# DS203-2025-S1: Exercise – 7 (Project)

- This Exercise (Project) carries 30 marks and 30% weightage
- It should preferably be done in groups
    - A group can have maximum 4 members
    - Team of '1' is strongly discouraged, but not barred.
    - **Identify one member of the team as the Group's 'representative'**
    - **The group representative should register the group using the link alongside**
- Read the instructions and the evaluation criteria carefully.
- Note and adhere to the submission requirements and deadlines carefully.
- **Submissions due by: November 07, 2025, 11:55pm**

## Evaluation Criteria

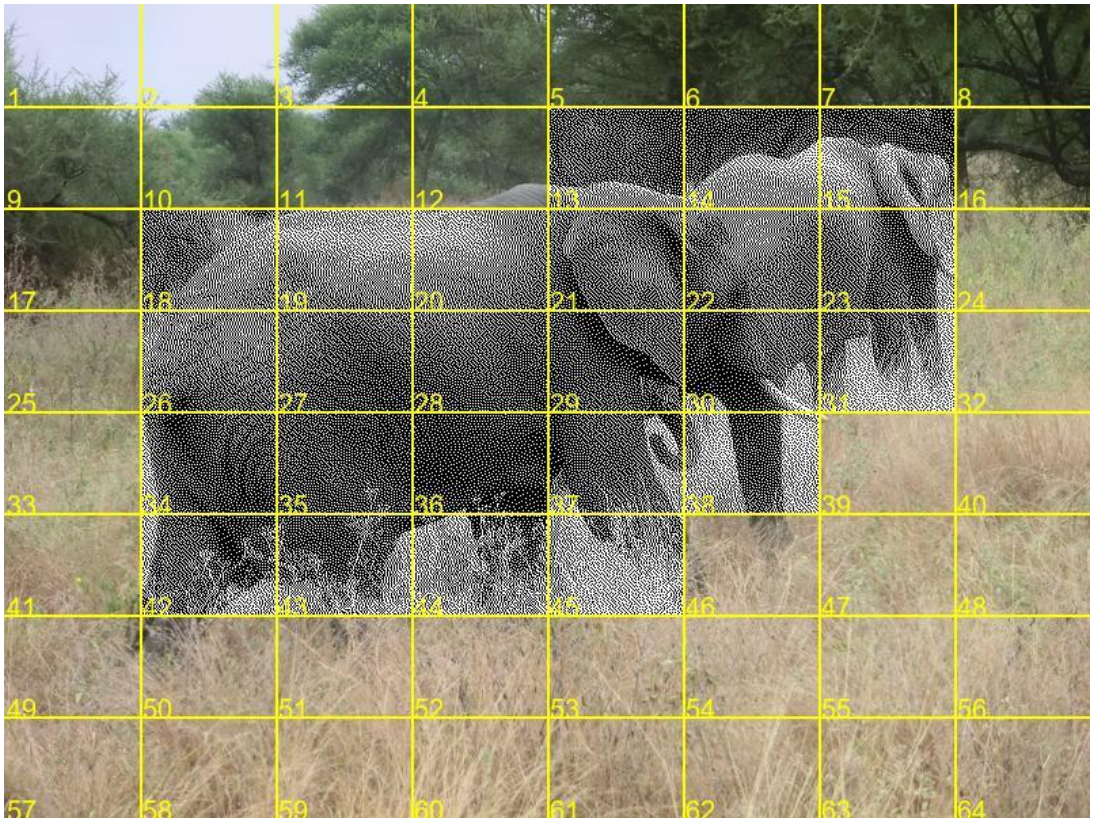
| Evaluation Criteria | Marks |
|---|---|
| **Process and Results**<br>• Feature Engineering / Feature Creation in terms of relevance to the problem<br>• Steps of Data Science: Have they been correctly applied, and backed up with proper metrics / reasons / explanations?<br>• Completeness and correctness of the results achieved<br><br>(Equal weightage will be given to all the above aspects) | 15 |
| **Documentation (Presentation and video)**<br>• Completeness and preciseness in the documentation of the process and results [5 marks]<br>• Design, formatting and readability of the slides. [3 marks]<br>• Documentation of 'Executive Summary', successes, failures, hurdles and learnings [2 marks]<br>• **Video (maximum 5 minutes) covering all aspects of the project: [5 marks]** | 15 |
| **Penalty**<br>• **If the presentation contains raw and hyped-up outputs generated using LLM tools like ChatGPT / Gemini etc.**<br>• **If a project is seen to be 'copied' from another team, both teams will get ZERO marks.** | **-10** |
| • Viva will be conducted, if deemed necessary, to ascertain originality of work, and to ascertain contribution by team members.<br>• The project will be evaluated **ONLY** by reviewing the video and presentation. Source code / Jupyter Notebook will **NOT** be reviewed to understand your work; they will be only used to *verify* the claims you have made in the presentation. Therefore, **if you forget to mention some part of your work / analysis in your presentation, it will be concluded that you have not done it**!<br>• If you do not submit your source and data files, your project becomes unverifiable and the submission will not be given any credit. | |

Follow the '**Reporting and Presentation Guidelines'** outlined later in this document

## Problem Description

Using the data contained in https://tinyurl.com/E7-IMAGES-DS203-2025-S1 machine learning models have to be created to identify regions that contain animals, birds, and insects, as outlined and illustrated below:

1. Before training, the images should be converted into 800 x 600 pixels, without distortion.
    a. You should only use those images that have 4:3 aspect ratio, or crop the images appropriately.
    b. Images larger than 800 x 600 should be scaled down to 800 x 600.
    **c. But, do not scale-up images smaller than 800 x 600**
2. The 800 x 600 image should be **conceptually** divided into an 8 x 8 grid and the grid cells numbered as shown in the image below
3. Using appropriate **hand-crafted** features, create ML model(s) to be able to identify grid cells as:
    a. 1 – 'contains wild-life
    b. 0 – 'does not contain wild-life'
4. Once the ML models are trained and tested, selected, and **saved into a pickle file** - any given set of images should be processed to generate the output outlined and shown below:
    a. Cells containing any wild-life should be highlighted so that they can be easily visually differentiated
    b. An output **CSV** file should be created with the following data: **ImageFileName**, columns **c01** through **c64**, with values being either **0** or **1** depending on whether the corresponding grid cell is flagged as 'does not contain wild-life' or 'contains wild-life'




## Dataset of images to be used:

- **https://tinyurl.com/E7-IMAGES-DS203-2025-S1**

All teams are expected to diligently follow Data Science steps while creating the ML models. **You are required to create a well-organized, complete yet non-verbose PRESENTATION that nicely captures EVERYTHING that is done in the project - including failures and challenges encountered**. **And yes, a crisp 5 minute video needs to be created, with contributions from all the team members!**

---

## Solution guidelines

- Doing justice to this project involves much work, including thoroughly researching into feature creation methods for image data.
- Do what it takes to submit well researched, well-designed, and well-reasoned solutions to the above problems – and effectively communicate them!
- Re-visit all major aspects of Data Science that you have learned so far, and check if they can be / need to be used to solve the posed problems.
- It will really help to work in a team, and divide work amongst the members – to do a good job.
- It is important to have an initial **overall solution design** in place to guide your steps, distribute work among team members, and ensure that you are not overwhelmed, and that you don't get lost!

## Reporting and Presentation Guidelines

1. As often mentioned, succinctly communicating your work and results is a very important part of the Data Science process. The most important submission are the presentation and a 5 minute movie – that summarizes your approach, work, results, achievements, learnings, and possibilities. Budget adequate time for this activity and design your presentation well; last minute work will invariably be shoddy.
2. **Include the names and roll numbers of all group members in the title slide. No credit will be given to members who are not mentioned in the title slide!**
3. Provide an executive overview (1-2 slides) at the start of the presentation.
4. DO NOT use verbose paragraphs, or storytelling, to explain your steps, observations, results, and recommendations. All these should be presented precisely and point-wise.
5. Summarize your observations and results using charts / Tables / metrics and explain them and draw conclusions from them. Merely including plots / Tables, with no associated comment, is not acceptable.
6. Slides should be well designed. Use as many slides as required to completely convey your work.
7. If you do not include something in your presentation, it will be deemed that you have not done it. **Source code / Notebooks will NOT be reviewed to *understand* your work**. But they may be checked to verify your claims.
8. Towards the end of the presentation, include slides that clearly answer all the questions posed in the ***Evaluation Criteria*** table. In addition to focussing the evaluator's attention, this will also ensure that you have covered all the expected points in the presentation.
9. Finally, include a slide or two outlining your learnings from this project, and your experiences and hurdles while doing the project.
10. Apply similar guidelines while creating the summary video.

## Submission Guidelines

1. All source code files, results consisting of images and the output CSV file, the final presentation and the final video will have to be submitted.
2. The submission process will be shared closer to the submission date.

## Queries and Clarifications

- Questions should be posted to Moodle discussion forum: **E7 – Project Queries**

oooOOOooo