

Trabajo Final - Covid-19

Miguel de la Cal Bravo
Félix Ángel Martínez Muela

Profesor/es:

ARTURO PERALTA MARTÍN-PALOMINO
RICARDO GARCÍA RÓDENAS

Introducción

Datos

Predicción

Clustering

Correlación

Conclusiones

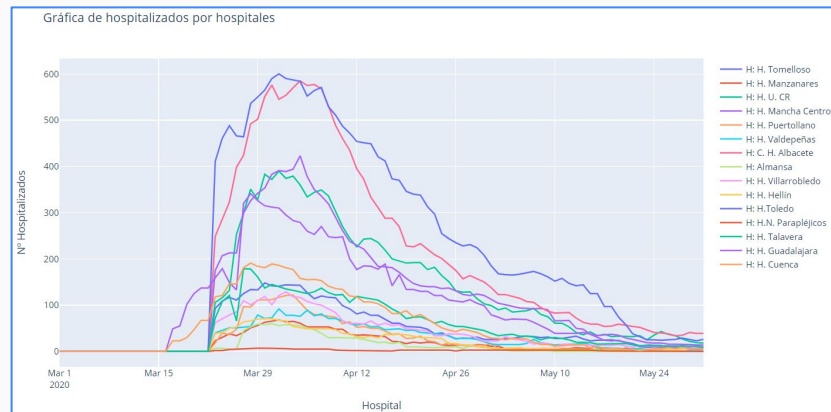
Datos

Hospitales:

- H. Tomelloso
- H. Manzanares
- H. U. CR
- H. Mancha Centro
- H. Puertollano
- H. Valdepeñas
- C. H. Albacete
- Almansa
- H. Villarrobledo
- H. Hellín
- H. Toledo
- H.N. Paraplégicos
- H. Talavera
- H. Guadalajara
- H. Cuenca

activos.csv
altas.csv
casos.csv
fallecidos.csv
hospitalizados.csv

<https://www.castillalamancha.es/actualidad/notasdeprensa>



Fecha	24-mar	25-mar	26-mar	27-mar	28-mar	29-mar	30-mar	31-mar	01-abr	02-abr	03-abr	04-abr	05-abr	06-abr	07-abr	08-abr	09-abr	10-abr
	M	X	J	V	S	D	L	M	X	J	V	S	D	L	M	X	J	V
Hospitalizados Totales																		
H. Tomelloso	109	117	111	125	134	133	148	142	141	144	144	143	127	114	120	117	116	99
H. Manzanares	30	38	34	41	51	56	63	65	68	64	65	61	53	53	53	53	48	48
H. U. CR	116	132	253	299	351	328	383	372	390	374	379	360	334	344	349	336	304	270
H. Mancha Centro	207	213	213	299	326	342	353	383	391	389	394	423	378	350	336	319	290	261
H. Puertollano	40	36	48	97	96	111	112	112	117	121	122	106	89	80	78	76	74	60
H. Valdepeñas	46	51	51	53	54	79	71	73	92	78	78	76	89	77	81	71	71	66
Ciudad Real	548	587	710	914	1012	1049	1130	1147	1199	1170	1182	1169	1070	1018	1017	972	903	804
C. H. Albacete	286	323	398	424	492	502	551	576	545	554	570	585	575	577	568	531	489	460
Almansa	7	5	5	45	53	59	58	60	56	58	57	56	51	47	39	30	30	30
H. Villarrobledo	70	79	85	110	101	110	119	101	121	129	118	117	108	103	100	92	84	66
H. Hellín	41	50	53	64	66	70	71	71	67	66	54	51	49	49	49	49	47	39
Albacete	404	457	541	643	712	741	799	808	789	807	799	809	783	776	756	702	650	595
H.Toledo	461	488	466	464	536	550	565	590	600	590	587	584	552	564	571	529	511	487
H.N. Paraplégicos	2	4	5	6	7	7	7	7	6	6	5	5	5	5	5	5	3	2
H. Talavera	108	121	66	179	178	161	137	145	140	135	132	129	125	129	137	127	122	123
Toledo	571	613	537	649	721	718	709	742	746	731	724	718	682	698	713	661	636	612
H. Guadalajara	179	149	133	320	341	326	315	312	310	295	283	279	260	253	270	248	246	248
Guadalajara	179	149	133	320	341	326	315	312	310	295	283	279	260	253	270	248	246	248
H. Cuenca	121	146	146	181	191	184	181	189	186	181	177	158	155	156	153	141	136	134
Cuenca	124	146	146	181	191	184	181	189	186	181	177	158	155	156	153	141	136	134
Total CLM	1826	1952	2067	2707	2977	3018	3134	3198	3230	3184	3165	3133	2950	2901	2909	2724	2571	2393
Diferencia día anterior																		
Ciudad Real	93	39	123	204	98	37	81	17	52	-29	12	-13	-99	-52	-1	-45	-69	-99
Albacete	47	53	84	102	69	29	58	9	-19	18	-8	10	-26	-7	-20	-54	-52	-55
Toledo	86	42	-76	112	72	-3	-9	33	4	-15	-7	-6	-36	16	15	-52	-25	-24
Guadalajara	20	-30	-16	187	21	-15	-11	-3	-2	-15	-12	-4	-19	-7	17	-22	-2	2
Cuenca	6	22	0	35	10	-7	-3	8	-3	-5	-4	-19	-3	1	-3	-12	-5	-2
CLM	252	126	115	640	270	41	116	64	32	-46	-19	-32	-183	-49	8	-185	-153	-178

Dataset realizado por Miguel de la Cal Bravo (también autor de este trabajo) donde se recogen los datos publicados en los informes diarios por Sanidad de Castilla-La Mancha

https://github.com/miguelcal97/dsi_prediccionCovid19

Predicción

ARIMA - autoregressive integrated moving average

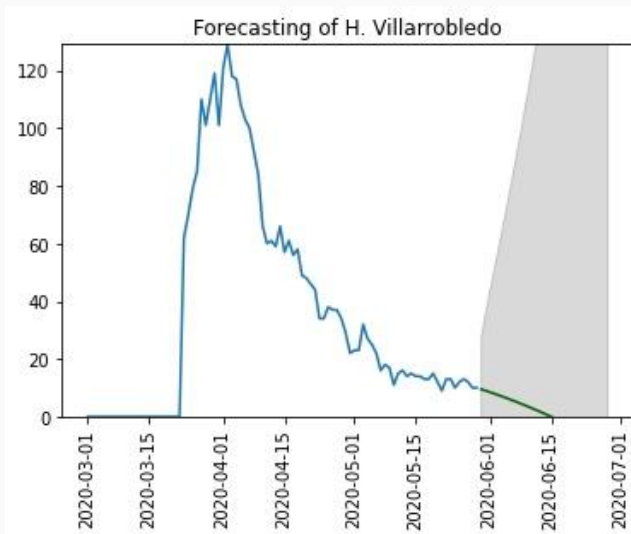
AR: Y_t depende solo de su retardo (lag)

MA: Y_t depende sólo de los errores de pronóstico retardado (lag)

Modelo dinámico { Datos del pasado ✓
Variables independientes ✗

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

Serie de tiempo no estacionaria y que no tenga ruido blanco aleatorio



Clustering

Clustering con series temporales

K-medias (o K-means)

Clusters =>

Patrones de evolución de hospitalizados
en los distintos hospitales de CLM



Gráfica de hospitalizados en los hospitales de Valdepeñas, Albacete y Toledo, donde se detecta un patrón similar de la curva de hospitalizados

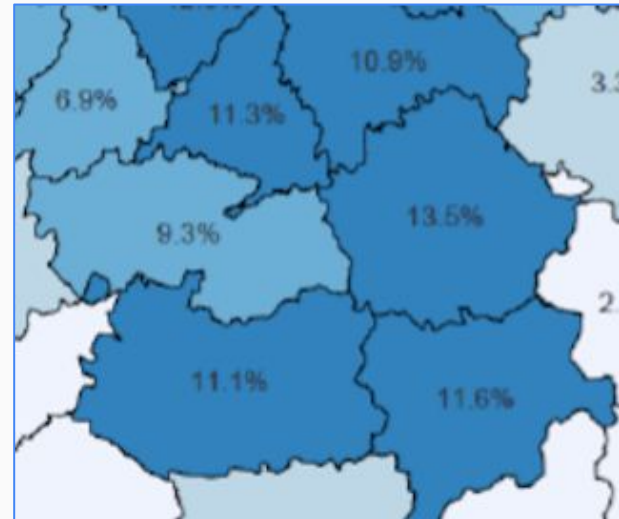
Correlación

Coeficiente de Pearson

Correlación cercana a 1

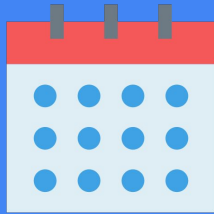
DTW: Dynamic Time Warping

Variables: activos, altas, casos, fallecidos y hospitalizados

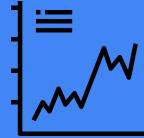


<https://www.encastillalamancha.es/castilla-la-mancha/cuatro-provincias-de-clm-doblan-la-media-espanola-de-contagiados-por-el-covid-19/>

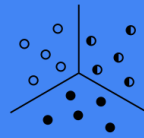
Conclusiones



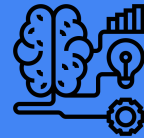
Series
Temporales



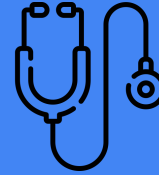
Análisis



K-means



ARIMA



Referencias

- <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>
- <https://pypi.org/project/pmdarima/>
- <http://alkaline-ml.com/pmdarima/1.6.1/index.html>
- <https://www.kaggle.com/egregori/clustering-time-series>
- <https://towardsdatascience.com/time-series-hierarchical-clustering-using-dynamic-time-warping-in-python-c8c9edf2fda5>
- https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima_model.ARIMA.html
- <https://www.castillalamancha.es/actualidad/notasdeprensa>

¡Gracias!

Contacto:

Universidad de Castilla la
Mancha - UCLM

FelixAngel.Martinez@alu.uclm.es
<https://github.com/FelixAngelMartinez>

MiguelLa.Cal@alu.uclm.es
<https://github.com/miguelcal97>

