

MIE 1624 Introduction to Data Science and Analytics – Winter 2020

Course Project

Deadline: Thursday, March 26, 11:59pm

Background

Data science, analytics, AI, big data are becoming widely used in many fields, that leads to the ever-increasing demand of data analysts, data scientists and other data professionals. Due to that, data science education is now a hot topic for educators and entrepreneurs.

In this project, your team will first re-design the course curriculum for “MIE1624: Introduction to Data Science and Analytics” course at University of Toronto, such that students acquire an introduction to the most relevant topics and skills in data science and AI. Second, your team will need to design a curriculum for a new “Master of Data Science and Artificial Intelligence” program at University of Toronto with focus not only on technical but also on business and soft skills, see, e.g., <http://www.rotman.utoronto.ca/Degrees/MastersPrograms/MMA>, http://smith.queensu.ca/grad_studies/mmai/, <http://mbai.kse.ua>, or <http://schulich.yorku.ca/programs/mmai/>, that contains optimal courses (and internships, projects, extra-curricular activities, etc.) for students to obtain necessary technical and business skills to pursue a successful career as data scientist, analytics and data manager, data analyst, business analyst, AI system designer, etc. Third, your team will need to develop your own analytics-based solution in education field, examples of ideas can be found below, but you are not limited to those.

Learning Objectives

- Develop the ability to work in a team on a consulting project. (You are required to work on the project in the same group as for your in-class presentation. Check the Quercus portal for the list of your group members.)
- Improve on skills and competencies required for performing a full cycle of data science and analytics workflow, i.e., data collection and pre-processing, applying algorithms to analyze data, trend identification, storytelling based on analytics (writing a consulting report and delivering an oral presentation).

Tools Allowed

- You can use Python libraries mentioned in-class as well as any other Python libraries you find during your research. Note that you can only use Python 3.
- For visualizing results in your report and presentation you may use Python or any other outside tool, e.g., Tableau, Power BI, IBM Watson Analytics, etc.

TO DO:

Finish the following four parts **based on your data analytics:**

Part 1 - Course curriculum design:

Your team has been tasked with redesigning the course curriculum for MIE 1624: Intro to Data Science and Analytics (**i.e., choose topics that should be covered in this course**). Topics selection can be based on skills required for data analyst/scientist/manager job postings from websites like <http://www.indeed.ca> (may need to do web-scraping), <http://www.glassdoor.com>, <http://www.linkedin.com>, <https://www.upwork.com>, 2018 Kaggle ML and Data Science Survey <http://www.kaggle.com/kaggle/kaggle-survey-2018>, 2019 Kaggle ML and Data Science Survey <http://www.kaggle.com/c/kaggle-survey-2019>, on curriculums of courses offered at online platforms, such as <http://CognitiveClass.ai>, Coursera, EdX, DataCamp, and/or any other data that your team can find or collect yourself.

Part 2 – Data Science program curriculum design:

Your task is to design a sequence of courses and curriculum for each of those that can be used by University of Toronto. The hope is that your team will help to **design a curriculum for a new technically- and business-oriented program: Master of Data Science and Artificial Intelligence**. Propose how many required and optional courses should be offered and design a curriculum for each course based on datasets from Part 1 and any other data that your team can find or collect (e.g., similar programs at other universities worldwide, current Emphasis in Analytics at UofT course curriculum, etc.). Propose what other activities should be part of the program, e.g., internships, projects, extra-curricular activities, etc.

Part 3 – Visualizations of course curriculum:

Your task is to visualize at least 5 courses that you have designed in Part 1 and 2 as well as the whole program structure. Identify skills that each course is teaching and include those in your visualizations. Identify sequence(s) in which skills need to be learned and visualize those in the form of a graph/network. Design visualization(s) that allows potential applicants of the Master of Data Science and Artificial Intelligence program at UofT to grasp program curriculum, program structure, acquired skills, etc., in one shot via your visualization(s).

Part 4 - Data Science education EdTech effort:

Now your team has become an experienced team in Data Science education, and you are concerned to advance AI education, namely to establish your own EdTech startup. **Propose and answer your own question related to design of educational efforts (something that helps you and others to find jobs) in data science.** It can be, but not limited to:

- Recommender system to help students to select elective courses at Emphasis in Analytics <http://gradstudies.engineering.utoronto.ca/professional-degrees/emphasis-in-analytics/>

- Online course delivery system that automatically adapts to learning patterns of an individual student and automatically adjusts a sequence of slides in online course for her/him.
- Which companies can be approached in Canada for internships in data science and why?
- Your own idea. You are advised to discuss your ideas with course instructor and course TAs prior to settling on one of those.

Note: the scope of the question is quite wide, and it is advised that you narrow it down based on your interests and expertise. Make the work truly yours.

Project Presentations

- Project presentations are open to the public and are scheduled for **Tuesday, March 31, 6:00-9:00pm, room BA 1160.**
- **Do not make your presentation overly technical.** Your audience is business-oriented and may know little about data science, people are interested in the insights that you got from your analysis and why your results can and should be used for decision-making.

What to Submit via Quercus:

1. Your Jupyter notebook with appropriate documentation for every step as well as the relevant data files. Comment out any data retrieval processes (e.g., from web scraping, downloading, APIs, etc.) in your code and replace it with code for reading the corresponding data from files. (**Submit all those data files together with your Jupyter notebook**). Make sure that your Jupyter notebook runs on <https://labs.cognitiveclass.ai> portal with the Python 3 kernel and that all needed data files are included in your submission. If the size of the data files exceeds Quercus's capacity, those should be stored on a cloud drive (e.g., Dropbox, Google Drive), and the link to the directory should be included in the notebook.
2. A 5 to 10-page consulting report in PDF and DOCX formats that summarizes your findings and results (all graphs should have axes appropriately labelled, all visual materials should be understandable and the graphics of sufficient quality to be easily readable.) This report should be business oriented and cover your problem more extensively than your presentation.
3. Your business-oriented presentation slides in PowerPoint and PDF formats. (Each group will present their findings and results during a 7 minute presentation with 1 minute for questions. Presentations will be timed and stopped after 7 minutes.)

Marking

- The project is worth 30 points (15 points for your analysis and report and 15 points for your business-oriented presentation).
- The presentation will be graded as follows (15 total marks):
 - 5 marks for organization and delivery (e.g. clarity, enthusiasm, poise)
 - 5 marks for content (e.g. proper visuals, high-level ideas, answering questions)

- 5 marks for the business pitch (e.g. recommendations, solution to the problem)
- The analysis in Jupyter notebook and the report will be graded as follows (15 total marks):
 - 4 marks for identifying the problem and searching the relevant data
 - 7 marks for the analysis (e.g. cleaning the data, visualizations, applying algorithms)
 - 4 marks for discussion and insight (e.g. how your analysis contributes to the problem, making a decision, storytelling)
- Every group member gets the same mark for the project. It is your responsibility to determine how you split the work inside your group. At least half of your group needs to be present during the project presentations to present the results.

Notes

- For the deliverables, consider the Jupyter notebook as what you would report to senior data scientists and machine learning engineers, and the consulting report and the presentation as what you would report to CEOs, VPs, PR managers, university officials, government officials, journalists, angel investors and venture capital people.
- The presentation would be a visual representation of the executive summary of your report.
- The audience for your presentation and report in particular is business-oriented and includes people who are interested in the insights you gathered from your analysis and how your results should be used for decision-making.