

MAIS 202 - PROJECT DELIVERABLE 1

Due: Tuesday, October 1st, 2019

Over the course of MAIS202, you will be completing a machine learning based project of your choice for the final project. At the end, you will have the option to either:

- 1. Demo your project by integrating it into a webapp (or something more advanced) **OR***
- 2. Present your work as an academic project through a poster*

McGill AI Society will be hosting a project fair where you will be showcasing your work! It will be awesome!

For both cases, you will end the project by writing a blog post about it.

Submission

This is an individual deliverable. All deliverables should be electronically submitted on Github and completed with the same academic integrity and standards expected at McGill University. Include appropriate citations.

To submit, create a repository under your own Github, you can name it whatever you want, and push your report there. List your repository link in this [spreadsheet](#). All of the code and report for this project should be found in this repository. Make sure to maintain it with properly documented README and structured code.

Submit this deliverable as “Data Selection Proposal.pdf”

Max length: 1 page

Deliverable Description

The first step of the project is to choose the dataset that you want to work with and propose your project idea.

1. Choose your dataset

You can choose any public dataset of your choice (don't forget to cite them!). There are also a couple of useful databases that are available: [Google Dataset Search](#) and [Kaggle](#). Explain the reasons why you choose this dataset. Furthermore, you can also look into creating your own custom dataset by scraping websites. If so, explain what kind of data you will be scraping.

2. Methodology

Describe how you plan on approaching the project. This should be a **high level** overview of your plans, and this will allow us to judge the feasibility of your project. Be as thorough as you can, so we can give you critical feedback.

i. Data Preprocessing

Is the dataset you chose feasible? What information provided is/are the most useful? How are you planning on preprocessing the dataset to extract this information?

Here are [slides](#) to our data preprocessing workshop from Fall 2018. The slides for this semester's data preprocessing workshop will be released once the workshop is given.

ii. Machine learning model

What do you want predict/estimate from this dataset? Propose a machine learning model/algorithm for it, and explain your reasoning. Have you considered other alternative models? What are the pros and cons?

iii. Final conceptualization

For demo purposes, we want you to be able to showcase your project! Indicate your choice for the final project and explain the details requested below. This is not final, you can always change through the semester, just keep us updated.

- Application

We want you to integrate your model in a simple landing-page webapp. For those of you who have experience, you are welcome to integrate your model in more sophisticated technologies (eg. mobile, hardware, webapps). Keep in mind that this is machine learning course, so we will not provide significant guidance for software dev related topics. Work with something that you are comfortable with :). Give the general idea of your application, and the technologies you planned on using. Here is an example of a simple [webapp](#) that uses Computer Vision to estimate people's age.

There will be no limitations to the product. You are free to build anything you want! :)

- Poster Presentation

We want you to prepare a poster presentation, just like what people find at NeurIPS, and other ML conferences. Since this is a more academic route, your goal for this project would be to try and match and/or beat the standard baselines for the output you are trying to predict. Do some research on previous explorations of the dataset and the accuracies/mean squared error that have been achieved with it (include the references in this deliverable). For other kinds of projects, do some research and explain the specific problem's accepted metrics. Report the average baseline results which you hope to beat (eg. predict x with y% accuracy, < x mean squared error, etc).