**Title**:Predicting the Severity of COVID-19 Cases in the Philippines

**Group Members:** Lauren D’Arinzo (lhd258), Elizabeth Combs (eac721), Angela Marie (Amber) Teng (at2507) - *responsible*, Paula Kiatkamolwong (kk4158), Steven Dornberg (dornbs01)

**Summary of Plans**:

*Chosen Project:* The team selected to complete project 3: “fighting COVID-19”. We will use a [dataset](https://docs.google.com/spreadsheets/d/16g_PUxKYMC0XjeEKF6FPUBq2-pFgmTkHoj5lbVrGLhE/edit?fbclid=IwAR29_EijSX1_bTS1oNW6nuYNjMvoN-HcP44PCEoM13JSfXD0gf1Uu2D2XoE#gid=0) from COVID-19 Philippines, which is a live source gathered by Filipino data scientists, and obtained from the Philippine Department of Health (DOH) records of COVID-19 reported cases. The dataset was compiled and is currently managed by the Philippine Data Science [Group](https://www.facebook.com/datasciencephilippines). We chose this dataset because it contains [patient-level data](https://public.tableau.com/profile/rozauro.cordon?fbclid=IwAR1EaH7YlwhyylXkE4reI3wwBlbayTcyrl1W3KbxDLTnRnHo5BU-Mc_6P0I#!/vizhome/COVID-19PHTracker/COVID-19?publish=yes), which would allow us to answer the research question: Given a presumed COVID-19 diagnosis and features like age, travel history, epidemiological virus links etc, can we predict a person’s health status in relation to COVID-19? Particularly, will they have a mild or severe case? The dataset also contains opportunities for generating additional features including the length of time between on-set of symptoms and their admittance to a hospital.

*Proposed Approach:* We propose to use multiclass classification to solve this question. Our baseline models will consist of Naive Bayes and Decision Tree, which are naturally extensible to the case of having more than two classes. We also plan to test whether using SVM, K-nearest neighbors, and neural networks improves our error rate.  
*Suggested Experiments:* In finetuning our model, we will perform hyperparameter tuning for the optimal algorithm. We can also use oversampling, undersampling techniques to better address our imbalanced dataset. Finally, with additional time, we could add additional data from other countries to our analysis, including data from this [source](https://docs.google.com/spreadsheets/d/e/2PACX-1vQU0SIALScXx8VXDX7yKNKWWPKE1YjFlWc6VTEVSN45CklWWf-uWmprQIyLtoPDA18tX9cFDr-aQ9S6/pubhtml#). Another dataset we could review is news articles from the original source data that could allow us to add more features.

*References:*

1. <https://public.tableau.com/profile/rozauro.cordon?fbclid=IwAR1EaH7YlwhyylXkE4reI3wwBlbayTcyrl1W3KbxDLTnRnHo5BU-Mc_6P0I#!/vizhome/COVID-19PHTracker/COVID-19?publish=yes>
2. <https://news.abs-cbn.com/news/03/15/20/list-health-departments-list-of-confirmed-covid-19-cases?fbclid=IwAR29_EijSX1_bTS1oNW6nuYNjMvoN-HcP44PCEoM13JSfXD0gf1Uu2D2XoE>
3. <https://www.facebook.com/groups/datasciencephilippines/?post_id=897213460732154>
4. <https://www.doh.gov.ph/covid-19/case-tracker>
5. <https://docs.google.com/spreadsheets/u/1/d/e/2PACX-1vQU0SIALScXx8VXDX7yKNKWWPKE1YjFlWc6VTEVSN45CklWWf-uWmprQIyLtoPDA18tX9cFDr-aQ9S6/pubhtml#>
6. <https://en.wikipedia.org/wiki/Multiclass_classification#Naive_Bayes>

*Other Data Sources:*

1. <https://covidtracking.com/data/>
2. <https://www.tableau.com/covid-19-coronavirus-data-resources>
3. <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>
4. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
5. <https://www.doh.gov.ph/covid-19/case-tracker>
6. <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>
7. <https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data>
8. <https://www.safegraph.com/dashboard/covid19-commerce-patterns?fbclid=IwAR2vMqEtNSincgmCaLSde-eZ2lLxTCoKbftXHXjhIhe8XepIXudXZQaPyVg>
9. <https://github.com/beoutbreakprepared/nCoV2019/blob/master/source_list.csv>
10. <https://towardsdatascience.com/5-datasets-about-covid-19-you-can-use-right-now-46307b1406a>
11. <https://docs.google.com/spreadsheets/u/1/d/e/2PACX-1vQU0SIALScXx8VXDX7yKNKWWPKE1YjFlWc6VTEVSN45CklWWf-uWmprQIyLtoPDA18tX9cFDr-aQ9S6/pubhtml#>