

After-Covid19

O papel da Enxeñería Civil e da Edificación nun novo escenario

MODELOS DE MACHINE LEARNING APLICADOS A LA COMPRESION DE LA COVID-19

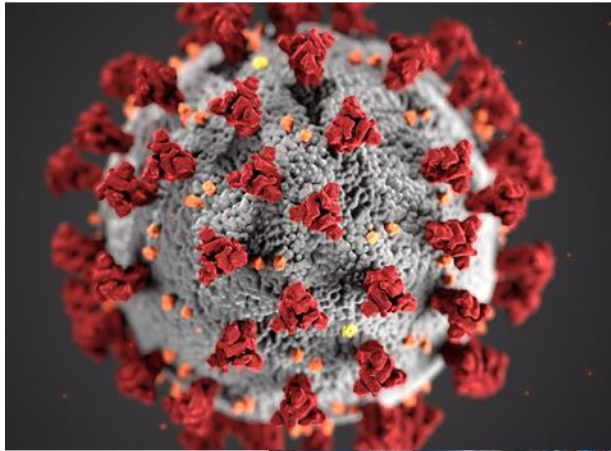
Manuel Andrés Soage Quintáns & GMNI

21 de maio de 2020

Modelos de Machine Learning aplicados a la compresión de la COVID-19

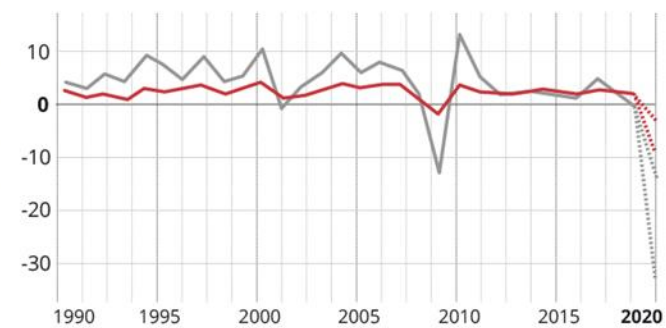
- Introducción.
- Base de datos Korea.
- SpeeWise Machine Learning (QRI AI LLc).
- Análisis de datos. Compresión del problema.
- Entrenamiento de Modelo de ML.
- Explicación del proceso infeccioso en Korea.
- Predicciones y otras aplicaciones.

Modelos de Machine Learning aplicados a la comprensión de la COVID-19



EL DESPLOME DEL COMERCIO MUNDIAL

Variaciones en porcentaje



- Crecimiento del PIB mundial
- Volumen de los intercambios globales
- Previsiones

LAVANGUARDIA Fuente: OMC

Modelos de Machine Learning aplicados a la comprensión de la COVID-19

Web con bases de datos gratuitas

Herramienta SaaS (Software as a Service) de la Compañía QRI



La base de datos "PatientInfor.csv" contiene datos de 3.338 pacientes, y 18 tipos de atributos como edad, país de origen o fecha de diagnóstico entre otros. 38.086 datos totales.

Base de datos de infectados por SARS-CoV-2 en Korea del Sur

Context

COVID-19 has infected more than 10,000 people in South Korea. KCDC (Korea Centers for Disease Control & Prevention) announces the information of COVID-19 quickly and transparently. We make a structured dataset based on the report materials of KCDC and local governments. Also, we analyze and visualize the data using various data mining or visualization techniques.

patient_id	global_num	sex	birth_year	age	country	province	city	disease	infection_case	infection_order	infected_by	contact_number	syn
1000000001	2	male	1964	50s	Korea	Seoul	Gangseo-gu	overseas inflow		1		75	
1000000002	5	male	1987	30s	Korea	Seoul	Jungnang-gu	overseas inflow		1		31	
1000000003	6	male	1964	50s	Korea	Seoul	Jongno-gu	contact with patient		2	2002000001	17	
1000000004	7	male	1991	20s	Korea	Seoul	Mapo-gu	overseas inflow		1		9	
1000000005	9	female	1992	20s	Korea	Seoul	Seongbuk-gu	contact with patient		2	1000000002	2	
1000000006	10	female	1966	50s	Korea	Seoul	Jongno-gu	contact with patient		3	1000000003	43	
1000000007	11	male	1995	20s	Korea	Seoul	Jongno-gu	contact with patient		3	1000000003	0	
1000000008	13	male	1992	20s	Korea	Seoul	etc	overseas inflow		1		0	
1000000009	19	male	1983	30s	Korea	Seoul	Songpa-gu	overseas inflow		2		68	
1000000010	21	female	1960	60s	Korea	Seoul	Seongbuk-gu	contact with patient		3	1000000003	6	
1000000011	23	female	1962	50s	China	Seoul	Seodaemun-gu	overseas inflow		1		23	
1000000012	24	male	1992	20s	Korea	Seoul	etc	overseas inflow		1		0	
1000000013	29	male	1938	80s	Korea	Seoul	Jongno-gu	contact with patient		4	1000000017	117	
1000000014	30	female	1952	60s	Korea	Seoul	Jongno-gu	contact with patient		5	1000000013	27	
1000000015	40	male	1943	70s	Korea	Seoul	Seongdong-gu	Seongdong-gu APT				8	
1000000016	56	male	1945	70s	Korea	Seoul	Jongno-gu	contact with patient		4	1000000017		
1000000017	83	male	1944	70s	Korea	Seoul	Jongno-gu	contact with patient		3	1000000003		
1000000018	111	male	2000	20s	Korea	Seoul	etc	etc					
1000000019	112	female	1941	70s	Korea	Seoul	Jongno-gu	contact with patient		5	1000000021		
1000000020	121	female	1944	70s	Korea	Seoul	Seongdong-gu	Seongdong-gu APT			1000000015		

PatientInfo.csv (356.2 KB) 18 of 18 columns Views

patient_id	# global_num	sex	# birth_year
the ID of the patient	the number given by KCDC	the sex of the patient	the birth year of the patient

female 55%

male 43%

Other (1) 2%

Modelos de Machine Learning aplicados a la comprensión de la COVID-19

“PatientInfor.csv” 3.338 pacientes, y 18 tipos de atributos.

Tratamiento de la base de datos eliminando atributos que no tienen causalidad, en la información, datos erróneos, etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	patient_id	global_num	sex	birth_year	age	country	province	city	disease	infection_case	infection_order	infected_by	contact_number
2	1000000001	2	male	1964	50s	Korea	Seoul	Gangseo-gu		overseas inflow			75
3	1000000002	5	male	1987	30s	Korea	Seoul	Jungnang-gu		overseas inflow	1		31

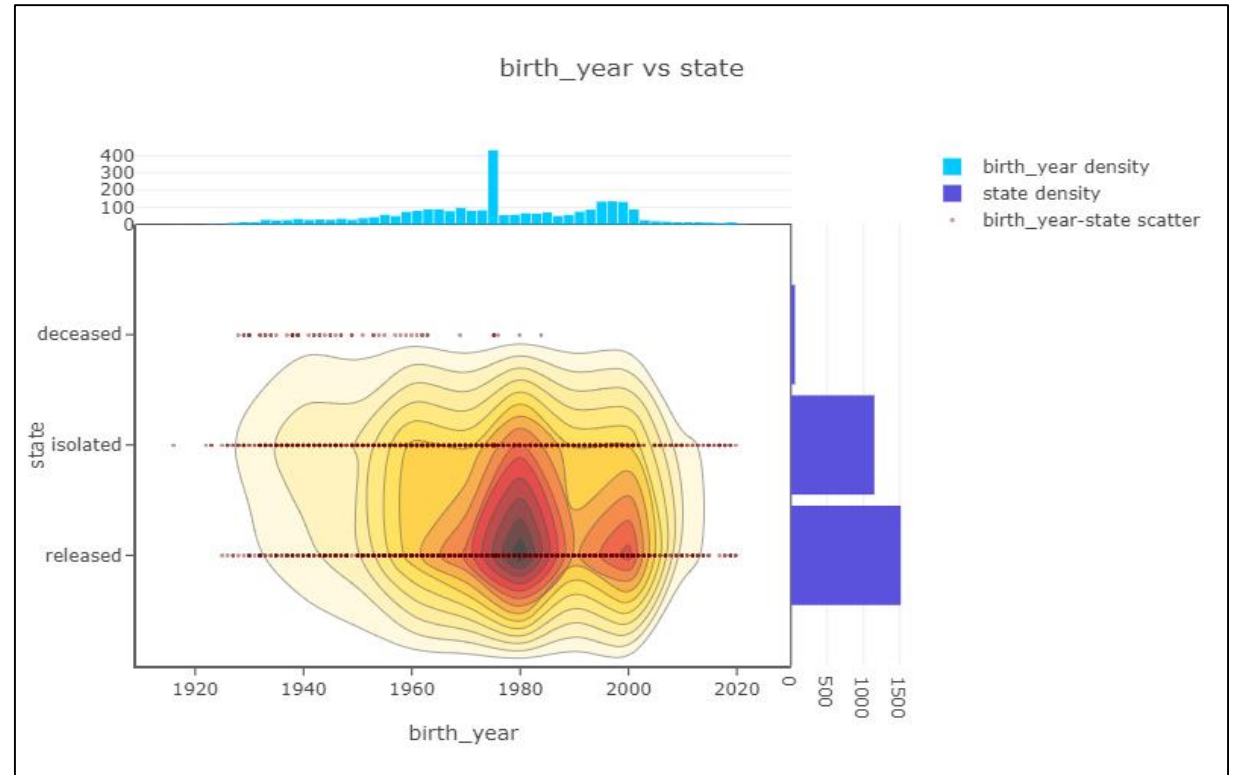
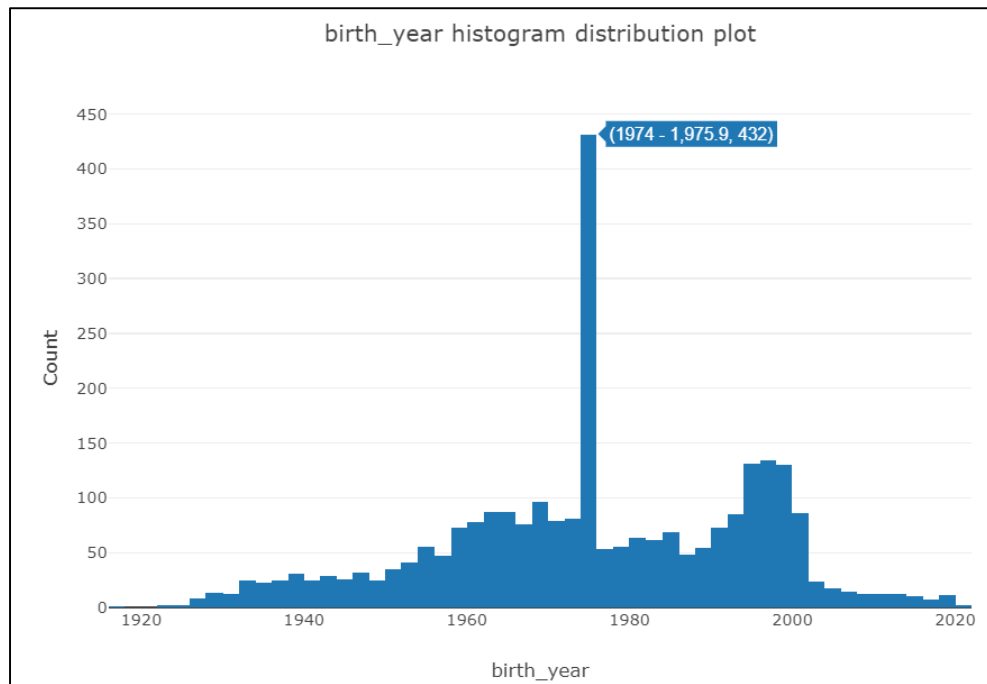
	A	B	C	D	E	F	G	H
1	sex	birth_year	country	province	city	infection_case	confirmed_date	state
2	male	1964	Korea	Seoul	Gangseo-gu	overseas inflow	1/23/2020	released
3	male	1987	Korea	Seoul	Jungnang-gu	overseas inflow	1/30/2020	released
4	male	1964	Korea	Seoul	Jongno-gu	contact with patient	1/30/2020	released
5	male	1991	Korea	Seoul	Mapo-gu	overseas inflow	1/30/2020	released
6	female	1992	Korea	Seoul	Seongbuk-gu	contact with patient	1/31/2020	released
7	female	1966	Korea	Seoul	Jongno-gu	contact with patient	1/31/2020	released
8	male	1995	Korea	Seoul	Jongno-gu	contact with patient	1/31/2020	released
9	male	1992	Korea	Seoul	etc	overseas inflow	2/2/2020	released
10	male	1983	Korea	Seoul	Songpa-gu	overseas inflow	2/5/2020	released
11	female	1960	Korea	Seoul	Seongbuk-gu	contact with patient	2/5/2020	released
12	female	1962	China	Seoul	Seodaemun-gu	overseas inflow	2/6/2020	released
13	male	1992	Korea	Seoul	etc	overseas inflow	2/7/2020	released
14	male	1938	Korea	Seoul	Jongno-gu	contact with patient	2/16/2020	released
15	female	1952	Korea	Seoul	Jongno-gu	contact with patient	2/16/2020	released
16	male	1943	Korea	Seoul	Seongdong-gu	Seongdong-gu APT	2/19/2020	isolated
17	male	1945	Korea	Seoul	Jongno-gu	contact with patient	2/19/2020	released
18	male	1944	Korea	Seoul	Jongno-gu	contact with patient	2/20/2020	released
19	male	2000	Korea	Seoul	etc	etc	2/20/2020	released
20	female	1941	Korea	Seoul	Jongno-gu	contact with patient	2/20/2020	released
21	female	1944	Korea	Seoul	Seongdong-gu	Seongdong-gu APT	2/20/2020	released

“Patient_Korea.csv” 3.312 pacientes y 8 atributos: sexo, año de nacimiento, país de origen, provincia, ciudad, tipo de infección, fecha de confirmación y estado.

Modelos de Machine Learning aplicados a la comprensión de la COVID-19



“Patient_Korea.csv” 3.312 pacientes y 8 atributos: sexo, año de nacimiento, país de origen, provincia, ciudad, tipo de infección, fecha de confirmación y estado. (70 fallecimientos del 20 de enero al 29 de abril de 2020)



Modelos de Machine Learning aplicados a la comprensión de la COVID-19

Comprensión de los datos en "Patient_Korea.csv"

Por ejemplo, al Ciudad de Seul es considerada como una "Provincia". La ciudad tienen una población de 9.77 millones de habitantes. Y cada distrito o "Gu" es considerado una "Ciudad". Hay 25 "Gu". (Madrid tiene 21 distritos).



Por ejemplo, Gangnam-Gu es famoso por la canción "Gangnam Style"



Cantante PSY.
"Gangnam Style"
3.606.899.955
visualizaciones en
YouTube

Modelos de Machine Learning aplicados a la comprensión de la COVID-19

Comprensión de los datos en "Patient_Korea.csv"

Qué se entiende por "estate" (?)

Isolated



Released



Deceased



Modelos de Machine Learning aplicados a la comprensión de la COVID-19

Comprensión de los datos en "Patient_Korea.csv"

Qué se entiende por "Infection Case" (?)

Overas inflow



Contact with patient



Shincheonji Church



Modelos de Machine Learning aplicados a la comprensión de la COVID-19

Se emplean 10 tipologías de algoritmos, cada uno con su topología.



Para cada algoritmo se entrenan 100 modelos, buscando el que consiga una mejor precisión en “test a ciegas”. (# 1.000 modelos)

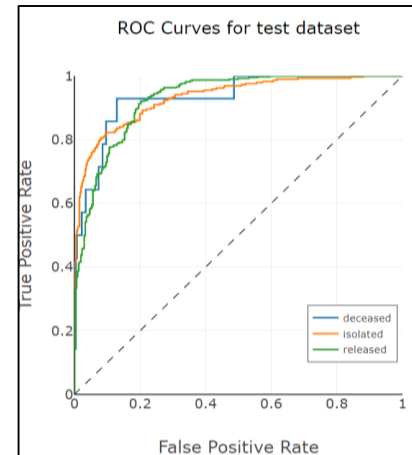
Select Models to Train

- Random Forest
- XGBoost
- Adaboost
- Gradient Boosting
- Extra Trees
- Decision Tree
- SVM
- KNN
- L2 Logistic Regression
- L1 Logistic Regression

COVID_19_Korea - Test #1

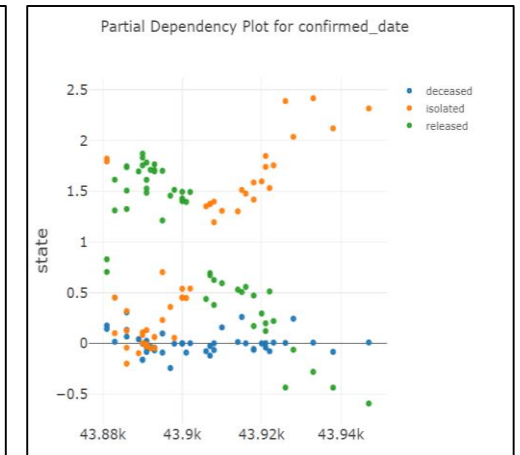
XGBoost SELECTED

	Acc	F1	Log Loss	Precision	Recall	Auc
Training	0.928	0.848	0.237	0.942	0.797	0.982
Validation	0.845	0.697	0.396	0.801	0.659	0.918
Test	0.855	0.656	0.377	0.904	0.624	0.931



Confusion Matrix for Test Dataset

	deceased	isolated	released	
deceased	2	5	7	14%
isolated		234	53	82%
released		33	344	91%
Precision	100%	86%	85%	Acc 86%
	deceased	isolated	released	Recall



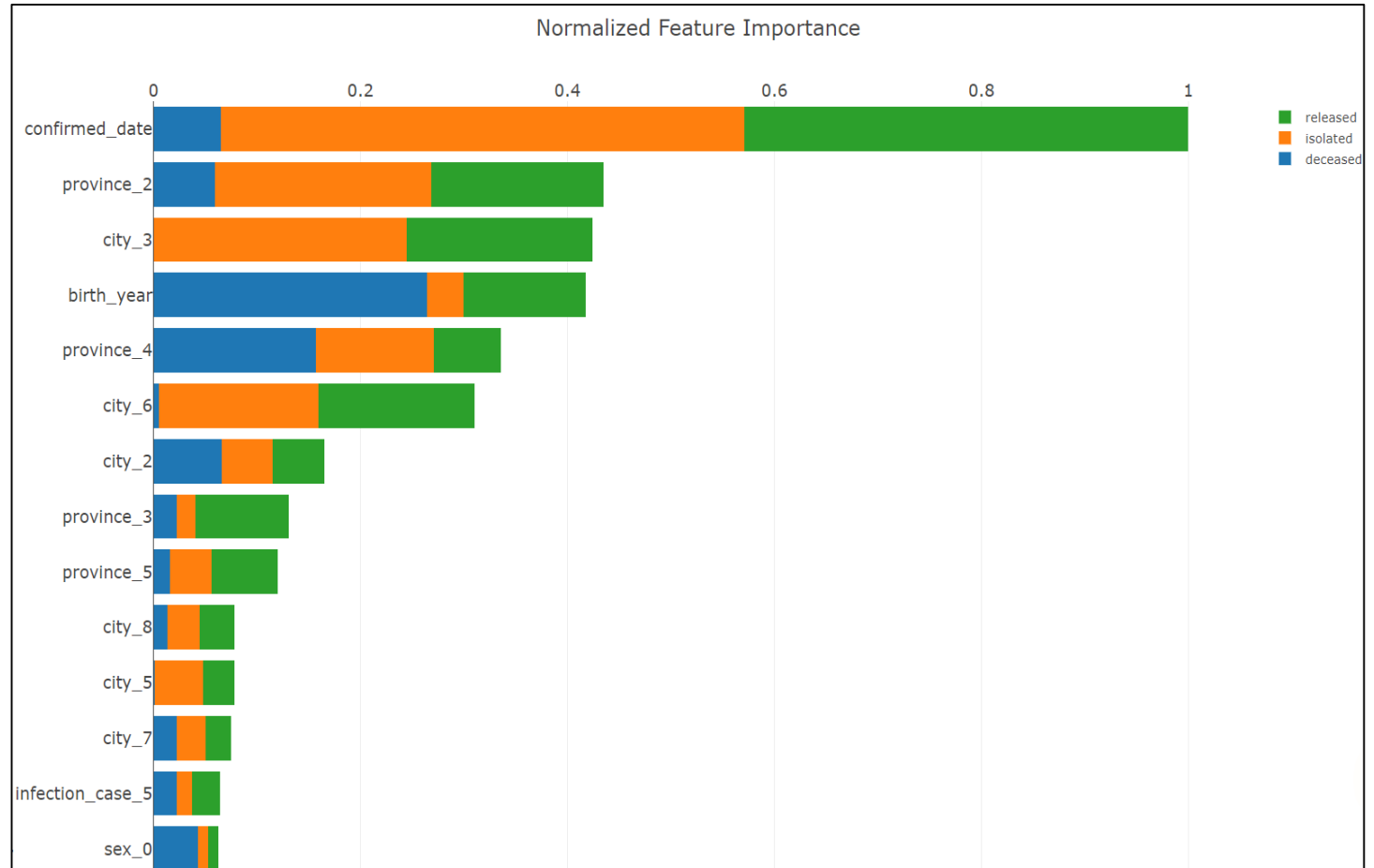
Modelos de Machine Learning aplicados a la compresión de la COVID-19



CONCLUSIONES:

- La fecha de confirmación de la enfermedad resulta fundamental para anticipar el colapso sanitario.
- En la mortalidad resulta fundamental la edad, el sexo y zona geográfica.
- Analizando los datos:
 - El aislamiento con alta y la hospitalización con alta depende de una predicción temprana.
 - En caso de hombres de avanzada edad en focos con carga viral elevada no se puede controlar la mortalidad mediante prevención únicamente.

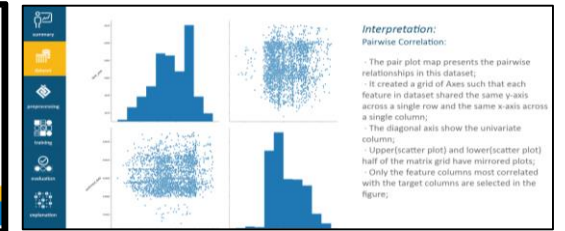
SML permite explicabilidad del modelo. Entender la física del proceso.



Modelos de Machine Learning aplicados a la comprensión de la COVID-19



SML permite predictibilidad y genera informe de relación de datos.



Predict_MASQ_FLSM_V8.csv with prediction results ↓

sex	birth_year	country	province	city	infection_case	confirmed_date	state_pred
male	1982	Spain	Seoul	Gangnam-gu	overseas inflow	43924	isolated
male	1960	Spain	Seoul	Gwanak-gu	Shincheonji Church	43887	released



Después de la estancia en Universidad de Seul y conocer la “secta” Fermín es diagnosticado el día 26 de febrero de 2020. Es ingresado pero se recupera. Yo voy a buscarlo para traerlo de vuelta y me quedo en casa de Peter pero también enfermo, se me diagnostica el día 3 de abril de 2020, me contagio tomándome un café en el aeropuerto de Madrid, también me recupero pero mediante aislamiento en casa de Peter.

Modelos de Machine Learning aplicados a la comprensión de la COVID-19



70 muertos!

	A	B	C	D	E	F	G	H
1	sex	birth_year	country	province	city	infection_case	confirmed_date	state_province
110	male	1984	Mongolia	Gyeonggi-do	Namyangju-si	etc	2/25/2020	deceased
286	male	1980	Korea	Gyeongsangbuk-do	Gyeongju-si	etc	2/22/2020	deceased
705	male	1976	Korea	Seoul	Mapo-gu	Guro-gu Call Center	3/19/2020	deceased
729	male	1969	Korea	Gyeongsangbuk-do	Gyeongsan-si		2/26/2020	deceased
731	female	1965	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/22/2020	deceased
774	female	1963	Korea	Daegu	Nam-gu		2/18/2020	deceased
813	male	1963	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/19/2020	deceased
814	female	1962	Korea	Daegu			3/11/2020	deceased
815	male	1962	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/22/2020	deceased
816	male	1961	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/19/2020	deceased
817	male	1960	Korea	Gyeongsangbuk-do	Gyeongsan-si		3/3/2020	deceased
818	male	1959	Korea	Gyeongsangbuk-do	Gyeongsan-si		2/25/2020	deceased
819	male	1958	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/22/2020	deceased
820	male	1957	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/20/2020	deceased
821	male	1955	Korea	Daegu			2/27/2020	deceased
822	male	1954	Korea	Daegu			2/28/2020	deceased
823	male	1953	Korea	Daegu			2/28/2020	deceased
824	male	1953	Korea	Ulsan	Nam-gu	etc	2/28/2020	deceased
825	male	1953	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	2/21/2020	deceased
826	female	1951	Korea	Daegu			3/1/2020	deceased
827	male	1951	Korea	Gyeongsangbuk-do	Gyeongsan-si		2/26/2020	deceased
828	male	1949	Korea	Daegu			2/24/2020	deceased
829	male	1949	Korea	Daegu			2/28/2020	deceased
830	male	1949	Korea	Gyeongsangbuk-do	Gyeongsan-si		3/15/2020	deceased
831	male	1947	Korea	Busan	Busanjin-gu	etc	3/13/2020	deceased
1008	male	1947	Korea	Gyeongsangbuk-do	Gyeongsan-si		2/28/2020	deceased
1121	male	1946	Korea	Daegu		Shincheonji Church	2/21/2020	deceased
1798	male	1945	Korea	Daegu			2/27/2020	deceased
1805	male	1945	Korea	Daegu			3/8/2020	deceased
2065	male	1944	Korea	Gyeongsangbuk-do	Gyeongsan-si		3/20/2020	deceased

La posibilidad de hacer análisis hace interesante el usar SML para estudiar “supuestos”. Se podría “periciar” la gestión realizada en la pandemia. Por ejemplo, en la base de datos de Korea los muertos son 70, pero los diagnósticos son en días muy variados. Suponiendo todo igual ya que no se puede cambiar el sexo o la edad a alguien, o no resulta complejo imaginarse movimientos de masas de gente de un lugar a otro, ¿si los test hubieran comenzado a nivel global de Korea le mismo día que se hizo el primer test positivo los muertos hubieran cambiado?

sex	birth_year	country	province	city	infection_case	confirmed_date	state_province
male	1963	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	43850	deceased
male	1962	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	43850	deceased
male	1961	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	43850	deceased
male	1958	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	43850	deceased
male	1957	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	43850	deceased
male	1953	Korea	Gyeongsangbuk-do	Cheongdo-gun	Cheongdo Daenam Hospital	43850	deceased
male	1946	Korea	Daegu		Shincheonji Church	1/20/2020	deceased
male	1945	Korea	Daegu			43850	deceased
male	1945	Korea	Daegu			43850	deceased
male	1939	Korea	Daegu			43850	deceased
male	1938	Korea	Daegu			43850	deceased
male	1938	Korea	Daegu			43850	deceased
female	1934	Korea	Daegu			43850	deceased
female	1934	Korea	Daegu			43850	deceased
male	1933	Korea	Daegu			43850	deceased
male	1930	Korea	Daegu			43850	deceased

17 muertos!!!!

Modelos de Machine Learning aplicados a la comprensión de la COVID-19



La predictividad no sirve para aplicarla a escenarios lejanos, pero sí para aplicarla a escenarios cercanos en el tiempo. Predicciones de alta frecuencia, con alimentación de datos constante y entrenamiento recurrente de los modelos.

Clusterización o catalogación de pacientes

TIPO A -- LEVES



TIPO B – MODERADO A GRAVE



TIPO C – SEVERO



After-Covid19

O papel da Enxeñaría Civil e da Edificación nun novo escenario

ENXEÑARÍA CIVIL
E EDIFICACIÓN
SUSTENTABLES
E INNOVADORAS



Agrupación estratéxica
Centro de Innovación Tecnolóxica
en Edificación e Enxeñaría Civil



UNIVERSIDADE DA CORUÑA



Campus
Sustentabilidade



XUNTA
DE GALICIA