



Munich Personal RePEc Archive

**Data Article. Data mining for the study
of the Epidemic (SARS- CoV-2)
COVID-19: Algorithm for the
identification of patients (SARS-CoV-2)
COVID 19 in Mexico**

Medel-Ramírez, Carlos and Medel-López, Hilario

Universidad Veracruzana / Instituto de Investigaciones y Estudios
Superiores Económicos y Sociales, Universidad Veracruzana /
Instituto de Antropología

3 June 2020

Online at <https://mpra.ub.uni-muenchen.de/100888/>
MPRA Paper No. 100888, posted 30 Jun 2020 10:01 UTC

Data Article.

Data mining for the study of the Epidemic (SARS- CoV-2) COVID-19: Algorithm for the identification of patients (SARS-CoV-2) COVID 19 in Mexico

Authors

Carlos Medel-Ramírez, Hilario Medel-López²

Affiliations

1. Universidad Veracruzana / Instituto de Investigaciones y Estudios Superiores Económicos y Sociales
2. Universidad Veracruzana / Instituto de Antropología

Corresponding author

Carlos Medel-Ramírez (cmedel@uv.mx)

Abstract

The importance of the working document is that it allows analyzing the information and status of the cases associated with (SARS-CoV-2) COVID-19 as open data at the municipal, state and national levels, with a daily registry of patients, according to age, sex, comorbidities, for the condition of (SARS-CoV-2) COVID-19 according to the following characteristics: a) Positive, b) Negatives, c) Suspects. Likewise, it presents information regarding the identification of an outpatient and / or hospitalized patient, attending to their medical development, identifying: a) Recovered, b) Deaths and c) Assets, in Phase 3 and Phase 4, at the national state and municipal level in Mexico, the data analysis is carried out by applying an algorithm of data mining, which provides the information, fast and timely, required for the estimation of Scenarios for Medical Care of the (SARS-CoV-2) COVID-19.

Keywords

(SARS-CoV-2) COVID-19, Algorithm (SARS-CoV-2) COVID-19, Mexico, identification of patients

Specifications Table

Subject	Infectious Diseases
Specific subject area	Information from the Viral Respiratory Diseases Epidemiological Surveillance System for (SARS-CoV-2) COVID-19 in Mexico
Type of data	Table Figure
How data were acquired	Government of Mexico. Health Secretary. Databases Covid-19 México https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico/resource/e8c7079c-dc2a-4b6e-8035-08042ed37165 Instruments: Software Orange Data Mining version 3.25.1 https://orange.biolab.si Make and model and of the instruments used: Algorithm for the identification of patients according to following characteristics: a) Positive, b) Negatives, c) Suspects. Likewise, it presents

	information regarding the identification of an outpatient and / or hospitalized patient, attending to their medical development, identifying: a) Recovered, b) Deaths and c) Assets
Data format	The information is presented in raw in CVS format, the Ministry of Health of Mexico since April 14, 2020 published the cases associated with (SARS-CoV-2) COVID-19 as open data. The data processing corresponds to the records on the epidemic (SARS-CoV-2) COVID-19 at 1 June 2020. The treatment of the information is carried out through the application software for data mining Orange version 3.25.1, in which the algorithm for the analysis of information is filtered to present the current scenario in Mexico of the SARS-CoV-2 (COVID 19).
Parameters for data collection	The information is presented at the municipal, state and national levels, with a daily registry of patients, according to age, sex, comorbidities, for the condition of (SARS-CoV-2) COVID-19 according to the following characteristics: a) Positive, b) Negatives, c) Suspects. Likewise, it presents information regarding the identification of an outpatient and / or hospitalized patient, attending to their medical development, identifying: a) Recovered, b) Deaths and c) Assets.
Description of data collection	This information is filtered to present the current scenario in Mexico of the SARS-CoV-2 (COVID 19) in a fast and timely manner, to support public decision-making in health matters.
Data source location	Institution: Universidad Veracruzana / Instituto de Investigaciones y Estudios Superiores Económicos y Sociales Country: México
Data accessibility	Raw data can be retrieved from the Github repository https://github.com/CMedelR/dataCovid19/edit/master/README.md

Value of the Data

- The Algorithm for the identification of patients (SARS-CoV-2) COVID 19 in Mexico allows to analyze at the municipal, state and national level, the registry of patients, according to age, sex, comorbidities, for condition of (SARS-CoV-2) COVID-19 according to the following characteristics: a) Positive, b) Negative, c) Suspicious, as well as presenting information on the identification of an outpatient and / or hospitalized patient, attending to their medical development, identifying: a) Recovered, b) Deaths and c) Assets, in Phase 3 and Phase 4, in a fast and timely manner, to support public decision-making in health matters.
- Taking into account their strategic roles in public health and researchers can use the data from this study to identify the action scenario for decision-making in the combat of (SARS-CoV-2) COVID 19 in Phase 3 and Phase 4.
- The importance of data analysis is that it allows identifying the cases (SARS-CoV-2) COVID-19 in Mexico is concentrated on a daily for patients of (SARS-CoV-2) COVID-19 and allows preparing action scenarios for making public health policy decisions to combat SARS-CoV-2) COVID-19 in Mexico.

Data Description

The source of information on the number of registered cases of (SARS-CoV-2) COVID-19 at 1 June 2020 for Mexico comes from the website <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico/resource/e8c7079c-dc2a-4b6e-8035-08042ed37165> by the Ministry of Health, with the participation of the National Council for Science and Technology (CONACYT), the Center for Research in Geospatial Information Sciences (CENTROGEO), the National Laboratory for Geo-Intelligence (GEOINT), the Data Laboratory of the National Laboratory for Geointelligence (DataLab), where the registry of COVID-19 cases (SARS-CoV-2) COVID-19 is concentrated, and is the official means of communication and information on the epidemic in Mexico.

The information of the cases (SARS-CoV-2) COVID-19 in Mexico is concentrated on a daily basis since April 19, 2020, communication and official information on the epidemic in Mexico, the data are presented at the municipal, state and national levels, with a daily registry of patients, according to age, sex, comorbidities, for the condition of (SARS-CoV-2) COVID-19 according to the following characteristics: a) Positive, b) Negatives, c) Suspects. Likewise, it presents information regarding the identification of an outpatient and / or hospitalized patient, attending to their medical development, identifying: a) Recovered, b) Deaths and c) Assets. The data processing corresponds to the records on the epidemic (SARS-CoV-2) COVID-19 at May 18, 2020. The treatment of the information is carried out through the application software for data mining and visual programming Orange Data Mining version 3.25.1. Orange Data Mining is a machine learning and data mining suite for data analysis through Python scripting and visual programming. [1]

According to (WHO, 2020) the (SARS-CoV-2) COVID-19 disease pattern presents 4 scenarios identified from the confirmation of Laboratory Diagnosis: a) Not Infected or b) Infected, in this finally, the following categories are observed, taking into account age and specific comorbidities in each case: a) Mild Infection, b) Moderate Infection, c) Severe Infection and d) Critical Infection.

Depending on the category observed in Patients who have a Confirmation of Infected, as in the case of a) or b) it can assume the character of Outpatient, so the strategy is isolation or "quarantine" at home, where the result it is hoped that he will recover. Regarding the Patients who have a Confirmation of Infected, in categories c) and d) they assume the character of Hospitalized Patient, with a probability of requiring care in Intensive Care Units and requiring Intubation, and where it is hoped to save as many patients as possible.

The importance of the research is that it allows identifying the action scenario for making public health policy decisions to combat CO(SARS-CoV-2) COVID-19, since they consider the following states of process in medical treatment, in order to carry out the Estimate of Scenarios for Medical Care of the (SARS-CoV-2) COVID-19 under the following premises of hospital care:

1. A patient with a positive (SARS-CoV-2) COVID-19 laboratory diagnosis can be considered: a) Outpatient, or b) Hospitalized.
2. If the (SARS-CoV-2) COVID-19 Positive patient is Hospitalized, the following should be considered: a) Enter the Intensive Care Unit or b) Do not enter the Intensive Care Unit.
3. If the (SARS-CoV-2) COVID-19 Positive patient is Hospitalized and Entered into the Intensive Care Unit, the following should be considered: a) The patient requires intubation or b) The patient does NOT require intubation.

Methods

The information is presented in raw in CVS format, the Ministry of Health of Mexico. The data processing corresponds to the records on the epidemic (SARS-CoV-2) COVID-19 at 1 June 2020. The treatment of the information is carried out through the application software for data mining Orange version 3.25.1, in which the algorithm for the analysis of information are developed and it is filtered to present the current scenario in Mexico of the SARS-CoV-2 (COVID 19). In this way, the algorithm that is presented allows us to project the requirements for the use of installed infrastructure in the face of the growing requirement for patient care Positive (SARS-CoV-2) COVID-19, allowing the identification of scenarios at the national, state and municipal levels. The construction of the algorithm is based on the following definitions.

Definition 1: Total Patients to consider in Model (SARS-CoV-2) COVID-19.- It is the number of total patients according to the confirmatory laboratory result or not of (SARS-CoV-2) COVID-19).

Be:

TP SARS-CoV-2 i j = Total patients according to (SARS-CoV-2) COVID-19 confirmatory laboratory result
Which consists of:

TP SARS-CoV-2 i j = (P + SARS-CoV-2 i j) + (P- SARS-CoV-2 i j) + (Px SARS-CoV-2 i j), where: i = State, j = Municipality

Of which:

P+ SARS-CoV-2 i j = Patient with a positive (SARS-CoV-2) COVID-19 result in the State, Municipality

P- SARS-CoV-2 i j = Patient with negative (SARS-CoV-2) COVID-19 result in the State, Municipality

Px SARS-CoV-2 i j = Patient with pending confirmation (SARS-CoV-2) COVID-19 in the State, Municipality

Definition 2: Identification of a suspected (SARS-CoV-2) COVID-19 case.- This is the patient who undergoes an initial qualification according to the initial diagnostic characteristics indicated in the case definitions for surveillance by the World Health Organization for primary care of (SARS-CoV-2) COVID-19 cases.

Be:

CsCOVID 19 (SARS-CoV-2) = Patient with initial classification as a suspected case of (SARS-CoV-2) COVID-19

Where:

Cs (SARS-CoV-2) COVID-19 = Cs (SARS-CoV-2) COVID-19 Type 1 + Cs (SARS-CoV-2) COVID-19 Type 2 + Cs (SARS-CoV-2) COVID-19 Type 3

Of which:

According to the World Health Organization, there are 3 categories (identified as Type 1, Type 2 and Type 3) to identify suspected cases of (SARS-CoV-2) COVID-19, defined below:

1. Cs (SARS-CoV-2) COVID-19 Type 1.- Is a patient with acute respiratory disease (fever and at least one sign / symptom of respiratory disease, with no other aetiology that fully explains the clinical presentation and a history of travel or residence in a country / area or territory that reports local transmission of COVID-19 disease during the 14 days prior to the onset of symptoms.
2. Cs (SARS-CoV-2) COVID-19 Type 2.- He is a patient with an acute respiratory disease, who has been in contact with a confirmed or probable COVID-19 case in the last 14 days before the onset of symptoms.

3. Cs (SARS-CoV-2) COVID-19 Type 3.- Is a patient with severe acute respiratory infection (fever and at least one sign / symptom of respiratory illness (eg cough, shortness of breath) and requiring hospitalization and without another etiology that fully explains the clinical presentation.

Definition 3: Total Patients to consider in the (SARS-CoV-2) COVID-19 Model .- It is the number of total patients according to the confirmatory laboratory result or not of (SARS-CoV-2) COVID-19).

Be:

TP SARS-CoV-2 i j = Total patients according to confirmatory laboratory result or not of (SARS-CoV-2) COVID-19

Which consists of:

TP SARS-CoV-2 i j = (P + SARS-CoV-2 i j) + (P- ARS-CoV-2 i j) + (Px ARS-CoV-2 i j) , where: i = State, j = Municipality

Of which:

P + SARS-CoV-2 i j = Patient with a positive (SARS-CoV-2) COVID-19 result in the State, Municipality

P- ARS-CoV-2 i j = Patient with negative (SARS-CoV-2) COVID-19 result in the State, Municipality

Px ARS-CoV-2 i j = Patient with pending confirmation (SARS-CoV-2) COVID-19 in the State, Municipality

Definition 4: Positive Patients for (SARS-CoV-2) COVID-19 i j.- It is the number of patients with laboratory results with positive confirmation for (SARS-CoV-2) COVID-19 i j .

It has:

P + SARS-CoV-2 i j = Patient with a positive (SARS-CoV-2) COVID-19 result in the State, Municipality

Definition 5.- Medical Treatment Strategy for a patient with positive laboratory confirmation for (SARS-CoV-2) COVID-19 i j .- It is the Action Plan in Medical Treatment for a patient with positive laboratory confirmation for SARS-CoV-2 in attention to your degree of infection and comorbidities present that is channeled to determine the Physician.

According to the Strategy of Medical Care required for Patients with a Positive SARS-CoV-2 Result, according to their degree of identified infection, they have the following.

Be:

ET P + SARS-CoV-2 i j = Medical Treatment Strategy P + SARS-CoV-2 i j

The medical treatment for a patient with a positive laboratory result for (SARS-CoV-2) COVID-19, based on the Medical Treatment Strategy (ETM P + SARS-CoV-2 ij), based on his degree of infection and present comorbidities, poses two action scenarios : i) Outpatient (SARS-CoV-2) COVID-19 patient or ii) Hospitalized (SARS-CoV-2) COVID-19 patient.

Be:

i) Outpatient COVID19 patient.

P + SARS-CoV-2 i j Outpatient = Positive (SARS-CoV-2) COVID-19 with Outpatient mode in the State, Municipality

ii) COVID19 Patient Hospitalized.

P + SARS-CoV-2 i j Hospitalized = Positive (SARS-CoV-2) COVID-19 with modality Hospitalized in the State, Municipality

where:

Depending on the degree of infection (I1, I2 or I3), the Hospitalized (SARS-CoV-2) COVID-19 Patient may require: i) Access to the Intensive Care Area without Intubation or ii) Access to the Intensive Care Area with Intubation.

Definition 6.- Patients with a Positive (SARS-CoV-2) COVID-19 Result Hospitalized with Access to the Intensive Care area.- It is the number of Patients with a Positive SARS-CoV-2 Result Hospitalized with Access to the Intensive Care area, according to its degree of infection.

Be:

$P + \text{SARS-CoV-2 } i j \text{ Hospital Intensive Care} = \text{Positive (SARS-CoV-2) COVID-19 with modality Hospitalized in the State, Municipality}$

Definition 7.- Patients with a positive (SARS-CoV-2) COVID-19 result Hospitalized with access to the Intensive Care Area with Intubation.- It is the number of Patients with a Positive (SARS-CoV-2) COVID-19 Result Hospitalized with Access to the Intensive Care area with Intubation.

Be:

$P + \text{SARS-CoV-2 } i j \text{ Hospital Intensive Care with Intubation} = \text{Positive (SARS-CoV-2) COVID-19 with Hospitalized modality and intubation in the State, Municipality.}$

Definition 8.- $P + \text{SARS-CoV-2 } i j \text{ Deaths.}$ - Deaths of Patients with a positive result for SARS-CoV-2. Deaths are all those positive to (SARS-CoV-2) COVID-19 where one is indicated in the data record (DATE_DEF other than the value "99-99-9999").

Definition 9.- (SARS-CoV-2) COVID-19 case fatality rate.- It is the proportion of people who die from (SARS-CoV-2) COVID-19 among the Patients with a positive (SARS-CoV-2) COVID-19 result in a given period and area.

Be:

$TL \text{ SARS-CoV-2 } i j = (\text{SARS-CoV-2) COVID-19 case fatality rate}$

Where:

$(\text{SARS-CoV-2) COVID-19 case fatality rate} = [(\text{Deaths of Patients with a Positive (SARS-CoV-2) COVID-19 Result in the State or Municipality}) / (\text{Total of Patients with a Positive (SARS-CoV-2) COVID-19 result in the State or Municipality})] \times 100$

Of which:

$DP + \text{SARS-CoV-2 } i j = \text{Deaths of Patients with a positive (SARS-CoV-2) COVID-19 result in the State / Municipality}$

And:

$P + \text{SARS-CoV-2 } i j = \text{Total Patients with a positive (SARS-CoV-2) COVID-19 result in the State / Municipality}$

So:

$TL \text{ SARS-CoV-2 } i j = [D P + \text{SARS-CoV-2 } i j / P + \text{SARS-CoV-2 } i j] \times 100$

The data processing corresponds to the records on the epidemic (SARS-CoV-2) COVID-19 at 1 June 2020. The treatment of the information is carried out through the application software for data mining Orange version 3.25.1, in which the algorithm for the information analysis are developed. (See Table 1 and Figure 1, below). According to information from the Ministry of Health, the following records are available at the national level:

1. The number of patients with a positive (SARS-CoV-2) COVID-19 result is 94,435 of which: a) 60,828 are Care Outpatients and b) 32,607 are Hospitalized patients.
2. Of the 32,607 Hospitalized (SARS-CoV-2) COVID-19 Positive patients: a) 3,074 patients enter the Intensive Care Unit; while b) 29,499 patients do not enter the Intensive Care Unit.
3. Only 1,712 Hospitalized (SARS-CoV-2) COVID-19 Positive patients admitted to the Intensive Care Unit required intubation; while 1,362 patients did not require intubation.
4. Likewise, to date 10,167 deaths from positive (SARS-CoV-2) COVID-19 patients have been registered nationwide, of which 1,595 deaths corresponded to Positive (SARS-CoV-2) COVID-19

patients who were in Intensive Care Intubation and 7,510 corresponded to (SARS-CoV-2) COVID-19 Positive patients who were not in Intensive Care.

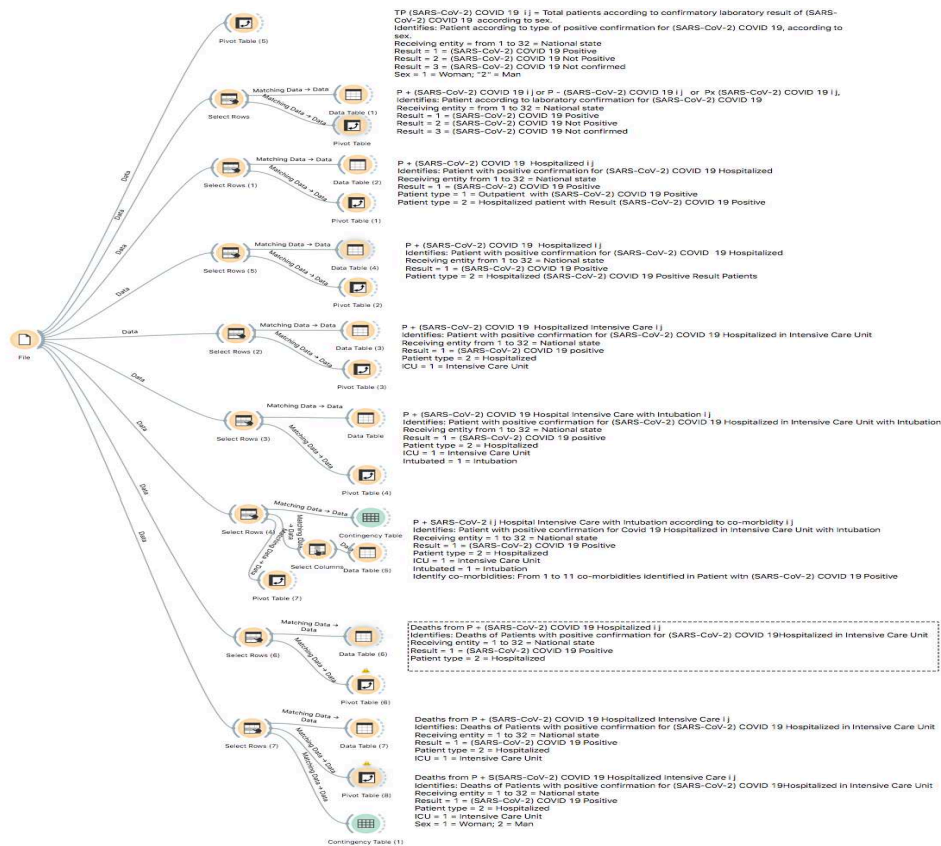
- The number of women with a positive (SARS-CoV-2) COVID-19 diagnosis is 40,693; while the number of men is 52,742. This information is presented in the following table 1.

Table 1. Total number of cases in Mexico as of June 1, 2020, According to Sex and Result at (SARS-CoV-2) COVID-1

Result	Sex		Total
	Women	Men	
Positive (SARS-CoV-2) COVID-1	40,693	52,742	93,435
No positive (SARS-CoV-2) COVID-1	78,702	71,455	150,157
Pending result	18,572	19,925	38,497
Total	137,967	144,122	282,089

Source: Government of Mexico. Health Secretary. Information from the Epidemiological Surveillance System for Viral Respiratory Diseases as of June 1, 2020.

Figure 1. Algorithm for the identification of patients (SARS-CoV-2) COVID 19 in Mexico in Orange Data Mining version 3.25.1



Source: Self made. With information provided by the Ministry of Health as of 1 June 2020 and Orange Data Mining version 3.25.1

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

References

[1] Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research* 14(Aug): 2349–2353. <https://dl.acm.org/doi/pdf/10.5555/2567709.2567736>

[2] Government of Mexico. Health Secretary. Databases Covid-19 México. <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico/resource/e8c7079c-dc2a-4b6e-8035-08042ed37165>

[3] Software Orange Data Mining version 3.25.1 <https://orange.biolab.si>

[4] World Health Organization. (2020). Laboratory testing for coronavirus disease (COVID-19) in suspected human cases. Interim guidance. 19 March 2020. Recuperado de: <https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117>

Reference to a dataset:

[6] Raw data can be retrieved from the Github repository <https://github.com/CMedeIR/dataCovid19/>