Mid-Term Project Requirements (First DL project)

0: **Project IDEA: No Late Submission of Project IDEA**

* 1 Page. Submission: word or written
* Can be handwritten and photocopied
* Items to include:
  1. Problem definition. What issue, business / practical problem you need to resolve
  2. What data you will use:
     + Source: link(s)
     + (Sample or at least list of columns/ features)
  3. Hypothesis (outcome expectations)
     + 3 of them (at least 2)
  4. What Deep Learning algorithms (methods) you intend to use
  5. What DL model you will train or reuse as transfer learning
  6. Outcome expectation (practical purpose of the project)

1: Individual or small team of up to three students

* It has to include elements of the following listed below, at least one of each:
* Data: CSV file, images / pictures, text data (can be PDF as well), timeseries (metrics)
* DL Frameworks: TensorFlow, PyTorch, Keras, Rapids, SckiKit (DL algorithms, such as MLP only), no LLMs for this project.
* Algorithms, such as deep learning method, procedures, functions
* As possible, use GPU (on Google Colab, your laptop/ workstation, GCP/AWS/Azure)

2.1. Examples of Open data-data sources are:

* Use one or more public data sources or collect your own data (data element). See Google Cloud Big Query public data sources.
* San Francisco Public Data
* New City Public Data
* (other public data sources)
* Openweathermap.com
* OpenStreetMap.org
* Part of Keras, TensorFlow and other frameworks
* UCLA, Univ of Toronto, MIT, other universities

2.3. Example of algorithms

* DL Regression (linear, non-linear, single/ multi variant)
* DL Classification
* CNN/RNN /LSTM models
* (LLMs / GenAI will be for the other, Final Project)

3: The project must solve a **problem** (have a business or practical problem); always start with the actual problem you need to solve.

Accuracy is not important as long achieves 50%; You can explain part of future work how you would improve accuracy and other elements.

Don’t present project which just compares multiple models and improves accuracy.

4: The project should have documented initial hypothesis about how you want to derive the outcomes. Through the course of the project refine the hypothesis, potentially drop some, add new ones.

5: The project should have the following sections:

* Short Abstract
* Review and validation of the output. Explain how it supports (confirms or disconfirms) your hypothesis. Evaluate loss and/or accuracy.
* Data requirements, data source(s) selection and data review (sample content or simplified data mining), data description (ex. data structure used, such as csv format, data frame, etc.)
* Python Notebook (or Scala Notebook)
* Description of chosen DL algorithm(s)
* Simplified review of the approach and process used to perform the project.
* High level diagram about your solution (showing individual components).
* Evaluation of the results, **conclusion**.
* References and publications used.
* (Future work) What you would have done differently if you would have more:
* Time
* Analytics programming skills
* Better data
* Better functions/ algorithms shipped with libraries, packages
* Upfront training on using certain DL technology
* More knowledge of GPU and/or cloud platforms
* (optional) Future enhancement areas to your project (ex. if you want to publish about at Pace Research Day in 2024 or as formal publication, such as IEEE.
* IEEE publications receive extra grade points.
* As outcome, evaluating accuracy / loss is not enough !!! Using the model, must predict or must classify and evaluate the results (quality of prediction, confusion matrix, etc.)

6: Format:

* Word document of at least 4 pages (with at least one diagram). When multiple team members, add additional page for 2nd and also for 3rd team member. IEEE Word format recommended but not mandatory.
* PowerPoint presentation to be able to present it in max 10 min (Title + 3 slides per individual project and 2 slides per person when working in team)
* Must be submitted via “Pace Portal”
* Project Idea: Early draft in format of Word document expected to be submitted earlier (minimum 1 page per individual project and minimum 2 pages per group project). No late submission on project idea.
* Suggested Word Template is IEEE