



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

**ANÁLISIS TÉCNICO DE LA CAPACIDAD DE PRODUCCIÓN DE ENERGÍA
ELÉCTRICA NETA EN ECUADOR PARA SATISFACER LA DEMANDA DEL
SECTOR PRODUCTIVO EN UN PERIODO DE 10 AÑOS.**

**Profesor
Ing. Mario González, Ph.D.**

**Autor
Nicolás Felipe Araque Araque
Paola María Mancheno Domínguez**

2023

RESUMEN

El presente proyecto tuvo como objetivo predecir el volumen de energía producida y la eficiencia en su suministro.

En primer lugar, se describieron los factores clave que contribuyen a mejorar la eficacia del servicio eléctrico, como la implementación de tecnologías eficientes y la optimización de la distribución de energía. A continuación, se comparó el rendimiento del modelo Random Forest con otros modelos de aprendizaje supervisado, demostrando su efectividad en la predicción con un alto coeficiente de determinación y un bajo error medio absoluto.

Además, se identificaron las empresas productoras y distribuidoras de energía eléctrica en el Ecuador, así como su capacidad actual de producción. Esto proporcionó una visión clara de los actores clave en el sector eléctrico. Asimismo, se evaluaron las tecnologías de generación de energía disponibles en el mercado y su potencial de aplicación en el contexto ecuatoriano.

Finalmente, se seleccionaron las opciones más viables y sostenibles para aumentar la capacidad de producción de energía eléctrica neta en el Ecuador. Se destacó la importancia de utilizar tecnologías renovables, mejorar la eficiencia energética y diversificar las fuentes de energía.

Los hallazgos presentados en el presente proyecto son valiosos para la toma de decisiones estratégicas en la optimización de la generación y distribución de energía en Ecuador, mejorando la eficiencia y competitividad del sector eléctrico.

Palabras clave: volumen de energía, eficiencia, predicción, modelo de aprendizaje automático, Random Forest, capacidad de producción, tecnologías de generación, competitividad.

ABSTRACT

The aim of this project was to predict the volume of energy production and its efficiency in the supply. Firstly, key factors contributing to improving the effectiveness of the electricity service were described, such as the implementation of efficient technologies and the optimization of energy distribution. Next, the performance of the Random Forest model was compared with other supervised learning models, demonstrating its effectiveness in prediction with a high coefficient of determination and low mean absolute error.

In addition, electricity-producing and distributing companies in Ecuador were identified, along with their current production capacity. This provided a clear understanding of the key players in the electricity sector. Furthermore, available energy generation technologies in the market were evaluated, considering their potential application in the Ecuadorian context.

Finally, the most viable and sustainable options to increase net electricity production capacity in Ecuador were selected. The importance of using renewable technologies, improving energy efficiency, and diversifying energy sources was highlighted.

The findings presented in this project are valuable for strategic decision-making in optimizing energy generation and distribution in Ecuador, ultimately enhancing the efficiency and competitiveness of the electricity sector.

Key words: energy volume, efficiency, prediction, machine learning model, Random Forest, production capacity, generation technologies, competitiveness.

ÍNDICE DEL CONTENIDO

1.	INTRODUCCIÓN	1
2.	REVISIÓN DE LITERATURA.....	3
3.	IDENTIFICACIÓN DEL OBJETO DE ESTUDIO	8
4.	PLANTEAMIENTO DEL PROBLEMA	9
5.	OBJETIVO GENERAL	10
6.	OBJETIVOS ESPECÍFICOS.....	10
7.	JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA	11
7.1.	RECOLECCION DE DATOS.....	12
7.2.	LIMPIEZA, PRE-PROCESAMIENTO Y/O TRANSFORMACIÓN DE DATOS	13
7.3.	IDENTIFICACIÓN Y DESCRIPCIÓN DE VARIABLES	16
7.4.	VISUALIZACIÓN DE VARIABLES	22
7.5.	SELECCIÓN DE MODELO ESTADÍSTICO	30
8.	RESULTADOS.....	35
8.1.	ANÁLISIS DE MODELO ESTADÍSTICO E INTERPRETACIÓN DE RESULTADOS	35
8.1.1.	MODELOS DE REGRESIÓN	35
8.1.2.	MODELO DE CLASIFICACIÓN.....	41
9.	PROPUESTA DE SOLUCIÓN	47
9.1.	IMPLICACIONES ORGANIZACIONALES.....	47
9.1.1.	IMPLICACIONES ORGANIZACIONALES SOBRE INNOVACIÓN EMPRESARIAL.....	48
9.2.	ESTRATEGIA ORGANIZACIONAL.....	50
9.3.	BENEFICIO ECONÓMICO.....	54
10.	CONCLUSIONES Y RECOMENDACIONES.....	57

10.1.	CONCLUSIONES	57
10.2.	RECOMENDACIONES	58
	REFERENCIAS	62

ÍNDICE TABLAS

Tabla 1. Diccionario variables base de datos 1	14
Tabla 2. Diccionario variables base de datos 2	15
Tabla 3. Conjunto de variables a usar para la base de datos 1.....	17
Tabla 4. Conjunto de variables a usar para la base de datos 3.....	17
Tabla 5. Estadísticos descriptivos de las variables en el conjunto de datos 1..	19
Tabla 6. Estadísticos descriptivos de las variables en el conjunto de datos 3..	19
Tabla 7. Resultados de los modelos de aprendizaje supervisado.....	35
Tabla 8. Informe de clasificación	45

ÍNDICE DE FIGURAS

Figura 1. Matriz de correlación base de datos 1.....	20
Figura 2. Matriz de correlación base de datos 3.....	21
Figura 3. Ejemplo de gráfico de cajas variables continuas base de datos 1	23
Figura 4. Gráfico densidad energía neta producida.	23
Figura 5. Gráfico de densidad variable predictora uso de Fueloil_gal.....	24
Figura 6. Gráfico de densidad variable predictora uso de Gas_Natural_KPC..	24
Figura 7. Gráfico de barras de centrales / tipo de centrales con mayor producción de energía.....	25
Figura 8. Gráfico de barras empresas que producen mayor cantidad de energía.	25
Figura 9. Gráfico de barras tipos de empresas que producen una mayor cantidad de energía.....	26
Figura 10. Ejemplo de gráfico de cajas variables continuas base de datos 3 ..	27
Figura 11. Gráfico densidad valor del indicador de eficacia a la conexión del servicio.	27
Figura 12. Gráfico densidad uso de Diesel2_gal.....	28
Figura 13. Gráfico de barras centrales con el indicador promedio más alto....	28
Figura 14. Gráfico de barras empresas con el indicador promedio más alto....	29
Figura 15. Gráfico de barras con los tipos de fuente de energía que en promedio se tuvieron mejores indicadores.	29
Figura 16. Gráfico de dispersión conjunto de datos 1.	37
Figura 17. Gráfico de dispersión conjunto de datos 3.	38
Figura 18. Importancia de las características conjunto de datos 1.....	39
Figura 19. Importancia de las características conjunto de datos 3.....	40
Figura 20. Importancia de las características conjunto de datos 3, modelo clasificación.	42
Figura 21. Gráfico de dispersión, clasificación de empresas por volumen de energía neto producido y valor de indicador de eficacia de conexión al servicio.	44

1. INTRODUCCIÓN

La producción y el consumo de energía son elementos clave para el desarrollo económico de un país. En el caso de Ecuador, la energía eléctrica es una de las principales fuentes de energía utilizadas en la producción industrial y en la vida cotidiana de los ciudadanos. Sin embargo, existe una creciente preocupación sobre la capacidad de producción de energía eléctrica neta para abastecer la demanda del sector productivo y de la población en general (Ministerio de Energía y Minas, 2023).

Según datos del Ministerio de Energía y Recursos Naturales No Renovables (MERNNR), la demanda de energía eléctrica en Ecuador ha aumentado de forma constante en los últimos años. En el 2019, la demanda total de energía eléctrica fue de 28.277 GWh, mientras que en el 2020 se registró una demanda de 28.990 GWh, lo que representa un aumento del 2,5% respecto al año anterior. Además, se prevé que la demanda de energía eléctrica en el país continúe aumentando en los próximos años (Ministerio de Energía y Recursos Naturales No Renovables, 2021).

Sin embargo, la capacidad de producción de energía eléctrica neta en el país es limitada y no ha logrado mantenerse al mismo ritmo que la demanda. Según un informe del Banco Interamericano de Desarrollo (BID), en los últimos años se ha registrado una brecha entre la oferta y la demanda de energía eléctrica en el país. Esta brecha se ha traducido en problemas de suministro en algunas regiones del país, lo que ha afectado negativamente a la economía y a la calidad de vida de los ciudadanos (Banco Interamericano de Desarrollo, 2018).

Además, la matriz energética del país depende en gran medida de los combustibles fósiles, lo que genera impactos ambientales y de sostenibilidad a largo plazo. Según el MERNNR, en el 2020, el 62,6% de la energía eléctrica generada en el país provino de fuentes térmicas, mientras que el 35,5% provino de fuentes hidroeléctricas y el 1,9% de otras fuentes (Ministerio de Energía y Recursos Naturales No Renovables, 2021).

Ante esta problemática, se requiere una evaluación rigurosa de las opciones disponibles para aumentar la capacidad de producción de energía eléctrica neta en el país, tomando en cuenta aspectos técnicos y económicos. Un enfoque analítico permitirá identificar las soluciones más viables y sostenibles para la organización y el país en su conjunto.

En este sentido, CAPSTONE es un proyecto de investigación que se lleva a cabo en el contexto de la Maestría en Inteligencia de Negocios y Ciencia de Datos en Ecuador a través del cual se busca estudiar la problemática organizacional de la falta de capacidad de producción de energía eléctrica neta en Ecuador y proponer soluciones viables y sostenibles para abordarla. A través de una revisión exhaustiva de la literatura académica y de fuentes oficiales pertinentes, se busca obtener información y datos relevantes para entender la magnitud del problema y las posibles soluciones, así como también se realizará una evaluación de las prácticas actuales en la producción y distribución de energía eléctrica en el país y finalmente se realizarán análisis estadísticos y modelos de predicción para evaluar diferentes escenarios y estrategias.

El resultado final del proyecto será un conjunto de recomendaciones y estrategias para aumentar la capacidad de producción de energía eléctrica en Ecuador, lo que permitirá abastecer la creciente demanda del sector productivo y mejorar la competitividad de las empresas en el país.

2. REVISIÓN DE LITERATURA

La energía es una necesidad básica para el desarrollo económico y social, y se puede obtener de diversas fuentes primarias como los recursos naturales. Según el Panel Intergubernamental sobre Cambio Climático (IPCC, 2011), la energía obtenida directamente de los recursos naturales se denomina fuente primaria, mientras que el resultado de la transformación de la energía obtenida a partir de los recursos presentes en la naturaleza se denomina fuente secundaria.

A su vez, las fuentes de energía se clasifican en renovables y no renovables. Las renovables son aquellas cuyas reservas se producen indefinidamente, y las no renovables son aquellas cuyas reservas se agotan. Entre las fuentes de energía primaria figuran las no renovables, como el petróleo, carbón y gas natural, y las renovables, como el viento, agua, sol y biomasa (IPCC, 2011).

A nivel mundial, el consumo de energía secundaria procedente de refinerías de petróleo y centrales de generación eléctrica aumenta continuamente (IPCC, 2011). Es importante tener en cuenta que el uso de fuentes no renovables de energía puede tener impactos negativos en el medio ambiente y la salud humana, por lo que se promueve el uso de fuentes renovables como alternativa más sostenible y limpia.

En Ecuador, la matriz energética depende en gran medida de los combustibles fósiles, principalmente petróleo y gas natural. Según el Ministerio de Energía y Recursos Naturales No Renovables de Ecuador, en 2020 el petróleo representó el 50,5% de la oferta de energía primaria del país, seguido del gas natural con el 34,1%, la biomasa con el 13,9% y el resto de las fuentes renovables con el 1,5% (Ministerio de Energía y Recursos Naturales No Renovables, 2021). De la misma manera, en el informe del Banco Interamericano de Desarrollo (BID) titulado "Ecuador: energía para el desarrollo sostenible", se menciona que la matriz energética del país está altamente concentrada en combustibles fósiles, siendo el petróleo y el gas natural los principales recursos utilizados para la generación

de energía eléctrica. Esto ha generado una vulnerabilidad en el suministro y un impacto ambiental significativo (BID, 2016).

En un estudio realizado por la Comisión Económica para América Latina y el Caribe (CEPAL) sobre la matriz energética en América Latina, se identifica que la dependencia de los combustibles fósiles sigue siendo muy elevada en la región, lo que representa un riesgo para la seguridad energética y la sostenibilidad a largo plazo (CEPAL, 2017). Además, se destaca la importancia de impulsar la transición hacia fuentes de energía renovable y mejorar la eficiencia energética para lograr un desarrollo más sostenible.

En cuanto a la generación de electricidad, según el mismo informe del Ministerio de Energía y Recursos Naturales No Renovables, en 2020 la hidroelectricidad representó el 62,7% de la generación total de electricidad, seguida de la generación térmica con el 34,5% y la generación eólica con el 2,7%. En este sentido, es importante destacar que Ecuador cuenta con un gran potencial de generación de energía renovable, especialmente de origen hidroeléctrico y eólico, lo que podría ayudar a diversificar la matriz energética del país y reducir su dependencia de los combustibles fósiles (Organización Latinoamericana de Energía, 2019).

Cabe mencionar que en general, la situación de la demanda energética en Ecuador ha sido influenciada por diversos factores, como el crecimiento económico y poblacional, la infraestructura disponible, las políticas energéticas, entre otros. Sin embargo, en las últimas décadas, la demanda energética en Ecuador ha ido en aumento debido principalmente al crecimiento económico y poblacional. Según datos del Ministerio de Energía y Recursos Naturales No Renovables de Ecuador, la demanda de energía eléctrica en el país aumentó de 12.034 GWh en 2008 a 23.393 GWh en 2019, lo que representa un incremento del 94% en 11 años. Se espera que esta tendencia de aumento en la demanda energética continúe en los próximos años (Ministerio de Energía y Recursos Naturales No Renovables, 2021).

En este sentido, el Plan Nacional de Energía 2016-2025 del Gobierno de Ecuador proyecta un aumento de la demanda de energía eléctrica a un promedio anual del 3,3% entre 2016 y 2025. Según este plan, se espera que la demanda de energía eléctrica alcance los 37.325 GWh en 2025 (Ministerio de Energía y Recursos Naturales No Renovables, 2016).

En este punto, es importante destacar que la pandemia de COVID-19 ha afectado la demanda de energía en el país y a nivel mundial en el año 2020, y se espera que esta situación tenga un impacto en la demanda energética en el corto plazo. Sin embargo, a largo plazo, se espera que la demanda de energía siga creciendo en línea con el desarrollo económico y poblacional del país (Ministerio de Energía y Recursos Naturales No Renovables, 2021).

Es por ello que surge la necesidad de mejorar la eficiencia energética en el sector productivo de Ecuador, con el fin de reducir el consumo de energía y mejorar la competitividad de las empresas a través de la búsqueda de oportunidades para implementar tecnologías y procesos más eficientes en sectores como la industria manufacturera y la construcción, como por ejemplo, un artículo publicado en la revista *Renewable and Sustainable Energy Reviews* muestra que la energía solar es una de las fuentes renovables de mayor potencial en América Latina, especialmente en países con alta irradiación solar como Ecuador, sin embargo, se requiere una inversión significativa en infraestructura y tecnología para aprovechar este potencial (Huacón-Tapia et al, 2019). Otro ejemplo de esta búsqueda de eficiencia energética se muestra en el "Plan Nacional de Energía 2021-2040" del Ministerio de Energía y Recursos Naturales No Renovables (2021), en donde se establecen metas específicas para la generación de energía eléctrica a partir de fuentes renovables, como la solar y la eólica.

Desde el punto de vista técnico y económico, se han realizado diversos estudios en el país para analizar la situación actual de la demanda energética y las posibles soluciones para garantizar un suministro estable y sostenible de energía. Algunos de estos estudios han utilizado metodologías como el análisis

de datos y modelos económicos para analizar la demanda energética y las fuentes de energía disponibles en el país, mismos que, ha considerado los siguientes puntos para su desarrollo (en el Anexo 1, se puede observar la matriz con el detalle de investigaciones similares):

Metodologías para el análisis de datos:

- Análisis de tendencias históricas: se analizan los patrones de consumo de energía en el pasado para predecir la demanda futura.
- Análisis de correlación: se identifican las relaciones entre la demanda de energía y otros factores, como el clima, la economía y la población, para predecir la demanda futura.
- Modelos de simulación: se utilizan modelos matemáticos para simular diferentes escenarios de demanda de energía y evaluar el impacto de las políticas y estrategias energéticas en el futuro.
- Análisis de costo-beneficio: se evalúa la relación entre los costos y beneficios de diferentes estrategias energéticas y se identifican las mejores opciones para la toma de decisiones.

Tipos de datos utilizados:

- Datos históricos de consumo de energía: se utilizan datos de consumo de energía en el pasado para predecir la demanda futura.
- Datos socioeconómicos: se utilizan datos de población, economía y otros factores socioeconómicos para predecir la demanda futura.
- Datos meteorológicos: se utilizan datos meteorológicos para predecir la demanda de energía en diferentes condiciones climáticas.
- Datos de políticas energéticas: se utilizan datos de políticas energéticas pasadas y actuales para evaluar el impacto de diferentes estrategias energéticas en el futuro.

Implicaciones gerenciales:

- Identificación de oportunidades para la eficiencia energética: los estudios de demanda energética pueden identificar oportunidades para mejorar la eficiencia energética y reducir los costos energéticos.
- Desarrollo de políticas energéticas: los estudios de demanda energética pueden proporcionar información valiosa para el desarrollo de políticas energéticas efectivas y sostenibles.
- Evaluación de inversiones energéticas: los estudios de demanda energética pueden ayudar a evaluar el impacto financiero y ambiental de las inversiones energéticas y a identificar las mejores opciones de inversión.

Como se observa, la demanda energética en Ecuador ha presentado un aumento constante en los últimos años debido al crecimiento económico y poblacional del país. Las fuentes de energía renovable, como la hidroeléctrica y la eólica, han cobrado mayor relevancia en la matriz energética del país y se espera que sigan siendo una opción atractiva en el futuro.

Es importante seguir evaluando la situación técnica y económica de la demanda energética en el país para poder identificar oportunidades de mejora y garantizar la sostenibilidad del sector energético en el largo plazo.

3. IDENTIFICACIÓN DEL OBJETO DE ESTUDIO

El objeto de estudio es la capacidad de producción de energía eléctrica neta en el Ecuador, su impacto en el desarrollo económico y la competitividad del país en el mercado global y cómo esta capacidad puede aumentarse para satisfacer la demanda del sector productivo (índice de efectividad del servicio de energía eléctrica en el Ecuador). La justificación para llevar a cabo este proyecto radica en la necesidad urgente de abordar la problemática de la falta de capacidad de producción de energía eléctrica neta en el Ecuador para abastecer la demanda del sector productivo.

Un enfoque analítico permitirá identificar las soluciones más viables y sostenibles para el país en su conjunto, tomando en cuenta aspectos técnicos y económicos. De esta manera, se podrá abordar eficazmente la problemática y diseñar un plan de acción que permita aumentar la capacidad de producción de energía eléctrica neta en el país, con miras a mejorar la calidad de vida de la población y contribuir al desarrollo sostenible del Ecuador.

Además, este proyecto se justifica en el marco de la Agenda 2030 para el Desarrollo Sostenible de la ONU, en la cual el Objetivo de Desarrollo Sostenible número 7 busca garantizar el acceso a una energía asequible, segura, sostenible y moderna para todos (Naciones Unidas, 2015). En este sentido, la realización del proyecto contribuirá a avanzar en el cumplimiento de los compromisos adquiridos por el Ecuador en el marco de esta agenda internacional.

Como se ha mostrado, la importancia de resolver la problemática de falta de capacidad de producción de energía eléctrica neta en Ecuador es crucial para el desarrollo económico y social del país, así como para garantizar la seguridad energética. Además, el aumento de la capacidad de producción de energía eléctrica neta puede contribuir al desarrollo económico y social del país, al garantizar un suministro de energía adecuado para el sector productivo y mejorar la competitividad en el mercado global.

4. PLANTEAMIENTO DEL PROBLEMA

En el Ecuador, la demanda de energía eléctrica ha venido creciendo en los últimos años debido al aumento de la población y a la mayor penetración de dispositivos electrónicos y electrodomésticos en los hogares, lo que ha llevado a un incremento en el consumo de energía eléctrica (Ministerio de Energía y Recursos Naturales No Renovables, 2021). Además, la matriz energética del país depende en gran medida de los combustibles fósiles, lo que genera impactos ambientales y de sostenibilidad a largo plazo (Aguilera, González, & Henríquez, 2019). Con base a lo mencionado, la problemática a ser estudiada es la falta de capacidad de producción de energía eléctrica neta en el Ecuador para abastecer la demanda del sector productivo.

Esta problemática es crítica para el Ecuador porque la falta de capacidad de producción de energía eléctrica neta puede afectar negativamente el desarrollo económico del país y limitar su competitividad en el mercado global. Además, la dependencia de combustibles fósiles puede generar volatilidad en los precios y afectar la seguridad energética del país (Cruz & Calderón, 2021).

Según el Ministerio de Energía y Recursos Naturales No Renovables de Ecuador, en el Plan de Generación, Transmisión y Distribución de Energía Eléctrica 2020-2025, se espera un déficit de energía eléctrica de aproximadamente 328 MW en el período 2020-2025. Esto se debe al aumento de la demanda de energía eléctrica en el país y a la limitada capacidad de producción de energía eléctrica. Esto se evidencia en el "Informe de la Situación del Sector Eléctrico Ecuatoriano" del Ministerio de Energía y Recursos Naturales No Renovables (2020), el cual muestra que la demanda de energía eléctrica en el país ha ido en aumento en los últimos años, con una tasa promedio de crecimiento anual del 3,2% entre 2015 y 2019. Sin embargo, la capacidad instalada de generación eléctrica solo ha aumentado en un 1,3% anual en promedio durante el mismo período.

La justificación para adoptar un enfoque analítico es que se requiere una evaluación rigurosa de las opciones disponibles para aumentar la capacidad de producción de energía eléctrica neta en el país, tomando en cuenta aspectos técnicos y económicos.

5. OBJETIVO GENERAL

Predecir el volumen de energía producida y el indicador de efectividad en el suministro mediante el uso de modelos de aprendizaje automático supervisado, específicamente el algoritmo Random Forest.

6. OBJETIVOS ESPECÍFICOS

- Describir los factores que permitan mejorar la eficacia del servicio del sector eléctrico a través de estrategias de optimización de la generación y distribución de energía.
- Comparar el rendimiento del modelo de aprendizaje automático supervisado Random Forest en la predicción del volumen de energía producida y la eficiencia en su suministro con otros modelos de aprendizaje supervisado.
- Identificar las empresas productoras y distribuidoras de energía eléctrica en el Ecuador y los factores que influyen en su capacidad de producción.
- Seleccionar las opciones más viables y sostenibles para aumentar la capacidad de producción de energía eléctrica neta en el Ecuador y proponer recomendaciones para su implementación.
- Analizar el posible beneficio económico de afectar los principales factores que aumenten el volumen de energía neta producida e incrementen el indicador de eficacia del servicio.

7. JUSTIFICACIÓN Y APLICACIÓN DE LA METODOLOGÍA

Para poder resolver la problemática planteada de la falta de capacidad para abastecer la demanda de energía eléctrica en Ecuador, así como validar los medios para cumplir con los objetivos propuestos en la investigación, se procederá a realizar un análisis cuantitativo con fuentes de datos secundarias de libre acceso para la población. Esta información permite identificar el volumen de producción de energía de las distintas empresas productoras a nivel nacional, así como de los factores que pueden tener una relación con un aumento del volumen de energía producida. Así mismo, se cuenta con información de un indicador que mide la eficacia de conexión al servicio otorgado por las distintas empresas, con ello se valida que el suministro de energía sea constante a los distintos sectores productivos y de consumo, facilitando el crecimiento de las distintas industrias, así como manteniendo un nivel de uso adecuado en los hogares (Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables, 2023).

Con estos datos como insumos, se procede a realizar un análisis exploratorio y descriptivo de los datos, para poder identificar las principales relaciones entre las variables presentes en el conjunto de datos, así como aquellas que, en función de nuestro problema de estudio, son de interés a ser predichas con el objetivo de determinar la metodología a usar para realizar el modelamiento del conjunto de datos, realizar las predicciones y conseguir una base estadística robusta para realizar recomendaciones de que factores modificar en función de incrementar el volumen neto de energía producida, además de mejorar el indicador de conexión al servicio. Según este planteamiento, cabe enfatizar que las variables objetivo del análisis son el volumen neto de energía producido, así como el indicador de conexión al servicio (Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables, 2023).

7.1. RECOLECCION DE DATOS

Como se mencionó previamente, la fuente de datos a usar es secundaria y proviene de la Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables. En su apartado de Reportes de Información Estadística del Sector Eléctrico, se pueden obtener archivos planos en formato de Excel de las variables anteriormente mencionadas como objetivos, además de un conjunto de datos que tienen relación con el volumen de energía neta producida, así como el indicador de eficacia de la conexión al servicio (Ministerio de Energía y Recursos Naturales No Renovables, 2023).

Los reportes se obtienen por años, en el primer caso se obtuvo una base de datos de “Energía Producida” desde el año 2009 al 2022, la cual contiene diversas variables relacionadas a la empresa, el tipo de empresa, la central, el tipo de energía producida y demás variables que generan un impacto sobre el volumen neto de energía producida.

Por otro lado, la segunda base de datos corresponde al indicador de eficacia de conexión del servicio, para la cual se obtienen datos desde el año 2009 al 2019 dado que este indicador únicamente se midió hasta este año, pese a lo mencionado, de igual forma, se considera de vital relevancia para poder asegurar un suministro de energía adecuado, además de garantizar una tarifa acorde al consumo de cada cliente.

En estas dos bases de datos, se tienen variables temporales, como el mes y el año de evaluación, el tipo de indicador y el detalle de la empresa que lo produjo, además del tipo de área, urbana y rural, en la que se encuentra la empresa. Finalmente, el valor del indicador mencionado que oscila como una variable continua entre 0 y 1, siendo 1 el valor de eficacia más alto (Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables, 2023).

7.2. LIMPIEZA, PRE-PROCESAMIENTO Y/O TRANSFORMACIÓN DE DATOS

Para el pre-procesamiento de los datos, así como para el modelamiento estadístico final la herramienta que se usará es Python. Inicialmente por su capacidad de poder integrar distintas funcionalidades, tanto para la transformación de variables, visualización de gráficos y correlaciones, así como para crear modelos estadísticos robustos que permitan predecir el comportamiento futuro de nuestras variables objetivo de forma adecuada (Python, 2023).

Para esto, inicialmente se importan las librerías a usar, que son: “pandas”, “numpy”, “matplotlib” y “seaborn”. Estas librerías permitirán usar los datos como una base de datos, DataFrame, bidimensional con una estructura columnar, la librería que permitirá esto es pandas. De igual forma se importa numpy en el caso de que debamos trabajar con datos en series, en el caso de modelos de regresión o “arrays” al momento de ejecutar las proyecciones. Las librerías matplotlib y seaborn permiten realizar gráficos estadísticos para las visualizaciones y el análisis descriptivo de los datos (Python, 2023).

El detalle del código paso a paso utilizado para la importación de las librerías se encuentra cargado en la plataforma GitHub desde Jupyter Notebook ([Tesla7757, 2023](#)).

Posteriormente, se cargan las bases de datos y se transforman cada una en un DataFrame de pandas como se muestra en la figura 2. Después, se procede a visualizar la forma de las bases de datos, basadas en su total de columnas para confirmar el número de observaciones, las variables presentes en cada base de datos y proceder a describir el contenido de estas.

Una vez cargadas las bases de datos, utilizando Jupyter Notebook, en el primer conjunto de datos, el cual contiene el volumen neto de energía producida, se mantiene un total de 143.498 observaciones y 29 variables. Mientras que, en el

DataFrame 2 se cuenta con el indicador de eficacia de conexión al servicio mantiene 5.248 observaciones en 8 variables.

Posteriormente, se verifican las variables que componen cada uno de los conjuntos de datos, así como las primeras observaciones registradas. Adicionalmente, procederemos a detallar para cada conjunto de datos el diccionario de variables en las tablas 1 y 2.

Tabla 1. Diccionario variables base de datos 1

VARIABLE	DESCRIPCIÓN
Año	Representa el año calendario al que corresponde la información.
Mes	Indica el mes del calendario al que corresponde la información.
Id_Mes	Es un identificador numérico asignado a cada mes.
Id_Empresa	Es el identificador de la empresa a la que pertenece la información.
Empresa	Nombre empresa productora o distribuidora de energía.
Tipo_Empresa	Tipo de empresa (Productora o distribuidora) Empresas productoras de energía: Estas empresas se dedican a la generación de energía en diversas formas, como hidroeléctrica, petrolera, termoeléctrica, eólica, solar, entre otras. Producen energía a partir de diferentes fuentes y la suministran al sistema eléctrico para su distribución y uso posterior. Empresas distribuidoras de energía: Estas empresas son responsables de la distribución de energía eléctrica a los usuarios finales. Adquieren la energía producida por las empresas generadoras y la entregan a través de sus redes de distribución para su consumo por parte de los usuarios residenciales, comerciales e industriales.
Id_Central	Número de identificador de la central
Central	Nombre de la central. Una central de energía eléctrica es una instalación diseñada para generar electricidad a gran escala. Estas centrales convierten diferentes fuentes de energía en electricidad utilizable para su distribución y consumo. Pueden ser: hidroeléctricas, térmicas, eólicas, etc.
Unidad	Número de identificación de la unidad dentro de la central, es decir, es la identificación asignada a una máquina o equipo que se usa en la central para generar electricidad.
Tipo_Central	Tipo de central en términos de método de producción de energía, como: hidroeléctricas, térmicas, eólicas, solares, etc.
Subtipo_central	Subdivisión del tipo de central
Tipo_Energia	El tipo de energía que produce: renovable o no renovable.
Tipo_Servicio	Tipo de servicio: público o no público
Energia_Bruta_MW	Energía bruta producida

Consumo_Auxiliare_ Unidades_MWh	El consumo de los generadores internos a la planta
Energia_Neta_MWh	Energía bruta - menos consumo interno en auxiliares de energía (Objetivo de maximización)
FUELOIL_gal	Consumo de Fuel para producción
Diesel2_gal	Consumo de Diesel para la producción
Nafta_gal	Consumo de nafta para la producción
Gas_Natural_kpc	Consumo de gas natural para producción
Biogas_mpc	Consumo de Biogás para la producción
Crudo_gal	Consumo de crudo para la producción
Residuo (gal)	Residuo producido al final del proceso
Bagazo_t	Consumo de Bagazo para la producción
LPG_gal	Consumo de LPG para producción
Observaciones	Observaciones de los usuarios
Energia_Entregada_Servicio_Publico_MWh	Energía entregada después de la producción al servicio no Público
Energia_Entregada_Servicio_No_Publico_MWh	Energía entregada después de la producción al servicio Público

Elaborado por: Los autores.

Referencia: Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables, 2023.

Tabla 2. Diccionario variables base de datos 2

VARIABLE	DESCRIPCIÓN
Año	Representa el año calendario al que corresponde la información.
Mes	Indica el mes del calendario al que corresponde la información.
Id_Empresa	Es el identificador de la empresa a la que pertenece la información.
Empresa	Nombre empresa productora o distribuidora de energía.
Indicador	El tipo indicador que mide la eficacia de conexión, se presentan dos: CSZU (Conexión del Servicio a Zona Urbana) CSZR (Conexión del Servicio a Zona Rural)
Tipo Area	Si se encuentra en un área rural o urbana
Valor del indicador	Índice asignado entre 0 y 1, mientras más cercano a uno, mejor es el desempeño.
Observaciones	Observaciones de los usuarios

Elaborado por: Los autores.

Referencia: Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables, 2023.

Posteriormente, se procede a combinar ambas bases de datos, usando la función “merge” presente en pandas, con el condicional “inner” para que nos devuelva un nuevo DataFrame que contenga el conjunto intersección de ambas bases de datos (Pandas, 2023).

Se usaron todas las columnas en común como criterios para ejecutar el cruce, las variables criterio son: "Id_Empresa", "Año", "Mes", "Empresa". Estas columnas fueron definidas para poder determinar el periodo y la empresa en la cual corresponden el nivel de energía producida y el índice de efectividad de conexión del servicio ([Tesla7757, 2023](#)).

De esta manera, se obtiene un nuevo DataFrame con 33 columnas, la suma de ambas en las distintas bases de datos excluyendo las duplicadas, y 26.732 observaciones que corresponden a la intersección de los dos conjuntos de observaciones.

Para finalizar el pre-procesamiento, en el nuevo conjunto de datos se eliminaron algunas variables que no tenían registros, es decir todos sus valores numéricos eran iguales a cero, por ejemplo, se elimina de ambos conjuntos de datos la variable "Observaciones" por su falta de valor analítico y se reemplazan las observaciones perdidas por 0, dado que en el diccionario de datos se indica que el no tener información notifica que no se produjo, usó o entregó energía, o sus fuentes, en ese periodo.

7.3. IDENTIFICACIÓN Y DESCRIPCIÓN DE VARIABLES

Las variables para predecir, como se ha venido mencionando en función de los objetivos de la investigación son, "Energia_Neta_MWh", que corresponde al volumen neto de energía producida y "Valor del indicador", que corresponde al indicador que mide la eficacia de la conexión del servicio. El resto de las variables en un modelo de regresión, serán las usadas para predecir el comportamiento de estas dos, exceptuando algunas variables que no generan un valor analítico. Todas las variables seleccionadas, se usan en la creación de los distintos modelos de predicción. De esta forma, los conjuntos a usar quedan conformados de la siguiente manera (tablas 3 y 4):

Tabla 3. Conjunto de variables a usar para la base de datos 1.

VARIABLE	TIPO DE VARIABLE	
Año	Independiente	Numérica
Id_Mes	Independiente	Numérica
Tipo_Empresa	Independiente	Categórica
Central	Independiente	Categórica
Unidad	Independiente	Categórica
Tipo_Central	Independiente	Categórica
Subtipo_central	Independiente	Categórica
Tipo_Energia	Independiente	Categórica
Tipo_Servicio	Independiente	Categórica
Sistema	Independiente	Categórica
FUELOIL_gal	Independiente	Numérica
Diesel2_gal	Independiente	Numérica
Nafta_gal	Independiente	Numérica
Gas_Natural_kpc	Independiente	Numérica
Biogas_mpc	Independiente	Numérica
Crudo_gal	Independiente	Numérica
Residuo (gal)	Independiente	Numérica
Bagazo_t	Independiente	Numérica
LPG_gal	Independiente	Numérica
Energia_Entregada_Servicio_No_Publico_MWh	Independiente	Numérica
Energia_Neta_MWh	DEPENDIENTE	Numérica

Elaborado por: Los autores.

Tabla 4. Conjunto de variables a usar para la base de datos 3.

VARIABLE	TIPO DE VARIABLE	
AÑO	Independiente	Numérica
Id_Mes	Independiente	Numérica
Tipo_Empresa	Independiente	Categórica
Central	Independiente	Categórica
Unidad	Independiente	Categórica
Tipo_Central	Independiente	Categórica
Subtipo_central	Independiente	Categórica
Tipo_Energia	Independiente	Categórica
Tipo_Servicio	Independiente	Categórica
Sistema	Independiente	Categórica
Energia_Neta_MWh	Independiente	Numérica
FUELOIL_gal	Independiente	Numérica
Diesel2_gal	Independiente	Numérica
Nafta_gal	Independiente	Numérica
Gas_Natural_kpc	Independiente	Numérica
Biogas_mpc	Independiente	Numérica
Crudo_gal	Independiente	Numérica
Residuo (gal)	Independiente	Numérica
Bagazo_t	Independiente	Numérica
LPG_gal	Independiente	Numérica
Energia_Entregada_Servicio_No_Publico_MWh	Independiente	Numérica
Indicador	Independiente	Numérica
Tipo Area	Independiente	Categórica
Energia_Entregada_Servicio_Publico_MWh	Independiente	Numérica
Valor del indicador	DEPENDIENTE	Numérica

Elaborado por: Los autores.

Una vez que se han determinado las variables, tanto la objetivo como las independientes, podemos proceder con el análisis descriptivo y exploratorio de las variables, en función de los objetivos planteados en este apartado, que obedecen a poder determinar el tipo de problema de ciencia de datos que más se ajusta a nuestra necesidad (regresión y/o clasificación), así como el algoritmo que mejor se ajuste a la naturaleza de los datos para tener un modelo robusto que permita proyecciones fiables.

Las medidas que se pueden usar principalmente para analizar el conjunto de datos pueden ser de tendencia central, como la media aritmética y la mediana; medidas de dispersión como la desviación estándar y el rango intercuartil. Adicionalmente, se pueden usar índices de correlación entre las distintas variables, para validar si existen altas relaciones entre variables independientes, todo esto a fin de poder descartar el uso modelos de regresión simple como la regresión lineal y validar que algoritmo de aprendizaje automático supervisado se podría usar esperando tener resultados más fiables, tomando en cuenta que ambas variables son continuas y se abordaría como un problema de regresión.

También, se pueden usar herramientas gráficas, como las cajas y los gráficos de densidad, para poder validar las medidas anteriormente mencionadas, el rango intercuartil, las desviaciones estándar y de suma importancia, la distribución de las variables continuas, en caso de que tengan una distribución normal. También se realizan gráficos de barras para las variables categóricas, de modo que se pueden observar si existen relaciones entre estas variables con nuestros objetivos, además de usar gráficos de dispersión para validar otro posible método para llegar a los objetivos planteados, por medio de un problema de clasificación.

Tabla 5. Estadísticos descriptivos de las variables en el conjunto de datos 1.

	Energía Bruta MWh	Consumo Auxiliare Unidades MWh	Energía Neta MWh	FUELOIL _gal	Diesel2 _gal	Nafta _gal	Gas Natural _kpc	Biogas _mpc	Crudo _gal	Residuo (gal)	Bagazo _t	LPG _gal	Energía Entregada Servicio No_Publico MWh
count	143498	96460	143498	143498	143498	143498	143498	143498	143498	143498	143498	143498	133803
mean	2612.3	53.4	2576.4	21839.0	16372.2	293.4	2030.8	1068.4	8824.8	3230.0	128.8	669.6	428.1
std	10002.8	309.5	9894.9	246864.7	115461.5	29528.1	20896.4	20944.0	42838.2	32888.3	2413.3	14026.9	1045.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	11.1	0.0	10.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	200.4	0.7	198.9	0.0	10.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	122.5
75%	676.6	19.2	665.4	0.0	11351.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	344.5
max	405420.9	9207.3	405420.9	5977915.0	6233316.0	4902689.0	660000.0	628145.1	1822235.0	3283604.0	132237.5	393681.7	57589.4

Elaborado por: Los autores.

Tabla 6. Estadísticos descriptivos de las variables en el conjunto de datos 3.

	Energía Bruta MWh	Consumo Auxiliare Unidades MWh	Energía Neta MWh	FUELOIL _gal	Diesel2 _gal	Crudo _gal	Residuo _gal	Energía Entregada Servicio Publico MWh	Valor Indicador
count	26732	22030	26732	26732	26732	26732	26732	25048	26732
mean	732.5	9.1	725.0	7911.6	7950.2	353.1	106.5	773.4	0.8
std	1494.7	35.8	1476.5	55470.2	63971.1	7069.6	4390.0	1332.2	0.2
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.6	0.0	0.6	0.0	0.0	0.0	0.0	73.5	0.8
50%	155.3	0.3	154.5	0.0	0.0	0.0	0.0	263.6	0.9
75%	787.7	2.7	777.6	0.0	546.4	0.0	0.0	951.6	1.0
max	26187.7	897.6	25877.0	1582078.0	1913923.0	214904.0	216891.9	20107.4	1.0

Elaborado por: Los autores.

Como se puede observar en las tablas 5 y 6, en la descripción de las variables numéricas se observan las medidas de centralidad y desviaciones estándar mucho más altas al valor de la media, además que los valores máximos de las variables, al estar muy separados en rango del tercer cuartil, dan a considerar que se tienen valores atípicos. Por tanto, se puede considerar que trabajamos con datos dispersos y altos valores atípicos, tanto en nuestras variables objetivo, como en nuestras variables independientes.

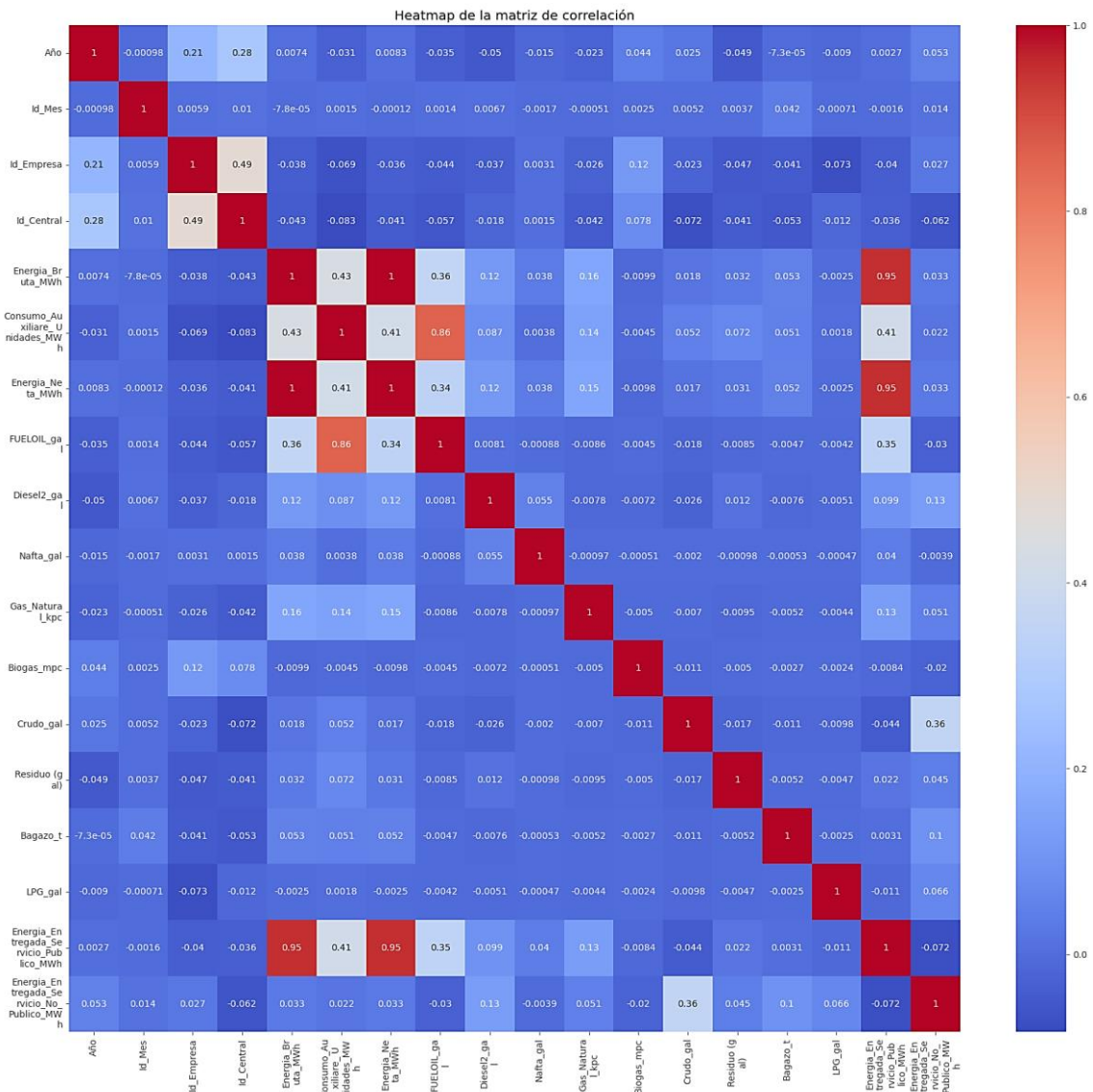


Figura 1. Matriz de correlación base de datos 1.

Elaborado por: Los autores.

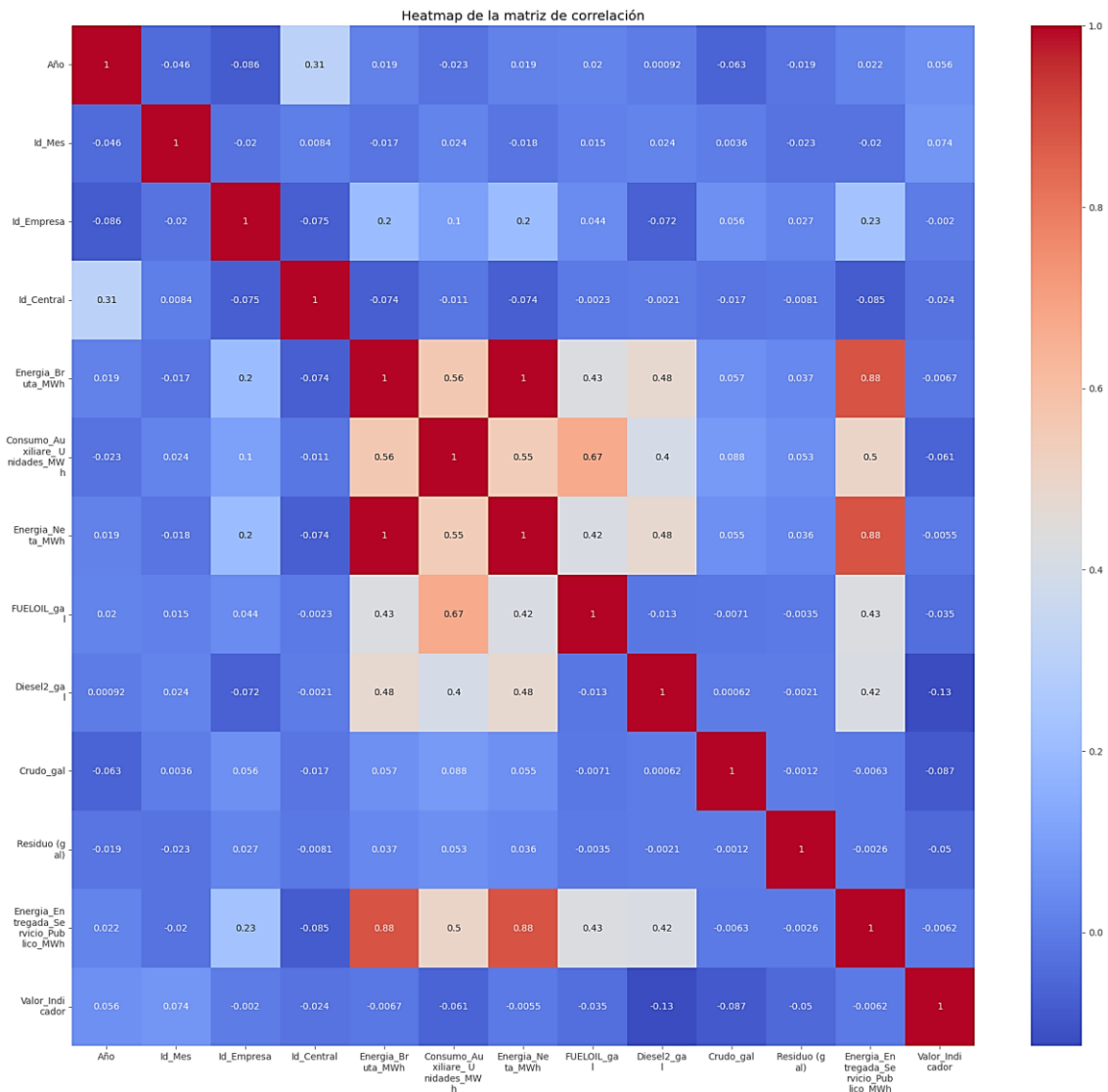


Figura 2. Matriz de correlación base de datos 3.

Elaborado por: Los autores.

En ambas matrices de correlación (figuras 1 y 2), se puede observar que, para la variable dependiente, tanto en la base de datos 1 como en la base de datos 3, no tienen una alta correlación con sus variables independientes. Por tanto, se puede descartar de manera muy acertada el uso de un modelo de regresión lineal simple para las predicciones estadísticas con este conjunto de datos.

Además, se puede observar en las figuras 1 y 2 que, para ambas bases de datos, la Energía Neta y la Energía Bruta presentan una correlación de 1, eso implica

que ambas variables están perfectamente correlacionadas. En otras palabras, cuando la energía bruta aumenta, la energía neta también aumenta en proporción exacta, y viceversa. Esto se debe a que, la energía neta es una proporción de la energía bruta, es decir, la energía bruta es la cantidad total de energía producida, mientras que la energía neta es la cantidad de energía que está disponible para su uso final después de restar la energía que se gastó en la generación, distribución y otras actividades asociadas.

En este sentido, cuando se habla de la energía bruta y la energía neta en la matriz de correlación, un valor de 1 indica una correlación positiva perfecta entre estas dos variables. En términos matemáticos, esto sugiere que la energía neta y la energía bruta son una combinación lineal una de la otra, es decir, se pueden expresar en la forma $\text{Energía Neta} = a * \text{Energía Bruta} + b$, donde 'a' es una constante que representa la proporción de energía bruta que se convierte en energía neta, y 'b' es la cantidad de energía que se pierde durante la producción, distribución y otros procesos.

De igual manera, se observó correlación alta de estas dos variables (energía bruta y energía neta) con la variable Energía entregada al servicio público, una relación de 0.95 en la base de datos 1 y 0.88 en la base de datos 3. La energía neta y la energía entregada al servicio público en dos bases de datos diferentes, esto indica una fuerte relación lineal positiva entre estas variables. En otras palabras, a medida que aumenta la cantidad de energía bruta o neta generada, la cantidad de energía entregada al servicio público también tiende a aumentar en gran medida.

7.4. VISUALIZACIÓN DE VARIABLES

Se inicia el análisis de visualización de las variables de la base de datos 1, en la que se busca predecir el volumen neto de energía producida. Se revisan primero las variables numéricas, y posteriormente las categóricas.

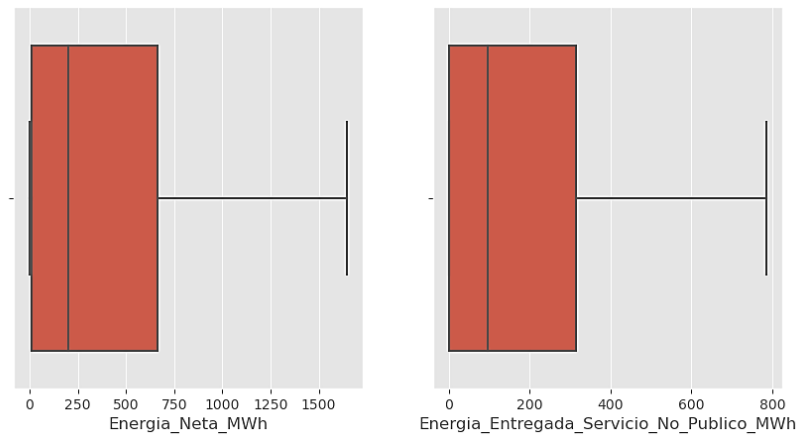


Figura 3. Ejemplo de gráfico de cajas variables continuas base de datos 1.

Elaborado por: Los autores.

Se observó de manera sencilla, que las variables continuas de conjunto de datos tienen valores muy dispersos. Especialmente el volumen de energía neto producido, tiene muchos valores atípicos que costarían modelar bajo un esquema de regresión lineal simple.

Pasamos a las gráficas de densidad, para poder ver cómo es la distribución de las variables continuas. Tanto de la variable objetivo como de las variables predictoras.

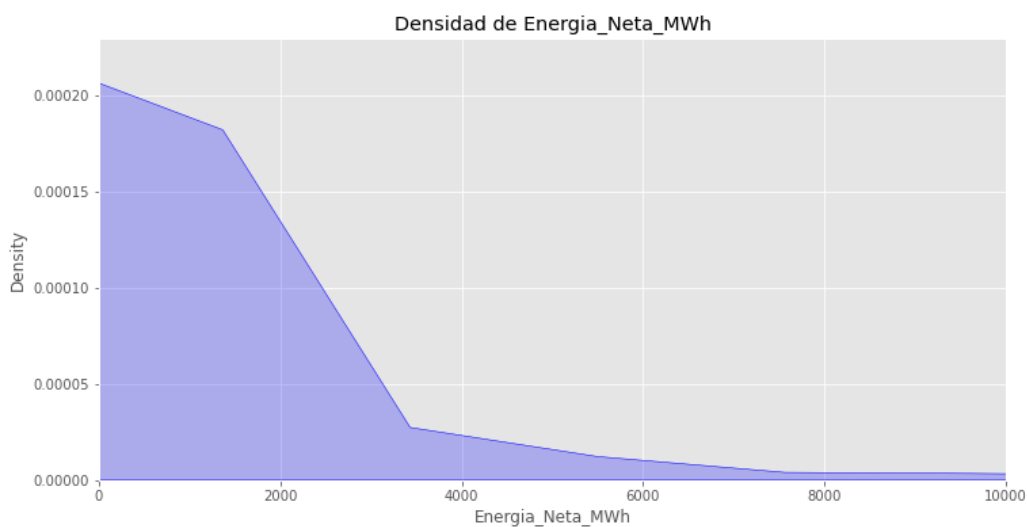


Figura 4. Gráfico densidad energía neto producida.

Elaborado por: Los autores.

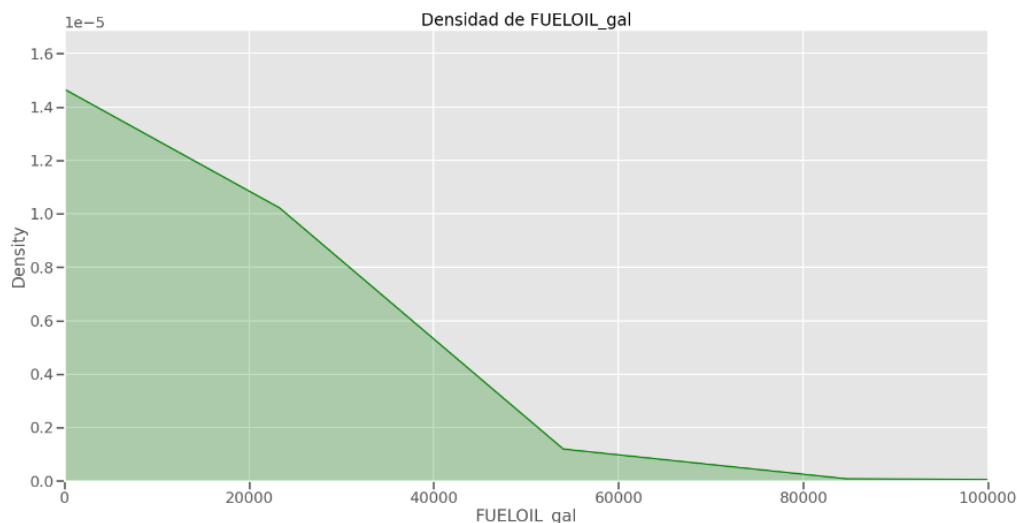


Figura 5. Gráfico de densidad variable predictora uso de Fueleoil_gal.

Elaborado por: Los autores.

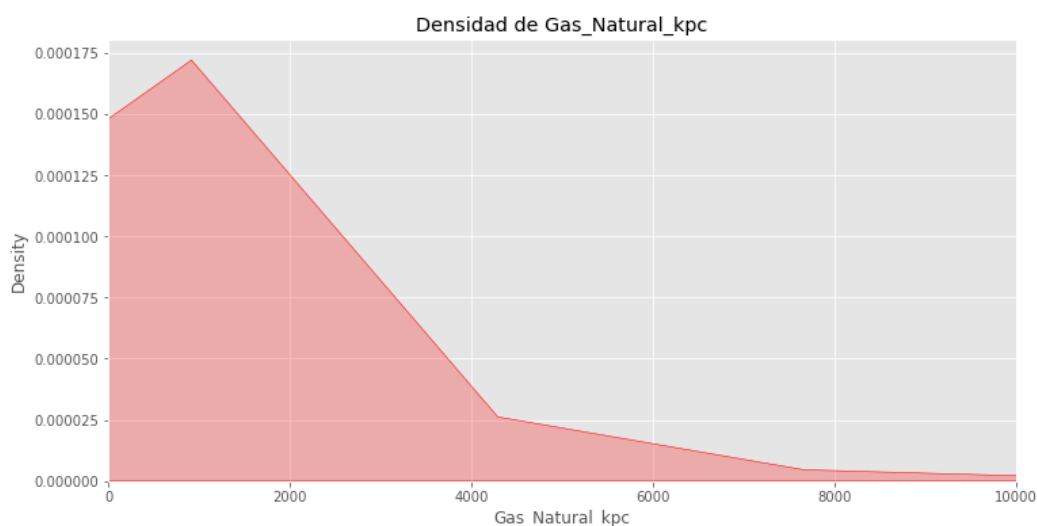


Figura 6. Gráfico de densidad variable predictora uso de Gas_Natural_KPC.

Elaborado por: Los autores.

En las variables continuas, se observa que existen muchos valores atípicos, tienen unos altos niveles de desviación y su distribución no es normal.

Por parte de las variables categóricas, tenemos los siguientes resultados.

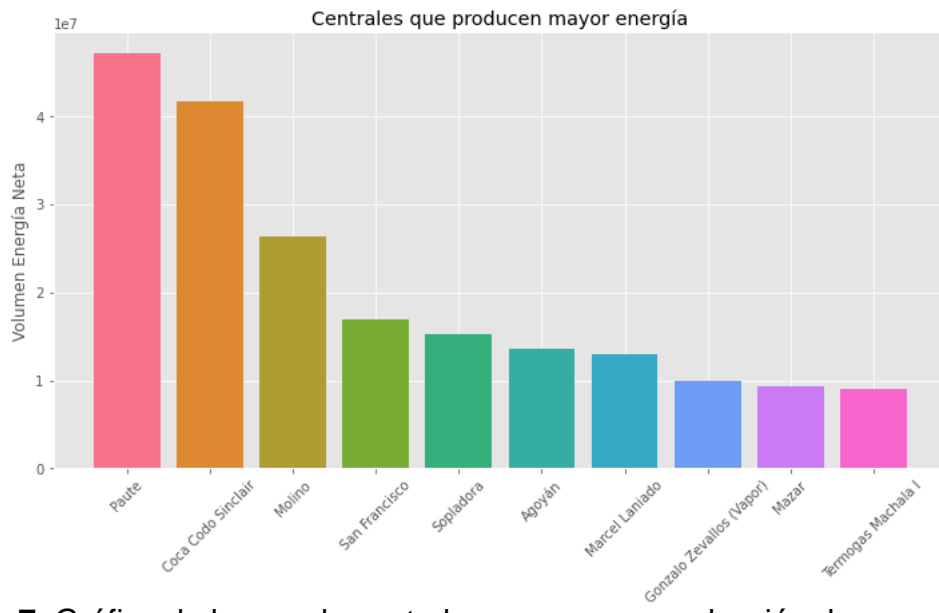


Figura 7. Gráfico de barras de centrales con mayor producción de energía.

Elaborado por: Los autores.

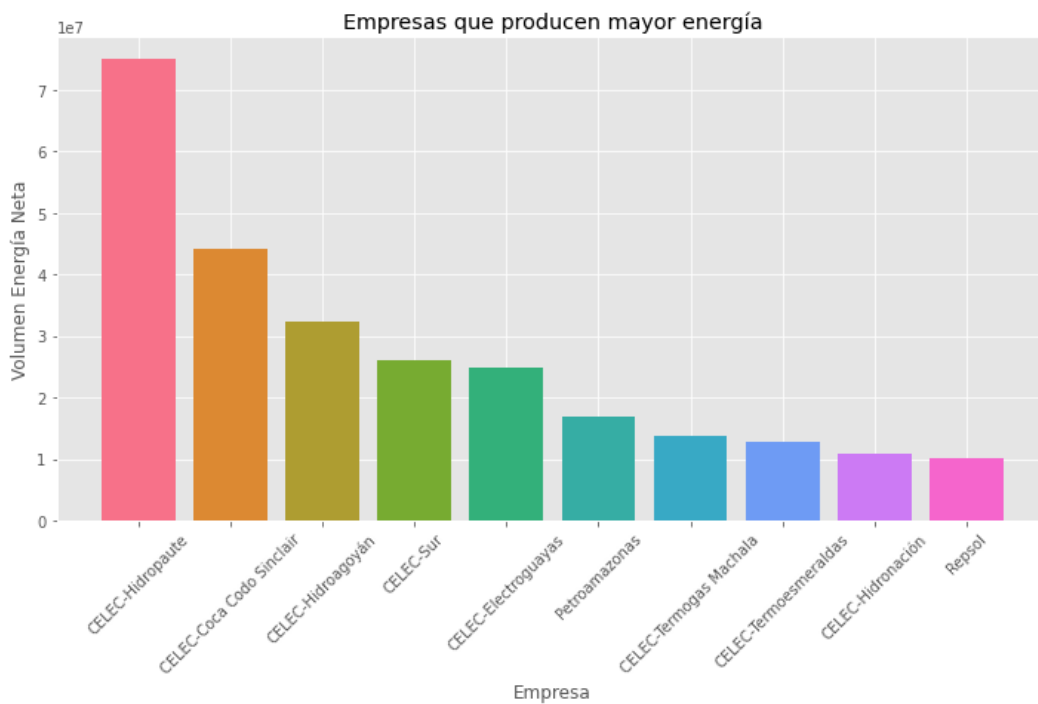


Figura 8. Gráfico de barras empresas que producen mayor cantidad de energía.

Elaborado por: Los autores.

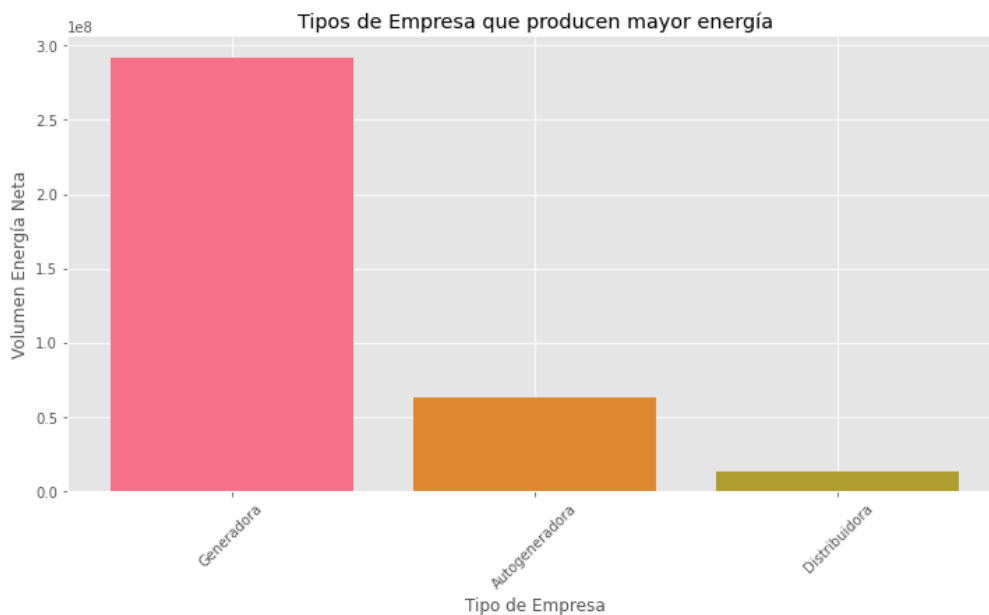


Figura 9. Gráfico de barras tipos de empresas que producen una mayor cantidad de energía.

Elaborado por: Los autores.

Se observa que algunas centrales mantienen una concentración elevada del volumen de energía que se produce a nivel total, así como que las empresas generadoras tienen una mayor representación en el total de la energía obtenida. Esto sirve para sustentar un esquema de clasificación, y validar que características componen a las empresas que mayores niveles de producción mantienen.

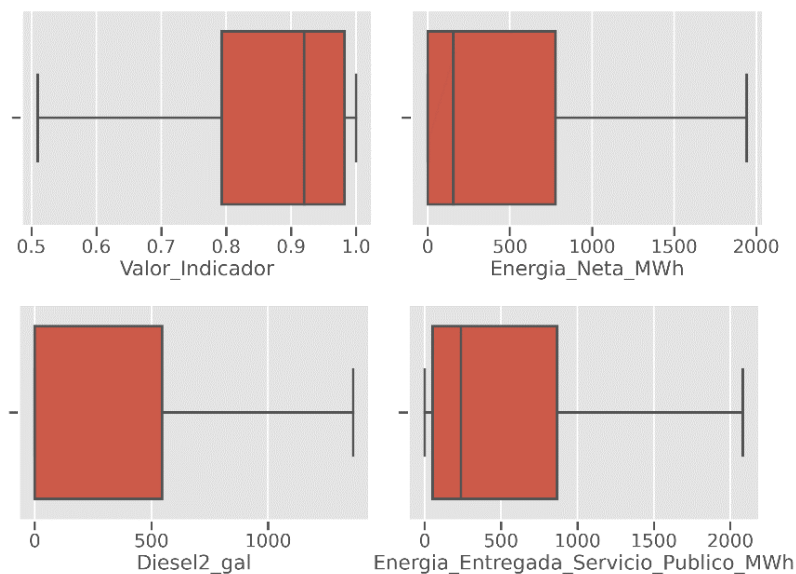


Figura 10. Ejemplo de gráfico de cajas variables continuas base de datos 3.

Elaborado por: Los autores.

De igual forma que con la base de datos 1, en el caso del conjunto de datos 3 se observaron muchos valores atípicos, alta variabilidad en los datos y poca simetría.

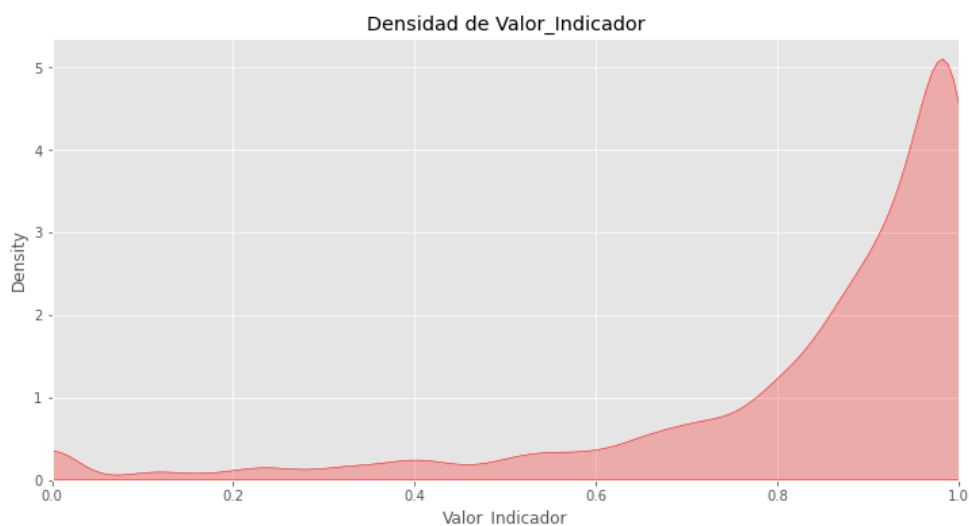


Figura 11. Gráfico densidad valor del indicador de eficacia a la conexión del servicio.

Elaborado por: Los autores.

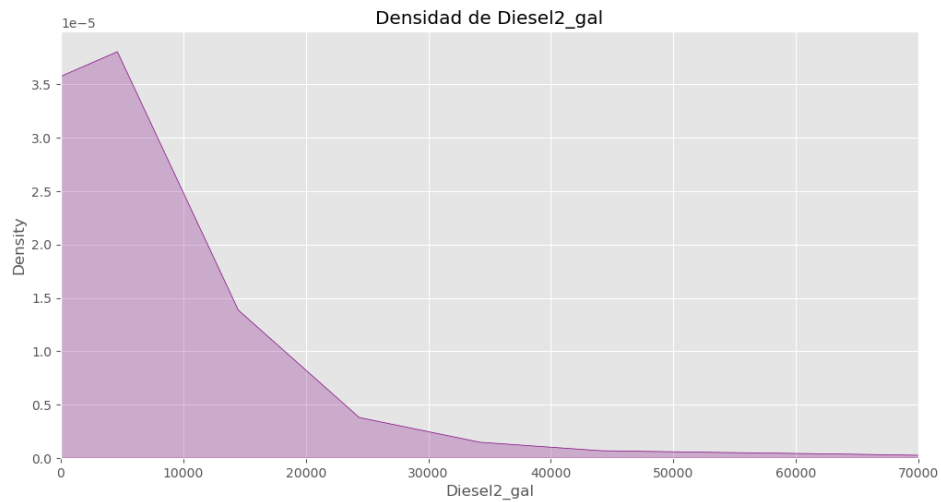


Figura 12. Gráfico densidad uso de Diesel2_gal.

Elaborado por: Los autores.

Conforme se observan las distribuciones en las figuras 11 y 12, en ningún caso alguna variable cumple dentro de la distribución normal.

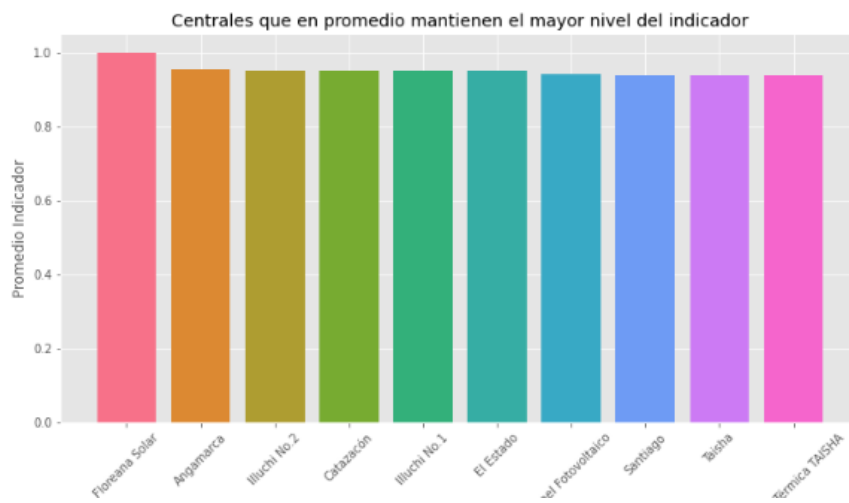


Figura 13. Gráfico de barras centradas con el indicador promedio más alto.

Elaborado por: Los autores.



Figura 14. Gráfico de barras empresas con el indicador promedio más alto.

Elaborado por: Los autores.

Para las variables categóricas, se obtuvo el promedio del valor del indicador en cada clasificación. Como se observa en las figuras 13 y 14 no hay una diferencia significativa en cuanto a los promedios de indicador de desempeño para cada tipo de central o empresa. A continuación, se presentan más comparaciones de las variables categóricas.

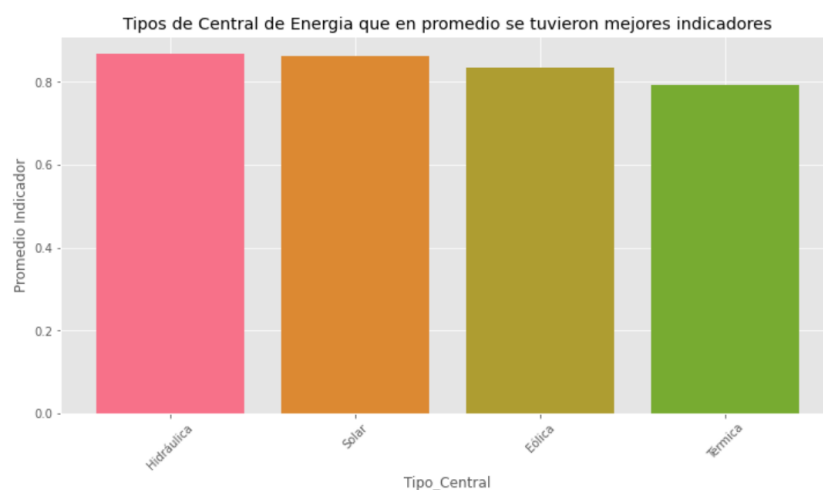


Figura 15. Gráfico de barras con los tipos de fuente de energía que en promedio se tuvieron mejores indicadores.

Elaborado por: Los autores.

En cuanto a la fuente de energía, se observa en la figura 15 que la energía hidráulica representa el mejor promedio de indicador de desempeño mientras que la energía térmica representa el promedio de indicador de desempeño más bajo.

Como en estos casos presentados no es tan evidente una diferencia entre los promedios de desempeño, de igual forma algunas empresas y centrales tienen índices en promedio más altos. Por tanto, se podría modelar estas relaciones en un problema de clasificación.

7.5. SELECCIÓN DE MODELO ESTADÍSTICO

En el marco de este proyecto, y tras una revisión bibliográfica, se evaluarán diversos modelos de aprendizaje automático con el objetivo de identificar el enfoque más adecuado y eficiente para abordar las necesidades específicas de la investigación. En particular, se dará especial consideración al uso del algoritmo Random Forest, el cual ha sido ampliamente respaldado por la literatura como uno de los mejores modelos, por ejemplo: Bueno, 2017; Li et al, 2018; Ahmad et al, 2020 son, por mencionar, algunos trabajos que respaldan el uso de Random Forest para esta aplicación. No obstante, se llevará a cabo una comparación de su rendimiento con otros enfoques previamente utilizados y documentados en la literatura, con el fin de garantizar una evaluación completa y precisa.

Los modelos a considerar son:

- RandomForestRegressor
- GradientBoostingRegressor
- LinearRegression

Cada uno de estos modelos presentan diferentes fortalezas y debilidades, y la eficacia de cada uno puede variar dependiendo de los datos específicos que se están analizando. Por lo tanto, el enfoque será pragmático: se implementará

cada uno de estos modelos y se evaluará su rendimiento mediante una serie de métricas de evaluación, como el error absoluto medio (MAE) y el coeficiente de determinación (R^2).

La elección del modelo final no se basará únicamente en estas métricas de rendimiento. También se tendrán en cuenta factores como la interpretabilidad del modelo, la velocidad de entrenamiento y predicción, y la capacidad del modelo para manejar nuevas entradas o cambios en los datos.

El objetivo es seleccionar el modelo que, en última instancia, proporcionará las predicciones más precisas y útiles para los datos de energía, ayudando a tomar decisiones informadas basadas en estas predicciones. A través de este enfoque riguroso y metódico, se espera encontrar el mejor modelo para las necesidades específicas del proyecto.

A continuación, se describen brevemente cada uno de los modelos a aplicar:

1. *Random Forest (Regressor y Classifier).*

Random Forest es un algoritmo de aprendizaje supervisado que se utiliza tanto para problemas de clasificación como de regresión. Es una técnica de ensamblado que combina múltiples árboles de decisión para obtener un resultado más preciso y estable (Breiman, 2001).

En este modelo, cada árbol del bosque se construye independientemente utilizando un subconjunto aleatorio de los datos. Cuando se necesita hacer una predicción, cada árbol da su propio resultado y la respuesta final se obtiene a través de la agregación de los resultados de todos los árboles, generalmente por votación en el caso de la clasificación y por promedio en el caso de la regresión (Breiman, 2001).

La razón por la que los bosques aleatorios son menos propensos al sobreajuste que un solo árbol de decisión es que cada árbol se entrena en un subconjunto distinto de los datos. Aunque cada árbol puede ser propenso al sobreajuste debido a su alta varianza, la agregación de los resultados reduce esta varianza y genera un modelo más robusto (Breiman, 2001).

$$RFf_i = \frac{\sum_{j \in \text{all trees}} normf_{ij}}{T}$$

En cuanto a la aplicabilidad de este modelo al presente proyecto se tiene:

- **Random Forest Regressor:** Este modelo es especialmente útil para predecir la demanda de energía eléctrica y los precios de la electricidad, que son variables continuas. Los Random Forest son capaces de modelar relaciones complejas y no lineales entre las características de entrada y la variable de salida, lo que podría ayudar a capturar los muchos factores que influyen en los precios y la demanda de energía.
- **Random Forest Classifier:** Aunque la predicción de la demanda y los precios de la energía suelen tratarse de una regresión, se puede usar la clasificación Random Forest Classifier para predecir si la demanda de energía superará cierto umbral en un día determinado, o si los precios de la energía estarán por encima o por debajo de la media. Esto podría ser útil para la toma de decisiones estratégicas y la planificación de la producción.

2. Gradient Boosting Regressor

Gradient Boosting es un algoritmo de aprendizaje supervisado que puede ser utilizado tanto para problemas de regresión como de clasificación. Este algoritmo es una técnica de ensamblado que combina múltiples modelos débiles, en general árboles de decisión, para crear un modelo fuerte (Friedman, 2001).

A diferencia de Random Forests, que entrena cada árbol de decisión de manera independiente y paralela, Gradient Boosting entrena cada árbol de forma secuencial. Cada nuevo árbol se ajusta para corregir los errores cometidos por los árboles anteriores. Para ello, cada nuevo árbol se ajusta a los residuos, es decir, la diferencia entre las predicciones actuales y las verdaderas etiquetas (Friedman, 2002).

El "gradient" en Gradient Boosting se refiere al hecho de que el algoritmo utiliza el gradiente descendente para minimizar la función de pérdida. En cada iteración, el algoritmo se mueve en la dirección del gradiente negativo de la función de pérdida para reducir los errores (Natekin & Knoll, 2013).

3. Linear Regression

La regresión lineal es un método estadístico utilizado para predecir una variable dependiente, a partir de una o más variables independientes (Montgomery et al., 2012). La variable dependiente es la que se pretende predecir o estimar, mientras que las variables independientes son las que se utilizan para hacer la predicción.

El modelo de regresión lineal asume una relación lineal entre las variables independientes y la variable dependiente. Esta relación se representa mediante la ecuación:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Donde,

- Y es la variable dependiente.
- X_1, X_2, \dots, X_n son las variables independientes.
- β_0 es el término de intercepción.
- $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de las variables independientes, que miden el cambio en la variable dependiente por cada cambio de una unidad en la variable independiente correspondiente.

- ε es el término de error, que es una variable aleatoria que añade ruido a nuestro modelo (Montgomery et al., 2012).

Aunque es un método relativamente simple, puede ser muy eficaz para hacer predicciones (Freedman, 2009).

Es importante tener en cuenta que la regresión lineal asume que las variables tienen una relación lineal. Si esta suposición no se cumple, el modelo puede no ser eficaz para hacer predicciones (James et al., 2013).

En el caso del modelo de regresión lineal, las variables que fueron tomadas tienen de los más altos índices de correlación con las variables objetivo. Por ejemplo, para el volumen de energía producida, se incluyó el uso de Diesel y Gas natural, que tienen correlaciones de 0.12 y 0.15 con la variable objetivo, respectivamente. De igual manera, las variables distintas al objetivo, que tienen correlación igual a 1 fueren descartadas del modelo.

8. RESULTADOS

8.1. ANÁLISIS DE MODELO ESTADÍSTICO E INTERPRETACIÓN DE RESULTADOS

8.1.1. MODELOS DE REGRESIÓN

A continuación, se presentan los principales resultados obtenidos con los diferentes modelos de aprendizaje supervisado.

Tabla 7. Resultados de los modelos de aprendizaje supervisado.

Modelos	Error medio absoluto (MAE)	Coefficiente de determinación (r2_score)
Random Forest	435.23	0.90
Gradient Boosting Regressor	1785.70	0.69
Linear Regression	2935.80	0.35

Elaborado por: Los autores.

En el caso del rendimiento de estos modelos, como se observa en la tabla 5, es evidente que cada uno tiene una capacidad diferenciada para captar la variabilidad en los datos del proyecto:

- **Random Forest:** Este modelo presenta el mejor rendimiento entre los tres analizados. El error medio absoluto (MAE) es de 435.23, el MAE bajo indica que el modelo fue capaz de hacer predicciones muy cercanas al valor real en promedio, y el coeficiente de determinación es 0.90, lo que significa que el modelo puede explicar el 90% de la variabilidad de la variable de respuesta. Es un valor muy alto resultando en una muy buena precisión del modelo en la predicción de los datos. En este caso, Random Forest supera a los otros dos modelos, lo que indica que es muy capaz de capturar la complejidad de los datos en el proyecto.
- **Gradient Boosting Regressor:** El modelo Gradient Boosting tiene un error medio absoluto mayor que el Random Forest (1785.70), y su coeficiente

de determinación es de 0.69. Esto sugiere que el modelo puede explicar el 69% de la variabilidad en la variable de respuesta. Sin embargo, su MAE fue bastante más alto que el de Random Forest, lo que significa que, en promedio, sus predicciones no fueron tan precisas. Aunque este modelo puede ser poderoso y flexible al lidiar con diferentes tipos de datos y relaciones, su rendimiento en este caso es inferior al del Random Forest. Esto puede deberse a la naturaleza del conjunto de datos o a la necesidad de afinar más los parámetros del modelo.

- **Linear Regression:** Este modelo es el que tiene el peor rendimiento de los tres, con un error medio absoluto de 2935.80 y un coeficiente de determinación de 0.35. Esto significa que el modelo solo puede explicar el 35% de la variabilidad en la variable de respuesta y su MAE fue el más alto de los tres modelos, lo que indica que sus predicciones en promedio estuvieron bastante lejos del valor real. Esto puede ser debido a que la relación entre las características y la variable de respuesta no es lineal.

Para la evaluación adicional de los modelos, tanto en el conjunto de datos 1 como en el 3, se trazan gráficos de dispersión que comparan las observaciones con las predicciones del modelo de regresión. Mientras la línea de ajuste tenga un grado de inclinación más cercano a 45%, es decir que su pendiente positiva este más cerca de 1, se tienen predicciones más precisas.

- **Gráfico de dispersión (scatter plot) para evaluar el conjunto de datos 1, con el ajuste por feature importance.**

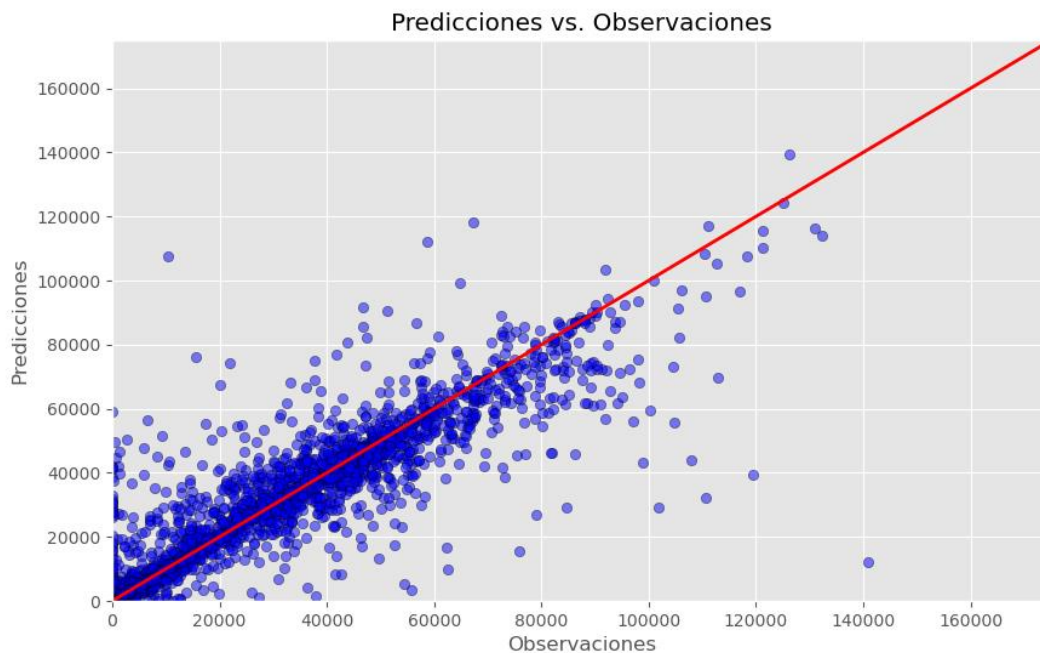


Figura 16. Gráfico de dispersión conjunto de datos 1.

Elaborado por: Los autores.

Como se visualiza en la gráfica anterior, en el eje de las “y” se tienen las predicciones del modelo ensemble random forest en un periodo de 13 años, para el volumen de energía producida. Mientras que, en el eje de las “x” se tienen las observaciones de la producción real desde el 2009 al 2022. Se puede evidenciar claramente que las predicciones, con los valores observados, son muy similares, por ese motivo es que el grado de inclinación de la recta de regresión es de casi 45 grados. Esto evidencia de forma visual el alto rendimiento del modelo en los datos obtenidos.

- **Gráfico de dispersión (scatter plot) para evaluar el conjunto de datos 3, con el ajuste por feature importance.**

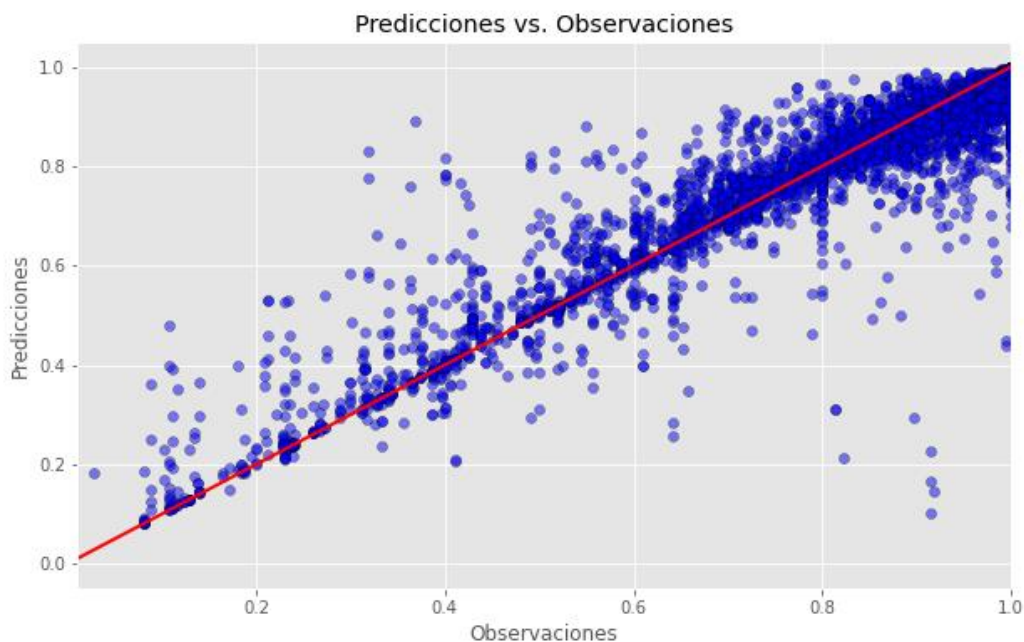


Figura 17. Gráfico de dispersión conjunto de datos 3.

Elaborado por: Los autores.

De la misma forma que en la base 1, en el gráfico de dispersión actual, se puede visualizar la relación entre las predicciones usadas en el modelo random forest, eje "y", y las observaciones en el eje "x", en la cual existen muchos puntos de intersección generando que la recta de regresión sea cercana a 45 grados, en un periodo de 10 años. Esto demuestra el alto nivel de precisión del modelo y su fiabilidad para poder prescribir escenarios futuros.

Es importante destacar, según el diccionario de datos existían algunas observaciones en cero que representaban información de un índice muy bajo de efectividad al servicio, y otros que indican que no se reportó información. Por ese motivo, al momento de evaluar las predicciones mediante el método gráfico, se decidió retirar de la visualización para que se pueda valorar correctamente la precisión del modelo en donde se tienen indicadores específicos para cada empresa.

Para el análisis prescriptivo, se identifica el "Feature Importance" ó "Importancia de las Características" de los modelos de regresión. Con esto, se puede

determinar qué factores se pueden modificar para incrementar el volumen de energía neta producida. Se ordenan primero las importancias de la más importante a la menos de forma descendente.

Cabe mencionar que, debido a la alta correlación observada entre las variables de energía bruta, energía neta y energía entregada al servicio público en el apartado 7.3, estas fueron eliminadas de los conjuntos de entrenamiento y prueba utilizados para el modelado. Esta decisión se tomó para evitar la multicolinealidad, que puede dificultar la interpretación del modelo y disminuir la estabilidad de las estimaciones de los parámetros. Al eliminar estas variables altamente correlacionadas, se busca mejorar la capacidad del modelo para hacer predicciones precisas y robustas sobre datos no vistos.

Conjunto de datos 1:

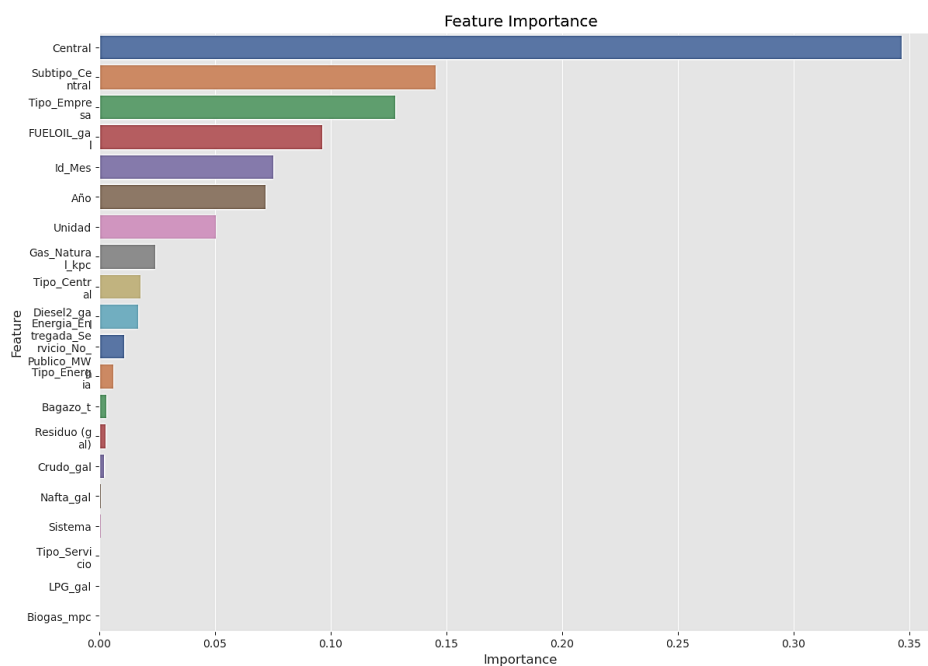


Figura 18. Importancia de las características conjunto de datos 1.

Elaborado por: Los autores.

Conjunto de datos 3:

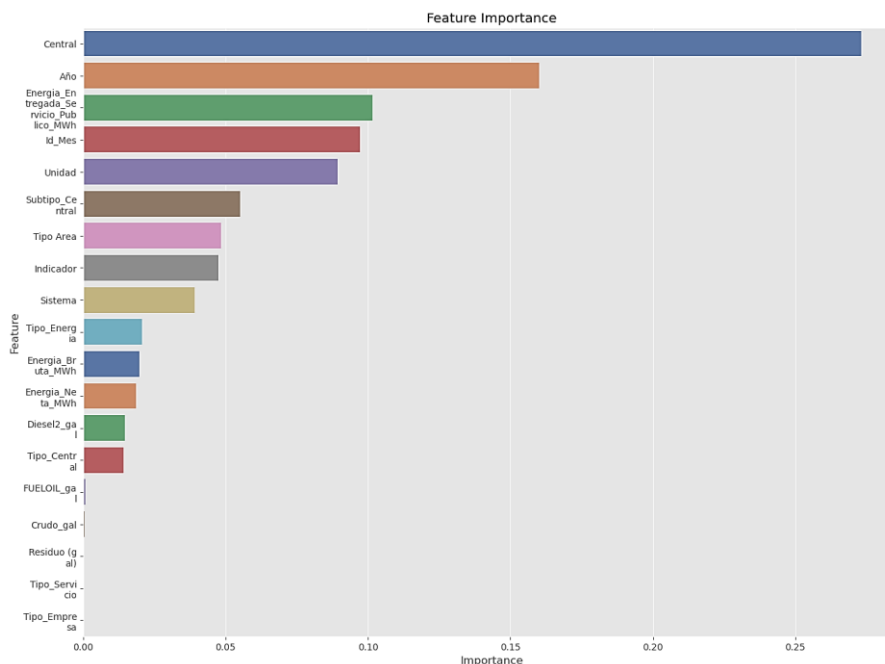


Figura 19. Importancia de las características conjunto de datos 3.

Elaborado por: Los autores.

En este caso, 'Central', 'Tipo_Empresa', 'Unidad', 'Subtipo_Central', y 'FUELOIL_gal' son las cinco características más importantes en la predicción del modelo, lo que indica que estas variables tienen un gran impacto en la producción y eficiencia del suministro de energía.

Con base a estos resultados obtenidos, cabe mencionar que de la visualización de las variables realizadas en el subcapítulo 7.4 se obtuvo que:

- Centrales con mayor producción de energía: Las tres principales centrales en términos de producción de energía son Paute, Coca Codo Sinclair, y Molino. Esto indica que estas centrales son altamente productivas y son esenciales para el suministro de energía en el país.
- Tipos de centrales que producen más energía: Las centrales hidráulicas son las que más energía producen, seguidas por las térmicas y las de

interconexión. Esto podría ser un reflejo de la disponibilidad de recursos (por ejemplo, agua para las centrales hidráulicas) o de la eficiencia de estos tipos de centrales.

- Empresas que producen la mayor cantidad de energía: Las empresas CELEC Hidropaute, CELEC Coca Codo Sinclair y CELEC Hidroagoyan, son las que más energía producen. Esto podría ser debido a que estas empresas tienen más centrales, centrales más eficientes, o una combinación de ambas.
- Tipos de empresas que producen una mayor cantidad de energía: Según los datos proporcionados, las empresas generadoras son las que producen la mayor cantidad de energía. Esto tiene sentido ya que estas empresas están directamente involucradas en la producción de energía, a diferencia de otras empresas que podrían estar más enfocadas en la distribución o venta de energía.
- Con el análisis de la importancia de los factores, podemos visualizar de manera sencilla que aspectos se deben impactar o incrementar su rendimiento para aumentar el volumen de energía producida, así como que métodos usar para tener un indicador de eficacia en la conexión del servicio más alto.

8.1.2. MODELO DE CLASIFICACIÓN

Se genera un nuevo potencial modelo de clasificación mediante el Random Forest Classifier con la agrupación de las empresas de acuerdo con su volumen de energía producida, así como a su indicador, para ello se identifica el "Feature Importance" ó "Importancia de las Características". Se ordenan primero las importancias de la más importante a la menos de forma descendente.

El análisis de la importancia de las características en el modelo Random Forest proporciona una visión adicional de los factores que contribuyen más significativamente a la producción de energía y a la eficiencia del suministro.

Conjunto de datos 3:

El análisis previo de la importancia de las características en el modelo condujo a una selección más precisa de las variables para el modelo de predicción. Las variables 'Tipo_Servicio' y 'Tipo_Empresa' fueron eliminadas del modelo debido a su baja importancia relativa (como se observa en la figura 20), mientras que las demás se mantuvieron.

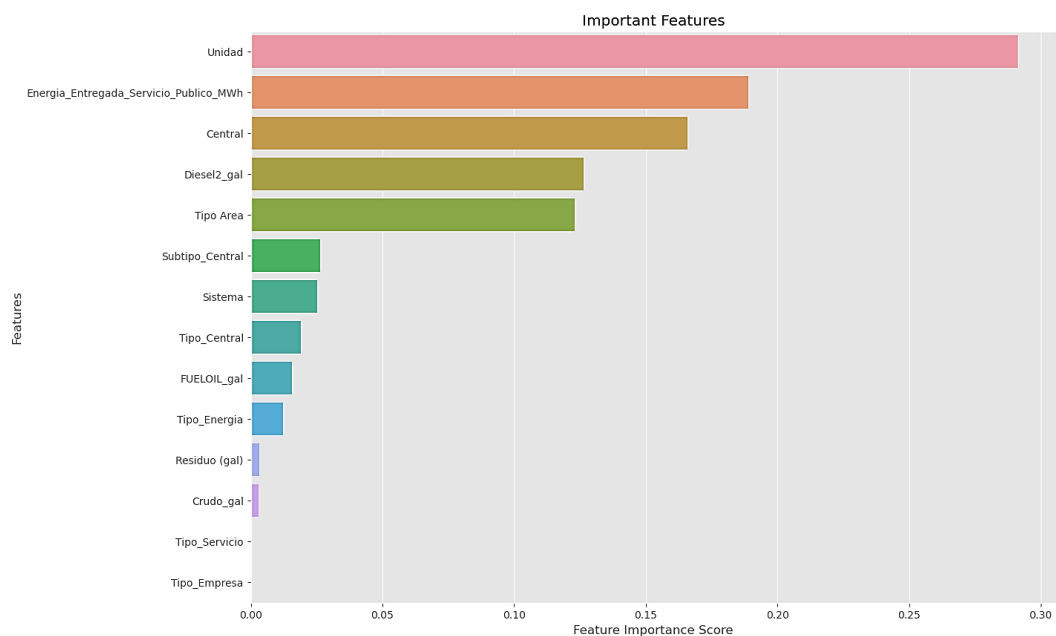


Figura 20. Importancia de las características conjunto de datos 3, modelo clasificación.

Elaborado por: Los autores.

Esta selección más centrada de las variables permite que el modelo se enfoque en las características que tienen un impacto más significativo en la producción y la eficiencia de la energía, y puede mejorar la precisión de las predicciones del modelo.

8.1.2.1. Informe de Clasificación

Utilizando el rango intercuartil como criterio, se genera un gráfico de dispersión que divide en el eje de las abscisas las empresas por su nivel de producción de energía neta, y en el eje de las ordenadas a las empresas por su indicador promedio. Tomando en cuenta la mediana para cada indicador, se pueden dividir a las empresas en aquellas que tienen un alto grado de producción energética y un valor de indicador alto, aquellas que tienen un alto grado de producción energética y un valor de indicador bajo, aquellas que tienen un bajo grado de producción energética y un valor de indicador alto; y aquellas que tienen un bajo grado de producción energética y un valor de indicador bajo (figura 21). De esta forma, se podría efectuar un modelo de clasificación para poder, en estos grupos, determinar cuáles son los factores que se podrían usar para predecir o lograr que las empresas mejoren su nivel de producción y a su vez su indicador de eficacia. De esta forma, en el análisis de la importancia de características, fácilmente se puede reconocer los factores de mayor impacto para empujar hacia el cuadrante de un alto índice de efectividad del servicio, así como un alto volumen de energía. Al existir factores relevantes como la “central”, el “subtipo de central” que son factores productivos que coinciden entre los tres modelos levantados y evaluados, se puede considerar que, si se impacta, las prescripciones generadas en los próximos escenarios serían altamente fiables.

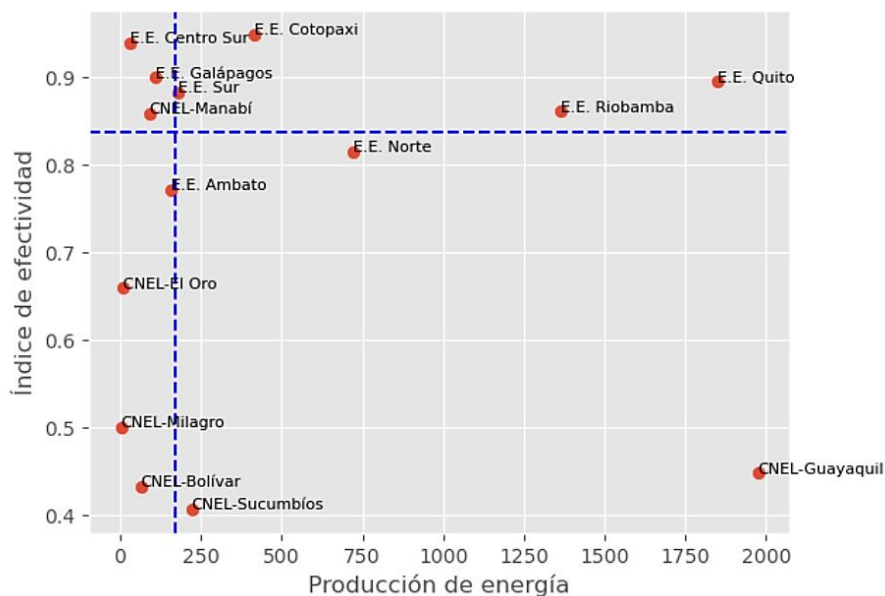


Figura 21. Gráfico de dispersión, clasificación de empresas por volumen de energía neto producido y valor de indicador de eficacia de conexión al servicio.

Elaborado por: Los autores.

El informe de clasificación proporciona un desglose detallado del rendimiento del modelo de clasificación:

- **Precisión:** Es una medida de cuán precisas son las predicciones del modelo. La precisión se calcula como el número de verdaderos positivos dividido por el total de verdaderos y falsos positivos.
- **Recall (Sensibilidad):** Esta métrica dice qué porcentaje de los casos positivos reales fueron capturados por el modelo. Es una medida de cuán completo es el modelo en la captura de los verdaderos positivos
- **F1-Score:** Esta es una métrica que combina la precisión y la sensibilidad en un solo número. El F1-score es la media armónica de la precisión y la sensibilidad.
- **Support:** Esta métrica da una idea de la distribución de las clases en tus datos.
- **Accuracy:** Esta métrica proporciona una medida general de cuántas predicciones fueron correctas, independientemente de la clase.

Tabla 8. Informe de clasificación

Clasificación	Precision	Recall	F1-score	Support
Alta producción y alta efectividad	0.89	0.79	0.84	39
Alta producción y baja efectividad	0.79	0.71	0.75	21
Baja producción y alta efectividad	0.75	0.60	0.67	30
Baja producción y baja efectividad	0.68	0.91	0.78	35
Accuracy			0.77	125
Macro avg	0.78	0.76	0.76	125
Weighted avg	0.78	0.77	0.77	125

Elaborado por: Los autores.

Este informe de clasificación muestra los resultados obtenidos de las predicciones de cada categoría de producción y efectividad. La precisión, recall y el f1-score son métricas importantes que nos ayudan a entender el rendimiento del modelo de clasificación:

Alta producción y alta efectividad: El modelo tiene una precisión del 89%, lo que significa que, de todas las predicciones que el modelo hizo para esta categoría, el 89% fueron correctas. El recall de 79% indica que el modelo identificó correctamente el 79% de los casos reales de esta categoría. El f1-score, que es una medida equilibrada de precisión y recall, es del 84%, lo que sugiere que de todas las observaciones que realmente pertenecen a esta categoría, el modelo solo pudo identificar correctamente el 84% de ellas.

Alta producción y baja efectividad: El modelo tiene una precisión del 79%, un recall del 71% y un f1-score del 75%. lo que indica que el modelo tiende a tener dificultades para clasificar correctamente las observaciones de esta categoría.

Baja producción y alta efectividad: El modelo tiene una precisión del 75%, un recall del 60% y un f1-score del 67%. Este parece ser el grupo donde el modelo tiene el desempeño más débil. En este caso, el recall es especialmente bajo. Esto indica que el modelo tiene problemas para identificar todas las instancias de esta categoría en los datos. Aunque la precisión es razonable, la baja

capacidad de recall sugiere que el modelo puede no ser muy útil para identificar estas unidades de producción en particular.

Baja producción y baja efectividad: El modelo tiene una precisión del 68%, pero un recall muy alto del 91% y un f1-score del 78%. Esto indica que el modelo es capaz de identificar la mayoría de las instancias de esta categoría en los datos, aunque puede hacer algunas predicciones incorrectas.

La precisión global del modelo es del 77%, lo que significa que el 77% de todas las predicciones del modelo son correctas. La media ponderada de la precisión, recall y f1-score también es del 77%, lo que indica que el modelo tiene un rendimiento equilibrado en todas las categorías.

9. PROPUESTA DE SOLUCIÓN

9.1. IMPLICACIONES ORGANIZACIONALES

El análisis de datos y la implementación de técnicas de Machine Learning en el presente proyecto han proporcionado una serie de perspectivas y herramientas valiosas para abordar la problemática organizacional identificada, que en este caso es la eficiencia y producción de energía. A continuación, se detallan algunas formas en que este análisis ayuda a resolver estos problemas a fin de establecer una propuesta de solución:

✓ Entender los factores que afectan la producción de energía:

Mediante el análisis exploratorio de datos y la identificación de las características más importantes en los modelos, se ha identificado los factores que tienen un impacto más significativo en la producción de energía. Por ejemplo, se ha encontrado que el tipo de central, el subtipo de central y el tipo de empresa tienen una gran influencia en la eficiencia y la producción de energía. Esta información puede ayudar a las organizaciones a tomar decisiones más informadas sobre dónde invertir recursos, cómo optimizar la producción y qué áreas podrían necesitar más atención o mejoras.

✓ Predecir la producción de energía:

Los modelos de regresión han permitido predecir la cantidad de energía que se producirá. En particular, el modelo Random Forest ha demostrado ser capaz de capturar la mayoría de la variabilidad en los datos, lo que significa que puede hacer predicciones precisas sobre la producción de energía en base a las características dadas. Este tipo de predicciones es útil para la planificación de la producción de energía, la gestión de la demanda y el ajuste de las operaciones en función de las predicciones.

✓ Clasificar las empresas por su eficiencia y producción de energía:

El modelo de clasificación ha demostrado ser útil para categorizar las empresas según su volumen de producción de energía y eficiencia, este modelo aún es una herramienta para identificar las empresas que están operando eficientemente y las que podrían necesitar mejoras.

✓ **Facilitar la toma de decisiones basada en datos:**

El análisis de datos y la aplicación de modelos de Machine Learning contribuyen a resolver la problemática organizacional proporcionando una mayor comprensión de los factores que afectan la eficiencia y la producción de energía, permitiendo predecir la producción de energía y clasificar las empresas por su eficiencia y producción. Estos hallazgos e insights pueden informar las decisiones organizacionales, optimizar las operaciones y mejorar la eficiencia en general. De modo, que se puede hacer uso de la analítica prescriptiva, validando los factores que tienen mayor impacto sobre el volumen de energía producida y el índice de eficacia del servicio, para poder mejorar ambos indicadores.

9.1.1. IMPLICACIONES ORGANIZACIONALES SOBRE INNOVACIÓN EMPRESARIAL

El uso de la analítica de datos, así como de técnicas que permiten realizar modelos predictivos y prescriptivos del volumen de energía producido y del indicador de eficacia del servicio, permite conocer que factores pueden modificarse para tener mucho mejores resultados a nivel nacional, además de permitir conocer qué métodos se pueden usar para realizar una explotación energética mucho más eficiente y sostenible a lo largo del tiempo, todo esto basado en la innovación usando métodos que probablemente, no eran ejecutados en el contexto ecuatoriano. Entre las principales implicaciones sobre la innovación, encontramos las siguientes:

✓ **Integración de energías renovables:**

Mediante la identificación de las características importantes, se pudo determinar qué factores como la central eléctrica, y el subtipo de central ejercen un fuerte impacto sobre el volumen de energía producida. Con ello, se validó que las centrales hidroeléctricas son las que tienen un mayor impacto en el volumen de energía producida, seguidas de la Térmica MCI y Térmica Turbovapor. De esta forma, se podría considerar impactar de manera positiva a las centrales hidroeléctricas para una mayor integración entre las empresas productoras y distribuidoras, además de impulsar otras fuentes alternativas como la energía eólica y fotovoltaica, que actualmente no generan un gran volumen, pero pueden generar muchos beneficios en el futuro.

✓ **Cumplimiento y eficiencia energética:**

Con el uso de la analítica predictiva, al tener modelos con un nivel de precisión tan elevados, se puede realizar una planificación en el mediano y largo plazo de los niveles de producción de energía y la eficacia del servicio otorgado por el sector energético al productivo. De esta manera, se puede asegurar el suministro para las distintas industrias, mantener esquemas mucho más efectivos de abastecimiento y sostener tarifas competitivas en el transcurso del tiempo.

✓ **Tarifificación dinámica:**

La analítica de datos y los algoritmos de inteligencia artificial permiten a planificar su esquema de producción, así como ajustar sus factores más importantes para poder llegar a un volumen de energía determinado. Con esto, se pueden dinamizar las tarifas, a fin de hacerlas más competitivas, conforme la demanda y requerimientos de las diversas industrias. Se pueden establecer franjas de tarifas más competitivas de acuerdo a volúmenes de consumos determinados, los cuales pueden ser más competitivos con el uso de fuentes energéticas más eficientes.

9.2. ESTRATEGIA ORGANIZACIONAL

Con base a lo mencionado, en el contexto del problema organizacional identificado, la capacidad de predecir de manera precisa la producción de energía y la eficiencia del suministro basándose en las características más relevantes se propone la siguiente estrategia organizacional:

1. Invertir en las centrales más productivas: Dado que se identificó que las centrales Paute, Coca Codo Sinclair, y Molino son las más productivas, se podría considerar asignar más recursos para mejorar estas instalaciones, expandir su capacidad o replicar sus prácticas de operación en otras centrales, se puede considerar incluso el cierre de centrales que no son eficientes.

2. Desarrollo de proyectos hidroeléctricos: Se observó que las centrales hidroeléctricas son las que más energía producen. Por lo tanto, podría ser beneficioso para la organización explorar oportunidades para el desarrollo de más proyectos de este tipo, siempre y cuando sean factibles desde el punto de vista ambiental y económico.

3. Enfocarse en las empresas más eficientes: CELEC Hidropaute, CELEC Coca Codo Sinclair y CELEC Hidroagoyán fueron identificadas como las empresas que más energía producen. La organización podría analizar las prácticas y operaciones de estas empresas para entender su eficiencia y aplicar estos conocimientos a otras empresas bajo su paraguas.

4. Mejorar los modelos predictivos: A través de la mejora continua de los modelos de regresión y clasificación empleados, la organización puede mejorar la precisión de sus predicciones sobre la producción y eficiencia de la energía. Los resultados obtenidos del modelo de aprendizaje automático (Random Forest) mostraron una precisión significativamente alta, lo que respalda la decisión de usar esta técnica para predecir la producción de energía. Esto permite hacer un uso más eficiente de los recursos, mejorando el rendimiento de la producción de energía. La estrategia de la organización también puede ahora

centrarse en perfeccionar este modelo y hacer uso de la predicción en tiempo real para optimizar sus operaciones.

5. Formación y desarrollo del personal: Con los conocimientos adquiridos sobre qué factores influyen más en la producción y eficiencia de la energía, la organización puede desarrollar programas de formación y desarrollo para su personal, centrándose en estas áreas clave.

6. Implementación de herramientas de Inteligencia Artificial (IA): Basándose en los resultados de este análisis, la organización puede buscar implementar herramientas de IA más robustas para mejorar la eficiencia y productividad de las operaciones. Esto podría incluir la implementación de algoritmos de aprendizaje automático para optimizar las operaciones en tiempo real, o el uso de IA para predecir y resolver problemas antes de que ocurran.

7. Mejorar las medidas de eficiencia energética: Esto podría incluir la implementación de tecnología de ahorro de energía, la mejora de los procesos operativos, o la capacitación del personal en prácticas de eficiencia energética.

8. Expansión de la interconexión: El análisis mostró que las centrales de interconexión también son importantes en términos de producción de energía. Esto indica que la expansión de las redes de interconexión puede ser beneficioso para la organización. Esta estrategia puede facilitar un suministro de energía más eficiente y confiable, y permitir a la organización aprovechar las oportunidades de mercado tanto a nivel local como internacional.

9. Estrategia de gestión de riesgos: Utilizando el análisis de datos para identificar áreas de riesgo potencial y desarrollar estrategias de mitigación. Por ejemplo, si se identifica que ciertos tipos de centrales son particularmente vulnerables a las fluctuaciones del mercado o a las condiciones climáticas, se pueden desarrollar e implementar estrategias de gestión de riesgos para tratar estos problemas.

11. Participación activa en la Política Energética: La capacidad de entender y predecir la producción y eficiencia energética puede dar a la organización una voz más fuerte en el desarrollo de políticas energéticas a nivel local y nacional. Esto puede incluir la promoción de políticas que fomenten el uso de fuentes de energía más eficientes o la adopción de tecnologías más limpias y sostenibles.

13. Alianzas Estratégicas y Colaboraciones: Los resultados de este análisis muestran que ciertas tecnologías o prácticas resultan en una mayor eficiencia energética, la organización podría buscar asociarse con empresas que se especializan en estas áreas.

14. Innovación en Tecnología Energética: Los hallazgos de este análisis pueden usarse para impulsar la innovación en tecnología energética dentro de la organización. Esto implica el desarrollo de nuevos métodos o tecnologías para mejorar la eficiencia energética, o la adaptación de las tecnologías existentes para mejorar su rendimiento.

Por otro lado, entre las opciones específicas para aumentar la capacidad de producción de energía eléctrica neta en Ecuador se tiene:

1. Promover la inversión en energías renovables: Como se puede validar en el análisis de la importancia de las características, es importante tomar en consideración el subtipo de central como un factor de alta relevancia para incrementar el volumen de energía producida, así como tener un indicador de efectividad de conexión al servicio más alto. Por ello se deben incentivar políticas y programas que fomenten la inversión en energías renovables, como la energía solar, eólica, hidroeléctrica y geotérmica. Esto implica la construcción de nuevas plantas de generación de energía renovable y la expansión de las existentes.

2. Modernizar y optimizar las centrales hidroeléctricas existentes: Las centrales hidroeléctricas son una importante fuente de energía en Ecuador. Se

deben realizar inversiones para modernizar y optimizar estas centrales, mejorando su eficiencia y capacidad de generación.

3. Fomentar la cogeneración y la generación distribuida: La cogeneración consiste en aprovechar el calor residual de los procesos industriales para generar energía. Se debe promover la implementación de sistemas de cogeneración en industrias y centros de producción. Además, la generación distribuida permite que los consumidores generen parte de su propia energía, utilizando fuentes como paneles solares o turbinas eólicas.

4. Implementar políticas de eficiencia energética: Es importante promover la eficiencia energética en todos los sectores, incluyendo la industria, el comercio y los hogares. Esto implica la adopción de tecnologías más eficientes, la implementación de sistemas de gestión energética y la concientización sobre el uso responsable de la energía.

5. Fomentar la investigación y desarrollo de tecnologías energéticas: Se debe impulsar la investigación y desarrollo de nuevas tecnologías energéticas en el país. Esto incluye la exploración de nuevas fuentes de energía, como la energía mareomotriz o la energía de biomasa, y el desarrollo de tecnologías más eficientes y sostenibles.

6. Mejorar la infraestructura de transmisión y distribución: Para aprovechar al máximo la capacidad de generación de energía, es necesario contar con una infraestructura de transmisión y distribución eficiente y robusta. Se deben realizar inversiones para mejorar y expandir la red eléctrica, asegurando una distribución confiable y eficiente de la energía en todo el país.

Como se puede observar, es este caso de estudio, la estrategia organizacional propuesta se refiere a la forma en que una organización puede abordar la eficiencia y la producción de energía a través del aprendizaje automático y el análisis de datos.

La implementación de la inteligencia artificial y el aprendizaje automático en este caso es un claro ejemplo de innovación tecnológica. Esta innovación no solo permitirá a la organización obtener un modelo con alta precisión, sino que también ha proporcionado una mejor comprensión de los factores que más afectan su producción de energía. Esta nueva perspectiva podría permitirles innovar en su enfoque de la producción de energía, tal vez cambiando la forma en que mantienen su equipo o ajustando su uso de recursos en base a las condiciones climáticas previstas.

Los resultados obtenidos y la innovación aplicada tienen un impacto directo en la competitividad de la empresa en el sector energético. Con la capacidad de predecir con precisión la producción de energía, pueden planificar y administrar de manera más eficiente su operación. Esto puede traducirse en costos operativos más bajos y un servicio más confiable para sus clientes, ambos factores claves para su competitividad en el mercado.

9.3. BENEFICIO ECONÓMICO

Basándonos en los factores discutidos anteriormente, existen varias formas en las que el aumento del volumen de energía neta producida e incremento en la eficacia del servicio pueden generar beneficios económicos:

- ✓ **Mayor producción de energía:** Al optimizar los factores que influyen en la generación de energía, como la utilización de tecnologías eficientes y la maximización de la capacidad de generación, se puede aumentar el volumen de energía neta producida. Esto resulta en un mayor suministro de energía disponible para su venta, lo que puede generar ingresos adicionales para las empresas del sector eléctrico. Por ejemplo, En un estudio realizado por IRENA (Agencia Internacional de Energías Renovables), se analizó el impacto económico de la energía renovable en varios países. Se encontró que el aumento de la capacidad de generación

de energía renovable tuvo beneficios económicos significativos como en Dinamarca, en donde se estimó que el incremento en la capacidad de energía eólica llevó a una reducción del 8% en los precios de la electricidad y a un ahorro de aproximadamente 1.7 mil millones de dólares al año en importaciones de combustibles fósiles (IRENA, 2019).

- ✓ Mejora en la eficiencia operativa: Al enfocarse en factores como la gestión de la demanda, la reducción de pérdidas de energía y la optimización de los procesos de distribución, se puede mejorar la eficiencia operativa. Esto implica una utilización más efectiva de los recursos y una reducción de los costos operativos, lo que puede contribuir a la rentabilidad y eficiencia económica de las empresas. En otro estudio realizado por el Banco Mundial, se examinaron los beneficios económicos de la mejora en la eficiencia del servicio eléctrico en países de bajos ingresos. Se descubrió que una mejora del 10% en la eficiencia de la distribución de energía podría aumentar el ingreso de las empresas en un promedio de 2.7%, lo que se traduce en un aumento de aproximadamente 3 mil millones de dólares en ingresos adicionales en estos países (Banco Mundial, 2016).
- ✓ Tarifas competitivas: Un aumento en la eficacia del servicio puede permitir a las empresas ofrecer tarifas más competitivas a los consumidores. Esto puede atraer a nuevos clientes y fomentar la retención de los existentes, lo que se traduce en un aumento en los ingresos y una mayor participación en el mercado.
- ✓ Reducción de costos: Al implementar medidas para mejorar la eficiencia energética y reducir las pérdidas, se pueden disminuir los costos de operación y mantenimiento. Esto contribuye a una mayor rentabilidad y puede reflejarse en tarifas más bajas para los consumidores.

Es importante destacar que el beneficio económico específico dependerá de las condiciones y características propias de cada empresa y mercado eléctrico.

Además, es necesario considerar aspectos adicionales como la inversión inicial requerida, los incentivos gubernamentales y las políticas regulatorias que pueden afectar la viabilidad económica de las mejoras propuestas.

10. CONCLUSIONES Y RECOMENDACIONES

10.1. CONCLUSIONES

De los modelos de regresión evaluados (Random Forest, Gradient Boosting Regressor, Linear Regression), el Random Forest ha demostrado ser el más eficaz en términos de precisión y capacidad de explicar la variabilidad en los datos con un nivel de precisión del 96%. Este alto nivel de precisión indica que el modelo es una herramienta valiosa para la planificación y toma de decisiones en la organización.

El modelo de clasificación ha demostrado una precisión aceptable en la categorización de la producción y efectividad de las unidades de generación de energía. Sin embargo, se observa una dificultad para clasificar correctamente las unidades de baja producción y alta efectividad.

Se realizó la identificación de las empresas productoras y distribuidoras de energía eléctrica en el Ecuador, lo que proporciona una visión clara de los actores clave en el sector eléctrico y su capacidad actual de producción. Esto proporcionó información valiosa para la planificación estratégica en el sector energético.

Se identificaron factores clave que pueden mejorar la eficacia del servicio del sector eléctrico, como la implementación de tecnologías de generación más eficientes, la optimización de la distribución de energía y el uso de modelos de predicción para mejorar la planificación y gestión del suministro eléctrico.

Se presentaron opciones viables y sostenibles para aumentar la capacidad de producción de energía eléctrica neta en el Ecuador, como la implementación de tecnologías renovables, la mejora de la eficiencia energética y la diversificación de fuentes de energía.

Los resultados del modelo de aprendizaje automático permiten a una organización prever con precisión su producción de energía, lo que puede conducir a mejoras significativas en la eficiencia operativa y la planificación de recursos. Esto también puede resultar en costos operativos más bajos.

10.2. RECOMENDACIONES

Dado que los modelos de aprendizaje supervisado dependen en gran medida de los datos de entrada, es fundamental contar con datos precisos y actualizados sobre variables relevantes, como la producción de energía, características de las centrales, condiciones climáticas, entre otros. Esto permitirá mejorar la precisión y confiabilidad de los modelos.

Se recomienda realizar análisis más detallados y evaluaciones de eficiencia para identificar oportunidades de mejora en el consumo de combustible y la gestión de los recursos disponibles, ya que, las variables identificadas como más importantes en la producción de energía, como la central, el tipo de empresa y el consumo de combustible, pueden ser utilizadas para optimizar el uso de recursos.

Dado que las centrales hidráulicas fueron identificadas como las que producen más energía en este caso de estudio, se recomienda considerar estrategias de diversificación en la matriz energética. Esto podría incluir la exploración e implementación de otras fuentes de energía renovable, como la solar o la eólica, para reducir la dependencia de una sola fuente y promover la sostenibilidad.

Es importante continuar explorando tecnologías de generación de energía eléctrica disponibles en el mercado y evaluar su potencial de aplicación en el contexto ecuatoriano. Esto permitirá diversificar las fuentes de energía y

aumentar la capacidad de producción neta, contribuyendo así a la seguridad y sostenibilidad del suministro eléctrico en el país.

Las recomendaciones propuestas, como la implementación de energías renovables y la modernización de infraestructuras eléctricas, pueden ayudar a mejorar la capacidad de producción de energía eléctrica neta en el Ecuador. Estas recomendaciones deben ser evaluadas en función de las condiciones y recursos específicos del país para garantizar su viabilidad y sostenibilidad a largo plazo.

Para mejorar la competitividad empresarial en el sector energético, se recomienda fomentar la colaboración entre empresas, centros de investigación y entidades gubernamentales. Esto puede facilitar la transferencia de conocimientos, promover la innovación en tecnologías y prácticas de generación de energía, y estimular el desarrollo de soluciones más eficientes y sostenibles.

Dado que el sector energético está en constante evolución, se recomienda realizar evaluaciones periódicas de los modelos y estrategias implementadas. Esto permitirá identificar cambios en las condiciones del mercado, nuevas tecnologías o regulaciones, y adaptar las estrategias organizacionales en consecuencia.

En el caso de realizar un análisis similar en el futuro, es importante considerar las limitaciones identificadas en este proyecto, como la disponibilidad de datos completos y actualizados (el presente proyecto se centró en un contexto geográfico y temporal específico, lo que limita la aplicabilidad de los resultados a otros lugares o períodos de tiempo), la consideración de factores no considerados como: aspectos sociales, ambientales o políticos que también podrían influir en los resultados y la mejora de la metodología utilizada (el rendimiento del modelo depende de la presencia de valores atípicos, de la configuración adecuada de los hiperparámetros, como el número de árboles en

el bosque, la profundidad máxima de los árboles y el número mínimo de muestras requeridas para dividir un nodo).

Para ello se recomienda ampliamente:

- Recopilar datos detallados y actualizados sobre la producción de energía y la eficiencia del suministro es fundamental para un análisis preciso. Esto puede incluir datos relacionados con la calidad y continuidad del suministro.
- Considerar datos climáticos relevantes, como la velocidad del viento, la temperatura y la humedad. Estos datos pueden ayudar a identificar patrones y tendencias estacionales, así como evaluar el impacto de las condiciones climáticas en la eficiencia del servicio.
- Un enfoque de análisis de series de tiempo puede ser beneficioso para analizar la evolución de la producción de energía y la eficiencia del suministro a lo largo del tiempo. Esto implica identificar patrones estacionales, tendencias a largo plazo y posibles cambios en el comportamiento de las variables.
- Considerar las limitaciones de los modelos de aprendizaje automático, dado que en el caso del modelo ensemble, el random forest usado a pesar de tener el más alto índice de precisión por encima de los demás algoritmos efectuados, al incrementar el número de estimadores no se tiene un incremento muy notable en el rendimiento de las predicciones, lo que podría limitar las capacidades de predicción del modelo en escenarios más complejos.
- En caso de que se desee aplicar estos algoritmos a otro conjunto de datos similar a los usados en este estudio, se recomienda hacer un análisis descriptivo a profundidad, el cual permitirá conocer las relaciones entre los datos, la distribución de las variables, tanto numéricas como categóricas, y así poder tener una visión clara de que esquema analítico

usar para obtener predicciones fiables. Cabe destacar que todos los modelos usados, tanto de regresión como de clasificación, son algoritmos supervisados, no se realiza una comparativa contra algoritmos de aprendizaje automático no supervisado, dado que, dentro de la literatura de investigación, las técnicas utilizadas son las que brindan un rendimiento más elevado.

REFERENCIAS

- Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables. (2023, January 1). Reportes de Información Estadística del Sector Eléctrico (SISDAT).
<http://reportes.controlrecursosyenergia.gob.ec/>.
- Aguilera, R., González, E., & Henríquez, F. (2019). Análisis del sistema eléctrico ecuatoriano y sus principales desafíos para el 2030. *Revista Iberoamericana de Energía*, 2(4), 37-53.
- Ahmad, T., Chen, H., Guo, Y., Zafar, M. H., & Alazab, M. (2020). Intelligent intrusion detection in low-power IoTs by using RF and DL algorithms. *IEEE Transactions on Industrial Informatics*.
- Arévalo, V., Sandoval, E., & Ochoa, D. (2019). Residential energy demand analysis in Ecuador. *Energy Reports*, 5, 852-856.
<https://doi.org/10.1016/j.egyr.2019.06.003>
- Banco Interamericano de Desarrollo (BID). (2016). Ecuador: energía para el desarrollo sostenible. Recuperado el 23 de abril de 2023 de <https://publications.iadb.org/publications/spanish/document/Ecuador-Energia-para-el-Desarrollo-Sostenible.pdf>
- Banco Interamericano de Desarrollo. (2018). Ecuador: Acelerando la Eficiencia Energética en el Sector Productivo. Recuperado el 23 de abril de 2023 de <https://publications.iadb.org/publications/spanish/document/Ecuador-Acelerando-la-Eficiencia-Energetica-en-el-Sector-Productivo.pdf>
- Banco Mundial. (2016). *Where There Is No Power: Increasing the Impact of Electrification on Income and Well-being*. Recuperado el 6 de junio de 2023 de: <https://openknowledge.worldbank.org/handle/10986/25090>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. DOI: 10.1023/A:1010933404324. Recuperado el 28 de mayo de 2023 de: <https://link.springer.com/article/10.1023/A:1010933404324>
- Bueno, B. (2017). Predicción de la demanda eléctrica horaria utilizando técnicas de Machine Learning. Trabajo Fin de Grado. Universidad

- Politécnica de Madrid, E.T.S.I. Industriales. Recuperado de:
https://oa.upm.es/47648/1/TFG_BLANCA_MARTIN_BUENO.pdf
- Cañizares, R., González-Longatt, F., & Pérez-Arriaga, I. (2019). Analysis of the Industrial Energy Demand in Ecuador. *Energies*, 12(18), 3527.
- Cevallos, C., Ortega, M., & Espinosa, M. (2019). Evaluación técnica y económica del potencial de energía solar en la industria ecuatoriana. *Revista de la Facultad de Ingeniería, Universidad Central del Ecuador*, 33(1), 1-10. <https://doi.org/10.29166/rfing.uc.33.1.2019.1>
- Comisión Económica para América Latina y el Caribe. (2017). Balance preliminar de las economías de América Latina y el Caribe 2017. Santiago de Chile: CEPAL.
- Cruz, V., & Calderón, L. (2021). Propuesta de diversificación de la matriz energética del Ecuador. *Revista Internacional de Investigación en Ciencias Administrativas y Turismo*, 2(1), 61-72.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232. Recuperado el 3 de junio de 2023 de: <https://www.jstor.org/stable/2699986>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. Recuperado el 3 de junio de 2023 de: [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Freedman, D. A. (2009). *Statistical models: theory and practice*. Cambridge university press. Recuperado el 3 de junio de 2023 de <https://www.cambridge.org/core/books/statistical-models/A3A1C6A4A9A6ABAE42B26C3E78A36C37>
- García, A., Rosales, J., Sánchez, C., & Vidal, R. (2020). Technical and economic analysis of wind energy in Ecuador. *Heliyon*, 6(10), e05232. <https://doi.org/10.1016/j.heliyon.2020.e05232>
- Huacón-Tapia, D., Díaz-González, F., Ceballos-Silva, A., & Castro-Rodas, M. (2019). Solar energy potential in Latin America and the Caribbean: A review of the literature. *Renewable and Sustainable Energy Reviews*, 102, 146-159. <https://doi.org/10.1016/j.rser.2018.12.014>

- IPCC (2011). Energía y desarrollo sostenible. Informe Especial del IPCC. Recuperado el 22 de abril de 2023 de https://www.ipcc.ch/site/assets/uploads/2018/03/SRREN_SPMs_ES.pdf
- IRENA. (2019). Renewable Power Generation Costs in 2018. Recuperado el 7 de junio de 2023 de: <https://www.irena.org/publications/2019/May/Renewable-Power-Generation-Costs-in-2018>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer. Recuperado el 2 de junio de 2023 de: <https://www.springer.com/gp/book/9781461471370>
- Jaramillo, C., López, C., & Guzmán, L. (2020). Technical and economic evaluation of photovoltaic solar energy in the Sierra region of Ecuador. *Energies*, 13(9), 2171. <https://doi.org/10.3390/en13092171>
- Li, X., Huan, H., Wang, X., & Zhao, Q. (2018). Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BPNN optimized by fireworks algorithm. *Energy Conversion and Management*, 165, 681-695.
- Ministerio de Energía y Recursos Naturales No Renovables de Ecuador. (2020). Plan de Generación, Transmisión y Distribución de Energía Eléctrica 2020-2025. Recuperado el 20 de abril de 2023 de <https://www.rekursosyenergia.gob.ec/plan-de-generacion-transmision-y-distribucion-de-energia-electrica-2020-2025/>
- Ministerio de Energía y Recursos Naturales No Renovables. (2016). Plan Nacional de Energía 2016-2025. Quito, Ecuador.
- Ministerio de Energía y Recursos Naturales No Renovables. (2020). Informe de la Situación del Sector Eléctrico Ecuatoriano. Recuperado el 20 de abril de 2023 de <https://www.rekursosyenergia.gob.ec/informe-de-la-situacion-del-sector-electrico-ecuadoriano/>
- Ministerio de Energía y Recursos Naturales No Renovables. (2021). Estadísticas del sector eléctrico. Recuperado el 27 de abril de 2023, de <https://www.rekursosyenergia.gob.ec/wp->

content/uploads/downloads/2021/06/Estadisticas-Sector-Elctrico-Mayo-2021.pdf

- Ministerio de Energía y Recursos Naturales No Renovables. (2016). Plan Nacional de Energía 2016-2025. Quito, Ecuador.
- Ministerio de Energía y Recursos Naturales No Renovables. (2020). Informe de la Situación del Sector Eléctrico Ecuatoriano. Recuperado el 23 de abril de 2023 de <https://www.rekursosyenergia.gob.ec/informe-de-la-situacion-del-sector-electrico-ecuadoriano/>
- Ministerio de Energía y Recursos Naturales No Renovables. (2021). Estadísticas del sector eléctrico. Recuperado el 27 de abril de 2023, de <https://www.rekursosyenergia.gob.ec/wp-content/uploads/downloads/2021/06/Estadisticas-Sector-Elctrico-Mayo-2021.pdf>
- Ministerio de Energía y Recursos Naturales No Renovables. (2021). Matriz energética. Recuperado el 23 de abril de 2023, de <https://www.rekursosyenergia.gob.ec/matriz-energetica/>
- Ministerio de Energía y Recursos Naturales No Renovables. (2021). Plan Nacional de Energía 2021-2040. Recuperado el 22 de abril de 2023 de <https://www.rekursosyenergia.gob.ec/plan-nacional-de-energia-2021-2040/>
- Ministerio de Energía y Recursos Naturales No Renovables de Ecuador. (2020). Plan de Generación, Transmisión y Distribución de Energía Eléctrica 2020-2025. Recuperado el 21 de abril de 2023 de <https://www.rekursosyenergia.gob.ec/plan-de-generacion-transmision-y-distribucion-de-energia-electrica-2020-2025/>
- Ministerio de Energía y Minas. (2023). Plan Maestro de Electricidad. Recuperado el 23 de abril de 2023 de <https://www.rekursosyenergia.gob.ec/>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons. Recuperado el 3 de junio de 2023 de: <https://www.wiley.com/en->

us/Introduction+to+Linear+Regression+Analysis%2C+5th+Edition-p-9780470542811

- Muñoz, L., & Villacrés, D. (2019). Análisis de la demanda energética en la ciudad de Quito. *Revista EURE - Revista de Estudios Urbano Regionales*, 45(136), 1-25. <https://doi.org/10.4067/S0250-71612019000300001>
- Naciones Unidas. (2015). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible. Objetivo de Desarrollo Sostenible número 7: Energía asequible y no contaminante. Recuperado el 24 de abril de 2023, de <https://www.un.org/sustainabledevelopment/es/energy/>
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7, 21. Recuperado el 1 de junio de 2023 de: <https://doi.org/10.3389/fnbot.2013.00021>
- Ochoa, G., & Paredes, D. (2019). Evaluación de la eficiencia energética en el sector industrial en la ciudad de Cuenca. *Revista de la Facultad de Ingeniería, Universidad de Cuenca*, 1(1), 1-14. <http://www.ingenieria.ucuenca.edu.ec/revista/index.php/revista/article/view/10>
- Organización Latinoamericana de Energía. (2019). Transición energética y renovables en América Latina y el Caribe. Recuperado el 27 de abril de 2023, de <https://www.olade.org/wp-content/uploads/2019/12/Transici%C3%B3n-Energ%C3%A9tica-y-Renovables-en-Am%C3%A9rica-Latina-y-el-Caribe-2019.pdf>
- Pandas. (2023, May 2). *pandas.DataFrame.merge*. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.merge.html>.
- Pérez, J., & García, D. (2020). Evaluación técnica y económica de la implementación de un sistema de gestión energética en una fábrica de muebles en la ciudad de Ambato. *Revista Tecnológica-ESPE*, 33(1), 25-32. <https://doi.org/10.37811/rte.v33i1.376>
- Python. (2023, May 2). Python About. <https://www.python.org/about/>

- Rosero-Mantilla, C., Torres-Bejarano, Á., Cisneros, E. A., & Silva-Sánchez, C. (2020). Evaluación técnica y económica de la generación de energía eléctrica a partir de residuos sólidos urbanos en la ciudad de Quito. *Revista Tecnológica-Escuela Politécnica Nacional*, 43(1), 71-80. <https://doi.org/10.37039/rt.v43i1.115>
- Rueda, J., & Ramírez, J. (2018). Evaluación técnica y económica de la generación distribuida en el sector residencial en Quito. *Revista de la Facultad de Ingeniería, Universidad de las Américas*, 15(2), 1-12. <https://doi.org/10.17163/ings.n15.2018.01>
- Sánchez, F., & Maldonado, M. (2018). Evaluación técnica y económica de la implementación de medidas de eficiencia energética en hoteles de la ciudad de Quito. *Revista Tecnológica-ESPE*, 31(4), 47-55. <https://doi.org/10.37811/rte.v31i4.298>
- Tapia, P., & Valencia, D. (2020). Evaluación técnica y económica de la instalación de paneles solares en viviendas unifamiliares en la ciudad de Cuenca. *Revista Científica de la Facultad de Ingeniería, Universidad de Cuenca*, 28(3), 89-99. <https://doi.org/10.32870/rci-ucuenca.v28i3.2019>

Código Proyecto:

Tesla7757 (2023).

Proyecto_Capstone_Final_Mancheno_Paola_Araque_Nicolás.ipynb:

Github repository [Accessed 2023, June].

https://github.com/Tesla7757/Capstone_Master_DataScience/blob/96f51253f34785c1615a8ed6e3da3c7873bf660c/Proyecto_Capstone_Final_Mancheno_Paola_Araque_Nicol%C3%A1s.ipynb

ANEXOS

Anexo 1

Tabla 1. Matriz con el detalle de investigaciones similares.

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
<i>Análisis de la demanda energética en el sector industrial de Ecuador.</i>	Datos de consumo energético del sector industrial, datos de producción industrial y datos de precios de la energía.	Análisis de regresión múltiple y análisis de series de tiempo.	Se encontró que el consumo energético en el sector industrial está altamente relacionado con la producción industrial y el precio de la energía, y se identificaron oportunidades de mejora en la eficiencia energética en algunos subsectores industriales.	Adopción de tecnologías más eficientes, la optimización del uso de la energía, la identificación de oportunidades de inversión y la implementación de políticas de gestión de energía.	Cañizares, R., González-Longatt, F., & Pérez-Arriaga, I. (2019). Analysis of the Industrial Energy Demand in Ecuador. <i>Energies</i> , 12(18), 3527.
<i>Análisis técnico y económico de la energía eólica en Ecuador.</i>	Datos de viento, datos de costos de inversión y operación de parques eólicos, datos de precios de la energía y datos de emisiones de gases de efecto invernadero.	Modelización de la velocidad del viento, análisis de costo-beneficio y evaluación del impacto ambiental.	Se encontró que la energía eólica podría ser una opción viable para la diversificación de la matriz energética del país, y se identificaron barreras técnicas, económicas y regulatorias que deben superarse para su implementación a gran escala. Se concluyó que los beneficios ambientales y económicos justifican la inversión en energía eólica en el país.	Identificación de oportunidades de inversión en proyectos de energía eólica, la promoción de políticas y regulaciones para fomentar el desarrollo de energías renovables, y la planificación de la infraestructura	Sánchez, C., Vidal, R., García, A., & Rosales, J. (2020). Technical and economic analysis of wind energy in Ecuador. <i>Heliyon</i> , 6(10), e05232.

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
				necesaria para la integración de la energía eólica en la red eléctrica.	
<i>Evaluación técnica y económica de la energía solar fotovoltaica en la región Sierra de Ecuador.</i>	Datos de radiación solar, datos de costos de inversión y operación de sistemas fotovoltaicos, datos de precios de la energía y datos de emisiones de gases de efecto invernadero.	Modelización de la radiación solar, análisis de costo-beneficio y evaluación del impacto ambiental.	Se encontró que la energía solar fotovoltaica podría ser una opción rentable para la generación de energía en la región Sierra de Ecuador, especialmente en zonas con alta radiación solar y altos costos de conexión a la red eléctrica. Se identificaron también barreras técnicas, económicas y regulatorias que deben ser superadas para su implementación a gran escala.	Identificación de oportunidades de inversión en proyectos de energía solar, la promoción de políticas y regulaciones para fomentar el desarrollo de energías renovables, y la planificación de la infraestructura necesaria para la integración de la energía solar en la red eléctrica.	Jaramillo, C., López, C., & Guzmán, L. (2020). Technical and economic evaluation of photovoltaic solar energy in the Sierra region of Ecuador. <i>Energies</i> , 13(9), 2171.
<i>Análisis de la demanda energética residencial en Ecuador.</i>	Datos de consumo energético residencial, datos socioeconómicos y demográficos de los hogares y datos	Análisis de regresión múltiple y análisis de varianza.	Se encontró que el consumo energético residencial en Ecuador está altamente relacionado con el nivel socioeconómico de los hogares y el precio de la energía. Se identificaron también	Identificación de oportunidades para la implementación de programas de eficiencia energética en hogares, la promoción de	Arévalo, V., Sandoval, E., & Ochoa, D. (2019). Residential energy demand analysis in Ecuador. <i>Energy</i>

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
	de precios de la energía.		oportunidades de mejora en la eficiencia energética en los hogares, especialmente en los de bajos ingresos. Estos resultados pueden ser útiles para el diseño de políticas públicas dirigidas a mejorar la eficiencia energética residencial en el país.	políticas y regulaciones para fomentar la adopción de tecnologías más eficientes en hogares, y la planificación de la infraestructura necesaria para satisfacer la demanda energética residencial de manera sostenible.	Reports, 5, 852-856.
<i>Análisis de la relación entre producción y consumo de energía eléctrica en Ecuador</i>	Datos de producción y consumo de energía eléctrica	Análisis de regresión lineal, análisis de series de tiempo y análisis de correlación.	Se identificó una relación positiva entre la producción de energía eléctrica y el consumo de energía eléctrica en Ecuador, así como una relación negativa entre el consumo de energía eléctrica y la tasa de interés.	Los resultados del estudio pueden ser útiles para el diseño de políticas energéticas que promuevan una mayor producción y consumo de energía eléctrica en Ecuador, así como para la toma de decisiones empresariales en el sector energético. Planificación y el	Morales, M., Tandazo, N., & Cabrera, J. (2017). Análisis de la relación entre producción y consumo de energía eléctrica en Ecuador. Revista de Gestión de la Energía y del Medio Ambiente, 14(3), 17-31.

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
				diseño de infraestructuras energéticas en el país.	
<i>Estudio sobre el potencial de la energía solar en la industria ecuatoriana.</i>	Datos de consumo energético de empresas industriales, datos de radiación solar y tarifas eléctricas.	Análisis de costo-beneficio y evaluación del potencial de la energía solar mediante el uso de un modelo de simulación	El estudio mostró que el uso de paneles solares puede reducir significativamente los costos de energía en la industria, con un potencial de ahorro de hasta el 40%. Además, se identificó que la falta de financiamiento es uno de los principales obstáculos para la adopción de energía solar en la industria.	El estudio sugiere que la implementación de políticas públicas para promover la adopción de energía solar, como la creación de incentivos fiscales y programas de financiamiento, puede ser efectiva para fomentar el uso de energía renovable en la industria.	Cevallos, C., Ortega, M., & Espinosa, M. (2019). Evaluación técnica y económica del potencial de energía solar en la industria ecuatoriana. Revista de la Facultad de Ingeniería, Universidad Central del Ecuador, 33(1), 1-10.

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
<i>Estudio sobre el análisis de la demanda energética en la ciudad de Quito.</i>	Datos de consumo de energía eléctrica y térmica de hogares y edificios, datos climáticos, y datos socioeconómicos.	Análisis de correlación y modelos de regresión múltiple.	El estudio mostró que el clima, la densidad de población, el tamaño de los hogares, el nivel socioeconómico y la edad de las viviendas son factores significativos en la demanda energética residencial y comercial en la ciudad de Quito. Además, se encontró que la demanda de energía eléctrica y térmica ha aumentado en los últimos años y se espera que siga aumentando en el futuro.	Los resultados del estudio pueden ayudar a los responsables de la toma de decisiones a desarrollar políticas energéticas y programas de eficiencia energética para reducir el consumo de energía en la ciudad.	Muñoz, L., & Villacrés, D. (2019). Análisis de la demanda energética en la ciudad de Quito. Revista EURE - Revista de Estudios Urbano Regionales, 45(136), 1-25.
<i>Evaluación de la eficiencia energética en el sector industrial en la ciudad de Cuenca.</i>	Datos de consumo energético de empresas industriales, datos de producción, datos climáticos, y tarifas eléctricas.	Análisis de costo-beneficio y evaluación de la eficiencia energética mediante el uso de un modelo de simulación.	El estudio encontró que la eficiencia energética en el sector industrial en la ciudad de Cuenca es baja, y que existen grandes oportunidades de mejora mediante la implementación de tecnologías y prácticas más eficientes en el uso de energía. Además, se identificó que la implementación de medidas de eficiencia energética puede reducir los costos de energía en un 10-30%.	El estudio sugiere que la implementación de políticas públicas para fomentar la eficiencia energética, como la creación de incentivos fiscales y programas de capacitación, puede ser efectiva para mejorar la eficiencia energética en el sector industrial en	Ochoa, G., & Paredes, D. (2019). Evaluación de la eficiencia energética en el sector industrial en la ciudad de Cuenca. Revista de la Facultad de Ingeniería, Universidad de Cuenca, 1(1), 1-14.

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
				la ciudad de Cuenca.	
<i>Evaluación técnica y económica de la generación distribuida en el sector residencial en Quito.</i>	Datos de consumo energético de hogares, datos de radiación solar, y tarifas eléctricas.	Análisis de costo-beneficio y evaluación del potencial de la generación distribuida mediante el uso de un modelo de simulación.	El estudio encontró que la generación distribuida en el sector residencial en Quito es viable desde el punto de vista técnico y económico, y que puede ser una alternativa atractiva para reducir los costos de energía y aumentar la resiliencia del sistema eléctrico.	El estudio sugiere que la implementación de políticas públicas para fomentar la generación distribuida puede ser efectiva para mejorar la resiliencia del sistema eléctrico y reducir los costos de energía en el sector residencial en Quito.	Rueda, J., & Ramírez, J. (2018). Evaluación técnica y económica de la generación distribuida en el sector residencial en Quito. Revista de la Facultad de Ingeniería, Universidad de las Américas, 15(2), 1-12.
<i>Evaluación técnica y económica de la generación de energía eléctrica a partir de residuos sólidos urbanos en la</i>	Datos de generación de residuos sólidos, datos climáticos y tarifas eléctricas.	Análisis de costo-beneficio y evaluación de la viabilidad técnica y económica mediante el uso de un modelo de simulación.	El estudio encontró que la generación de energía eléctrica a partir de residuos sólidos urbanos en la ciudad de Quito es viable desde el punto de vista técnico y económico, y que puede ser una alternativa atractiva para reducir los costos de energía y mitigar los efectos negativos de la generación de residuos sólidos. Además, se	El estudio sugiere que la implementación de tecnologías de generación de energía a partir de residuos sólidos puede ser efectiva para reducir los costos de energía y mitigar los efectos	Rosero-Mantilla, C., Torres-Bejarano, Á., Cisneros, E. A., & Silva-Sánchez, C. (2020). Evaluación técnica y económica de la generación de energía eléctrica a partir de residuos

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
<i>ciudad de Quito.</i>			identificó que la implementación de políticas públicas para promover la generación de energía a partir de residuos sólidos, puede ser efectiva para fomentar su adopción.	negativos de la generación de residuos sólidos en la ciudad de Quito.	sólidos urbanos en la ciudad de Quito. Revista Tecnológica-Escuela Politécnica Nacional, 43(1), 71-80.
<i>Evaluación técnica y económica de la instalación de paneles solares en viviendas unifamiliares en la ciudad de Cuenca.</i>	Datos de consumo energético, datos de irradiación solar y tarifas eléctricas.	Análisis de costo-beneficio y evaluación de la eficiencia energética mediante el uso de un modelo de simulación.	El estudio encontró que la instalación de paneles solares en viviendas unifamiliares en la ciudad de Cuenca puede ser rentable desde el punto de vista económico y puede reducir significativamente los costos de energía. Además, se identificó que la implementación de políticas públicas para fomentar la adopción de tecnologías solares en hogares puede ser efectiva para promover la adopción de tecnologías solares.	El estudio sugiere que la implementación de tecnologías solares en hogares puede ser efectiva para reducir los costos de energía y mejorar la eficiencia energética en la ciudad de Cuenca.	Tapia, P., & Valencia, D. (2020). Evaluación técnica y económica de la instalación de paneles solares en viviendas unifamiliares en la ciudad de Cuenca. Revista Científica de la Facultad de Ingeniería, Universidad de Cuenca, 28(3), 89-99.
<i>Evaluación técnica y económica de la</i>	Datos de consumo energético, datos de tarifas eléctricas y datos de	Análisis de costo-beneficio y evaluación de la eficiencia	El estudio encontró que la implementación de medidas de eficiencia energética en hoteles de la ciudad de Quito puede	El estudio sugiere que la implementación de medidas de	Sánchez, F., & Maldonado, M. (2018). Evaluación técnica y

Estudio	Tipo de datos utilizados	Metodologías para el análisis de datos	Resultados	Implicaciones gerenciales	Referencia
<i>Implementación de medidas de eficiencia energética en hoteles de la ciudad de Quito.</i>	ocupación de los hoteles.	energética mediante el uso de un modelo de simulación.	reducir significativamente los costos de energía y mejorar la sostenibilidad ambiental de los establecimientos. Además, se identificó que la implementación de políticas públicas para fomentar la adopción de medidas de eficiencia energética en hoteles puede ser efectiva para promover su adopción.	eficiencia energética en hoteles puede ser efectiva para reducir los costos de energía y mejorar la sostenibilidad ambiental en la ciudad de Quito.	económica de la implementación de medidas de eficiencia energética en hoteles de la ciudad de Quito. Revista Tecnológica-ESPE, 31(4), 47-55.
<i>Evaluación técnica y económica de la implementación de un sistema de gestión energética en una fábrica de muebles en la ciudad de Ambato.</i>	Datos de consumo energético, datos de tarifas eléctricas y datos de producción de la fábrica.	Análisis de costo-beneficio y evaluación de la eficiencia energética mediante el uso de un modelo de simulación.	El estudio encontró que la implementación de un sistema de gestión energética en la fábrica de muebles en la ciudad de Ambato puede ser rentable desde el punto de vista económico y puede reducir significativamente los costos de energía.	El estudio sugiere que la implementación de sistemas de gestión energética en empresas puede ser efectiva para reducir los costos de energía y mejorar la eficiencia energética en la ciudad de Ambato.	Pérez, J., & García, D. (2020). Evaluación técnica y económica de la implementación de un sistema de gestión energética en una fábrica de muebles en la ciudad de Ambato. Revista Tecnológica-ESPE, 33(1), 25-32.

Elaborado por: Los autores

