Estudiando la predicción de casos de Covid-19 utilizando RNN

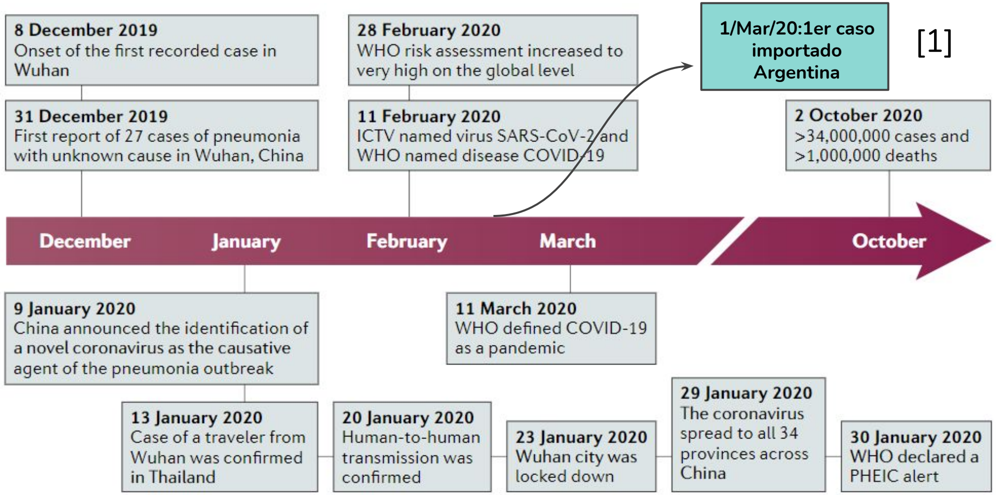
Denise Stefania Cammarota
Aprendizaje Automático y redes neuronales artificiales
Instituto Balseiro
denisescammarota@gmail.com

Introducción

SARS-COV-2 ⟶ COVID-19

(Severe Acute Respiratory Syndrome Coronavirus 2)

(Coronavirus Disease 2019)



[1]

Argentina

1.424.533

TOTAL DE AFECTADOS

128.576

INFECTADOS ACTIVOS

1.257.227

RECUPERADOS

38.730

FALLECIDOS

Día 30/nov/20

[1] Hu B., Hua G. et al., *Characteristics of SARS-CoV-2 and COVID-19*, https://www.nature.com/articles/s41579-020-00459-7
[2] Ministerio de Salud, Información epidemiológica, https://www.argentina.gob.ar/salud/coronavirus-COVID-19/sala-situacion

Motivación

Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study

Sourabh Shastri*, Kuljeet Singh, Sachin Kumar, Paramjit Kour, Vibhakar Mansotra[1]

Department of Computer Science & IT, University of Jammu, Jammu & Kashmir, India

https://doi.org/10.1016/j.chaos.2020.110227

Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India

Parul Arora[a,*], Himanshu Kumar[b], Bijaya Ketan Panigrahi[a]

[a] *Department of Electrical Engineering, Indian Institute of Technology, Delhi, New Delhi, India*
[b] *Infosys, Pune, India*

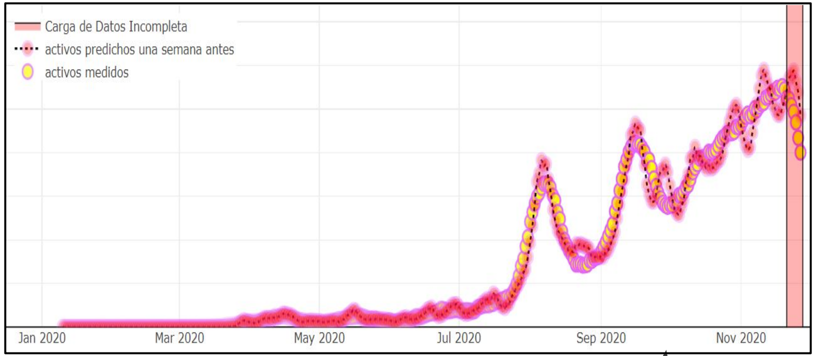
https://doi.org/10.1016/j.chaos.2020.110017

Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM

Farah Shahid, Aneela Zameer*, Muhammad Muneeb

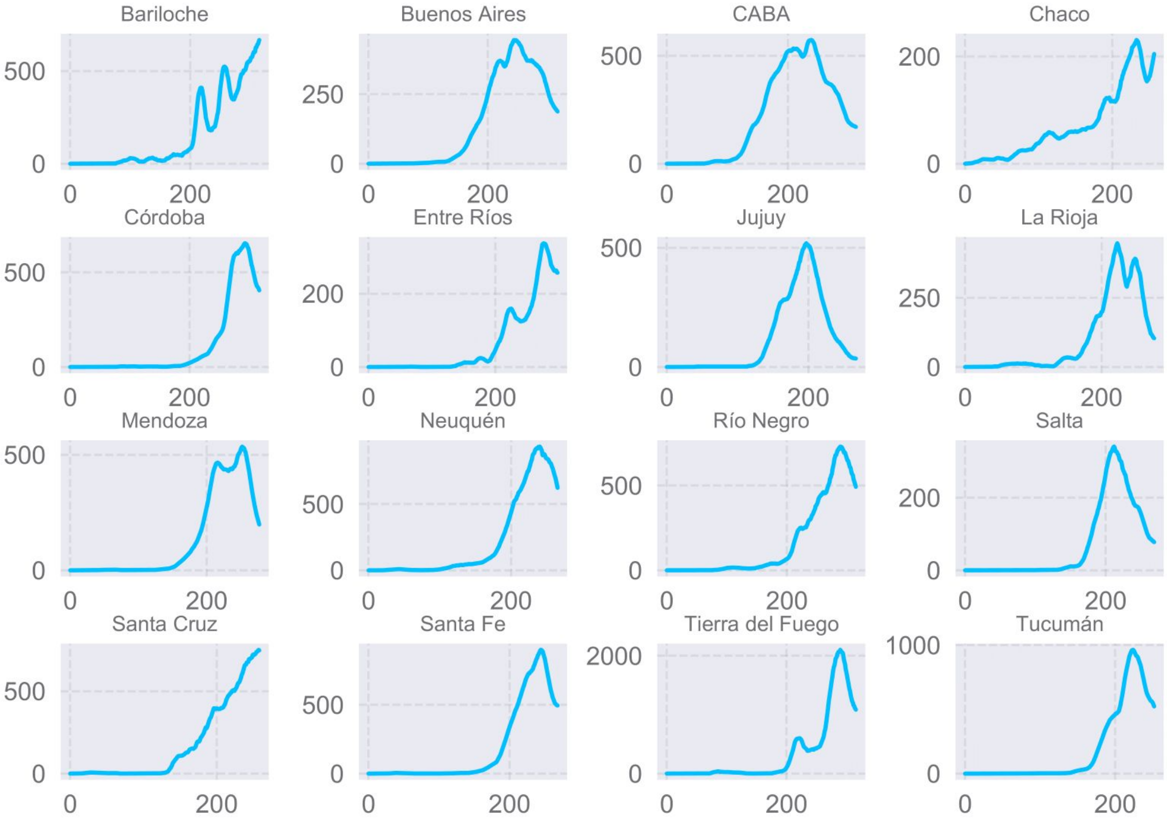
Department of Computer & Information Sciences, Pakistan Institute of Engineering & Applied Sciences (PIEAS), Nilore, Islamabad 45650, Pakistan

https://doi.org/10.1016/j.chaos.2020.110212



+ Riesgómetro Covid-19 Argentina: https://droyktton.github.io/loscoihues/index.html

Datos



Incidencia acumuladas Casos/100 mil habitantes hasta el 25/nov

16 sets de infectados / 100 mil habitantes

Calculados a partir de los datos oficiales del Ministerio de Salud:

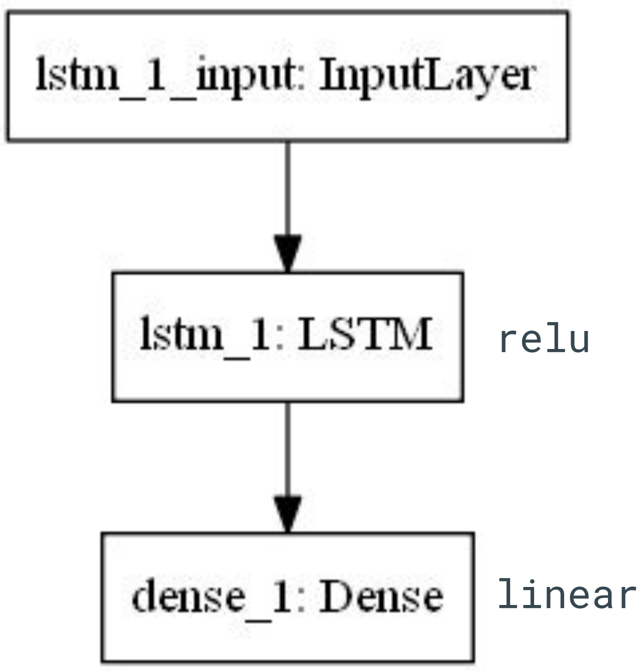
http://datos.salud.gob.ar/dataset?groups=covid-19

Los registros correspondientes a los últimos 10 días pueden ser incompletos por un retraso en la carga de datos

Preprocesado similar al del ej 1 práctica 7

Primera estrategia: LSTM

- Preprocesado de los datos similares al ejercicio 1 de la Práctica 7
- 80% train, 20% test
- Con un l = 3 en todos los casos
- Usando MaxMinScaler con un feature_range = (0,1)
- Intentando con una red determinada para cada ciudad analizada
- Por ejemplo, miremos los casos de Bariloche, CABA o Buenos Aires.



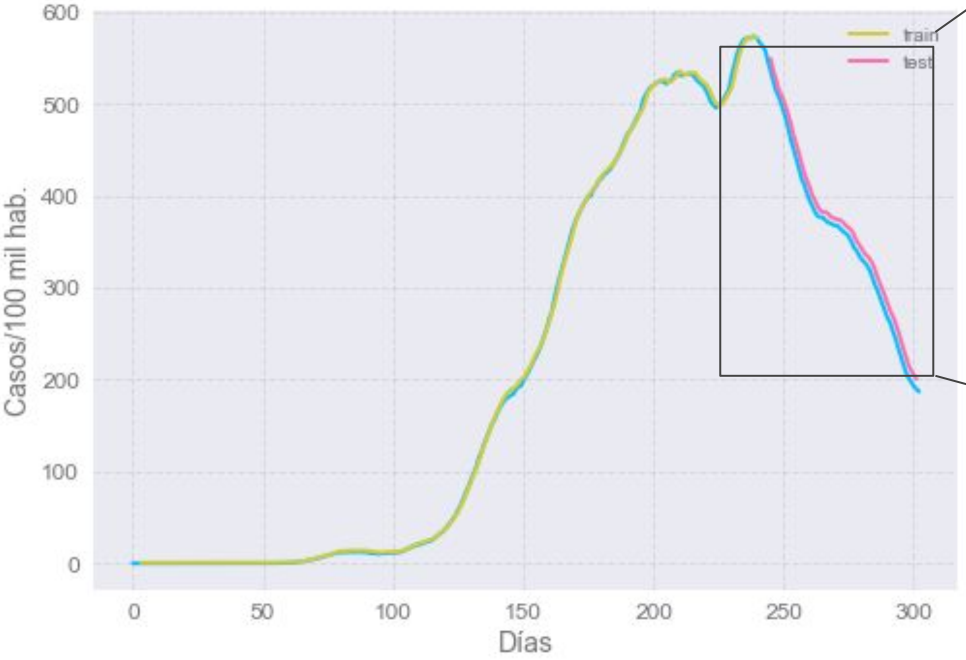
Bariloche

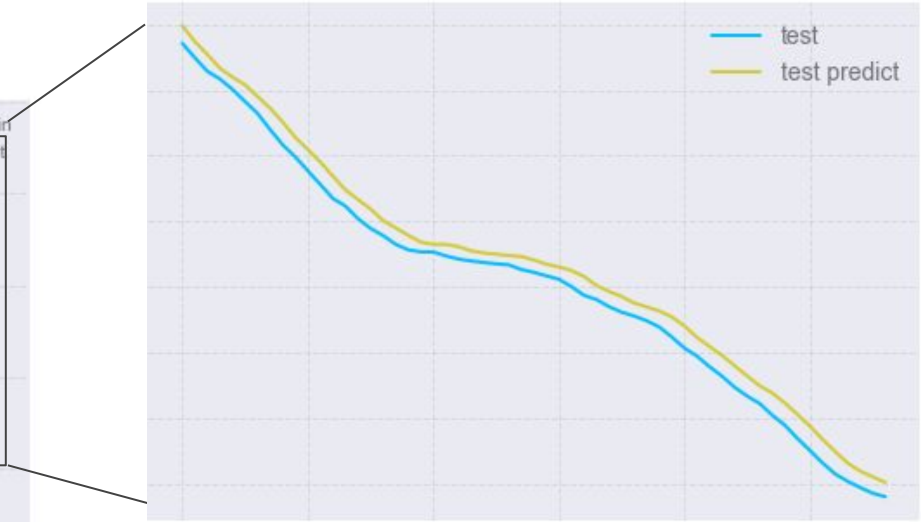
epochs = 500, batch_size = 16, neurons = 8
optimizer = adam, loss = mse

train mse: 11 train r2: 0.998
test mse: 133 test r2: 0.98

train mae squared: 58
test mae squared: 457

CABA





epochs = 500, batch_size = 8, neurons = 8
optimizer = adam, loss = mse

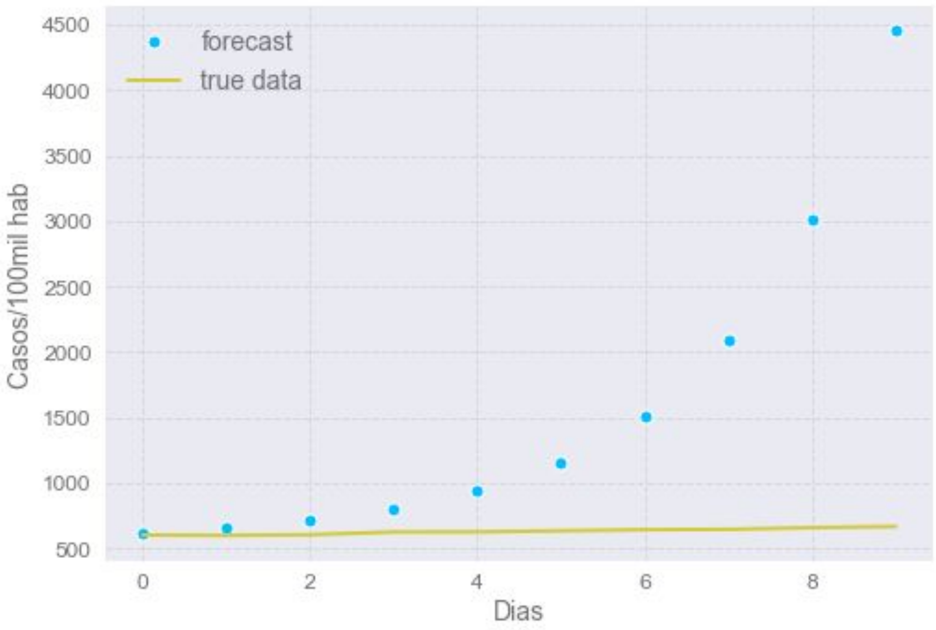
train mse: 8
test mse: 173

train r2: 0.999
test r2: 0.978

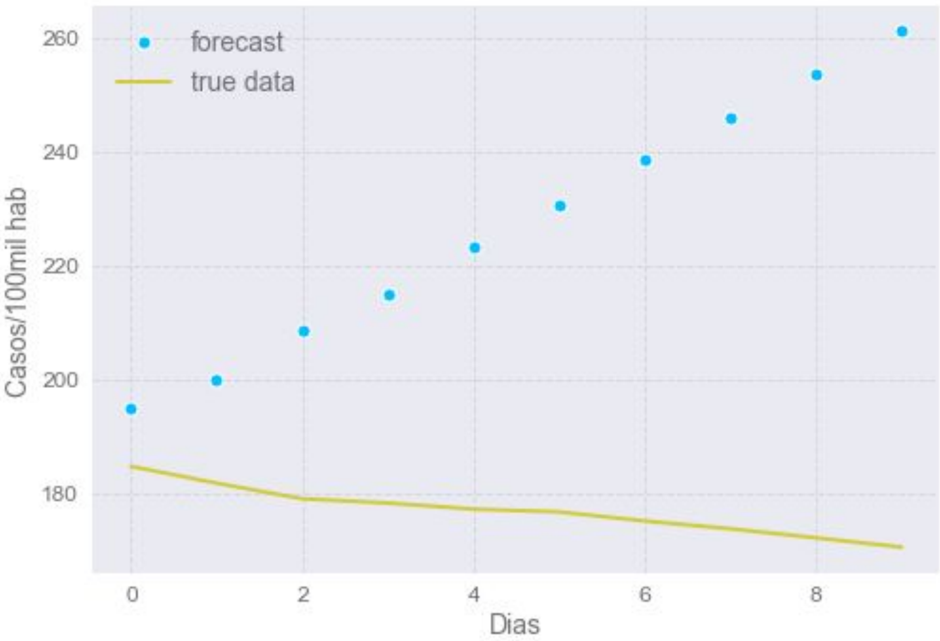
train mae squared: 188
test mae squared: 365

Forecasts

Bariloche



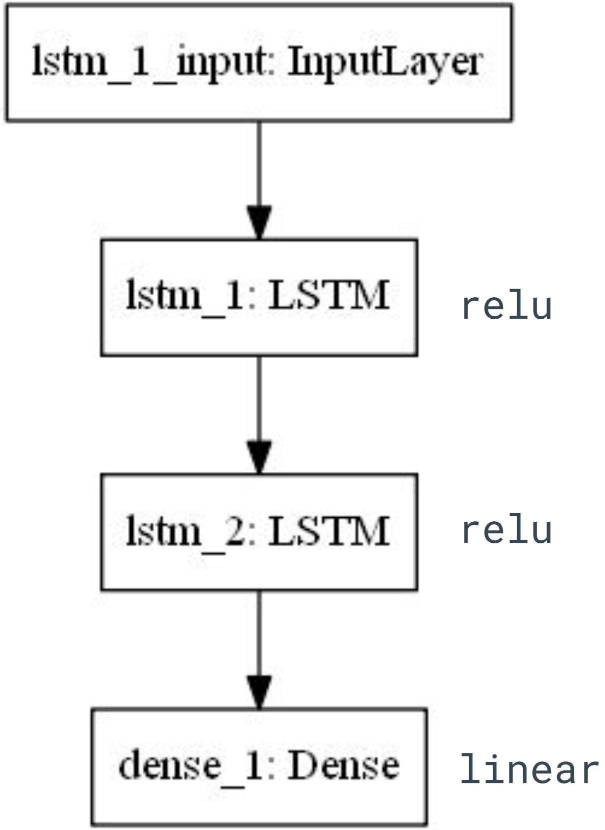
CABA



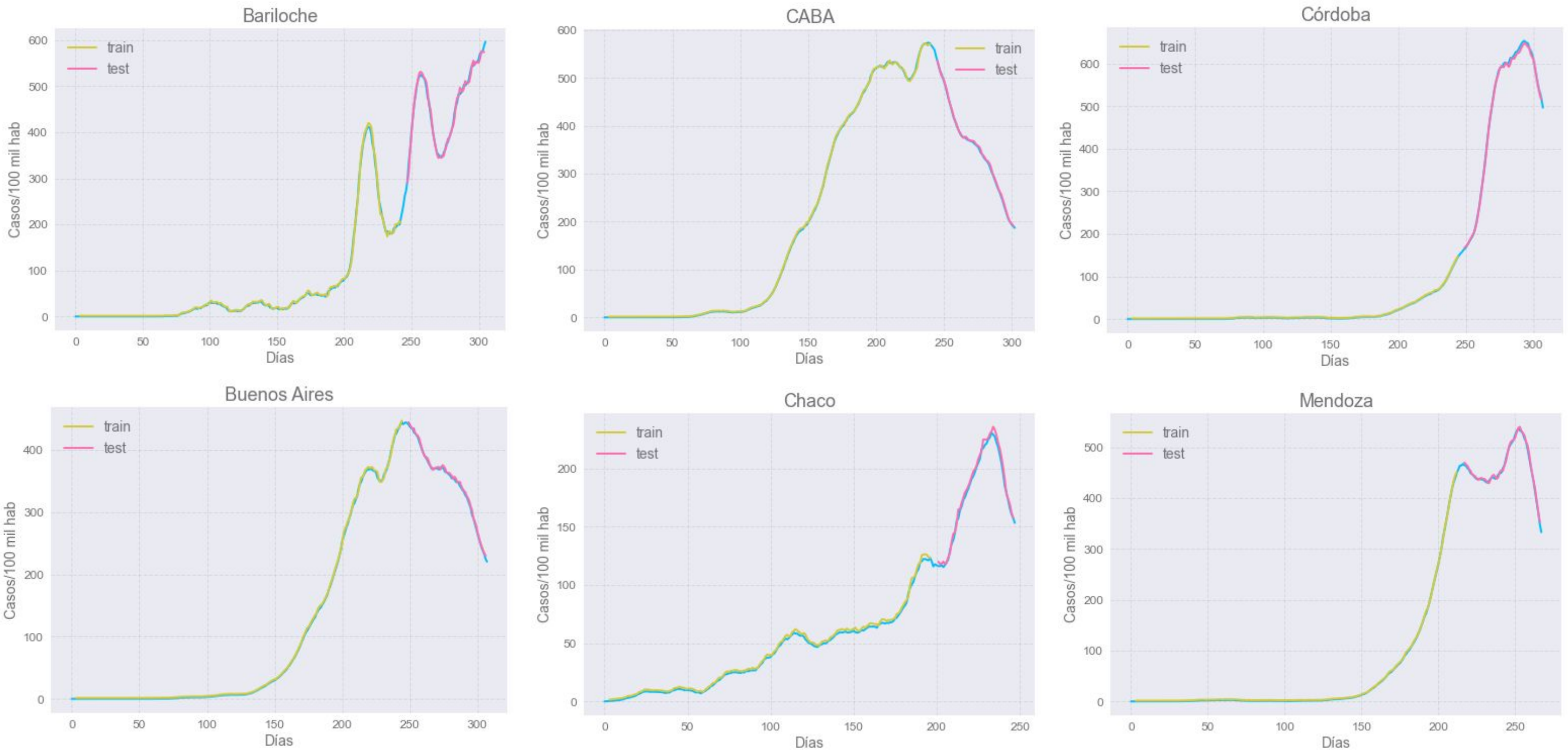
Segunda estrategia: Stacked LSTM

- Datos de 16 localidades de Argentina
- 80% train, 20% test de cada localidad, como en la bibliografía

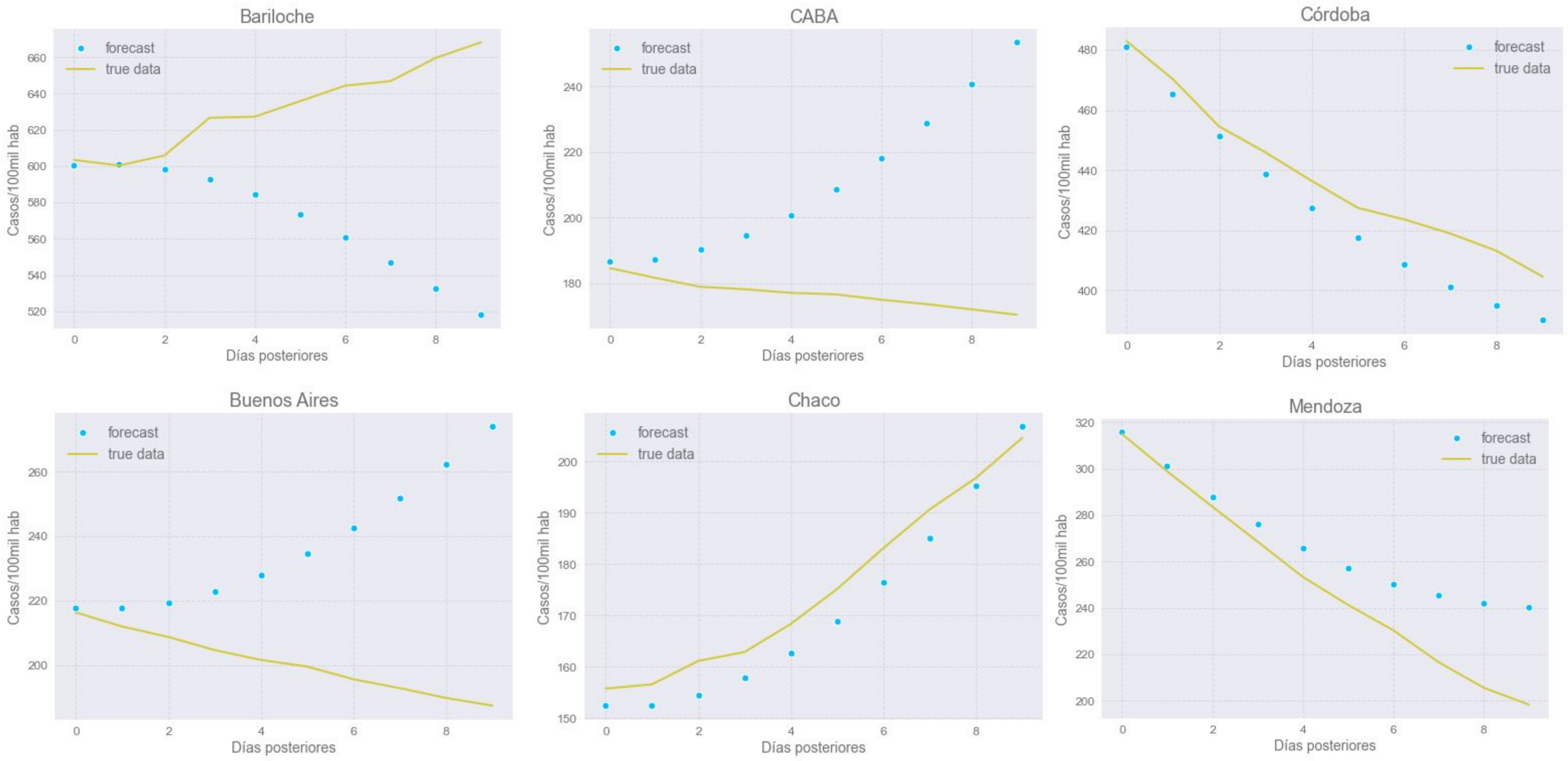
```
epochs = 1000
batch_size = 256
neurons = 128
optimizer = adam
loss = mse
```



Resultados sobre los datos de train y test



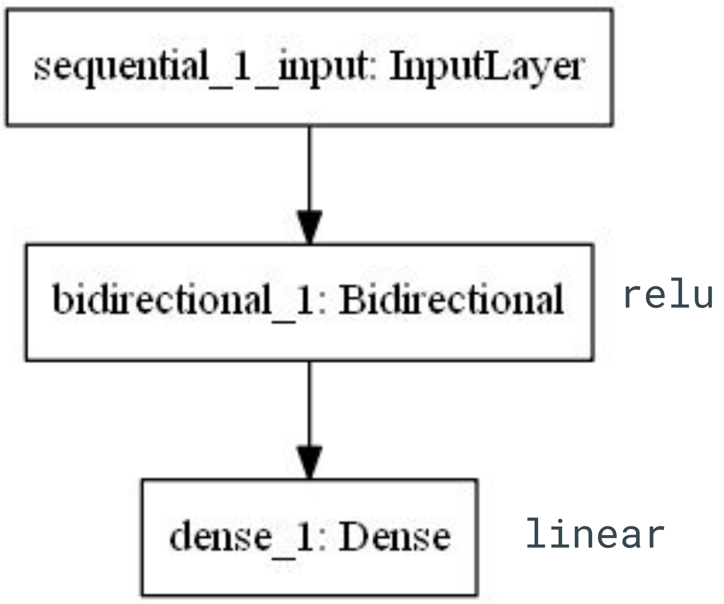
Resultados de forecast



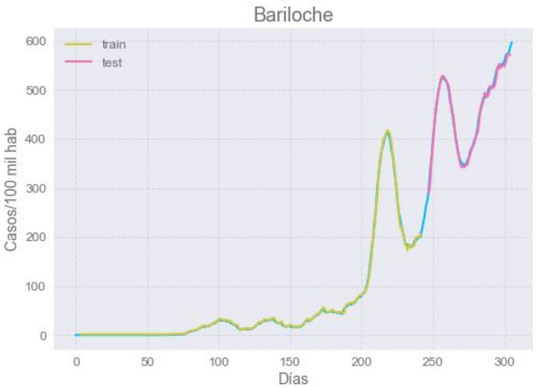
Tercera estrategia: Bidirectional LSTM

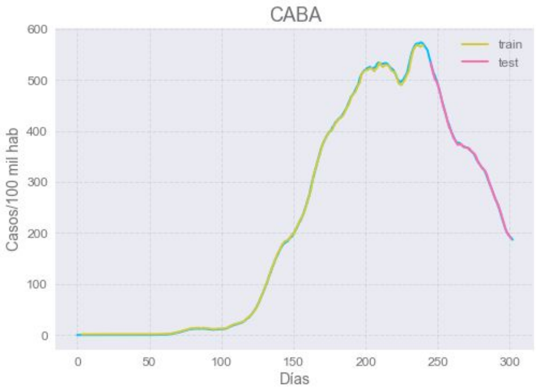
- Datos de 16 localidades de Argentina
- 80% train, 20% test de cada localidad, como en la bibliografía

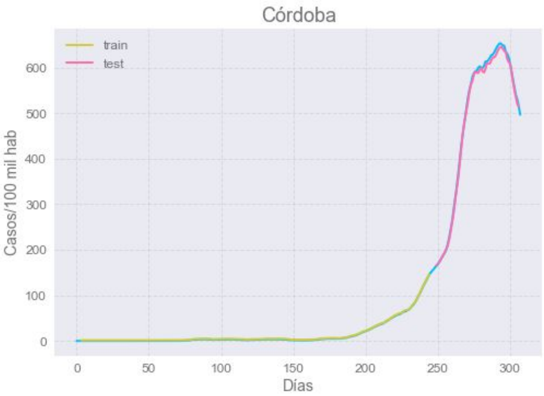
```
epochs = 1000
batch_size = 256
neurons = 128
optimizer = adam
loss = mse
```

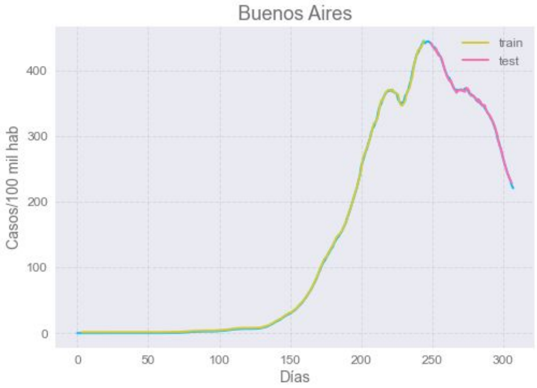


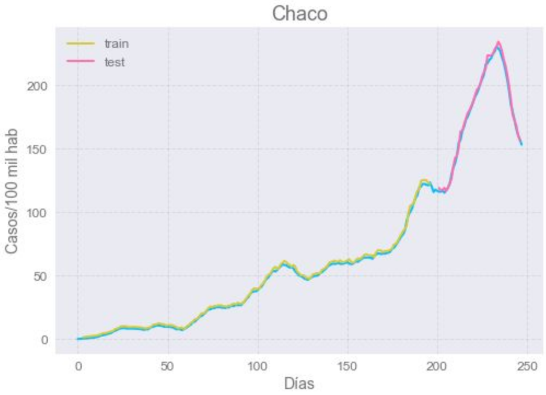
Resultados sobre los datos de train y test

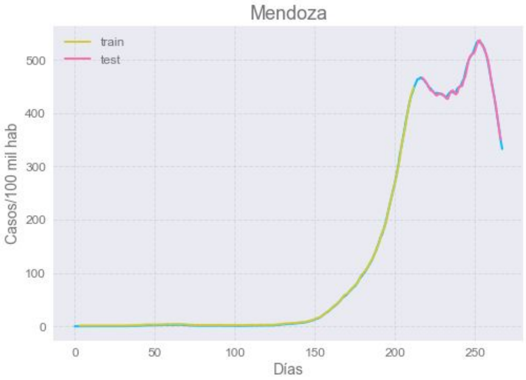




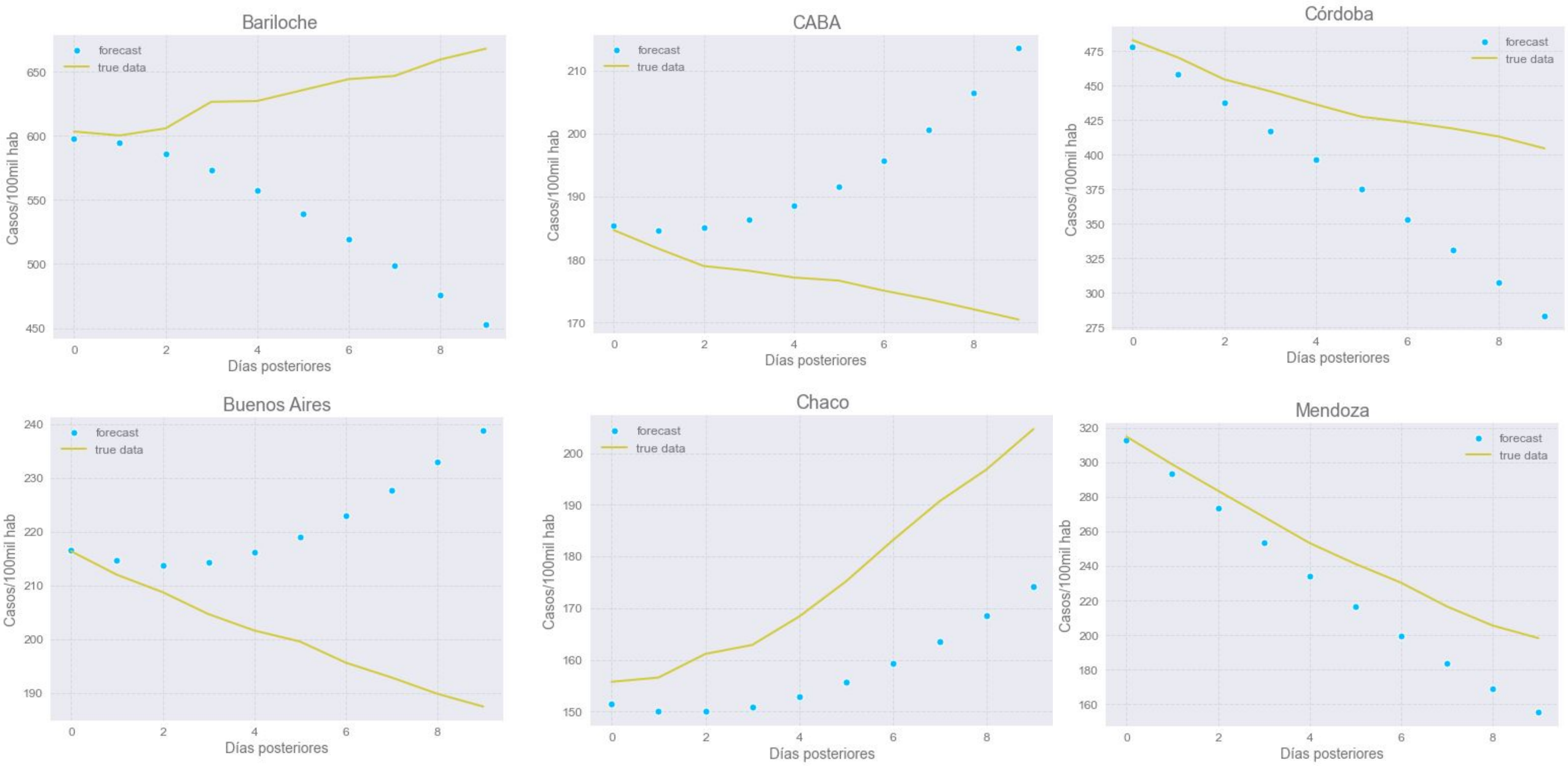








Resultados de forecast



Comparaciones

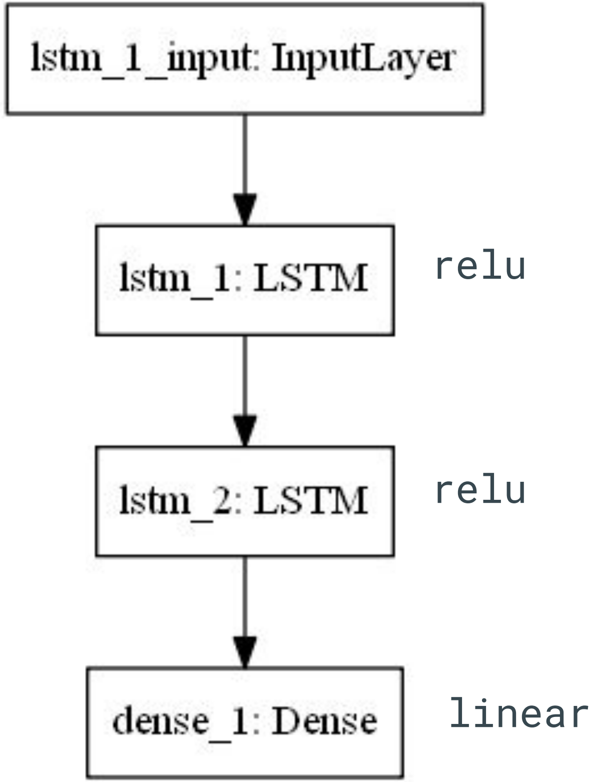
Lugar	MSE train	MSE test	R2 train	R2 test	MAE train	MAE test	Forecast MSE
CABA	8/4/6	173/8/5	0.999/0.999/ 0.999	0.978/0.999/ 0.999	188/2/2	364/2/2	3e4/2e4/4e3
Bs. As.	10/4/3	170/10/6	0.999/0.999/ 0.999	0.944/0.997/ 0.998	91/2/1	362/3/2	3e4/2e4/7e3
Bariloche	11/15/13	133/55/62	0.998/0.998/ 0.998	0.977/0.99/ 0.989	58/3/2	457/6/6	2e7/6e4/1e5
Córdoba	2/2	27/50	0.997/0.998	0.999/0.998	1/1	4/6	1e3/4e4
Chaco	4/3	13/9	0.996/0.997	0.989/0.994	2/1	3/2	2e2/4e3
Mendoza	2/2	14/13	0.999/0.999	0.991/0.992	1/1	3/3	4e2/6e3
Santa Fe	2/2	300/241	0.999/0.999	0.976/0.983	1/1	16/14	1e5/3e5
Río Negro	5/4	100/117	0.999/0.999	0.992/0.992	2/2	8/8	3e4/9e4

LSTM   Stacked LSTM   Bidirectional LSTM

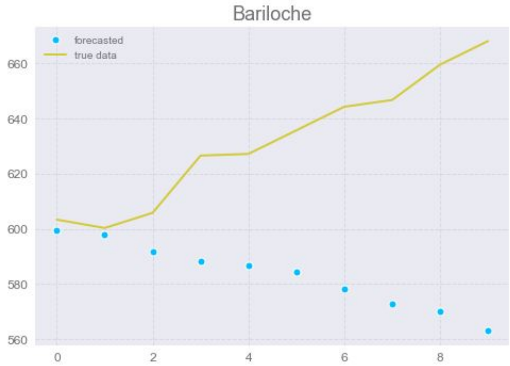
Cuarta estrategia

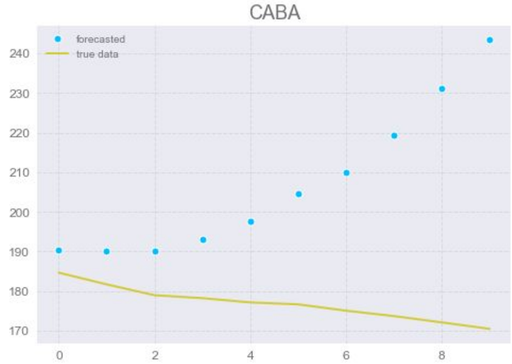
- Datos de 16 localidades de Argentina
- 80% train, 20% test de cada localidad, como antes
- 10 días -> 10 días posteriores

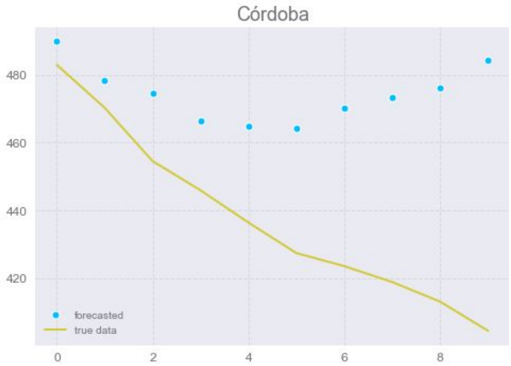
```
epochs = 1000
batch_size = 256
neurons = 32,32,10
optimizer = adam
loss = mse
```

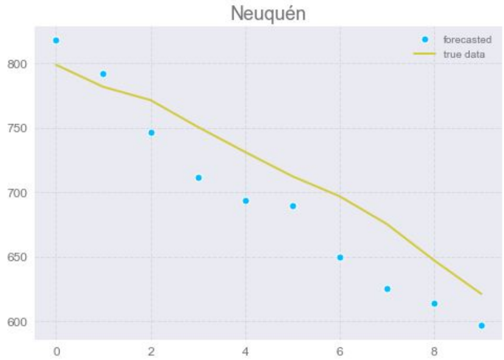


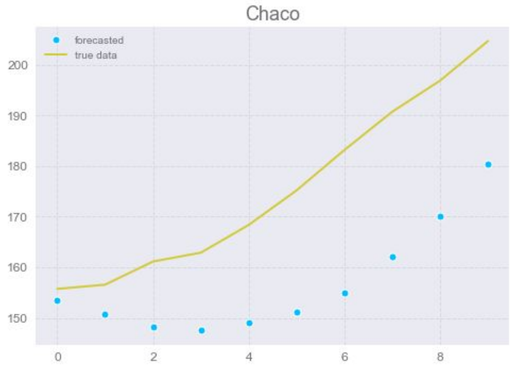
Resultados de forecast

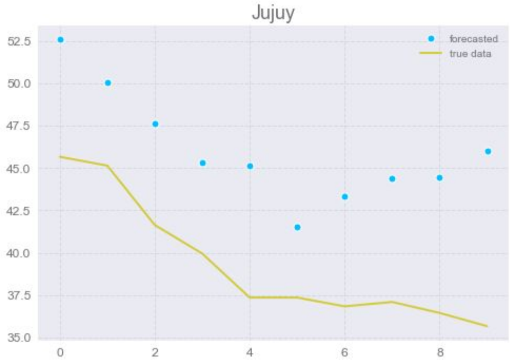












- Los valores de errores del forecast resultan ser menores, en algunos casos llega a funcionar mejor
- Para el resto de los indicadores, los resultados son tan buenos como los anteriores

Conclusiones y trabajo a futuro

Conclusiones

- Se estudia el uso de RNN para predicción a corto plazo de casos de covid-19 en Argentina.
- Se comparan las distintas estrategias para realizar estas predicciones.
- Se analiza la posibilidad de predecir casos con varios días de anticipación.
- Comparación con conclusiones/predicciones de la bibliografía.

Trabajo a futuro

Se podrían estudiar muchas cosas más relacionadas con el tema:

- Estudiar la posibilidad de usar datos de otros países/ usar más datos (https://arxiv.org/pdf/2005.04809v2.pdf)
- Utilizar otros modelos de predicción seq2seq (https://towardsdatascience.com/which-models-to-use-for-epidemic-prediction-25b22932c4ca)

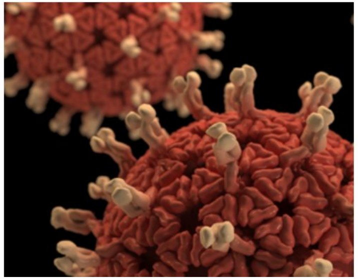
Otros problemas

Help us better understand COVID-19

Kaggle's platform is currently hosting a variety of challenges focused on better understanding COVID-19 . Join one of the below challenges to help the global community and health organizations stay informed and make data driven decisions.

View the contributions the community has already made on this page.

https://www.kaggle.com/covid19

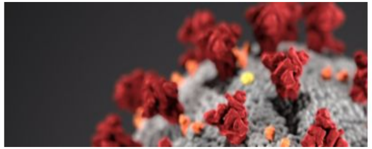


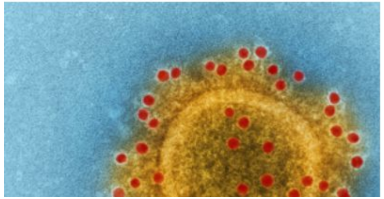
Use NLP to answer key questions from the scientific literature

Answer 9 key questions using natural language processing to help the world to understand COVID-19 faster. This challenge includes over 200,000 scholarly articles about COVID-19 and related coronaviruses. It was pulled together by the White House Office of Science and Technology Policy and several other partner institutions.

Forecast COVID-19 cases and fatalities

This challenge asked data scientists to predict the number of cases and





Use exploratory analysis to answer research questions that support frontline responders

Contribute research to some of the most pressing questions posed by frontline responders in healthcare and public policy in the Roche

COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images

Linda Wang[1,2,3*], Zhong Qiu Lin[1,2,3], and Alexander Wong[1,2,3]

efforts of the research community, in this study we introduce COVID-Net, a deep convolutional neural network design tailored for the detection of COVID-19 cases from chest X-ray (CXR) images that is open source and available to the general public. To the best of the authors' knowledge, COVID-Net is one of the first open source network designs for COVID-19 detection from CXR images at the time of initial release. We also introduce COVIDx, an open access benchmark dataset that we generated comprising of 13,975 CXR images across 13,870 patient patient cases, with the largest number of publicly available COVID-19 positive cases to the best of the authors' knowledge. Furthermore, we investigate how COVID-Net makes predictions using an

https://arxiv.org/pdf/2003.09871.pdf
(May 2020)

Fin