Adrien Philardeau-Planche

MAIS 202 Project Deliverable 1

1. Dataset:

World Happiness Report: “Happiness scored according to economic production, social support, etc.” by Sustainable Development Solutions Network.
https://www.kaggle.com/unsdsn/world-happiness
I chose this dataset because it offered a few different interesting opportunities for machine learning models. For example, one could train a model to estimate how much one specific factor affects overall happiness, and which factor is the most important. Also, a trained model could predict happiness of any group if it was inputted its data.

2. Methodology:

i. The data is a CSV file with 158 countries and scores for 6 different factors, along with the happiness rank and score of each country. The data is thus quite good and feasible, and probably won’t need too much preprocessing. A potential problem could be the relatively small amount of data, with approximately only 1000 data points.
ii. It would be interesting to have a predictor of happiness. Using regression, we could input the six data points and then figure out the level of happiness of a group of people with those attributes have achieved. Because we have the data for happiness already, we can use supervised learning, and because we want to predict, we should use regression, so the two best options here would probably be a neural network or a random forest algorithm. Unsupervised learning, more specifically dimensionality reduction algorithms, could be used but probably don’t have enough data to train an accurate model.
iii. I would choose a simple landing-page webapp. The user would simply have to input their scores for a handful of categories and the algorithm would spit out their calculated happiness level and actionable steps on how to improve it. Web design technologies would be adequate for this task.