Project Proposal

MIS 545 Summer 2020

Ryan Straight

Abstract

Assignment submitted to MIS 545 Summer 2020 on 2020-07-29. R packages and versions used in this document: R (Version 4.0.0; R Core Team, 2020) and the R-package *papaja* (Version 0.1.0.9942; Aust & Barth, 2020). Any is generated using the [LaTeX Equation Editor](https://latexeditor.lagrida.com/).

*Word count:* X

Project Proposal

Given the timeliness of COVID-19, the illness caused by the SARS-CoV2, and the considerable amount of tremendously granular data available about confirmed cases, school and business closings, travel restrictions, et cetera, this will be the main focus of the project. The main goal of this project is to determine what, if any, predictors for success and financial loss mitigation exist within the data avaialble.

## Business Value

The potential financial hardship caused by school closings and resultant educational delays on the future workforce cannot be overstated. The amount of future earnings decreased by even a scant four months of higher educational closures might not seem detrimental but “the loss in marginal future earnings would be 2.5 percent per year over a student’s working life” (Vegas, 2020). Applied to the 76 million students and presuming a working-life of 45 years at an average annual income of $53,490, the Brookings Institute estimates an annual loss of $1,337 per student, resulting in a staggering $2.5 trillion for those four months of lost education (2020).

## Data Source

There is no shortage of sources for COVID-19 data. In an effort to ensure the data used in this project is up-to-date at time of publishing and will be consistently formatted throughout, the data will be pulled using the COVID19 package (Guidotti & Ardia, 2020). This data is combined with the “COVID-19 Lockdown dates by country” data set (jcyzag, 2020).

## Data Mining Techniques

TBD but will consist of:

* 3 different data mining techniques
  + two predictive (decision tree, naive Bayes, SVM, neural network)
  + one descriptive (cluster analysis, associate rule mining)

# References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>

Guidotti, E., & Ardia, D. (2020). COVID-19 Data Hub. *Journal of Open Source Software*, *5*(51), 2376. <https://doi.org/10.21105/joss.02376>

jcyzag. (2020). COVID-19 Lockdown dates by country. Retrieved from <https://kaggle.com/jcyzag/covid19-lockdown-dates-by-country>

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Vegas, V. C., Harry Patrinos. (2020). The COVID-19 cost of school closures. *Brookings*. Retrieved from <https://www.brookings.edu/blog/education-plus-development/2020/04/29/the-covid-19-cost-of-school-closures/>