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:

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

Supervisor:

* Tutor

Septiembre, 2023

**Declaration of Authorship**

MRC Consultants and Transaction Advisers

Signed:

Date: September, 2023

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

Peter Drucker

**Abstract**

Global energy demand for energy consumption is increasing day by day, and it seems complicated for most countries to meet energy demand with total energy production.

In an increasingly tense world, BRICS countries－Brazil, Russia, India, China and South Africa－are very complementary with the member countries being the world's major energy producers as well as major fuel consumers.

The Brazilian energy sectors have witnessed numerous technological changes and has evolved to become a global leader in clean technology sales, both to the domestic and foreign market. A lot of factors contributed to the innovative activities in its electricity sector which includes both government and FDI contribution.

This work focuses on modeling (….) and predicting the energy consumption in Brazil ….  
  
More specifically, (…)

**ÍNDICE**

[**1. Introducción 4**](#_heading=h.gjdgxs)

[Motivación y objetivos 4](#_heading=h.30j0zll)

[Estructura del documento 4](#_heading=h.1fob9te)

[**2.  Análisis descriptivo de la demanda de energía eléctrica 4**](#_heading=h.3znysh7)

[Análisis descriptivo del conjunto de datos. 4](#_heading=h.2et92p0)

[Factores económicos / no económicos que influyen en la demanda. 4](#_heading=h.tyjcwt)

[Exploración y limpieza del conjunto de datos (tratamiento de los valores nulos).](#_heading=h.3dy6vkm) 4

Selección de variables relevantes para el estudio. [4](#_heading=h.3dy6vkm)

[Preprocesamiento de datos. 4](#_heading=h.1t3h5sf)

[**3. Desarrollo del modelo de Machine Learning 4**](#_heading=h.4d34og8)

[Introducción a los modelos desarrollados 4](#_heading=h.2s8eyo1)

[Selección el modelo óptimo 4](#_heading=h.17dp8vu)

[**4. Análisis y visualización de los resultados 4**](#_heading=h.3rdcrjn)

[Estimaciones y predicciones de la demanda de energía sectorial](#_heading=h.26in1rg) 4

Evaluación del rendimiento de los modelos estimados y análisis de los errores. [4](#_heading=h.26in1rg)

[Visualización e interactiva aplicación 4](#_heading=h.lnxbz9)

[**5. Conclusiones 4**](#_heading=h.35nkun2)

[Conclusiones obtenidas 4](#_heading=h.1ksv4uv)

[Limitaciones y futuras líneas investigaciones 4](#_heading=h.44sinio)

[**6. Recomendaciones de negocio. 4**](#_heading=h.2jxsxqh)

[**Referencias bibliográficas/ Anexos**](#_heading=h.z337ya)

**1.Introducción**

### Motivación y objetivos

En el ámbito de la ciencia de datos y el análisis de datos, la exploración y modelado de series temporales desempeñan un papel fundamental en la comprensión de patrones y tendencias en datos que evolucionan a lo largo del tiempo. Estas series temporales, que consisten en secuencias ordenadas de observaciones registradas en intervalos temporales definidos, se encuentran en diversas disciplinas, desde finanzas hasta biología y más allá.

El presente proyecto tiene como objetivo abordar la tarea de analizar y predecir patrones en series temporales, específicamente centrándose en el contexto del consumo de energía con un enfoque sectorial en Brasil. La motivación detrás de este proyecto radica en la importancia de comprender las dinámicas subyacentes y desarrollar modelos precisos para predecir el comportamiento futuro de estas series.

El estudio del consumo a nivel sectorial es esencial para el desarrollo de políticas energéticas efectivas y sostenibles. Estas políticas influyen en la planificación de fuentes de generación, el uso eficiente de recursos y la toma de decisiones en el ámbito energético. Conocer y prever la demanda de energía en diferentes sectores permitirá una asignación óptima de recursos y la implementación de medidas adecuadas para garantizar un suministro energético estable y sostenible.

Objetivos:

* El objetivo central de este proyecto es desarrollar un modelo de estimación de la demanda de energía en diferentes sectores de Brasil.

Los objetivos específicos son:

* Recopilación y Organización de Datos: Recolectar y organizar los datos pertinentes relacionados con la demanda de energía en los distintos sectores de Brasil. La calidad y coherencia de los datos son esenciales para asegurar resultados precisos y confiables.
* Análisis Exploratorio de Datos (EDA): Realizar un análisis exploratorio de datos para comprender las tendencias, patrones y relaciones presentes en los datos de la demanda energética sectorial. Esta exploración inicial proporcionará información valiosa para orientar los análisis posteriores.
* Desarrollo de Modelo de Estimación: Diseñar y construir un modelo de estimación de la demanda de energía sectorial utilizando técnicas de aprendizaje automático. Este modelo debe ser capaz de prever la demanda de energía en diferentes sectores a mediano y largo plazo.
* Validación y Evaluación del Modelo: Evaluar la precisión y eficacia del modelo utilizando técnicas de validación cruzada y métricas apropiadas para evaluar la calidad de las predicciones. Ajustar y refinar el modelo según sea necesario.
* Visualización y Comunicación de Resultados: Crear visualizaciones informativas, como gráficos y diagramas, que permitan comprender fácilmente las tendencias y proyecciones de la demanda de energía en los diversos sectores. Estas representaciones visuales facilitarán la comunicación de los resultados a los interesados y tomadores de decisiones.
* El proyecto se llevará a cabo utilizando el lenguaje de programación Python, y se implementará un modelo de estimación de demanda de energía sectorial para Brasil. Los resultados se comunicarán a través de un informe final y se resumirán en un tablero interactivo. El cronograma de trabajo detallado abarcará desde la recopilación de datos hasta la presentación de los resultados y su posterior implementación.

**2. Análisis Descriptivo de la Demanda de Energía.**

En esta sección, se llevará a cabo un análisis detallado de la demanda de energía eléctrica en el contexto de Brasil. El objetivo es obtener una comprensión profunda de los datos disponibles y las tendencias históricas en el consumo de energía.

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 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.

### Análisis descriptivo del conjunto de datos.

**Suministro Total, Transformación Total y Consumo Total en el período de estudio.**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

* **Total Supply (Suministro Total):** La línea azul representa la tendencia del Suministro Total de energía. Inicialmente, en la década de 1970, el suministro total estaba en torno a los 65,000-70,000 unidades. A lo largo de los años, hay una tendencia creciente, con algunos altibajos, llegando a un pico en los años más recientes (aproximadamente en 2021) en alrededor de 280,000 unidades.
* **Total Transformation (Transformación Total):** La línea naranja representa la tendencia de la Transformación Total de energía. En 1970, la transformación total estaba cerca de -10,000 unidades. A lo largo de los años, hay una tendencia creciente en la transformación total, llegando a un pico positivo alrededor de 2021.
* **Total Consumption (Consumo Total):** La línea verde muestra la tendencia del Consumo Total de energía. Comienza en alrededor de 60,000 unidades en 1970 y muestra un aumento gradual en los primeros años. Sin embargo, a partir de alrededor de 1980, hay un aumento más significativo en el consumo total, alcanzando un pico en alrededor de 2021 en alrededor de 240,000 unidades.

**Suministro de energía primaria en el período de estudio.**

Gráfico

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

A lo largo del tiempo, Brasil ha mostrado una diversificación en sus fuentes de energía primaria. En la década de 1970, el país dependía en gran medida de la quema de leña, así como del petróleo y sus derivados. Con el paso de los años, se ha producido un cambio significativo en la matriz energética, incorporando fuentes como el gas natural, la hidroenergía y el carbón.

En el **primer gráfico**, que representa el suministro de energía primaria a lo largo de los años, se destacan varias tendencias clave. En las décadas iniciales, el petróleo (OIL) y el gas natural (NATURAL GAS) fueron las principales fuentes de energía primaria, con un suministro que continuó aumentando a lo largo del tiempo. A partir de la década de 1980, se observa un marcado aumento en la contribución del gas natural, lo que sugiere su creciente importancia en la matriz energética del país.

El **segundo gráfico**, que muestra la contribución porcentual de cada fuente de energía primaria con respecto al suministro total, revela aún más detalles sobre la transformación de la matriz energética de Brasil. A medida que avanzan los años, el gas natural emerge como un componente vital, mostrando un aumento constante en su porcentaje de contribución. Esta tendencia sugiere que Brasil ha dependido cada vez más del gas natural para satisfacer sus necesidades energéticas.

Aunque se ha observado una diversificación en las fuentes de energía, el petróleo y sus derivados, así como el gas natural, siguen siendo componentes fundamentales en la matriz energética de Brasil. Ambos siguen representando una parte significativa del suministro total.

La energía nuclear y otras fuentes no especificadas de energía primaria también tienen una presencia notable en la tabla. Sin embargo, su participación en el suministro total sigue siendo relativamente baja.

En resumen, los gráficos resaltan cómo Brasil ha transitado hacia una matriz energética más sostenible y diversificada a lo largo del tiempo, con un aumento en la participación de la hidroenergía y un crecimiento moderado en el uso del gas natural, así como un decrecimiento moderado en el uso del petróleo. Estos cambios indican una estrategia de diversificación energética en Brasil en busca de una mayor seguridad energética y una menor dependencia de los combustibles fósiles, aunque es fundamental seguir avanzando hacia fuentes de energía verdaderamente limpias y renovables.

**Evolución de Importación y Exportación de Fuentes Primaria en el período de estudio.**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

De 1970 a 2000, Brasil experimentó una transformación en su posición energética, pasando de ser un país importador de energía a convertirse en un importante exportador neto de energía en algunos aspectos. Aquí hay un resumen de cómo evolucionó la situación energética de Brasil durante ese período:

* **Hasta la década de 1970:**

Brasil dependía en gran medida de la importación de petróleo y derivados para satisfacer sus necesidades energéticas. La economía brasileña estaba en crecimiento, lo que aumentaba la demanda de energía y, por lo tanto, la dependencia del petróleo importado.

* **Década de 1970:**

En la década de 1970, Brasil comenzó a enfrentar problemas con la inflación y la deuda externa, que se agravaron por el alto costo de las importaciones de petróleo. Como respuesta a estos desafíos, Brasil inició programas para reducir su dependencia del petróleo importado y diversificar su matriz energética. Década de 1980:

Durante esta década, Brasil implementó políticas para promover la producción interna de petróleo y gas. Se descubrieron y desarrollaron importantes yacimientos de petróleo en la cuenca de Campos, lo que contribuyó a aumentar la producción de petróleo en el país. Década de 1990:

Brasil continuó avanzando en la explotación de sus recursos energéticos, incluidos el petróleo y la energía hidroeléctrica. El país también comenzó a desarrollar biocombustibles, especialmente etanol derivado de la caña de azúcar, como una alternativa más sostenible a los combustibles fósiles.

* **Para la década de 2000**:

Brasil había avanzado en la producción interna de petróleo y gas, lo que le permitió reducir significativamente sus importaciones de petróleo. El país también dependía en gran medida de la energía hidroeléctrica, con la construcción de grandes represas que le proporcionaron una fuente significativa de energía renovable. Además, Brasil se convirtió en un exportador neto de petróleo y productos energéticos, gracias a su creciente producción interna y la explotación de sus recursos naturales. En resumen, a lo largo de las décadas de 1970 a 2000, Brasil pasó de ser un país importador de energía a ser un exportador neto en ciertos aspectos, particularmente en lo que respecta al petróleo y sus derivados. La diversificación de su matriz energética y el desarrollo de sus recursos naturales jugaron un papel crucial en esta transformación.

**Suministro de energía secundaria en el período de estudio.**

Gráfico, Histograma

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

* Primer Gráfico - Energía Secundaria (Cantidades Absolutas):

El **primer gráfico** de área apilada muestra la evolución de la energía secundaria suministrada en Brasil a lo largo del tiempo en cantidades absolutas. Cada área coloreada representa un tipo de fuente de energía secundaria, y la altura total de la pila en cada año representa el suministro total de energía secundaria.

En las décadas iniciales (1970-1980), la mayoría de las fuentes de energía secundaria experimentaron un crecimiento constante. La electricidad (ELECTRICITY) y los derivados del petróleo, como el diésel (DIESEL OIL) y la gasolina (GASOLINE/ALCOHOL), fueron las principales contribuyentes al suministro de energía secundaria.

A medida que avanzan los años, hay un aumento significativo en el suministro de electricidad y productos derivados del petróleo, como el gasóleo y la gasolina. Sin embargo, a partir de mediados de la década de 1990, se observa una tendencia decreciente en el suministro de derivados del petróleo y un aumento en la contribución de otras fuentes de energía secundaria, como el carbón (COKE) y la biomasa (CHARCOAL).

* Segundo Gráfico - Porcentaje de Energía Secundaria:

El **segundo gráfico** de área apilada muestra la composición de la matriz de energía secundaria de Brasil en términos de porcentaje de cada fuente con respecto al suministro total de energía secundaria en cada año.

Durante las primeras décadas, los derivados del petróleo y la electricidad dominaron la matriz de energía secundaria, con un enfoque particular en el diésel y la gasolina. Sin embargo, a medida que avanzan los años, se pueden observar varios cambios en la composición de la matriz energética.

A finales de la década de 1990 y principios de la década de 2000, hay una disminución en la participación de los derivados del petróleo, mientras que otras fuentes de energía, como la electricidad y el carbón, mantienen una presencia significativa. Además, a medida que avanzan los años, se observa un aumento en la participación de fuentes de energía más limpias y renovables, como la biomasa y la electricidad.

En resumen, los gráficos muestran una transformación en la matriz de energía secundaria de Brasil a lo largo del tiempo. Se ha producido una disminución gradual en la dependencia de los derivados del petróleo y un aumento en la participación de fuentes de energía más diversificadas, incluidas las renovables. Esto sugiere una tendencia hacia una matriz energética más sostenible y diversificada en Brasil, con un enfoque en la reducción de la dependencia de los combustibles fósiles.

**Consumo de energía por sectores en el período de estudio***.*

Gráfico

Descripción generada automáticamente

Gráfico

Descripción generada automáticamente

* Gráfico 1: Energy Consumption by Sector Over the Years

Este gráfico representa la distribución del consumo de energía en diferentes sectores a lo largo de los años. Cada área coloreada en el gráfico corresponde a un sector específico, como Agricultura, Industria, Transporte, Residencial, Comercial y otros. La altura de cada área en un año determinado muestra la cantidad total de energía consumida por ese sector en ese año.

* Agricultura: Durante los primeros años, la agricultura tenía una participación significativa en el consumo de energía, pero a medida que avanzó el tiempo, su contribución disminuyó.
* Industria: La industria es uno de los principales consumidores de energía y su contribución ha sido constante, con algunos aumentos notables en ciertos años.
* Transporte: El consumo de energía en el sector del transporte ha aumentado de manera constante a lo largo de los años, lo que indica un crecimiento en la movilidad y la demanda de combustibles.
* Residencial: La energía utilizada en los hogares ha aumentado, posiblemente debido al aumento de la población y las necesidades de calefacción, refrigeración y electrónica.
* Comercial y Otros: Este sector incluye edificios comerciales y otros usos diversos de energía. Su contribución generalmente se mantiene constante.
* Gráfico 2: Percentage of Energy Consumption by Sector Over the Years

Este gráfico muestra la proporción de consumo de energía que cada sector representa en relación con el total. En lugar de valores absolutos, se centra en la distribución relativa en forma de porcentaje.

* A lo largo del tiempo, la industria sigue siendo una parte significativa del consumo total de energía.
* El transporte también ha aumentado en términos de porcentaje, lo que indica un aumento en la importancia y demanda del transporte en el consumo de energía.
* El sector residencial muestra una disminución relativa en comparación con otros sectores, lo que sugiere un cambio en la composición del consumo de energía.
* El sector agrícola ha disminuido constantemente en términos de porcentaje.
* Comercial y Otros se mantiene relativamente constante en términos de porcentaje.

Exploración y limpieza del conjunto de datos (tratamiento de los valores nulos).

El dataset que hemos utilizado para este trabajo se trata del balance de la matriz energética desde 1970 a 2021 de Brasil descargado de la página web de OLADE.

Los datos de la matriz energética los tenemos contenidos en un archivo Excel con múltiples hojas llamado "Brazil\_Energy balance matrix.xlsx".

Realizamos acciones como redondear valores, renombrar columnas, eliminar espacios y corregir nombres. Además, agregamos el año como una columna. Convertimos las unidades de medida de Ktoe (kilotoneladas equivalentes de petróleo) a Mtoe (megatoneladas equivalentes de petróleo) en todas las columnas numéricas. Los valores numéricos se redondean a dos decimales. Se eliminan los encabezados de fila que contienen unidades de medida, y se renombran las columnas 'Unnamed: 0' a 'SECTOR' y 'OTHER PRIMARY\_x000d\_' a 'OTHER PRIMARY'. Además, se corrige el nombre del sector 'COKE PLANTS AND BLAST FURNACES\_x000d\_' a 'COKE PLANTS AND BLAST FURNACES'. Estas transformaciones aseguran que los datos estén en un formato coherente y listos para su análisis posterior. Los datos procesados se almacenan en un nuevo diccionario llamado modified\_dict, que reemplaza el original. Finalmente, se guarda el diccionario limpio en un nuevo archivo Excel llamado "Brazil\_Energy balance matrix\_cleaned.xlsx", preservando las hojas originales para su análisis y visualización posteriores. Este proceso asegura la calidad y coherencia de los datos antes de su utilización en análisis más profundos.

Después de esta conversión empezamos a analizar nuestro dataset.

**3. Desarrollo del modelo de predicción**

En la sección de "Desarrollo del modelo de Machine Learning," se abordará la parte central de este proyecto de análisis de energía en Brasil. El objetivo principal en esta fase es construir un modelo capaz de realizar pronósticos precisos sobre el consumo energético en el país. En este contexto, se ha elegido el modelo ARIMA (Autoregressive Integrated Moving Average) como enfoque para abordar esta tarea. La elección de ARIMA se basa en diversas razones fundamentales que respaldan su idoneidad para modelar y predecir series temporales, como la demanda de energía.

Esas razones consisten en:

1. **Adecuado para Series Temporales**: ARIMA es ampliamente reconocido y utilizado en la modelización de series temporales, lo que lo hace altamente relevante para el pronóstico de datos que varían con el tiempo, como el consumo de energía.
2. **Capacidad para Capturar Patrones Temporales**: El modelo ARIMA es eficaz para capturar patrones, tendencias y estacionalidades en los datos temporales, lo que lo hace apropiado para predecir la demanda de energía que puede mostrar variaciones a lo largo del año.
3. **Estacionalidad y Tendencias**: Permite la identificación y tratamiento de componentes estacionales y tendencias en los datos, lo que es crucial para el análisis de la demanda de energía, que a menudo exhibe patrones estacionales y cambios a lo largo del tiempo debido a factores económicos y estacionales.
4. **Comportamiento de Series no Estacionaria**s: ARIMA puede manejar series temporales no estacionarias mediante la diferenciación, lo que es común en datos energéticos que pueden mostrar cambios a lo largo del tiempo.
5. **Base en Principios Estadísticos**: Este modelo se basa en principios estadísticos sólidos, lo que lo convierte en una elección confiable y respaldada por la teoría.

Introducción a los modelos desarrollados

**Selección del modelo óptimo**

Los modelos ARIMA se aplicarán a todos los sectores de consumo energético en este análisis. Sin embargo, antes de implementar el modelo, es esencial realizar una selección cuidadosa de los parámetros óptimos (p, d, q) para garantizar que el modelo se ajuste adecuadamente a las características de las series temporales no estacionarias de consumo de energía en Brasil.

Hemos decidió no escalar las variables de nuestro dataframe debido a las siguientes casuísticas:

1. **Naturaleza de las Series Temporales:** Las series temporales ya representan datos en función del tiempo, lo que significa que las unidades y las magnitudes de las variables pueden tener sentido en su forma original. En el caso del consumo de energía, las unidades generalmente son medidas físicas como en nuestro caso, megatoneladas equivalentes de petróleo (Mtoe), y estas unidades son inherentemente relevantes para la interpretación de los datos.

1. **Conservación de la Interpretabilidad:** Al no escalar las variables, los valores en las predicciones del modelo ARIMA se mantendrán en la misma escala y unidades que los datos originales. Esto facilita la interpretación de los resultados, ya que los pronósticos seguirán siendo comprensibles y significativos en términos del consumo real de energía, lo que es fundamental para los responsables de la toma de decisiones y los expertos en energía.
2. **Sensibilidad al Contexto:** Escalar las variables puede afectar la magnitud de los coeficientes en el modelo y, en consecuencia, las estimaciones de las predicciones. En un contexto como el análisis de energía, donde las cifras reales son críticas para la planificación y la política, realizar un escalado podría introducir cierta opacidad en la interpretación de los resultados.
3. **Enfoque Tradicional de ARIMA:** El modelo ARIMA se desarrolló originalmente para trabajar con series temporales en su forma original, sin la necesidad de escalamientos adicionales. Esta metodología se basa en la idea de que los patrones y las relaciones inherentes en los datos se pueden modelar sin requerir escalas uniformes.

Para asegurar que nuestros modelos ARIMA sean robustos y precisos en la predicción de la demanda de energía, se realizarán análisis exhaustivos para determinar la mejor combinación de parámetros. Esto permitirá que el modelo se ajuste de manera óptima a las variaciones y patrones específicos presentes en los datos de consumo de energía de Brasil, lo que, a su vez, respaldará decisiones fundamentadas en políticas y planificación energética.

En el proceso de selección del modelo óptimo en el contexto de modelos ARIMA aplicados al análisis de series temporales de consumo de energía para cada sector en Brasil, se utilizan varias funciones fundamentales. Estas funciones desempeñan un papel crítico en la evaluación y ajuste de los modelos para garantizar que sean apropiados para las características específicas de los datos, las cuales son:

1. **Función MAPE (Error Porcentual Absoluto Medio):**

La función **MAPE** se utiliza para calcular el Error Porcentual Absoluto Medio (MAPE) entre las observaciones reales y las predicciones generadas por un modelo ARIMA. El MAPE proporciona una medida de la precisión de las predicciones expresada como un porcentaje del error con respecto a las observaciones reales. Un MAPE más bajo indica predicciones más precisas y es una métrica esencial para la selección del modelo óptimo.

1. **Función test\_stationarity (Prueba de Dickey-Fuller para Estacionariedad):**

La función test\_stationarity se emplea para evaluar si una serie temporal es estacionaria o no. La estacionariedad es un requisito fundamental para aplicar modelos ARIMA, ya que estos modelos asumen que la estadística de la serie no cambia con el tiempo.El Test de Dickey-Fuller comprueba si el valor p es lo suficientemente bajo como para rechazar la hipótesis nula de no estacionariedad.

1. **Función residcheck (Evaluación de Residuos):**

La función residcheck se utiliza para evaluar los residuos generados por un modelo ARIMA. Los residuos son las diferencias entre las observaciones reales y las predicciones del modelo. Esta función verifica si los residuos cumplen con condiciones deseables, como ser ruido blanco (sin patrones discernibles), tener una media cercana a cero y una varianza constante. También evalúa la autocorrelación, normalidad y estacionariedad de los residuos para garantizar que el modelo haya capturado adecuadamente la información de los datos.

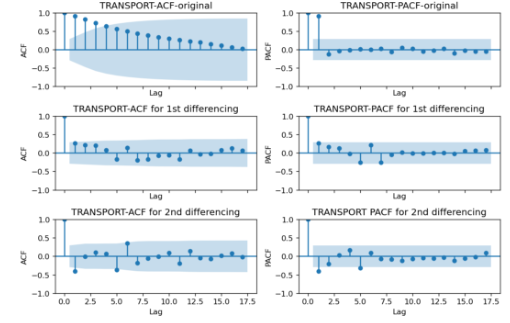
1. **Función order\_aic\_bic (Selección de Parámetros p y q):**

Esta función realiza un bucle a través de diferentes valores de 'p' (orden de autoregresión) y 'q' (orden de media móvil) en un rango de 0 a 4. Para cada combinación de 'p' y 'q', se ajusta un modelo ARIMA a los datos y se calculan los valores de AIC y BIC. Los resultados se almacenan en una lista junto con los valores de 'p' y 'q', y luego se presenta una tabla de modelos ordenados según los valores más bajos de AIC y BIC.

**Búsqueda del valor “d”:**

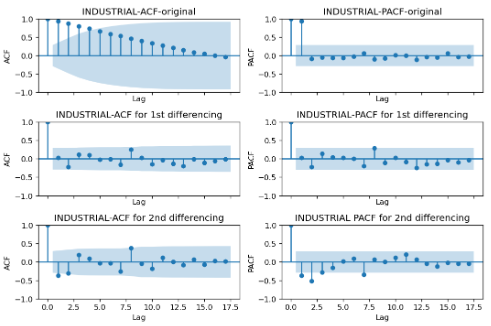
Realizamos una búsqueda para la selección apropiada del valor de 'd' (orden de diferenciación) para cada sector. Para esto, utilizaremos la prueba de estacionariedad de Dickey-Fuller hasta que se alcanza la estacionariedad (cuando el p-valor es menor que 0.05).

Los cuáles nos dan los siguientes resultados para cada sector:

**TRANSPORT:**

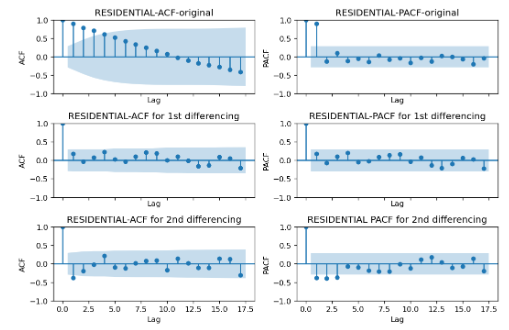
* Test Statistic (Estadístico de Prueba) - 2.756864 (1st diff), -3.444595 (2nd diff)
* p-value (Valor p): 0.064704 (1st diff), 0.009529 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**INDUSTRIAL:**

****

* Test Statistic: -5.317197 (1st diff), -3.912924 (2nd diff)
* p-value: 0.000005 (1st diff), 0.001942 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**RESIDENTIAL:**

****

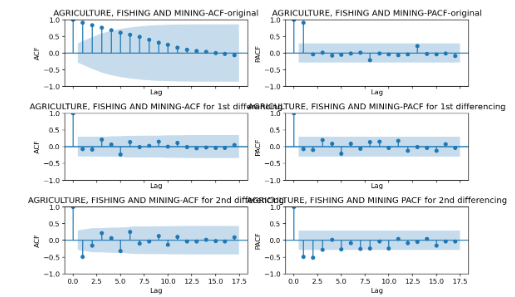
* Test Statistic: -5.425666 (1st diff), -7.081189 (2nd diff)
* p-value: 0.000003 (1st diff), 4.661825e-10 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**COMMERCIAL, SERVICES, PUBLIC:**

****

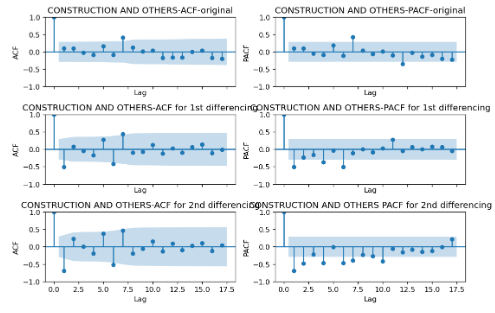
* Test Statistic: -5.187659 (1st diff), -5.083537 (2nd diff)
* p-value: 0.000009 (1st diff), 0.000015 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**AGRICULTURE, FISHING AND MINING:**

****

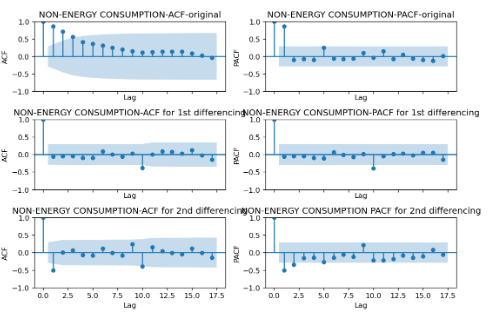
* Test Statistic: -6.949322e+00 (1st diff), -6.809117e+00 (2nd diff)
* p-value: 9.785090e-10 (1st diff), 2.138701e-09 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**CONSTRUCTION AND OTHERS:**

****

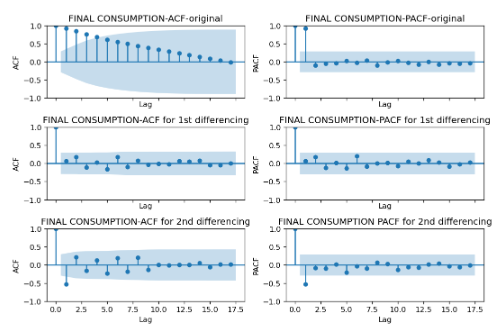
* Test Statistic: -5.843043 (1st diff), -5.001523 (2nd diff)
* p-value: 3.738493e-07 (1st diff), 0.000022 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**NON-ENERGY CONSUMPTION:**

****

* Test Statistic: -2.172822 (1st diff), -2.724450 (2nd diff)
* p-value: 0.216257 (1st diff), 0.069902 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

**FINAL CONSUMPTION:**

****

* Test Statistic: -1.457273 (1st diff), -4.365760 (2nd diff)
* p-value: 0.554532 (1st diff), 0.000341 (2nd diff)
* Conclusión: La segunda diferenciación (2nd diff) hace que la serie sea estacionaria con un p-value menor que 0.05. Se selecciona d = 2.

Para todos los sectores, la segunda diferenciación (d = 2) hace que las series sean estacionarias, lo que indica que se aplicará una diferenciación de orden 2 en la modelización ARIMA de cada sector para lograr la estacionariedad requerida.