Overview

* Objective: To help WHO monitoring progress by visualizing the dataset of Covid-19 vaccination progress.

Methodology

* Feature Engineering Techniques
* Used ExtraTreesClassifier:
  + This class implements a meta estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

Dataset

* How many features - 15
* Size of the dataset - **21628 observations and 15 attributes**
* What kind of data – Categorical (Ex: Country) and Numerical (Ex: People\_Vaccinated)
* Balanced or imbalanced – Balanced
* Distribution of Training set, validation set, testing set - NA
* Missing data and Preprocessing challenges –
* Imputed 0 in-place of Null
* Column Date has been split into Day, Month and Year columns
* Converted numerical attributes to integer

Results

Conclusion

* People from all the parts of the world are educating themselves and willingly taking the vaccines in most parts of the world. Also these vaccines have been proved effective against COVID-19 (till now). If the rate of people taking the vaccine continues to grow, then all the countries can vaccinate their people before the end of this year.

Future Work

* Analyze the effectiveness of vaccine with the death ratio and survival ratio of COVID-19 patient dataset

Feature Engineering Techniques

* Features removed -

“iso\_code","source\_website” –> Redundant data

* Feature creation –

“Year”, “month”, “day”

* Feature ranking –

Used ExtraTreesClassifier for feature ranking, top 3 numerical features are:

1. people\_fully\_vaccinated
2. total\_vaccinations\_per\_hundred
3. people\_fully\_vaccinated\_per\_hundred