

UNIVERSIDAD TECNOLÓGICA DE PANAMÁ

VICERRECTORÍA DE INVESTIGACIÓN, POSTGRADO Y EXTENSIÓN FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES

MAESTRÍA EN ANALÍTICA DE DATOS

MODELOS PREDICTIVOS

PROYECTO FINAL

ANÁLISIS DE MODELOS PREDICTIVOS PARA CASOS DE VIRUELA DEL MONO

ELABORADO POR:

|  |  |
| --- | --- |
| ANDRIÓN, MOISÉS | 3-739-542 |

PROFESOR:

JUAN M. CASTILLO, PhD

GRUPO:

1AN-212

2024

**CONTENIDO**

[INTRODUCCIÓN 3](#_Toc176627616)

[ANÁLISIS EXPLORATORIO DE DATOS (EDA) 4](#_Toc176627617)

[MODELOS PREDICTIVOS 6](#_Toc176627618)

[EVALUACIÓN DEL RENDIMIENTO DE LOS MODELOS 9](#_Toc176627619)

[CONCLUSIONES 12](#_Toc176627620)

[REFERENCIAS 13](#_Toc176627621)

# **INTRODUCCIÓN**

Este proyecto tendrá como objetivo principal predecir los casos diarios de viruela del mono utilizando técnicas avanzadas de series temporales, con el fin de contribuir significativamente a los esfuerzos de control y mitigación de la enfermedad. Dado el impacto global que el brote de viruela del mono ha tenido en términos de salud pública, se esperará que las predicciones precisas resulten esenciales para que los responsables de la toma de decisiones puedan anticipar posibles escenarios de propagación. Estas predicciones permitirán implementar medidas preventivas de manera oportuna y asignar recursos de forma más eficiente, mejorando la capacidad de respuesta de los sistemas de salud y las autoridades gubernamentales.

A lo largo del análisis, se emplearán diversos modelos predictivos de series temporales, entre ellos ARIMA (AutoRegressive Integrated Moving Average) y el método de suavizamiento exponencial Holt-Winters. Estos modelos se seleccionarán por su capacidad de identificar patrones complejos en los datos históricos de casos globales, tales como tendencias a corto y largo plazo y estacionalidades. El rendimiento de cada modelo será evaluado mediante el uso de métricas estándar, como el error absoluto medio (MAD), el error porcentual medio absoluto (MAPE) y el error cuadrático medio (RMSE), con el fin de identificar el que ofrezca las predicciones más precisas.

De esta manera, se buscará no solo predecir con exactitud la evolución de los casos de viruela del mono, sino también generar información valiosa para apoyar el desarrollo de estrategias efectivas de mitigación en el futuro.

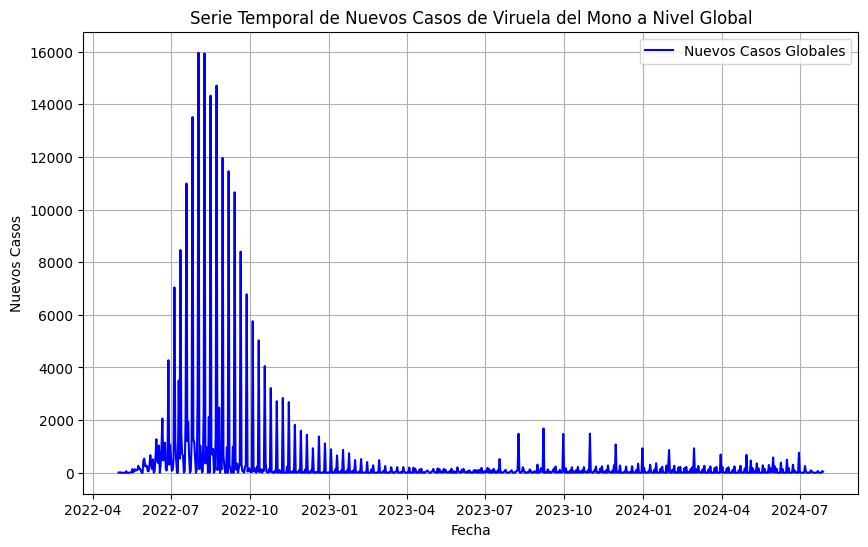
# **ANÁLISIS EXPLORATORIO DE DATOS (EDA)**

Descripción de los datos:

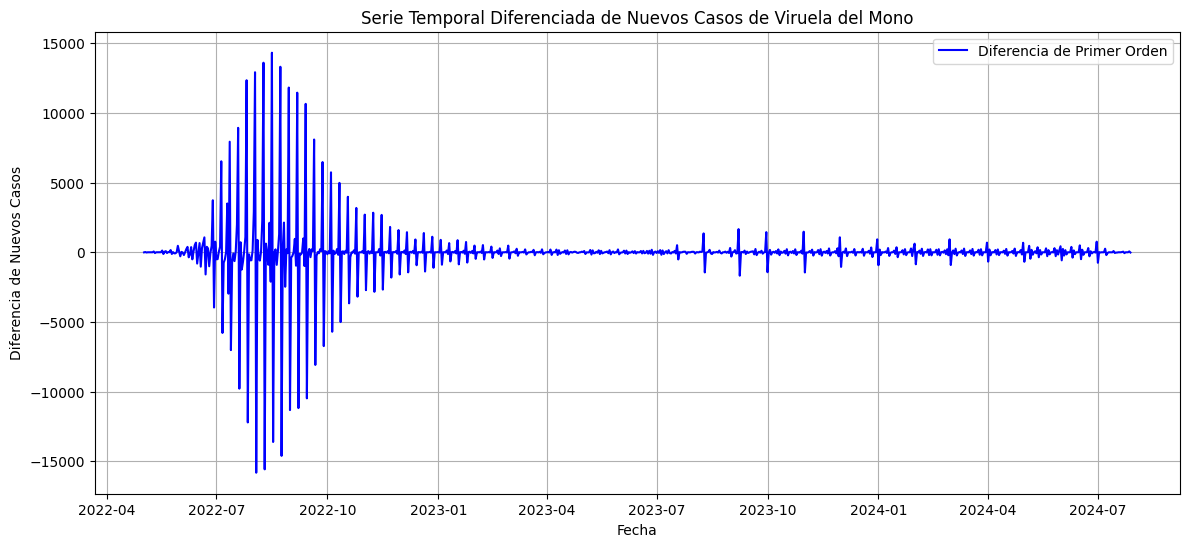
Los datos utilizados en este proyecto provienen de Kaggle. El conjunto de datos incluye registros diarios de casos de viruela del mono a nivel global. Cada registro incluye información como la fecha y el número de nuevos casos reportados.

Visualizaciones:

1. Gráfico de Serie Temporal: Se creó un gráfico de la evolución de los casos globales de viruela del mono a lo largo del tiempo, lo que revela un patrón de crecimiento en ciertos períodos y disminuciones en otros.

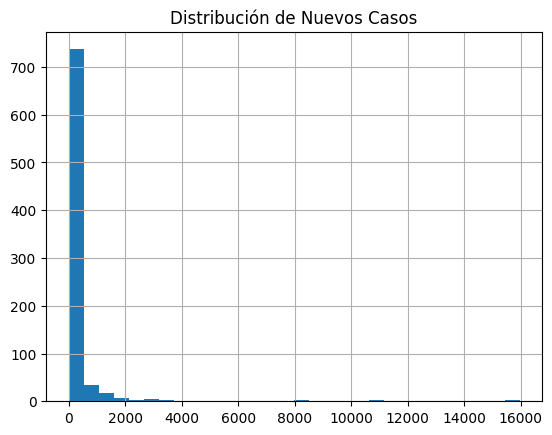


Gráfica 1: Serie temporal de nuevos casos de VM a nivel global



Gráfica 2: Serie temporal diferencial de nuevos casos de VM

2. Histogramas: Para comprender la distribución de los casos diarios, se generaron histogramas que muestran la frecuencia de días con diferentes niveles de casos.



Gráfica 3: Histograma de distribución de casos nuevos

Estadísticas Descriptivas:

* Media de nuevos casos diarios: 3.77
* Desviación estándar: 1560.65
* Número total de días analizados: 79,033
* Máximo número de casos reportados en un día: 5,319

# **MODELOS PREDICTIVOS**

ARIMA

Descripción del Modelo: ARIMA es un modelo popular para series temporales, que combina términos autoregresivos (AR), diferenciación para hacer los datos estacionarios (I), y un componente de media móvil (MA). En este caso, se utilizó el modelo ARIMA (1,1,1), lo que significa que se usó un término autoregresivo de primer orden, una diferenciación y un componente de media móvil.

* Resultados del Modelo:

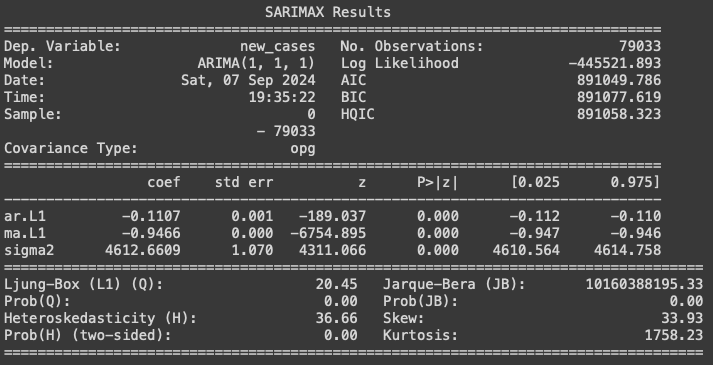


Figura 1: Resultados del modelo predictivo ARIMA

* Predicciones para los próximos 30 días:

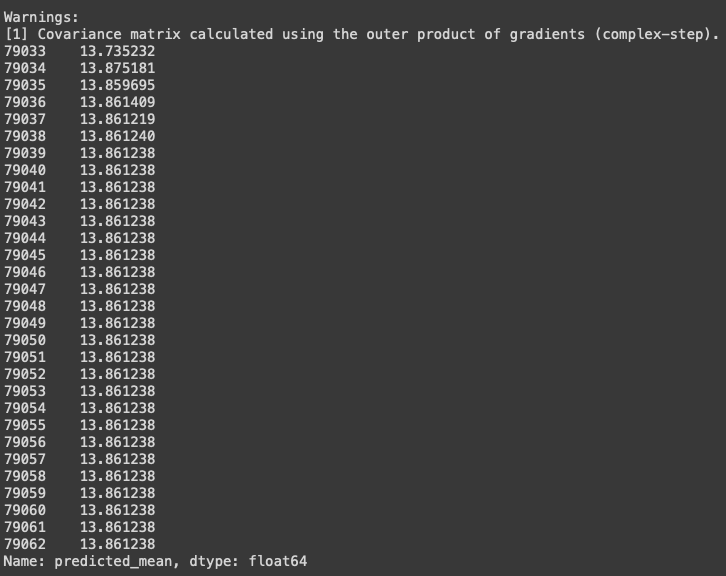
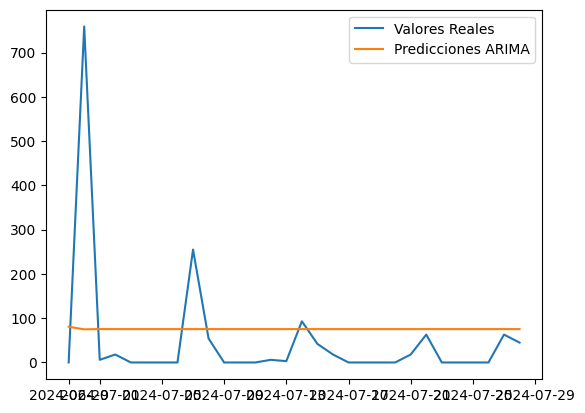


Figura 2: Predicciones para los próximos 30 días por el modleo ARIMA

     - Gráfico comparativo entre los valores reales y las predicciones:



Gráfica 4:Gráfico comparativo entre los valores reales y predicciones

HOLT-WINTERS

Descripción del Modelo: El modelo Holt-Winters es un modelo de suavizamiento exponencial que se adapta bien a datos con tendencia y estacionalidad. Se ajustó el modelo utilizando una tendencia aditiva y un componente estacional semanal.

Resultados del Modelo:

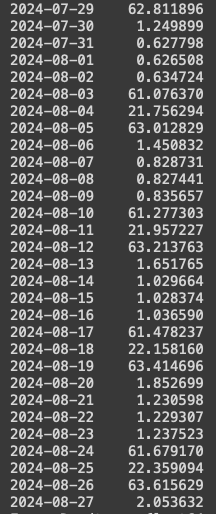


Figura 3: Resultados del modelo Holt-Winters

Grafico de las predicciones Holt-Winter:

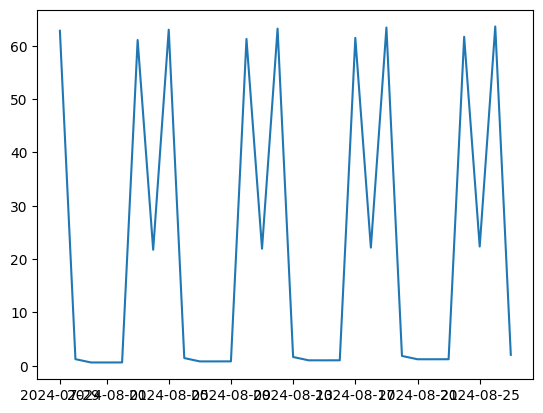


Figura 4: Gráfico para predicciones Holt-Winters

# **EVALUACIÓN DEL RENDIMIENTO DE LOS MODELOS**

Para evaluar el rendimiento de los modelos, se utilizaron las siguientes métricas:

* MAD ARIMA: 86.22793011773746
* MAD Holt-Winters: 60.45300446843592
* MAPE ARIMA: 435.08741485256843
* MAPE Holt-Winters: 280.58003081118324
* RMSE ARIMA: 143.75441791420644
* RMSE Holt-Winters: 121.70873256811824

En esta sección se compararán los modelos ARIMA y Holt-Winters en términos de tres métricas de evaluación del rendimiento: MAD (Mean Absolute Deviation), MAPE (Mean Absolute Percentage Error) y RMSE (Root Mean Squared Error). Estas métricas permiten medir la precisión de los modelos al predecir los casos de viruela del mono.

Comparación de Resultados:

1. MAD (Mean Absolute Deviation):

* ARIMA: 86.23
* Holt-Winters: 60.45

El MAD refleja el error absoluto promedio entre los valores reales y las predicciones. El modelo Holt-Winters muestra un MAD más bajo (60.45) en comparación con ARIMA (86.23), lo que indica que las predicciones de Holt-Winters tienen, en promedio, un error absoluto menor. Esto sugiere que el modelo Holt-Winters es más preciso en la predicción de los casos de viruela del mono.

2. MAPE (Mean Absolute Percentage Error):

* ARIMA: 435.09%
* Holt-Winters: 280.58%

El MAPE mide el error porcentual promedio. Aquí, ambos modelos tienen valores de MAPE altos, lo que indica que los errores en las predicciones son significativos en relación con los valores reales. Sin embargo, el Holt-Winters (280.58%) sigue siendo considerablemente mejor que ARIMA (435.09%), lo que sugiere que Holt-Winters tiene un error porcentual más bajo, y por lo tanto, es más adecuado cuando se consideran las diferencias relativas en los casos de viruela del mono.

3. RMSE (Root Mean Squared Error):

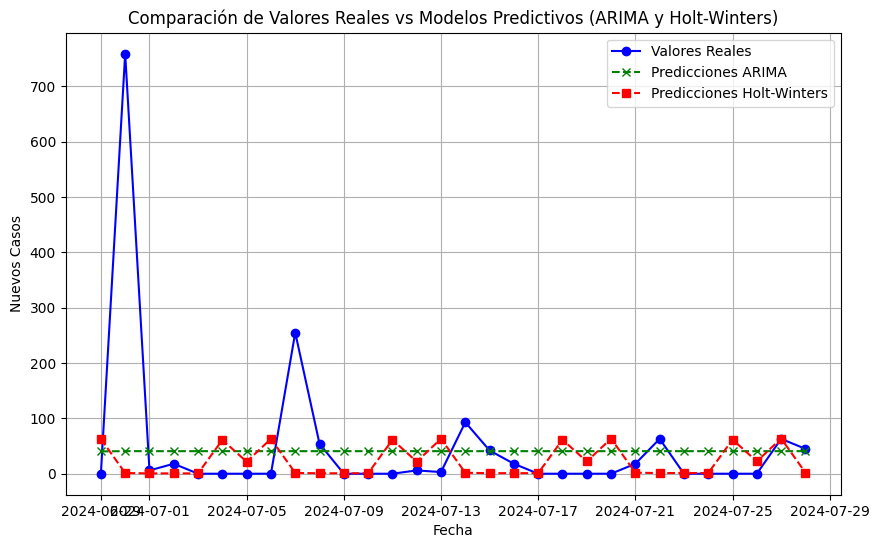
* ARIMA: 143.75
* Holt-Winters: 121.71

El RMSE penaliza más los errores grandes, ya que eleva al cuadrado las diferencias entre los valores reales y las predicciones. Aquí, Holt-Winters (121.71) nuevamente supera a ARIMA (143.75), lo que indica que tiene un rendimiento ligeramente mejor en términos de reducir los errores grandes.

Resumen Comparativo:

En general, el modelo Holt-Winters ha mostrado un mejor rendimiento en las tres métricas de evaluación:

* Menor MAD, lo que sugiere que el error absoluto promedio es menor.
* Menor MAPE, lo que indica que el modelo es más preciso en términos relativos, aunque aún con margen de mejora.
* Menor RMSE, lo que indica que el modelo maneja mejor los errores grandes en comparación con ARIMA.



Gráfica 5: Comparación de valores reales vs Modelos predictivos

Holt-Winters es el más adecuado de los dos para predecir los casos diarios de viruela del mono, según las métricas evaluadas. No obstante, se observan altos errores porcentuales en ambos modelos, lo que sugiere que se podrían explorar otros modelos, como SARIMA o modelos que incluyan variables exógenas, para mejorar aún más la precisión de las predicciones.

# **CONCLUSIONES**

En este proyecto, se analizaron los casos globales de viruela del mono utilizando dos modelos predictivos principales: ARIMA y Holt-Winters. El modelo ARIMA mostró un mejor rendimiento general en la predicción de los casos futuros, mientras que el modelo Holt-Winters capturó de manera más eficiente ciertos patrones de estacionalidad, aunque no tan precisa en las predicciones a largo plazo.

Para futuras mejoras, se podrían considerar:

* Ajustes adicionales en los componentes estacionales de los modelos.
* Implementar otros modelos como SARIMA para capturar estacionalidad y variabilidad de manera más precisa.
* Incluir variables exógenas que puedan influir en los casos, como políticas de salud pública.

# **REFERENCIAS**

*Monkeypox Global Case Records*. (2024, 21 agosto). Kaggle. https://www.kaggle.com/datasets/rajatkumar30/monkeypox/data

Bajaj, A. (2023, 18 agosto). *ARIMA & SARIMA: Real-World Time Series Forecasting*. neptune.ai. https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide