Term Project

A. Overview

1. Team Formation
2. Project subject
3. Used data
4. System Architecture
5. Project Repository
6. Creating project manager account
7. Create project's webpage
8. Project delivery
9. Project presentation

1. Team Formation

- We will have 5 teams.
- We will have **three groups** with **5 members** and **two groups** with **four members.**
- Teams will be given alphabetic names: Team-A, Team-B, Team-C, Team-D, Team-E.

1. Used data

The data used is provided by Amazon for applications reviews. The data contains these fields (columns): application ID, reviewer name, review text, and reviewer rating, review and review date. The data contains 111,143 reviews each in a row (record). The reviews text has sentimental about how the reviews think about the application, and the reviewers rating is an expression of how the reviewer like the application on the scale 0 to 1, with increments of 0.1 (10 levels).

2. Requirements

The main functional requirement is to have a recommendation from a machine learning model when presented with a text review.

3. Subject Of The Project

- A multi-class neural network model was prepared for predicting the output reviews. The model was trained on 1555 records and tested on 667 records. Preprocessed and some feature extraction operations were performed on the data before using it to prepare the model. The performance of the model is very bad (overall accuracy: 60%). Basically the the model failed to predict the right classes in the testing phase and tends to predict only one class.
- The goal of this project is to re-design and re-train the model to give better performance.
- The project is about a new cycle for the model where:
  - the requirement analysis will be revisited and validated with stockholders (you can use your imagination to capture the stakeholders' requirements)
  - formulate the specification from the requirements
  - put an architecture for the system
  - plane a remodeling for the system
  - implement the new model
  - Verify the model's outputs against the requirements
  - document the process step by step
  - manage the process step by step

- communicate the results step by step
- deliver the product in a presentation

- The platform used to prepare the model is Microsoft Azure ML Studio. Yo can easily create a free account using your RIT email account.
- The model is published and you can use this link to access it:
  https://gallery.cortanaintelligence.com/Experiment/DSCI644-TermProject

4. Project's webpage

A webpage has to be developed to communicate the project's purpose, requirements, goals, architecture, and any other relevant information. You can use GitHub to host your webpage since a team member can create a repository for it and add the other members of the team as collaborators. This template is for your convenience:
https://github.com/a-elsaid/https-github.com-a-elsaid-dsci644_demo_webpage.git

5. Project Management

The team shall use Trello (trello.com) to manage the progress of the project. The team shall create cards for each step in the project and assign the to team members. (Please see attached picture for a quick start).

6. Software Life-Cycle Model & Deliverables

The project will use OpenUp process. For more information of OpenUp process (http://www.se.rit.edu/~swen-261/topics/OpenUP%20Phases.html).
We will use four sprints starting from week 4:

a. **Inception Phase (1 week-Week 4 [7 Sep - 13 Sep])**
   Deliverables: Detailed project proposal and key requirements

b. **Elaboration Phase & Planning (5 weeks-Week 5-9 [14 Sep - 18 Oct])**
   Deliverables
   - ***Github repository (please add me as a collaborator: a-elsaid)***
   - ***Project manager account (use Trello… please add me to the team: aae8800@rit.edu)***
   - ***Project's webpage***
   - ***Architecture and other designs***

c. **Construction Phase (4 weeks-Week 9-12 [12 Oct - 8 Nov])**
   Deliverables: The working product!

d. **Transition Phase (1+ week-Week 13-15 [9 Nov - 24 Nov])**
   Deliverables: Team reflection, presentation, and other wrap-up materials.

7. Remarks

The preprocessing and features extraction operations performed on the data are not ideal. The model used is not ideal too. The challenge is to find a better method(s) to process the data and to pick a suitable model for the problem. The problem was initially defined as a multi-class problem but you as a data scientist and AI architect can redefine the problem and propose a better model accordingly.
Most of the question which will come up during the project are research questions, like which text preprocessing method should be used, and what ML model should be used, or what hyper parameters should be used. Research questions are open-ended questions which should be

answered through trail and error, studying papers and reviewing articles and forums about how others approached the similar problems. Using Azure AI Studio is also an opportunity to learn about the tool, which is simple to use and offers most of what you would need for ML.

8. Dates *(Note: "the end of week X" means Sunday-11:59PM of week X)*:

A. Team formation, Detailed project proposal, & key requirements : By the end of week 6.
B. GitHub repository & Trello accounts: By the end of week 7.
C. Project's webpage: By the end of week 8.
D. Projects Architecture & Design: By end of week 9.
E. Project Progress Report/Demo: By the end of each sprint as listed in the above table.
F. Final Project Deliverables: By the end of week 13
G. Project Presentation: In weeks 14-15.

9. Remarks

- The preprocessing and features extraction operations performed on the data are not ideal.
- The model used is not ideal too.
- The challenge is to find a better method(s) to process the data and to pick a suitable model for the problem.
- The problem was initially defined as a multi-class problem but you as a data scientist and AI architect can redefine the problem and propose a better model accordingly.
- Most of the question which will come up during the project are research questions, like which text preprocessing method should be used, and what ML model should be used, or what hyper parameters should be used.
- Research questions are open-ended questions which should be answered through trail and error, studying papers and reviewing articles and forums about how others approached the similar problems.
- Using Azure AI Studio is also an opportunity to learn about the tool, which is simple to use and offers most of what you would need for ML.
- A considerable part of the evolution will be based on the effort done by each member in the team (which should be reflected on the Trello Board).