues.	0% - 39%	Less than 50% of section marks

📝 Data Pre-processing & EDA (15%)

Subtask	5 – Excellent	4 – Very Good	3 – Good	2 – Satisfactory	0 – Needs Improvement
Data Cleaning (5%)	All issues handled with clear justification.	Mostly complete with minor omissions.	Basic cleaning with some issues.	Minimal cleaning with significant gaps.	No meaningful cleaning performed.
Feature Engineering (5%)	Creative and relevant new features added.	Relevant features with some improvements.	Basic features with limited relevance.	Simple features with weak justification.	No feature engineering attempted.
Visualisations & Insights (5%)	High-quality visuals with deep insights.	Clear visuals with good interpretations.	Basic visuals and surface-level insights.	Poor visuals with unclear interpretations.	No visuals or insights provided.

📊 Hypothesis Testing & ANOVA (15%)

Subtask	5 – Excellent	4 – Very Good	3 – Good	2 – Satisfactory	0 – Needs Improvement
Hypothesis Formulation (5%)	Clear, relevant, and testable hypotheses.	Mostly clear with minor gaps.	Basic hypotheses with limited depth.	Vague or partially relevant hypotheses.	No hypotheses formulated.

Test Application (5%)	Correct and well-justified test selection.	Appropriate tests with minor issues.	Basic tests with some errors.	Misapplication of tests.	No tests applied.
Result Interpretation (5%)	Insightful conclusions with supporting data.	Good interpretation with minor gaps.	Basic conclusions with limited analysis.	Superficial or unclear interpretations.	No interpretation provided.

🤖 Machine Learning Model Development (30%)

Subtask	5 – Excellent	4 – Very Good	3 – Good	2 – Satisfactory	0 – Needs Improvement
Implementation (15%)	Clean, efficient, and well-documented code.	Functional code with minor issues.	Working code but lacks clarity.	Incomplete or poorly structured code.	Code not functional or missing.
Evaluation Metrics (10%)	Comprehensive use with clear analysis.	Adequate metrics with slight oversights.	Basic metrics without deep interpretation.	Limited metrics with errors.	No metrics reported.
Feature Importance (5%)	Insightful analysis with strong conclusions.	Good analysis with minor omissions.	Basic analysis with limited insights.	Weak or unclear analysis.	No analysis provided.

🔄 K-Fold Cross-Validation & Hyperparameter Tuning (10%)

Subtask	5 – Excellent	4 – Very Good	3 – Good	2 – Satisfactory	0– Needs Improvement
Cross-Validation (5%)	Correct k-fold with strong rationale.	Mostly correct with minor issues.	Basic application with some mistakes.	Minimal use without justification.	No cross-validation used.
Hyperparameter Tuning (5%)	Effective tuning improving performance.	Reasonable tuning with slight gaps.	Basic tuning with marginal gains.	Limited tuning attempts.	No tuning applied.

🖥️ Interactive Dashboard Development (20%)

Subtask	5 – Excellent	4 – Very Good	3 – Good	2 – Satisfactory	0 – Needs Improvement
Functionality (10%)	Fully functional with all features.	Mostly complete with minor bugs.	Basic dashboard missing some features.	Limited functionality with issues.	No dashboard or non-functional.
User Experience (5%)	Intuitive and user-friendly interface.	Good design with slight usability issues.	Basic interface with usability gaps.	Poor design affecting usability.	No consideration for user experience.
Model Integration (5%)	Seamless integration of models and outputs.	Models integrated with minor issues.	Basic integration with errors.	Limited or flawed integration.	No integration attempted.

📚 Academic Writing & References (10%)

Subtask	5 – Excellent	4 – Very Good	3 – Good	2 – Satisfactory	0 – Needs Improvement
Academic Writing (5%)	Clear, well-structured, and professionally written with excellent grammar and flow.	Mostly clear with minor grammatical issues.	Adequate writing with noticeable errors.	Basic structure with significant errors.	Poor writing with major issues.
References (5%)	Properly formatted, comprehensive references covering all sources.	Mostly correct with slight formatting errors.	Basic referencing with several omissions.	Few references with incorrect formatting.	No references or serious citation issues.

How to get help

Student would be required to attend a coursework specification discussion session on Wednesday 26th February 2025 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.