## Project 2 (150 points)

## Supervised Non-Linear Classification Models

## Implementing a Naïve Bayes Classifier

## **Due Date and Time**

* Week 5, Sunday at 11:59 p.m. ET.

## **Instructions**

1. Review Chapter 8, pages 350-354 in the course textbook.
2. OPTIONAL: Review Naïve Bayes Classifier: <https://uc-r.github.io/naive_bayes>
3. OPTIONAL: Review the video: Naive Bayes Classification with R: <https://www.youtube.com/watch?v=RLjSQdcg8AM>
4. Write R source code that uses the **Naïve Bayes algorithm** on a dataset of your choice.
5. You may use one of the built-in R datasets (“Iris” for example) or find another dataset on your own. You may use any R packages to perform your analysis. You do not need to use the ones in the tutorials above.
6. Provide brief comments in your source code to document what you have done and how the algorithm is working.

## **Recommended Dataset Resources**

* The R dataset packages: <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html>
* Geospatial Datasets: <https://www.mlhub.earth>
* Datasets for Classification: <https://www.kaggle.com/tags/classification>, <https://www.kaggle.com/tags/binary-classification> and <https://www.kaggle.com/tags/multiclass-classification>
* Other datasets: <https://www.kaggle.com/datasets>

## **Academic Integrity**

The algorithms and programming techniques used in this assignment are very common Data Science problem solving approaches. You will find examples and source code all over the internet. Please do not copy code or use pre-existing examples that you find. Select your own, unique analysis on a new dataset of your choice. If you have questions about the use of acceptable examples and datasets, please contact your instructor for further clarification.

## **Submission Instructions**

Upload your **Source code** to the **Project 2** submission area.