

Predicción de defunciones por COVID-19 en la región de Lima con redes neuronales en MATLAB

Pronóstico de series de tiempo de muertes por el virus y sus
comorbilidades

Integrantes: Fiorella Calderón Camacho

Aramis Palacios Aguirre

Ariana Villafuerte Kanagusuku

Asesor: Victor Ayma Quirita

Contenido:

- Introducción
- Marco Teórico
- Metodología
- Resultados
- Conclusiones

Contenido:

- **Introducción**
- Marco Teórico
- Metodología
- Resultados
- Conclusiones

Introducción:

Planteamiento del problema: Se han alcanzado niveles bajos de eficacia en la predicción de defunciones por COVID-19 en Perú

teleSURtv.net NOTICIAS VIDEOS MULTIMEDIA OPINIÓN BLOGS PROGRAMAS EMPLEOS

Noticias > Latinoamérica y el Caribe

Perú sobrepasa los 60.000 decesos por coronavirus



El pico más alto de fallecidos reportado en Perú se reportó el pasado 10 de abril, cuando se registraron 314 decesos. | Foto: EFE

🔍 | ☰ Menú | **GESTIÓN** | Muertes en Perú por COVID-19 subirán en 60% hasta fin de año,...

Muertes en Perú por COVID-19 subirán en 60% hasta fin de año, prevé universidad de EE.UU.

El Instituto de Métricas y Evaluaciones de Salud (IHME) de la Universidad de Washington indicó que en el escenario más probable, las muertes acumuladas a nivel mundial aumentarán en 1.9 millones de aquí a final de año.

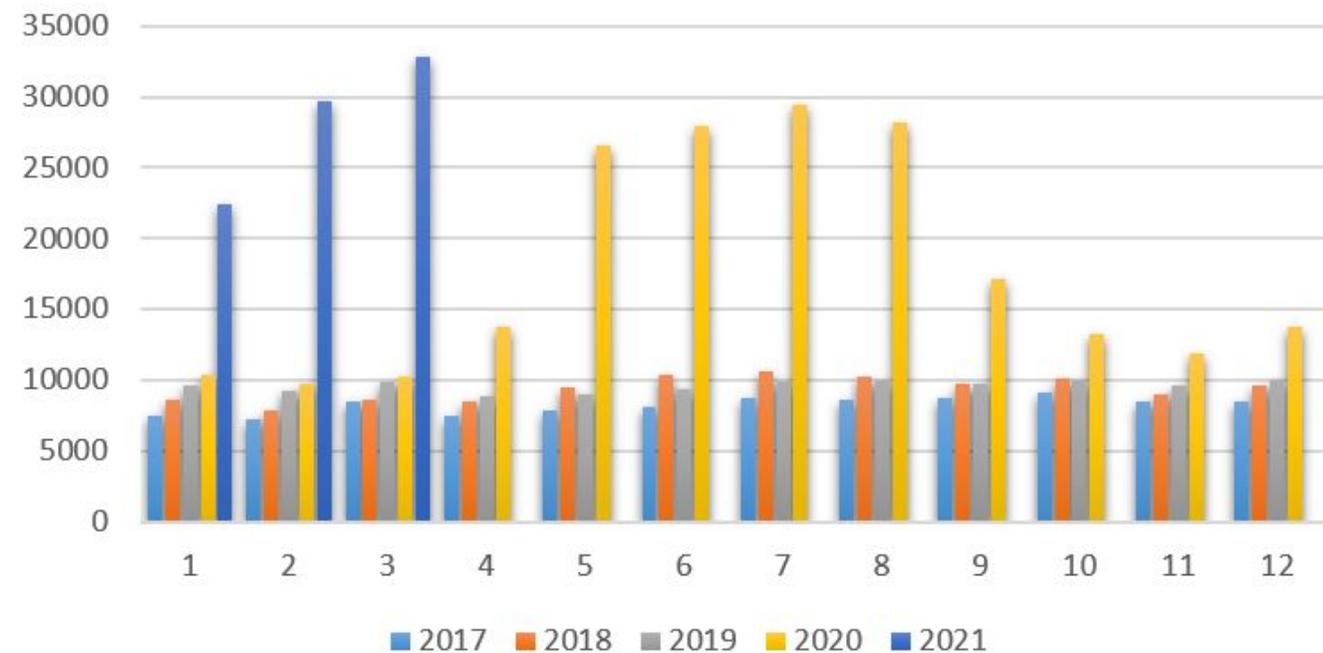


Introducción:

Motivación

La herramienta a desarrollar podría ser de utilidad al gobierno para estar preparados ante una posible ola de contagios y ver la mortandad que tendrá la COVID-19 en un futuro.

Defunciones por mes entre 2017-2021



Introducción:

Objetivos

- General:

Desarrollar un modelo de predicción de muertes por COVID-19 en la región de Lima con los datos sobre defunciones brindadas por SINADEF mediante redes neuronales

- Específicos:

- Determinar el nivel de precisión del modelo desarrollado mediante el error cuadrático medio
- Conocer la variación de defunciones por Covid-19 debido a comorbilidades de los fallecidos mediante series de tiempo

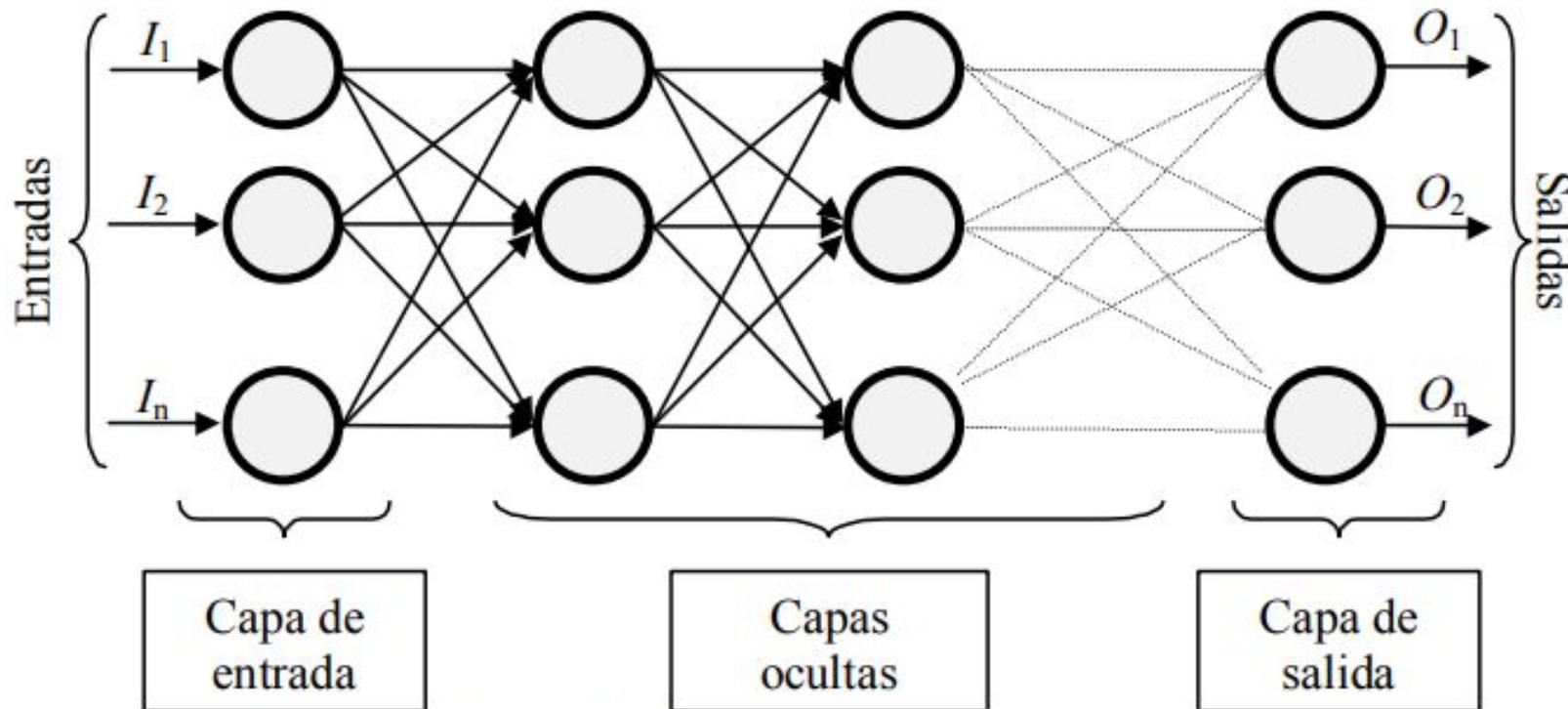
Contenido:

- Introducción
- **Marco teórico**
- Metodología
- Resultados
- Conclusiones

Proyecto

Marco teórico:

- **Red neuronal:** "Es un sistema de computación compuesto por un gran número de elementos simples, elementos de procesos muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas" (Matich, 2001).



Proyecto

Marco teórico:

- **NARNET (Nonlinear Autoregressive Neural Network):** Es una una red de ciclo cerrado que puede realizar predicciones multipasos donde los valores futuros de una serie de tiempo son predichos a partir de los valores pasados de la serie.

El proceso autorregresivo NARNET se puede describir de la siguiente manera:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_n y_{t-n} + \varepsilon_t$$

Proyecto

Marco teórico:

- **Matlab:**

- Función de entrenamiento:

Retropropagación de regularización bayesiana, "*trainbr*", que actualiza los pesos de los nodos, minimiza los errores cuadráticos y determina la combinación correcta.

→ Argumentos:

- "*feedbackDelays*"
- "*hiddenSizes*"
- "*feedbackMode*"
- "*trainFcn*"

Proyecto

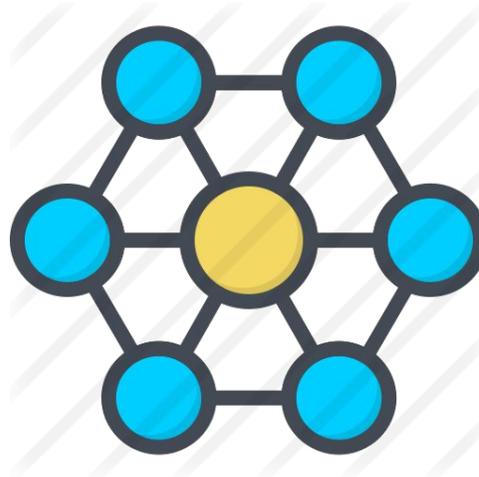
Contenido:

- Introducción
- **Metodología**
- Resultados
- Conclusiones

Metodología:

- **Propuesta de solución:**

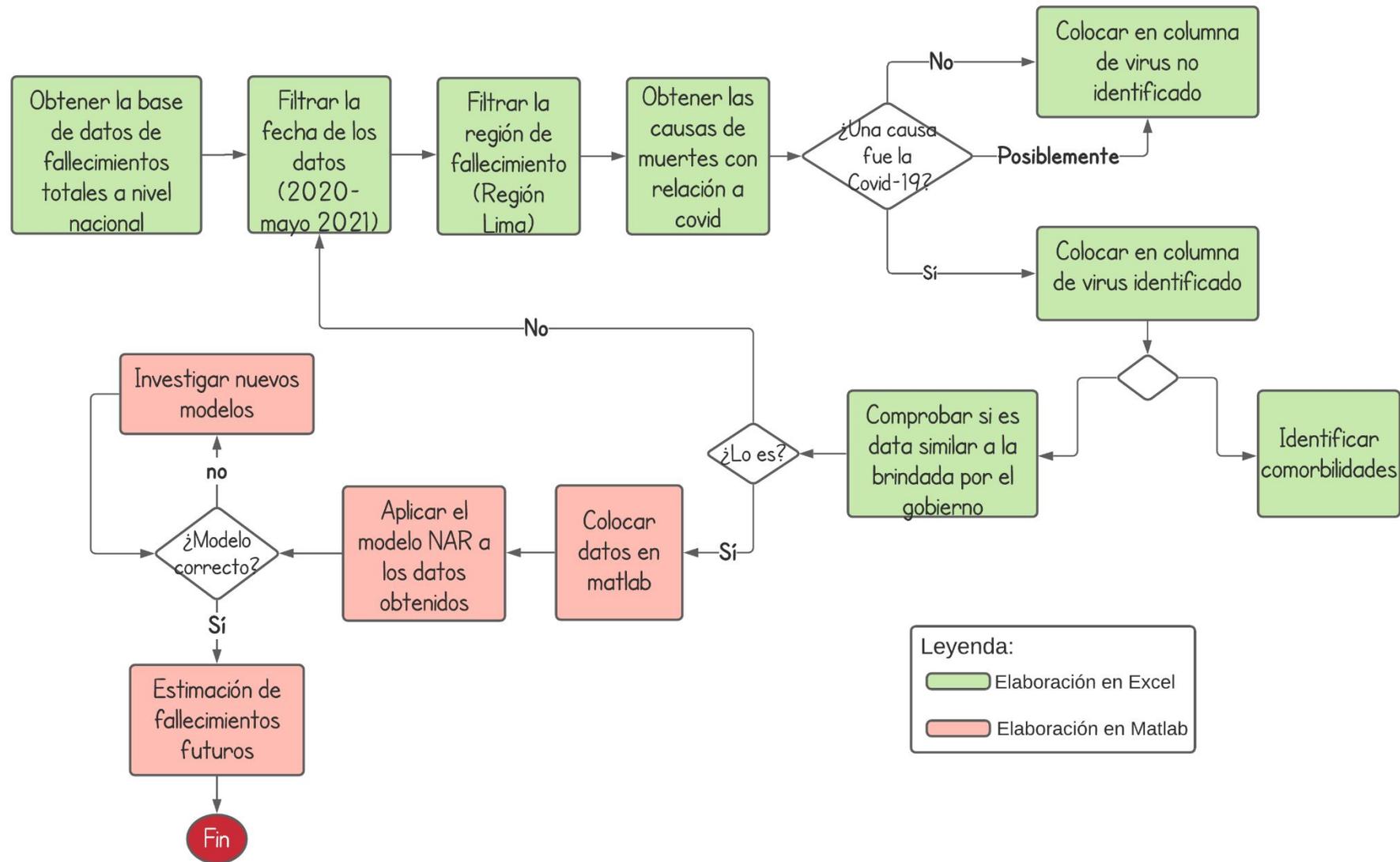
Implementación de una red neuronal no lineal autorregresiva (NAR) a través de Matlab.



Proyecto

Metodología:

● Procedimiento



Metodología:

- **Implementación realizada:**

- Mejoramiento de la base de datos para el modelo NAR
- Herramientas de Matlab para implementar la red neuronal en base a la data trabajada



Proyecto

Metodología:

- Mejoramiento de la base de datos para el modelo NAR

												Total	162,080
												Confirmado	141,910
Nº	SEXO	EDAD	TIEMPO EDAD	PAIS	DEPARTAMEN	PROVINCIA	DISTRITO DOMICILIO	FECHA	AÑO	MES	TODAS LAS CAUSAS	VIRUS IDENTIFICADO	
36209	MASCULINO	75	AÑOS	PERU	LIMA	LIMA	PUENTE PIEDRA	21/03/2020	2020	3	SEPSIS NO ESPECIFICAD	0	
36660	FEMENINO	29	AÑOS	PERU	LIMA	LIMA	BREÑA	28/03/2020	2020	3	NEUMONIA. SOSPECHA	0	
36819	FEMENINO	77	AÑOS	PERU	LIMA	LIMA	VILLA EL SALVADOR	30/03/2020	2020	3	INSUFICIENCIA RESPIRA	0	
36900	FEMENINO	91	AÑOS	PERU	LIMA	LIMA	CIENEGUILLA	15/03/2020	2020	3	INSUFICIENCIA RESPIRA	0	
36938	FEMENINO	81	AÑOS	PERU	LIMA	LIMA	SAN MIGUEL	10/03/2020	2020	3	COVID 19 SEPSIS SIN RE	1	
37212	MASCULINO	30	AÑOS	PERU	LIMA	LIMA	LA VICTORIA	1/04/2020	2020	4	NEUMONIA POR COVID	1	
37509	FEMENINO	65	AÑOS	PERU	LIMA	LIMA	LA MOLINA	31/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
37713	MASCULINO	85	AÑOS	PERU	LIMA	LIMA	JESUS MARIA	27/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
37872	MASCULINO	35	AÑOS	PERU	LIMA	LIMA	LURIGANCHO	31/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
37951	FEMENINO	83	AÑOS	PERU	LIMA	LIMA	BREÑA	29/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
38028	MASCULINO	59	AÑOS	PERU	LIMA	LIMA	CHORRILLOS	31/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
38065	FEMENINO	86	AÑOS	PERU	LIMA	LIMA	LA VICTORIA	30/03/2020	2020	3	NEUMONÍA - SOSPECHA	0	
38318	MASCULINO	60	AÑOS	PERU	LIMA	LIMA	LIMA	30/03/2020	2020	3	SHOCK SEPTICO NEUMO	1	
38325	MASCULINO	60	AÑOS	PERU	LIMA	LIMA	EL AGUSTINO	30/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
38353	MASCULINO	53	AÑOS	PERU	LIMA	LIMA	SAN ISIDRO	30/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	
39106	MASCULINO	86	AÑOS	PERU	LIMA	LIMA	SURQUILLO	27/03/2020	2020	3	NEUMONIA - PROBABLE	0	

Metodología:

Obtención de datos de comorbilidades

			Total	162,078						
			Confirmado	141,910	11,843	10,278	5,934	555	1,030	
FECHA	AÑO	MES	TODAS LAS CAUSAS	VIRUS IDENTIFICADO	HTA	Diabete	Obesida	Asma	Cáncer	
21/03/2020	2020	3	SEPSIS NO ESPECIFICAD	0	0	0	0	0	0	0
28/03/2020	2020	3	NEUMONIA. SOSPECHA	0	0	0	0	0	0	0
30/03/2020	2020	3	INSUFICIENCIA RESPIRA	0	0	0	0	0	0	0
15/03/2020	2020	3	INSUFICIENCIA RESPIRA	0	0	0	0	0	0	0
10/03/2020	2020	3	COVID 19 SEPSIS SIN RE	1	0	0	0	0	0	0
1/04/2020	2020	4	NEUMONIA POR COVID	1	0	0	0	0	0	0
31/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	0	0	0	0
27/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	0	0	0	0
31/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	1	0	0	0
29/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	0	0	0	0
31/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	0	0	0	0
30/03/2020	2020	3	NEUMONÍA - SOSPECHA	0	0	0	0	0	0	0
30/03/2020	2020	3	SHOCK SEPTICO NEUMC	1	0	0	0	0	0	0
30/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	0	0	0	0
30/03/2020	2020	3	INSUFICIENCIA RESPIRA	1	0	0	0	0	0	0
27/03/2020	2020	3	NEUMONIA - PROBABLE	0	0	0	0	0	0	0

Metodología:

- Herramientas de Matlab para implementar el modelo en base a la data trabajada
- Se programaron seis redes neuronales para las siguientes series de tiempo (450 días desde el 08 de marzo del 2020 al 31 de mayo del 2021) en la región Lima:
 - decesos por COVID-19
 - decesos por COVID-19 relacionados con:
 - Asma,
 - Cáncer,
 - Diabetes,
 - Obesidad e
 - Hipertensión Arterial
 - Cálculo del RMSE para medir la precisión de la predicción

Proyecto

Contenido:

- Introducción
- Metodología
- **Resultados**
- Conclusiones

Resultados:

Experimento: construcción de 6 redes neuronales

- Los argumentos usados para esta red neuronal son los siguientes: `feedbackDelays = 1:22`, `hiddenSizes = 10`, `trainFcn = "trainbr"` y `feedbackMode = 'open'`
- El primer 85% de la data de las series de tiempo se usó para entrenar la red y el porcentaje restante de data se usó para validar la predicción y poner a prueba la red
- Una vez creada y entrenada la red de ciclo cerrado, se procedió con la predicción multipasos

Resultados:

Objetivo: Medir la precisión de la predicción y proyectar la serie de tiempo en el futuro

- Calcular los errores cuadráticos medios entre la data de los últimos 30 días de la serie de tiempo original y los últimos 30 días de la data pronosticada por la red neuronal.
- Realizar un pronóstico adicional de 60 días a partir del 31 de mayo del 2021

Proyecto

Resultados:

Fig. 2. Serie de tiempo original para los decesos por COVID-19

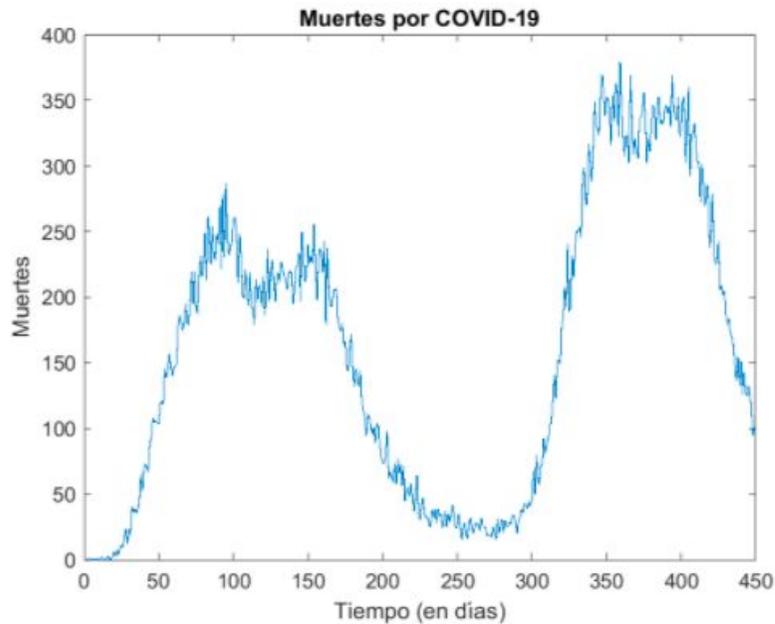


Fig. 3. Predicción de los últimos 30 días de la serie de tiempo de los decesos por COVID-19

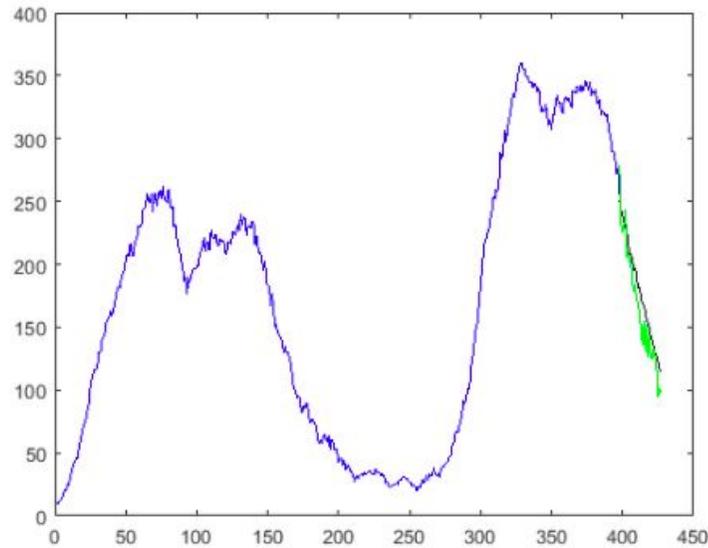
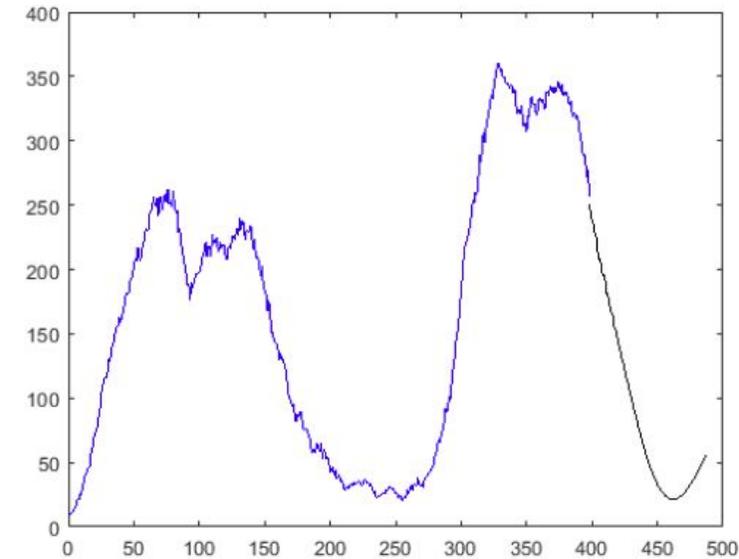


Fig. 4. Predicción de 60 días de la serie de tiempo de los decesos por COVID-19



En la Fig. 3 el valor del error cuadrático medio calculado es 49.3381.

Resultados:

Fig. 5. Serie de tiempo original para los decesos por COVID-19 relacionados con asma

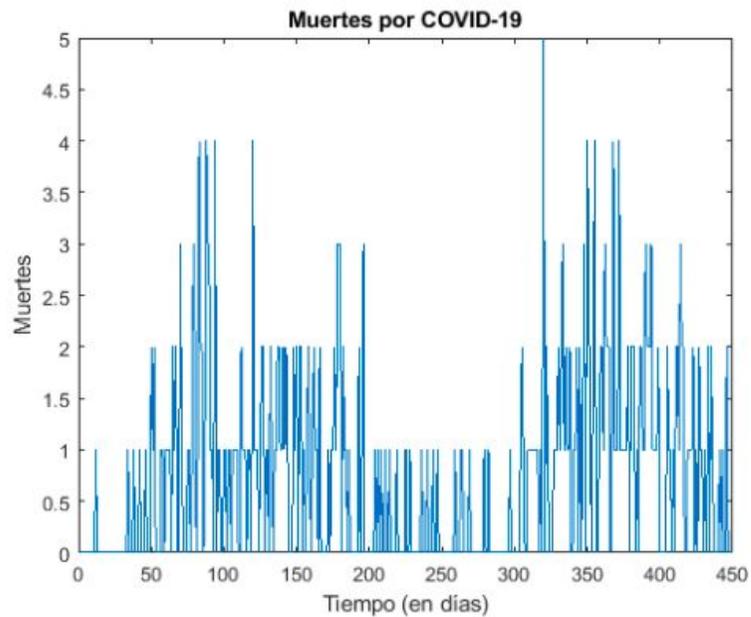


Fig. 6. Predicción de los últimos 30 días de la serie de tiempo de los decesos por COVID-19 relacionados con asma

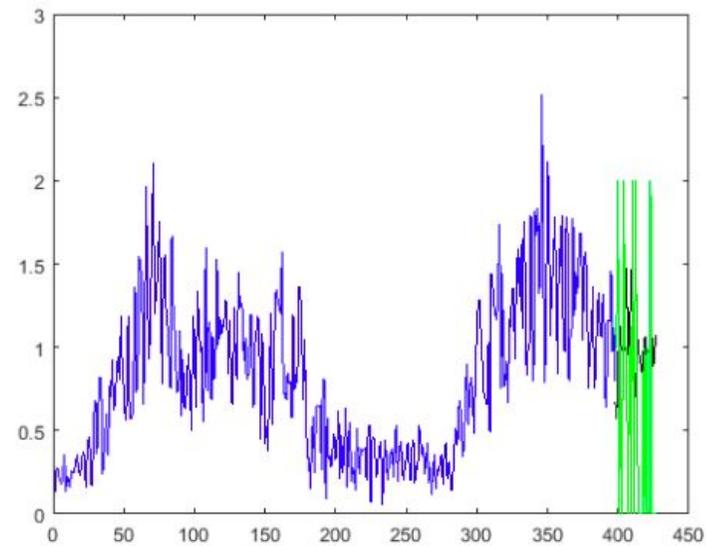
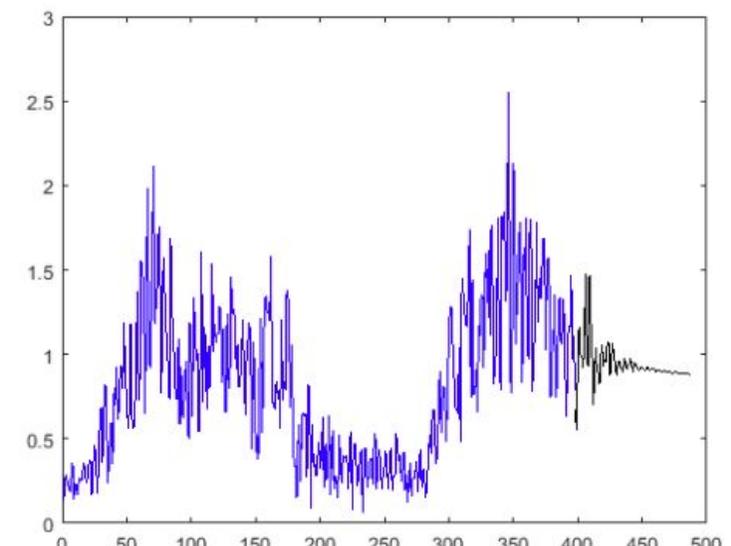


Fig. 7. Predicción de 60 días de la serie de tiempo de los decesos por COVID-19 relacionados con asma



En la Fig. 6 el valor del error cuadrático medio calculado es 0.9060.

Proyecto

Resultados:

Fig. 8. Serie de tiempo original para los decesos por COVID-19 relacionados con cáncer

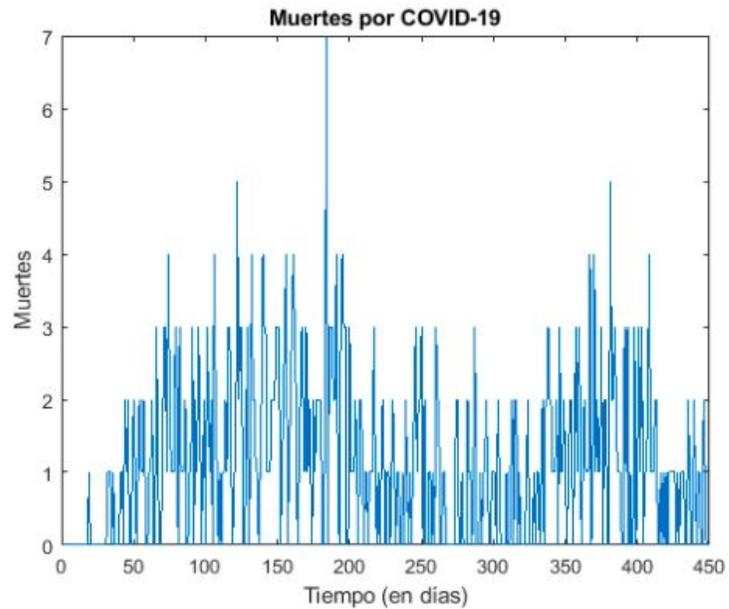


Fig. 9. Predicción de los últimos 30 días de la serie de tiempo de decesos por COVID-19 relacionados con cáncer

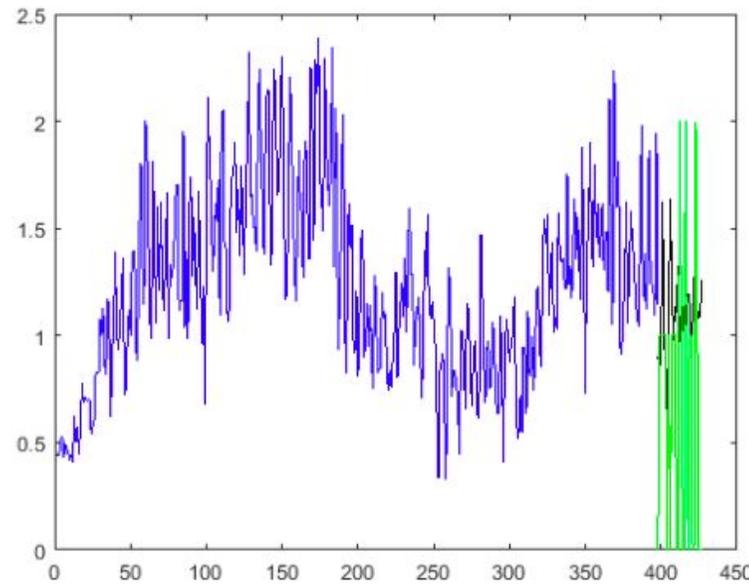
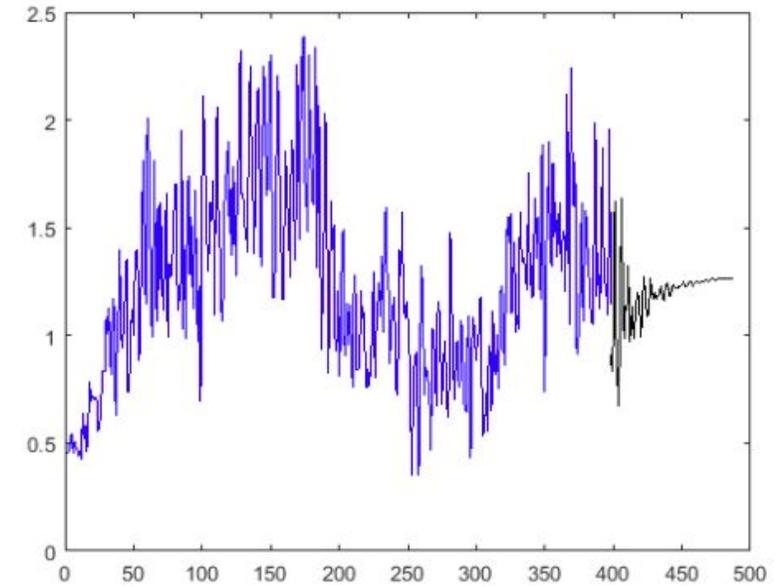


Fig. 10. Predicción de 60 días de la serie de tiempo de los decesos por COVID-19 relacionados con cáncer



En la Fig. 9 el valor del error cuadrático medio calculado es 0.7368.

Proyecto

Resultados:

Fig. 11. Serie de tiempo original para los decesos por COVID-19 relacionados con diabetes

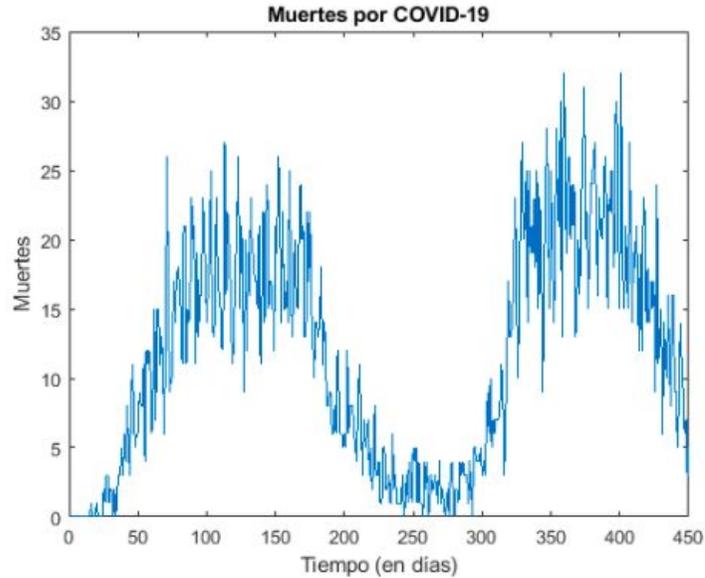


Fig. 12. Predicción de los últimos 30 días de la serie de tiempo de los decesos por COVID-19 relacionados con diabetes

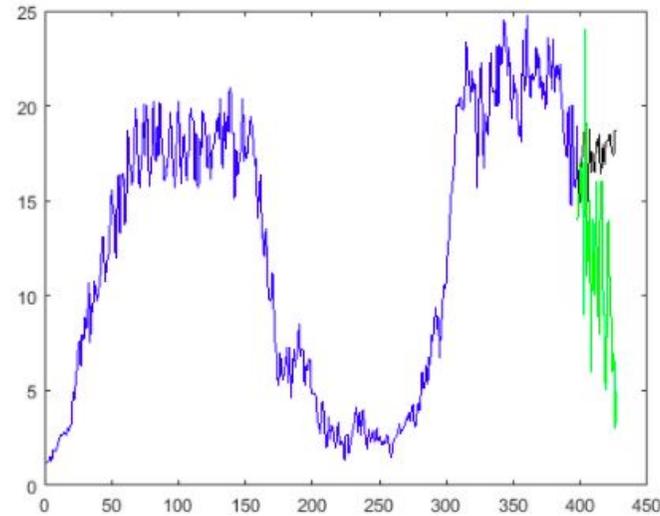
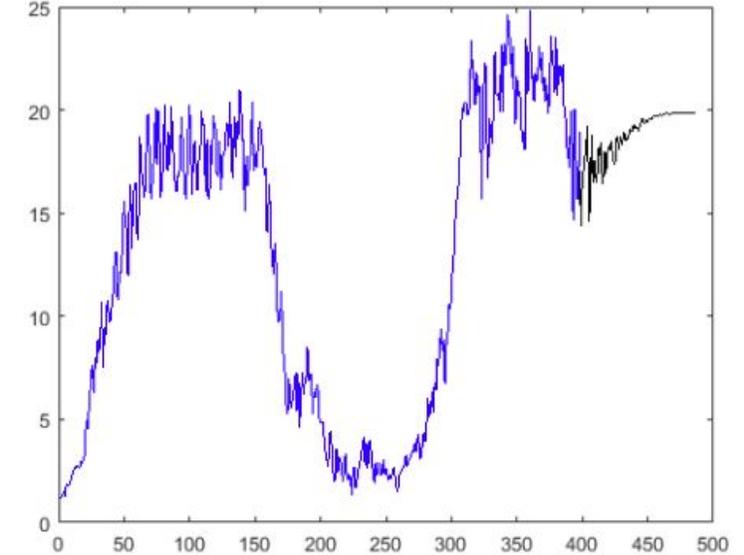


Fig. 13. Predicción de 60 días de la serie de tiempo de los decesos por COVID-19 relacionados con diabetes



En la Fig. 12 el valor del error cuadrático medio calculado es 5.5499.

Proyecto

Resultados:

Fig. 14. Serie de tiempo original para los decesos por COVID-19 relacionados con obesidad

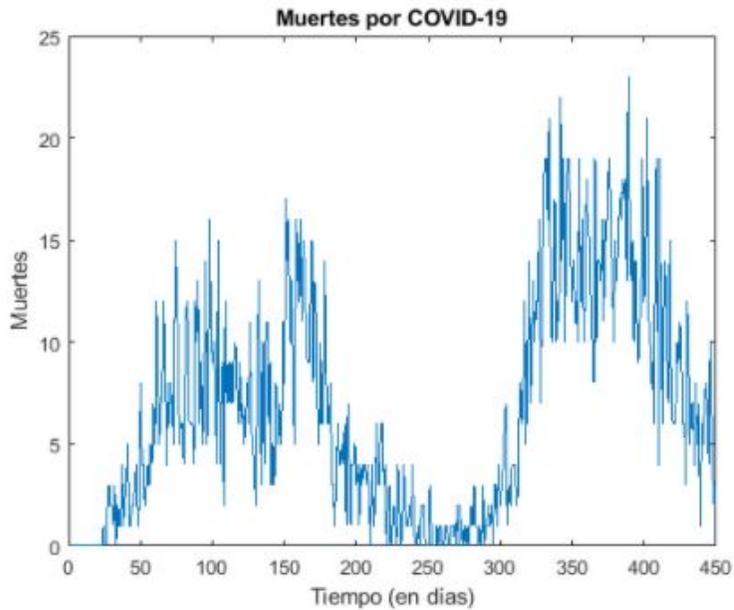


Fig. 15. Predicción de los últimos 30 días de la serie de tiempo de los decesos por COVID-19 relacionados con obesidad

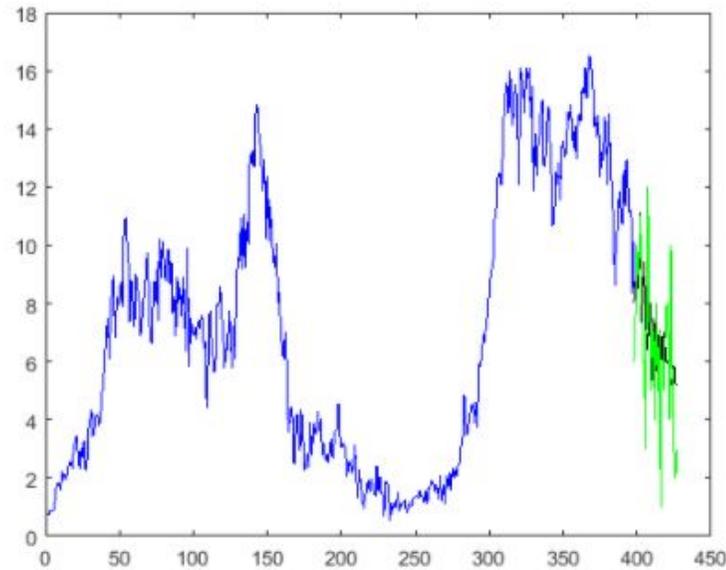
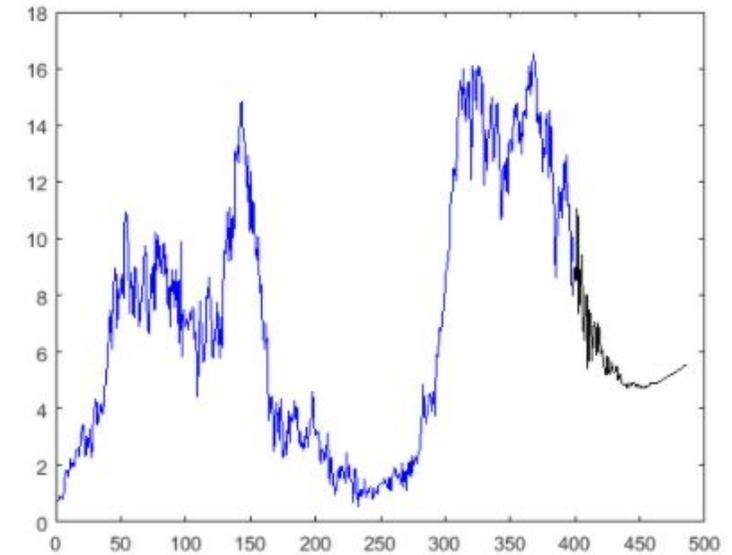


Fig. 16. Predicción de 60 días de la serie de tiempo de los decesos por COVID-19 relacionados con obesidad



En la Fig. 15 el valor del error cuadrático medio calculado es 3.5305.

Proyecto

Resultados:

Fig. 17. Serie de tiempo original para los decesos por COVID-19 relacionados con Hipertensión Arterial

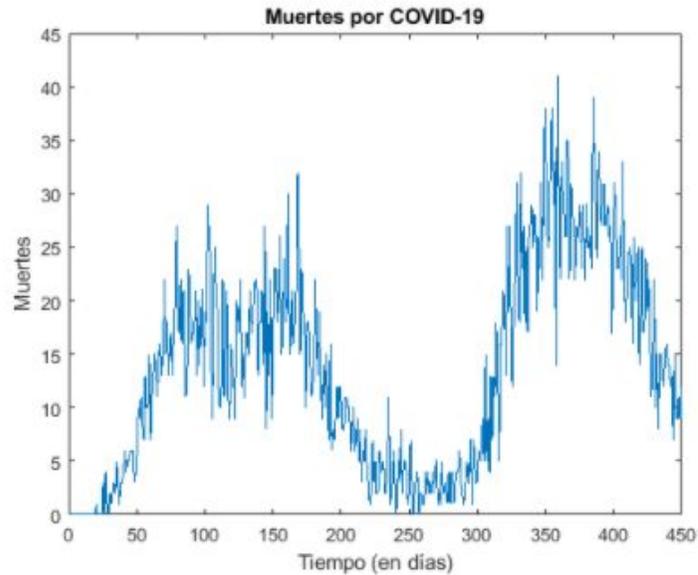


Fig. 18. Predicción de los últimos 30 días de la serie de tiempo de los decesos por COVID-19 relacionados con Hipertensión Arterial

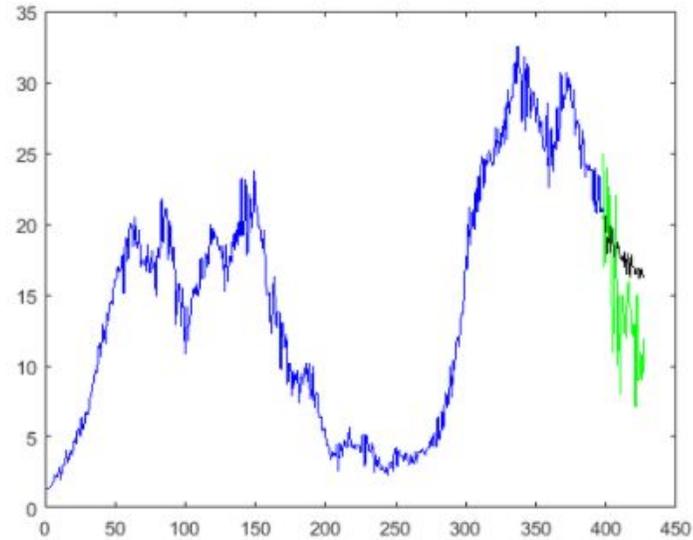
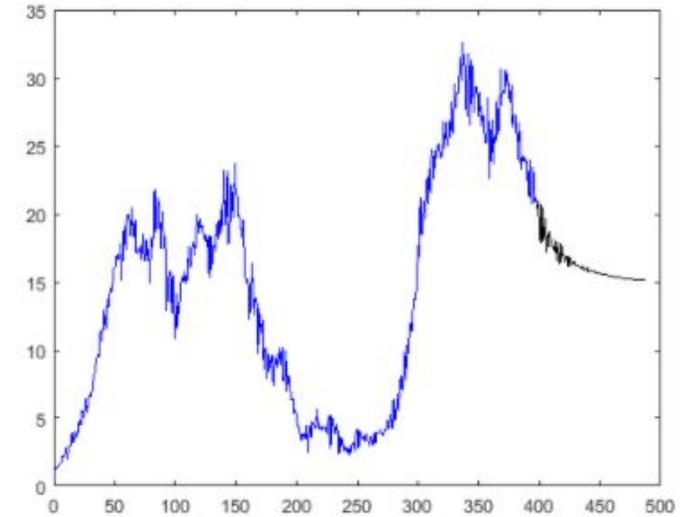


Fig. 19. Predicción de 60 días de la serie de tiempo de los decesos por COVID-19 relacionados con Hipertensión Arterial



En la Fig. 18 el valor del error cuadrático medio calculado es 4.4594.

Proyecto

Contenido:

- Introducción
- Metodología
- Resultados
- **Conclusiones**

Conclusiones:

- Se puede obtener un nivel satisfactorio de precisión en la predicción de series de tiempo mediante redes neuronales.
- Algunos modelos se desarrollaron mejor que otros en términos de precisión:
 - Precisión más baja:
 - Decesos por COVID-19 relacionados con diabetes
 - Decesos por COVID-19 relacionados con Hipertensión Arterial
- Las técnicas computacionales como las redes neuronales pueden tener un gran impacto en la creación de políticas para lograr el bienestar de toda una población.

Proyecto

Contenido:

- Introducción
- Metodología
- Resultados
- Conclusiones
- **Comentarios finales**

Comentarios finales:

- Escalabilidad del proyecto: con más variables de entrada y adecuadas a los datos de más regiones, puede ser lo suficientemente preciso para aprender las diversas dinámicas en diferentes puntos del país.
- Recomendaciones: una red neuronal que incluya más inputs adquiere un mayor peso, pues es más robusta y está menos sesgada de la realidad. Ej. medidas gubernamentales, feriados, etc.
- Un agradecimiento especial al profesor Víctor Ayma por el asesoramiento del presente trabajo.