A picture containing shape

Description automatically generated

# Coursework Specification

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

|  |  |
| --- | --- |
| **Module Code** | CSI\_7\_PPA\_2425 |
| **Module Title** | Python programming for AI and Visualisation |
| **Lecturer** | Dr Ali Salimian |
| **% of Module Mark** | 100% |
| **Distributed** | 01/Nov/2024 |
| **Submission Method** | Submit online via this Module’s Moodle site |
| **Submission Deadline** | 9th/Dec/2024 for code submission  Presentations 16th Dec 2024 |
| **Release of Feedback & Marks** | Feedback and provisional marks will be available in the Gradebook on Moodle |

## Coursework Aim:

You will complete this course work while being part of a team of 2 or 3. You will be working directly with your team mates to deliver on the requirements of the coursework. You will be submitting a file containing the complete coding you and your team have carried out. Your code can be identical to your teammates BUT not to the other teams. However, during the presentation part, you will be required to present your code and discuss the steps you carried out to complete the tasks.

**Part A\_1 (50%)**

Coding and submission of the code (All team members will share the same mark from this section).

You will be provided with a Data set associated with chemical compositions of volcanic rocks.

**You are NOT required to learn or know much about the physics of the data**, however, you will be required to use the data to produce specific visualisations, perform exploration of data distribution, and demonstrate your skills to implement some machine learning models to perform classifications, given the data you have.

The code should be structured into four segments:

1. Data preprocessing
2. Visualisation of the data
3. ML Classification (two techniques)
4. DL classification

**Part B\_ (50%)**

Presentation of your work (max 20 minutes)

* Part B\_1 (10%):

Group/ Team efficiency

This will be evaluated on assessing your team’s overall performance and collaboration of the team members toward delivering the objectives of the coursework. It is expected that all team members provide a similar level of competence in presenting their task, keeping up with the time allowed and organisation of the team during their presentation. (All team members will share the same mark from this section)

* Part B\_2 (40%):

Presentation, you will be required to present your submitted work. The presentation should be structured as follows:

1. Presenting the code associated with data import and preprocessing. (5 min)
2. Presenting the code associated with visualisation. (5 min)
3. Presenting the code on ML model: (5 min)
4. Presenting the code on DL model: (5 min)

**Each team member should be able to present the entire presentation prepared by their Team.**

**How ever on the day of your team’s presentation, each member will be selected randomly by the examiner to present either of the 4 sections of the presentation.**

(Team members will be marked individually for this section)

## Coursework Details:

|  |  |
| --- | --- |
| **Type:** | Code |
| **Word Count:** | Complete Python code file submission is required.  You can add all your comments in the notebook file you submit. |
| **Presentation:** | * A presentation of the code is required. |
| **Referencing:** | NA |
| **Regulations:** | Make sure you understand the [University Regulations](http://www.lsbu.ac.uk/__data/assets/pdf_file/0008/84347/academic-regulations.pdf) on expected academic practice and academic misconduct.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 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. |

## Learning Outcomes

This coursework will fully or partially assess the following learning outcomes for this module.

You will demonstrate your ability to code in python and present your code.

You will demonstrate your ability to assess a classification model’s performance.

You will demonstrate your ability to utilise SciKit learn to implement Machine learning models and assess the performance of your models via correct metrics.

You will develop your presentation skills and enhance your teamwork capacity.

## 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.

The Course work in total requires you to complete 2 tasks:

A: coding (30 marks)

1. Correct commenting on your code (5 marks)
2. Data Import (5marks)
3. Visualising the data (5 marks)
4. Regression analysis (5 marks)
5. Classification (5 marks)
6. Model performance evaluation (5 marks)

B\_1: Team efficiency (10 marks)

1. Time keeping by all members (5 marks)
2. Equal competence among team members (5 marks)

B\_2: Presentation (60 marks)

1. Time keeping (5 marks)
2. Quality of engaged presentation (5 marks)
3. Correct description of the code and approach toward coding (50 marks)

## Total: 100 Marks

## How to get help

We will discuss this Coursework Specification in class. However, if you have related questions, please contact me [salimiaa@lsbu.ac.uk] as soon as possible.

## Resources

The Files required for this coursework associated will be supplied to you separately.

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.

Details of the moderators for this coursework specification are below:

|  |  |
| --- | --- |
| **Moderated (internal)** | [Name, date] |
| **Moderated (CSI lead)** | [Name, date] |
| **Signed off by (HoD/DHoD)** | [Name, date] |

----------------For Internal use by CSI lead only-----------------

|  |  |
| --- | --- |
| **Changes required to CW?** | Yes, No \* |
| **Examples of good practice** |  |

**\* if changes are required, moderator to complete the below:**

|  |  |
| --- | --- |
| **List of changes required** | [These needs to be met before signoff can be achieved] |
| **ML Response** | [ML response, date] |
| **Moderator Response** | [ML response, date] |

**Background:**

This coursework involves analysing geochemical data obtained from volcanic rock samples, focusing on three distinct locations within a crystal structure in the rocks. The data has been generated using two different analytical techniques:

1. **Electron Microprobe Analysis (EMPA)** - Quantifies specific elements, including oxides and compounds.
2. **Laser Ablation Spectroscopy** - Provides a broader spectrum of elemental analysis.

The objective is to classify the age of the volcanic rocks based on the data from these techniques. EMPA data is complete for all samples, but some samples lack Laser Ablation results. The measurements have been taken in various labs across the globe, adding variability to the dataset.

**Objectives:**

1. **Data Preprocessing**:
   * Handle missing values (particularly in the Laser Ablation dataset).
   * Perform data cleaning and normalization as necessary.
   * Decide on imputation strategies for missing data in the Laser Ablation results.
2. **Data Handling**:
   * Create and manage data frames in Python.
   * Explore data filtering techniques to separate EMPA and Laser Ablation data.
   * Structure the data for machine learning workflows, choosing appropriate data columns for analysis.
3. **Feature Selection**:
   * Decide on feature selection strategies to determine which columns from EMPA and Laser Ablation datasets are relevant for classification.
   * Analyze correlations among elements and assess their impact on classification accuracy.
4. **Machine Learning and Deep Learning**:
   * **Classification Task**: Develop models to classify the age of the volcanic rocks.
   * Use **separate models for EMPA** and **Laser Ablation data** to compare their performance.
   * Experiment with multiple machine learning algorithms and at least one deep learning model.
   * Evaluate model performance, focusing on accuracy, precision, recall, and F1-score.
5. **Data Visualization**:
   * Produce a variety of visualizations to analyze and present findings.
   * Suggested visualizations:
     + **Bar Charts**: To show distributions of key elements across samples.
     + **Pie Charts**: To illustrate proportions of various elements.
     + **Scatter Plots**: To explore correlations between pairs of elements.
     + **Heatmaps**: To visualize correlations between elements.
   * Create visualizations that compare the EMPA and Laser Ablation data distributions.

**Dataset Description:**

* **EMPA Elements**:
  + NiO, F, CaO, SiO₂, Cr₂O₃, Na₂O, TiO₂, V₂O₃, MnO, MgO, ZnO, FeO, Al₂O₃, K₂O, Li₂O
* **Laser Ablation Elements**:
  + Mg#, Li, Be, B, Mg, Si, Ca, Ca, Sc, Ti, V, Cr, Mn, Co, Ni, Cu, Rb, Sr, Y, Zr, Nb, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, Pb, Th, U

**Submission Requirements:**

1. **Data Preprocessing Notebook**:
   * Document the data preprocessing steps, including handling missing values, normalizing data, and feature selection processes.
2. **Model Development Notebook**:
   * Develop and train machine learning and deep learning models for the classification task.
   * Report on the choice of models, hyperparameters, and evaluation metrics.
3. **Visualization Report**:
   * Include the data visualizations with interpretations of key insights.
   * Highlight comparisons between EMPA and Laser Ablation datasets.
4. **Summary Report**:
   * A brief summary of the analysis, methods used, challenges faced, and final conclusions.
   * Discuss the comparative efficacy of EMPA and Laser Ablation data in classifying volcanic rock age.

This coursework will enable students to practice essential data science skills, including data preprocessing, feature engineering, machine learning modeling, and data visualization, while gaining hands-on experience in handling real-world geological data.

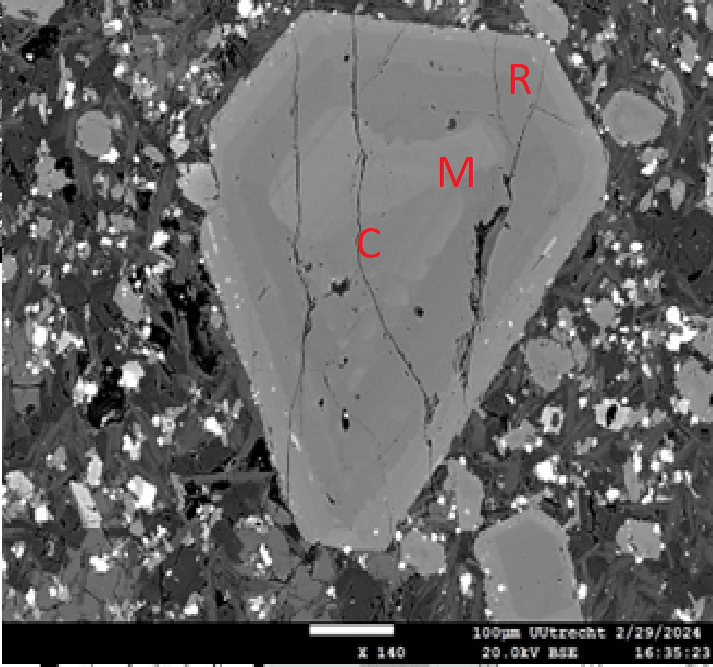


Figure 1: The locations of a single crystal that were analysed.

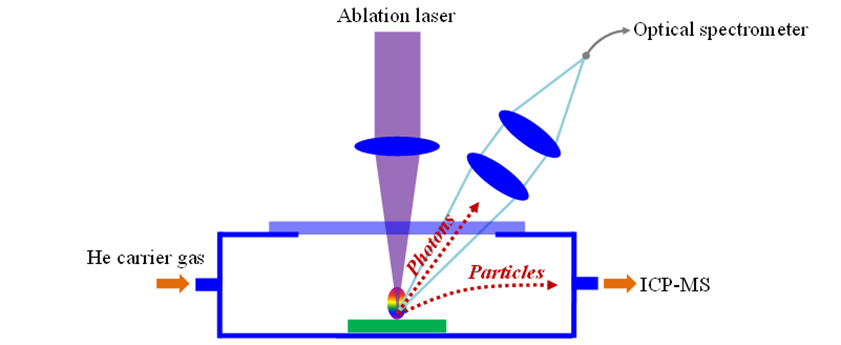


Figure 2. Laser Ablation method of data collection.

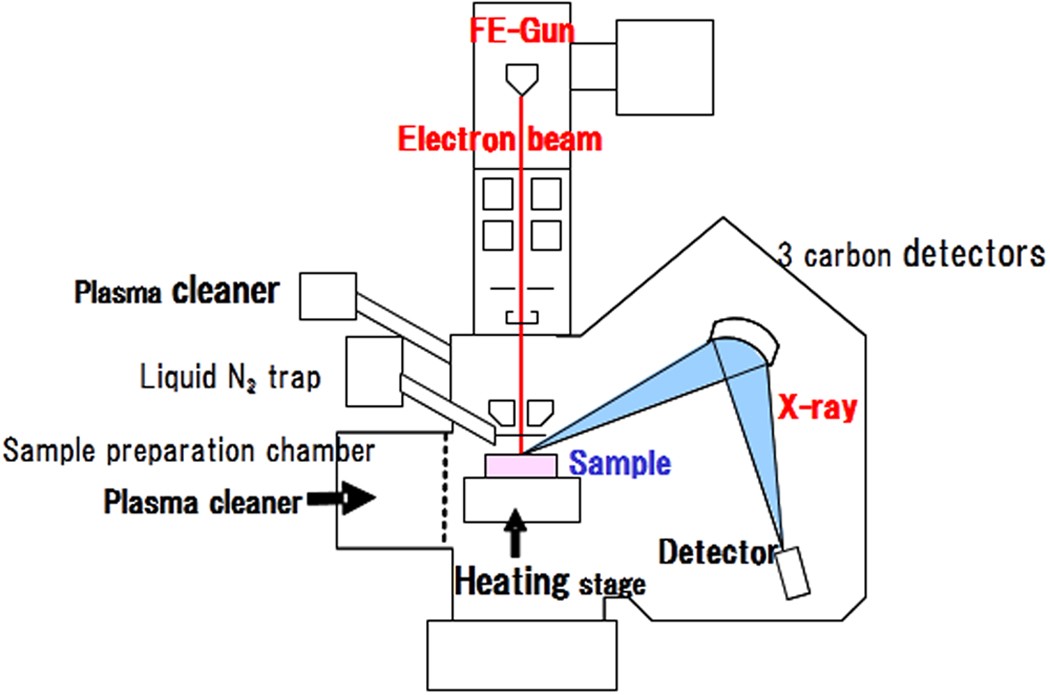


Figure 3. EMPA method of data collection.