



Title: Caracterización de casos SARS-CoV-2 en México utilizando minería de datos

Authors: LUNA-RAMÍREZ, Enrique, SORIA-CRUZ, Jorge, VELARDE-MARTÍNEZ, Apolinar y TAYA-ACOSTA, Edgar Aurelio

Editorial label ECORFAN: 607-8695

BECORFAN Control Number: 2020-05

BECORFAN Classification (2020): 111220-0005

Pages: 8

RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.

143 – 50 Itzopan Street
La Florida, Ecatepec Municipality
Mexico State, 55120 Zipcode
Phone: +52 1 55 6159 2296
Skype: ecorfan-mexico.s.c.
E-mail: contacto@ecorfan.org
Facebook: ECORFAN-México S. C.

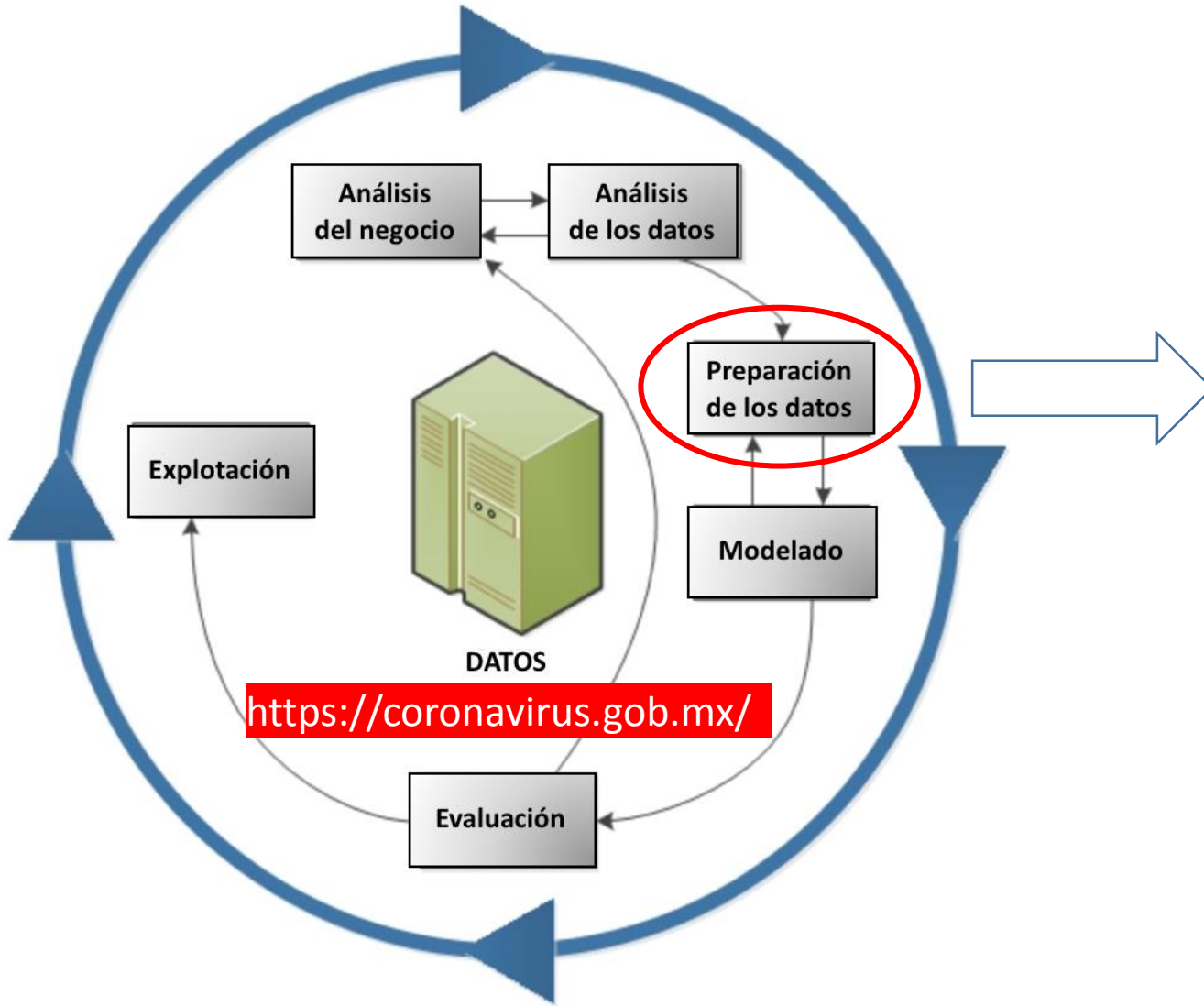
Twitter: @EcorfanC

www.ecorfan.org

Holdings

Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic
Spain	El Salvador	Republic
Ecuador	Taiwan	of Congo
Peru	Paraguay	Nicaragua

Introducción



CRISP-DM

WEKA: Herramienta para *Machine Learning*

<https://www.cs.waikato.ac.nz/ml/weka/>

Relation: 200924COVID19MEXICO v3
Instances: 1048575
Attributes: 21
Sum of weights: 1048575

Attributes

All None Invert Pattern

No.	Name
5	<input type="checkbox"/> TIPO_PACIENTE
6	<input type="checkbox"/> FALLECIO
7	<input type="checkbox"/> INTUBADO
8	<input type="checkbox"/> NEUMONIA
9	<input type="checkbox"/> EDAD
10	<input type="checkbox"/> EMBARAZO
11	<input type="checkbox"/> DIABETES
12	<input type="checkbox"/> EPOC
13	<input type="checkbox"/> ASMA
14	<input type="checkbox"/> INMUSUPR
15	<input type="checkbox"/> HIPERTENSION
16	<input type="checkbox"/> CARDIOVASCULAR
17	<input type="checkbox"/> OBESIDAD
18	<input type="checkbox"/> RENAL_CRONICA
19	<input type="checkbox"/> TABAQUISMO
20	<input type="checkbox"/> RESULTADO
21	<input type="checkbox"/> UCI

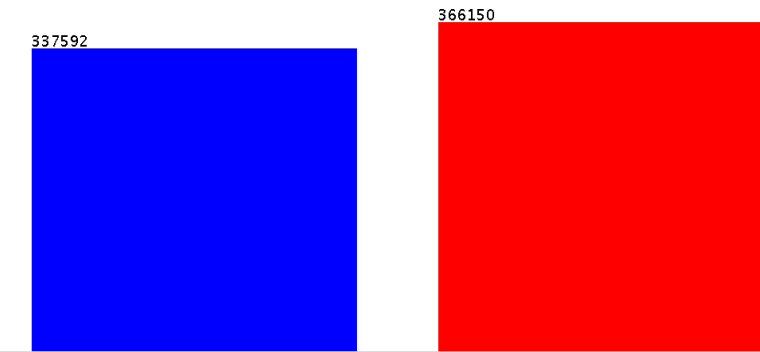
Variables significativas

Metodología

Segmentación de casos positivos
por sexo

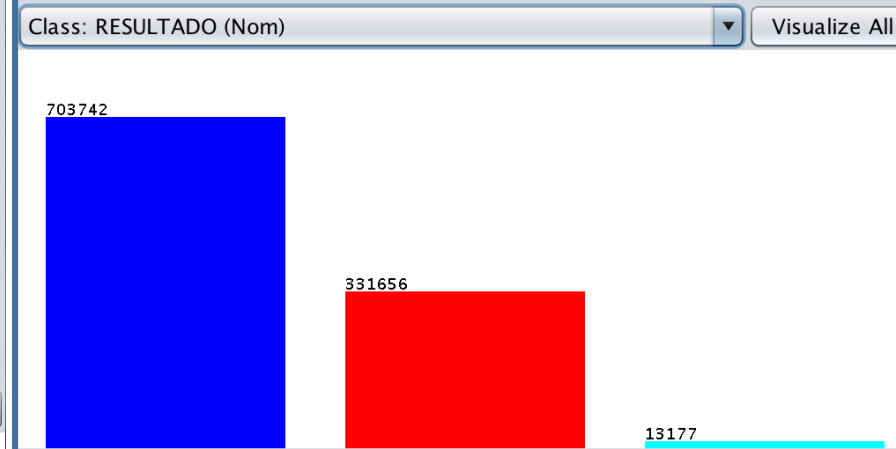
No.	Label	Count	Weight
1	Mujer	337592	337592.0
2	Hombre	366150	366150.0

Class: SEXO (Nom) Visualize All



No.	Label	Count	Weight
1	Positivo	703742	703742.0
2	No Positivo	331656	331656.0
3	Pendiente	13177	13177.0

Class: RESULTADO (Nom) Visualize All

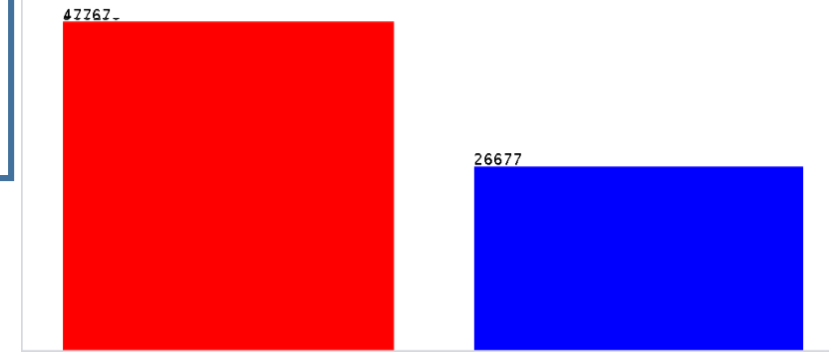


Identificación de los casos positivos
(1,048,575 casos publicados)

Segmentación de decesos
por sexo

No.	Label	Count	Weight
1	Hombre	47767	47767.0
2	Mujer	26677	26677.0

Class: FALLECIO (Nom) Visualize All



Metodología

Trabajo relacionado

Meng *et al.* (2020): muestra de 366 pacientes COVID-19 → red neuronal densamente conectada orientada a predecir la probabilidad de que un paciente avance a un estado de alto riesgo

Qjidaa *et al.* (2020): muestra de 566 personas en áreas rurales → DSS orientado al diagnóstico temprano de la enfermedad COVID-19

Roy *et al.* (2020): repositorio abierto → plataforma on-line orientada a predecir casos positivos y decesos por COVID-19

=== Run information ===

Scheme: weka.classifiers.rules.PART -C 0.25 -M 2
Relation: 0 200924COVID19 MEXICO Positivos-weka.filters
Instances: 703742
Attributes: 18

ENTIDAD_UM
SEXO
TIPO_PACIENTE
FALLECIO
INTUBADO
NEUMONIA
EDAD
EMBARAZO
DIABETES
EPOC
ASMA
INMUSUPR
HIPERTENSION
CARDIOVASCULAR
OBESIDAD
RENAL_CRONICA
TABAQUISMO
RESULTADO

Test mode: 10-fold cross-validation

=== Summary ===

Correctly Classified Instances	648100	92.0934 %
Incorrectly Classified Instances	55642	7.9066 %
Kappa statistic	0.5343	
Mean absolute error	0.101	
Root mean squared error	0.2373	
Relative absolute error	53.3687 %	
Root relative squared error	77.1552 %	
Total Number of Instances	703742	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0.970	0.490	0.944	0.970	0.956	0.540
	0.510	0.030	0.665	0.510	0.577	0.540
Weighted Avg.	0.921	0.442	0.914	0.921	0.916	0.540

=== Confusion Matrix ===

a	b	<-- classified as
610144	19154	a = No
36488	37956	b = Si

Modelo 1: Clasificación por fallecidos en México

Resultados

TIPO_PACIENTE = Ambulatorio AND
NEUMONIA = No AND
EDAD <= 55 AND
RENAL_CRONICA = No AND
DIABETES = No AND
EPOC = No AND
EDAD <= 43 AND
INMUSUPR = No: No (295180.0/336.0)

INTUBADO = Si AND
EDAD > 48 AND
EDAD > 61 AND
ENTIDAD_UM = Edo. de Mexico AND
ASMA = No AND
EPOC = No AND
SEXO = Hombre: Si (632.0/56.0)

Muestra de dos reglas en el Modelo 1

Primera regla involucra 42% de los casos positivos (295,180 de 703,742 casos con solo el 0.1 % de casos mal clasificados.

Una persona con COVID-19, sin mayores padecimientos como neumonía, una enfermedad renal crónica, diabetes y EPOC, saldrá adelante de la enfermedad siempre que no exceda los 43 años.

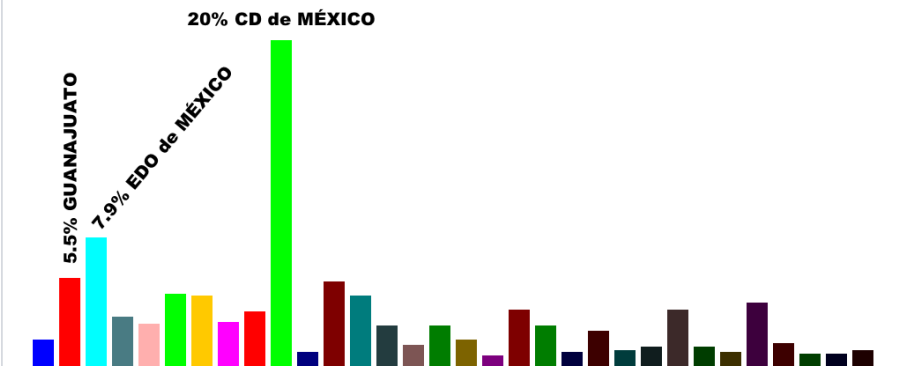
Segunda regla indica que cuando un paciente (hombre) mayor a 48 años es intubado, su enfermedad se complicará hasta la muerte, aún cuando no tenga otros padecimientos como asma o EPOC. (9% de casos mal clasificados)

Resultados

Name: ENTIDAD_UM
Missing: 0 (0%)
Distinct: 32
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Hidalgo	11745	11745.0
2	Guanajuato	38773	38773.0
3	Edo. de Mexico	55899	55899.0
4	San Luis Potosi	21735	21735.0
5	Baja California	18466	18466.0
6	Veracruz	31289	31289.0
7	Puebla	30943	30943.0
8	Michoacan	19085	19085.0
9	Sonora	23940	23940.0

Class: ENTIDAD_UM (Nom) Visualize All



Segmentación de casos positivos por Estado

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 200924COVID19 Guanajuato Positivos-weka.filters
Instances: 38773
Attributes: 16

- SEXO
- TIPO_PACIENTE
- FALLECIO
- INTUBADO
- NEUMONIA
- EDAD
- EMBARAZO
- DIABETES
- EPOC
- ASMA
- INMUSUPR
- HIPERTENSION
- CARDIOVASCULAR
- OBESIDAD
- RENAL_CRONICA
- TABAQUISMO

Test mode: 10-fold cross-validation

=== Summary ===

Correctly Classified Instances	36501	94.1403 %
Incorrectly Classified Instances	2272	5.8597 %
Kappa statistic	0.4379	
Mean absolute error	0.0839	
Root mean squared error	0.2098	
Relative absolute error	63.177 %	
Root relative squared error	81.4316 %	
Total Number of Instances	38773	

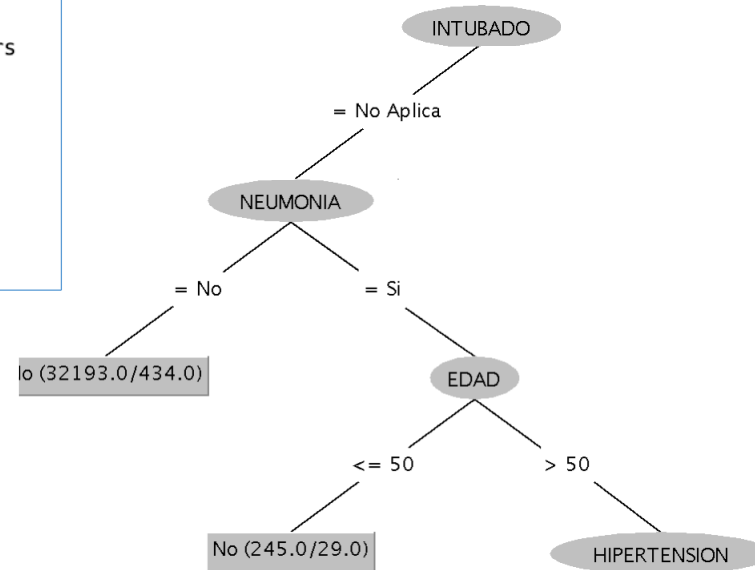
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
Weighted Avg.	0.986	0.642	0.952	0.986	0.969	0.462
	0.358	0.014	0.669	0.358	0.466	0.462
	0.941	0.598	0.932	0.941	0.933	0.462

=== Confusion Matrix ===

a	b	<-- classified as
35510	491	a = No
1781	991	b = Si

Modelo 2: Clasificación por fallecidos en Gto



Mejor regla en el Modelo 2

REGLA:

En el Estado de Gto., un paciente ambulatorio (INTUBADO = No Aplica), sin un padecimiento de neumonia, no estará en riesgo de muerte

=== Run information ===

Scheme: weka.classifiers.rules.PART -C 0.25 -M 2
 Relation: 200924COVID19 Guanajuato Positivos-weka.filters
 Instances: 38773
 Attributes: 16
 SEXO
 TIPO_PACIENTE
 FALLECIO
 INTUBADO
 NEUMONIA
 EDAD
 EMBARAZO
 DIABETES
 EPOC
 ASMA
 INMUSUPR
 HIPERTENSION
 CARDIOVASCULAR
 OBESIDAD
 RENAL_CRONICA
 TABAQUISMO
 Test mode: 10-fold cross-validation

=== Summary ===

Correctly Classified Instances	36409	93.903 %
Incorrectly Classified Instances	2364	6.097 %
Kappa statistic	0.4369	
Mean absolute error	0.0814	
Root mean squared error	0.2128	
Relative absolute error	61.2752 %	
Root relative squared error	82.6083 %	
Total Number of Instances	38773	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
	0.983	0.626	0.953	0.983	0.968	0.453
Weighted Avg.	0.374	0.017	0.623	0.374	0.467	0.453
	0.939	0.583	0.930	0.939	0.932	0.453

=== Confusion Matrix ===

a	b	<-- classified as
35373	628	a = No
1736	1036	b = Si

Modelo 3: Clasificación por fallecidos en Gto

Resultados

(a)	(b)
INTUBADO = No Aplica AND NEUMONIA = No AND EPOC = No AND EDAD <= 50: No (24198.0/91.0)	INTUBADO = No Aplica AND NEUMONIA = No AND EPOC = No AND EDAD <= 65: No (5723.0/144.0)
(c)	(d)
INTUBADO = No Aplica AND NEUMONIA = No AND SEXO = Mujer AND TABAQUISMO = No AND INMUSUPR = No AND RENAL_CRONICA = No AND EPOC = No: No (984.0/50.0)	INTUBADO = Si AND EDAD > 42 AND EPOC = No AND EDAD <= 80 AND TABAQUISMO = No AND DIABETES = Si AND INMUSUPR = No AND OBESIDAD = No: Si (136.0/8.0)

Muestra de cuatro reglas en el Modelo 3

Las reglas de los incisos (a), (b) y (c), en su conjunto, en buena medida equivalen a la regla del Modelo 2

(a) adiciona que la persona no debe exceder de 50 años, lo que domina a la condición del inciso (b)

(c) adiciona que la persona debe ser mujer que no fuma y no padece inmunosupresión, ni EOPC, ni una enfermedad renal crónica

Conclusiones

- ❖ Construcción de diferentes modelos para extraer conocimiento de los datos que publica el Gobierno Federal de México sobre la enfermedad COVID-19
- ❖ Los modelos fueron construidos básicamente como modelos de clasificación, mediante el algoritmo C4.5, y evaluados con la técnica de validación cruzada (cross-validation)
- ❖ En los modelos, se observaron diversas reglas que en buena medida pueden ser utilizadas para predecir escenarios futuros sobre la evolución de la enfermedad COVID-19
- ❖ Como trabajo futuro, se tiene considerado construir modelos de predicción más robustos mediante el uso de redes neuronales artificiales

Referencias

Chauhan, H. & Chauhan, A. (2013). "Implementation of decision tree algorithm c4.5".
Intnl. Journal of Scientific and Research Publications, Volume 3, Issue 10, October 2013).

Meng, L., Dong, D., Li, L., Niu, M., Bai, Y., Wang, M., Qiu, X., Zha, Y. & Tian, J. (2020).
"A Deep Learning Prognosis Model Help Alert for COVID-19 Patients at High Risk of Death: A Multi-center Study".
IEEE Journal of Biomedical and Health Informatics, accepted for future publication,
but has not been edited in an issue of this journal.

Qjidaa, M., Mechbal, Y., Ben-fares, A., Amakdouf, H., Maaroufi, M., Alami, B. & Qjidaa, H. (2020).
"Early detection of COVID19 by deep learning transfer Model for populations in isolated rural areas".
Proceedings of the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), pp. 1-5.

Roy, S., Pal, M.N., Bhattacharyya, S. & Lahiri, S. (2020).
"Implementation of an Informative Website – "Covid19 Predictor", Highlighting COVID-19 Pandemic Situation in India".
Proceedings of the 2020 International IOT, Electronics and Mechatronics Conference (IEMTRONICS).



ECORFAN®

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BECORFAN is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- (www.ecorfan.org/ booklets)