UN Youth Hackathon

The impact on the SDGs of the COVID-19 pandemic

UNAMOR:
Barajas Cervantes Alfonso
Cerritos Lira Carlos
Cota Martinez Guillermo Oswaldo
Padilla Robles Artemio Santiago
Ruiz Puga Ingrid Pamela

The SDGs

The SDGs

The Sustainable Development Goals (SDGs) are 17 goals with **169 targets** that **all UN Member States** have agreed to **work towards achieving by the year 2030**.

The SDGs

We explored the four main SDGs

However we will focus on SDG 3: Good Health and hell being.

Dataset used

For this analysis we used:

- Mexico's National System of Epidemiologic Surveillance at Mexico City (SINAVE) (datos.cdmx.gob.mx)
- Sustainable Development Report 2021 (sdgindex.org)
- GoogleCloudPlatform/covid-19-open-data: Datasets of daily time-series data related to COVID-19 for over 20,000 distinct locations around the world. (github.com)
- COVID-19 cases and deaths by WHO - Dataset - UN Youth Hackathon Datahub (officialstatistics.org)
- US-based emergency declarations due to COVID-19 by Temple University - Dataset - UN Youth Hackathon Datahub (officialstatistics.org)
- COVID-19 national public responses in European countries by ECDC (2020-2021) - Dataset - UN Youth Hackathon Datahub (officialstatistics.org)

GOBIERNO DE LA CIUDAD DE MÉXICO | SECRETARÍA DE SALUD

SUSTAINABLE DEVELOPMENT REPORT

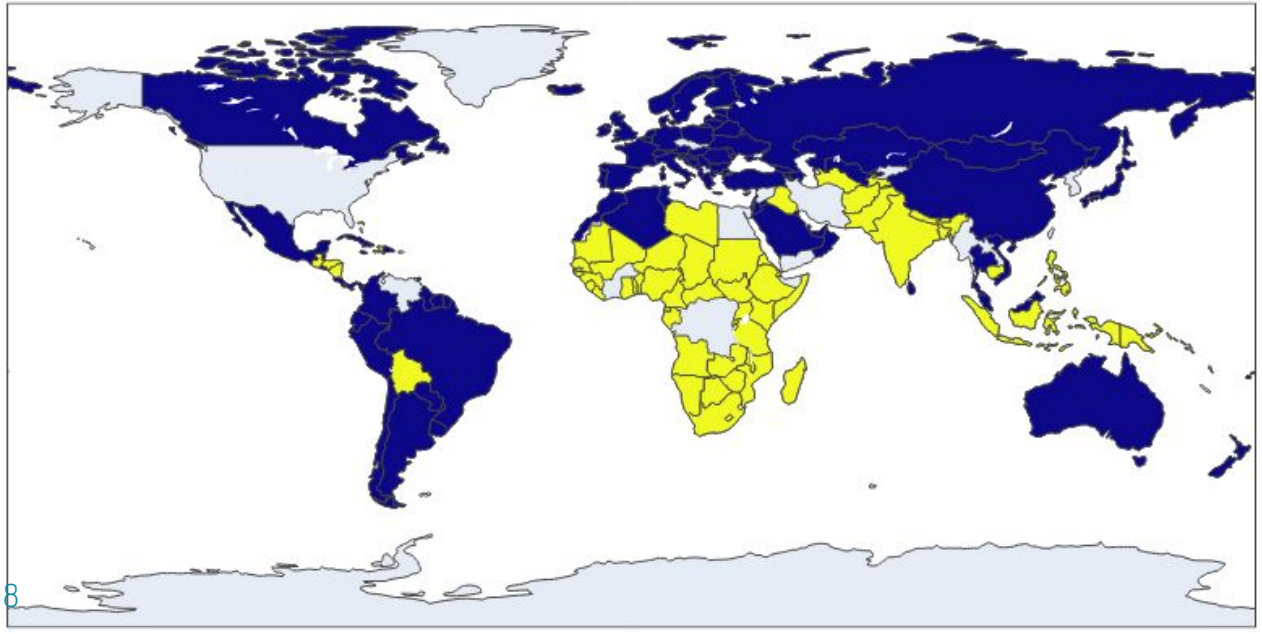
UN YOUTH HACKATHON

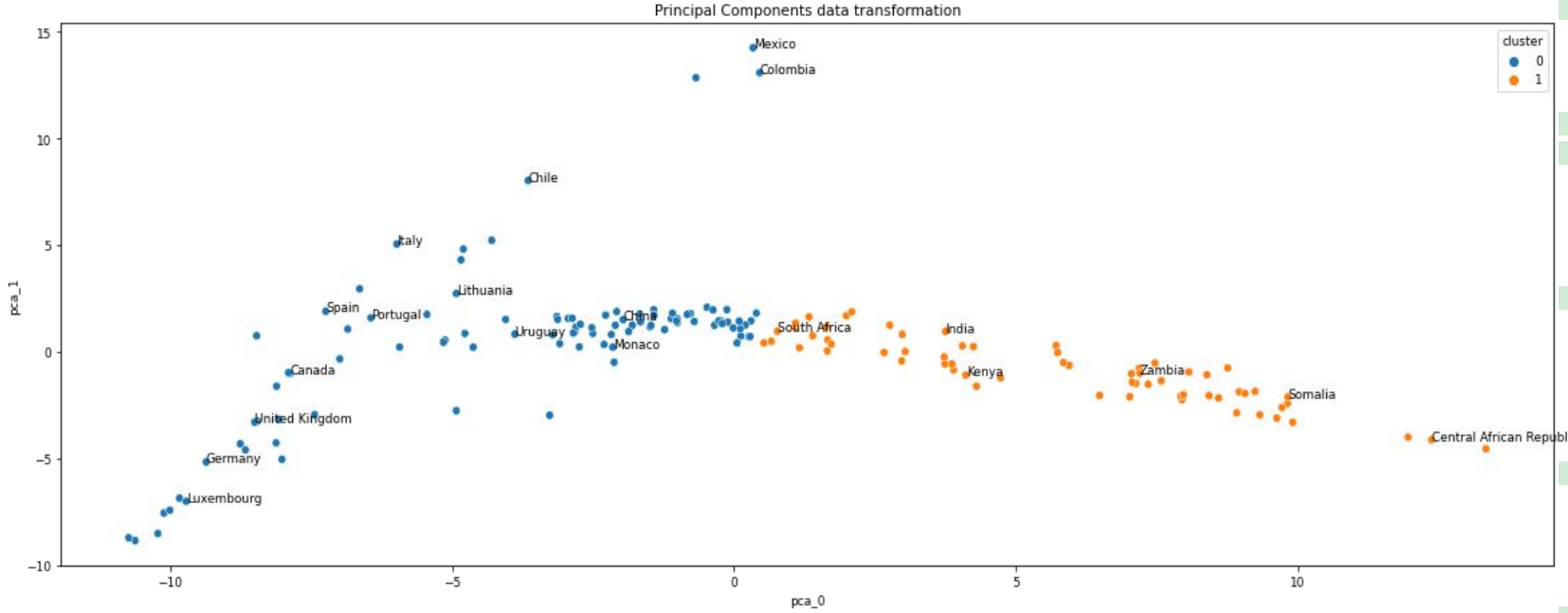
The impact of the pandemic on the SDGs

Clustering

Clustering on indicators

For a better understanding on the pandemic effect on SDGs, a **clustering** was made over **indicators** of demographic, health, economic, finance, among other sectors in 2020.





For the cluster creation, a principal component analysis was made together with an Agglomerative clustering.

The hyperparameters were optimized using the Silhouette coefficient

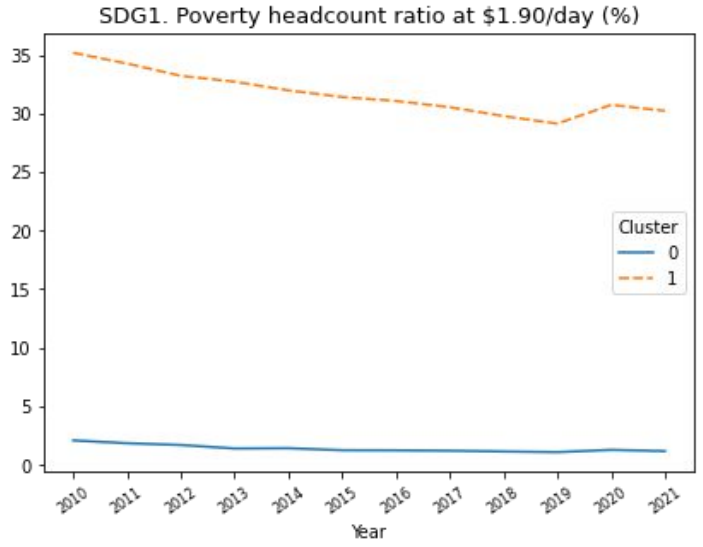
Preprocessing:

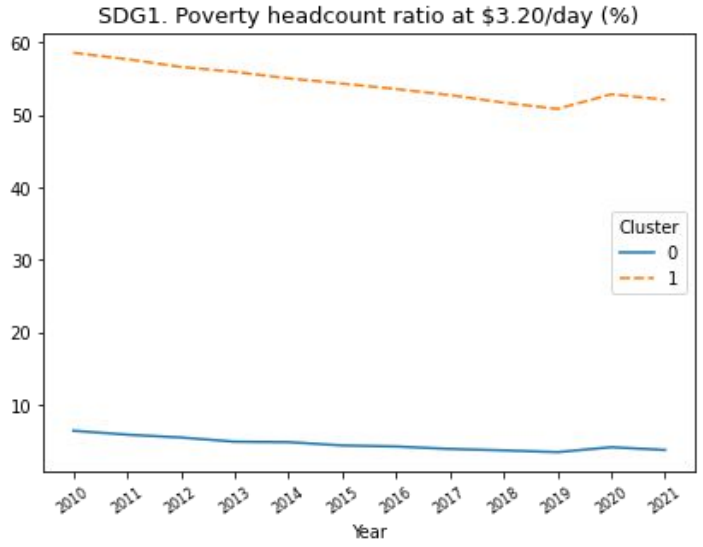
- Normalization
- PCA (2 components)

Clustering

- Agglomerative clustering
- Cosine distance
- 2 clusters
- Single linkage
- Silhouette coefficient: 0.77

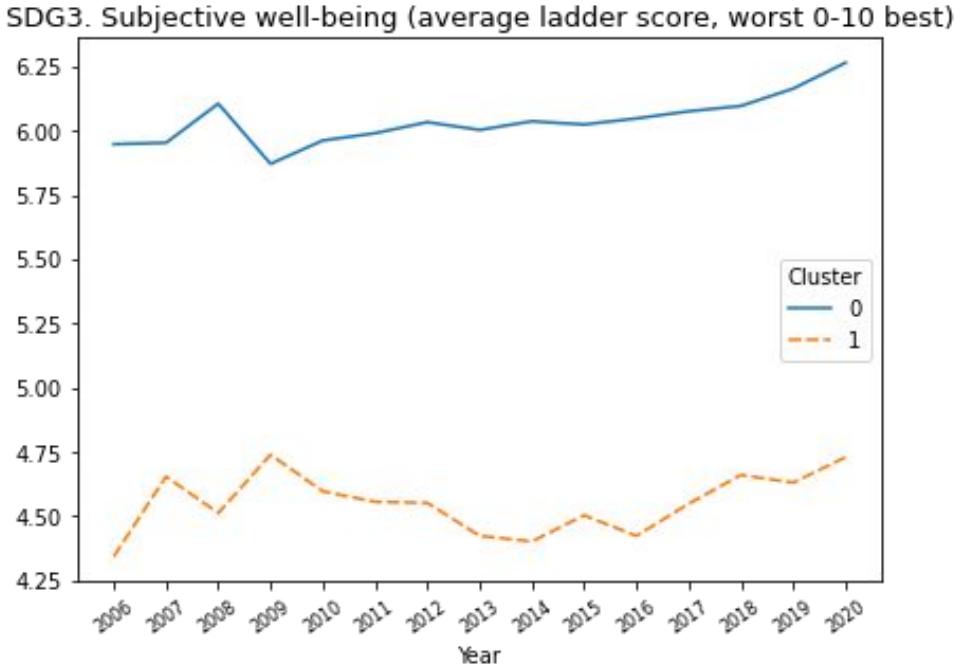
SDGs by cluster





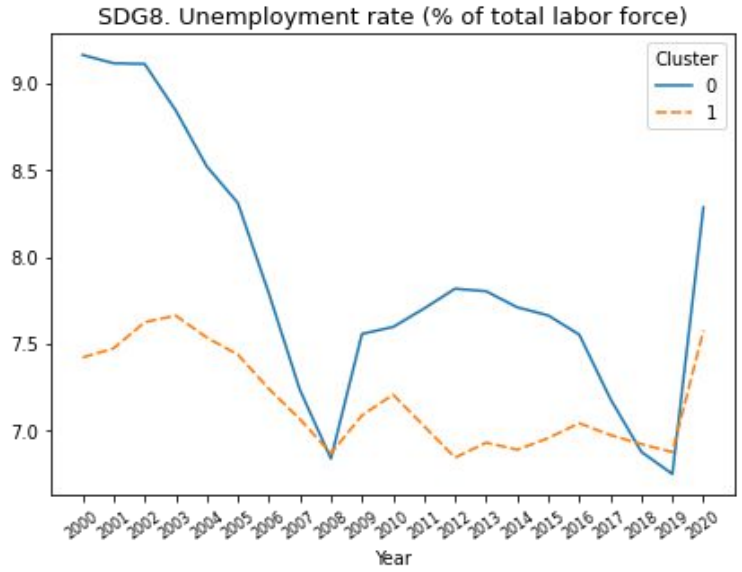
SDG1 Indicators:

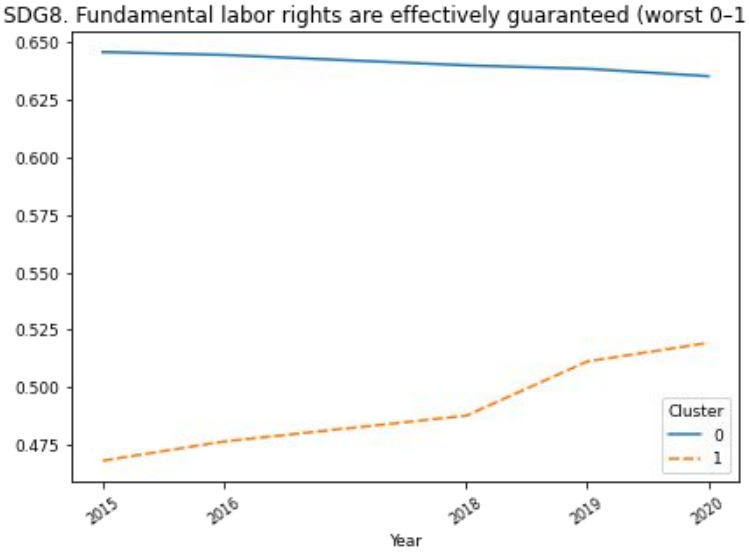
Poverty headcount ratio is significantly **higher** in cluster 0 than in cluster 1, this cluster had an **decrease** in 2020, at the start of the pandemic.



SDG3 well-being indicator:

This indicator saw an **increase** in 2020. There's a **margin** of approximately 2% **between the two clusters**, being the subjective well-being indicator of cluster 0 greater than in cluster 1.





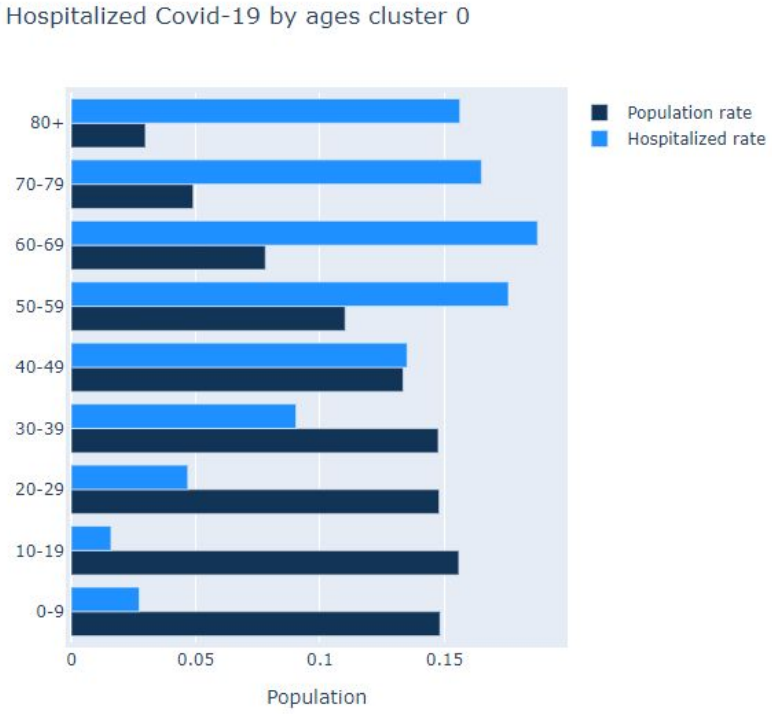
SDG8 Indicators:

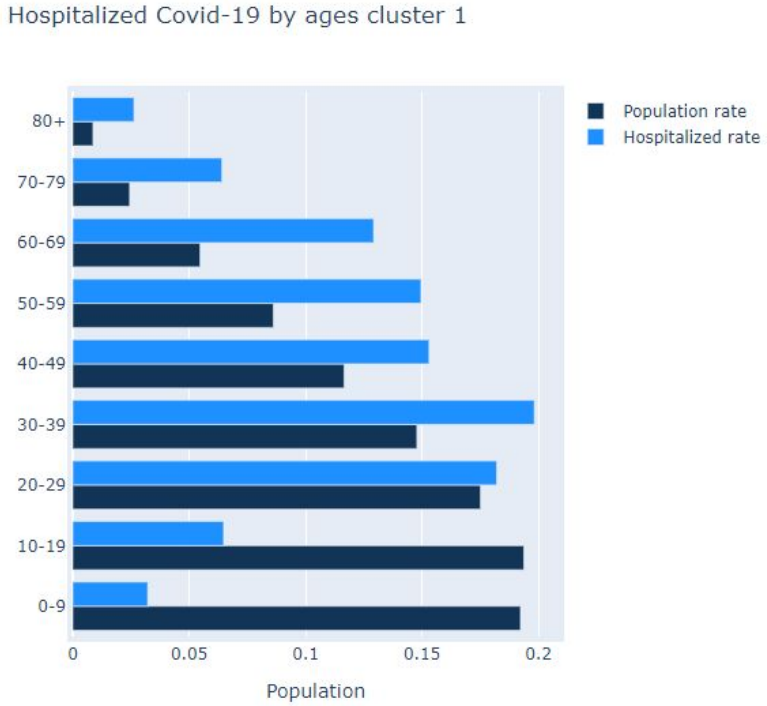
In 2020 cluster 0 saw a significant **increase** of **unemployment** rate (around %1.5), while cluster 1 saw this increase in lesser quantity (around %1). A small **increase** in fundamental **labor rights** guarantee of cluster 1 was observed during 2020, on the other hand, cluster 0 saw a small decrease in this indicator.

Population by cluster

Hospitalized cases

A notable difference is that countries belonging to **cluster 0**, a greater number of **adults over 70 years** of age were hospitalized than in **cluster 1**, where adults between **20 to 40 years** of age have a higher number of hospitalizations.



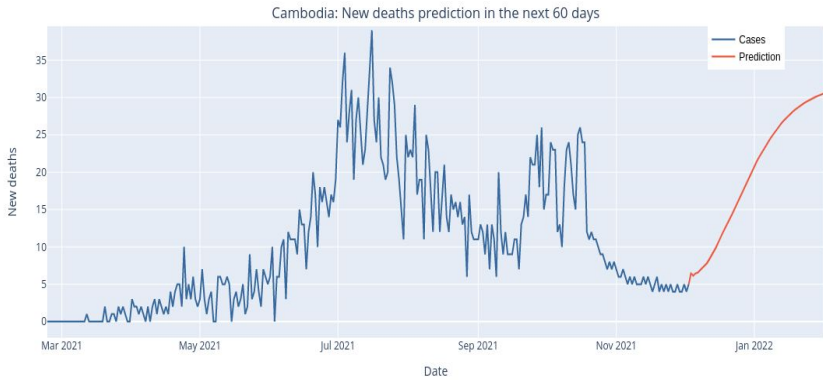


Using a Recurrent Neural Network to predict new cases and deaths by COVID-19

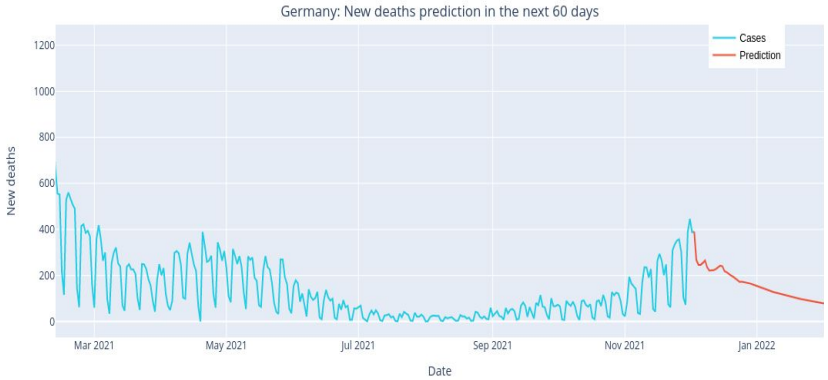
Predictions of new cases and deaths were compared between countries of different income categories.

There is **significant** difference between the trend of deceased values.

Low-income country: **Cambodia**



High-income country: **Germany**



Hypothesis derived from data

Hypothesis

Based on the Exploratory Data Analysis and Supervised Machine Learning insights…

"The pandemic has affected more SDGs in *low-income* countries as opposed to the *high-income* countries"

Proposed Solution

Identification of population at risk of hospitalization and death

using machine learning

The problem

Problem: SDGs in low income countries were the most affected by the pandemic, and there is still a long road ahead before the pandemic is over.

Solution: A machine learning model which is cheap, fast, effective and reliable for:

- Predict if someone with COVID will likely need to be hospitalized.
- Predict if someone with COVID will likely die.

Dataset

For this solution we use the database of Mexico's National System of Epidemiologic Surveillance at Mexico City that has data about symptoms, comorbidities, test results, hospitalizations and deceased, among others.

This methodology could be adapted to particular data from other countries.

GOBIERNO DE LA CIUDAD DE MÉXICO | PORTAL DE DATOS ABIERTOS

Machine Learning for everyone

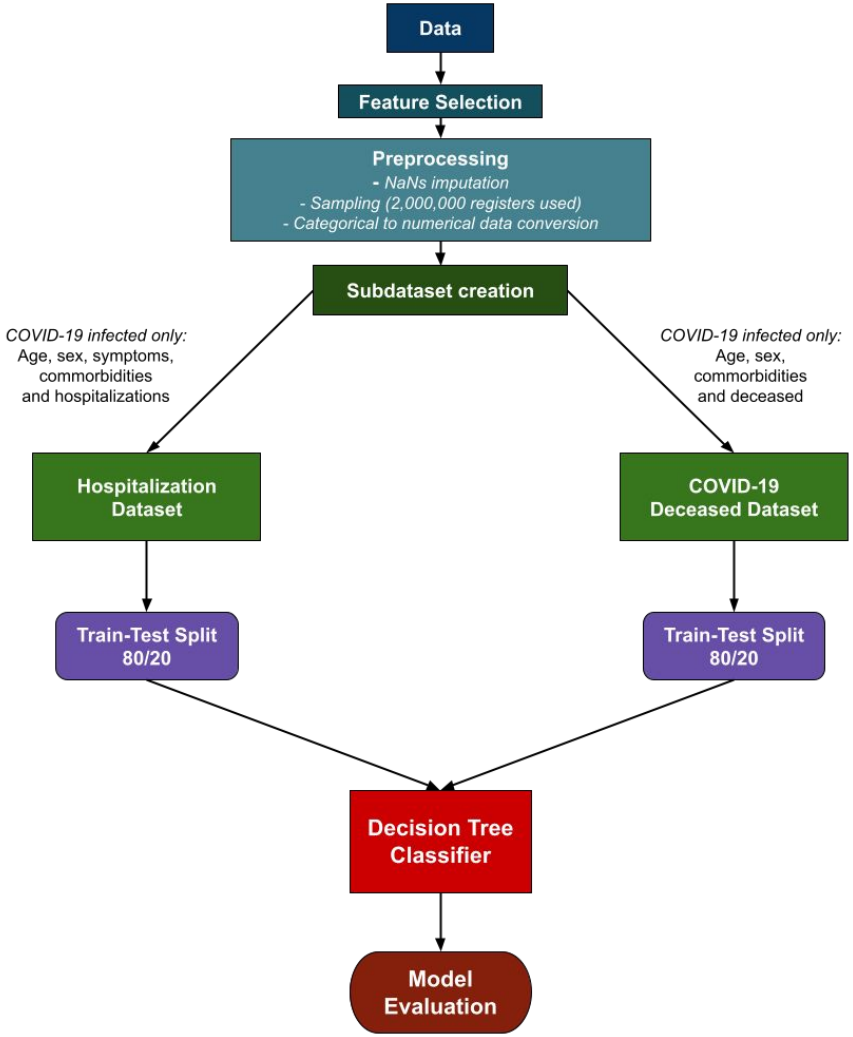
As seen in the data analysis previously made, the **pandemic** has particularly affected the SDGs of **vulnerable communities** and countries. Those vulnerable communities are unlikely to be able to widely apply machine learning algorithms that need computers in the time of need.

Therefore we propose the following **algorithms**:

- Decision Trees

- May be applicable on the field with no electricity nor internet

Solution Architecture



Evaluation

Our dataset is highly unbalanced:

- ≈ 20% of the registers are positive to COVID-19
- ≈ 10% of COVID-19 cases need hospitalization
- ≈ 5% of COVID-19 cases died

Therefore it **could be deceiving to use the accuracy**, in this way the **F1-score, precision** or **recall** could be better indicators.

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Prediction of hospitalization for COVID-19 infected

To know how likely someone who has COVID-19 is going to need to be hospitalized **can prevent diseases**, loss of family economic support and can **help** the people to prevent such a tragic event.

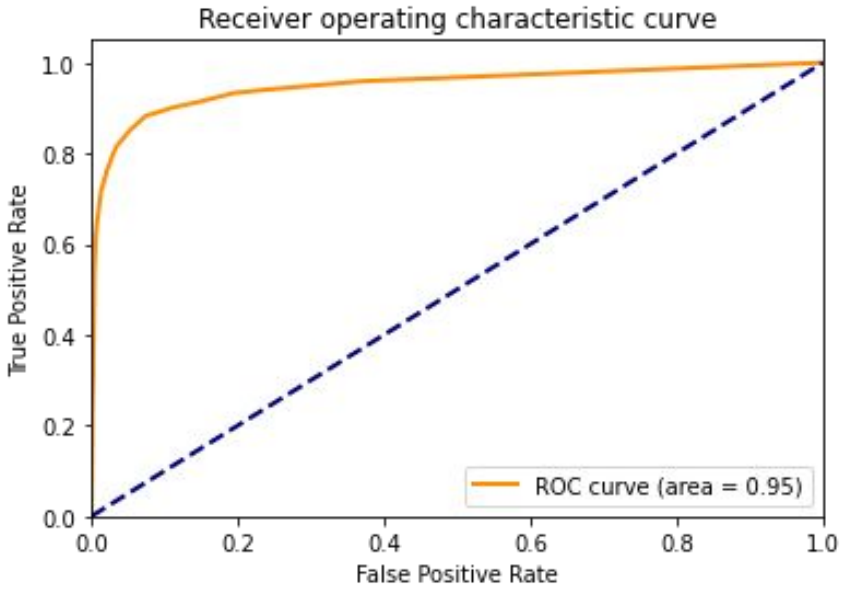
- Inputs: Age, sex, symptoms, comorbidities
- Outputs:
  - Classification
  - Likelihood of requiring hospitalization

Prediction of hospitalization for COVID-19 infected *Results*

Decision tree:

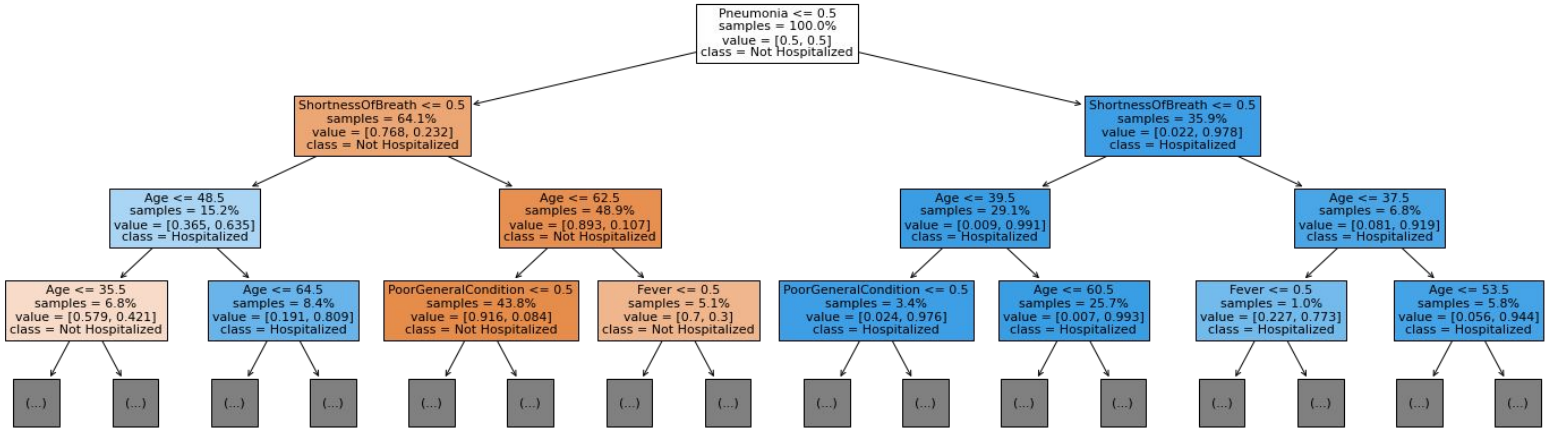
- Depth: 5 levels

Accuracy: 0.922
Precision: 0.570
Recall: 0.881
f1: 0.692



Prediction of hospitalization for COVID-19 infected *Results*

Decision tree:



Prediction of death for COVID-19 infected

Knowing how likely is someone to die if they get infected with COVID-19 can **incentivize** the vulnerable population to take **greater measures** of protection and to avoid reckless actions.

- Inputs: Age, sex, comorbidities
- Outputs:
  - Classification
  - Likelihood of death

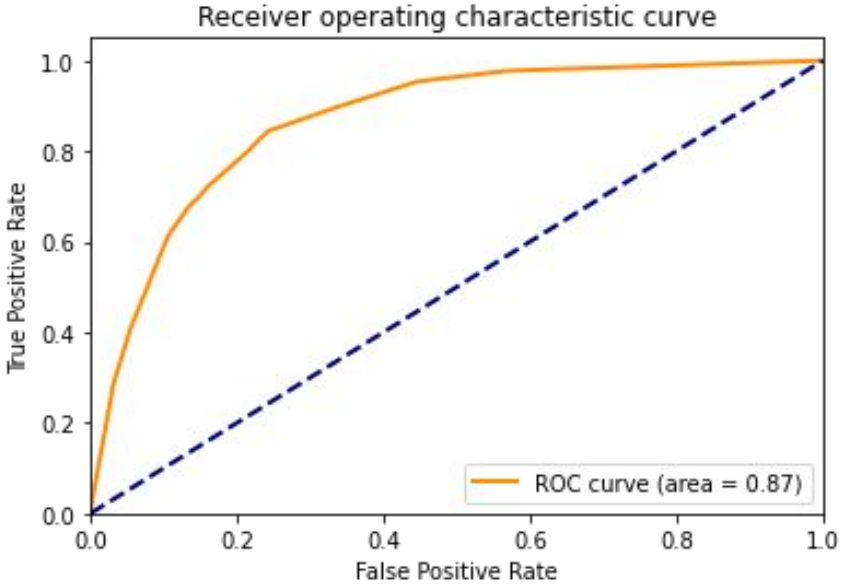
Prediction of death for COVID-19 infected

Results

Decision tree:

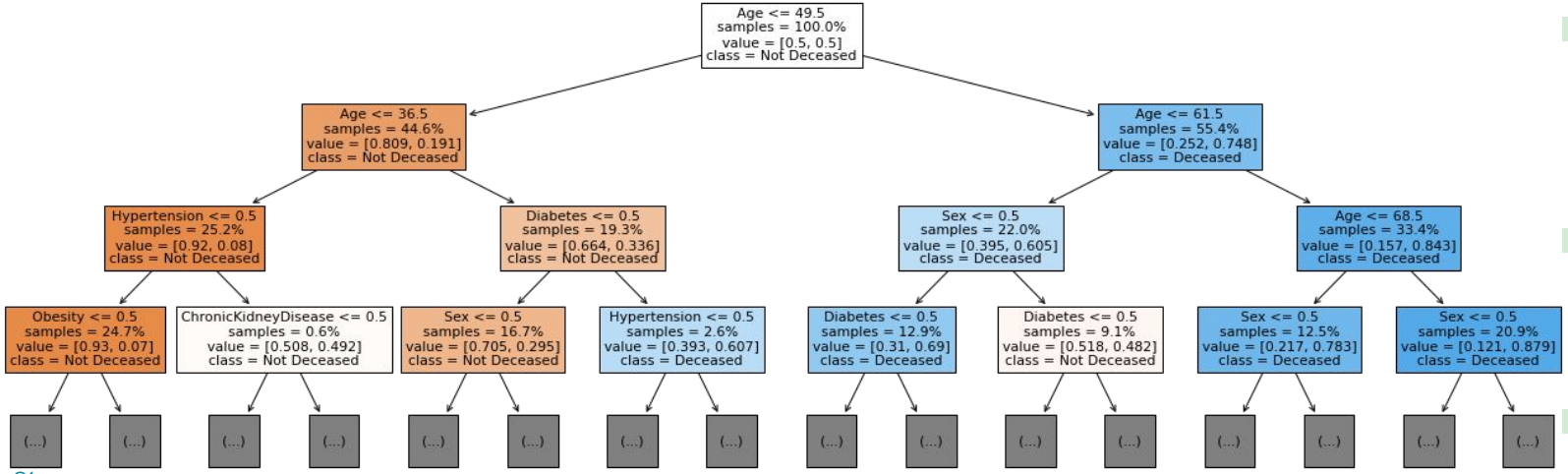
- Depth: 5 levels

Accuracy: 0.769
Precision: 0.147
Recall: 0.830
f1: 0.249



Prediction of death for COVID-19 infected *Results*

Decision tree:



Conclusions

Conclusions

- The pandemic has **affected** countries in a ***different*** way
- **High income countries** seem to be **less affected** than **low income countries** with regards to reaching the SDGs
- **Machine Learning** algorithms can provide **fast, cheap and reliable predictions** to mitigate the effects of the pandemic and help the most vulnerable communities reach the SDGs, particularly SDGs 1, 3, 4 and 8.
- We have **implemented** machine learning models that can be used for both, **vulnerable communities** without access to electricity or internet.

Future Work

Future Work

- This same methodology can be applied to the specific data of **other countries** in order to capture the particular nature of the pandemic in other communities with different characteristics than Mexico City.

- Our models can be tuned in regard with the precision-recall trade-off in order to **adapt to particular circumstances** and medical resources.

- Rural communities can print diagrams of the decision trees to **apply** in **remote localities**.

- The decision trees could be implemented as quizes in order to **attract** a wider **interest** of the population and have a greater impact