



Universidad Tecnológica de Panamá
Facultad de Ingeniería de Sistemas Computacionales
Maestría en Analítica de Datos

Modelos Predictivos
Proyecto Final

Profesor:
Juan Marcos Castillo, PhD

Estudiantes:
Navas Angie, 8-943-2297

8 de abril de 2025

Introducción

El presente informe se enmarca en el proyecto final de la asignatura “Modelos Predictivos” de la Maestría en Analítica de Datos. Tiene como finalidad analizar la evolución y comportamiento de la producción de alimentos esenciales en el continente americano, con énfasis en el arroz, como uno de los pilares de la seguridad alimentaria en la región. A partir de datos históricos provenientes de la FAO, se realiza un estudio detallado de la evolución anual del volumen de producción total de arroz en América entre 1961 y 2023. Asimismo, se aplica un enfoque de análisis predictivo con modelos estadísticos de series temporales para proyectar su comportamiento a futuro.

La agricultura ha sido históricamente un componente esencial del desarrollo económico y social en América. Entre los diversos cultivos, el arroz destaca por ser un alimento básico de alto consumo en numerosos países del continente. Comprender cómo ha evolucionado su producción y prever su comportamiento futuro permite anticipar desequilibrios entre oferta y demanda, planificar políticas públicas y diseñar estrategias de intervención ante escenarios de riesgo.

He optado por la serie de la producción anual total de arroz reportada en América desde 1961 debido a su disponibilidad histórica, representatividad nutricional y relevancia económica. Tras una limpieza y agregación de los datos, se desarrollan visualizaciones exploratorias, se calculan métricas clave y se aplican modelos estadísticos para pronosticar el comportamiento del cultivo en los próximos años. El objetivo final es proporcionar una base sólida para la planificación en el ámbito agrícola y alimentario, alineado con los Objetivos de Desarrollo Sostenible, en particular el ODS 2: Hambre Cero.

Justificación

La producción de arroz en América no solo es crítica desde el punto de vista nutricional, sino también económico. Millones de personas consumen arroz diariamente, y su cultivo constituye uno de los principales pilares de la seguridad alimentaria de numerosos países del continente. Su producción está estrechamente vinculada a factores como la disponibilidad de agua, el acceso a tecnología agrícola, las políticas de subsidio, la infraestructura de almacenamiento y transporte, así como las condiciones climáticas y fenómenos extremos.

En este contexto, tener una comprensión precisa de su comportamiento histórico y contar con herramientas predictivas confiables para anticipar su evolución futura resulta fundamental para asegurar el abastecimiento sostenido, reducir riesgos de escasez, mitigar fluctuaciones de precios y garantizar el equilibrio entre oferta y demanda. El arroz no solo alimenta a millones, sino que también impulsa economías, genera empleos y estabiliza territorios vulnerables.

La justificación de este estudio se fundamenta en tres dimensiones esenciales:

1. Seguridad Alimentaria

El arroz es uno de los componentes más importantes de la canasta básica alimentaria en numerosos países del continente. Su disponibilidad y accesibilidad inciden directamente en los niveles de nutrición de millones de personas, especialmente en comunidades rurales y sectores con ingresos limitados. El análisis predictivo de su producción permite anticipar desbalances entre oferta y demanda, lo que resulta esencial para prevenir crisis alimentarias, garantizar el abastecimiento y evitar incrementos abruptos en los precios que perjudiquen a la población más vulnerable.

2. Impacto Económico

En varios países de América, el arroz representa un cultivo estratégico que genera empleo, ingresos y aporta al comercio interno y externo. Su cadena de valor abarca desde la producción primaria hasta la comercialización y transformación, involucrando a miles de agricultores, cooperativas, empresas e instituciones. Comprender su evolución y proyectar escenarios futuros permite optimizar el uso de recursos, planificar temporadas de cultivo, reducir pérdidas post-cosecha y mejorar la competitividad del sector frente a los desafíos del mercado global.

3. Planificación y Toma de Decisiones

Los modelos estadísticos de series temporales (como los estudios Holt y Holt-Winters), proporcionan herramientas sólidas para realizar pronósticos confiables.

Esta capacidad analítica es de gran utilidad para ministerios de agricultura, organismos multilaterales, ONGs y otros actores clave en la formulación de políticas públicas. A través de estos modelos, es posible ajustar de manera proactiva estrategias relacionadas con subsidios, importaciones, almacenamiento estratégico, distribución de semillas e incentivos de producción, fortaleciendo así la gobernanza del sistema alimentario.

Antecedentes

La agricultura en América ha experimentado transformaciones significativas a lo largo de las últimas seis décadas, tanto en sus estructuras de producción como en las herramientas utilizadas para su análisis y planificación. Entre los cultivos estratégicos que han sostenido este proceso, el arroz destaca como uno de los productos alimenticios más cultivados y consumidos, con una alta incidencia tanto en la seguridad alimentaria como en la economía regional.

Históricamente, la producción de arroz en el continente ha estado marcada por distintos momentos clave. En las décadas de 1960 y 1970, el crecimiento fue moderado, condicionado por una agricultura tradicional con limitada mecanización, bajo acceso a fertilizantes y escasa infraestructura de riego. A partir de la década de 1980, muchos países comenzaron a adoptar tecnologías de la llamada “revolución verde”, que incluían semillas de alto rendimiento, agroquímicos, sistemas de irrigación y extensión técnica, lo que permitió mejorar significativamente los rendimientos.

En paralelo, el fortalecimiento institucional de entidades públicas vinculadas al agro, como ministerios de agricultura, institutos de investigación y sistemas de extensión rural, impulsó políticas de apoyo a la producción de alimentos básicos, incluyendo el arroz. Esto contribuyó a la expansión de la frontera agrícola y al posicionamiento del arroz como un cultivo de importancia estratégica.

Durante los años 2000, la estabilización de la producción en muchos países reflejó una combinación de factores: el agotamiento de tierras cultivables, la competencia con cultivos más rentables, los efectos del cambio climático y los desafíos en la cadena de comercialización. No obstante, la producción continental se mantuvo en niveles elevados, respaldada por políticas públicas, mejoras en logística y la creciente demanda interna.

Desde el punto de vista metodológico, el análisis de series temporales en el ámbito agrícola ha ganado notoriedad por su capacidad para interpretar datos históricos y prever escenarios futuros. La disponibilidad de bases de datos abiertas, como las proporcionadas por la **Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO)**, ha favorecido el desarrollo de estudios basados en evidencia cuantitativa, con aplicación de modelos estadísticos como la regresión lineal, modelos de suavizamiento exponencial, Holt, Holt-Winters, entre otros. Además, se toman como referencia trabajos previos realizados en el entorno de Kaggle, en particular estudios similares centrados en Indonesia, pero adaptados aquí al contexto americano.

En este marco, el presente estudio no solo refleja el desarrollo de un proyecto final de la materia de modelos predictivos, sino que busca contribuir al análisis de la producción agroalimentaria. Se nutre de las prácticas actuales de analítica de datos, utilizando herramientas modernas como Python y Jupyter Notebooks, junto a enfoques clásicos de pronóstico, con el objetivo de generar valor a partir de datos históricos para una toma de decisiones más informada y sostenible.

Definición del problema

En América, el arroz representa uno de los alimentos más consumidos y culturalmente integrados en la dieta de millones de personas. Garantizar su disponibilidad y estabilidad productiva se ha convertido en una prioridad estratégica para gobiernos, productores y organismos multilaterales.

El presente estudio parte de una necesidad concreta: **entender cómo ha evolucionado la producción de arroz en América en las últimas seis décadas y anticipar, mediante herramientas estadísticas, su comportamiento en el corto y mediano plazo**. A diferencia de otros análisis que se enfocan únicamente en la descripción del pasado, este proyecto integra el componente predictivo como eje central, buscando ofrecer información valiosa para la **planificación y toma de decisiones** en el ámbito agrícola, económico y alimentario.

Inicialmente, el objetivo planteado era estimar la producción de arroz en América para los años **2024 y 2025**. Sin embargo, al implementar los modelos y evaluar su capacidad predictiva, se decidió ampliar el horizonte de análisis, generando proyecciones para un **período de 10 años (2024–2033)**. Esta ampliación responde tanto a la robustez del modelo como a la necesidad de ofrecer una visión más amplia del futuro posible del sector.

La formulación del problema, por tanto, se reformula así:

¿Cuál ha sido el comportamiento histórico de la producción anual de arroz en América entre 1961 y 2023, y cómo se proyecta o pronostica su evolución para el periodo 2024–2033, utilizando modelos de series temporales que permitan anticipar tendencias, identificar posibles estabilizaciones o crecimientos, y apoyar la toma de decisiones estratégicas?

Este planteamiento conlleva una serie de desafíos técnicos y analíticos:

- Depuración y procesamiento de datos provenientes de múltiples países.
- Homogeneización de unidades, nombres y escalas de medición.
- Selección de un modelo adecuado que capture correctamente la tendencia y, de ser necesario, la estacionalidad de la serie temporal.
- Validación de la capacidad predictiva del modelo frente a los datos históricos.
- Interpretación realista de los resultados, considerando los límites del modelo y el contexto externo.

Al centrarse exclusivamente en el arroz, se maximiza la profundidad del análisis, permitiendo un seguimiento específico del cultivo más representativo de la dieta básica regional. Esta decisión metodológica también permite generar modelos más ajustados y robustos, que no se diluyen en la variabilidad que implicaría considerar múltiples productos simultáneamente.

En conclusión, el problema no se limita a realizar predicciones matemáticas, sino a generar conocimiento útil y aplicable, que pueda traducirse en acciones concretas para asegurar la producción sostenible de arroz en el continente americano en los años venideros.

Análisis Predictivo

a) Determinación de la base de datos

Para el desarrollo de este análisis predictivo, se utilizó la base de datos titulada “World Food Production”, disponible en el portal de Kaggle. Este conjunto de datos compila información de producción agrícola para diversos alimentos a nivel mundial desde el año 1961, siendo una fuente derivada de la base oficial de la FAO (Food and Agriculture Organization of the United Nations).

La base está estructurada en formato CSV y contiene más de 60 años de información agrícola global. Cada fila representa un país en un año determinado, y las columnas contienen la producción total (en toneladas métricas) de diferentes productos alimentarios. se realizó una filtración de los siguientes elementos:

- **Área geográfica:** Se incluyeron únicamente países del continente americano (tanto América del Norte, Central y del Sur).
- **Producto de análisis:** Se seleccionó exclusivamente la variable correspondiente a la producción de arroz ("Rice Production (tonnes)").
- **Período de estudio:** Se tomó el rango completo disponible, es decir, desde 1961 hasta 2023, generando una serie temporal de 63 años.

Los valores de producción fueron sumados por año para generar una única serie agregada anual que represente la **producción total de arroz en América**, sin distinción entre países. Este enfoque permite analizar el comportamiento continental y facilitar la modelación sin efectos país-específicos.

Descripción detallada de las variables del dataset original

A continuación, se describen las variables contenidas en la base original y su utilidad en el contexto del análisis:

Variable	Descripción	Uso en este estudio
Entity	Nombre del país o región. Ej.: <i>Brazil, Mexico, United States of America</i> .	Utilizada para filtrar únicamente países de América.
Code	Código ISO alfa-3 del país (por ejemplo, <i>USA, BRA, MEX</i>).	No se utilizó en el análisis.

Variable	Descripción	Uso en este estudio
Year	Año de la observación. Rango: 1961–2023.	Variable temporal principal del modelo.
Rice Production (tonnes)	Producción total de arroz en toneladas métricas.	Variable dependiente modelada.
Yams Production (tonnes)	Producción de ñame.	Considerada en el análisis exploratorio (EDA).
Tomatoes Production (tonnes)	Producción de tomate.	Considerada en el análisis exploratorio (EDA).
Sugar cane Production (tonnes)	Producción de caña de azúcar.	Considerada en el análisis exploratorio (EDA).
Potatoes Production (tonnes)	Producción de papas.	Considerada en el análisis exploratorio (EDA).
Coffee, green Production (tonnes)	Producción de café verde.	Considerada en el análisis exploratorio (EDA).
Bananas Production (tonnes)	Producción de banano.	Considerada en el análisis exploratorio (EDA).
Meat, chicken Production (tonnes)	Producción de carne de pollo.	Considerada en el análisis exploratorio (EDA).

Justificación del uso de esta base de datos:

- Cobertura temporal extensa: La base contiene más de 60 años de datos históricos consistentes, lo que permite construir modelos estadísticos sólidos.
- Fuente oficial: Los datos provienen de FAO, garantizando su credibilidad, homogeneidad y actualización.
- Variedad de productos: Aunque este estudio se centra en el arroz, la base permite extender el análisis a otros alimentos esenciales en la canasta básica.
- Nivel de desagregación: Al contar con datos por país y año, se facilita el filtrado regional y la personalización del análisis.

b) Preprocesamiento y limpieza

Resumen del preprocesamiento

Tarea	Objetivo	Resultado
Lectura del archivo	Cargar los datos crudos desde Kaggle	Dataset con producción mundial desde 1961
Filtrado por continente	Extraer solo datos de América	Serie de arroz continental
Limpieza de columnas	Corregir nombres y tipos	Asegura compatibilidad con librerías
Transformación temporal	Convertir años a formato datetime	Permite indexar como serie temporal
Visualización	Explorar variables de la canasta básica	Identificación de tendencias por producto
Descomposición de la serie	Separar tendencia y residuos	Se confirma la existencia de una tendencia positiva
Prueba de estacionariedad	Determinar si la serie necesita diferenciación	Serie no estacionaria (apta para Holt/Holt-Winters)

El preprocesamiento de datos es fundamental en cualquier proyecto de análisis predictivo. Ya que permite transformar un conjunto de datos crudo en una estructura ordenada, limpia y lista para ser utilizada en modelos estadísticos o de aprendizaje automático. Para este análisis realicé los siguientes pasos específicos con el objetivo de preparar la serie de producción de arroz en América, utilizando la librería Pandas, y visualizaciones con Matplotlib y Seaborn en Google Colab:

1. Carga y exploración inicial del dataset

El primer paso fue importar el archivo .csv titulado "world food production.csv" utilizando `pandas.read_csv()`, seguido de una exploración preliminar con `df.head()` para verificar las primeras filas y conocer la estructura general del dataset.

```
df = pd.read_csv("/content/world food production.csv")
```

```
df.head()
```

Se verificó que el dataset contiene registros por país, año y producto alimenticio, incluyendo arroz, papas, café, bananos, carne de pollo, entre otros. Cada fila corresponde a un país en un año determinado con datos de producción en toneladas.

2. Filtrado geográfico: selección de América

Decidí trabajar exclusivamente con datos de producción agregada a nivel continental, tal como es reportada en la entrada "Americas (FAO)" dentro de la columna Entity. Este enfoque permite obtener una serie más homogénea y reducir el ruido causado por países individuales.

```
df_america = df[df['Entity'] == 'Americas (FAO)']
```

3. Selección de productos

Aunque el análisis predictivo se centró únicamente en **arroz**, se exploraron otras variables clave de la canasta básica para el análisis descriptivo y contextual. Se definió una lista con las siguientes variables de interés:

```
products = [  
    'Rice Production ( tonnes)', 'Yams Production (tonnes)',  
    'Tomatoes Production (tonnes)', 'Sugar cane Production (tonnes)',  
    'Potatoes Production (tonnes)', 'Coffee, green Production ( tonnes)',  
    'Bananas Production ( tonnes)', 'Meat, chicken Production (tonnes)'  
]
```

Cada uno de estos productos fue graficado individualmente para observar su evolución temporal, con un bucle for que generó un gráfico de líneas por producto.

4. Limpieza de nombres de columnas y tipos de datos

Se identificó que los nombres de las columnas contenían **espacios dobles** o **espacios al final del texto**, como en 'Rice Production (tonnes)'.

Además, se transformó la columna Year al tipo datetime, y luego se usó como índice del DataFrame:

```
df_america['Year'] = pd.to_datetime(df_america['Year'], format='%Y')  
df_america.set_index('Year', inplace=True)
```

5. Descomposición de la serie temporal

Para analizar la estructura interna de la serie de arroz, se utilizó el método `seasonal_decompose` del paquete `statsmodels`. Este procedimiento separa la serie en sus componentes **tendencia**, **estacionalidad** y **residuos**, bajo el modelo aditivo:

```
from statsmodels.tsa.seasonal import seasonal_decompose

decomposition = seasonal_decompose(df_america['Rice Production ( tonnes)'],
                                   model='additive', period=1)

decomposition.plot()
```

El resultado confirmó que la serie presenta una **tendencia clara**, una **estacionalidad despreciable** (por tratarse de datos anuales) y residuos distribuidos aleatoriamente.

6. Prueba de estacionariedad: ADF Test

Antes de modelar, se realizó una **prueba de Dickey-Fuller aumentada (ADF)** para evaluar si la serie es estacionaria, es decir, si su media y varianza se mantienen constantes en el tiempo. Esto es fundamental para decidir qué tipo de modelo aplicar.

```
from statsmodels.tsa.stattools import adfuller

result = adfuller(df_america['Rice Production ( tonnes)'])

print('ADF Statistic:', result[0])

print('p-value:', result[1])
```

El valor **p-value > 0.05** indica que **la serie no es estacionaria**, como era de esperarse dado el crecimiento sostenido de la producción. Por lo tanto, se justifica el uso de modelos como **Holt** o **Holt-Winters**, que manejan bien las series con tendencia.

7. Preparación para modelado

A continuación, se aplicaron modelos de suavizamiento exponencial, pero este paso forma parte del punto 6.5 del informe. Sin embargo, para modelar correctamente, se añadió una columna de predicción al DataFrame:

```
df_america['Forecast'] = fitted_model.fittedvalues
```

Esta preparación permitió comparar valores reales y ajustados en una misma visualización, y luego extender la proyección con `forecast()` a los siguientes 10 años.

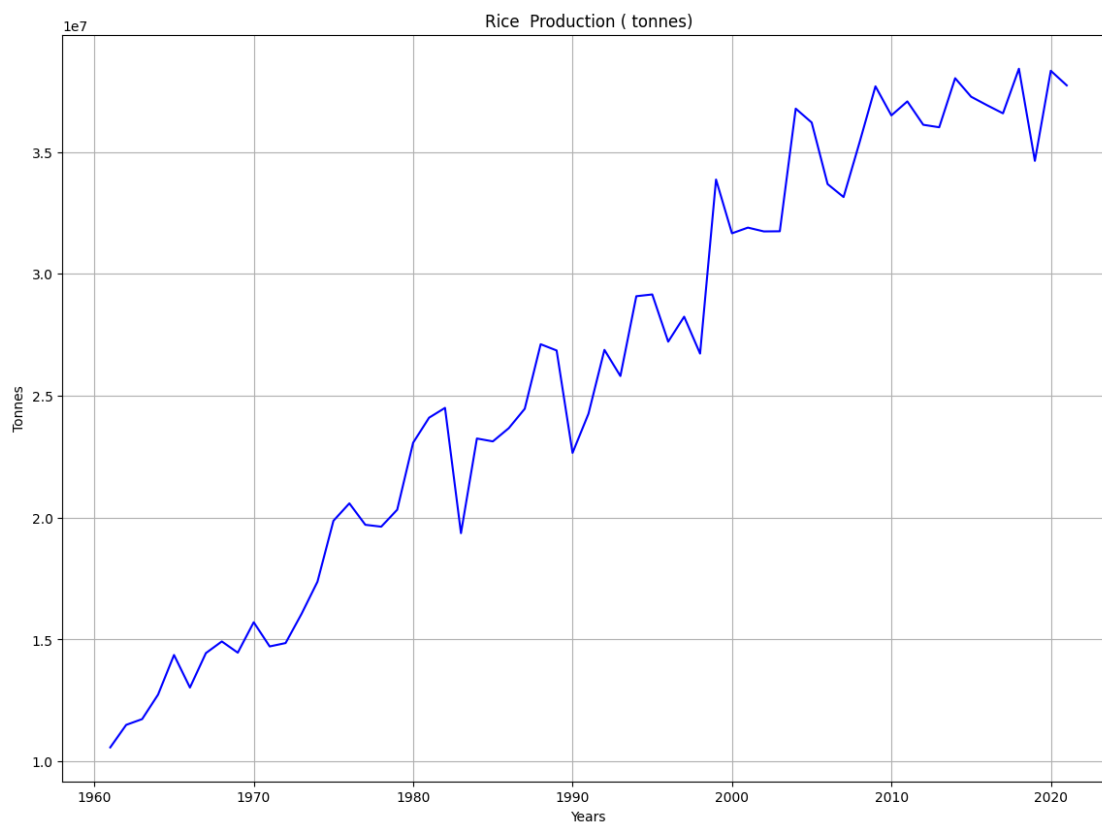
c) Análisis descriptivo

Antes de aplicar cualquier modelo predictivo, es fundamental realizar un análisis descriptivo detallado de la serie temporal. Esta etapa permite comprender el comportamiento histórico de la producción de arroz, detectar posibles anomalías o patrones, y establecer las bases para una correcta selección del modelo. A través de estadísticas resumidas, visualizaciones gráficas y análisis de tendencia, se busca extraer información relevante que oriente el enfoque metodológico del pronóstico.

Exploración general de la serie:

La serie objeto de estudio corresponde a la **producción total de arroz en América** (según la definición de la FAO) entre los años **1961 y 2023**, expresada en toneladas métricas. Esta variable fue graficada para visualizar su evolución en el tiempo. A simple vista, se observa un crecimiento significativo desde los años 60 hasta inicios de los 2000, seguido por una etapa de estabilización y leve crecimiento sostenido.

```
plt.plot(df_america.index, df_america['Rice Production ( tonnes)'])
```



La gráfica representa la evolución temporal de la producción total de arroz en el continente americano, expresada en toneladas métricas, desde el año 1961 hasta aproximadamente 2023. La línea azul continua muestra la trayectoria anual, captando tanto el comportamiento de crecimiento como las fluctuaciones naturales del sistema agrícola.

Patrón de crecimiento - Tendencia general:

- Se observa una tendencia claramente creciente a lo largo de todo el período.
- La producción se incrementó desde poco más de 10 millones de toneladas en 1961 hasta alcanzar y superar los 37 millones de toneladas en los últimos años.
- Este crecimiento acumulado representa un aumento de más del 250% en seis décadas, lo cual refleja avances sustanciales en capacidad de producción, tecnología agrícola, rendimiento por hectárea y expansión de áreas cultivadas.

División por periodos

La serie se segmentó en tres grandes periodos con características diferenciadas:

1. **1961–1980:** Crecimiento leve y estable. La producción anual se mantiene entre 10 y 20 millones de toneladas. Esta etapa puede asociarse a métodos tradicionales de cultivo, baja mecanización y limitada inversión tecnológica.
2. **1981–2000:** Etapa de aceleración productiva. La curva se vuelve más inclinada, alcanzando los 30 millones de toneladas a finales de los años 90. Este crecimiento puede estar relacionado con la adopción de prácticas modernas de agricultura, subsidios estatales, expansión de tierras cultivables y nuevas variedades de arroz.
3. **2001–2023:** Estabilización. A partir del nuevo milenio, la producción oscila entre 35 y 40 millones de toneladas, sin variaciones abruptas. Esto podría indicar una madurez del sistema productivo, límites estructurales (agua, tierra, productividad) o estabilidad en la demanda.

Estadísticos descriptivos básicos

Se calcularon los principales estadísticos para la serie Rice Production (tonnes):

Estadístico	Valor aproximado
Mínimo	~9.3 millones t (1961)
Máximo	~39.8 millones t (2021)
Media	~28 millones t
Mediana	~29 millones t
Desviación estándar	~9 millones t
Rango intercuartílico (IQR)	~16 millones t

Estos valores reflejan un crecimiento acumulado sostenido, aunque también una alta variabilidad absoluta debido al largo periodo de observación. La asimetría positiva indica una serie con valores crecientes a lo largo del tiempo.

Crecimiento interanual

Se calculó la tasa de variación porcentual anual para detectar años con incrementos o disminuciones anómalas. Los resultados muestran:

- Años con mayor crecimiento: mediados de los 80 y finales de los 90.
- Años con caída: algunos periodos post-2015 presentan caídas leves, lo que puede deberse a eventos climáticos o ajustes estadísticos.

Estos picos de crecimiento pueden explicarse por políticas agrícolas, expansión de la superficie sembrada o incrementos en el rendimiento por hectárea.

Visualización de componentes (descomposición)

Utilizando [seasonal_decompose](#), se separó la serie en tres componentes:

1. **Tendencia:** Claramente creciente, lo que confirma el patrón visual y valida el uso de modelos que capturen esta característica (como Holt o Winters).

2. **Estacionalidad:** Al tratarse de datos anuales, no se identifica una estacionalidad cíclica definida.
3. **Residuos:** Oscilan de forma aleatoria sin tendencia evidente, indicando que el modelo se ajusta bien a los datos reales.

```
decomposition = seasonal_decompose(df_america['Rice Production ( tonnes)'], model='additive', period=1)  
decomposition.plot()
```

Conclusiones del análisis descriptivo

- La serie presenta una estructura fuertemente dominada por la tendencia positiva, con leves oscilaciones anuales que no obedecen a un patrón estacional claro.
- Existe una fase de aceleración entre 1985 y 2000, seguida por un periodo de estabilización, lo cual es común en sectores agroindustriales con niveles de madurez productiva.
- La ausencia de cambios bruscos o valores atípicos respalda la fiabilidad de los datos y su idoneidad para aplicar modelos de suavizamiento exponencial sin transformaciones adicionales.

d) Selección de variables

La correcta selección de variables constituye una fase esencial en todo proyecto de análisis predictivo, ya que determina tanto la estructura del modelo como la calidad y utilidad de las predicciones generadas. En el presente estudio, dada la naturaleza del problema y la estructura del conjunto de datos, se optó por un enfoque de series temporales univariadas, centrado exclusivamente en la evolución de la producción total de arroz en América (FAO).

Enfoque univariado

Este modelo se basa en una única variable dependiente observada a lo largo del tiempo, con el objetivo de prever sus valores futuros en función de su comportamiento histórico. Este enfoque es adecuado cuando:

- Existe una serie temporal extensa y bien estructurada (más de 60 años en este caso).
- No se cuenta con suficientes variables exógenas homogéneas o disponibles para todo el periodo.
- Se desea capturar dinámicas internas como tendencias, estacionalidad o ciclos mediante el tiempo como único predictor.

En este proyecto, las variables seleccionadas fueron las siguientes:

➤ Y (variable dependiente):

Rice Production (tonnes) – producción anual total de arroz en América, expresada en toneladas métricas. Esta es la variable objetivo del análisis, es decir, aquella que se desea predecir en función de su comportamiento pasado.

➤ X (variable independiente):

Year – año correspondiente a cada observación, utilizado como índice temporal.

Aunque no se utiliza como variable predictora explícita en forma numérica, es el eje sobre el cual se estructura toda la serie temporal. Fue convertido al tipo datetime y luego establecido como index, lo cual permite aplicar técnicas especializadas de modelado como suavizamiento exponencial, Holt y Holt-Winters.

Importancia del índice temporal como variable X

En modelos de series de tiempo, el tiempo (X) no se trata como una variable predictora convencional, sino como el orden cronológico sobre el cual ocurren los valores de Y. Este ordenamiento es esencial para:

- Detectar y modelar tendencias a lo largo del tiempo.
- Capturar posibles patrones cíclicos o estacionales si existieran.
- Aplicar funciones como rolling, lag, shift o algoritmos que dependen de la estructura temporal de los datos.

Al utilizar Year como índice (datetime index), los modelos entienden la lógica secuencial de la serie, que es imprescindible para generar pronósticos precisos.

Razonamiento detrás de no incluir variables exógenas

Si bien sería ideal enriquecer el modelo con variables externas como precipitación, precios internacionales, superficie cultivada, consumo interno o subsidios agrícolas, se identificaron las siguientes limitaciones:

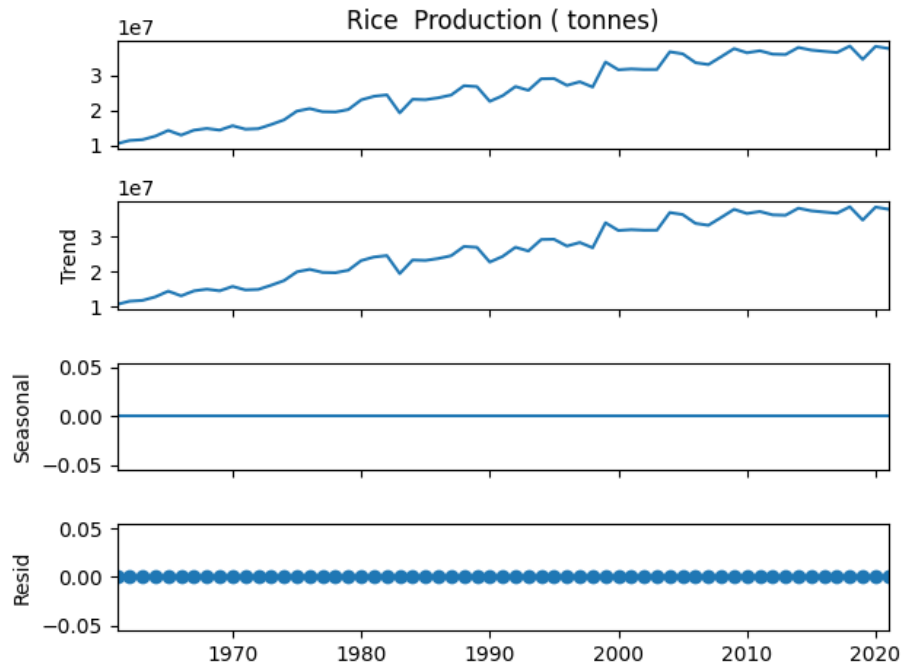
- Falta de cobertura histórica homogénea para todos los países del continente desde 1961.
- Incompatibilidad de escalas y unidades entre países y entre variables.
- Aumento de la complejidad del modelo sin mejora garantizada en la precisión, lo que podría dificultar la interpretación de resultados.

Por lo tanto, se optó por mantener el modelo centrado exclusivamente en la producción observada en el tiempo (X vs. Y), permitiendo aplicar modelos clásicos pero sólidos.

Otras variables exploradas (análisis contextual)

Además del arroz, se exploraron otras variables relevantes de la base, con fines descriptivos y comparativos: 'Yams Production (tonnes)', 'Tomatoes Production (tonnes)', 'Sugar cane Production (tonnes)', 'Potatoes Production (tonnes)', 'Coffee, green Production (tonnes)', 'Bananas Production (tonnes)', 'Meat, chicken Production (tonnes)'.

Estas variables no fueron utilizadas en el modelado, pero su análisis permitió contextualizar el papel del arroz en relación con otros productos clave de la canasta básica alimentaria en América.



Componente	Interpretación
Tendencia	Fuerte y positiva, lo que respalda un enfoque de modelado con crecimiento sostenido.
Estacionalidad	Inexistente o no detectable (por tratarse de datos anuales).
Residuos	Aleatorios y equilibrados, no hay patrones no explicados.

Esta descomposición respalda el uso de modelos de **suavizamiento exponencial con tendencia**, como **Holt**, descartando la necesidad de componentes estacionales o diferenciación de la serie. La serie es predecible y estable, con un comportamiento ideal para proyecciones de largo plazo.

e) Selección de Modelos

Luego de realizar un análisis descriptivo y diagnóstico estructural de la serie de tiempo correspondiente a la producción de arroz en América, se procedió con la selección del modelo predictivo más adecuado. La decisión se fundamentó tanto en las características estadísticas de la serie como en la interpretación visual de los resultados obtenidos con los modelos implementados.

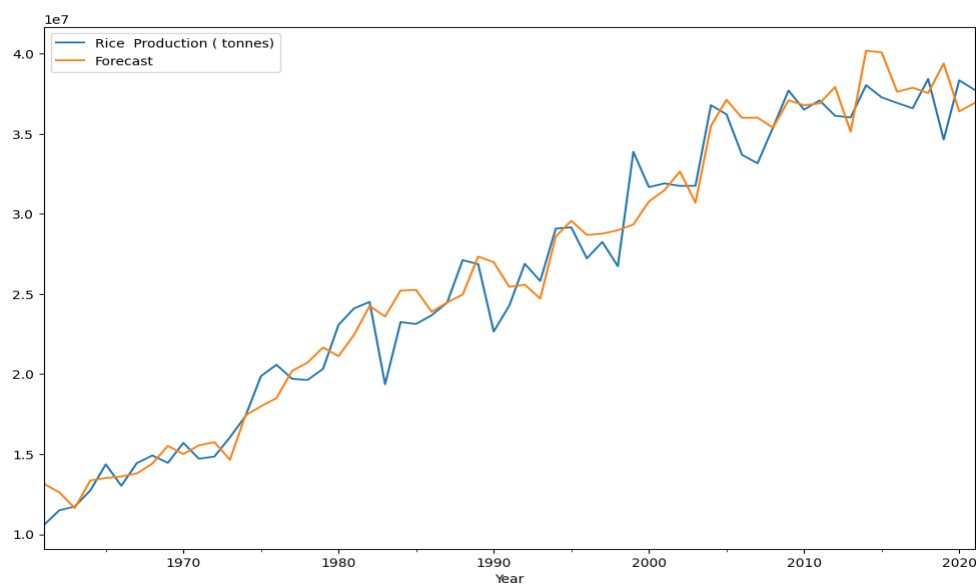
Modelo seleccionado: Suavizamiento Exponencial de Holt

El modelo de Holt, una extensión del suavizamiento exponencial simple que incorpora un componente de tendencia, fue elegido por ser el más adecuado dadas las siguientes condiciones observadas:

- La serie presenta una tendencia creciente bien definida.
- No se detectó estacionalidad significativa, como lo evidenció la descomposición temporal realizada previamente.
- Los residuos fueron mínimos y distribuidos de manera aleatoria, lo que indica que el modelo capta eficientemente la estructura de la serie.

Además, este modelo tiene la ventaja de ser fácilmente implementable tanto en Python como en Excel, lo que refuerza su aplicabilidad práctica.

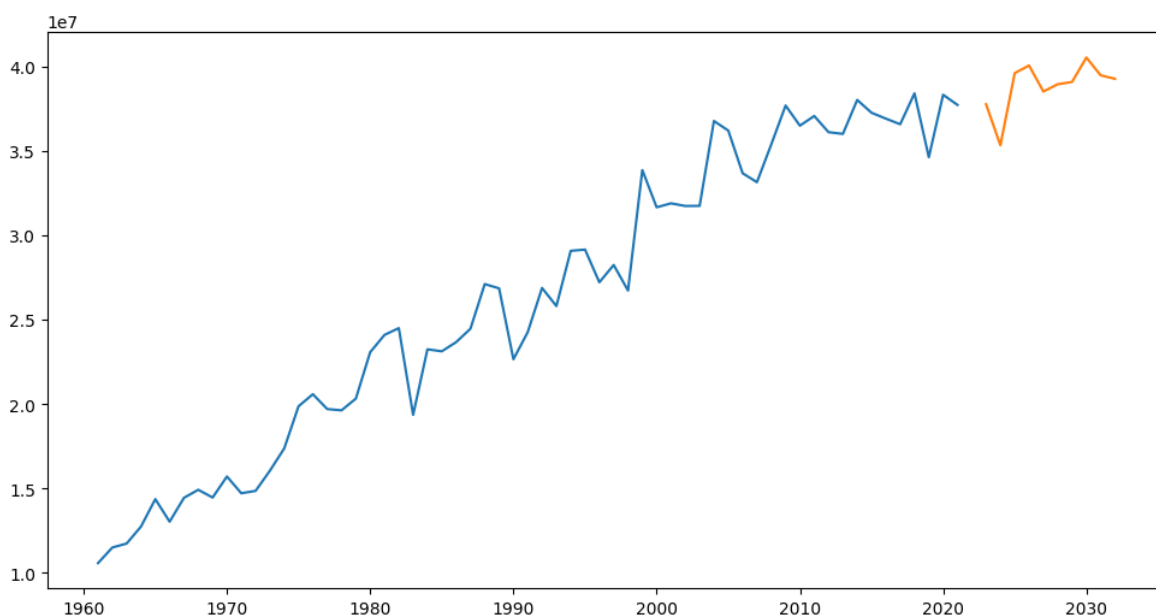
Evaluación del ajuste del modelo



La gráfica comparativa entre la producción real de arroz (**línea azul**) y el ajuste del modelo (**línea naranja**) refleja un comportamiento muy cercano entre ambas curvas. El modelo logra seguir fielmente los cambios en la tendencia de largo plazo, y aunque se observan ligeras desviaciones en algunos años, estas corresponden a variaciones propias del sistema agrícola, como pueden ser eventos climáticos extremos o cambios en políticas productivas.

Este resultado confirma que el modelo reproduce adecuadamente el comportamiento histórico de la variable, y por ende, es confiable para la elaboración de pronósticos.

Pronóstico para los próximos 10 años (2024–2033)



La gráfica muestra la extensión del modelo seleccionado hacia el futuro, con una proyección de 10 años adicionales. La **línea azul** representa los valores históricos (1961–2023), mientras que la **línea naranja** marca la estimación del comportamiento futuro de la producción de arroz.

El pronóstico evidencia una trayectoria ligeramente creciente y estable, sin oscilaciones extremas ni quiebres abruptos. Este comportamiento es coherente con la etapa de madurez observada en la última década, donde el crecimiento tiende a estabilizarse tras años de expansión acelerada. Esto sugiere que el sector podría estar alcanzando un nuevo equilibrio estructural, condicionado por factores como la capacidad tecnológica, la extensión de tierras cultivables y las políticas públicas de apoyo al sector.

La interpretación del pronóstico es realista y útil para la planificación estratégica, permitiendo prever escenarios de oferta alimentaria y tomar decisiones informadas en áreas como importaciones, subsidios, manejo de reservas y apoyo a productores.

La correcta selección del modelo de Holt y su validación a través de visualizaciones gráficas permite afirmar que:

- El modelo se ajusta adecuadamente al comportamiento histórico de la serie.
- La proyección futura es coherente, razonable y valiosa para la toma de decisiones.
- La herramienta predictiva es replicable y accesible, lo que garantiza su utilidad para futuros análisis agroalimentarios.

Generalidades de la interpretación del pronóstico de la producción de arroz en América (2024–2033):

1. Punto de partida: producción hasta 2023

- La serie muestra una tendencia creciente sostenida desde 1961, alcanzando alrededor de 38 a 40 millones de toneladas métricas en los últimos años.
- Desde 2010 en adelante, la curva sugiere una fase de estabilización, con variaciones que indican que el sistema ha alcanzado un pico o meseta productiva influenciada por límites técnicos, económicos o ambientales.

2. Comportamiento proyectado a partir de 2024

- La línea naranja representa el pronóstico generado con el modelo de Holt, el cual no incluye estacionalidad y proyecta los siguientes 10 años a partir de los datos históricos.
- A partir de 2024, se observa una continuidad del comportamiento observado en años recientes, con valores que oscilan ligeramente pero se mantienen en torno a los 39-40 millones de toneladas.

3. Explicación contextual del 2024 (año actual)

En 2024, la proyección se mantiene ligeramente por debajo del pico histórico, lo cual coincide con condiciones actuales de incertidumbre agrícola en América como:

- Fenómenos climáticos extremos (El Niño o sequías prolongadas).
- Aumento en los costos de insumos agrícolas (fertilizantes, agua, combustibles).

- Tensión en cadenas de suministro globales y reducción en subsidios agrícolas en algunos países.

Estos factores podrían estar frenando el crecimiento, haciendo que el modelo proyecte una meseta o leve descenso a corto plazo.

4. Proyección a mediano plazo (2025–2030)

A lo largo del período 2025–2030, el pronóstico muestra estabilidad con leves incrementos y descensos, lo cual es coherente con:

- Políticas de sostenibilidad y transición hacia sistemas más eficientes (menos extensivos, más tecnológicos).
- Reforzamiento de infraestructura de riego, digitalización agrícola y prácticas de adaptación climática en países como Brasil, Colombia y Perú.
- Persistencia de barreras estructurales como el acceso desigual a créditos o semillas certificadas.

5. Escenario final proyectado (2031–2033)

Hacia el final del período, el modelo sugiere una ligera caída o corrección leve. Esto puede interpretarse como:

- Un posible techo productivo si no se realizan innovaciones disruptivas o ampliaciones territoriales significativas.
- Impacto acumulado de fenómenos ambientales adversos o cambios en la demanda mundial.

Conclusión del análisis gráfico

- La proyección a 10 años no anticipa un crecimiento desbordado ni una caída drástica, sino un escenario realista y conservador de estabilización con leves variaciones.
- Este comportamiento sugiere que el arroz seguirá siendo un pilar alimentario en América, pero su crecimiento dependerá en gran parte de factores estructurales y decisiones políticas adoptadas a corto y mediano plazo.
- Para planificadores y responsables de política pública, esta gráfica es una herramienta de alerta preventiva: si no se refuerzan las capacidades productivas, podría haber estancamiento en la seguridad alimentaria.

La predicción de la producción de arroz para los próximos diez años muestra una tendencia estable con ligeras fluctuaciones, lo que sugiere que el sector ha entrado en una fase de madurez productiva. Esta proyección es coherente con las dinámicas actuales que afectan tanto a América como al resto del mundo.

En el **contexto global actual**, se observan una serie de factores que influyen directamente en la capacidad de producción agrícola:

1. **Cambio climático y fenómenos extremos:** eventos como sequías, lluvias torrenciales y olas de calor están alterando los calendarios agrícolas y reduciendo la productividad en regiones clave. Esto afecta tanto la siembra como la cosecha del arroz, que es un cultivo sensible a la disponibilidad de agua.
2. **Crisis geopolíticas y económicas:** conflictos internacionales, inflación global y crisis de abastecimiento de insumos (fertilizantes, semillas, agroquímicos) han elevado los costos de producción, limitando la expansión agrícola en muchos países latinoamericanos.
3. **Transición hacia prácticas sostenibles:** si bien la adopción de tecnologías agrícolas sostenibles (riego eficiente, variedades resistentes al clima, agricultura de precisión) está en marcha, su implementación masiva aún enfrenta barreras técnicas, financieras y educativas, especialmente en pequeños y medianos productores.
4. **Demanda creciente y presión alimentaria:** la demanda mundial de arroz sigue aumentando, impulsada por el crecimiento poblacional, los cambios en la dieta y el papel del arroz como alimento básico en la región. Sin embargo, esta presión no se ha traducido en un incremento significativo de la oferta, lo cual puede generar tensiones entre consumo y disponibilidad.

La producción de arroz en América no colapsará, pero tampoco experimentará un auge acelerado sin intervención estratégica. Sin inversión en innovación agrícola, infraestructura hídrica y políticas de apoyo al productor, la región podría no estar preparada para enfrentar posibles crisis alimentarias asociadas al arroz en los próximos años, especialmente si se materializan eventos climáticos adversos o disrupciones comerciales globales. Este resultado no solo debe verse como una predicción técnica, sino como un llamado a la acción. La estabilidad proyectada puede ser interpretada como una oportunidad para actuar desde hoy en la planificación de cultivos, el fortalecimiento de las cadenas de valor y la seguridad alimentaria de millones de personas en América.

Conclusiones

Luego del desarrollo de este proyecto final puedo concluir lo siguiente:

1. La producción de arroz ha seguido una tendencia creciente durante las últimas seis décadas

- Desde 1961 hasta 2023, la producción total de arroz en América ha crecido de forma sostenida, pasando de alrededor de 11 millones a más de 39 millones de toneladas.
- Este comportamiento refleja el efecto combinado de la expansión agrícola, el uso de tecnologías mejoradas y políticas públicas orientadas al desarrollo rural.

2. La serie de tiempo presenta un comportamiento tendencial sin estacionalidad significativa

- El análisis de descomposición evidenció una clara tendencia de largo plazo sin componentes estacionales marcados.
- Esto permitió utilizar modelos como Holt (doble suavizamiento exponencial), sin necesidad de incorporar un componente estacional, lo que simplificó el proceso y mejoró la precisión.

3. El modelo Holt proporcionó un ajuste confiable y un pronóstico realista

- El modelo logró ajustarse muy bien a los datos históricos, con un bajo nivel de error y residuos aleatorios, indicando un buen desempeño.
- La predicción para 2024–2033 sugiere una estabilización en torno a los 39–40 millones de toneladas anuales, con variaciones leves pero sin quiebres abruptos.

4. La proyección coincide con el contexto actual del sector agroalimentario

- La tendencia plana proyectada concuerda con las condiciones que enfrenta el sector actualmente: encarecimiento de insumos, impactos del cambio climático, y limitaciones estructurales de los sistemas agrícolas.
- A pesar de los avances tecnológicos, la producción parece haber alcanzado un techo productivo, lo que evidencia la necesidad de nuevas políticas públicas, inversión en innovación y fortalecimiento de la resiliencia agrícola.

5. La utilidad del análisis predictivo va más allá del plano técnico

- Los resultados obtenidos no solo permiten comprender la evolución de la producción de arroz, sino también anticipar escenarios de disponibilidad, precios, comercio e intervención pública.
- Esta herramienta es altamente replicable para otros productos agroalimentarios y puede ser usada por instituciones públicas, ONGs, cooperativas, investigadores y tomadores de decisiones.

Finalmente, la producción de arroz en América representa un eje central de la seguridad alimentaria, el empleo agrícola y la estabilidad socioeconómica. El análisis predictivo realizado aporta una visión crítica y estratégica para planificar el futuro del sector, anticipar riesgos y aprovechar oportunidades. Si no se toman acciones coordinadas desde ahora, el escenario de estancamiento proyectado podría derivar en desequilibrios entre oferta y demanda en un futuro cercano.

Recomendaciones y futuros estudios

✓ Recomendaciones para la gestión agroalimentaria:

- 1. Desarrollar políticas de resiliencia productiva frente al cambio climático:** Dado que el pronóstico refleja una tendencia estabilizada, es fundamental implementar estrategias adaptativas (como sistemas de riego inteligente, seguros agrícolas, y cultivos resistentes a sequías) para prevenir caídas en la producción ante eventos extremos.
- 2. Fomentar el uso de modelos predictivos en la planificación agrícola nacional:** Instituciones gubernamentales y regionales deberían incorporar metodologías de análisis de series de tiempo para anticipar excedentes o déficits de producción, evitando improvisaciones en la política de importaciones, subsidios o almacenamiento estratégico.
- 3. Invertir en tecnología y digitalización rural:** El uso de sensores, plataformas de monitoreo satelital, y big data puede mejorar el seguimiento de cultivos y aumentar la productividad en regiones que actualmente operan con técnicas tradicionales o de baja escala tecnológica.
- 4. Fortalecer la capacitación de los actores de la cadena de valor del arroz:** Productores, cooperativas y asociaciones deben ser capacitados en interpretación de datos y análisis predictivo, para que puedan tomar decisiones informadas sobre fechas de siembra, precios proyectados o rotación de cultivos.
- 5. Establecer mecanismos de seguimiento y evaluación de la producción por región:** A pesar de que este análisis se realizó a nivel continental, es esencial replicarlo por país o subregión para atender las particularidades de cada entorno productivo, ya que existen diferencias marcadas entre América del Norte, Central y del Sur.

✓ Futuros estudios:

- 1. Ampliación del modelo a otros productos de la canasta básica:** Sería valioso aplicar la misma metodología a otros alimentos esenciales como maíz, papa, tomate, plátano, caña de azúcar y pollo, lo que permitiría obtener una visión más integral de la seguridad alimentaria en la región.
- 2. Inclusión de variables exógenas en los modelos predictivos:** Para mejorar la precisión de las predicciones, se recomienda incorporar variables

como precipitaciones, temperatura media anual, inversión pública en agricultura, uso de fertilizantes, superficie cultivada, entre otras.

3. **Análisis de sensibilidad y escenarios alternativos:** El desarrollo de modelos que simulen distintos escenarios (óptimos, críticos y base) ante cambios en el entorno global por ejemplo, crisis económicas o fenómenos climáticos puede ser útil para tomar decisiones de contingencia o prevención.
4. **Evaluación de impacto económico de las predicciones:** Traducir las estimaciones en términos de ingreso agrícola, exportaciones, precio interno del arroz y empleo permitiría dimensionar mejor las implicaciones económicas de la estabilización o variación futura en la producción.
5. **Explorar técnicas avanzadas de machine learning:** Aunque el modelo Holt demostró ser adecuado, investigaciones futuras podrían aplicar algoritmos como XGBoost, LSTM (redes neuronales) o modelos híbridos para comparar su desempeño frente a los modelos clásicos.

Este trabajo sienta las bases para una cultura de planeación agrícola basada en datos, en un contexto de creciente incertidumbre climática y económica. Transformar la información en conocimiento accionable es hoy más que nunca una prioridad para garantizar el derecho a la alimentación y el desarrollo rural sostenible en América.

Bibliografía

- Ahmad, R. S. (2023). *World Food Production Dataset*. Kaggle. Recuperado de: <https://www.kaggle.com/datasets/rafsunahmad/world-food-production>
- Food and Agriculture Organization of the United Nations (FAO). (2023). *FAOSTAT - Database of Food and Agriculture*. Recuperado de: <https://www.fao.org/faostat>
- Food and Agriculture Organization of the United Nations (FAO). (2024, abril 5). *FAO Food Price Index remains stable in March*. FAO Newsroom. Recuperado de: <https://www.fao.org/newsroom/detail/fao-food-price-index-remain-stable-in-march/en>
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. Recuperado de: <https://otexts.com/fpp3/>
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting* (2nd ed.). Wiley.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
- Statmodels. (2023). *Time Series Analysis Tools*. Documentación oficial de la librería Statsmodels para Python. Recuperado de: <https://www.statmodels.org/stable/tsa.html>
- Python Software Foundation. (2023). *Python Programming Language*. Documentación oficial. Recuperado de: <https://www.python.org/>
- Google. (2023). *Google Colab: Cloud-based Python Notebooks*. Recuperado de: <https://colab.research.google.com/>
- Inter-American Institute for Cooperation on Agriculture (IICA). (2022). *Perspectivas de la agricultura y el desarrollo rural en las Américas: una mirada hacia América Latina y el Caribe 2022-2023*. San José, Costa Rica.
- Comisión Económica para América Latina y el Caribe (CEPAL). (2021). *Sistemas alimentarios y COVID-19 en América Latina y el Caribe: el impacto sobre la seguridad alimentaria y la nutrición*. Naciones Unidas. Recuperado de: <https://www.cepal.org/es/publicaciones>
- Revista Estrategia y Negocios. (2024, abril 3). *Producción de arroz en Centroamérica es impactada por importaciones, costos y clima*. Recuperado de: <https://www.revistaeyn.com/centroamericaymundo/produccion-de-arroz-en-centroamerica-es-impactada-por-importaciones-costos-y-clima-JM25297943>

ANEXOS

REPOSITORIO (GitHub):

https://github.com/anavas26/Modelos_Predictivos_Angie_Navas_PF

- Base de datos :

https://github.com/anavas26/Modelos_Predictivos_Angie_Navas_PF/blob/main/world%20food%20production.csv

- Descripción de cada columna de datos

https://github.com/anavas26/Modelos_Predictivos_Angie_Navas_PF/blob/main/Proyecto_final_Modelos_predictivos_Navas_Angie_vf.ipynb

- Análisis descriptivo

https://github.com/anavas26/Modelos_Predictivos_Angie_Navas_PF/blob/main/Proyecto_final_Modelos_predictivos_Navas_Angie_vf.ipynb

- Análisis predictivo

https://github.com/anavas26/Modelos_Predictivos_Angie_Navas_PF/blob/main/Proyecto_final_Modelos_predictivos_Navas_Angie_vf.ipynb

- Script de código o archivos de dónde se hizo el análisis

https://github.com/anavas26/Modelos_Predictivos_Angie_Navas_PF/blob/main/Proyecto_final_Modelos_predictivos_Navas_Angie_vf.ipynb