Learning Outcome

The Machine Learning project is a venue for students to achieve the learning outcomes below:

- LO1. Design and evaluate informed search algorithms and knowledge representations for problem solving
- LO2. Collaboratively build systems that consider a number of paths or strategies in order to improve its performance in achieving its goal in less amount of computing time, or by some other metric of performance
- LO4. Perform comparative analyses of algorithms for problem space search and Machine Learning using real-world datasets
- LO5. Articulate ideas and present results in correct technical written and oral English

The goal of this project is for the student to be able to perform a Machine Learning analysis given a specific data set. The student should be able to identify patterns, questions from the data set, and perform specific operations that will help them understand the data set. They should also be able to analyze the results. In this approach the student should be able to use existing AI/ML techniques in coming up with a scientific inquiry.

Instructions

1. Form a group composed with maximum of 4 members.
2. Download Anaconda from https://www.anaconda.com/
3. Go to Kaggle.com and create an account.
4. Find an existing benchmark dataset of at least 1,000 samples from the site and download it. The dataset must be suitable for use in classifying the (some) samples in just 2 classes. If you are currently working on specific Machine Learning technique, choose something else (for your own scientific growth).
5. Identify a problem statement or question that can be solved or answered by analyzing the chosen dataset.
6. Perform Machine Learning tasks so that an analysis can be derived from the said data set. Scikit-learn in Python will be used to perform the Machine Learning tasks.
7. Compare their Confusion Matrices, Accuracy Scores, Precision and Recall.
8. Write a paper following the Springer format. The template for the Springer format can be found in the latter part of this specifications.

Final Deliverables

The Machine Learning project has two components – a paper and a presentation. Prepare a paper in Springer format containing the following:

1. **Introduction**
    - Introduce the problem statement/question.

- You may start with defining/describing the domain of the problem/question. For example, the problem statement is recognizing the emotion of a person. You may start with defining emotions, what are the different emotions, and how are emotions expressed. You may also give some statistics related to the topic.
- After the introduction of the domain, discuss the motivation. Why the problem or question must be solved/answered.
- End the introduction by formally stating the problem statement and the approach. Then, enumerate or explain the possible benefits if the problem is solved.

2. <Name> Dataset

- Describe the dataset
- Show the demographics of the participant (if applicable)
- Enumerate and describe the features
- Enumerate and describe the labels
- Show the distribution of the classes

3. Methodology

- For the discussion of the methodology, follow the pipeline that you created.
- Before the discussion, include a figure of the pipeline from RapidMiner or create a flowchart based on the pipeline you created using Scikit-learn.
- Below the figure, discuss each step or process in the pipeline. You may create a subsection for each step or process. For each process:
  - What is the process?
  - What is the purpose of the process?
  - What is the input and output?
  - What are the hyperparameters used?

4. Results and Analysis

- Discuss the performances.
- You can have 3 subsections for this:
  - First subsection for Decision Trees. Compare the Accuracy Scores, F-measures, Precisions and Recalls of the models crated given the hyperparameters used. Illustrate the comparison using a graph. Give some hypotheses why the performance is decreasing or increasing as you modify the hyperparameters. End with stating the optimal hyperparameters based on the experiments.

- Second subsection for Neural Networks. Do the same as the first subsection.
- Third subsection for the comparison and analysis of the results using the 2 classification techniques. Compare the Confusion Matrices, Accuracy Scores, F-measures, Precisions and Recalls of the best model of Decision Tree and Neural Network. Give some hypotheses why Decision Tree is better than Neural Network for the dataset, or vice versa.

5. Conclusions and Recommendations

- Summarize what you did in 1-2 sentences.
- Briefly discuss the best model and its hyperparameters.
- Give some recommendations to improve the performances.

6. Contributions

- Enumerate the contributions of each member.

7. References (Follow the APA format)

8. Appendix A. Contribution of Members

Name	Contributions

Notes:

- Formatting requirements:
    - Follow the Documentations Format shown in the next page.

Submission Policy:

- Submit the presentation slides and .pdf documentation of your Machine Learning project on **December 13, 2019** (2359).
- Submission within 1 hour after the deadline (i.e. December 14, 0000 - 0100) will incur a 25% deduction on the grade.
- Submission more than 1 hour after the deadline (i.e. starting December 14, 0101) will not be accepted.
- The presentation can be scheduled thru Canvas.
- Reused works are not allowed unless you are going to change the approach.
- **Plagiarized works will automatically be given a grade of 0.0 for the course.**

Documentation Format

Title

First Author[1][0000-1111-2222-3333] and Second Author[2][1111-2222-3333-4444]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany
lncs@springer.com

Abstract. The abstract should summarize the contents of the paper in short terms, i.e. 150-250 words.

Keywords: First Keyword, Second Keyword, Third Keyword.

1 First Section

1.1 A Subsection Sample

Please note that the first paragraph of a section or subsection is not indented. The first paragraphs that follows a table, figure, equation etc. does not have an indent, either.
Subsequent paragraphs, however, are indented.

Sample Heading (Third Level). **Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.**

Sample Heading (Forth Level). *The contribution should contain no more than four levels of headings. The following Table 1 gives a summary of all heading levels.*

Table 1. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Run-in Heading in Bold.** Text follows	10 point, bold
4th-level heading	*Lowest Level Heading.* Text follows	10 point, italic

Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

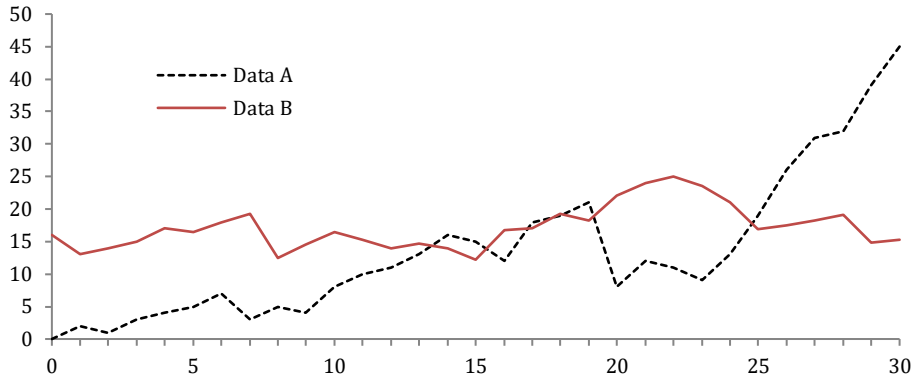

Fig. 1. A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro button chooses the correct format automatically.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1], an LNCS chapter [2], a book [3], proceedings without editors [4], as well as a URL [5].

References

1. Author, F. (2010). Contribution title. *9th International Proceddings on Proceedings* (pp. 1-2). Location: Publisher.
2. Author, F. (2016). Article title. *Joanna 2(5)*, 99-110.
3. Author, F., & Author, S. (2016). Title of a proceddings paper. In F. Editor, & S. Editor (Ed.), *Conference 2016, LNCS. 9999*, pp. 1-13. Heidelberg: Springer.
4. Author, F., Author, S., & Author, T. (1999). *Book title* (2nd ed.). Location: Publisher.
5. *LNCS Homepage*. (2016, 11 21). Retrieved from LNCS: https://www.springer.com/lncs

Appendix A. Contribution of Members

Table 2. Contribution of Members.

Name	Contributions
Member1	Contribution1, Contribution2
Member2	Contribution2
Member3	Contribution2
Member4	Contribution1