The Gender Leadership Comparison in COVID-19 Data

Harry Chalfin

December 2020

https://github.com/hchalfin/DATA1030_final_project

1 Introduction

In the spring of 2020, some western media outlets noticed that nations with a female head of state seemed to be faring better in battling the COVID-19 pandemic than those with a male head of state. If true, this could potentially have major implications for our understanding of what kind of leader is most effective at managing a global crisis. Of course, there are many factors at play other than a leader's gender which may influence how well or poorly a nation fares in its pandemic response. In this report, we investigate the degree to which the gender of a nation's leader is linked to its success in battling the COVID-19 pandemic.

We use data from *https://covid.ourworldindata.org*, focusing specifically on the cumulative and new daily per capita case counts, death tolls, and tests performed in 167 countries around the world as of December 1, 2020. Of these countries, only 18 (10.8%) have female heads of state. This is a classification problem in that the goal is to build a model based on these data features which can be used to predict the gender of a nation's leader based simply on its per capita case, death, and testing statistics. Each of the 167 nations corresponds to a row in a scatter matrix with 7 feature columns: new cases smoothed per million, total cases per million, new deaths smoothed per million, total deaths per million, total tests per thousand, new tests smoothed per thousand, and tests per case. [1] The target variable is the gender of the nation's leader.

[1] Here, "new" refers to the positive tests recorded on December 1, 2020, while "total" refers to the cumulative value. The inclusion of the tests per case feature is somewhat redundant, but we have chosen to include it to provide extra cover for some missing values.

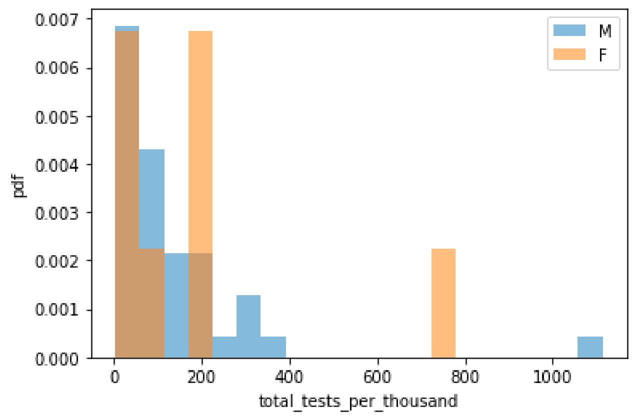

Figure 1: A histogram of the respective probability distribution functions (pdf's) for total tests performed per thousand citizens. The bin width is approximately 56. Male-led countries portrayed in blue, female-led in orange.

2 Exploratory Data Analysis

All seven features in our scatter matrix are continuous and have a wide range of values. Our target variable leader_gender is categorical (male or female).

The bar plot in Figure 1 shows how countries with female leaders tended to provided their citizens with slightly more testing than those with male leaders. The violin plots in Figure 2 and the heat maps in Figures 3 and 4 appear to undermine the narrative that female-led countries have fared considerably better than their male-led counterparts in this pandemic. The violin plots and the two heat maps appear to show female-led nations having only slightly more success than their male-led counterparts. (Consider that there are approximately 10 male-led countries for every female-led country.) Though there are many female-led countries including New Zealand, Norway, and Iceland which have had a largely successful response, there are also two female-led countries whose responses have been quite poor: Bolivia and Belgium. The narrative that female leaders have been stronger than male leaders in battling the pandemic likely originated from American journalists who noticed a poor pandemic response in their home country and many other male-led countries in the Americas and western Europe, particularly in comparison to the more successful responses of the primarily female-led countries of northern Europe. This perspective, however, fails to capture the low case and death totals reported in the many male-led developing nations of Africa and Asia. It is possible that these statistics are unreliable and under-reported, but the data suggest that this media narrative is inconclusive at best.

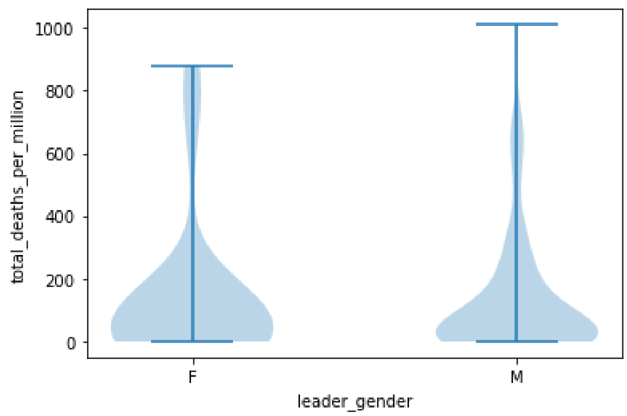

Figure 2: Violin plots of the total deaths per million citizens in female-led and male-led countries.

3 Methods

3.1 Data Preprocessing

With only 167 countries in this data set, the sample size is quite small. So we have chosen to allocate 60% of the points for our training set, 20% for our validation set, and 20% for our test set. Using per capita statistics, we presume that the data are independent and identically distributed – there is no inherent reason to expect certain countries' pandemic statistics to be linked. [2] Since all data features are either cumulative or specific to December 1, 2020, the details of the time series statistics leading up to December 1, 2020, are unimportant in this study.

The data are imbalanced – only 18 of these 167 countries are female-led. Therefore, it is appropriate to use stratified folds. To provide ample cross-validation, we iterate through 10 unique random states and extract one unique test set at each iteration; we then split the remaining data into four stratified folds and extract unique training and validation indices from each fold.

Since all the features are continuous with wide-ranging distributions, it is appropriate to use the StandardScaler on each. The LabelEncoder is an effective tool to use on the categorical target variable. These encoders are used to apply the fit_transform method to the training sets and the transform method to the validation and test sets. (It is important to apply the fit method only to the training set because we do not want to introduce an unnecessary bias when we

[2]Certain factors such as geographical proximity and frequency of travel between two countries might play a role in such a relationship, but it is difficult to say for sure so we use the simplifying assumption that the data are independent and identically distributed.

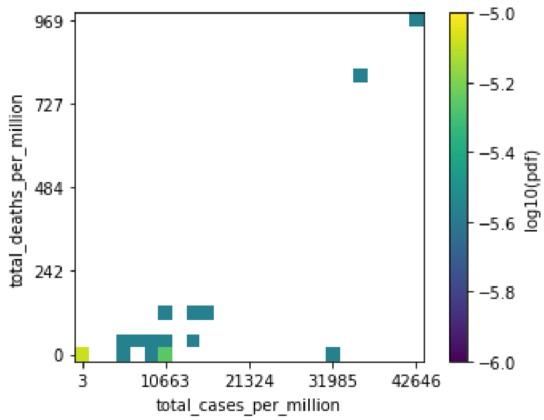

Figure 3: Heat map of the total cases per million citizens and the total deaths per million citizens for female-led countries. Scaled as a normalized probability distribution function (pdf) on a log_{10} color scale. Bin size is approximately 1775.

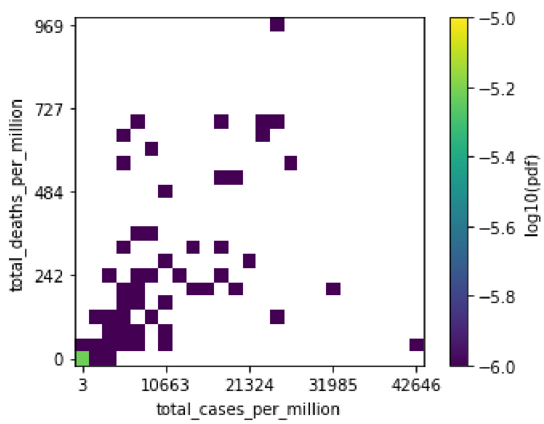

Figure 4: Heat map of the total cases per million citizens and the total deaths per million citizens for male-led countries. Scaled as a normalized probability distribution function (pdf) on a log_{10} color scale. Bin size is approximately 1775.

use the validation and test sets to optimize our model.)

3.2 Machine Learning Pipeline

One we have our training (X_train and y_train), validation (X_CV and y_CV), and test (X_test and y_test) sets, we input these pandas DataFrames into a reduced_feature_general function. This allows us to handle missing values by building numerous cousin models on the various subsets of the data, with each cousin corresponding to a unique missingness pattern. For each individual subset, we determine the ideal set of hyperparameters by finding the best score on the validation subset. In this study, we choose the accuracy_score metric, as it is quite simple and easy to work with. Once we have found the optimal set of hyperparamaters, we then use the model to predict the test subset. We combine all these prediction test subsets together to form a full prediction test set, which we then compare to the actual test set and compute an accuracy score accordingly. As we are looping through 10 random states, with four folds each, we compute a total of 40 scores. We then compute the mean and standard deviation of this set of 40 scores.

3.3 ML Algorithms

3.3.1 RidgeClassifer

The foremost hyperparameter of the RidgeClassifier is the regularization strength alpha. In the param_grid, we allow alpha to range from 10^{-4} to 10^{2}, in powers of 10.

3.3.2 RandomForestClassifier

Two important hyperparameters of the RandomForestClassifier are max_features (the number of features to consider when looking for the best split) and the max_depth (the maximum depth of the tree). When max_features takes on a value less than 1, scikit-learn interprets it as the fraction of all the features to consider. In the param_grid, we allow max_features to range from 0.2 to 1.0, in intervals of 0.2. we allow the max_depth to be 1, 3, 5, 10, 20, or None (no limit).

3.3.3 KNeighborsClassifier

Two important hyperparameters of the KNeighborsClassifier are n_neighbors (the number of neighbors) and weights (how important each neighboring point is). We allow the number of neighbors to take on a range of values from 1 to 13 (it is difficult to go above 13 while still being confident that each data point will have enough neighbors). The weights values can be either uniform (each point within the neighboring window gets an equal importance) or distance (points farther away are of diminishing reciprocal importance).

3.3.4 DecisionTreeClassifier

Like the RandomForestClassifier, the DecisionTreeClassifier is a tree-based classifier. Therefore it has a very similar set of hyperparameters to tune. In this project, we choose max_features and max_depth just as we did for the RandomForestClassifier, and we also choose identical ranges of hyperparameter values.

4 Results

Algorithm	mean score	std. deviation
RidgeClassifier	0.864	0.049
RandomForestClassifier	0.863	0.043
KNeighborsClassifier	0.848	0.048
DecisionTreeClassifier	0.847	0.043

Overall, the results of these models cannot consistently predict the correct gender more than about 86% of the time. The RidgeClassifier and the RandomForestClassifier are definitely the best, while the KNeighborsClassifier and DecisionTreeClassifier are worse by about 2%. All four models have a standard deviation of between 4 and 5%, so the four models are all well within one standard deviation of one another. However, since 89.2% of the countries in question are male-led, the simple baseline model which predicts 'male' every time regardless of the data in the scatter matrix would actually fare better than all four models here – more than half a standard deviation better than the two best models here, and about a full standard deviation better than the other two. This indicates that none of these models could meaningfully distinguish between the female-led and male-led countries based solely on the pandemic statistics analyzed here.

5 Outlook

5.1 A special note on Belgium

On October 1, 2020, The Belgian government underwent a transfer of power from female Prime Minister Sophie Wilmès to male Prime Minister Alexander De Croo. At the time, COVID-19 levels in Belgium were rising rapidly, and they continued to rise dramatically during subsequent weeks. Therefore it is difficult to neatly categorize Belgium as a "female-led" or "male-led" country during the period after September 30. The De Croo government, however, has more women than any other in Belgian history. It is notable though that De Croo had criticized Wilmès earlier in the pandemic: "Our country, our economy, our companies cannot afford new generalized lockdowns."

5.2 Conclusion

More broadly, the example of Belgium illustrates the difficulties with the immense oversimplification of the exercise of trying to assign credit or blame to world leaders of different genders in their handling of the COVID-19 pandemic merely by analyzing the nation's COVID-19 statistics. A country's statistics, such as case loads and death tolls, are the result of a very wide variety of factors other than who the nation's most powerful government official is, much less their gender. Among these factors are the response from other individuals occupying positions of power in those countries, as well as a nation's overall culture, population density, and economy.

Nevertheless, this study suggests that, even when such an analysis is performed, it is very difficult to meaningfully distinguish between the countries run by women and those run by men based on the COVID-19 statistics alone. (It is conceivable, however, that a more rigorous analysis could improve the model somewhat.) The western media narrative that suggests otherwise appears to be the result of cherry-picking certain countries (often based on geographic, cultural, and economic similarities) for comparison. For example, the female-led countries of Germany, Finland, and New Zealand have respectively been more successful in their responses than their male-led neighbors of France, Sweden, and Australia. Such media narratives are somewhat misleading because they ignore certain counterexample countries which refute the narrative. In particular, the female-led countries of Belgium (which had been hit hard even before the transfer of power) and Bolivia, as well as the many male-led countries on the African continent, largely refute this media narrative.

6 References

https://covid.ourworldindata.org/data/owid-covid-data.csv