**Nasa Space Apps Challenge**

**Team Farachat :**

***BENZMANE Asmaa***

***ESSAKHI Zineb***

***BOUCHOUAL Imane***

Our two proposed algorithms have the answer of all these challenge’s questions :

**Do geographic or temporal patterns from COVID-19 disease mapping reveal insights into human factors that may be related to the spread of the disease?**

**Could human activities that impact the environment play an indirect role in furthering COVID-19 spread? Are certain activities correlated with specific disease presentations or increased severity?**

***Humain factors***

Although there are many, we propose some of the human factors (as explanatory variables)that explain the degree of virus spread and that can be used in our data base, as follows:

1. Measurement of polution
2. The rate of destruction of natural habitats by humans
3. Animal trafficking

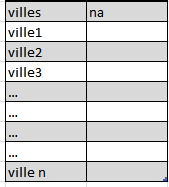
… etc

*Note that all programs are executed in R programming language*

***Approach of the algorithm 1***

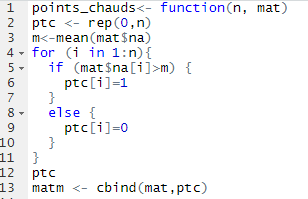
In the first algorithm, we propose an approach that can be adapted to any country and which consists of entering data (the number of daily cases of covid 19 contamination = na) for each city in a given country and which makes it possible to predict hotspots (the cities most affected by the virus) and non-hot spots (the cities least affected by the virus).

1. Our dataset consists of two columns as follows:

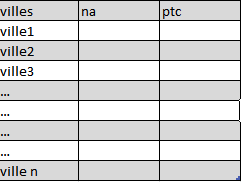


2-We have constructed a simple and efficient function to predict hotspots cities(class 1) and non-hotspots cities(class 0) in a given country. The data set above is adapted to each country, in other words we can apply our  « points\_chauds » function to any country, we just need to have the number of daily cases of covid 19 contamination.

n= number of cities in the given country



1. At the end of the algorithm and using cbind we will modify the table by adding a new column that describes the hotspots cities(1) and non hotspots cities (0).



**Remark :**

In the second algorithm we’ll use the result of that first algo

***Approach of algorithm 2***

The data table we have contains the information for a specific day and we will process it after making a prediction about the nature of each city : hotspot or not (algorithm 1).

Our table is adapted to any country:

It is enough to know the number of explanatory variables p and the number of cities n in a given country and of course the realizations of the explanatory variables.

In this algorithm we're going to tackle the most important step:

Divide our data set into two subsets: one for learning (70%) and one for testing (30%)

After we do this we can apply a method of machine learning ; for example the vector support machines, we'll end up with two classes:

for example for the first variable X1:

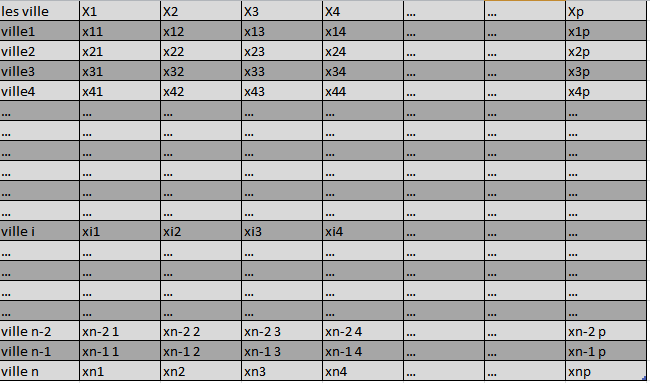
* 1 (the class of polluted points) and 0 (the class of unpolluted points)

And finally we propose to use the cmm matrix whose lens is to see the relationship between each human factor (explanatory variable) and hotspots

for example: the relationship between hotspots and polluted cities. This will allow us to identify the human factors that actually contribute to the spread of Covid 19 virus and therefore act in advance in order to reduce the risk of contaminations.

We will at the end of this algorithm modify our data set by removing the variables that do not influence the spread of the virus i.e. irrelevant variables .

1. First, our dataset consists of p columns and n rows as follows:



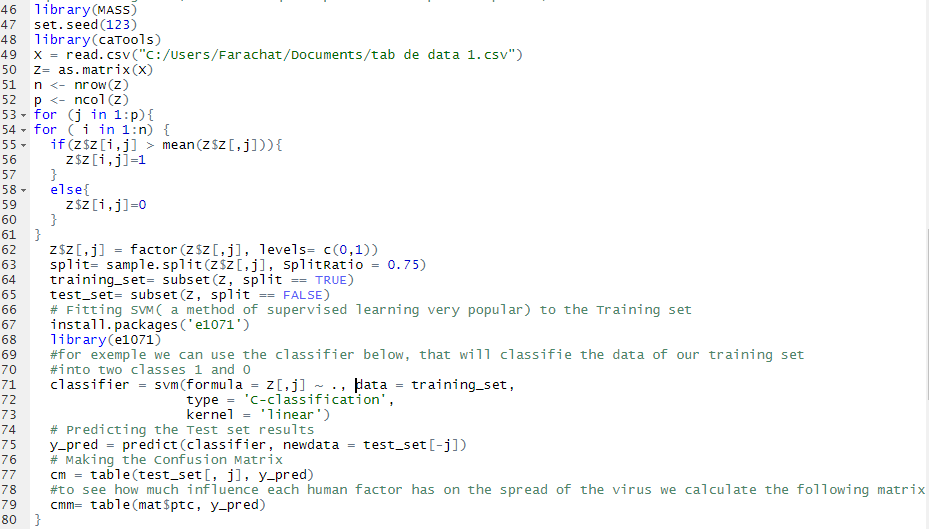
1. 1) Step 1 - We read our data set

2)Step 2 –we divide our dataset in two subsets:

* one for learning
* the other for testing.

Throughout the program, our data set will be noted : X

X1 is a vector that contains the pollution level measurements for all cities in the table and others factors Xj {j=2,…,p} could be any human activities :



* The result of analysing this cmm table will give us the relevant variables, i.e. the variables that really explains the spread of the virus, we assume that there is m relevant variables.
* we will calculate a percentage noted « s » in order to use it for the « COVID-19 Heroes » virtual application we’ve built and which we will use to give allergies to any user of our application :

after having found the m relevant variables we modify the previous table which will contain only these m variables.

Each user is in a specific city so :

for a fixed city i, we will execute the loop below :

