Review: Final Team Project Introduction

The Final Team Project for this course will assess the application of machine learning methods and programming techniques introduced in this course. You will be assigned to a group by your instructor. You and your teammate(s) will work closely to brainstorm a problem, find a dataset, clean and pre-process it, and perform a machine learning project on the data using Python. For this project, you will use Python in Google Colab, write a report, and prepare a recorded video presentation including visuals based on your report.

When you finish this project, you should have had a meaningful experience that you can describe in the future. If the project were something simple, like putting together a jigsaw puzzle or following a recipe exactly: every step is determined, you simply need to do the steps in order.

A better learning experience, just like a great story, will include a challenge to overcome, and the story of what you tried, in order to accomplish it. When thinking about a project, choose one which you will be able to explain well at the end, not just to your instructor, but to other people in the future. STAR (Situation, Task, Action, Result) is a format for answering interview questions:

- Situation
- Task
- Action
- Result

Instructions:

Datasets:

- You will be responsible for finding a dataset for this project. Several free data resources' links are provided within the **Resources** ⇲ **(https://usd-msads.github.io/resources/)** page on the MS-ADS GitHub website. Your dataset must have at least 10,000 data points.

Problem Definition:

- Include a clear statement describing the initial state of a problem by answering the following questions:
  - What is the problem you are going to solve?
  - Why does the problem need to be solved?
  - What aspect of the problem will a machine/deep learning model solve?
- Ensure your problem is interesting and understandable.

EDA and Pre-Processing:

- Include a clear discussion that ensures all steps are clearly explained and addresses the following:
  - How did you make sure that you are ready to apply machine/deep learning models?
  - How did you pick the best model?
  - How can you define and refine various feature variables that you may potentially use for the modeling?
  - Handling missing values or human error (if needed).
  - Identifying outliers.
- Include figures and tables that are highly insightful and are carefully tailored to the project tasks.

Modeling Methods, Validation, and Performance Metrics:

- Perform modeling using the training dataset.
- Evaluate the model(s) using the test dataset and validating as well.
- Ensure all modeling methods are well-motivated, correctly implemented, and, to the extent appropriate, span the range of methods discussed in this course.
- Cross-validation and/or held-out test sets are used in accordance with best practices to assess model performance.
- Performance metrics are carefully tailored to the project objectives.

Modeling Results and Findings:

- Discuss the results comparing different models and explain the differences and the challenges.
- Ensure all project objectives are fully met, findings are clearly presented, and question(s) are technically addressed in the report.
- Include tables/graphs comparing the different models, including their characteristics, performance and accuracy.

Project Timeline:

- Module 2 (by the end of Week 2): The course instructor will group students into teams of two to three members. Each team should select and introduce a dataset of their choice. Canvas, USD Email, or Slack can be used to find prospective team members.
- Module 4 (by the end of week 4): Your team will submit a status report of the project. The team representative will need to submit the "Team Project Status Update Form." This form includes describing the dataset(s) your team has chosen.
- Module 7 (by the end of Week 7): Each team should submit deliverables for the course project in the final week:
  1. **Report:** One PDF document containing a final report explanation for analyses, interpretations (7-10 pages without appendix), and appendix (see below).
  2. **Appendix:** Include PDF and HTML documents for submission with code, comments, and results. **Attach these documents to your final report for submission.**
  3. **Video Recording:** One video recording of the technical project presentation with all team members' contributions. Visuals are required in the video

recording presentation.

*It is critical to note that **no extensions will be given** for any of the final project due dates for any reason, and final projects submitted after the final due date will not be graded.

Examples:

- Feature Engineering: Develop ways to rank features. Do you think the final performance is a factor of the feature engineering, or the strength of the underlying algorithm?
- Using Kernels: Use a more complex nonlinear kernel, such as the pyramid match kernel for images.
- Using Neural Networks: What architecture do you plan to use? You should develop a network from scratch. You can easily download a trained neural net model for something like image classification. What would you learn from this exercise?
- Drum Sounds (expanded from class example). How can you get more examples? Many sets of drum samples exist, and could collect these online. There is likely to be some problems, such as covariate shift. What are the features? The raw data are .wav files. How will you preprocess or trim the files, if necessary? The spectrum of each sound will be computed. Each sound will be represented by the spectral power.

Project Deliverables and Submission Format:

- Write and submit your final project report in APA 7 style (here is a **Sample Professional Paper (https://sandiego.instructure.com/courses/19462/files/3233796?wrap=1)**). The final report (without Appendices), including text and selected tables/graphs, should be 7-10 pages.
  - Appendix: Attach the PDF and HTML documents for submission with code, comments, and results, to the report.
- Prepare a recorded video presentation of your project using a screencasting tool or you can use your USD licenses to Zoom to record your screen and provide a voice narration. Ensure that the sound quality of your video is good and each member presents an equal portion of the presentation. Export the video file to an mp4 format. Visuals are required in your presentation.
  - You may use any recording software you wish. View the **Recording Video Presentation and Submission Guidelines for MS-ADS Students (https://sandiego.instructure.com/courses/19462/files/3233684?wrap=1)** guide for additional recording instructions.
- Submit the final PDF and HTML documents and video file on the final project submission page of Canvas. You will use the naming convention Final Project-Team Number.pdf (e.g., **Final Project-Team 1.pdf**). <u>**Only one member of your team will need to submit these deliverables.**</u>
- Submit the peer evaluation form below individually.

NOTE: Team members may not get the same grade on the Final Team Project, depending on each team member's level of contribution.

To understand how your work will be assessed, view the assignment rubric on the **Final Team Project (https://sandiego.instructure.com/courses/19462/assignments/303424)** page.