Modelos de Predicción de Series Temporales: Gripe y COVID-19

Autor · · · Javier de Dios González

Tutor · · · Oriol Ramos Terrades

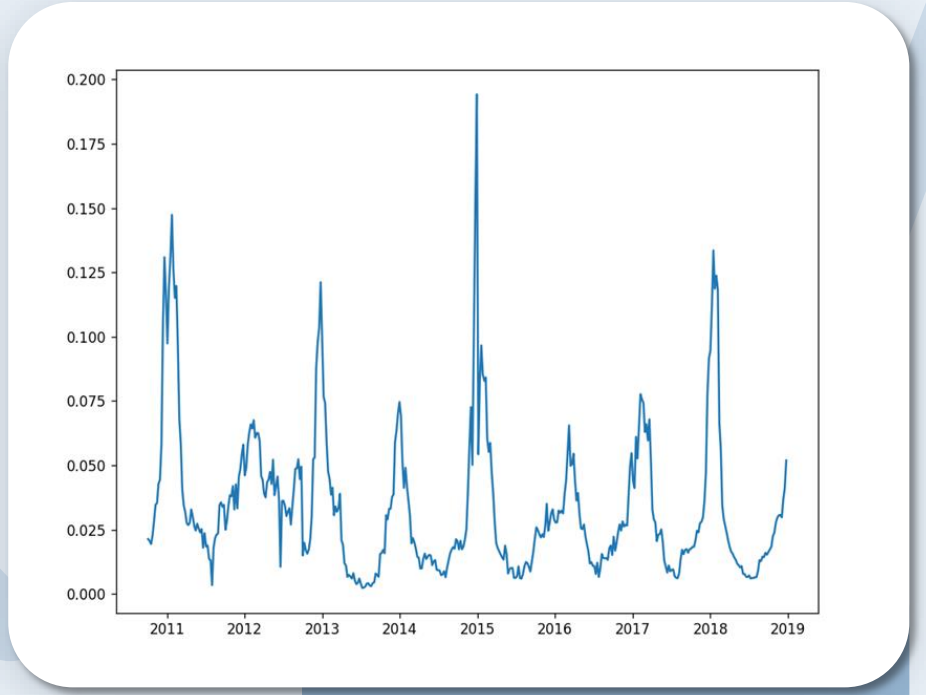
2021-22

Universitat Autònoma de Barcelona

ÍNDICE

Introducción

1
2
3
4
5
6
7

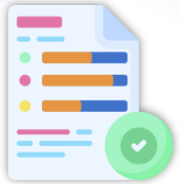


Serie temporal

Gripe y Covid-19

ARIMA

SARIMA



Resultados

Comparación

Estado del arte

1
2
3
4
5
6
7

Michael J Paul, Mark Dredze, David Broniatowski,
"Twitter improves influenza forecasting"

David J. McIver, John S. Brownstein,
"Wikipedia Usage Estimates Prevalence of
Influenza-Like Illness in the United States in Near Real-Time"

Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel,
Lynnette Brammer, Mark S Smolinski, Larry Brilliant,
"Detecting influenza epidemics using search engine query data"

- ARIMA
- LTSM
- Seq2Seq
- Transformer

Datos

1
2
3
4
5
6
7

“ILI”
Influenza-like illness

Gripe

Covid-19

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

- Fuente
  - WHO/NREVSS
  - ILINET
- Área de supervisión
  - Nacional
  - HHS Regions
  - Censo
  - Estatal
- Temporada
  - 1997
  - 2022

Datos

1
2
3
4
5
6
7

Gripe

Covid-19

ILINetSTATES.csv

ILINET

Estatal

2010 — 2018

Columna
REGION TYPE
REGION
YEAR
WEEK
%WEIGHTED ILI
%UNWEIGHTED ILI
AGE 0-4
AGE 25-49
AGE 25-64
AGE 5-24
AGE 50-64
AGE 65
ILITOTAL
NUM OF PROVIDERS
TOTAL PATIENTS

Filas: 23012

Datos

1
2
3
4
5
6
7

Gripe

Covid-19

Columna

REGION

Filas: 22529

+	YEAR	Ej.	2010
	WEEK	+	40
	DATE		2010-10-04

÷	ILITOTAL	
	TOTAL PATIENTS	+ ε
	ILIRATIO	

Datos

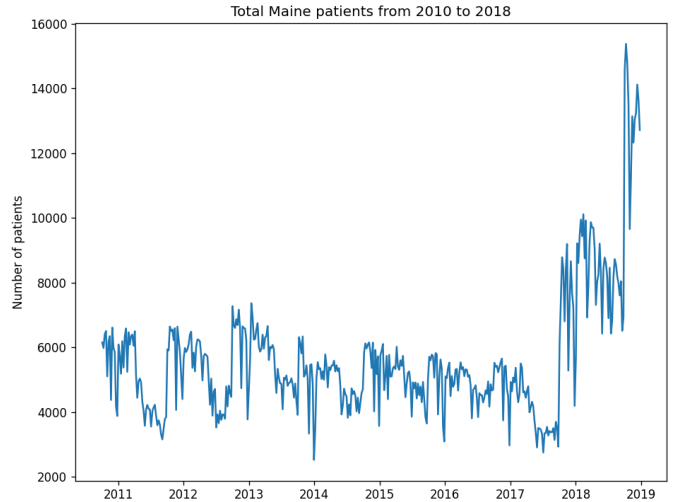
1
2
3
4
5
6
7

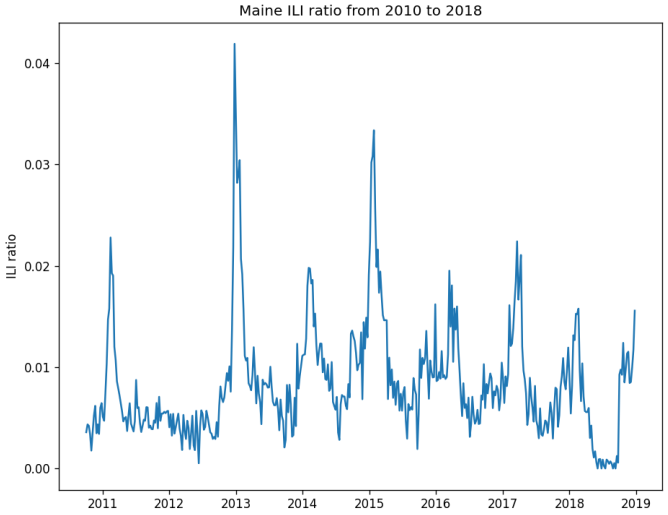
Gripe

Covid-19

Columna
REGION
YEAR
WEEK
ILITOTAL
TOTAL PATIENTS
DATE
ILIRATIO

Filas: 22529





Datos

1
2
3
4
5
6
7

Gripe

Covid-19

Columna
REGION
YEAR
WEEK
ILITOTAL
TOTAL PATIENTS
DATE
ILIRATIO

Filas: 22529

DATE	ILIRATIO
2010-10-04	0.003577
2010-10-11	0.004349
2010-10-18	0.004217
2010-10-25	0.003384
2010-11-01	0.001765
…	
2018-11-26	0.008434
2018-12-03	0.008536
2018-12-10	0.009985
2018-12-17	0.011723
2018-12-24	0.015572

dataset

serie temporal

Datos

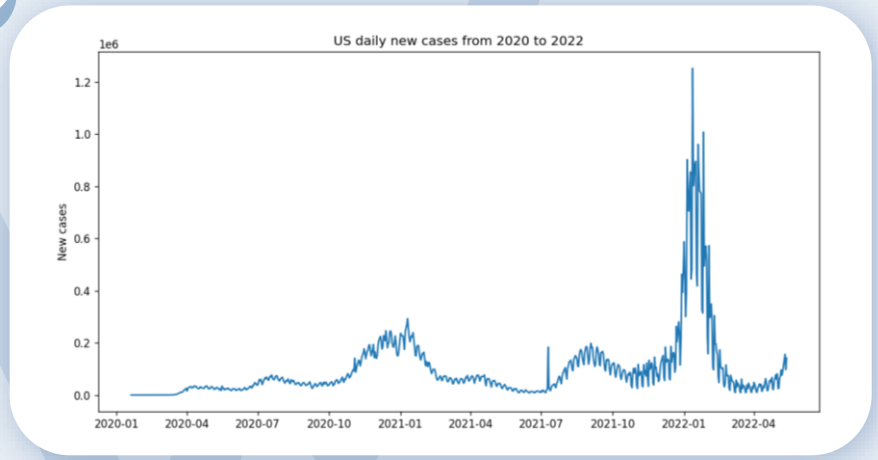
1
2
3
4
5
6
7

Gripe

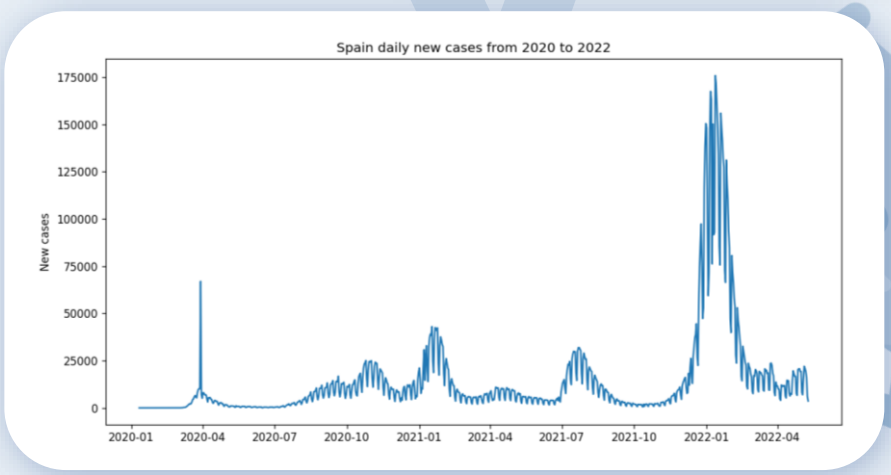
Covid-19

Columna
Date_reported
Country_code
Country
WHO_region
New_cases
Cumulative_cases
New_deaths
Cumulative_deaths

Filas: 21550



World Health Organization



Datos

1
2
3
4
5
6
7

Gripe

Covid-19

Columna

Date_reported

Country

New_cases

Filas: 21550

Date_reported	New_cases
2020-01-11	1
2020-01-12	1
2020-01-13	1
2020-01-14	1
2020-01-15	1
...	
2022-05-06	20828
2022-05-07	19355
2022-05-08	15993
2022-05-09	6054
2022-05-10	3638

dataset

serie temporal

Modelos

$$\underbrace{AR}_{p}\ \underbrace{I}_{d}\ \underbrace{MA}_{q}$$

1
2
3
4 ARIMA
SARIMA
5
6
7

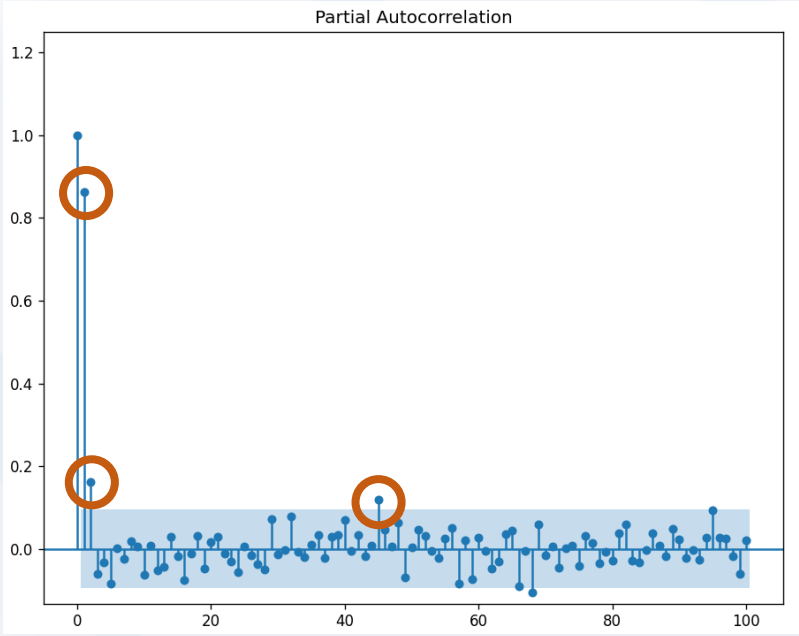
Modelos

1
2
3
4
5
6
7

ARIMA

SARIMA

$$\underbrace{AR}_{p}\ \underbrace{I}_{d}\ \underbrace{MA}_{q}$$



$p = 3$

$d =$

$q =$

Modelos

1
2
3
4 ARIMA
SARIMA
5
6
7

ARIMA

p d q

p = 3

d = 0

q =

ADF TEST

ADF Statistic: -4.570520099360388
p-value: 0.00014633251823777013
Critical Values:
1%: -3.445721386098794
5%: -2.868316661451884
10%: -2.5703797268320376

Modelos

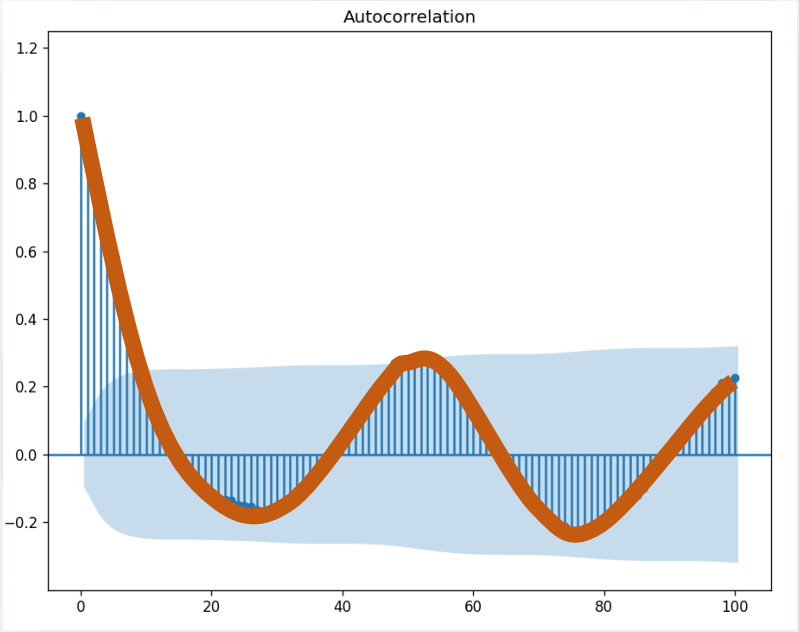
1
2
3
4
5
6
7

ARIMA

SARIMA

AR I MA

p d **q**



p = 3

d = 0

q = ?

Modelos

1
2
3
4
5
6
7

ARIMA

SARIMA

$$S\ ARIMA\ p\ d\ q\ x\ P\ D\ Q$$

Modelos

1
2
3
4
5
6
7

ARIMA

SARIMA

$$S\ ARIMA\ p\ d\ q\ x\ P\ D\ Q$$

$$x = 52$$

$$D < 1 \qquad d + D < 2$$

Resultados

1
2
3
4
5
6
7

ARIMA

pmdarima

Train: 66%
Test: 33%

Gripe

Covid-19

Resultados

1
2
3
4
5
6
7

ARIMA

pmdarima

Train: 66%
Test: 33%

p = 3
d = 0
q = ?

Gripe

Covid-19

```
ARIMA(1,0,1): AIC=-2502.616, Time=0.08 sec
ARIMA(0,0,0): AIC=-1781.321, Time=0.02 sec
ARIMA(1,0,0): AIC=-2496.283, Time=0.02 sec
ARIMA(0,0,1): AIC=-2045.896, Time=0.08 sec
ARIMA(2,0,1): AIC=-2502.681, Time=0.10 sec
ARIMA(2,0,0): AIC=-2503.956, Time=0.04 sec
ARIMA(3,0,0): AIC=-2502.858, Time=0.04 sec
ARIMA(3,0,1): AIC=-2500.715, Time=0.06 sec
ARIMA(2,0,0): AIC=-2513.092, Time=0.09 sec
ARIMA(1,0,0): AIC=-2509.250, Time=0.06 sec
ARIMA(3,0,0): AIC=-2513.549, Time=0.11 sec
ARIMA(3,0,1): AIC=-2510.789, Time=0.15 sec
ARIMA(2,0,1): AIC=-2511.336, Time=0.09 sec
```

Best model: ARIMA(3,0,0)
Total fit time: 0.955 seconds

Resultados

1
2
3
4
5
6
7

ARIMA

pmdarima

Gripe

Covid-19

Train: 66%

Test: 33%

$p = 3$

$d = 0$

$q = ?$

ARIMA Results

Dep. Variable:	y	No. Observations:	285
Model:	SARIMAX(3, 0, 0)	Log Likelihood	1261.775
Date:	Mon, 04 Jul 2022	AIC	-2513.549
Time:	18:06:51	BIC	-2495.287
Sample:	0	HQIC	-2506.228
	- 285		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0011	0.000	2.947	0.003	0.000	0.002
ar.L1	0.7551	0.039	19.120	0.000	0.678	0.833
ar.L2	0.2156	0.076	2.826	0.005	0.066	0.365
ar.L3	-0.0945	0.054	-1.740	0.082	-0.201	0.012
sigma2	8.287e-06	3.68e-07	22.515	0.000	7.57e-06	9.01e-06

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	1958.40
Prob(Q):	0.93	Prob(JB):	0.00
Heteroskedasticity (H):	2.53	Skew:	2.01
Prob(H) (two-sided):	0.00	Kurtosis:	15.20

Best model: ARIMA(3,0,0)
Total fit time: 0.955 seconds

Resultados

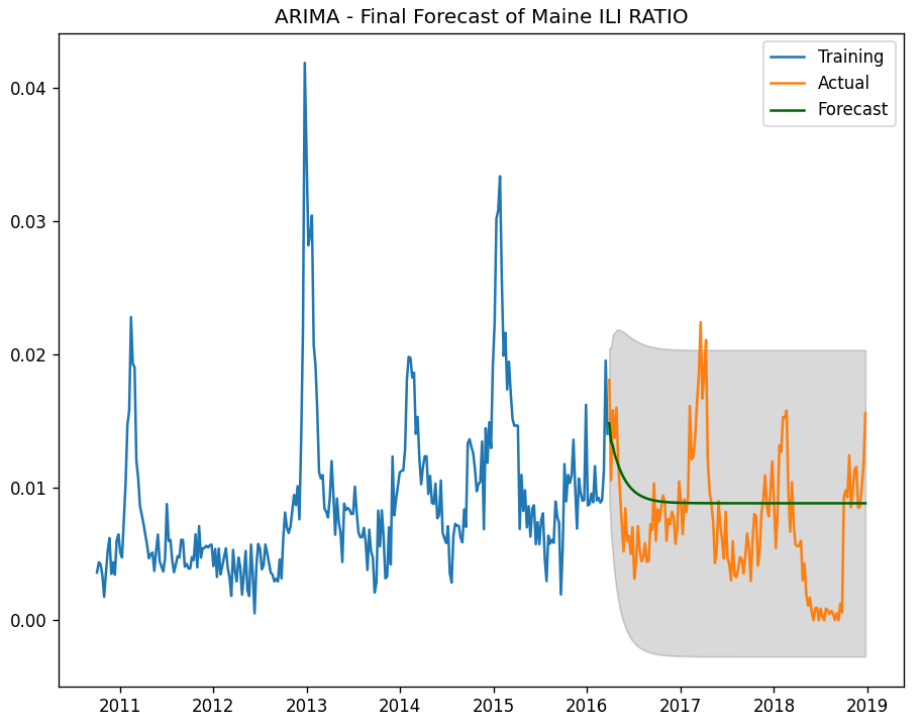
1 2 3 4 **5** 6 7

ARIMA

pmdarima

Train: 66%
Test: 33%

p = 3
d = 0
q = ?



Gripe

Covid-19

Best model: ARIMA(3,0,0)
Total fit time: 0.955 seconds

Resultados

1
2
3
4
5
6
7

SARIMA

pmdarima

Train: 66%
Test: 33%

Gripe

Covid-19

```
ARIMA(1,0,0)(0,1,0)[52]: AIC=-1895.071, Time=2.41 sec
ARIMA(0,0,0)(0,1,0)[52]: AIC=-1611.793, Time=0.67 sec
ARIMA(1,0,0)(1,1,0)[52]: AIC=-1956.496, Time=37.00 sec
ARIMA(2,0,0)(1,1,1)[52]: AIC=-1930.597, Time=30.40 sec
ARIMA(3,0,0)(2,1,2)[52]: AIC=-1977.621, Time=165.19 sec
ARIMA(2,0,1)(2,1,2)[52]: AIC=-1972.857, Time=121.42 sec
ARIMA(1,0,1)(2,1,2)[52]: AIC=-1981.786, Time=102.56 sec
ARIMA(1,0,1)(1,1,2)[52]: AIC=-1942.125, Time=34.02 sec
ARIMA(1,0,1)(2,1,1)[52]: AIC=-1973.920, Time=101.92 sec
ARIMA(1,0,1)(1,1,1)[52]: AIC=-1930.285, Time=25.80 sec
ARIMA(0,0,1)(2,1,2)[52]: AIC=-1822.646, Time=64.05 sec
ARIMA(1,0,2)(2,1,2)[52]: AIC=-1970.469, Time=144.72 sec
ARIMA(2,0,2)(2,1,2)[52]: AIC=-1967.996, Time=106.83 sec
```

Best model: SARIMA(1,0,1)(2,1,2)[52]
Total fit time: 1907.721 seconds

Resultados

1
2
3
4
5
6
7

SARIMA

pmdarima

Gripe

Covid-19

Train: 66%

Test: 33%

SARIMAX Results

Dep. Variable:	y	No. Observations:	285
Model:	SARIMAX(1, 0, 1)x(2, 1, [1, 2], 52)	Log Likelihood	998.893
Date:	Mon, 04 Jul 2022	AIC	-1981.786
Time:	18:38:57	BIC	-1954.178
Sample:	0 - 285	HQIC	-1970.653
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0003	0.001	0.426	0.670	-0.001	0.002
ar.L1	0.8627	0.035	24.592	0.000	0.794	0.931
ma.L1	-0.1562	0.075	-2.087	0.037	-0.303	-0.010
ar.S.L52	-0.4533	3.261	-0.139	0.889	-6.845	5.938
ar.S.L104	-0.1604	0.453	-0.354	0.723	-1.048	0.727
ma.S.L52	-0.4345	3.154	-0.138	0.890	-6.616	5.747
ma.S.L104	-0.0870	2.637	-0.033	0.974	-5.256	5.082
sigma2	8.972e-06	1.48e-06	6.053	0.000	6.07e-06	1.19e-05

Ljung-Box (L1) (Q):	0.47	Jarque-Bera (JB):	993.14
Prob(Q):	0.49	Prob(JB):	0.00
Heteroskedasticity (H):	0.86	Skew:	1.64
Prob(H) (two-sided):	0.51	Kurtosis:	12.57

Best model: SARIMA(1,0,1)(2,1,2)[52]
Total fit time: 1907.721 seconds

Resultados

1
2
3
4
5
6
7

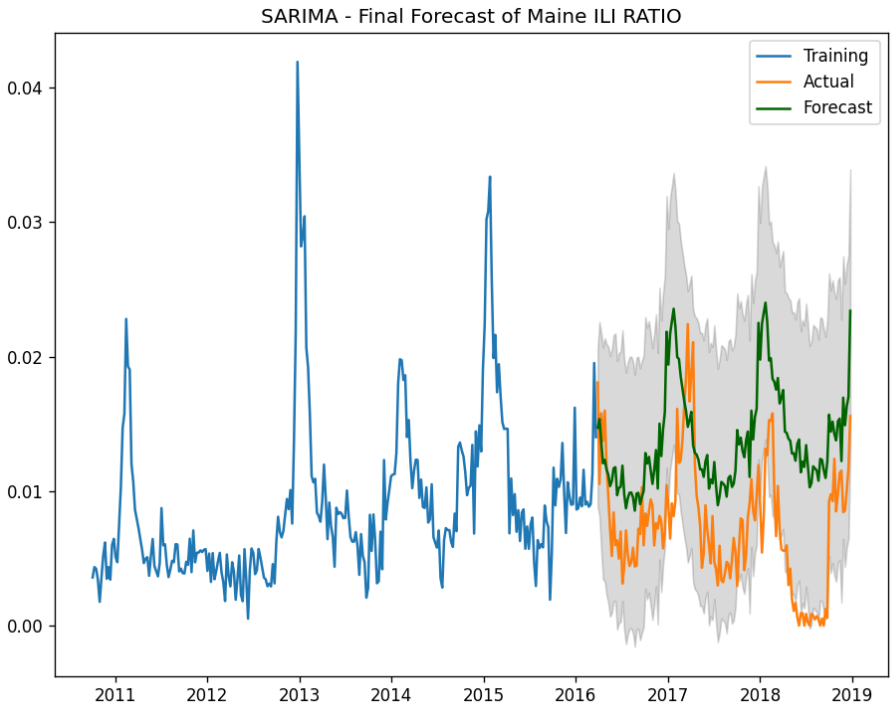
SARIMA

pmdarima

Train: 66%
Test: 33%

Gripe

Covid-19



Best model: SARIMA(1,0,1)(2,1,2)[52]
Total fit time: 1907.721 seconds

Resultados

1
2
3
4
5
6
7

ARIMA

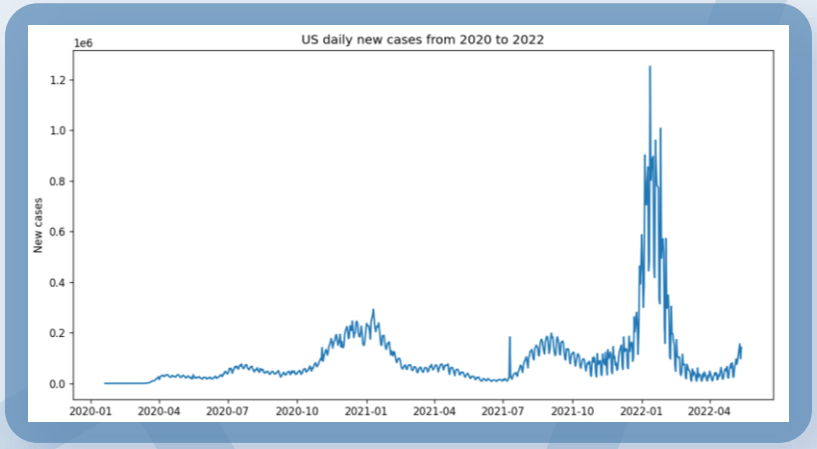
pmdarima

Gripe

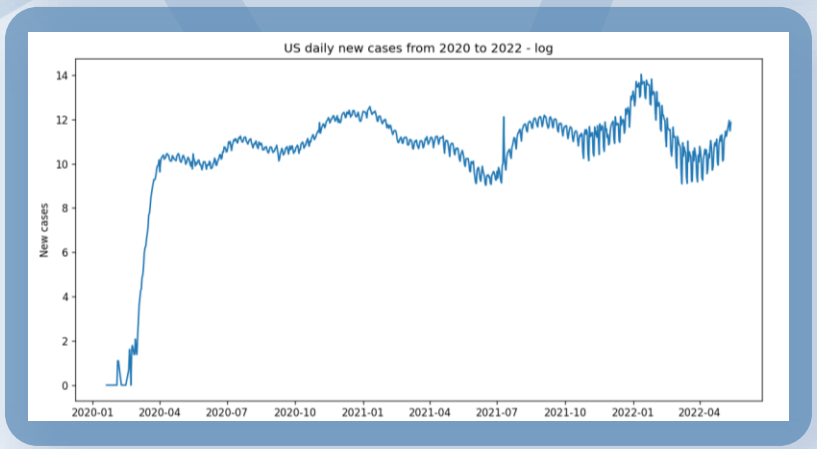
Covid-19

Best model: ARIMA(1,0,3)
Total fit time: 2.492 seconds

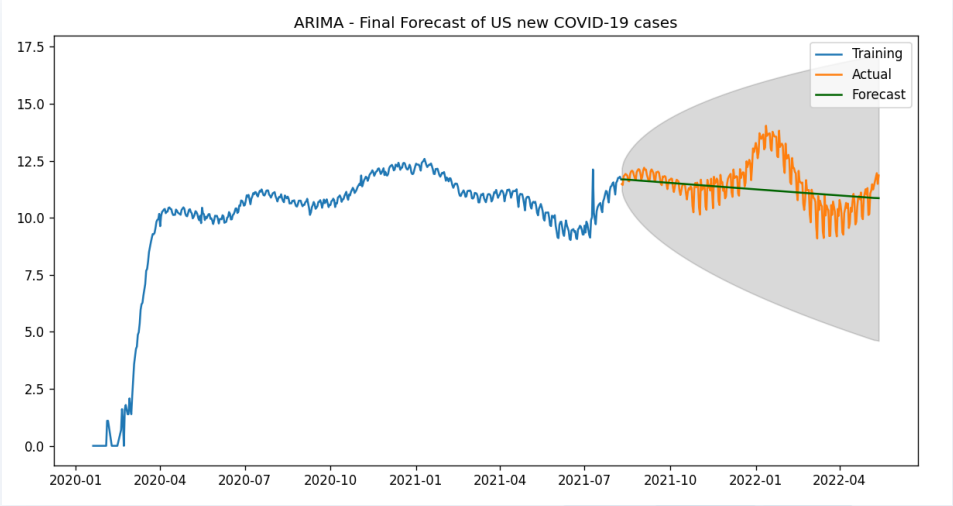
1.



2.



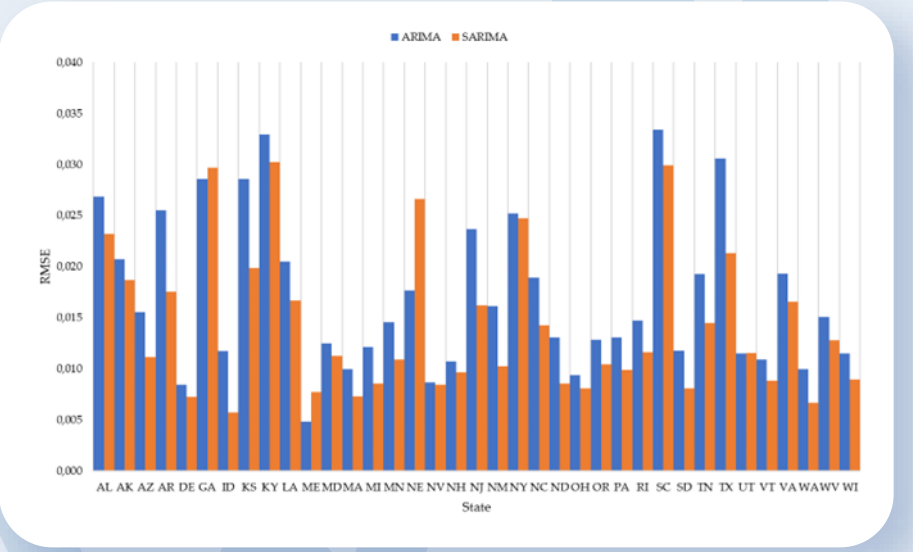
3.



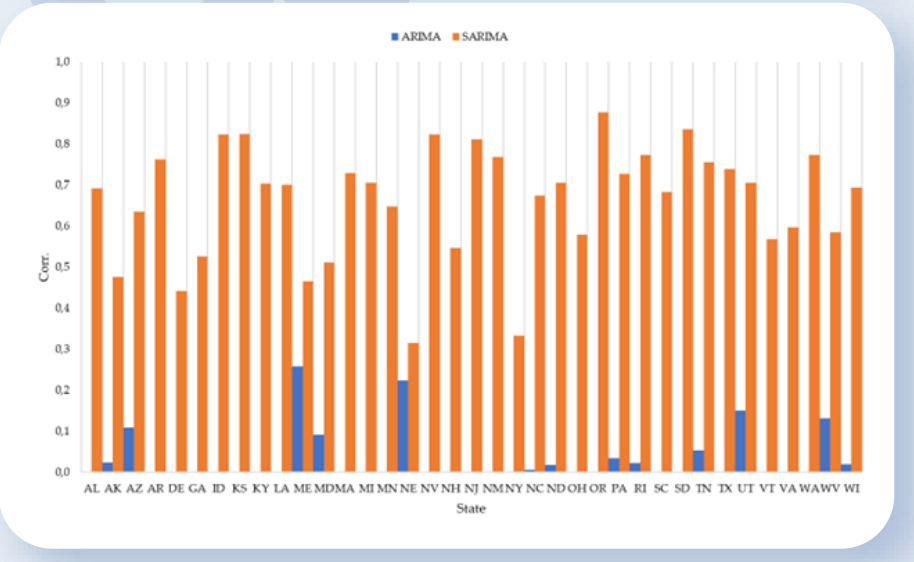
Discusión de los resultados

1 2 3 4 5 **6** 7

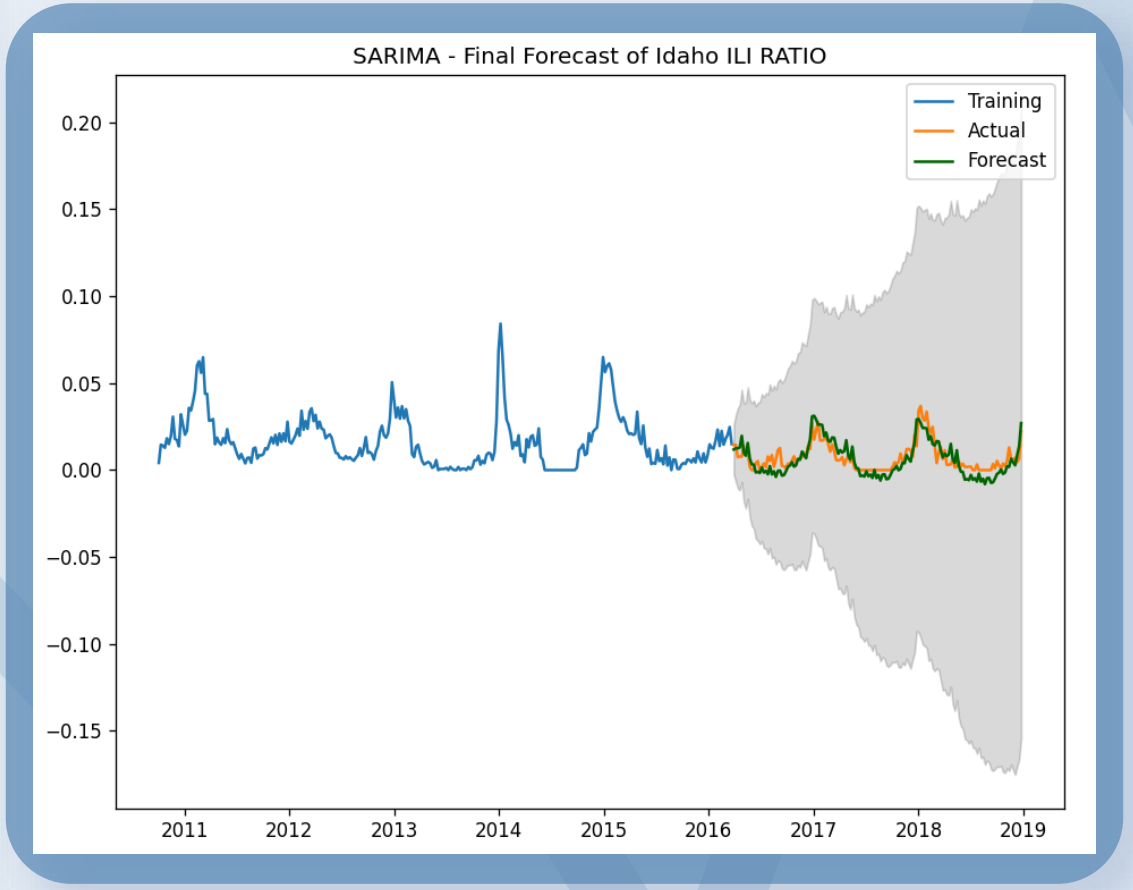
RMSE



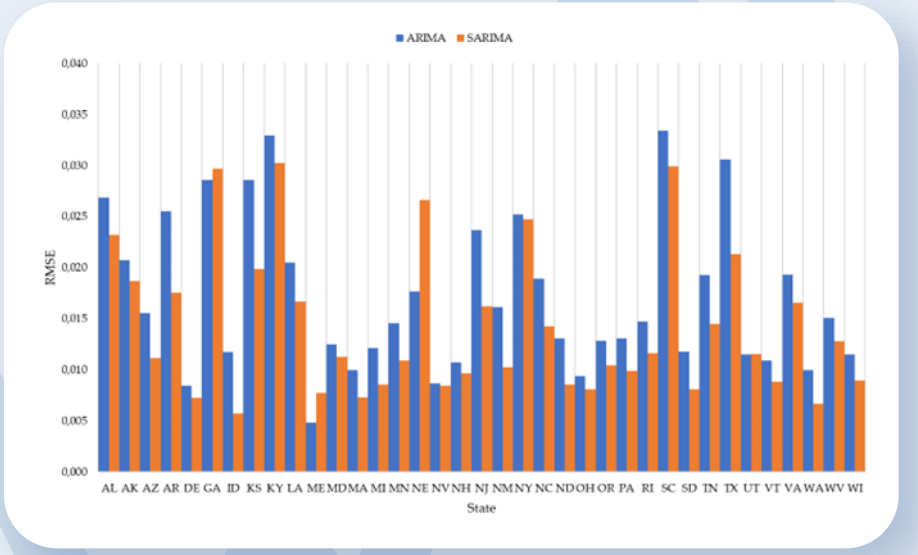
Corr.



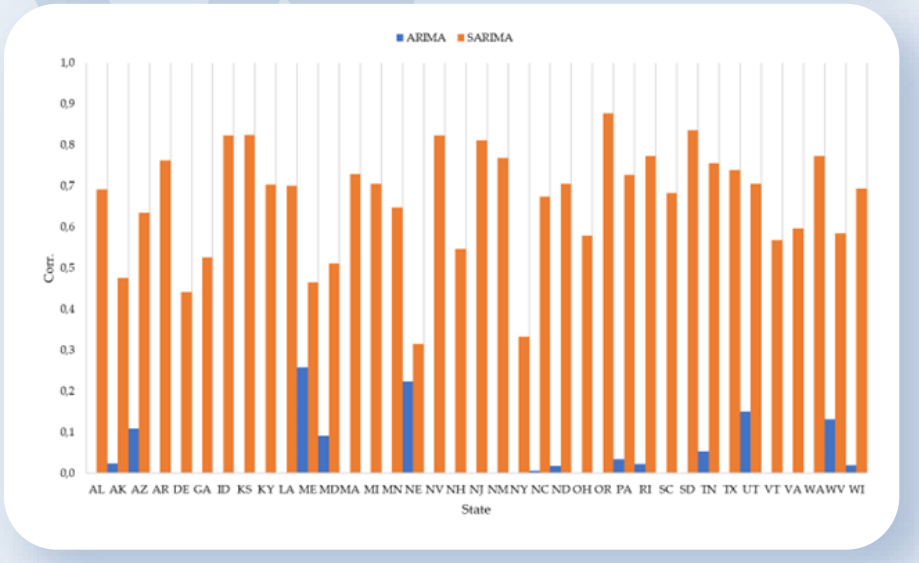
Discusión de los resultados



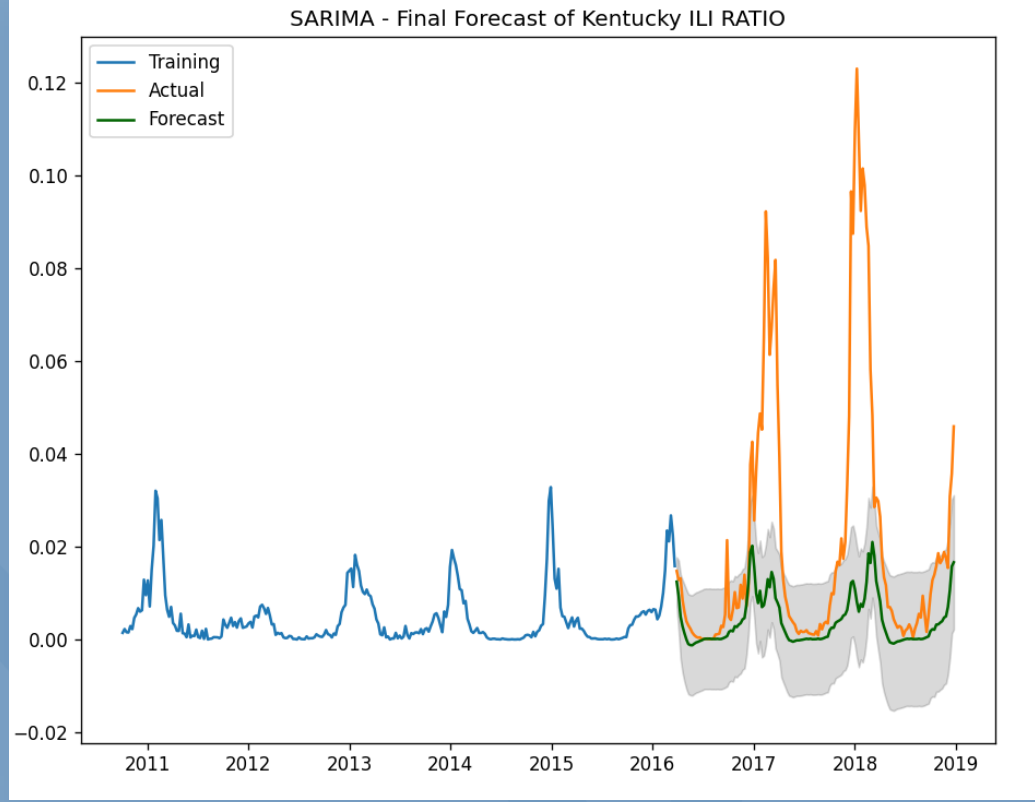
RMSE



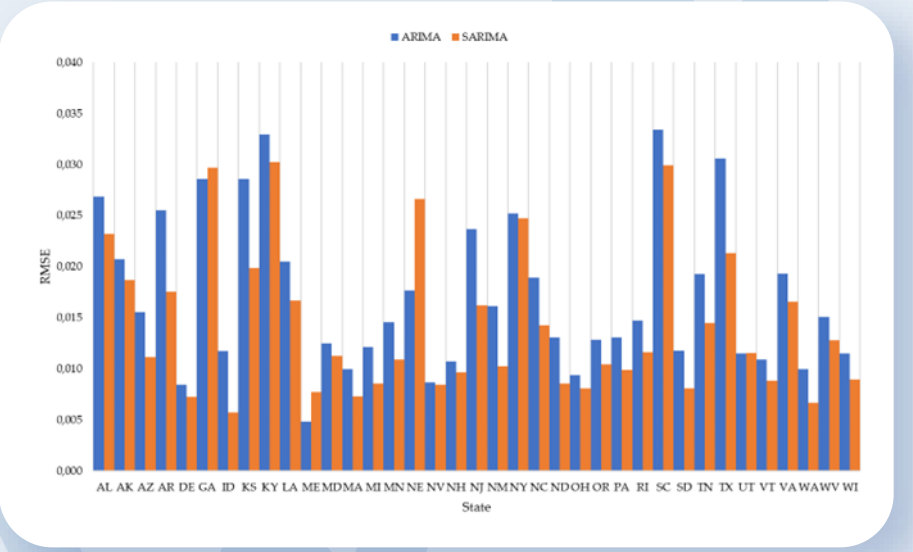
Corr.



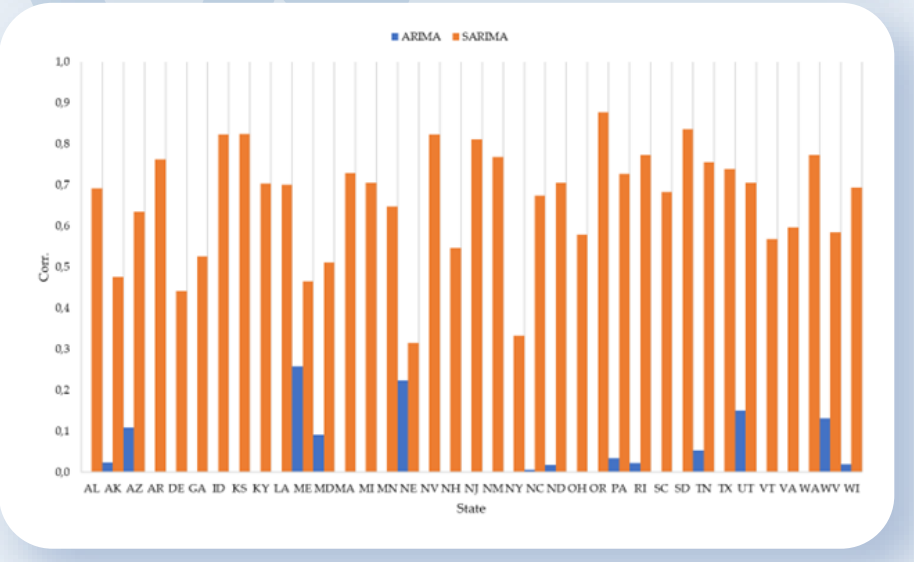
Discusión de los resultados



RMSE



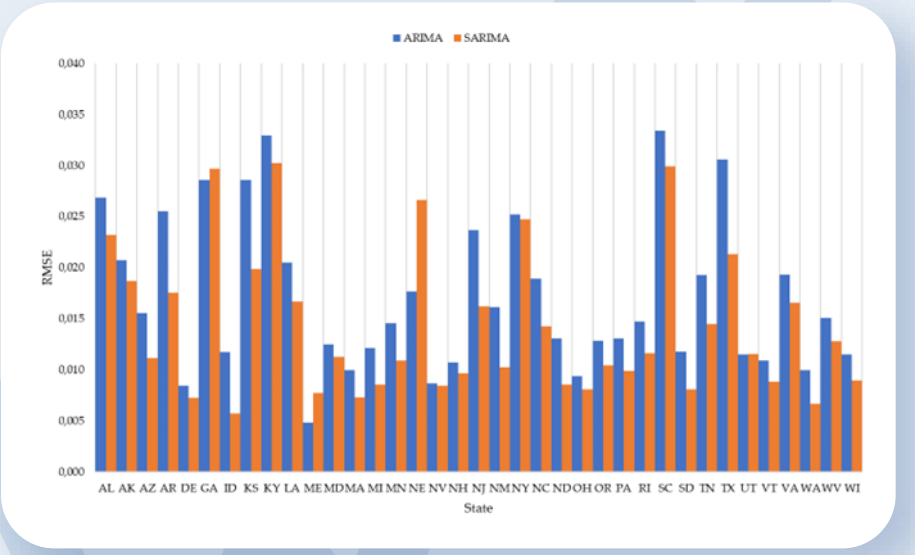
Corr.



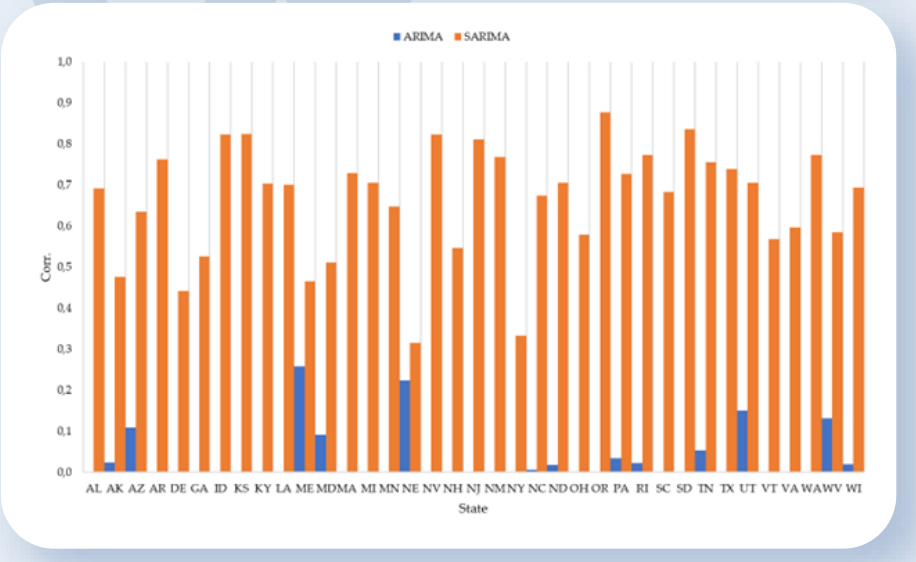
Discusión de los resultados

Modelo	Pearson Correlation
ARIMA	0,769 (+0%)
LTSM	0,924 (+19,9%)
Seq2Seq	0,920 (+19,5%)
Transformer	0,928 (+20,7%)
ARIMA TFG	-0,051
SARIMA TFG	0,662

RMSE



Corr.



Conclusiones

1 2 3 4 5 6 7

Gripe y Covid-19

ARIMA

SARIMA

Resultados

Comparación

GRACIAS POR SU ATENCIÓN