

# INTELIGENCIA ARTIFICIAL EN LAS ORGANIZACIONES

## PRÁCTICA 1

(Aplicación de RNA)



Grado en Ingeniería Informática

Campus de Colmenarejo

Curso 2019/2020

### Autores

Paula Gilabert Robles - 100363657@alumnos.uc3m.es

Alvaro Gonzalez Muñoz - 100363552@alumnos.uc3m.es

Eduardo Ureña Toledano - 100329937@alumnos.uc3m.es

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Parte I</b>	<b>3</b>
2.1 Planteamiento y desarrollo	3
2.2 Resultados	4
2.2.1 Modelo 1 (10 capas ocultas; 30 días)	4
2.2.1 Modelo 2 (10 capas ocultas; 100 días)	6
2.2.1 Modelo 3 (15 capas ocultas; 30 días)	9
2.2.1 Modelo 4 (15 capas ocultas; 15 días)	12
2.3 Análisis de resultados y conclusiones	14
<b>3. Parte II</b>	<b>15</b>
3.1 Planteamiento y desarrollo	15
3.2 Resultados	15
3.3 Análisis de resultados y conclusiones	19

# 1. Introducción

En esta práctica realizaremos los diferentes modelos de redes de neuronas que nos ayudarán a predecir el avance del coronavirus en los dos países que hemos seleccionado, estos son España e India.

En la primera parte de la práctica utilizaremos un modelo de regresión para predecir los casos confirmados que se van a encontrar en los próximos 3 días siguientes a los datos que nos han otorgado utilizando los datos de todos los países del mundo.

En la segunda parte utilizaremos los datos de confirmados y de muertes para predecir en España e India, mediante series temporales, cuáles serán los resultados para los próximos 15 días, utilizando los datos de los respectivos países desde que se inició la pandemia.

## 2. Parte I

### 2.1 Planteamiento y desarrollo

La primera parte de la práctica consiste en realizar un modelo de regresión cuyo objetivo será predecir los casos confirmados en los siguientes 3 días con respecto a los datos otorgados.

Hemos decidido hacer 4 modelos diferentes y ponerlos a prueba para escoger el que más se adecue a los resultados esperados.

Para todos los modelos hemos utilizado un training time de 1000 y un 0.4 de tasa de aprendizaje, ya que a pesar de haber hecho alguna prueba cambiando dichos valores, los resultados han sido mucho peores. Dependiendo del modelo hemos variado el número de capas ocultas y la cantidad de días que se usan para el entrenamiento, buscando así encontrar, de diferente manera, los mejores resultados posibles en cuanto a la predicción de los contagios en India y España, y valorar como afecta el cambio de dichos parámetros.

En todos los modelos, hemos entrenado el modelo de dos maneras: mediante "Percentage split" y mediante "Cross-validation". Todos los "Percentage split" han sido realizados con un 66% de training, y todos excepto el cuarto modelo se han entrenado con 10 Folds, habiéndose entrenado este último con 20 Folds, debido a que esos valores son los que minimizan los errores.

Para el primer modelo hemos utilizado 10 capas ocultas y lo hemos entrenado con los datos de los últimos 30 días.

Para el segundo modelo hemos utilizado 10 capas ocultas y lo hemos entrenado con los datos de los últimos 100 días.

Para el tercer modelo hemos utilizado 15 capas ocultas y lo hemos entrenado con los datos de los últimos 30 días.

Para el cuarto modelo hemos utilizado 15 capas ocultas y lo hemos entrenado con los datos de los últimos 15 días.

## **2.2 Resultados**

### *2.2.1 Modelo 1 (10 capas ocultas; 30 días)*

- Día 1:
  - Percentage split:

■ Correlation coefficient	0.849
■ Mean absolute error	49806.7895
■ Root mean squared error	92904.5107
■ Relative absolute error	31.2675 %
■ Root relative squared error	52.3872 %
  - Cross-validation:

■ Correlation coefficient	0.9849
■ Mean absolute error	41115.6174
■ Root mean squared error	140813.7546
■ Relative absolute error	19.9123 %
■ Root relative squared error	19.8748 %
- Día 2:
  - Percentage split:

■ Correlation coefficient	0.9494
■ Mean absolute error	36361.7739
■ Root mean squared error	54614.0585
■ Relative absolute error	22.6491 %
■ Root relative squared error	30.5051 %
  - Cross-validation:

■ Correlation coefficient	0.9884
■ Mean absolute error	41806.7013
■ Root mean squared error	117147.941
■ Relative absolute error	20.0616 %
■ Root relative squared error	16.3995 %

- Día 3:

- Percentage split:

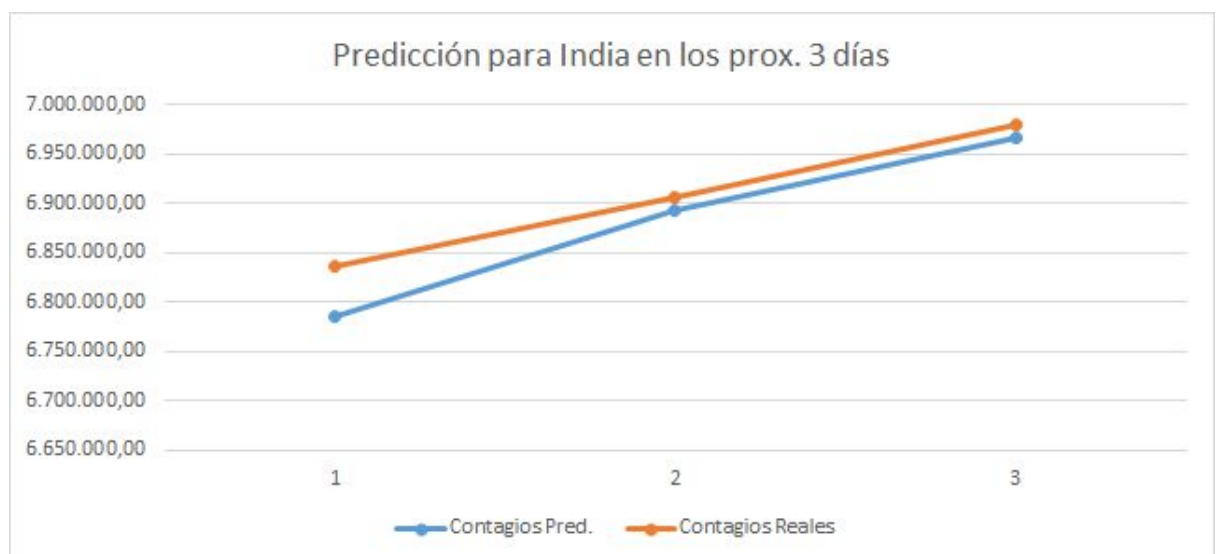
■ Correlation coefficient	0.9505
■ Mean absolute error	36474.6074
■ Root mean squared error	53993.4003
■ Relative absolute error	22.5433 %
■ Root relative squared error	29.877 %

- Cross-validation:

■ Correlation coefficient	0.9892
■ Mean absolute error	45693.2346
■ Root mean squared error	110307.4627
■ Relative absolute error	21.7241 %
■ Root relative squared error	15.3138 %

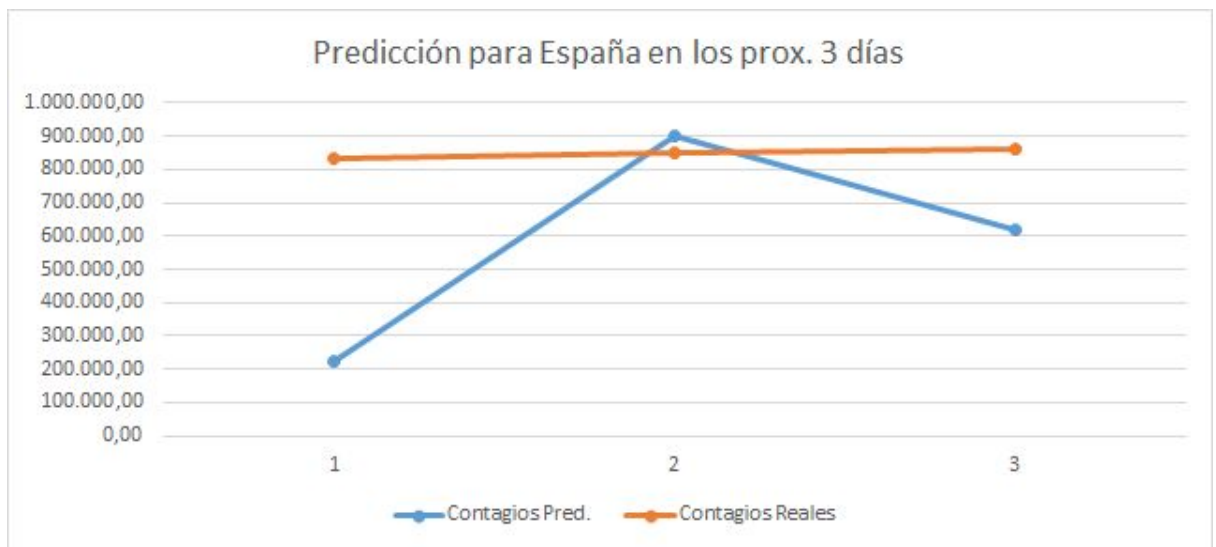
- India:

Contagios Pred.	Contagios Reales
6.785.465	6.835.655
6.892.299	6.906.151
6.967.267	6.979.423



- España:

Contagios Pred.	Contagios Reales
225.156	835.901
901.161	848.324
619.899	861.112



### 2.2.1 Modelo 2 (10 capas ocultas; 100 días)

- Día 1:
  - Percentage split:
    - Correlation coefficient 0.8556
    - Mean absolute error 122435.8998
    - Root mean squared error 742220.837
    - Relative absolute error 54.4346 %
    - Root relative squared error 76.9458 %
  - Cross-validation:
    - Correlation coefficient 0.9327
    - Mean absolute error 45423.725

■ Root mean squared error	210310.5855
■ Relative absolute error	27.9759 %
■ Root relative squared error	36.6 %

- Día 2:

- Percentage split:

■ Correlation coefficient	0.7849
■ Mean absolute error	130135.4297
■ Root mean squared error	811011.167
■ Relative absolute error	57.404 %
■ Root relative squared error	83.5053 %

- Cross-validation:

■ Correlation coefficient	0.8669
■ Mean absolute error	115742.3412
■ Root mean squared error	330136.8644
■ Relative absolute error	70.6241 %
■ Root relative squared error	57.0454 %

- Día 3:

- Percentage split:

■ Correlation coefficient	0.847
■ Mean absolute error	128306.1787
■ Root mean squared error	767638.8794
■ Relative absolute error	56.1558 %
■ Root relative squared error	78.4978 %

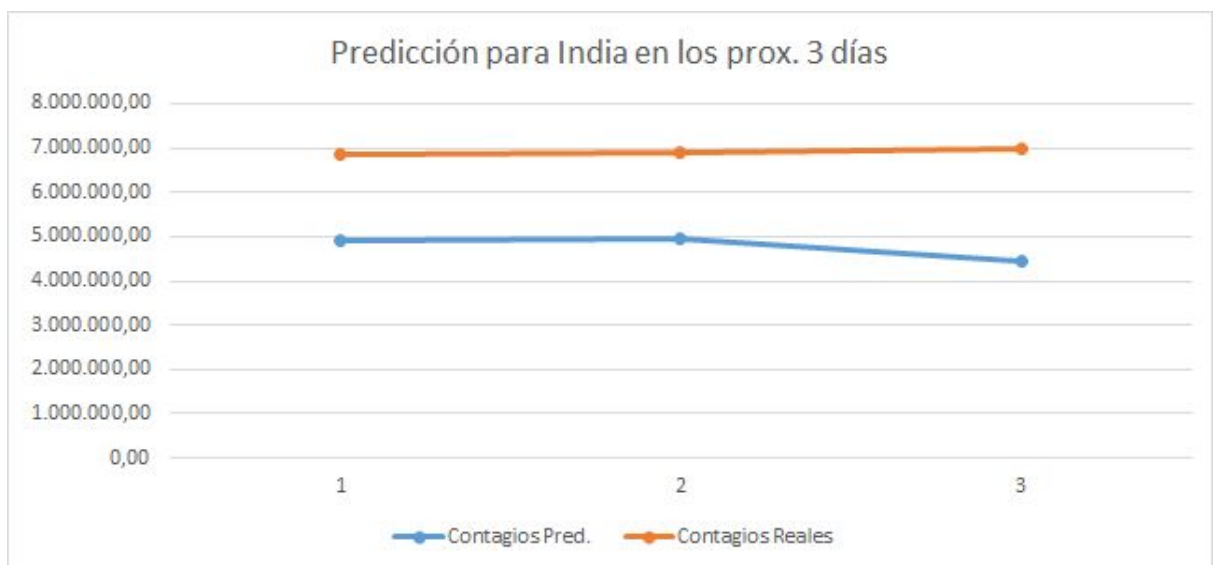
- Cross-validation:

■ Correlation coefficient	0.9283
■ Mean absolute error	50098.8279
■ Root mean squared error	219724.5152

- Relative absolute error 30.2871 %
- Root relative squared error 37.6959 %

- India:

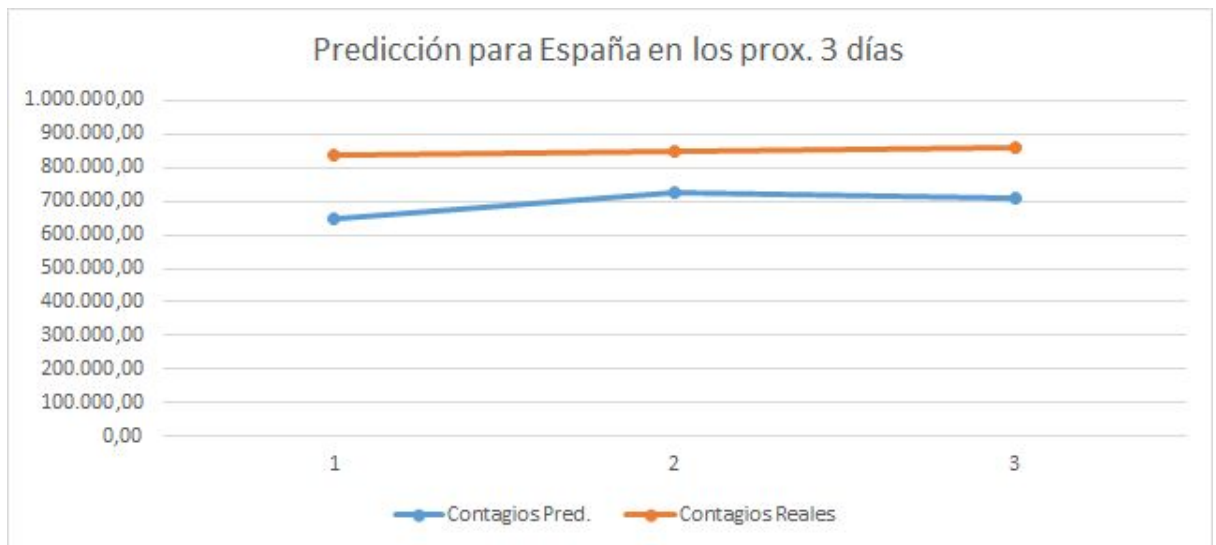
Contagios Pred.	Contagios Reales
4.924.181	6.835.655
4.946.580	6.906.151
4.440.004	6.979.423



- España:

Contagios Pred.	Contagios Reales
649.333	835.901
724.104	848.324
709.472	861.112





### 2.2.1 Modelo 3 (15 capas ocultas; 30 días)

- Día 1:
  - Percentage split:
    - Correlation coefficient      0.9754
    - Mean absolute error      31701.9072
    - Root mean squared error      37953.3423
    - Relative absolute error      19.9017 %
    - Root relative squared error      21.4012 %
  - Cross-validation:
    - Correlation coefficient      0.9823
    - Mean absolute error      46797.6358
    - Root mean squared error      156796.8768
    - Relative absolute error      22.6641 %
    - Root relative squared error      22.1306 %
- Día 2:
  - Percentage split:
    - Correlation coefficient      0.9774
    - Mean absolute error      32548.8572
    - Root mean squared error      39158.7866

- Relative absolute error 20.2741 %
- Root relative squared error 21.8724 %

- Cross-validation:

- Correlation coefficient 0.9951
- Mean absolute error 35052.7464
- Root mean squared error 73467.2064
- Relative absolute error 16.8206 %
- Root relative squared error 10.2847 %

- Día 3:

- Percentage split:

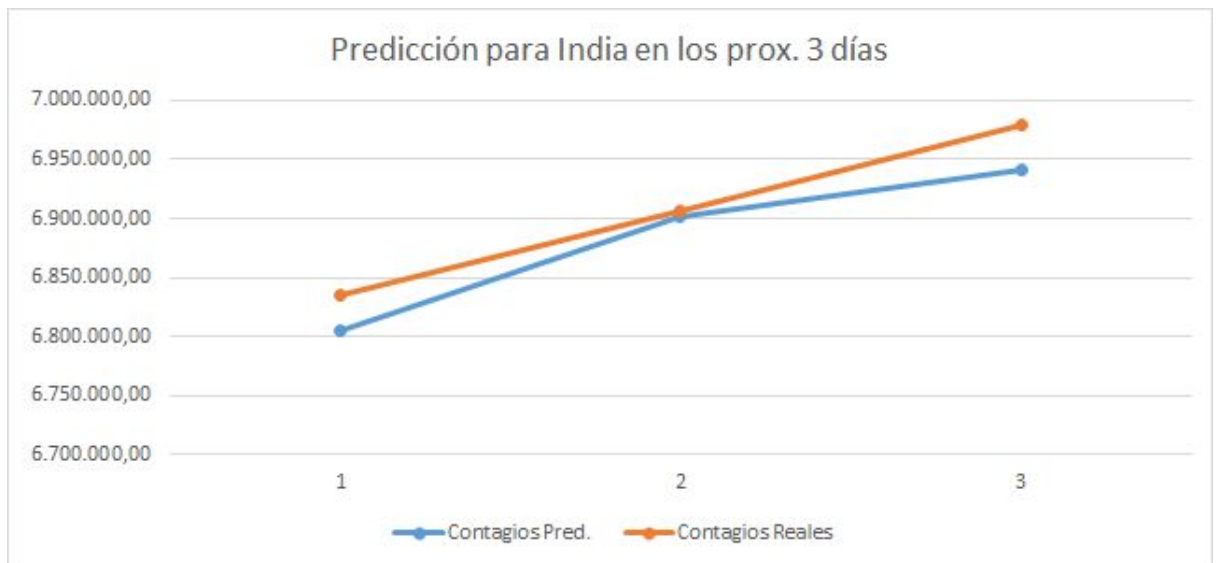
- Correlation coefficient 0.9782
- Mean absolute error 27745.5701
- Root mean squared error 33749.0575
- Relative absolute error 17.1483 %
- Root relative squared error 18.6749 %

- Cross-validation:

- Correlation coefficient 0.9937
- Mean absolute error 35701.0102
- Root mean squared error 87244.8646
- Relative absolute error 16.9735 %
- Root relative squared error 12.1121 %

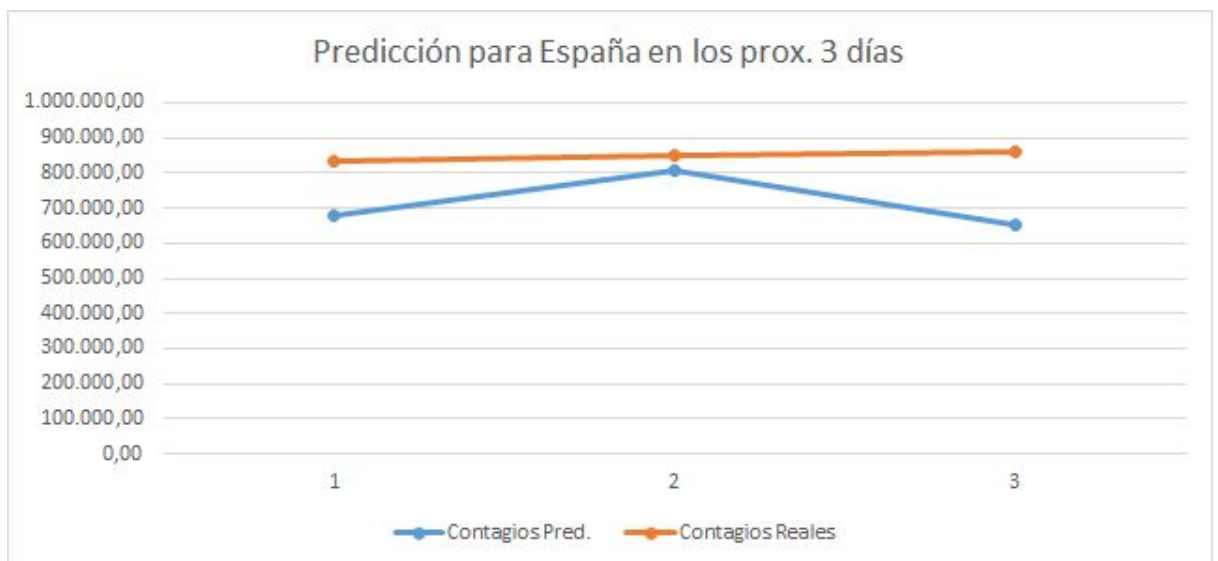
- India:

Contagios Pred.	Contagios Reales
6.804.379	6.835.655
6.902.161	6.906.151
6.941.338	6.979.423



- España:

Contagios Pred.	Contagios Reales
681.043	835.901
804.964	848.324
651.559	861.112



### 2.2.1 Modelo 4 (15 capas ocultas; 15 días)

- Día 1:

- Percentage split:

■ Correlation coefficient	0.99
■ Mean absolute error	32540.5953
■ Root mean squared error	43447.0245
■ Relative absolute error	20.4282 %
■ Root relative squared error	24.499 %

- Cross-validation:

■ Correlation coefficient	0.9799
■ Mean absolute error	45819.5959
■ Root mean squared error	142208.2805
■ Relative absolute error	22.2484 %
■ Root relative squared error	20.1133 %

- Día 2:

- Percentage split:

■ Correlation coefficient	0.988
■ Mean absolute error	30924.3897
■ Root mean squared error	46460.8167
■ Relative absolute error	19.2623 %
■ Root relative squared error	25.951 %

- Cross-validation:

■ Correlation coefficient	0.9725
■ Mean absolute error	58674.2211
■ Root mean squared error	165640.6469
■ Relative absolute error	28.2256 %
■ Root relative squared error	23.2364 %

- Día 3:

- Percentage split:

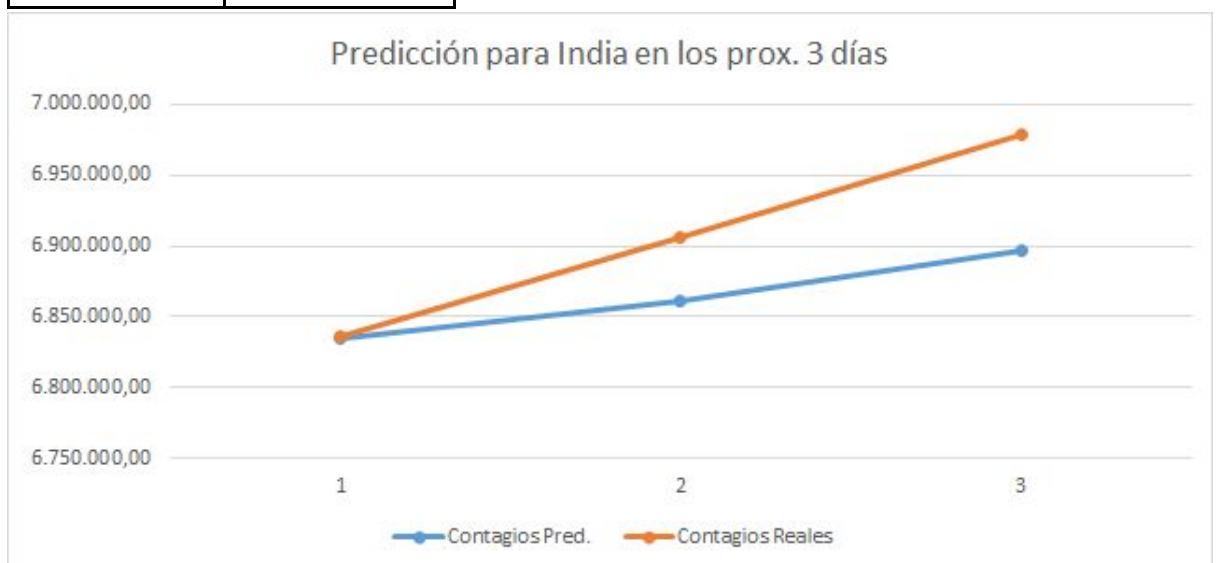
■ Correlation coefficient	0.989
■ Mean absolute error	31728.2698
■ Root mean squared error	45986.2575
■ Relative absolute error	19.6098 %
■ Root relative squared error	25.4463 %

- Cross-validation:

■ Correlation coefficient	0.6836
■ Mean absolute error	215101.6559
■ Root mean squared error	753602.2289
■ Relative absolute error	102.5199 %
■ Root relative squared error	104.8401 %

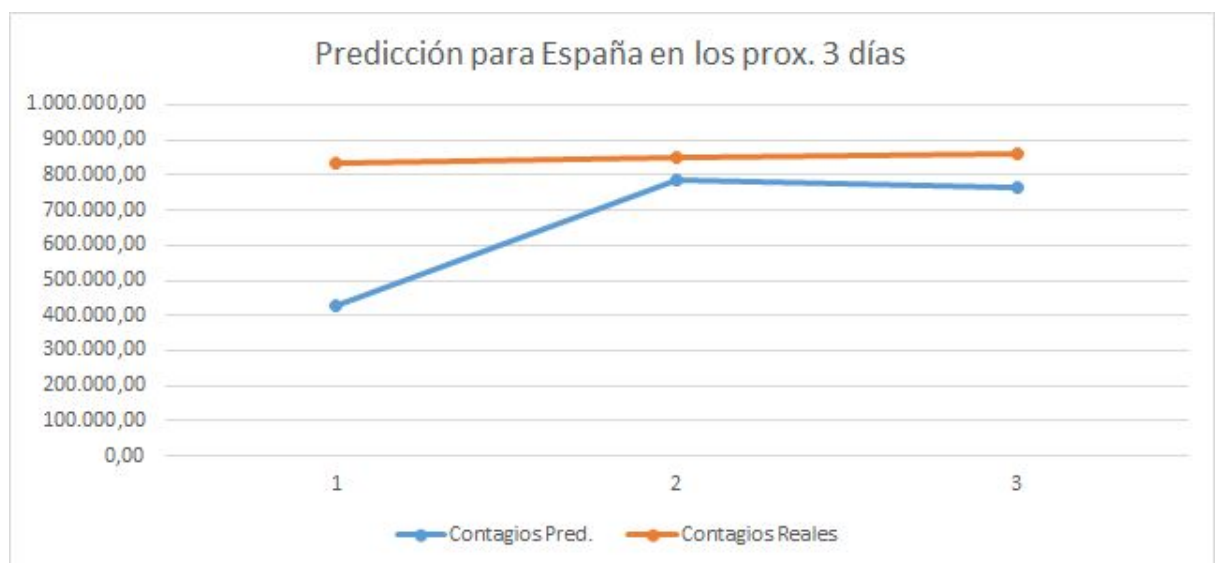
- India:

Contagios Pred.	Contagios Reales
6.835.290	6.835.655
6.860.708	6.906.151
6.896.638	6.979.423



- España:

Contagios Pred.	Contagios Reales
428.378	835.901
787.721	848.324
762.380	861.112



## **2.3 Análisis de resultados y conclusiones**

En conclusión, si tenemos en cuenta los datos de ambos países, el mejor modelo es el 3 debido a que los porcentajes de error son menores que en el resto de modelos. El porcentaje de error máximo de este modelo es el del día 1 con Cross-validation con un valor de 22.6641 %, el cual es el menor de los porcentajes máximos del resto de modelos.

Por otra parte el modelo 1 nos ha ofrecido el mejor resultado para India de todos los modelos, pero el peor para España, que no se aproxima a la realidad en ningún aspecto, por lo tanto lo hemos descartado, el modelo 2 ha predicho bien el crecimiento de los contagios en ambos países, pero la diferencia entre los casos reales y los estimados es demasiado grande por lo que también lo hemos descartado. Y por último el modelo 4, que nos ha otorgado un crecimiento diferente al real pero hay días en los que la diferencia de datos es demasiado grande, siendo el tercer día de India y el primer día de España, los demás días se acercan bastante a la realidad pero el modelo 3 sigue siendo el que ha dado mejores resultados.

## 3. Parte II

### 3.1 Planteamiento y desarrollo

Para poder evaluar los resultados de las predicciones hemos decidido entrenar a la RNA con todos los datos de cada país menos los de los últimos 15 días, para así predecirlos y compararlos con los reales.

### 3.2 Resultados

- España

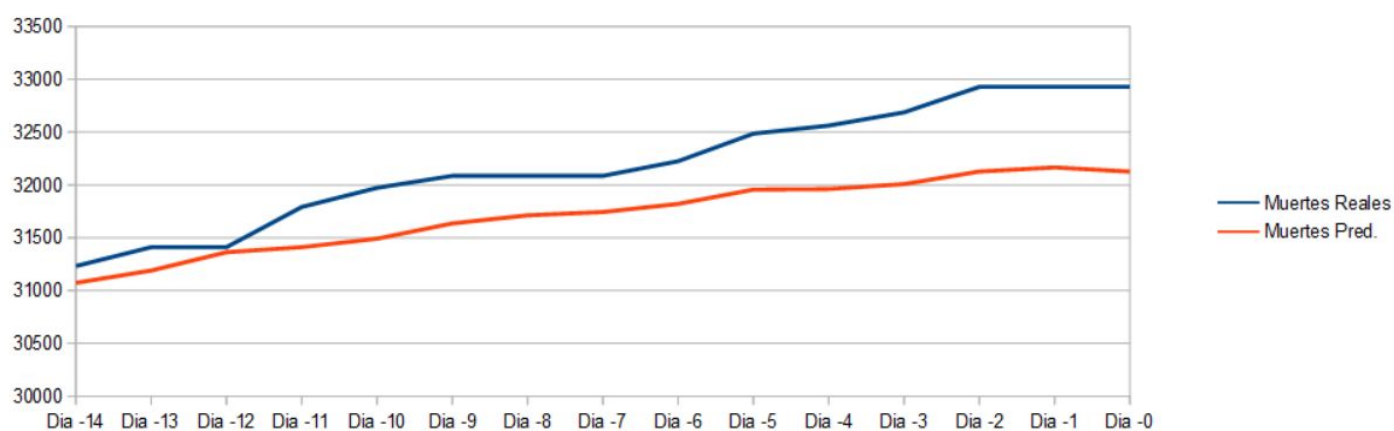
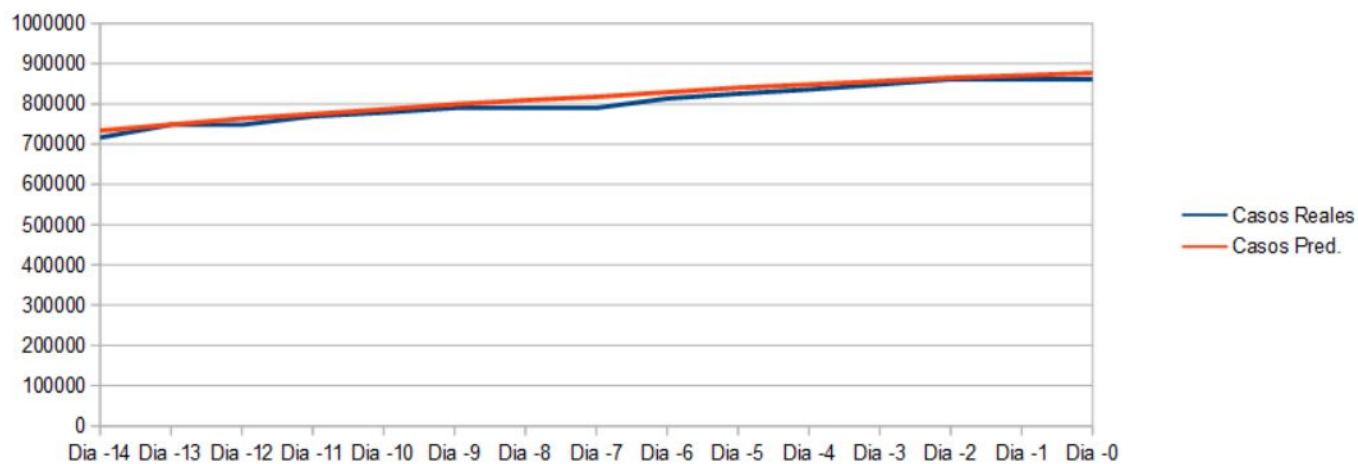
Valores reales de los últimos 15 días de casos y muertes:

Días	Casos	Muertes
Día -14	716481	31232
Día -13	748266	31411
Día -12	748266	31411
Día -11	769188	31791
Día -10	778607	31973
Día -9	789932	32086
Día -8	789932	32086
Día -7	789932	32086
Día -6	813412	32225
Día -5	825410	32486
Día -4	835901	32562
Día -3	848324	32688
Día -2	861112	32929
Día -1	861112	32929
Día 0	861112	32929

Resultados de la predicción con *lag* mínimo 14 y máximo 22, hemos elegido estos datos porque son los que proporcionaban unos valores más parecidos a los reales.

Días	Casos	Muertes
Día -14	733781.7 769	31072.89 4
Día -13	748623.7 87	31189.61 41
Día -12	763732.2 15	31363.06 69
Día -11	774835.1 594	31410.79 38
Día -10	786651.7 62	31491.15 33
Día -9	799268.3 86	31635.55 82
Día -8	809139.4 237	31712.41 78
Día -7	817634.0 723	31743.43 51
Día -6	829139.1 795	31820.93 8
Día -5	840344.5 438	31956.15 2
Día -4	848264.1 104	31961.68 4
Día -3	856531.7 509	32008.36 66
Día -2	864957.7 589	32127.18 77
Día -1	871014.5 027	32166.57 54
Día 0	876931.7 513	32126.62 07





## ● India

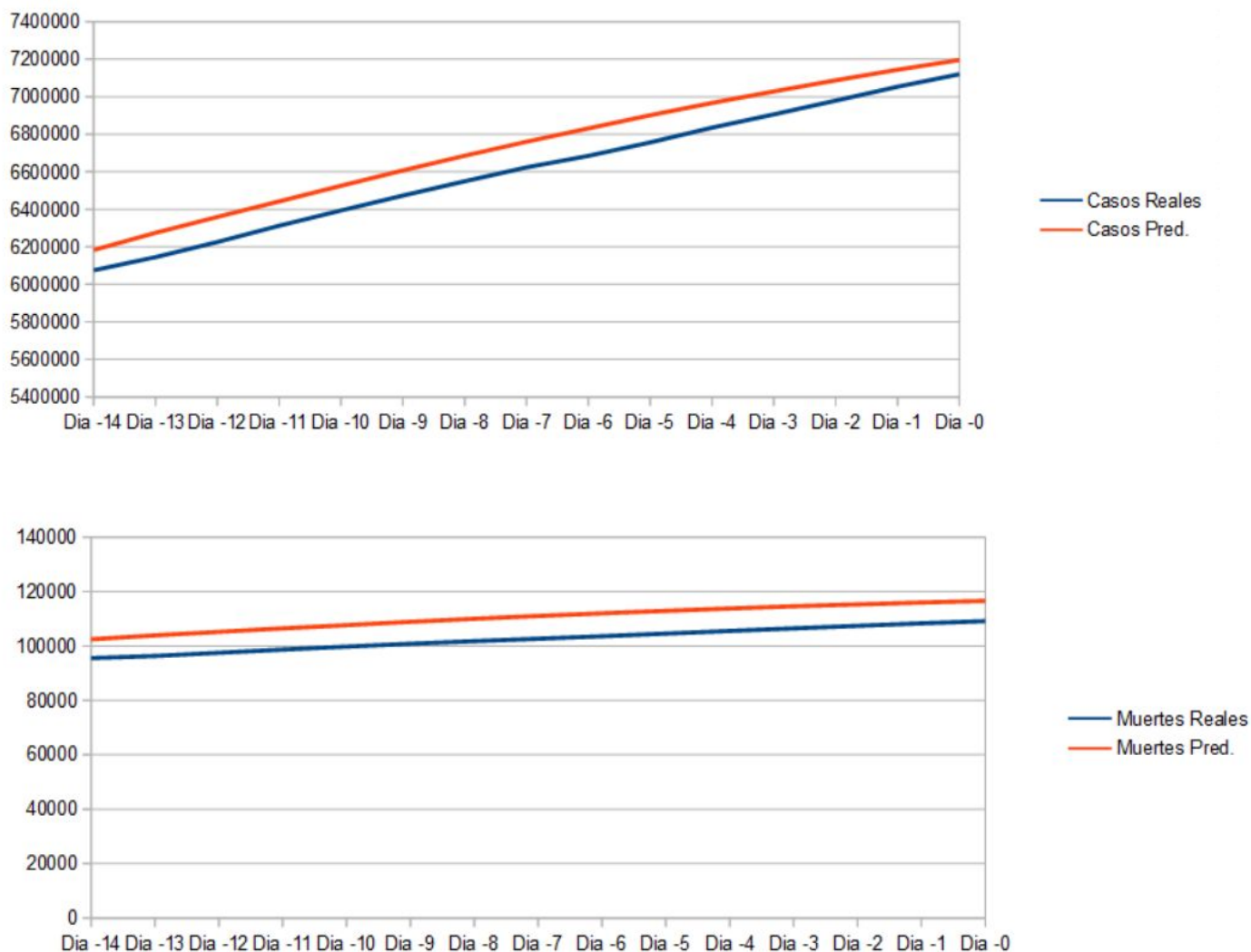
Valores reales de los últimos 15 días de casos y muertes:

Días	Casos	Muertes
Dia -14	6074702	95542
Dia -13	6145291	96318
Dia -12	6225763	97497
Dia -11	6312584	98678
Dia -10	6394068	99773
Dia -9	6473544	100842
Dia -8	6549373	101782
Dia -7	6623815	102685

Día -6	6685082	103569
Día -5	6757131	104555
Día -4	6835655	105526
Día -3	6906151	106490
Día -2	6979423	107416
Día -1	7053806	108334
Día 0	7120538	109150

Resultados de la predicción con *lag* mínimo 1 y máximo 30, hemos elegido estos datos porque son los que proporcionaban unos valores más parecidos a los reales.

Días	Casos	Muertes
Día -14	6183216,1057	102464,0434
Día -13	6274842,7964	103907,2562
Día -12	6359364,2855	105217,1342
Día -11	6442649,7293	106484,7184
Día -10	6525580,7186	107664,7941
Día -9	6607143,8069	108899,5625
Día -8	6685749,4502	110004,4338
Día -7	6760608,2573	111039,7559
Día -6	6831655,7951	112005,7768
Día -5	6901889,5302	112939,8983
Día -4	6967336,6205	113804,2278
Día -3	7028656,373	114615,2217
Día -2	7087834,0461	115299,4657
Día -1	7144128,3329	116008,2446
Día 0	7196094,5052	116595,6742



### **3.3 Análisis de resultados y conclusiones**

Como podemos observar en las diferentes gráficas que hemos obtenido, tenemos en color azul los datos reales y en rojo los datos predichos por el programa.

Empezando por España, podemos observar que la predicción de los contagios es muy similar a la realidad, en la mayor parte de la gráfica las líneas se superponen, por lo que la predicción es bastante certera.

Mientras que en las muertes encontramos mayores diferencias, el programa ha predicho menos muertes de las que en verdad han sido, basándose en los datos de los anteriores días, por lo que en los últimos 15 días las muertes en España han estado por encima de lo que se podía haber predicho mediante el programa. Aunque los datos se han alejado de los esperados, no ha sido una diferencia tan extrema en cuanto a números, la diferencia se encuentra entre 1000 y 1500 casos, siendo los totales más de 30000.

En cuanto a India, las dos tablas son bastante semejantes, siguen un crecimiento lento y uniforme, el modelo de aprendizaje ha predicho el crecimiento con bastante precisión, pero en ambas tablas el

número de muertes y contagios se encuentra por debajo de las que en verdad han sido, y al tratar con números más grandes que España, el error es mayor.

Para hacer una mejor comparación de los datos hemos calculado la diferencia de los reales y los predichos y después hemos hecho una media de dichas diferencias:

- Contagios en España:  $12257,54532 \approx 12258$
- Muertes en España:  $469,16949 \approx 470$
- Contagios en India:  $123987,29 \approx 123988$
- Muertes en India:  $7919,5471 \approx 7920$

Si tenemos en cuenta los valores que acabamos de calcular la predicción que mejor representa la realidad sería la de las muertes en España, después las de India, luego los contagios en España y por último los de India. Por lo tanto la red de neuronas ha predicho mejor los datos de España.