

# Aplicación de Deep Neural Network para la predicción de incidencias en COVID-19

Trabajo Final - Máster Universitario en Salud Digital

Autor: Samuel Paul Gallegos Serrano  
Director: Carlos Luis Sánchez Bocanegra  
Codirector: Gonzalo Aranda Pérez  
PRA: Francesc Saigi Rubió

# Índice

---

1. Introducción
2. Objetivos
3. Metodología
4. Resultados
5. Discusión
6. Conclusiones
7. Bibliografía



# Índice

---

1. **Introducción**
2. Objetivos
3. Metodología
4. Resultados
5. Discusión
6. Conclusiones
7. Bibliografía



# Introducción

---

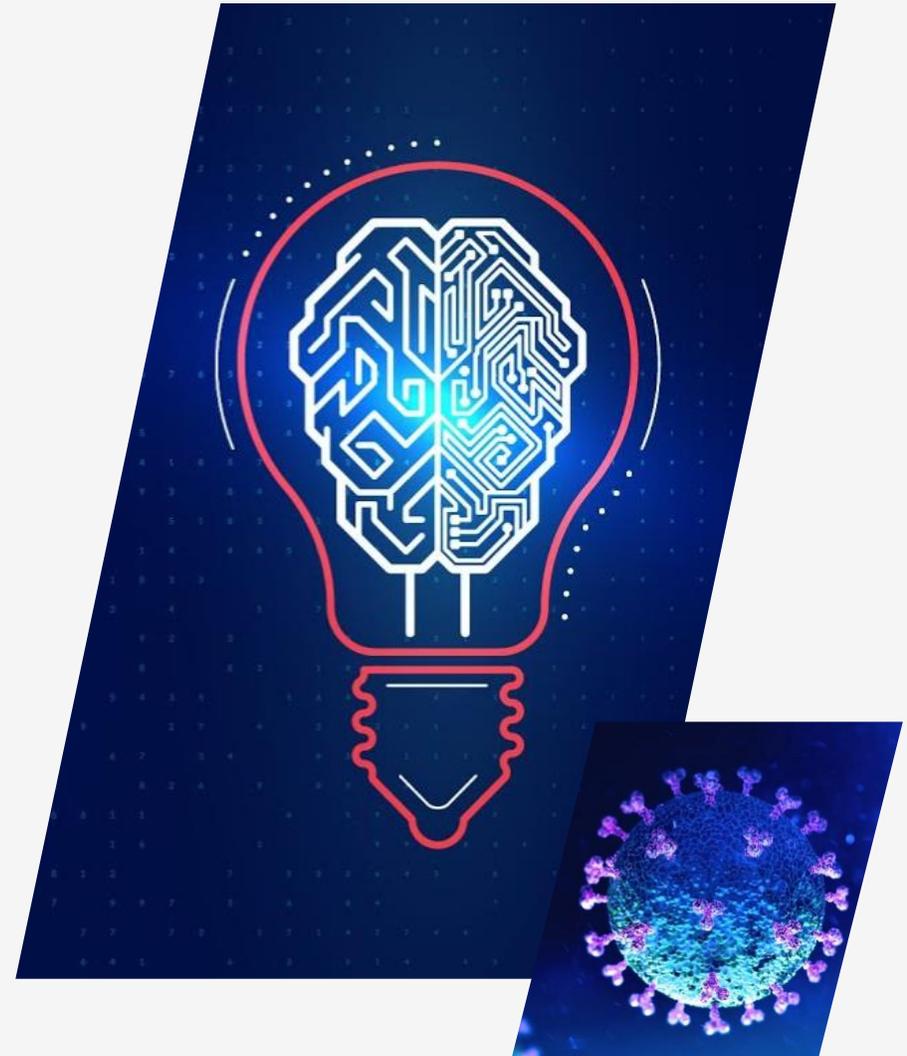
## ***Necesidad o problema a solventar con el proyecto***

Los sistemas de salud de todo el mundo están siendo **colapsados** por el **número creciente de casos de COVID-19** y la falta de recursos para hacerle frente.

Múltiples **países** han adoptado medidas de **distanciamiento social**, que se consideran clave para mermar la curva de contagios.

Diferentes **grupos de investigación** buscan **utilizar herramientas estadísticas y predictivas** que permitan encontrar soluciones para comprender mejor al virus, su propagación y **control**.

Es **crucial la planificación de la demanda** de atención asistencial con base en métodos predictivos.



# Introducción

---

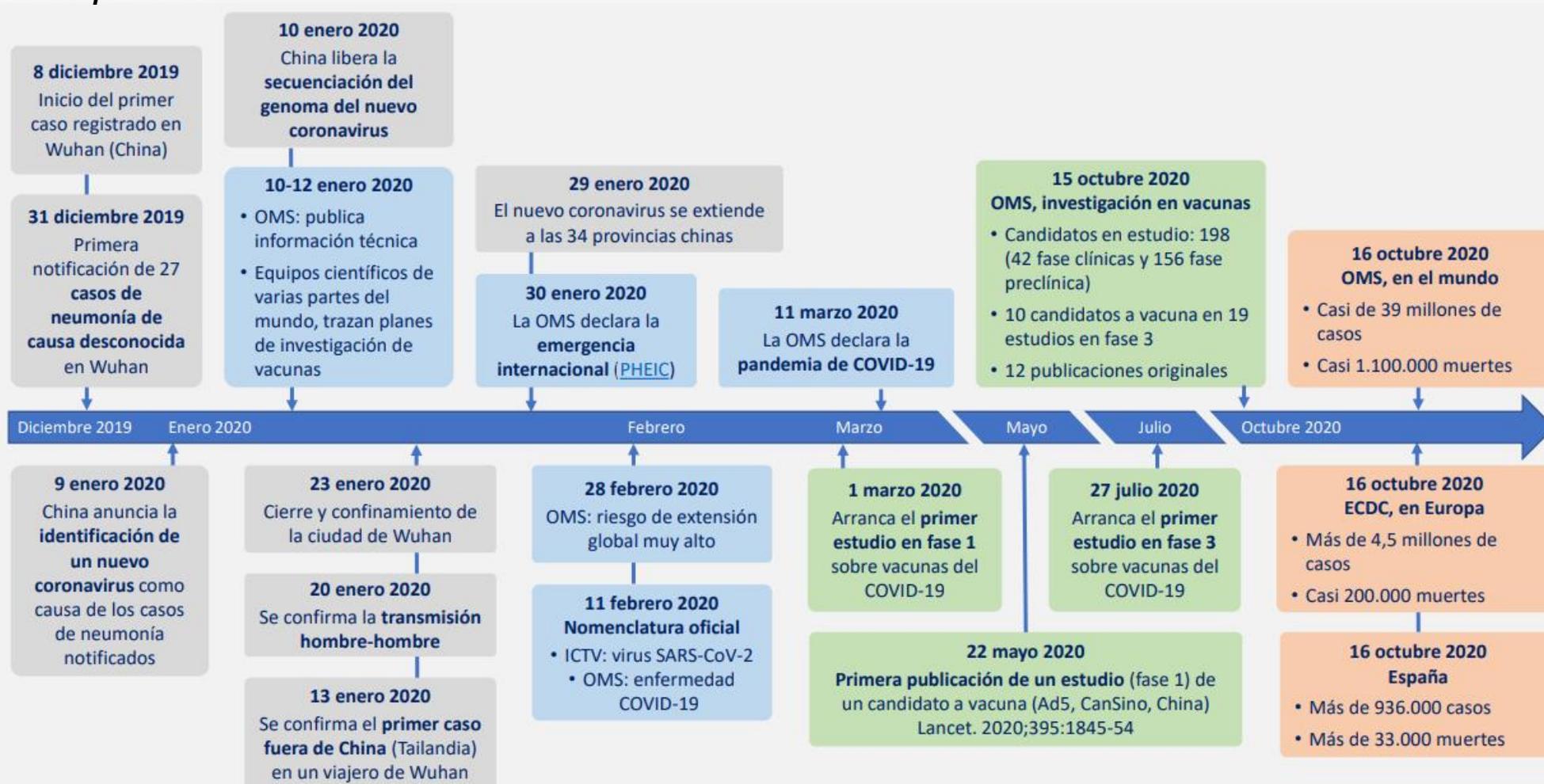
## *Necesidad o problema a solventar con el proyecto*

El presente trabajo utiliza un método de **Redes Neuronales Profundas** (del inglés, Deep Neural Networks, DNN) para la **predicción** de incidencias en **COVID-19** y lo **compara** con **métodos tradicionales de pronóstico**.



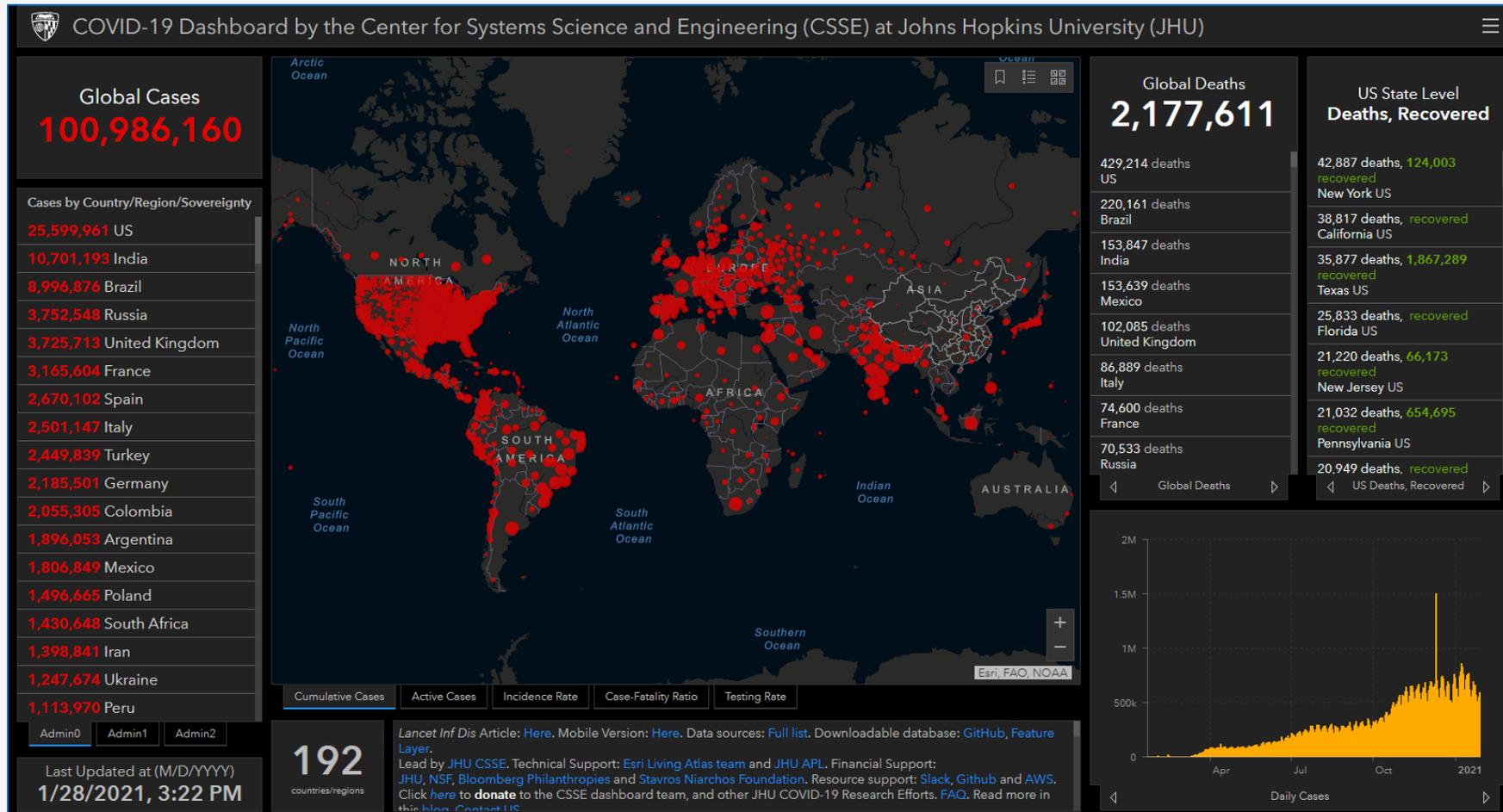
# Introducción

## Historia breve de la pandemia



# Introducción

## Historia breve de la pandemia



# Introducción

## Medidas contra el Covid-19

- 

**1** Quédate en casa si te sientes mal, aún si los síntomas son leves
- 

**2** Lava frecuentemente tus manos con agua y jabón o alcohol en gel
- 

**3** Cubre tu nariz y boca con un pañuelo desechable o el codo al estornudar o toser
- 

**4** Evita tocar tus ojos, nariz y boca
- 

**5** Mantén una distancia física mínima de 1 metro con otras personas
- 

**6** Evita aglomeraciones y permanecer en espacios poco ventilados
- 

**7** Utiliza un barbijo de tela cuando no puedas mantener el distanciamiento físico recomendado
- 

**8** Utiliza barbijo quirúrgico si perteneces a grupos de riesgo (por ej. edad avanzada, enfermedades pre-existentes)
- 

**9** Limpia y desinfecta regularmente las superficies de alto contacto

# Introducción

---

## *Medidas contra el Covid-19*

Después de un arduo trabajo de investigación y desarrollo, la **Unión Europea (UE)** tiene previsto dar su **visto bueno** a la **vacuna de Pfizer** el 21 de diciembre de 2020 y la de **Moderna**, el 12 de enero de 2021.



# Introducción

---

## *Recopilación de información para Covid-19*

Todos los **sistemas de predicción** requerirán de acceso a **grandes bases de datos**.

El 19 de noviembre de 2020 la **OMS** y otras agencias de la **ONU** emitieron una **“Declaración Conjunta sobre Protección de Datos y Privacidad en Respuesta a COVID-19”**

Our World  
in Data



statista 



GitHub

worldometers  
real time world statistics

# Introducción

## Recopilación de información para Covid-19

### Data on COVID-19 (coronavirus) by Our World in Data

Our complete COVID-19 dataset is a collection of the COVID-19 data maintained by [Our World in Data](#). It is updated daily and includes data on confirmed cases, deaths, hospitalizations, and testing, as well as other variables of potential interest.

📄 Download our complete COVID-19 dataset : [CSV](#) | [XLSX](#) | [JSON](#)

We will continue to publish up-to-date data on confirmed cases, deaths, hospitalizations, testing, and vaccinations, throughout the duration of the COVID-19 pandemic.

#### Our data sources

- **Confirmed cases and deaths:** our data comes from the [COVID-19 Data Repository by the Center for Systems Science and Engineering \(CSSE\) at Johns Hopkins University \(JHU\)](#). We discuss how and when JHU collects and publishes this data [here](#). The cases & deaths dataset is updated daily. *Note: the number of cases or deaths reported by any institution—including JHU, the WHO, the ECDC and others—on a given day does not necessarily represent the actual number on that date. This is because of the long reporting chain that exists between a new case/death and its inclusion in statistics. This also means that negative values in cases and deaths can sometimes appear when a country corrects historical data, because it had previously overestimated the number of cases/deaths. Alternatively, large changes can sometimes (although rarely) be made to a country's entire time series if JHU decides (and has access to the necessary data) to correct values retrospectively.*
- **Hospitalizations and intensive care unit (ICU) admissions:** our data comes from the [European Centre for Disease Prevention and Control \(ECDC\)](#) for a select number of European countries; the [government of the United Kingdom](#); the [COVID Tracking Project](#) for the United States; the [COVID-19 Tracker](#) for Canada. Unfortunately, we are unable to provide data on hospitalizations for other countries: there is currently no global, aggregated database on COVID-19 hospitalization, and our team at *Our World in Data* does not have the capacity to build such a dataset.
- **Testing for COVID-19:** this data is collected by the *Our World in Data* team from official reports; you can find further details in our post on COVID-19 testing, including our [checklist of questions to understand testing data](#), information on [geographical and temporal coverage](#), and [detailed country-by-country source information](#). The testing dataset is updated around twice a week.
- **Vaccinations against COVID-19:** this data is collected by the *Our World in Data* team from official reports.
- **Other variables:** this data is collected from a variety of sources (United Nations, World Bank, Global Burden of Disease, Blavatnik School of Government, etc.). More information is available in our [codebook](#).

Commit	Message	Time
edomt	Vaccination update	1 hour ago
..		
ecdc	Move hospital grapher dataset	29 days ago
excess_mortality	Update README.md	4 hours ago
jhu	Automated JHU update	10 hours ago
latest	Vaccination update	1 hour ago
testing	Data update	yesterday
vaccinations	Vaccination update	1 hour ago
who	Copy WHO data over to /who	10 months ago
README.md	Add change for Sweden c&d data to README	22 hours ago
owid-covid-codebook.csv	Add all vax metrics to complete dataset	2 days ago
owid-covid-data-last-updated-timestamp.txt	Vaccination update	1 hour ago
owid-covid-data.csv	Vaccination update	1 hour ago
owid-covid-data.json	Vaccination update	1 hour ago
owid-covid-data.xlsx	Vaccination update	1 hour ago

## scientific data

Explore Content ▾ Journal Information ▾ Publish With Us ▾

nature > scientific data > data descriptors > article

Data Descriptor | [Open Access](#) | Published: 08 October 2020

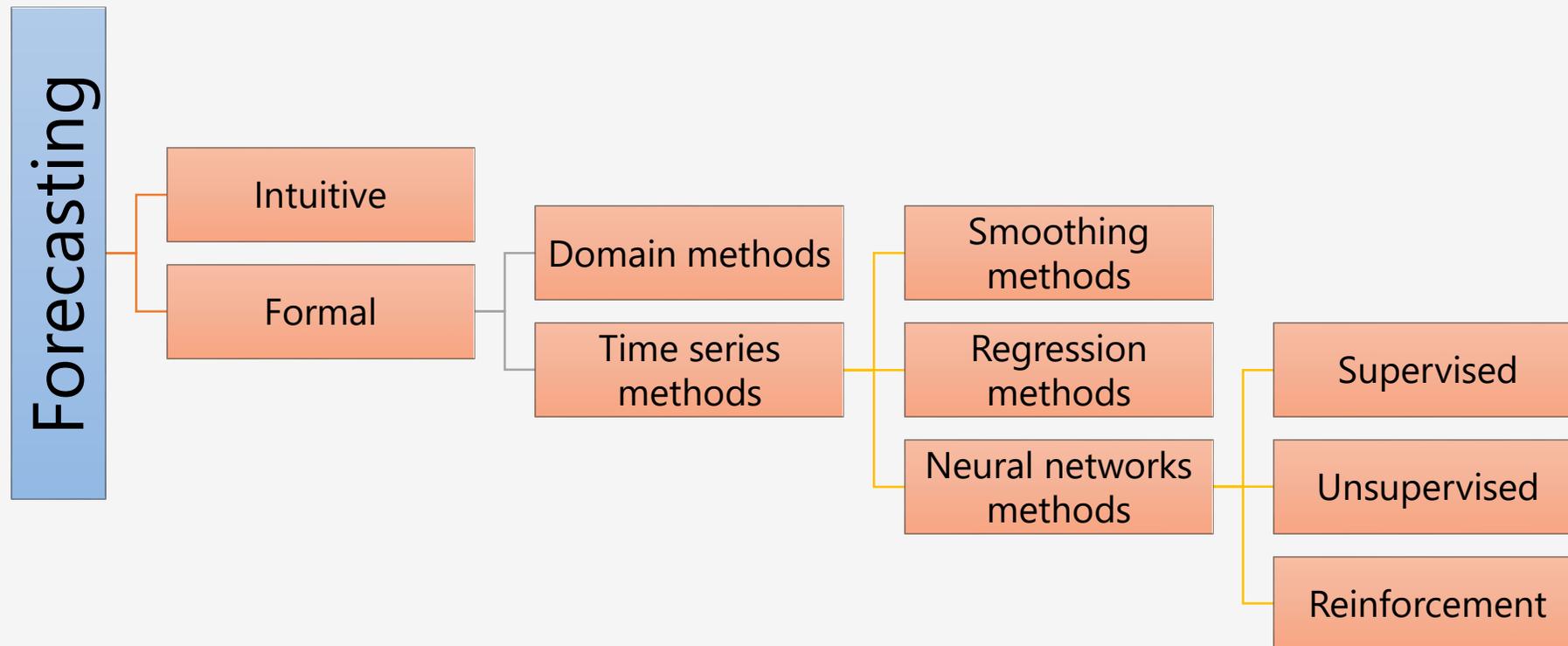
### A cross-country database of COVID-19 testing

Joe Hasell [✉](#), Edouard Mathieu, Diana Beltekian, Bobbie Macdonald, Charlie Giattino, Esteban Ortiz-Ospina, Max Roser & Hannah Ritchie

# Introducción

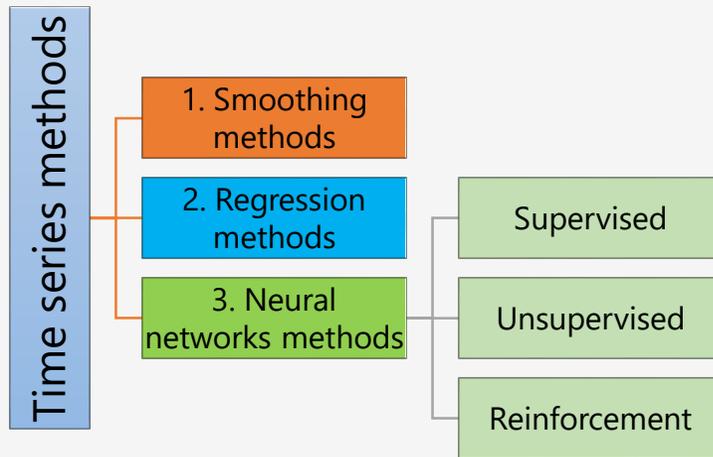
---

## *Estado del arte de los métodos predictivos para Covid-19*



# Introducción

## Estado del arte de los métodos predictivos para Covid-19



1. Los **métodos de suavizado** producen una serie temporal que elimina ruidos o comportamientos no deseables de los datos para hacer un patrón más visible.

Un ejemplo muy utilizado es el **método de Holt y el método de Winter**, que pueden aplicarse a conjuntos de datos con tendencias variables.

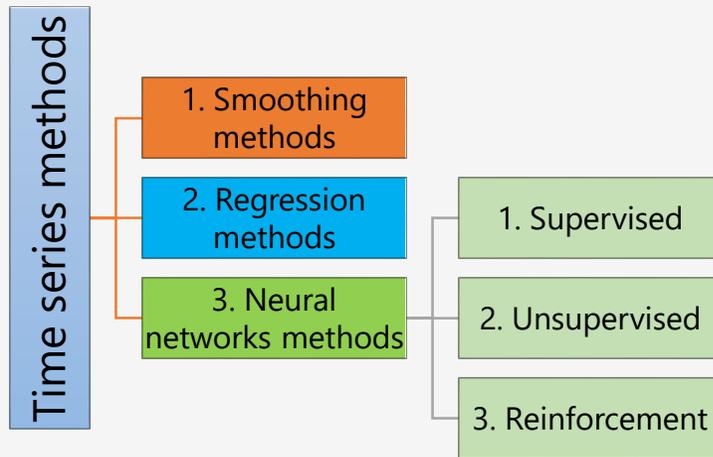
2. Los **métodos de regresión** buscan una relación entre una variable dependiente, con respecto a una o más variables independientes. Así mismo determinan el posible impacto de las variables independientes en la dependiente. En el caso que el tiempo sea una de estas variables, se debe tomar en cuenta que la serie puede exhibir estacionalidad. Los más comúnmente usados son:

**Métodos Autorregresivo Integrado de Media Móvil (ARIMA)**, el cual es una técnica estadística que utiliza datos de series temporales con el fin de encontrar patrones que predicen el futuro. Sustancialmente, sus predicciones se explican por datos pasados y no por variables independientes. Utiliza procesos autorregresivos y de medias móviles.

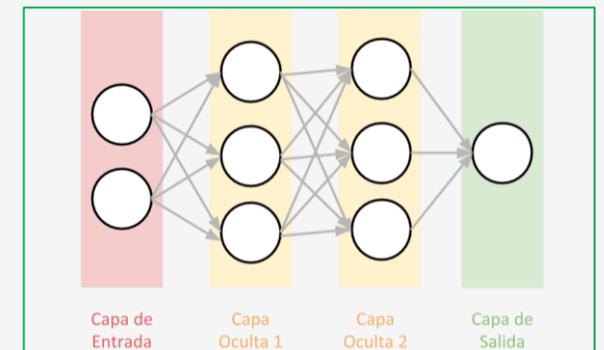
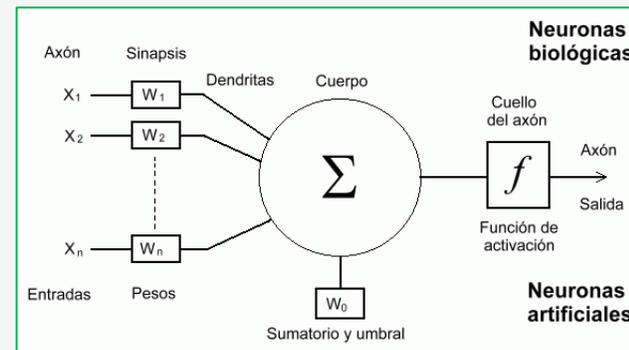
**Mínimos Cuadrados Ordinarios**, el cual es un método para encontrar los parámetros poblacionales en un método de regresión lineal, en otras palabras, consiste en la obtención de una recta que disminuya la suma de las distancias entre las respuestas observadas en la muestra y las respuestas del método. El parámetro resultante se expresa en una fórmula matemática sencilla.

# Introducción

## Estado del arte de los métodos predictivos para Covid-19



3. Los **métodos de redes neuronales** (del inglés, Neural Network, NN) son aquellos que utilizan el Aprendizaje Automático (del inglés, Machine Learning, ML), rama de la Inteligencia Artificial (del inglés, Artificial Intelligence, AI), para simular el comportamiento de su homólogo biológico.(30)



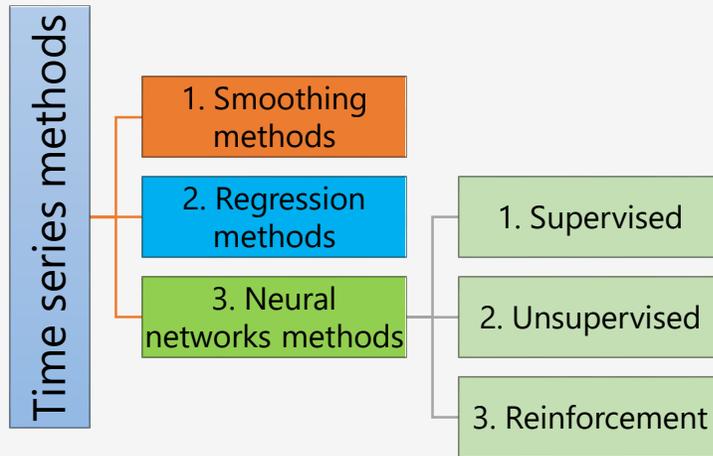
**1. Supervisadas**, cuando se aprenden funciones o relaciones que se asocian a los datos entrada y de salida, por tanto se ajustan a una serie de ejemplos que conocemos la salida deseada, usadas principalmente para temas predictivos.

**2. No supervisado**, en donde el interés de la causalidad de los datos de entrada y salida no es de suma importancia, si no tener conocimiento de la estructura de los datos, por ejemplo, en segmentación del mercado por tipo de clientes.

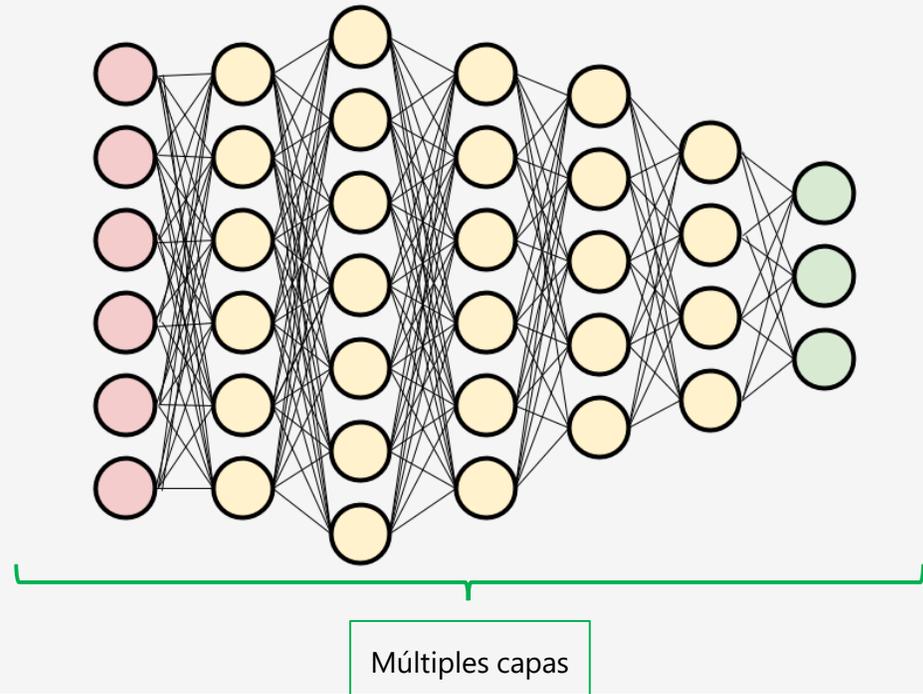
**3. Por refuerzo**, en la que existe una comparación de las entradas y salidas deseadas y si estas salidas no se corresponden a lo deseado, se ajustarán hasta lograr por probabilidad el mejor desempeño posible, usado por ejemplo para vehículos autónomos.

# Introducción

## Estado del arte de los métodos predictivos para Covid-19



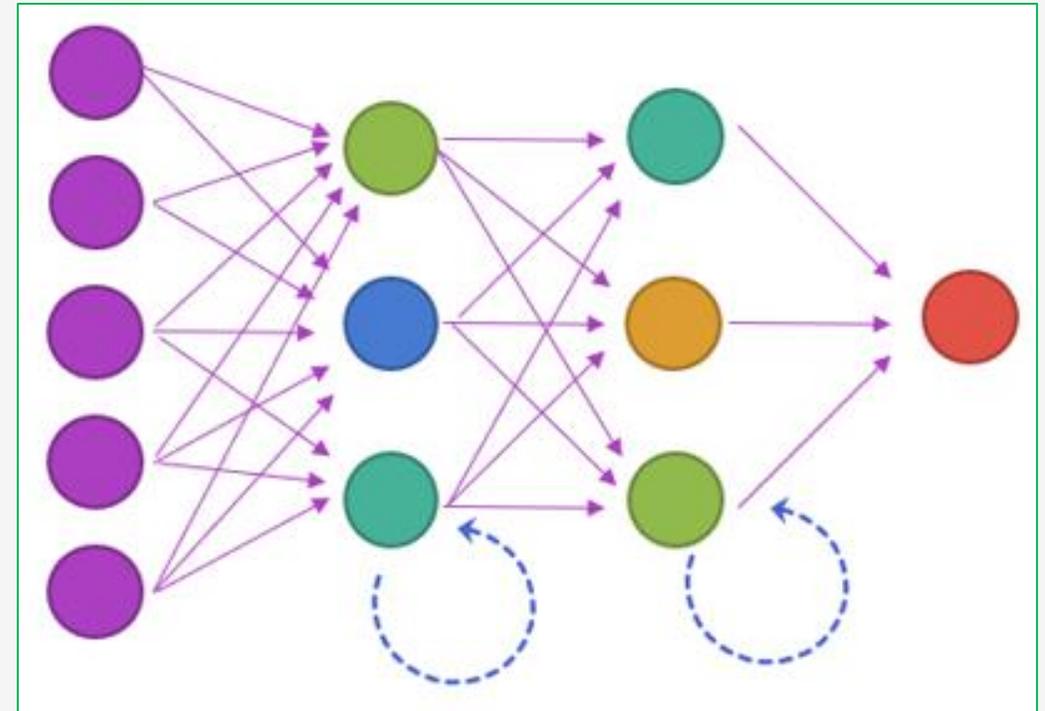
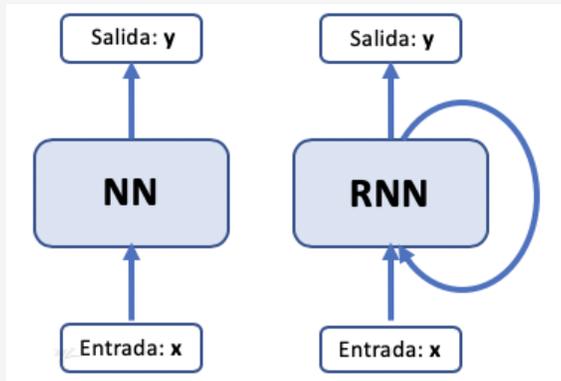
Las **Redes Neuronales Profundas** (del inglés, Deep Neural Networks, DNN) son una rama del Aprendizaje Automático (del inglés, Machine Learning, ML) **El número de capas es lo que determina la profundidad del método**, más capas, más profundidad y por tanto más neuronas para procesar los datos introducidos.



# Introducción

## Red neuronal a usar para predicción de datos Covid-19

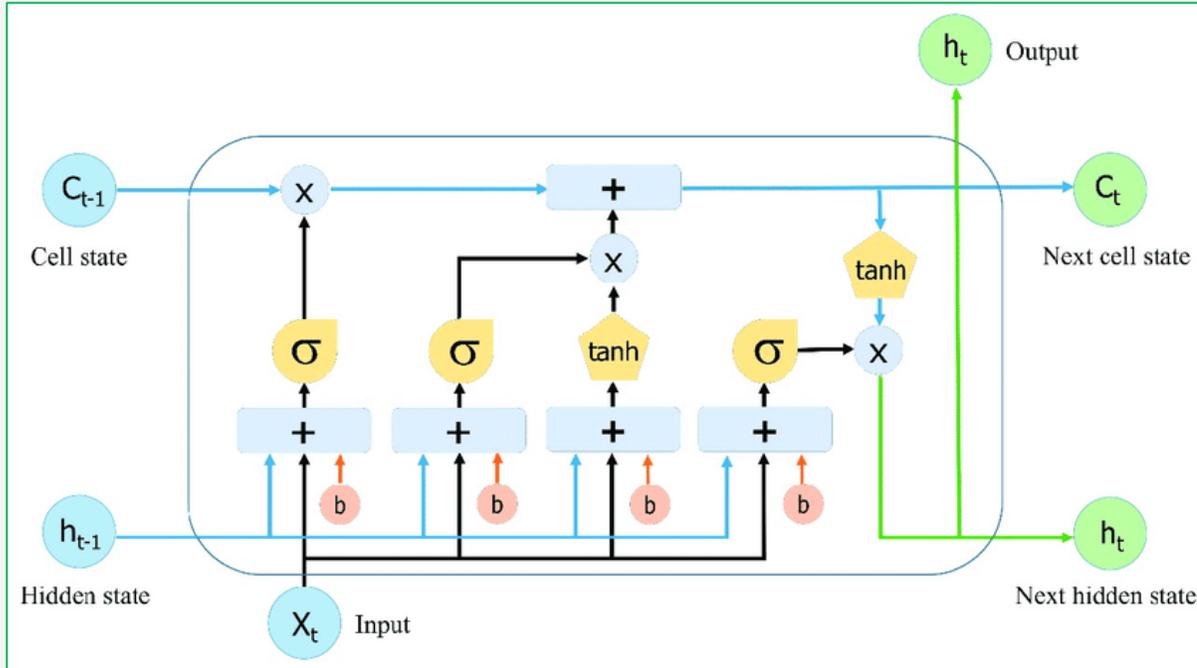
**Memoria a Largo – Corto Plazo** (del inglés, Long – Short Term Memory, LSTM). Son redes del tipo **aprendizaje supervisado**, integran **bucles de realimentación** (del inglés, Recurrent Neural Network, RNN), permitiendo que la información persista durante algunos pasos o épocas de entrenamiento (del inglés, epochs). Finalmente esta información pasa a las salidas de las capas y sus resultados se incrustan (del inglés, embedding) nuevamente en los datos de entrada.



El método **LSTM** está organizado estructuralmente en cadena, pero a diferencia de una red neuronal recurrente estándar, **tiene cuatro capas** que interactúan con un método de comunicación único.

# Introducción

## Red neuronal a usar para predicción de datos Covid-19



Inputs:	Outputs:	Nonlinearities:	Vector operations:
$X_t$ Current input	$C_t$ New updated memory	$\sigma$ Sigmoid layer	$\times$ Scaling of information
$C_{t-1}$ Memory from last LSTM unit	$h_t$ Current output	$\tanh$ Tanh layer	$+$ Adding information
$h_{t-1}$ Output of last LSTM unit		$b$ Bias	

Estructura de la red neuronal LSTM. Fuente: Giha Lee.

# Índice

---

1. Introducción
2. **Objetivos**
3. Metodología
4. Resultados
5. Discusión
6. Conclusiones
7. Bibliografía

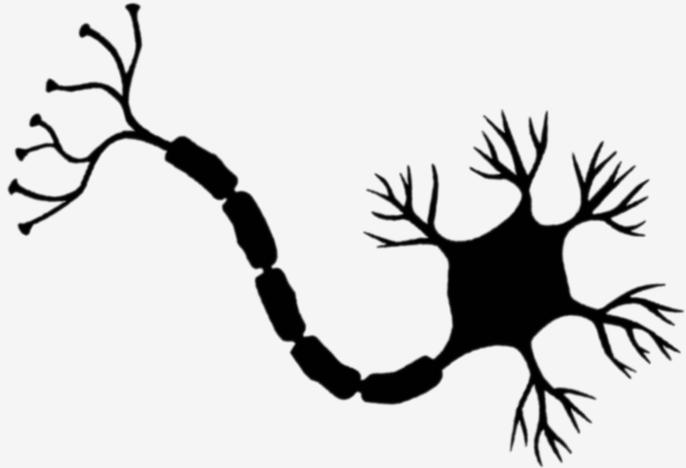


# Objetivos

---

## Objetivo principal

OP. Entrenar el método de Redes Neuronales Profundas (del inglés, Deep Neural Networks, DNN) con la base de datos de COVID-19 para comprobar su efectividad para **predecir y tomar decisiones en la pandemia.**



---

## Pregunta investigable

PI. ¿**Pueden** los métodos de Redes Neuronales Profundas (del inglés, Deep Neural Networks, DNN) **ayudar en la toma de decisiones epidemiológicas** y logísticas en la pandemia actual de COVID-19?

## Objetivos secundarios

OS1. Realizar una **revisión bibliográfica** para conocer el estado del arte en predicción de series temporales en datos de salud y en la pandemia de COVID-19.

OS2. Realizar un **análisis de los métodos predictivos y su efectividad** en relación con los datos de COVID-19.

OS3. **Identificar las bases** de datos más confiables en relación con COVID-19.

OS4. **Identificar los métodos** que logren una efectividad predictiva considerable con las bases de datos de COVID-19 en Italia.

OS5. **Entrenar el método** de Redes Neuronales Profundas (del inglés, Deep Neural Networks, DNN) y comprobar su efectividad con la base de datos de COVID-19 en Italia.

OS6. **Probar el método** de Redes Neuronales Profundas (del inglés, Deep Neural Networks, DNN) con datos específicos de movilidad en Barcelona, Madrid y Málaga para evidenciar su utilidad en España ante la pandemia de COVID-19.

# Índice

---

1. Introducción
2. Objetivos
3. **Metodología**
4. Resultados
5. Discusión
6. Conclusiones
7. Bibliografía



# Metodología

---



Se toma como referencia la **metodología KDD** (del inglés, Knowledge Discovery in Databases).

# Índice

---

1. Introducción
2. Objetivos
3. Metodología
4. **Resultados**
5. Discusión
6. Conclusiones
7. Bibliografía



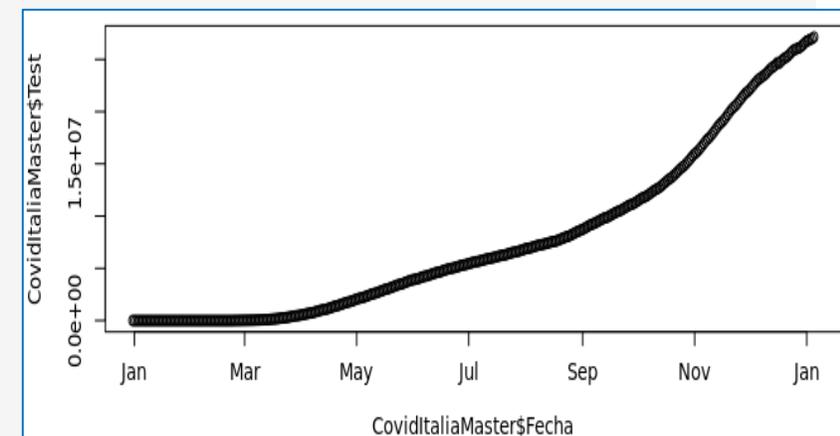
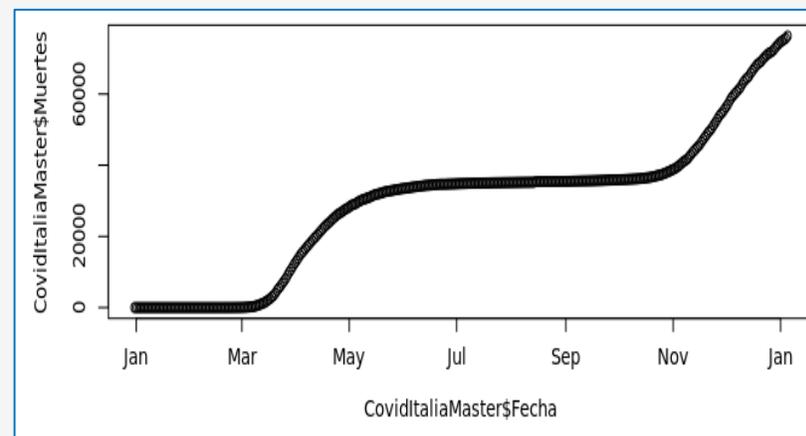
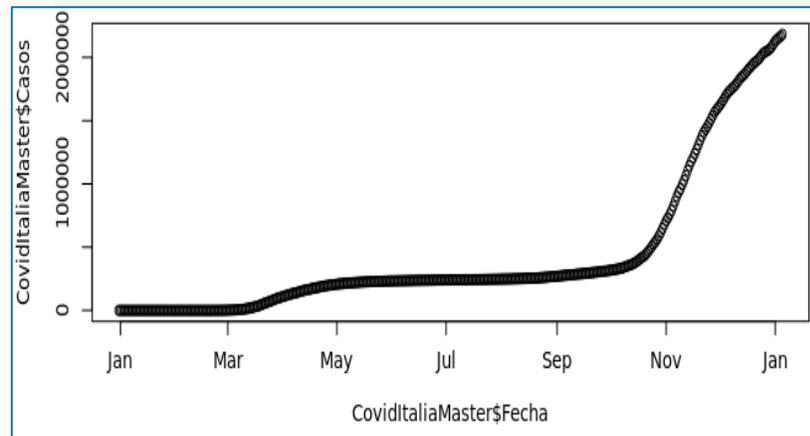
# Resultados

Se realiza un **análisis exploratorio** de los datos, entre el **31/12/2019 y 05/01/2021** (372 días).

- Cuatro variables (Fecha, Casos, Muertes y Test).
- Casos 2.181.619
- Muertes 76.329
- Test 27.139.378

```
> summary(CovidItaliaMaster)
      Fecha      Casos      Muertes      Test
Min.   :2019-12-31 00:00:00  Min.   :      0  Min.   :      0  Min.   :      0
1st Qu.:2020-04-01 18:00:00  1st Qu.: 114075  1st Qu.: 13725  1st Qu.: 571280
Median :2020-07-03 12:00:00  Median : 241302  Median : 34844  Median : 5574820
Mean   :2020-07-03 12:00:00  Mean   : 435684  Mean   : 29461  Mean   : 7736272
3rd Qu.:2020-10-04 06:00:00  3rd Qu.: 325893  3rd Qu.: 35990  3rd Qu.: 11799165
Max.   :2021-01-05 00:00:00  Max.   :2181619  Max.   : 76329  Max.   : 27139378
```

Análisis exploratorio de datos COVID-19 en Italia



Número de **casos, muertes y test** de COVID-19 en Italia

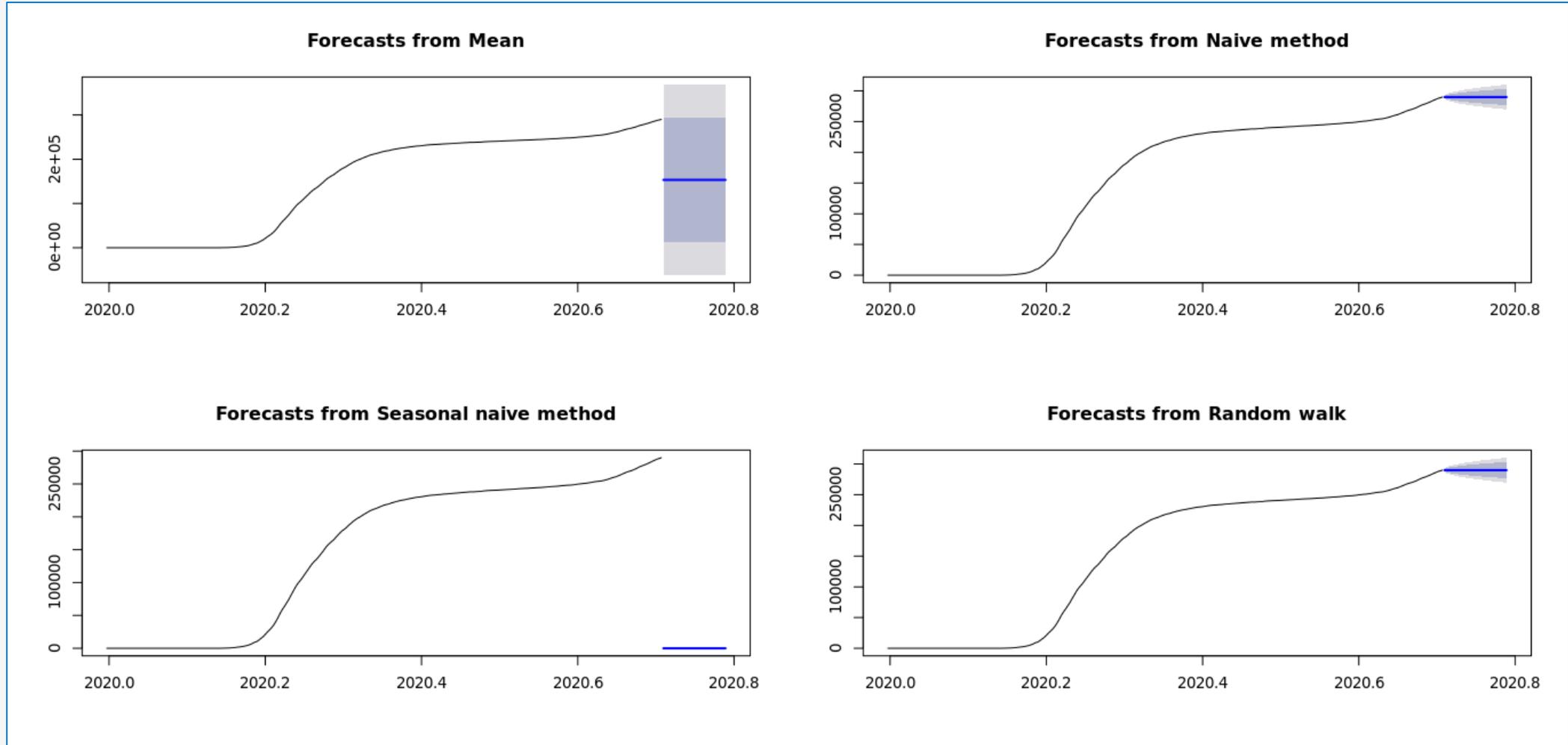
# Resultados

Al analizar los datos de casos, muertes y test de COVID-19 en Italia con **métodos pronósticos clásicos (Mean, Naive, RWF y SNaive)** a 30 días, se observó que estos no cuentan con pronóstico adecuado, evidenciado por pronósticos lineales en las gráficas y con errores cuadráticos medios (RMSE) elevados (>1.000).

```
> #Casos
> Casos_Covid_Prueba <- window(Casos_Covid_TS, start=c(2020,260), end= c(2021,5))
> accuracy(Casos_Covid_Meanf_Method,Casos_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1  Theil's U
Training set -8.861452e-12 108968.2 100966.3    -Inf      Inf  NaN  0.9930188      NA
Test set      1.696384e+05 171443.3 169638.4  52.26265  52.26265  NaN  0.8693707  49.80486
> accuracy(Casos_Covid_Naive_Method,Casos_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1  Theil's U
Training set  1119.653  1885.031  1120.795  4.494693  4.495236  NaN  0.9779350      NA
Test set      32945.567 41243.564 32945.567  9.695944  9.695944  NaN  0.8693707  11.36344
> accuracy(Casos_Covid_Snaive_Method,Casos_Covid_Prueba)
              ME      RMSE      MAE  MPE  MAPE  MASE      ACF1  Theil's U
Training set   NaN      NaN      NaN NaN  NaN  NaN      NA      NA
Test set      322935.6 323887.3 322935.6 100  100  NaN  0.8693707  94.73026
> accuracy(Casos_Covid_RWF_Method,Casos_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1  Theil's U
Training set  1119.653  1885.031  1120.795  4.494693  4.495236  NaN  0.9779350      NA
Test set      32945.567 41243.564 32945.567  9.695944  9.695944  NaN  0.8693707  11.36344
```

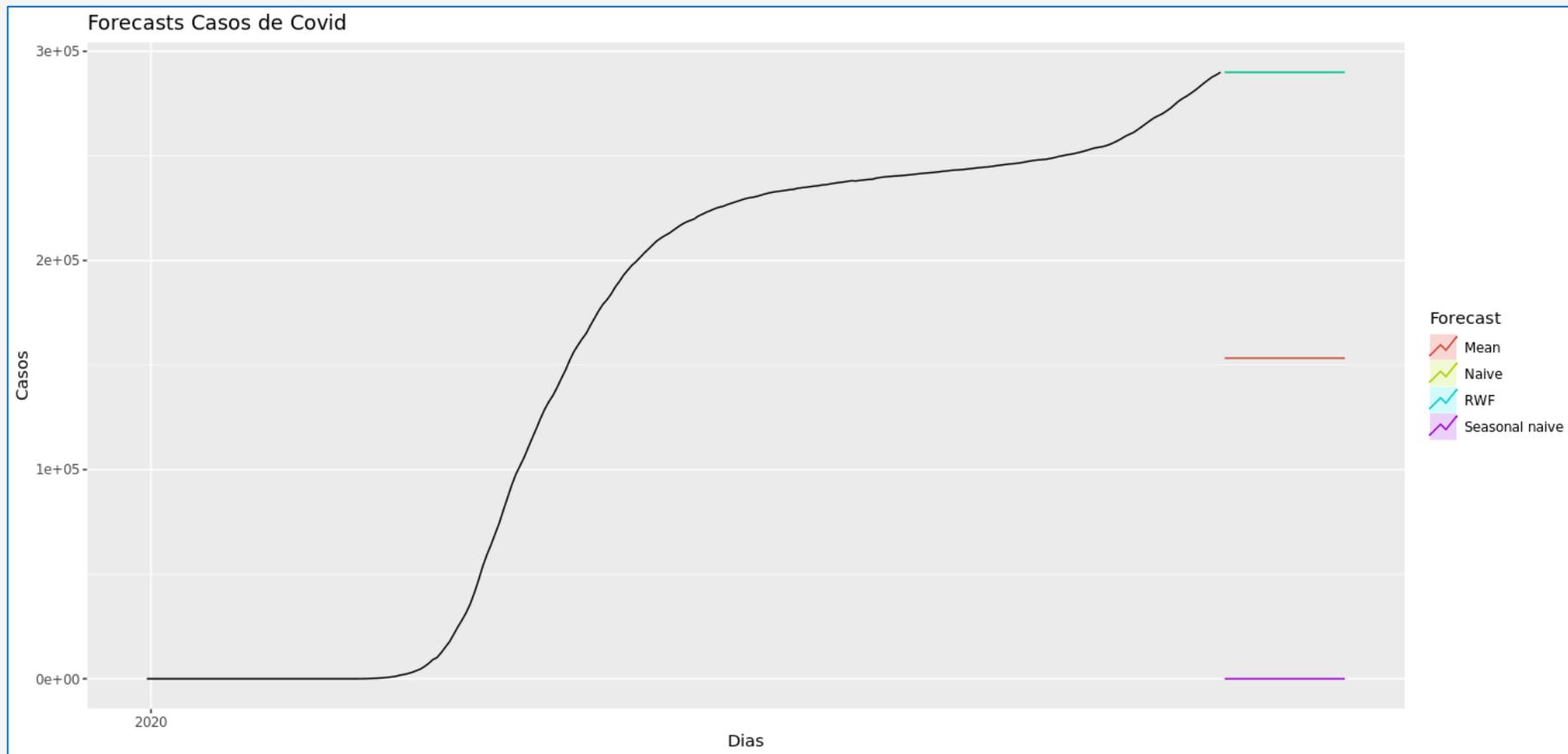
Casos COVID-19 en Italia. Error cuadrático medio

# Resultados



Casos COVID-19 en Italia. Métodos pronósticos Mean, Naive, RWF y SNaive

# Resultados

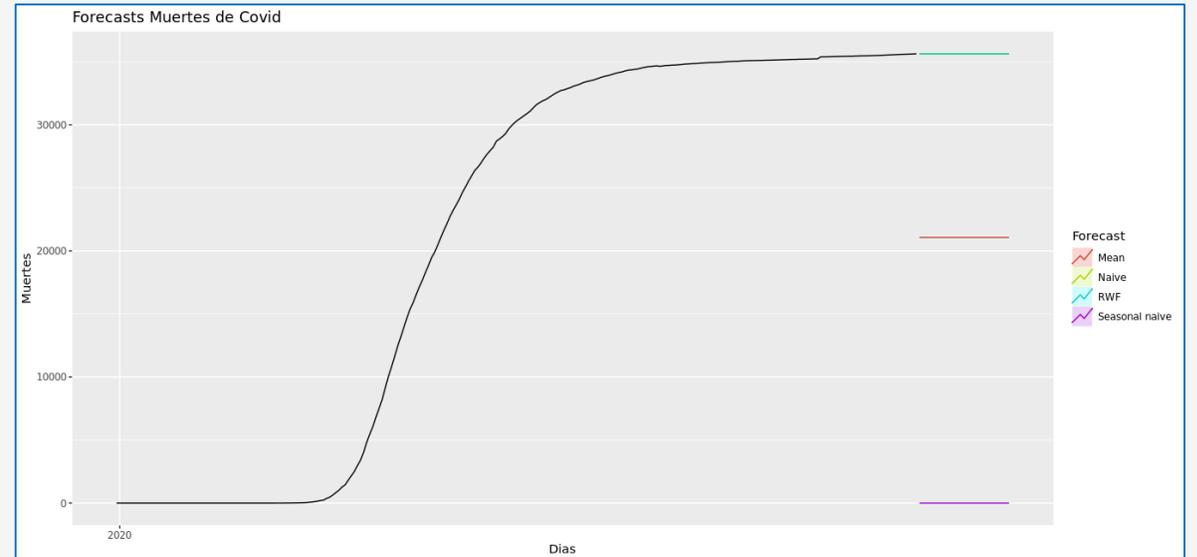
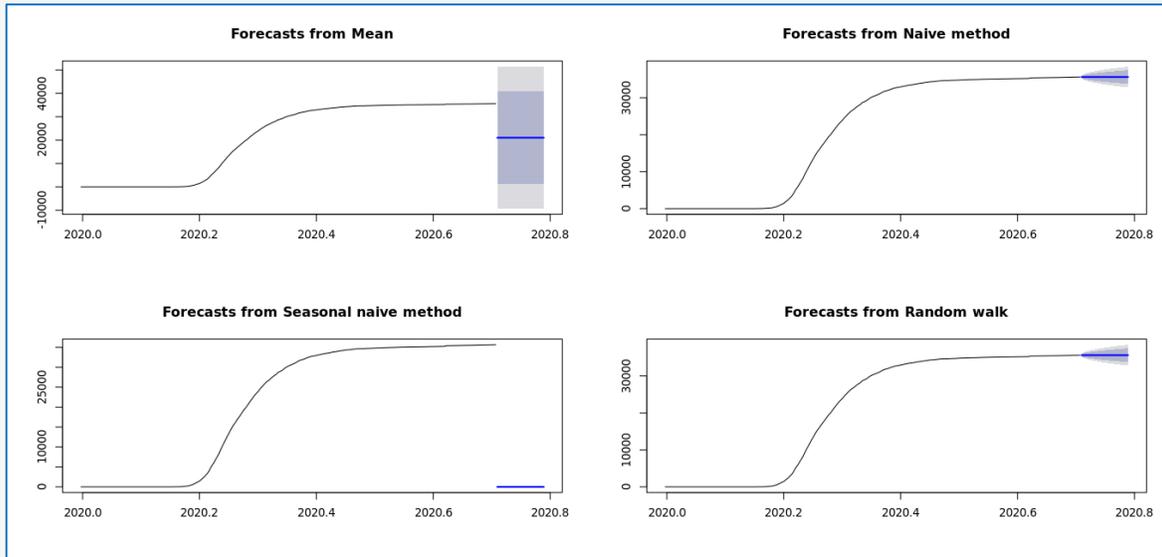


Casos COVID-19 en Italia. Métodos pronósticos Mean, Naive, RWF y SNaive

# Resultados

```

> #Muertes
> Muertes_Covid_Prueba <- window(Muertes_Covid_TS, start=c(2020,260), end= c(2021,5))
> accuracy(Muertes_Covid_Meanf_Method,Muertes_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1 Theil's U
Training set -6.629774e-13 15384.16 14413.82   -Inf      Inf  NaN  0.9945254      NA
Test set     1.486740e+04 14868.74 14867.40 41.37455 41.37455  NaN  0.8755116 524.6645
> accuracy(Muertes_Covid_Naive_Method,Muertes_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1 Theil's U
Training set 137.5792 262.1231 137.8185 4.8260719 4.8269323  NaN  0.9705402      NA
Test set    299.1000 359.9378 299.1000 0.8293306 0.8293306  NaN  0.8755116 12.84358
> accuracy(Muertes_Covid_Snaive_Method,Muertes_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1 Theil's U
Training set  NaN      NaN      NaN  NaN  NaN  NaN      NA      NA
Test set    35932.1 35932.66 35932.1 100 100  NaN  0.8755116 1267.545
> accuracy(Muertes_Covid_RWF_Method,Muertes_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1 Theil's U
Training set 137.5792 262.1231 137.8185 4.8260719 4.8269323  NaN  0.9705402      NA
Test set    299.1000 359.9378 299.1000 0.8293306 0.8293306  NaN  0.8755116 12.84358
    
```

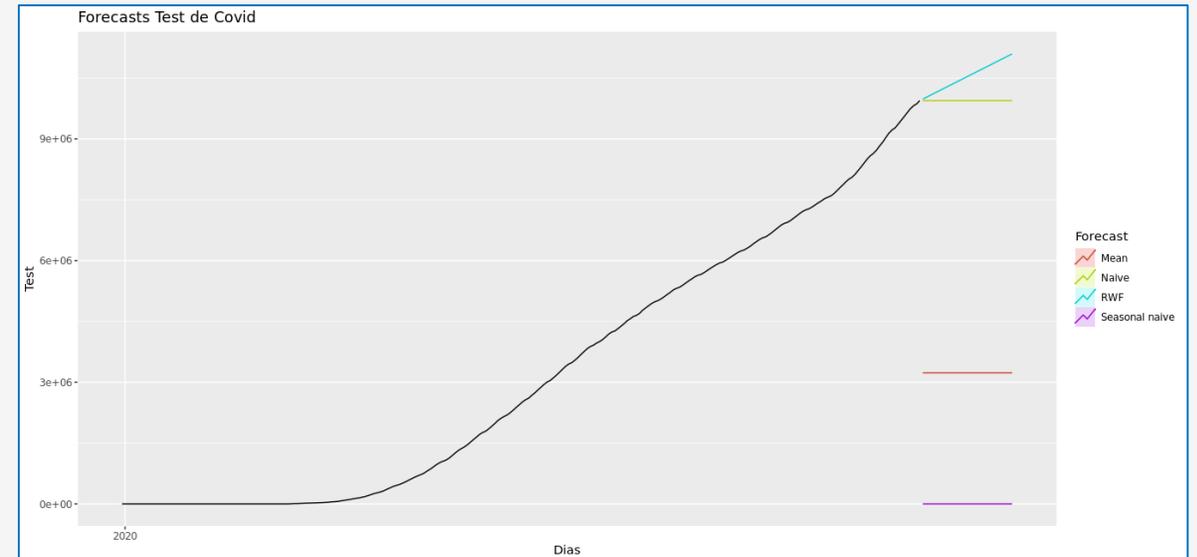
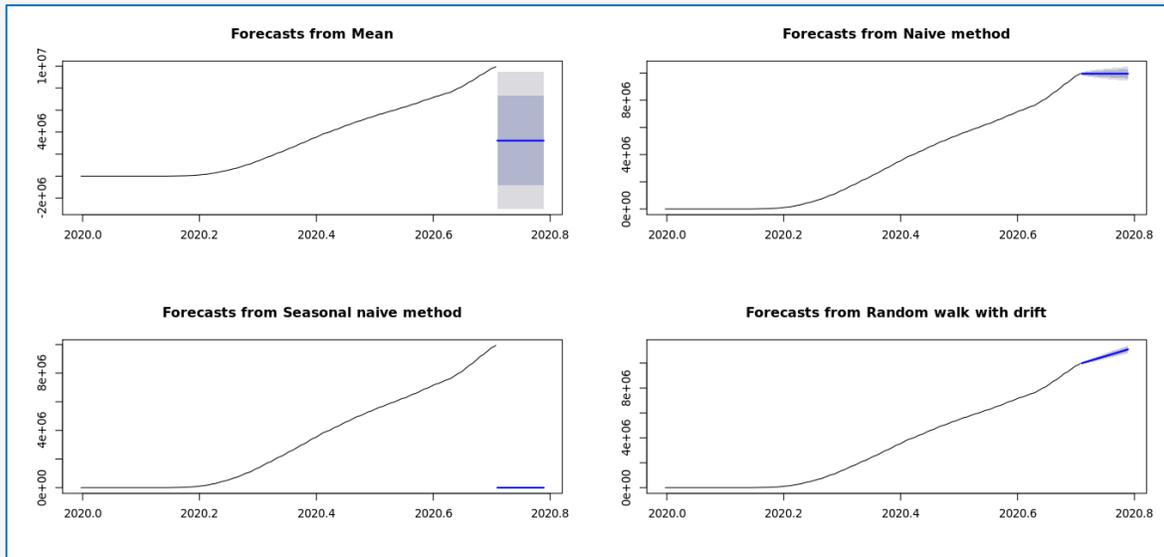


Muertes COVID-19 en Italia. Métodos pronósticos Mean, Naive, RWF y SNaive y error cuadrático medio

# Resultados

```

> #Test
> Test_Covid_Prueba <- window(Test_Covid_TS, start=c(2020,260), end= c(2021,5))
> accuracy(Test_Covid_Meanf_Method,Test_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE MASE      ACF1 Theil's U
Training set 1.929633e-10 3147004 807588      -Inf      Inf  NaN 0.9891057      NA
Test set     8.227767e+06 8275049 8227767 71.63937 71.63937  NaN 0.8941289  77.6415
> accuracy(Test_Covid_Naive_Method,Test_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE MASE      ACF1 Theil's U
Training set 38393.61 48023.01 38393.61 4.020951 4.020951  NaN 0.9137181      NA
Test set     1514251.20 1755070.72 1514251.20 12.699960 12.699960  NaN 0.8941289  15.79885
> accuracy(Test_Covid_Snaive_Method,Test_Covid_Prueba)
              ME      RMSE      MAE MPE MAPE MASE      ACF1 Theil's U
Training set      NaN      NaN      NaN NaN NaN NaN      NA      NA
Test set     11458195 11492195 11458195 100 100  NaN 0.8941289  108.1316
> accuracy(Test_Covid_RWF_Method,Test_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE MASE      ACF1 Theil's U
Training set -1.026310e-10 28846.84 24584.21      -Inf      Inf  NaN 0.9137181      NA
Test set     9.191503e+05 1072166.51 919150.30 7.699346 7.699346  NaN 0.8901087  9.641579
    
```

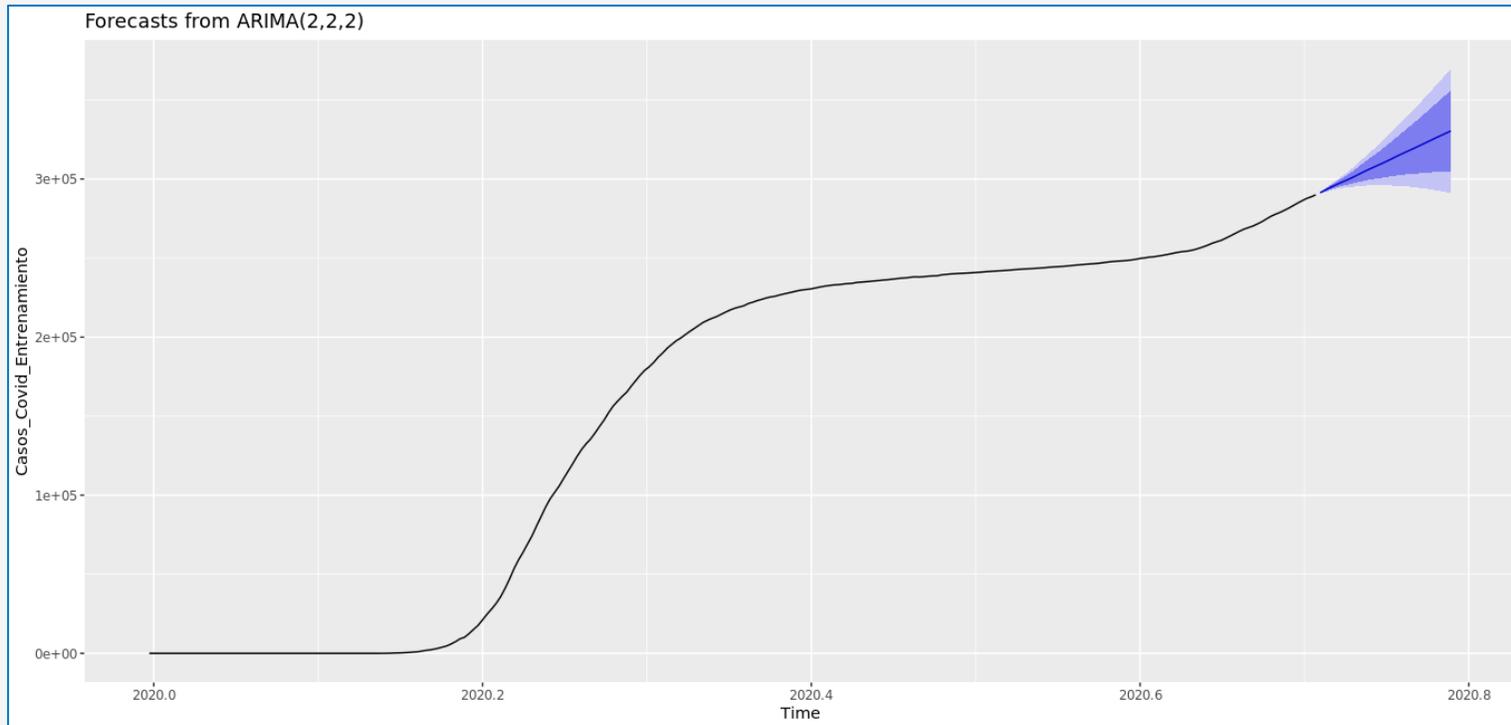


Test COVID-19 en Italia. Métodos pronósticos Mean, Naive, RWF y SNaive y error cuadrático medio

# Resultados

Enseguida se procedió a utilizar el **método Arima** en los datos de casos, muertes y test de COVID-19 en Italia. Los resultados mostraron pronósticos con un horizonte de pronóstico de 30 días más exactos, evidenciados con errores cuadráticos medios (RMSE) muy inferiores a los métodos anteriores (<280).

```
> accuracy(Casos_Covid_Arima_Method,Casos_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set  6.994205  275.6991  162.6715  1.180259  2.947222  NaN  -0.02169692      NA
Test set     11955.816490 18399.5200 11955.8165  3.418847  3.418847  NaN   0.83968728  4.931572
```



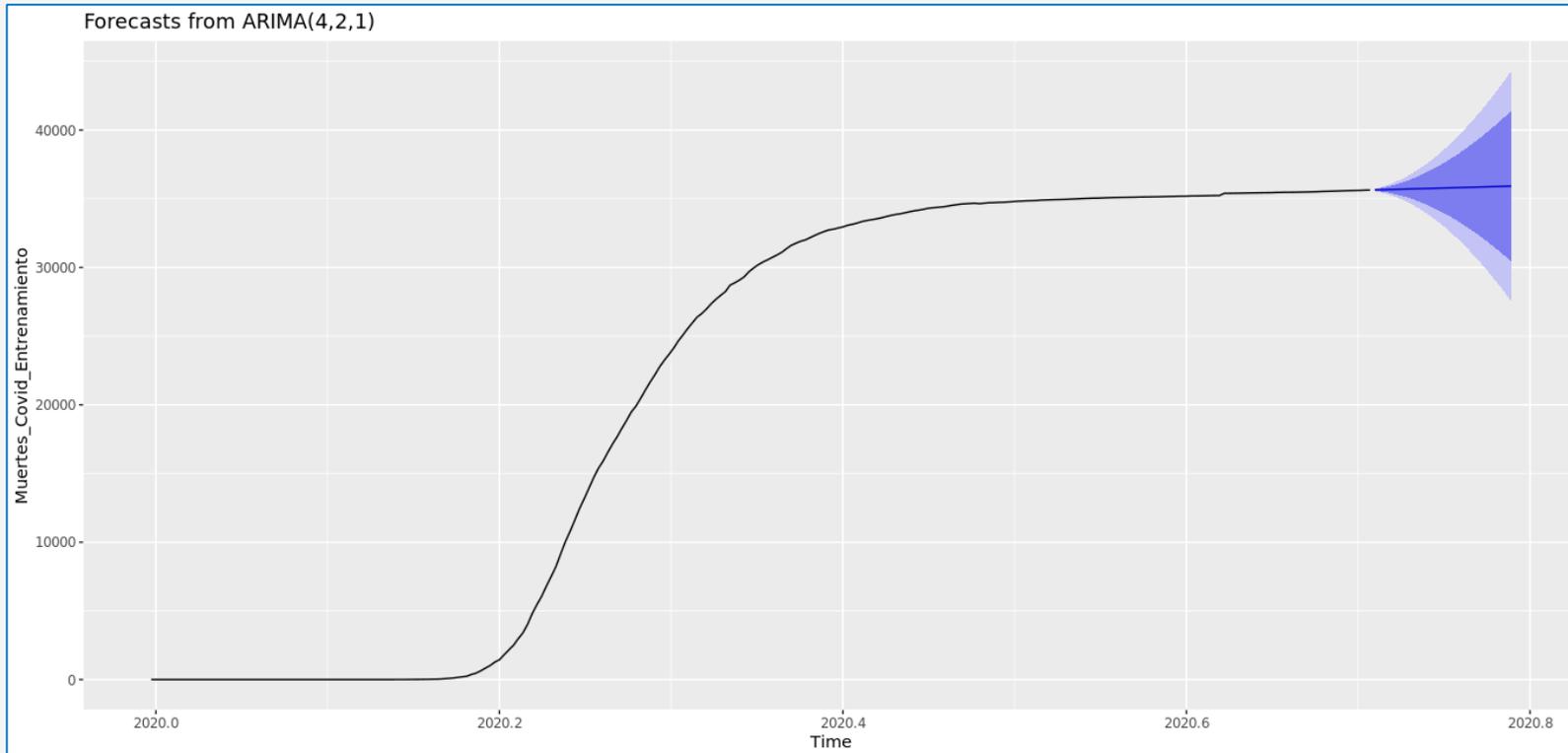
Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020.7096	291346.8	290989.4	291704.3	290800.1	291893.6
2020.7123	292808.4	292066.0	293550.9	291672.9	293943.9
2020.7151	294287.2	293145.0	295429.5	292540.3	296034.2
2020.7178	295694.4	294144.5	297244.3	293324.1	298064.7
2020.7205	296996.5	295007.1	298985.9	293954.0	300039.0
2020.7233	298230.9	295735.8	300725.9	294415.0	302046.7
2020.7260	299474.0	296385.5	302562.6	294750.5	304197.5
2020.7288	300788.4	297023.9	304553.0	295031.0	306545.8
2020.7315	302184.3	297689.2	306679.4	295309.7	309059.0
2020.7342	303618.7	298369.9	308867.4	295591.4	311646.0
2020.7370	305028.7	299019.8	311037.5	295839.0	314218.3
2020.7397	306373.9	299595.4	313152.4	296007.1	316740.7
2020.7425	307659.7	300085.4	315234.1	296075.8	319243.7
2020.7452	308928.9	300515.4	317342.5	296061.5	321796.3
2020.7479	310230.1	300927.0	319533.3	296002.1	324458.1
2020.7507	311586.2	301349.6	321822.9	295930.6	327241.9
2020.7534	312982.6	301783.3	324181.9	295854.8	330110.4
2020.7562	314380.4	302202.4	326558.4	295755.8	333005.0
2020.7589	315744.3	302575.5	328913.0	295604.4	335884.1
2020.7616	317064.3	302887.3	331241.2	295382.5	338746.0
2020.7644	318359.6	303147.3	333571.9	295094.4	341624.8
2020.7671	319663.1	303381.3	335945.0	294762.2	344564.1
2020.7699	320998.9	303613.4	338384.4	294410.0	347587.8
2020.7726	322367.7	303850.6	340884.8	294048.2	350687.2
2020.7753	323749.1	304080.6	343417.6	293668.8	353829.4
2020.7781	325117.0	304282.3	345951.7	293253.1	356980.9
2020.7808	326456.6	304440.3	348473.0	292785.6	360127.7
2020.7836	327773.0	304554.8	350991.1	292263.9	363282.0
2020.7863	329085.3	304640.2	353530.4	291699.8	366470.8
2020.7890	330413.4	304714.3	356112.5	291110.0	369716.7

Casos COVID-19 en Italia. Método Arima y error cuadrático medio.

# Resultados

```
> accuracy(Muertes_Covid_Arima Method, Muertes_Covid_Prueba)
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.0284767	46.75533	23.62429	1.4575929	1.8489719	NaN	0.001280063	NA
Test set	151.3053454	192.49214	151.30535	0.4192698	0.4192698	NaN	0.853033562	6.863596



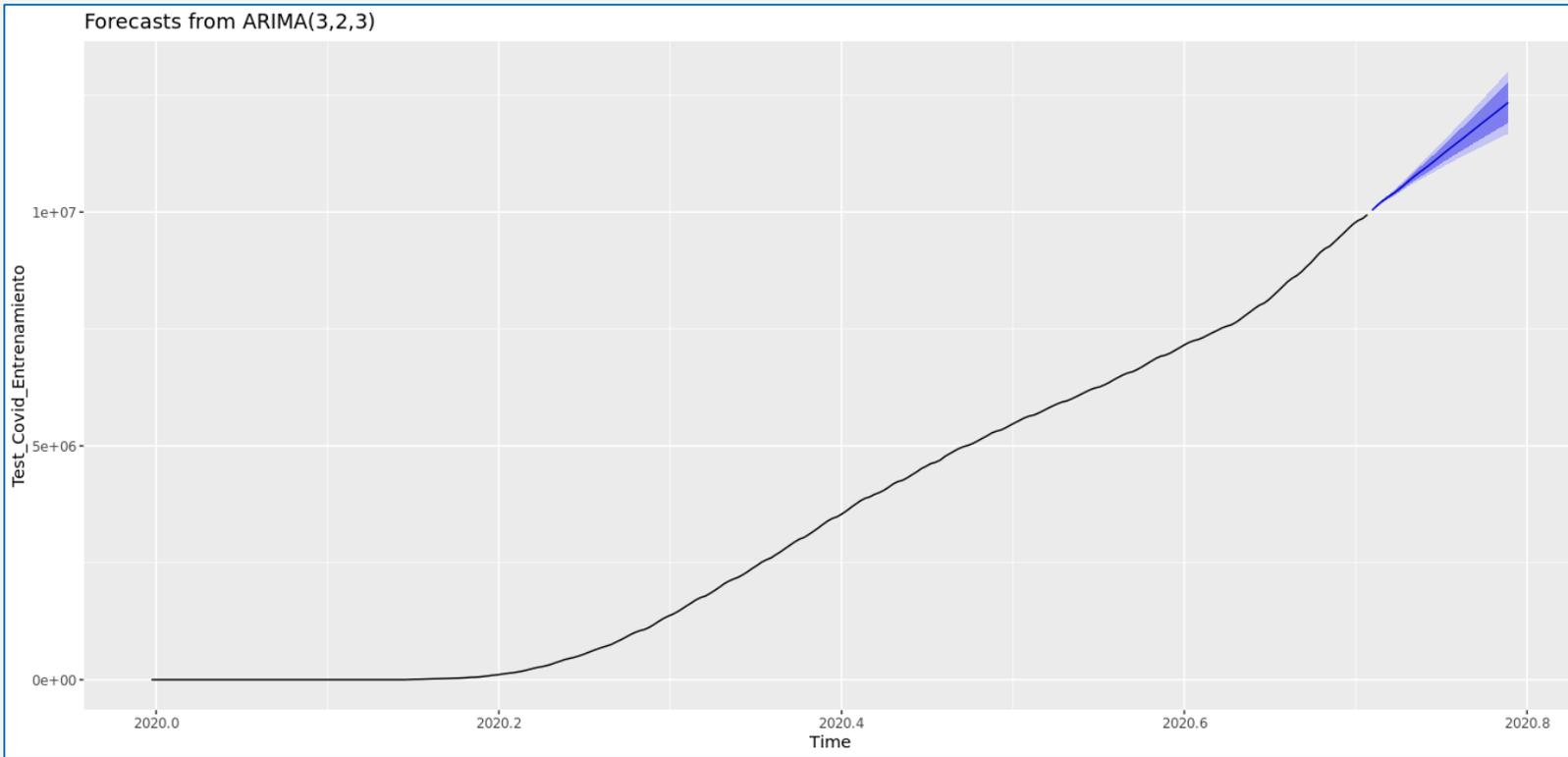
Forecasts:

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020.7096	35641.43	35580.69	35702.17	35548.53	35734.33
2020.7123	35651.18	35538.90	35763.46	35479.46	35822.89
2020.7151	35661.39	35494.40	35828.38	35406.00	35916.78
2020.7178	35670.80	35436.76	35904.84	35312.87	36028.73
2020.7205	35680.20	35364.26	35996.13	35197.01	36163.38
2020.7233	35689.88	35282.04	36097.71	35066.15	36313.61
2020.7260	35699.55	35190.18	36208.92	34920.53	36478.56
2020.7288	35709.08	35087.23	36330.93	34758.04	36660.12
2020.7315	35718.64	34973.46	36463.81	34578.99	36858.29
2020.7342	35728.24	34849.63	36606.85	34384.53	37071.96
2020.7370	35737.83	34715.76	36759.89	34174.72	37300.94
2020.7397	35747.39	34571.79	36922.99	33949.46	37545.31
2020.7425	35756.96	34417.93	37095.99	33709.09	37804.82
2020.7452	35766.53	34254.42	37278.64	33453.96	38079.11
2020.7479	35776.10	34081.38	37470.82	33184.25	38367.95
2020.7507	35785.66	33898.93	37672.40	32900.16	38671.17
2020.7534	35795.23	33707.26	37883.20	32601.96	38988.50
2020.7562	35804.79	33506.52	38103.06	32289.89	39319.70
2020.7589	35814.35	33296.86	38331.84	31964.19	39664.52
2020.7616	35823.91	33078.44	38569.39	31625.07	40022.75
2020.7644	35833.47	32851.40	38815.54	31272.79	40394.15
2020.7671	35843.03	32615.91	39070.15	30907.57	40778.49
2020.7699	35852.59	32372.09	39333.08	30529.63	41175.54
2020.7726	35862.14	32120.10	39604.18	30139.19	41585.10
2020.7753	35871.70	31860.08	39883.31	29736.46	42006.93
2020.7781	35881.25	31592.17	40170.33	29321.67	42440.83
2020.7808	35890.80	31316.50	40465.11	28895.01	42886.60
2020.7836	35900.36	31033.20	40767.51	28456.69	43344.02
2020.7863	35909.91	30742.41	41077.40	28006.90	43812.91
2020.7890	35919.46	30444.24	41394.67	27545.84	44293.08

Muertes COVID-19 en Italia. Método Arima y error cuadrático medio.

# Resultados

```
> accuracy(Test_Covid_Arima_Method, Test_Covid_Prueba)
              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1  Theil's U
Training set  744.0086  8312.056  5616.262  0.9635032  1.418403  NaN -0.1194013  NA
Test set     263887.2281 333135.479 263887.228 2.1843899  2.184390  NaN  0.8698945  2.953262
```

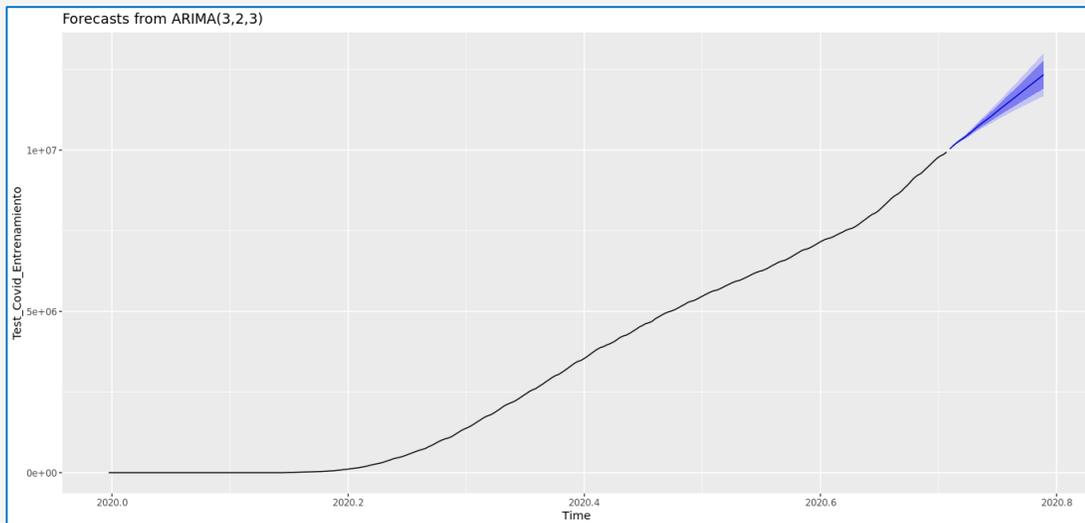
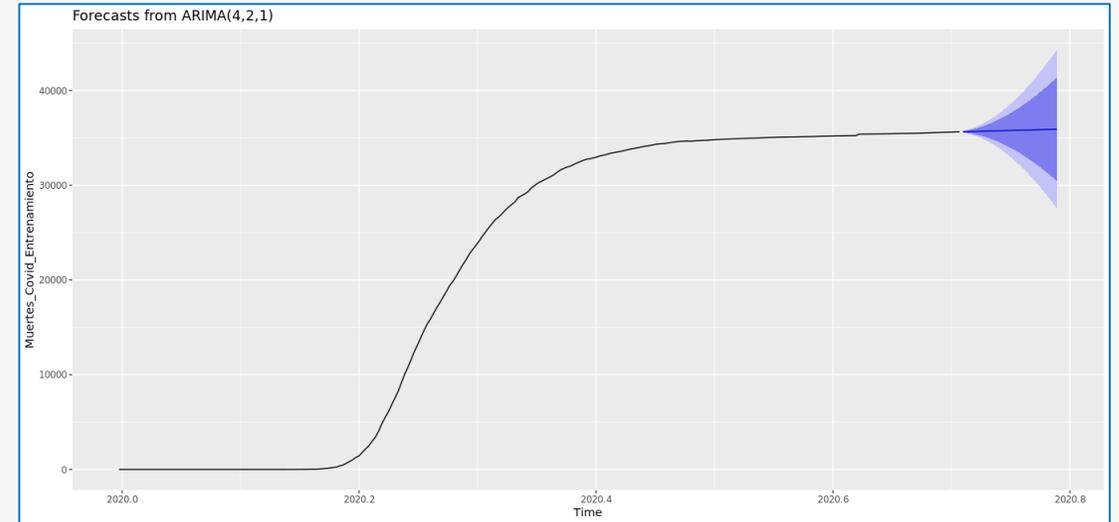
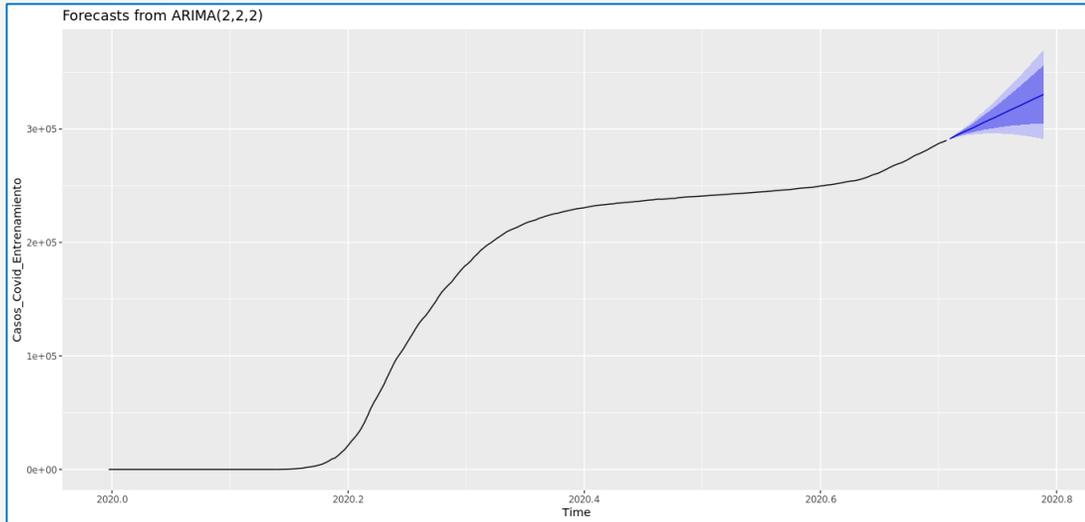


Forecasts:

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020.7096	10038723	10027903	10049543	10022175	10055271
2020.7123	10133652	10112721	10154584	10101640	10165664
2020.7151	10219098	10191202	10246994	10176434	10261761
2020.7178	10292434	10259293	10325574	10241750	10343118
2020.7205	10360592	10322622	10398562	10302522	10418663
2020.7233	10431598	10386885	10476310	10363216	10499979
2020.7260	10510575	10456199	10564951	10427414	10593736
2020.7288	10595768	10528965	10662570	10493601	10697934
2020.7315	10682149	10601996	10762302	10559565	10804733
2020.7342	10764619	10671371	10857867	10622008	10907230
2020.7370	10842027	10736366	10947688	10680432	11003622
2020.7397	10916559	10798553	11034565	10736084	11097034
2020.7425	10992020	10861031	11123008	10791690	11192350
2020.7452	11070638	10925470	11215807	10848622	11292654
2020.7479	11152220	10991740	11312700	10906787	11397652
2020.7507	11234559	11058034	11411084	10964588	11504531
2020.7534	11315505	11122698	11508312	11020633	11610378
2020.7562	11394213	11185067	11603359	11074352	11714074
2020.7589	11471544	11245906	11697181	11126460	11816627
2020.7616	11549037	11306479	11791595	11178076	11919998
2020.7644	11627834	11367734	11887934	11230045	12025623
2020.7671	11707947	11429651	11986243	11282329	12133565
2020.7699	11788543	11491539	12085548	11334315	12242772
2020.7726	11868631	11552577	12184685	11385268	12351994
2020.7753	11947778	11612432	12283124	11434911	12460645
2020.7781	12026251	11671351	12381152	11483478	12569025
2020.7808	12104714	11729909	12479518	11531499	12677928
2020.7836	12183686	11788545	12578827	11579370	12788002
2020.7863	12263250	11847325	12679175	11627148	12899352
2020.7890	12343075	11905963	12780188	11674570	13011581

Test COVID-19 en Italia. Método Arima y error cuadrático medio

# Resultados



Tomando en consideración los resultados anteriores y la **mejor predicción de los datos de COVID-19 en Italia por el método Arima**, se seleccionó dicho método para ser utilizado como **comparativo** estándar para medir la efectividad predictiva de la **red neuronal**.

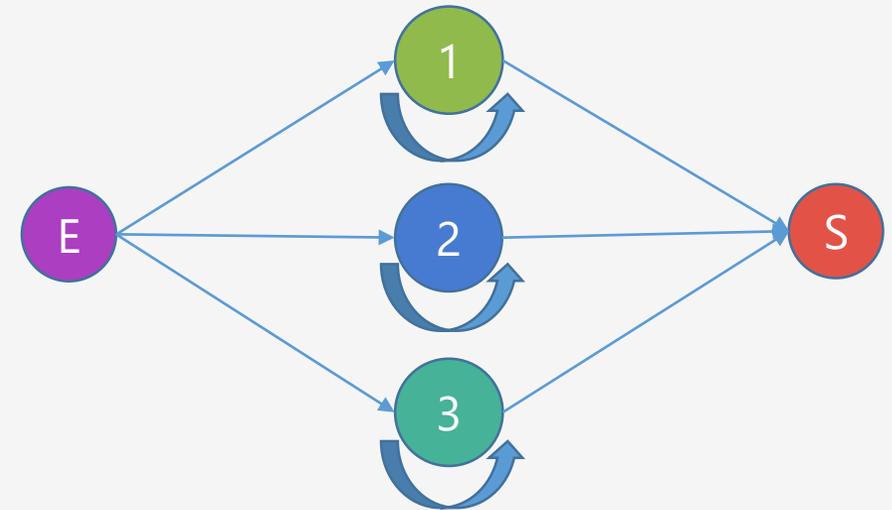
# Resultados

Finalmente se ejecutó la **red neuronal "Memoria a Largo – Corto Plazo"** (del inglés, Long – Short Term Memory, LSTM) con los datos de COVID-19 en Italia y **se procedió a su entrenamiento**. El método usado incluye tres neuronas, una capa y una salida.

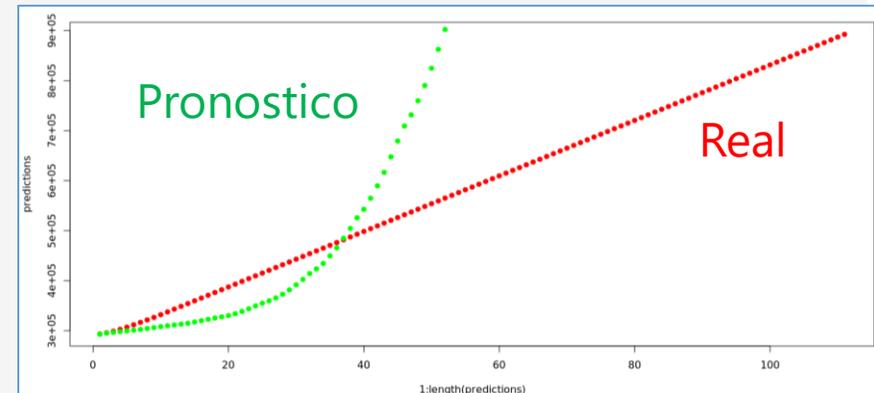
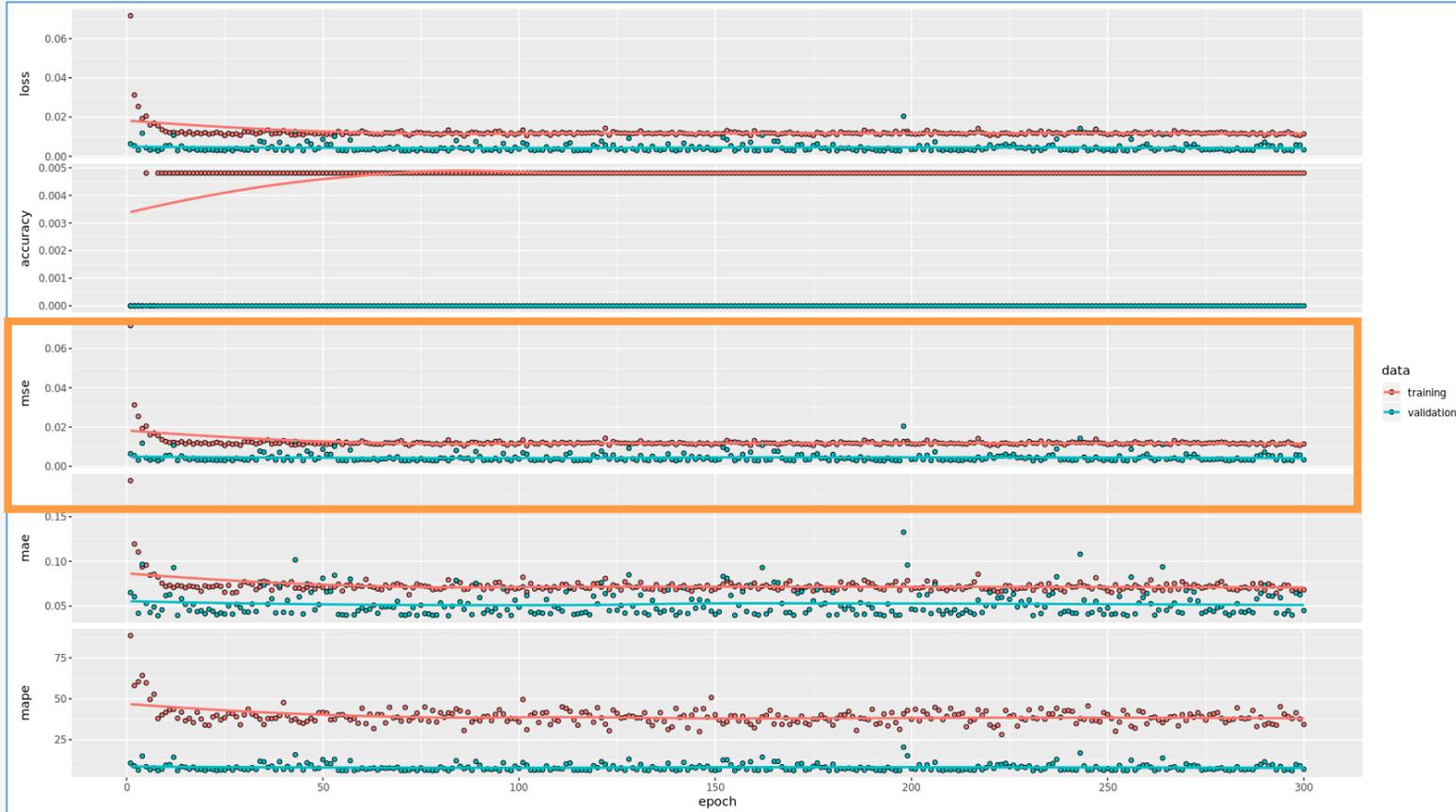
```
> summary(model)
Model: "sequential_25"
```

Layer (type)	Output Shape	Param #
lstm_27 (LSTM)	(1, 3)	60
dense_29 (Dense)	(1, 1)	4

Total params: 64  
Trainable params: 64  
Non-trainable params: 0

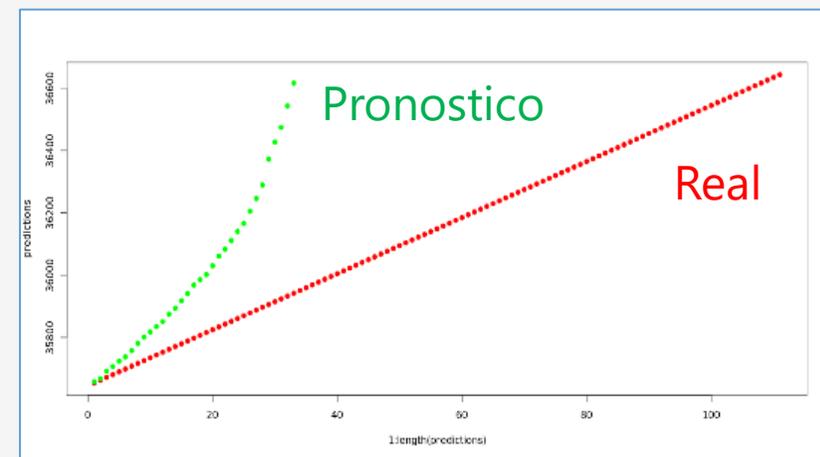
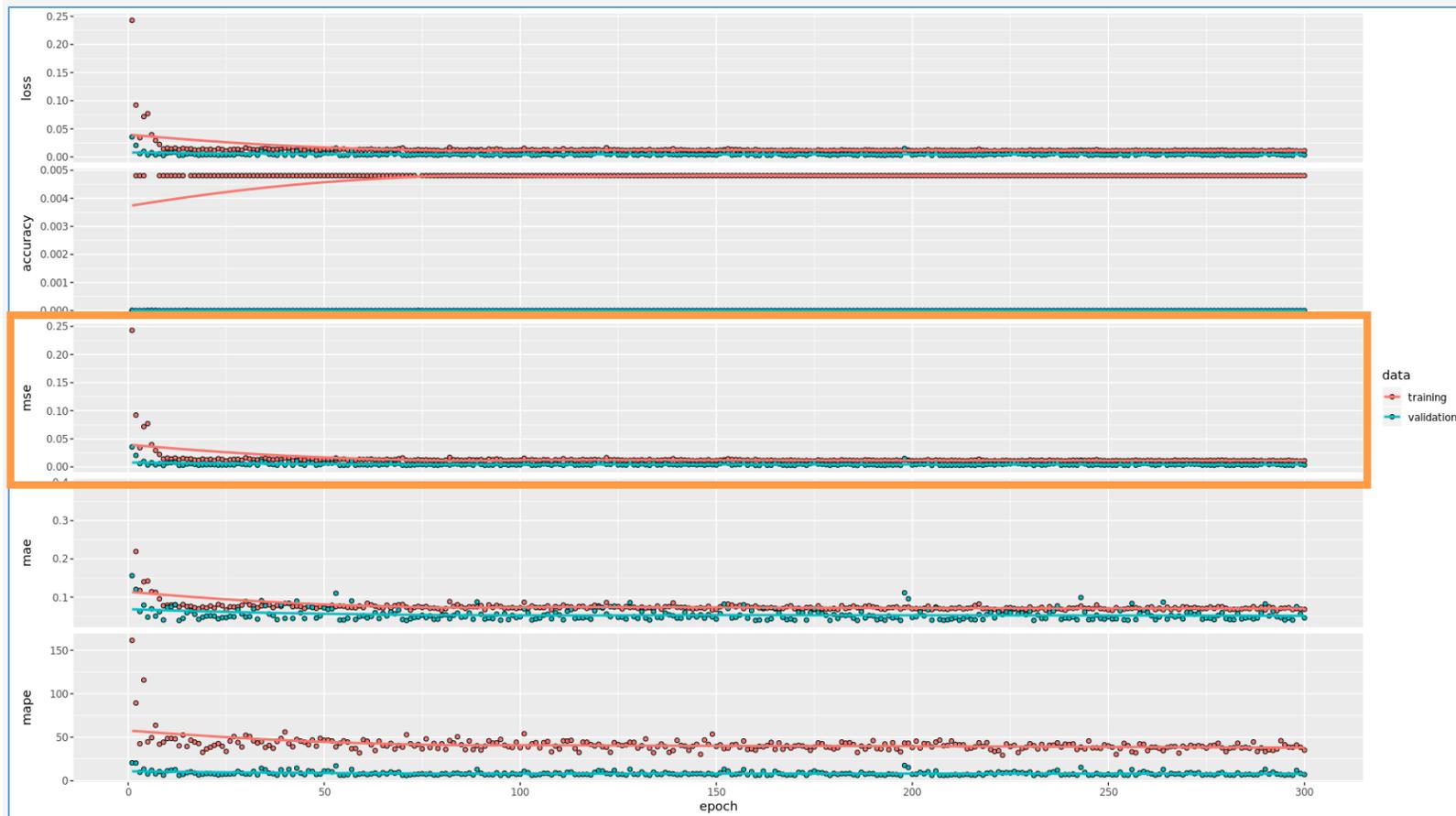


# Resultados



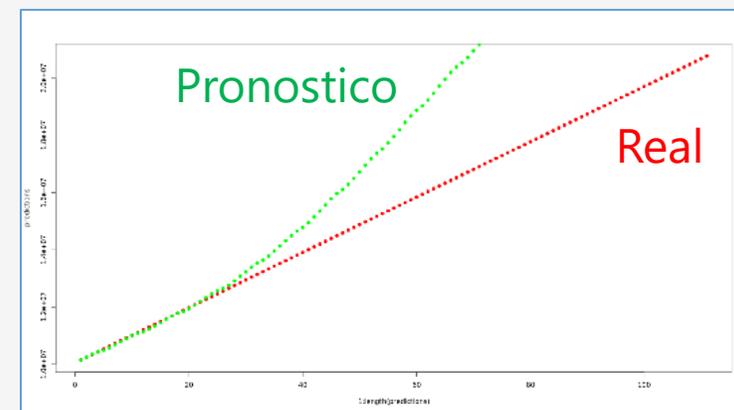
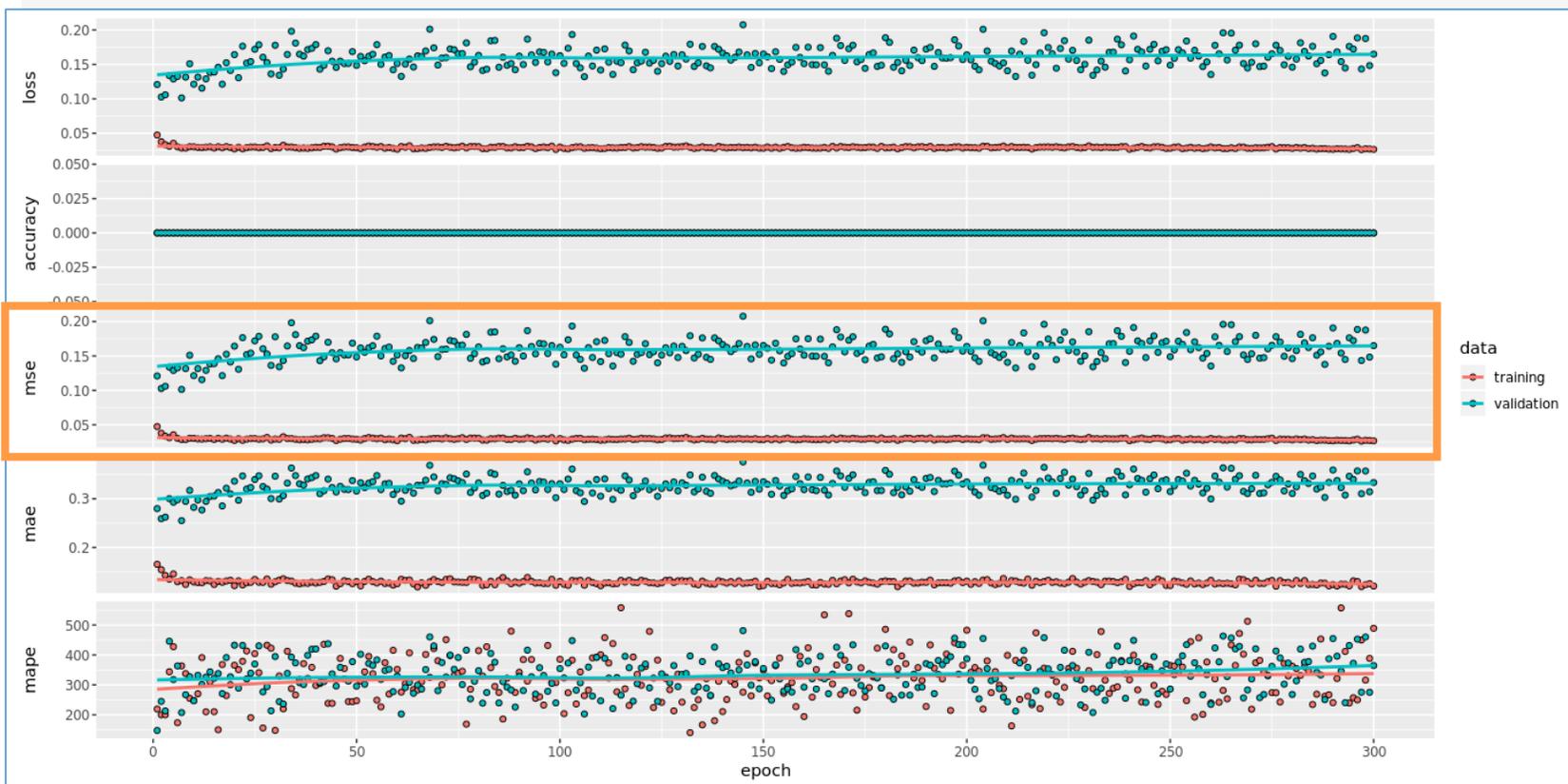
Predicción e Indicadores de calidad. LSTM en **casos** COVID-19 en Italia

# Resultados



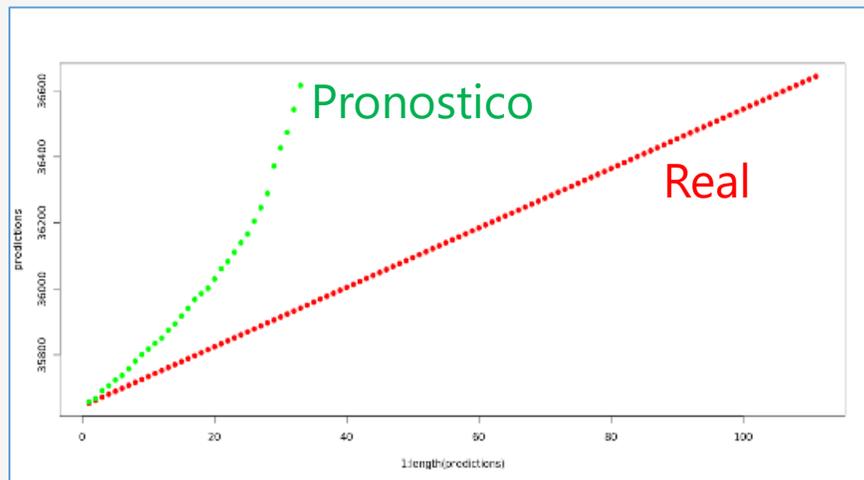
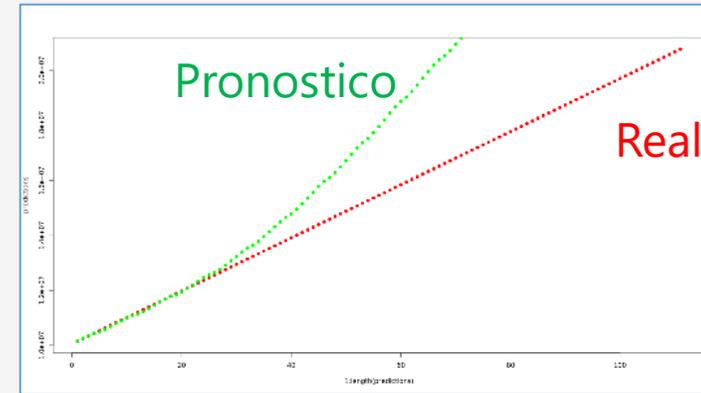
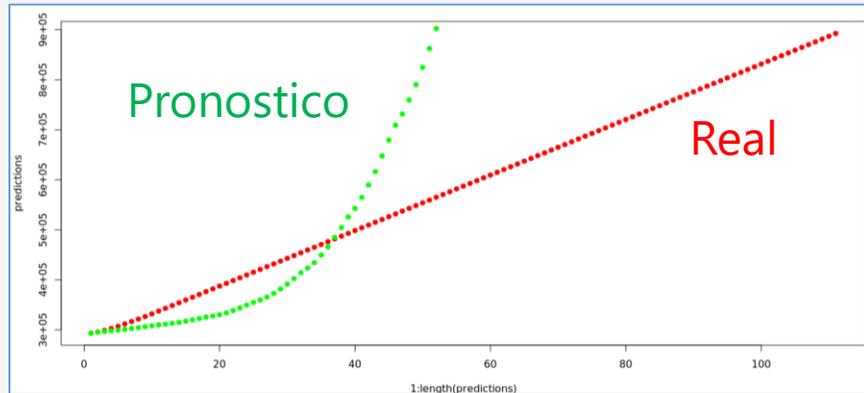
Predicción e Indicadores de calidad. LSTM en **muertes** COVID-19 en Italia

# Resultados



Predicción e Indicadores de calidad. LSTM en **test** COVID-19 en Italia

# Resultados



El **pronóstico de la red neuronal** “Memoria a Largo – Corto Plazo” (del inglés, Long – Short Term Memory, **LSTM**) para COVID-19 en Italia, **fue sumamente deficiente** con el modelo planteado.

No se realizaron más análisis con los datos de movilidad en Barcelona, Madrid y Málaga con el modelo planteado.

# Índice

---

1. Introducción
2. Objetivos
3. Metodología
4. Resultados
5. **Discusión**
6. Conclusiones
7. Bibliografía



## Discusión

---

- **Los datos de casos y de muertes cuentan con aumentos** a principios de **marzo** y **noviembre** de 2020, esto se correlaciona con la **expansión del virus inicialmente y con el periodo vacacional** de final de año.
- El método **Mean, Naive, Seasonal Naive y Random Walk Forecasts no mostraron una predicción adecuada** dado que obtuvieron un error cuadrático medio extremadamente elevado.
- El método **Arima ha demostrado ser un método pronóstico adecuado** para predecir los datos COVID-19 en Italia al estar por encima del 95% y con errores cuadráticos medios aceptables.
- El método **LSTM** fue probado con múltiples configuraciones de red neuronal, pero a pesar de aumentar o disminuir el número de neuronas y capas, los **resultados de los pronósticos no fueron satisfactorios**, incluso los datos de evaluación de los pronósticos mostraban una deficiente precisión.



# Índice

---

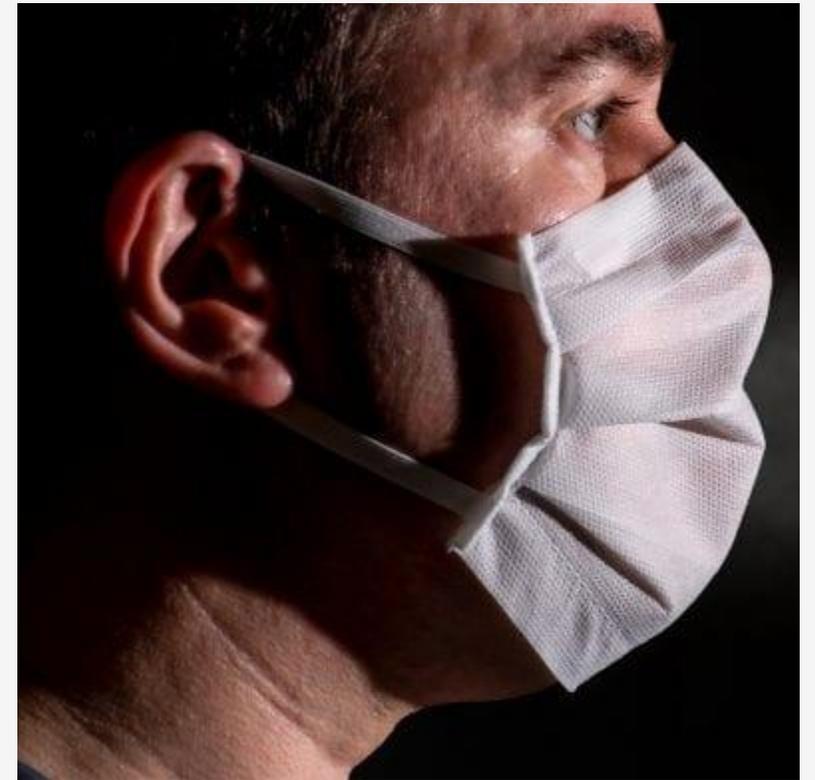
1. Introducción
2. Objetivos
3. Metodología
4. Resultados
5. Discusión
6. **Conclusiones**
7. Bibliografía



# Conclusiones

---

- Es evidente la **necesidad de estudios que permitan un análisis y predicción de la enfermedad** por medio de métodos probados y con alta precisión.
- **Arima ha demostrado ser efectivo y confiable** para predecir series temporales con datos de COVID-19 en Italia.
- **El método LSTM es posible que mejore su precisión con un mayor número de datos y variables** como podría ser el clima, la densidad de población, la movilidad, los recursos de salud de la región, datos demográficos y clínicos de los pacientes. Sería interesante en investigaciones futuras corroborar si el método aumenta su asertividad pronostica.
- **Este tipo de estudios** cuentan con el **potencial de apoyar a los tomadores de decisión** de los ministerios de salud a **planificar políticas de contención** para el virus y mecanismos de **preparación efectiva** para esta u otras pandemias.





# Aplicación de Deep Neural Network para la predicción de incidencias en COVID-19

Trabajo Final - Máster Universitario en Salud Digital

**Samuel Paul Gallegos Serrano**

Twitter: samuel\_gallegos

LinkedIn: <https://es.linkedin.com/in/samuel-paul-gallegos-serrano-918b7a35>

GitHub: <https://github.com/samuelgallegos/Forecasting-Methods-COVID-19-ARIMA-LSTM>

