

UNIVERSIDAD COMPLUTENSE DE MADRID  
FACULTAD DE COMERCIO Y TURISMO  
MASTER EN BIG DATA & DATA SCIENCE  
APLICACIONES AL COMERCIO, EMPRESA Y FINANZAS



TRABAJO FINAL DE MÁSTER  
2022 – 2023

---

**MODELO DE ESTIMACIÓN DE LA DEMANDA DE ENERGÍA O  
ELECTRICIDAD SECTORIAL DE UN PAÍS**

---

Autores: (GRUPO 1)

- Ruth Fernández Padilla
- Ángel Martínez Barrial
- Alejandro Minguez Bonache
- Eduardo Urrutia Rivas
- Nam Nguyen Thi
- Antonio Villamayor Delgado

Tutores:

- Carlos Ortega
- Santiago Mota.

Septiembre, 2023

**Derechos de autor**

MRC Consultores y Asesores de Transacciones

Firma:

Fecha: 21 de Septiembre, 2023

## Agradecimiento

Este proyecto ha sido llevado a cabo con el apoyo y la colaboración de muchas personas, a quienes expresamos nuestra sincera gratitud.

En primer lugar, nos gustaría extender nuestro más sincero agradecimiento a Jorge Colomer, Jefe de la Práctica de Aprendizaje Automático, Inteligencia de Negocios y Análisis de Datos, así como Director Comercial en la empresa MRC. A lo largo de todo este verano, ha sido una generosa fuente de inspiración y ayuda. Su orientación y apoyo no pueden ser exagerados.

También queremos reconocer el apoyo y la asistencia que recibimos de los profesores en este curso. Fueron lo suficientemente generosos como para compartir su tiempo y experiencia con nosotros.

De manera similar, nos gustaría expresar nuestra gratitud a los numerosos amigos y colegas que nos acompañaron en este viaje.

Finalmente, queremos agradecer a nuestras familias por su amor y apoyo. Siempre nos recuerdan nuestro valor, especialmente en momentos desafiantes.

*“Los datos no son solo información, son la clave para el conocimiento y la toma de decisiones informadas en la era digital”*

Peter Drucker

## **Resumen**

El objetivo de este proyecto es predecir el consumo de energía en Brasil hasta el año 2040 utilizando datos de matrices y balances publicados en el sitio web oficial de OLADE. Este trabajo comprueba que un modelo de aprendizaje automático puede producir resultados suficientemente precisos para un problema de pronóstico complejo. Hemos explorado varias técnicas de aprendizaje automático y desarrollamos un modelo basado en datos para la previsión del consumo de energía de Brasil. El conjunto de datos incluyó un registro de 52 años (1970-2021) y utilizamos un modelo ARIMA para entrenar los datos. El rendimiento del modelo se evaluó utilizando el error porcentual absoluto medio (MAPE - Mean Absolute Percentage Error) para garantizar la comparabilidad directa con las lecturas de energía en el conjunto de datos. Los resultados indican que los algoritmos de aprendizaje automático pueden predecir eficazmente el consumo de energía. Estos hallazgos se pueden aplicar a otros países de América del Sur para pronosticar su consumo de energía.

## INDICE

<b>1. Introducción .....</b>	<b>1</b>
<b>2. Fundamentos .....</b>	<b>2</b>
2.1 Series de tiempo.....	2
2.2 Características de las series temporales.....	2
2.3 Métodos de pronóstico de series temporales.....	2
2.4 ARIMA (p, d, q) .....	3
<b>3. Procesamiento de datos.....</b>	<b>4</b>
3.1. Datos.....	4
3.2 Preprocesamiento de datos.....	6
3.3 Análisis de datos y visualización de los datos originales.....	6
<b>4. Deploy Modelo ARIMA (univariado).....</b>	<b>8</b>
4.1 ARIMA para todos los sectores .....	8
4.2 Entrenamiento, validación y predicción del modelo.....	9
<b>5. Análisis y visualización de los resultados .....</b>	<b>13</b>
5.1 Predicción de distribución de energías por los sectores .....	13
5.2 Predicción de contribución de las fuentes de energía a consumo final .....	14
<b>6. Conclusiones .....</b>	<b>17</b>
<b>Referencias bibliográficas/ Anexos .....</b>	<b>18</b>

## 1. Introducción

La demanda mundial de energía para el consumo de energía aumenta día a día, y parece un desafío para la mayoría de los países satisfacer la demanda con la producción total de energía.

En un mundo cada vez más tenso, los países BRICS (Brasil, Rusia, India, China y Sudáfrica) son los principales productores de energía del mundo, así como los principales consumidores de combustible.

El sector energético brasileño ha sido testigo de numerosos cambios tecnológicos y ha evolucionado hasta convertirse en líder mundial en ventas de tecnologías limpias, tanto en el mercado interno como externo. Muchos factores contribuyeron a las actividades innovadoras en su sector eléctrico, que incluyen la contribución tanto del gobierno como de la IED.

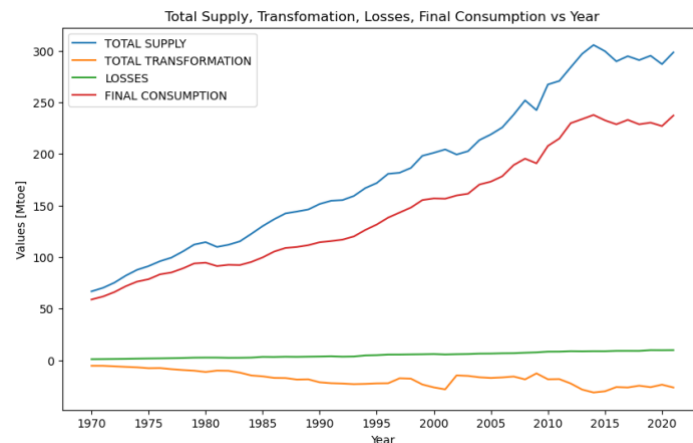


Figura 1: Abastecimiento total, transformación, pérdidas y consumo final por año

En los últimos años, la inteligencia artificial en general y el aprendizaje automático en concreto presentan soluciones a una enorme cantidad de problemas. Una de las aplicaciones prometedoras es predecir el consumo energético futuro utilizando series temporales históricas desarrollando un modelo de imitación.

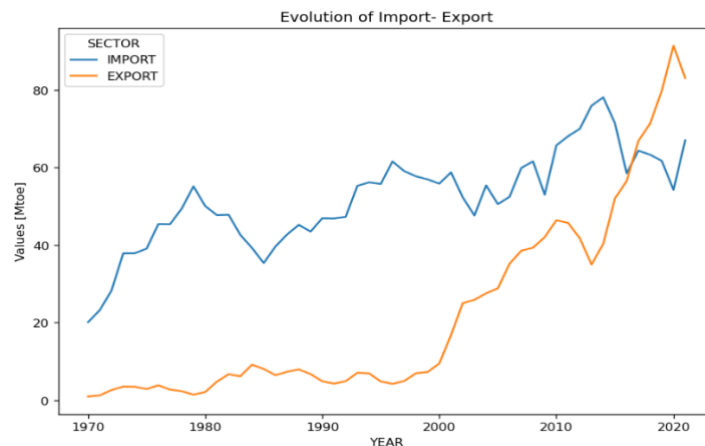


Figura 2: Evolución de importación y exportación de energía

Este proyecto tiene como objetivo probar si es posible aplicar el modelo ARIMA a los datos de series temporales. El objetivo es obtener resultados suficientemente buenos en la previsión del consumo de energía.

## 2. Fundamentos

### 2.1 Series de tiempo

Los datos de series de tiempo rastrean ciertos valores durante un período específico, generalmente con el tiempo como variable independiente, y a menudo son utilizados para pronósticos futuros. Las series de tiempo se puede clasificar en continuas y discretas según los intervalos de tiempo. Además, los datos de series de tiempo se pueden dividir en univariados (seguimiento de una sola variable a lo largo del tiempo) y multivariados (seguimiento de múltiples variables a lo largo del tiempo) según la cantidad de variables dependientes.

### 2.2 Características de las series temporales

Las series temporales tienen tres características claves y cruciales para realizar pronósticos precisos:

- **Autocorrelación:** Mide la relación entre el valor actual de una variable y sus valores pasados.
- **Estacionalidad:** Se refiere a variaciones regulares y de corto plazo que ocurren, como patrones semanales, mensuales o trimestrales.
- **Estacionariedad:** Implica que las propiedades estadísticas de una serie temporal permanecen constantes a lo largo del tiempo, con una media y varianza consistentes y una covarianza independiente del tiempo.

### 2.3 Métodos de pronóstico de series temporales

A lo largo de los años, numerosos estudios han evaluado el rendimiento de los métodos clásicos y de aprendizaje automático. Algunos métodos clásicos bien conocidos son los siguientes:

- Auto-Regressive Moving Average (ARMA)
- Auto-Regressive Integrated Moving Average (ARIMA)
- Seasonal Autoregressive Integrated Moving-Average (SARIMA)

Asimismo, también existen algoritmos modernos como:

- Multi-Layer Perceptron (MLP)
- Bayesian Neural Networks (BNN)
- Generalized Regression Neural Networks (GRNN).
- Recurrent Neural Networks (RNN)
- Long Short- Term Memory (LSTM)-Brownlee, 2016

## 2.4 ARIMA (p, d, q)

ARIMA significa media móvil integrada autorregresiva, y es una forma de modelar datos de series de tiempo para realizar pronósticos o sea, para predecir puntos futuros en la serie:

- **AR:** *Autoregresion*. Modelo que utiliza la relación de dependencia entre una observación y un cierto número de observaciones rezagadas.
- **I:** *Integration*. El uso de diferenciación de observaciones sin procesar, por ejemplo, restar una observación de una observación en el paso de tiempo anterior para hacer que la serie de tiempo sea estacionaria.
- **MA:** *Moving Average*. Un modelo que utiliza la dependencia entre una observación y un error residual de un modelo de promedio móvil aplicado a observaciones rezagadas.

Los parámetros del modelo ARIMA se definen de la siguiente manera:

- **p:** El número de observaciones de retraso incluidas en el modelo, también llamado orden de retraso.
- **d:** El número de veces que se diferencian las observaciones sin procesar, también llamado grado de diferenciación.
- **q:** El tamaño de la ventana de media móvil, también llamado orden de media móvil.

Un modelo ARIMA es aquel donde la serie temporal se diferencia al menos una vez para hacerla estacionaria y se combinan los términos AR y MA.

Entonces se obtiene la siguiente ecuación:

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

Para aplicar el modelo ARIMA, se deben seguir los siguientes pasos:

1. **Preparación de datos:** obtenga los datos de series temporales en orden.
2. **Estacionar los datos:** eliminar tendencias y estacionalidades.
3. **Identificar parámetros:** encontrar el orden correcto del modelo (p, d, q).
4. **Ajustar ARIMA:** aplique el modelo ARIMA a sus datos.
5. **Verifique los residuos:** asegúrese de que los residuos se parezcan al ruido blanco.
6. **Pronóstico:** utilice el modelo para predicciones futuras.
7. **Evaluar:** evaluar la precisión del pronóstico y ajustarlo si es necesario.
8. **Validar:** probar nuevos datos para confirmar la confiabilidad.
9. **Visualizar:** Comunique los resultados visualmente.
10. **Implementar:** ponga el modelo en acción si cumple con sus requisitos.



### 3. Procesamiento de datos

#### 3.1. Datos

El conjunto de datos utilizado para este estudio abarca el balance de la matriz energética de Brasil desde 1970 hasta 2021, obtenido del sitio web de OLADE (<https://www.olade.org/>). Este conjunto completo de datos está estructurado en un archivo Excel con varias hojas. Importamos este archivo de Excel directamente desde un repositorio en [GitHub](#) como un diccionario, y cada clave corresponde a un año específico dentro del período de 1970 a 2021, después de una limpieza de datos.

El Notebook también está organizado en un repositorio en [GitHub](#) ([TFM Group 1](#))

Cada año, los datos se organizan como un DataFrame, donde las columnas representan las fuentes de energía y las filas representan los sectores donde se distribuyen las energías.

	SECTOR	OIL	NATURAL GAS	COAL	HYDROENERGY	GEOTHERMAL	NUCLEAR	FIREWOOD
1	PRODUCTION	152.71343	26.37148	2.63308	31.12501	NaN	3.83575	26.01795
2	IMPORT	7.35921	14.79623	13.21631	NaN	NaN	NaN	NaN
3	EXPORT	65.72407	NaN	NaN	NaN	NaN	NaN	NaN
4	STOCK CHANGE	-1.00778	-3.59893	0.06782	NaN	NaN	NaN	NaN
5	UNUSED	NaN	1.21787	NaN	NaN	NaN	NaN	NaN

Figura 3: Primeras filas y columnas del DataFrame del año 2021

Todas las columnas del conjunto de datos son importantes para nuestro estudio debido a su función de proporcionar una visión integral del equilibrio de la matriz energética de Brasil de 1970 a 2021. Cada columna representa un aspecto específico de la producción y el consumo de energía y factores relacionados, incluidas diversas fuentes de energía y sectores.

Para conocer mejor la información que manejamos, vamos a explicar brevemente cada una de las variables que componen el set de datos:

- **SECTOR:** Indica el sector económico al que pertenecen los datos, como oferta total, consumo total (Agricultura, Industria, Transporte, Residencial, Comercial,...), etc. Cada fila corresponde a un sector y un año específico.
- **OIL:** La cantidad de energía consumida o producida a partir del petróleo en megatones equivalentes de petróleo (Mtoe).
- **NATURAL GAS:** La cantidad de energía consumida o producida a partir del gas natural en megatones equivalentes de petróleo (Mtoe).
- **COAL:** La cantidad de energía consumida o producida a partir del carbón en megatones equivalentes de petróleo (Mtoe).
- **HYDROENERGY:** La cantidad de energía hidroeléctrica consumida o producida en megatones equivalentes de petróleo (Mtoe).
- **GEOTHERMAL:** La cantidad de energía geotérmica consumida o producida en megatones equivalentes de petróleo (Mtoe).
- **NUCLEAR:** La cantidad de energía nuclear consumida o producida en megatones equivalentes de petróleo (Mtoe).

- **FIREWOOD:** La cantidad de energía consumida o producida a partir de leña en megatones equivalentes de petróleo (Mtoe).
- **SUGARCANE AND PRODUCTS:** La cantidad de energía consumida o producida a partir de caña de azúcar y sus productos en megatones equivalentes de petróleo (Mtoe).
- **OTHER PRIMARY:** La cantidad de energía consumida o producida a partir de otras fuentes primarias en megatones equivalentes de petróleo (Mtoe).
- **TOTAL PRIMARIES:** La suma total de todas las fuentes primarias de energía consumidas en el sector en un año específico.
- **ELECTRICITY:** La cantidad de energía consumida en forma de electricidad en megatones equivalentes de petróleo (Mtoe).
- **LPG:** La cantidad de energía consumida en forma de gas licuado de petróleo en megatones equivalentes de petróleo (Mtoe).
- **GASOLINE/ALCOHOL:** La cantidad de energía consumida en forma de gasolina y alcohol en megatones equivalentes de petróleo (Mtoe).
- **KEROSENE/JET FUEL:** La cantidad de energía consumida en forma de queroseno y combustible para aviones en megatones equivalentes de petróleo (Mtoe).
- **DIESEL OIL:** La cantidad de energía consumida en forma de aceite diesel en megatones equivalentes de petróleo (Mtoe).
- **FUEL OIL:** La cantidad de energía consumida en forma de aceite combustible en megatones equivalentes de petróleo (Mtoe).
- **COKE:** La cantidad de energía consumida en forma de coque en megatones equivalentes de petróleo (Mtoe).
- **CHARCOAL:** La cantidad de energía consumida en forma de carbón vegetal en megatones equivalentes de petróleo (Mtoe).
- **GASES:** La cantidad de energía consumida en forma de gases en megatones equivalentes de petróleo (Mtoe).
- **OTHER SECONDARY:** La cantidad de energía consumida en otras formas secundarias en megatones equivalentes de petróleo (Mtoe).
- **NON-ENERGY:** La cantidad de energía utilizada para fines no relacionados con la generación de energía en megatones equivalentes de petróleo (Mtoe).
- **TOTAL SECUNDARIES:** La suma total de todas las fuentes secundarias de energía consumidas en el sector en un año específico.
- **NON-ENERGY:** La cantidad de energía utilizada para fines no relacionados con la generación de energía en megatones equivalentes de petróleo (Mtoe).
- **TOTAL:** La suma total de todas las formas de energía consumidas en el sector en un año específico.
- **YEAR:** Indica el año al que corresponde la información de consumo de energía para ese sector en particular.

Si bien existen valores atípicos que pueden influir en los resultados del modelo, se han conservado para el análisis a fin de capturar posibles ideas valiosas o anomalías en los

datos energéticos. Comprender y considerar estos aspectos es crucial para un análisis de datos y un desarrollo de modelos sólidos.

La presencia de valores NaN (Not-a-Number) en la base de datos se debe principalmente a información faltante o no disponible. En este conjunto de datos, los valores de NaN son particularmente abundantes porque representan casos en los que los valores reales no están registrados o no están disponibles. Por lo tanto, estos valores faltantes se han representado como NaN para indicar la ausencia de datos.

## 3.2 Preprocesamiento de datos

Realizamos diferentes acciones como:

- Redondear valores, cambiar el nombre de columnas, eliminar espacios y corregir nombres. Además, agregamos el año como una nueva columna.
- Convertimos unidades de medida de Ktoe a Mtoe en todas las columnas numéricas.
- Los valores numéricos se redondearon a dos decimales.
- Se eliminaron los encabezados de fila que contenían unidades de medida y las columnas "Sin nombre: 0" pasaron a llamarse "SECTOR" y "OTHER PRIMARY\_x000d\_" pasó a llamarse "OTHER PRIMARY".
- Se corrigió el nombre del sector 'COKE PLANTS AND BLAST FURNACES\_x000d\_' fue corregido a 'COKE PLANTS AND BLAST FURNACES.'.

Estas transformaciones garantizan que los datos estén en un formato coherente y listos para un posterior análisis.

- Los datos procesados se almacenan en un nuevo diccionario llamado 'modified\_dict', que reemplaza al original.
- Finalmente, el diccionario limpio se guarda en un nuevo archivo Excel denominado “*Brazil\_Energy balance matrix\_cleaned.xlsx*” preservando las hojas originales para su posterior análisis y visualización.

Este proceso garantiza la calidad y coherencia de los datos antes de su utilización en análisis más profundos.

## 3.3 Análisis de datos y visualización de los datos originales

### 3.3.1 Diagrama de Sankey

Los diagramas de Sankey son herramientas valiosas para visualizar los flujos de energía y comprender los balances energéticos en sistemas complejos. Revisar los diagramas de Sankey en un contexto de balance energético es crucial para obtener conocimientos sobre los flujos de energía, identificar ineficiencias, garantizar un equilibrio entre la oferta y la demanda, y tomar decisiones informadas sobre políticas energéticas, mejoras de eficiencia y gestión ambiental. Proporcionan una representación concisa e intuitiva desde

el punto de vista visual de datos energéticos complejos, lo que facilita que los interesados comprendan y aborden los desafíos y oportunidades relacionados con la energía.

En nuestro Cuaderno, presentamos nuestro propio código para obtener el Diagrama de Sankey. Puede ser útil para futuras visualizaciones.

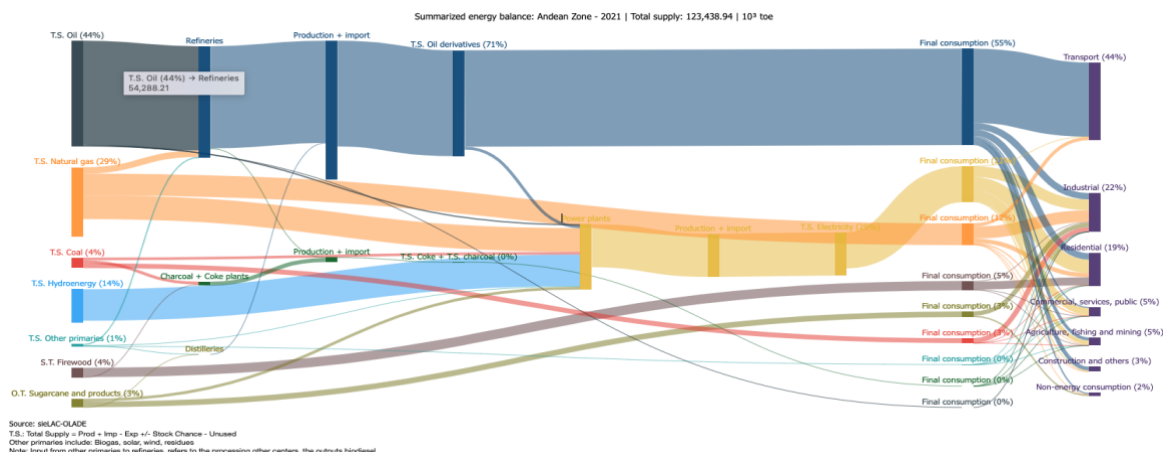


Figura 3: Diagrama de Sankey de la Matriz Energética de Brasil, 2021.

### 3.3.2 Energías primarias y secundarias

El suministro de energía depende de fuentes de energía tanto primarias como secundarias.

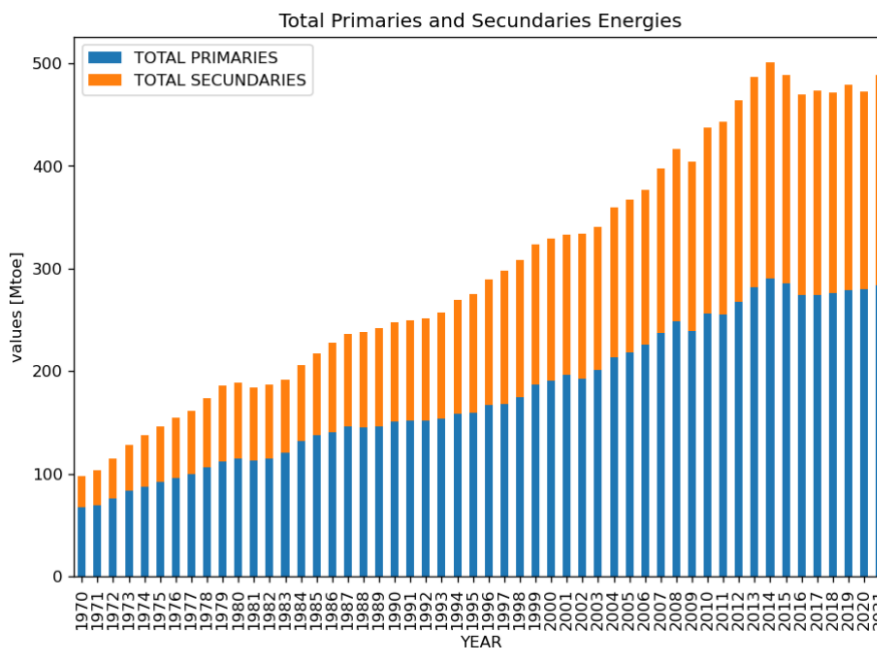


Figura 4: Oferta total de energías primarias y secundarias

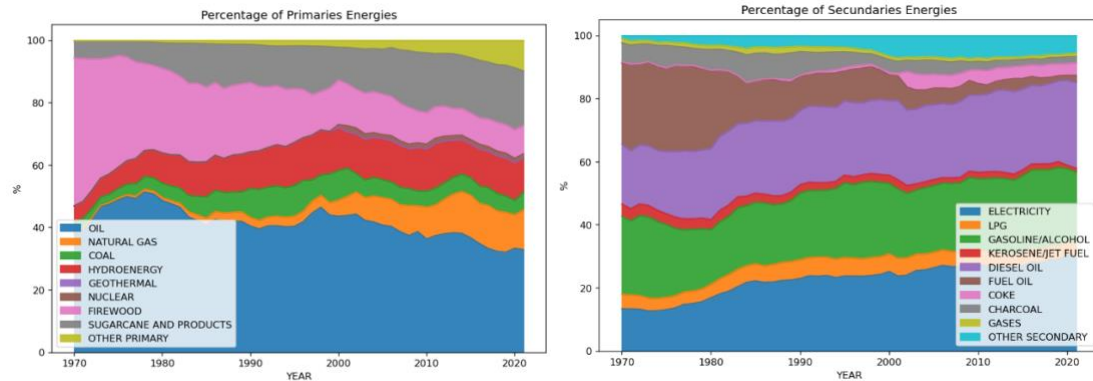


Figura 5: Fuentes de Energía primaria y secundaria por año

### 3.3.3 Distribución de consumo final

El consumo de energía final se distribuye entre varios sectores del país, incluidos el transporte, la industria, el residencial, los servicios públicos comerciales, la agricultura, la pesca y la minería, la construcción y el consumo no energético.

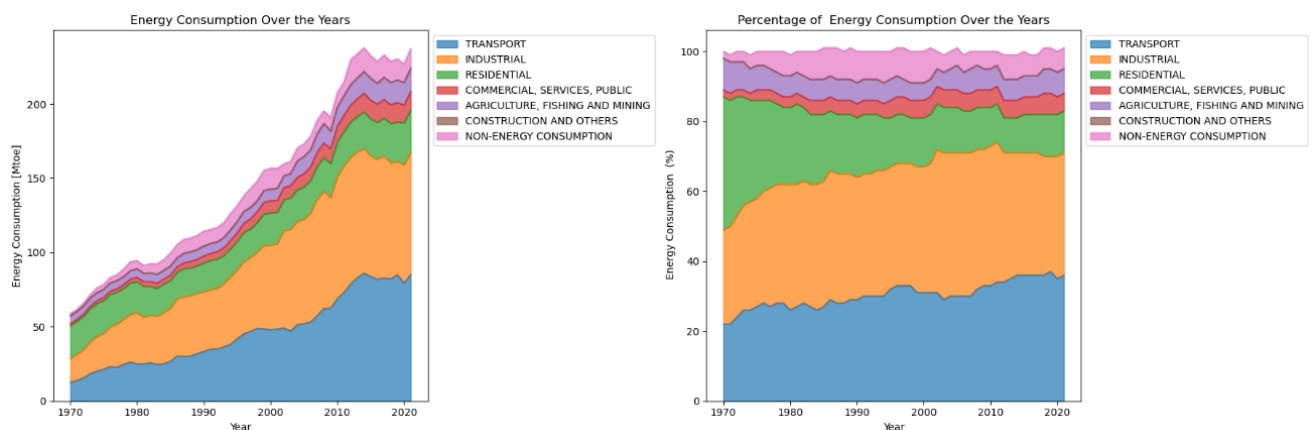


Figura 6: Tendencias del consumo de energía a lo largo del tiempo (Volumen - a, Porcentaje - b)

## 4. Deploy Modelo ARIMA (univariado)

Aplicamos el modelo ARIMA a todos los sectores, pero primero debemos realizar algunas comprobaciones para seleccionar los mejores parámetros (p, d, q).

En el análisis de series de tiempo, si necesitamos escalar los datos (es decir, estandarizarlos o normalizarlos) depende de las características específicas de los datos y de las técnicas de modelado que planeamos utilizar. En el cuaderno de código, utilizamos datos escalares y también debemos realizar una transformación inversa durante el proceso de predicción.

En esta sección, nos centraremos en modelar el problema univariado.

### 4.1 ARIMA para todos los sectores

**Paso 1:** Verificar manualmente para elegir d:

En este paso, primero verificamos la estacionariedad de una serie temporal realizando una prueba estadística, como la prueba Dickey-Fuller aumentada (ADF), y examinando el valor  $p$  asociado con la prueba. Si el  $p$ -valor es menor que 0,05, podemos considerar la serie como estacionaria; de lo contrario, hacemos que la serie sea estacionaria aplicando diferenciación. El orden mínimo de diferenciación es el valor de  $d$ .

Además, trazamos la función de autocorrelación (ACF, Auto-Correlation Function) y la función de autocorrelación parcial (PACF, Partial Auto-Correlation Function) en cada orden de diferenciación. Este paso, también, nos permite seleccionar los valores apropiados para ' $p$ ' y ' $q$ ' según las características de los gráficos ACF y PACF.

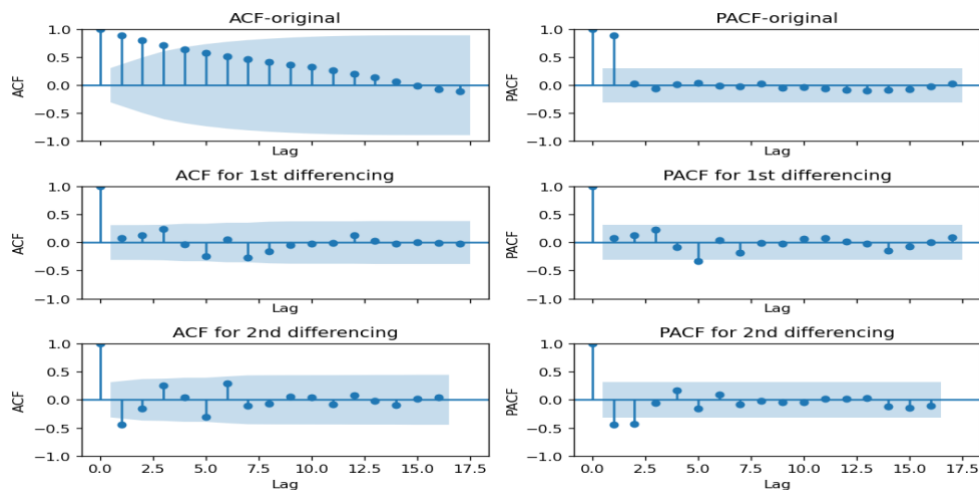


Figura 7: ACF- PACF

**Paso 2:** Encontrar la mejor combinación ( $p, q$ ) manualmente:

Este paso es opcional, ya que " $p$ " y " $q$ " también se pueden seleccionar encontrando los valores mínimos de AIC (Akaike Information Criterion) y BIC (Bayesian Information Criterion) si " $d$ " ya está determinado.

Los valores óptimos de " $p$ " y " $q$ " varían entre sectores, y la elección de los criterios (AIC o BIC) puede influir en la selección. Estas variaciones indican la complejidad de modelar el consumo de energía en diferentes sectores, y la selección del mejor modelo ARIMA puede depender del contexto específico y los objetivos del análisis.

## 4.2 Entrenamiento, validación y predicción del modelo

En nuestro proceso de modelado, asignamos una parte importante del conjunto de datos para fines de *training* o entrenamiento y *test* o validación. Específicamente, utilizamos el 90% de los datos para entrenamiento y el 10% para validación. Esta división nos permitió entrenar eficazmente nuestros modelos ARIMA y evaluar su rendimiento.

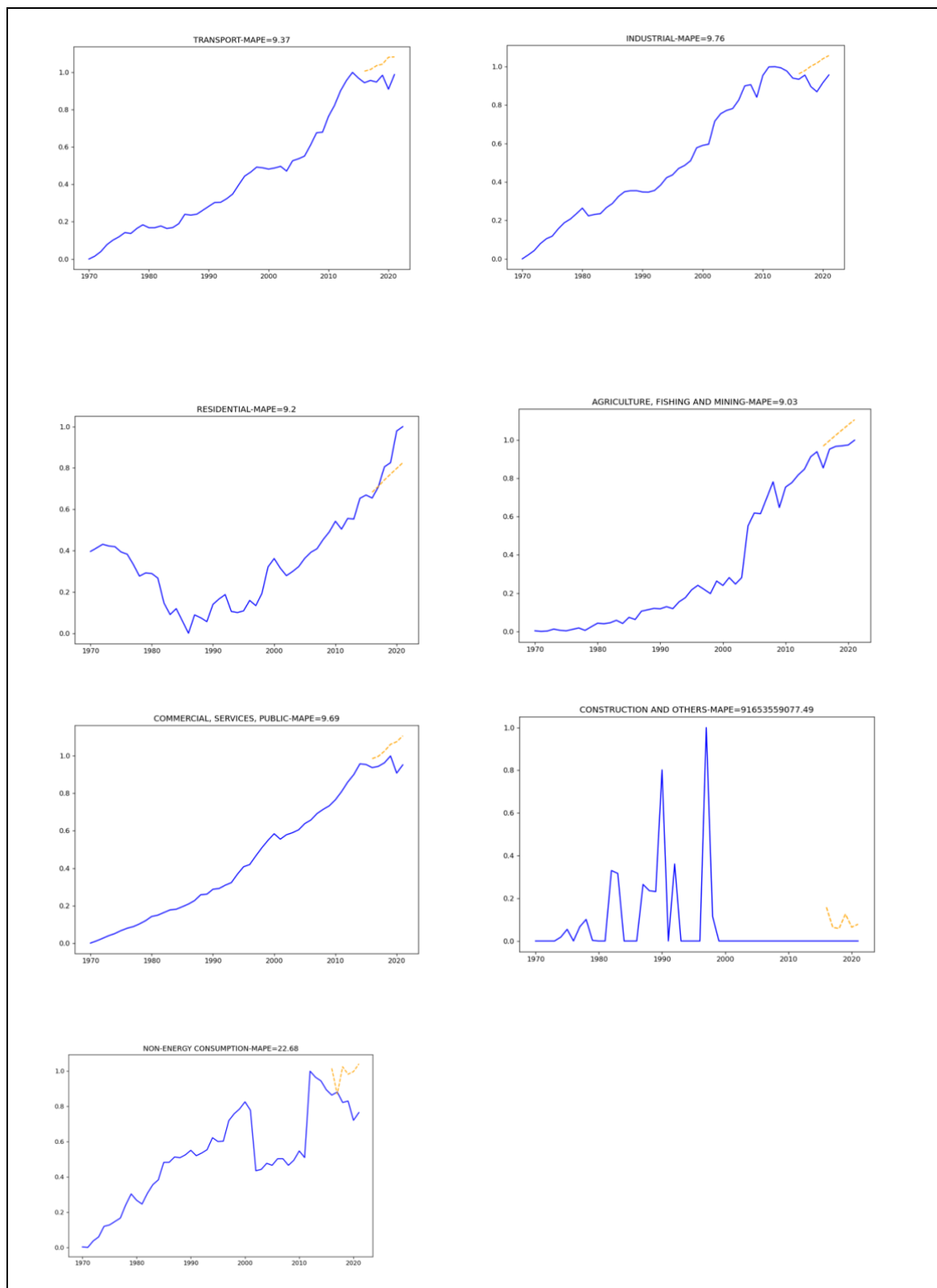
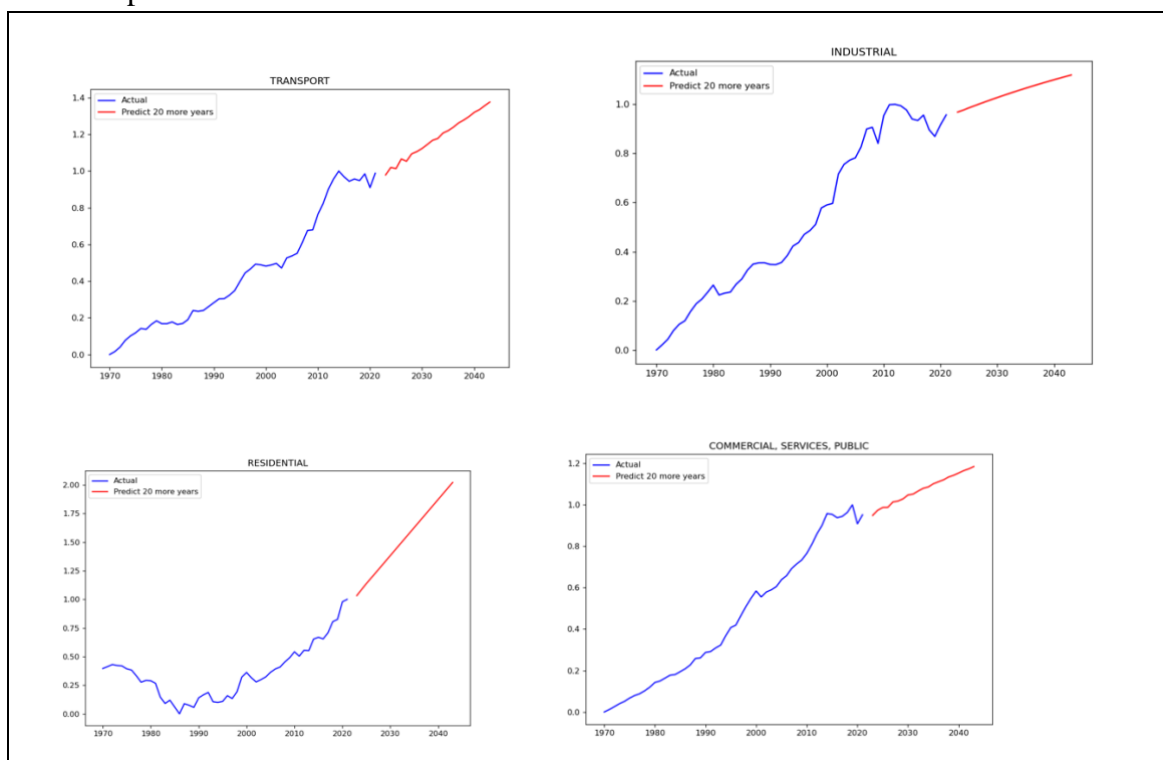


Figura 8: Resultados de validación

Para la fase de validación, utilizamos el conjunto de datos de validación para evaluar la precisión de las predicciones de nuestros modelos. Este proceso implicó comparar los valores predichos generados por nuestros modelos ARIMA con los datos reales en el

conjunto de validación. Luego, calculamos el Error Porcentual Absoluto Medio (MAPE) para cuantificar el nivel de precisión alcanzado. El MAPE proporcionó información valiosa sobre el rendimiento de los modelos, ayudándonos a comprender cuán bien pueden generalizar hacia datos no vistos. Si el valor del MAPE para cada sector es menor al 10%, podemos considerar que el modelo se ajusta bien. Sin embargo, en algunos sectores (como construcción y consumo no energético), carecemos de datos e información suficientes para emitir un juicio concluyente sobre el ajuste del modelo.

Estos valores MAPE proporcionaron información valiosa sobre el rendimiento de los modelos, ayudando a comprender qué tan bien podrían generalizarse a datos invisibles. Si bien la mayoría de los sectores alcanzaron valores MAPE razonables, es esencial abordar el valor atípico en el sector “CONSTRUCTION AND OTHERS”, donde MAPE parece excepcionalmente alto, lo que indica problemas potenciales con el desempeño del modelo para este sector.





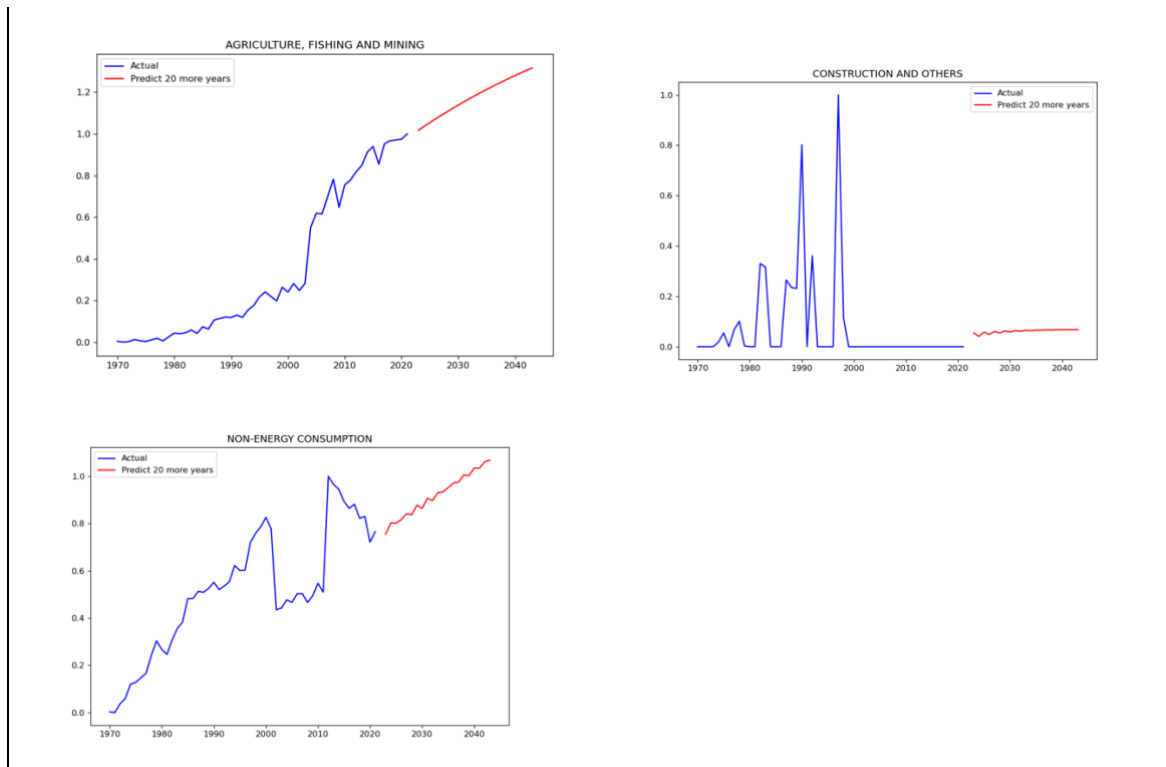


Figura 9: Resultados de predicción

En la fase de predicción, proyectamos el consumo futuro de energía para cada sector. Específicamente, pronosticamos las tendencias del consumo de energía para los próximos 20 años, lo que nos permite hacer predicciones informadas sobre el consumo de energía en los años venideros.

Siguiendo este enfoque de *training*-entrenamiento, *test*-validación y predicción, nos aseguramos de que nuestros modelos fueran sólidos y capaces de realizar pronósticos confiables para diversos sectores de consumo de energía. Estos pasos fueron esenciales en la preparación de nuestros datos para análisis posteriores y procesos de toma de decisiones.

## 5. Análisis y visualización de los resultados

### 5.1 Predicción de distribución de energías por los sectores

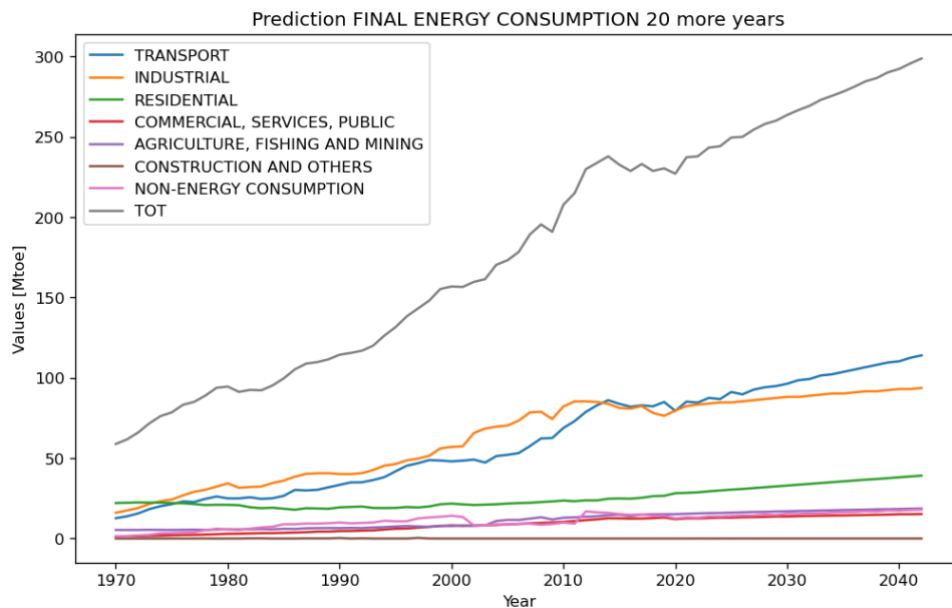


Figura 10: Predicción de “FINAL ENERGY CONSUMPTION” +20 años

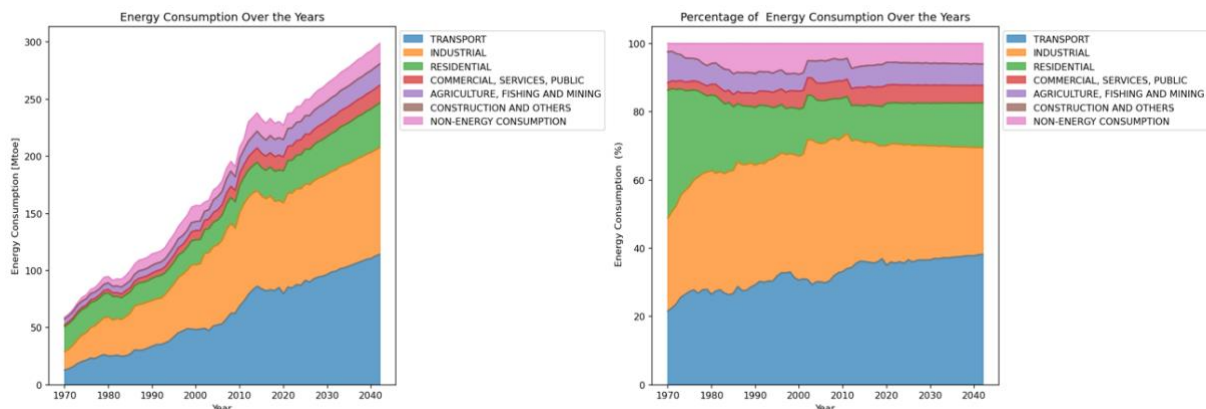


Figura 11: Predicción “ENERGY CONSUMPTION BY SECTOR” +20 años

Pronosticamos las tendencias del consumo de energía para los próximos 20 años. Para hacer esto, utilizamos los modelos ARIMA entrenados, y especificamos la cantidad de períodos que queremos pronosticar, lo que nos permite hacer predicciones informadas sobre el consumo de energía en los años venideros.

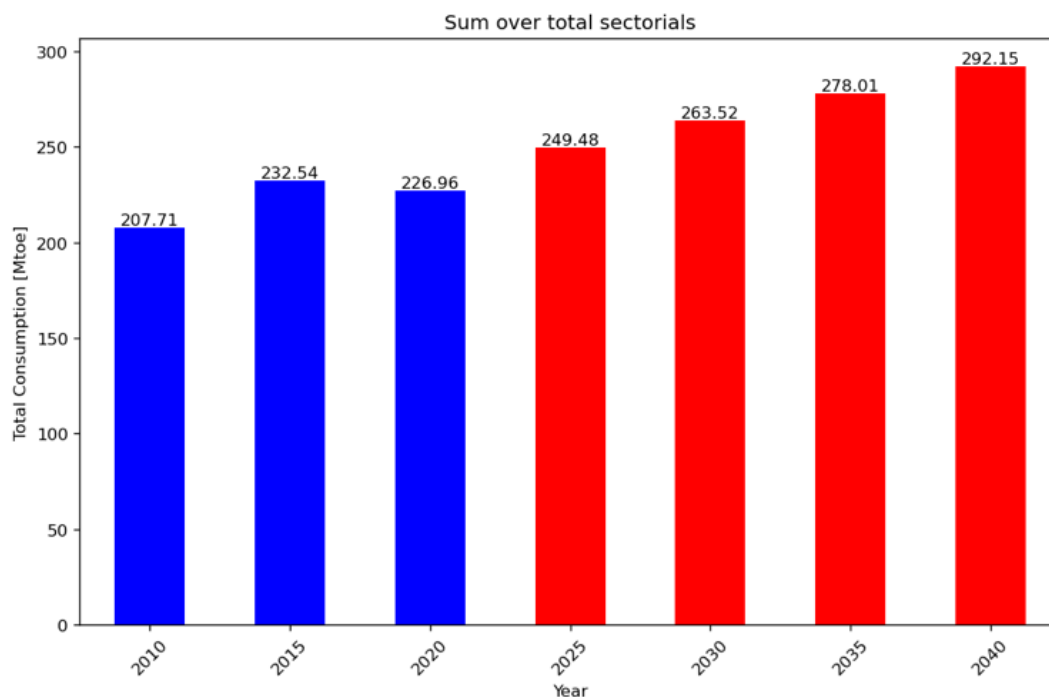


Figura 12: Predicción de “SUM OVER TOTAL SECTORIALS” +20 años

Puede que comparemos este resultado con la predicción publicada en el sitio web oficial ([eneroutlook](http://eneroutlook.org))

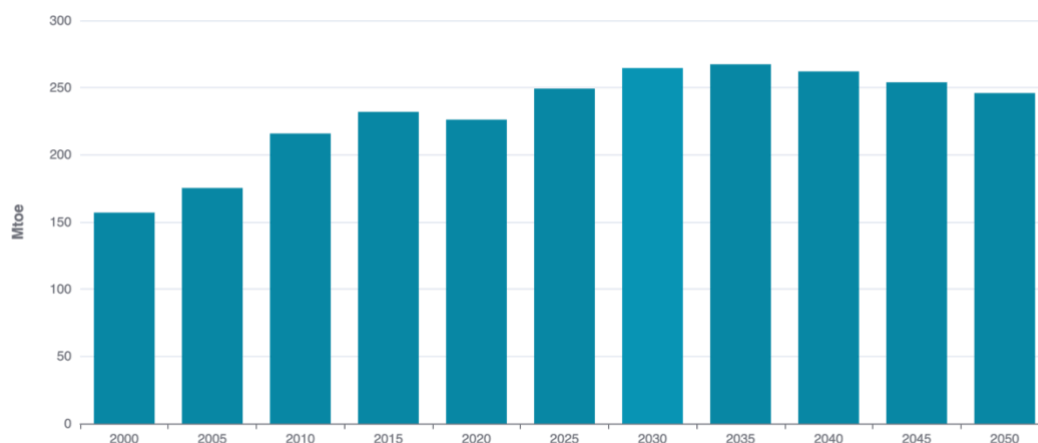


Figura 13: Resultado predicción en la página web eneroutlook

## 5.2 Predicción de contribución de las fuentes de energía a consumo final

En esta sección, hemos realizado una serie de pasos de preparación y análisis de datos para crear escenarios para el consumo futuro de energía en diferentes grupos de fuentes de energía. Los principales objetivos de este análisis fueron comprender los patrones

históricos de consumo de energía, clasificar las fuentes de energía en grupos significativos y estimar cómo estos grupos podrían contribuir al consumo de energía futuro.

Comenzamos seleccionando el sector “FINAL CONSUMPTION” de nuestro conjunto de datos y limpiando los datos para eliminar columnas innecesarias. A continuación, categorizamos las fuentes de energía en distintos grupos, como CARBON, PETROLEUM, NATURAL GAS, RENEWABLE, ELECTRICITY y OTHER. Esta agrupación nos permitió agregar el consumo de energía dentro de cada categoría.

```
# Clasify the energy sources:
group_CARBON=['COAL', 'FIREWOOD', 'COKE', 'CHARCOAL']
group_PETROLIUM= ['OIL', 'GASOLINE/ALCOHOL', 'KEROSENE/JET FUEL', 'DIESEL OIL', 'FUEL OIL']
group_NATURAL_GAS=['NATURAL GAS', 'LPG', 'GASES']
group_RENEWABLE=['SUGARCANE AND PRODUCTS']
group_ELECTRICITY= ['ELECTRICITY']
group_OTHER=['OTHER PRIMARY', 'OTHER SECONDARY', 'NON-ENERGY']
```

Se debe tener en cuenta que desde 1990 Brasil ha estado desarrollando un programa nuclear como una fuente de energía relativamente joven que ha realizado una contribución significativa a la generación de electricidad en el país.

Para comprender la distribución histórica de las fuentes de energía, calculamos la distribución porcentual de cada grupo en relación con el consumo total de energía para cada año. Estos porcentajes se redondearon a números enteros por simplicidad.

YEAR	group_CARBON		group_PETROLIUM		group_NATURAL_GAS		group_ELECTRICITY		group_RENEWABLE		group_OTHER		TOTAL	
	[Mtoe]	%	[Mtoe]	%	[Mtoe]	%	[Mtoe]	%	[Mtoe]	%	[Mtoe]	%	[Mtoe]	%
2012	32.59	14.0	94.41	41.0	22.43	10.0	40.62	18.0	17.89	8.0	21.84	10.0	229.77	100
2013	31.71	14.0	98.04	42.0	22.38	10.0	41.87	18.0	17.25	7.0	22.45	10.0	233.71	100
2014	32.41	14.0	101.04	42.0	22.28	9.0	43.02	18.0	16.16	7.0	22.86	10.0	237.76	100
2015	32.23	14.0	99.11	43.0	22.09	10.0	42.24	18.0	15.53	7.0	21.34	9.0	232.54	100
2016	29.82	13.0	96.53	42.0	21.22	9.0	42.17	18.0	17.57	8.0	21.34	9.0	228.65	100
2017	32.04	14.0	98.29	42.0	21.61	9.0	42.76	18.0	17.22	7.0	21.10	9.0	233.00	100
2018	33.18	15.0	95.91	42.0	21.94	10.0	43.49	19.0	13.25	6.0	20.89	9.0	228.67	100
2019	32.31	14.0	99.31	43.0	20.62	9.0	44.00	19.0	13.19	6.0	20.87	9.0	230.30	100
2020	31.75	14.0	92.49	41.0	19.22	8.0	43.70	19.0	18.09	8.0	21.71	10.0	226.96	100
2021	33.63	14.0	99.15	42.0	21.15	9.0	45.63	19.0	15.42	7.0	22.22	9.0	237.19	100

Tabla 1: Fuentes de energía en el consumo final durante los últimos 10 años

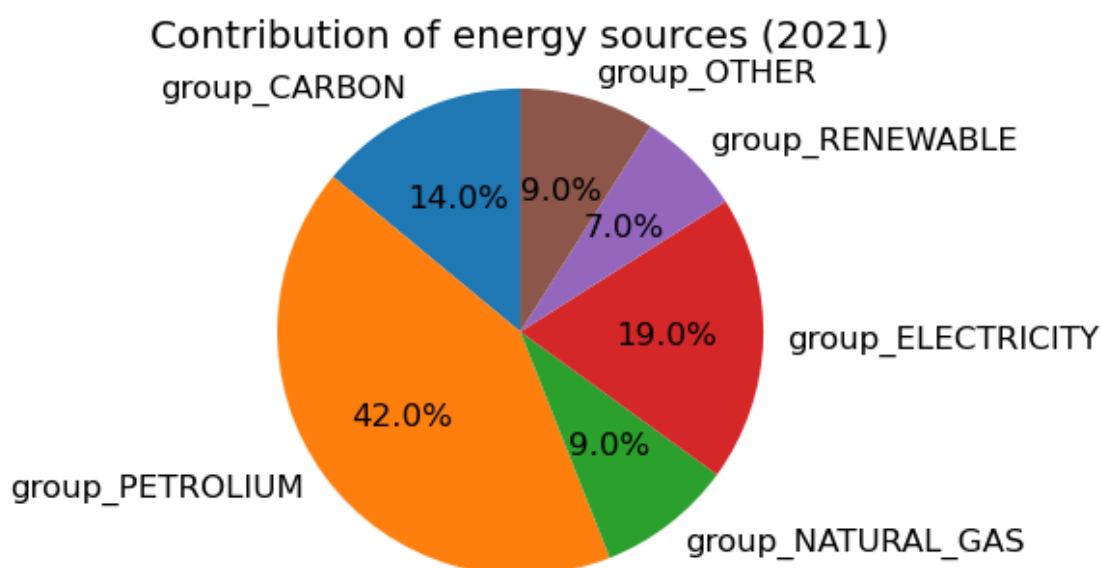


Figura 14: Contribución (%) de fuentes de energías al consumo final en Brasil (2021)

Luego, los datos resultantes se combinaron en un DataFrame completo, *final\_df*, que incluye valores históricos de consumo de energía (en Mtoe) y la distribución porcentual de las fuentes de energía para cada año. Este DataFrame proporciona una descripción clara de las tendencias del consumo de energía y la distribución de fuentes hasta el año actual.

Además, si asumimos que la contribución de las fuentes de energía se mantiene similar a la del año 2021, podemos estimar la distribución del consumo de energía en los diferentes sectores en el futuro. Estos escenarios se basan en el supuesto de que la distribución de las fuentes de energía seguirá siendo coherente con las tendencias históricas recientes.

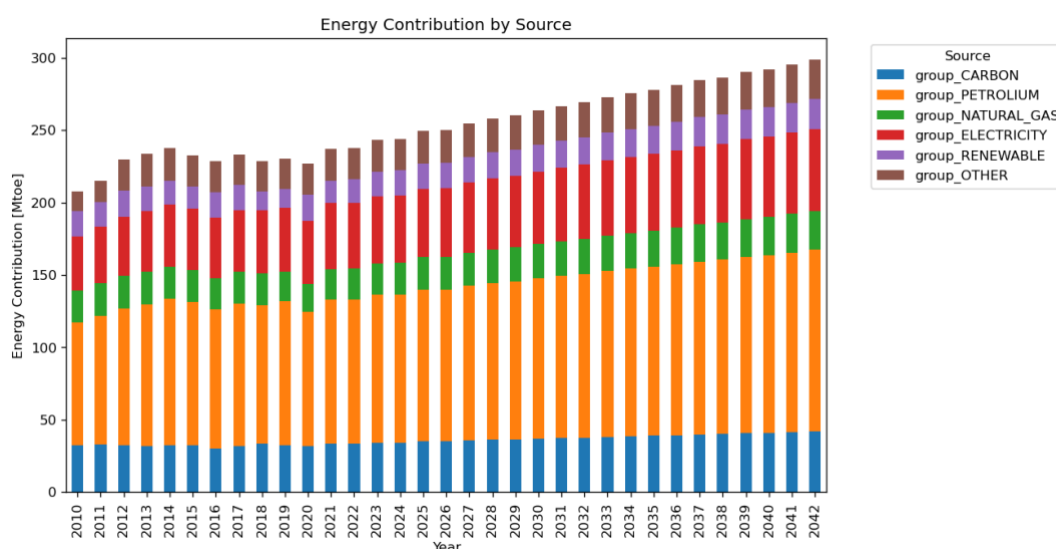


Figura 15: Predicción del consumo de energía por sector durante los próximos 20 años

Para escenarios futuros de consumo de energía (de 2022 a 2042), observamos varias tendencias clave:

- Fuentes basadas en carbono: El consumo de fuentes de energía basadas en carbono (CARBON) se mantiene relativamente estable a lo largo de los años, con una ligera tendencia al alza.
- Fuentes basadas en petróleo: El consumo de energía procedente de fuentes basadas en petróleo (PETRÓLEO) también muestra un aumento gradual, lo que refleja una demanda continua.
- Gas Natural: El consumo de gas natural (NATURAL GAS) mantiene un patrón de crecimiento consistente, contribuyendo significativamente al consumo total de energía.
- Electricidad: El consumo de electricidad (ELECTRICITY) aumenta constantemente durante el período de pronóstico, lo que indica su importancia en varios sectores.
- Energías renovables: El consumo de fuentes de energía renovables (RENEWABLE) experimenta un aumento continuo, aunque moderado, lo que pone de relieve el creciente énfasis en la energía sostenible.
- Otras fuentes: La energía de otras fuentes (OTHER) muestra una trayectoria ascendente constante, probablemente influenciada por diversas alternativas energéticas.

Estos escenarios pueden servir como herramientas valiosas para las partes interesadas involucradas en la planificación, formulación de políticas y toma de decisiones energéticas.

Asimismo, proporcionan información sobre posibles cambios en los patrones de consumo de energía, lo que permite estrategias y acciones informadas para abordar los desafíos energéticos futuros de manera efectiva.

## 6. Conclusiones

Este proyecto tenía como objetivo predecir el consumo de energía en Brasil utilizando un modelo ARIMA, el cual fue entrenado con un conjunto de datos que contiene un registro de 52 años de la matriz de balance energético publicada en el sitio web de OLADE. El conjunto de datos consiste en datos de series temporales recopilados anualmente. El modelo ARIMA fue desarrollado con el 90% de los datos para entrenamiento y el 10% para validación, y predice observaciones para un período adicional de 20 años, hasta el año 2042.

El modelo se construyó con éxito al seleccionar los parámetros (p, d, q) y utilizar el algoritmo de optimización del error porcentual absoluto medio (MAPE) para evaluar el rendimiento del modelo. Los resultados han demostrado un buen rendimiento cuando se aplican y pueden compararse con los resultados publicados en el sitio web oficial. Cabe

destacar que esta evaluación depende en gran medida de la elección de los parámetros ( $p$ ,  $d$ ,  $q$ ). Diferentes selecciones de parámetros podrían dar lugar a resultados diferentes.

### **Limitaciones e investigaciones futuras:**

En este proyecto, predecimos resultados futuros basados en registros históricos recopilados anualmente. Es posible que el conjunto de datos no sea lo suficientemente extenso cuando se utiliza un análisis univariado. Además, carecemos de información sobre otros factores que afectan al consumo de energía, como el crecimiento de la población, el PIB, el cambio climático, las crisis económicas y los avances tecnológicos en Brasil y América del Sur.

Para investigaciones futuras: Los resultados de la predicción pueden mejorarse aplicando otros modelos como LSTM, VAR, auto-ARIMA, entre otros. Además, se podría realizar una investigación para explorar el impacto de otros factores en el consumo de energía en Brasil y otros países.

## **Referencias bibliográficas/ Anexos**

1. Olade. (2021). Oferta y demanda/América Latina y el Caribe, de Olade website: <https://www.sielac.olade.org>
2. Ibge.gov. (2019). Proyecciones de población/Tablas, de Ibge.gov website: <https://www.ibge.gov.br>
3. PWC. (2022). The World in 2050, de PWC website: <https://www.pwc.com>
4. IMF. (2023). World Economic Outlook Database, de IMF website: <https://www.imf.org>
5. Knoema. (2022). Brazil Inflation Forecast 2019-2024 and up to 2060, de Knoema website: <https://knoema.com>
6. Our World in Data. (2019). Urbanization, de Our World in Data website: <https://ourworldindata.org>
7. United Nations. (2018). 2018 Revision of World Urbanization Prospects, de United Nations website: <https://population.un.org>