



Escuela de Administración
Escuela de Ingeniería Ciencia y Tecnología

Maestría en Business Analytics

Diseño de un modelo analítico de pronóstico del precio spot de energía en Colombia

Presentado por:

Cristian Arias García

Bogotá, D.C. 12 de mayo de 2023



Universidad del
Rosario

Escuela de Administración
Escuela de Ingeniería Ciencia y Tecnología

Maestría en Business Analytics

Diseño de un modelo analítico de pronóstico del precio spot de energía en Colombia

Presentado por:

Cristian Arias García

Bajo la dirección de:

John Pablo Calvo López

Bogotá, D.C. 12 de mayo de 2023

Contenido

Contenido	3
Declaración de originalidad y autonomía	5
Declaración de exoneración de responsabilidad	6
Lista de tablas	8
Resumen Ejecutivo.....	9
Palabras clave	9
Abstract.....	10
Keywords.....	10
1. Introducción	11
2. Objetivos	16
2.1. Objetivo General.....	16
2.2. Objetivos Específicos	16
3. Alcance.....	17
4. Contexto Sectorial y Organizacional.....	18
4.1. Contexto Sectorial.....	18
4.2. Contexto Organizacional.....	26
5. Estrategia Para la Solución.....	29
5.1. Comprensión del Negocio.....	30
5.2. Comprensión de los Datos	38
5.2.1. <i>Identificación y Definición de Variables</i>	38
5.2.2. <i>Base de Datos – Recolección de Variables</i>	44
5.3. Preparación de los Datos	45
5.3.1. <i>Número de Observaciones y Valores Ausentes</i>	45
5.3.2. <i>Variable Respuesta</i>	46
5.3.3. <i>Distribución Variables Exógenas</i>	48
5.3.4. <i>Análisis de Correlación</i>	49
5.3.5. <i>Estandarización de Variables</i>	50
5.4. Modelado.....	51
5.4.1. <i>K-Nearest Neighbor (KNN)</i>	51

5.4.2.	<i>Regresión Lineal</i>	52
5.4.3.	<i>Árboles de Decisión</i>	52
5.4.4.	<i>Random Forest</i>	53
5.4.5.	<i>Gradient Boosting Trees</i>	53
5.4.6.	<i>Maquina Soporte Vectorial (SVR)</i>	54
5.4.7.	<i>Redes Neuronales</i>	54
5.4.8.	<i>Modelos Híbridos</i>	55
5.4.9.	<i>Datos de Entrenamiento y Test</i>	56
5.5.	<i>Evaluación</i>	57
5.5.1.	<i>MSE</i>	57
5.5.2.	<i>MAPE</i>	58
5.5.3.	<i>R-Squared</i>	59
6.	<i>Resultados Solución</i>	60
6.1.1.	<i>K-Nearest Neighbor (KNN)</i>	61
6.1.2.	<i>Regresión Lineal (Ridge y Lasso)</i>	62
6.1.3.	<i>Árboles de Decisión</i>	64
6.1.4.	<i>Random Forest</i>	65
6.1.5.	<i>Gradient Boosting Trees</i>	66
6.1.6.	<i>Maquina Soporte Vectorial (SVR)</i>	67
6.1.7.	<i>Redes Neuronales (ANN y CNN)</i>	68
6.1.8.	<i>Modelos Híbridos (Voting Regressor)</i>	70
6.1.9.	<i>Comparación de Modelos</i>	70
7.	<i>Plan y Recomendaciones de Implementación</i>	74
8.	<i>Conclusiones y Recomendaciones</i>	76
9.	<i>Referencias Bibliográficas</i>	78
	<i>Anexos Técnico No.1</i>	83

Declaración de originalidad y autonomía

Declaro(amos) bajo la gravedad del juramento, que he(mos) escrito el presente Proyecto Aplicado Empresarial (PAE), en la propuesta de solución a una problemática en el campo de conocimientos del programa de Maestría por mi(nuestra) propia cuenta y que, por lo tanto, su contenido es original.

Declaro(amos) que he(mos) indicado clara y precisamente todas las fuentes directas e indirectas de información y que este PAE no ha sido entregado a ninguna otra institución con fines de calificación o publicación.

Cristian Arias García

Firmado en Bogotá, D.C. el 12 de mayo de 2023

Declaración de exoneración de responsabilidad

Declaro(amos) que la responsabilidad intelectual del presente trabajo es exclusivamente de su(s) autor(es). La Universidad del Rosario no se hace responsable de contenidos, opiniones o ideologías expresadas total o parcialmente en él.

Cristian Arias García

Firmado en Bogotá, D.C. el 12 de mayo de 2023

Lista de figuras

Figura 1. Plantas de Generación en Colombia	19
Figura 2. Generación y Capacidad Instalada	20
Figura 3. Agentes Sector Eléctrico Colombiano	22
Figura 4. Formación Precio Spot de Energía.....	24
Figura 5. Curvas de Oferta y Demanda de Energía.....	25
Figura 6. Metodología CRISP-DM	29
Figura 7. Análisis de Observaciones y Nulidad	46
Figura 8. Serie de Precio Spot de Energía en Colombia (2010-2022)	47
Figura 9. Distribución de Datos – Variable Objetivo.....	47
Figura 10. Distribución de Datos – Variable Exógenas	48
Figura 11. Análisis de Correlación Variables.....	50
Figura 12. Modelo Teórico Redes Neuronales.....	55
Figura 13. División Conjuntos de Datos (Entrenamiento – Test)	57
Figura 14. Predicción Modelo KNN	61
Figura 15. Predicción Modelo Regresión Ridge	62
Figura 16. Predicción Modelo Regresión Lasso.....	63
Figura 17. Predicción Modelo Decisión Tree.....	64
Figura 18. Predicción Modelo Random Forest.....	65
Figura 19. Predicción Modelo Gradient Boost.....	66
Figura 20. Predicción Modelo SVR	67
Figura 21. Predicción Modelo Red Neuronal ANN	68
Figura 22. Predicción Modelo Red Neuronal CNN	69
Figura 23. Predicción Modelo Híbrido.....	70
Figura 24. Comparación Predicciones Modelos.....	71
Figura 25. Comparación Modelos por Métrica Desempeño MSE	72

Lista de tablas

Tabla 1. Marco Conceptual	30
Tabla 2. Variables Identificadas	39
Tabla 3. Métricas de Desempeño Modelo KNN	61
Tabla 4. Métricas de Desempeño Modelos Regresión	63
Tabla 5. Métricas de Desempeño Modelo Decision Tree	64
Tabla 6. Métricas de Desempeño Modelo Random Forest	65
Tabla 7. Métricas de Desempeño Modelo Gradient Boost.....	66
Tabla 8. Métricas de Desempeño Modelo SVR	67
Tabla 9. Métricas de Desempeño Modelos Redes Neuronales	69
Tabla 10. Métricas de Desempeño Modelo Híbrido	70
Tabla 11. Resumen Métricas de Desempeño por Modelo.....	71

Resumen Ejecutivo

Diseño de un modelo analítico de pronóstico del precio spot de energía en Colombia

El precio de bolsa de energía es uno de los commodities con más volatilidad en el mercado colombiano, debido a los diferentes factores que pueden influir en su determinación, como la composición de la oferta y la demanda, el clima y precios de combustibles, lo que convierte su predicción en un desafío, lo cual genera incertidumbre en los agentes del mercado eléctrico colombiano para la toma de sus decisiones transaccionales en la compra y venta de energía. Por tanto, el objetivo de este proyecto es desarrollar un modelo analítico de pronóstico del precio spot de energía para la organización COSENIT con el fin de abordar la necesidad de sus clientes de tener mayor certeza en sus decisiones estratégicas y comerciales en el mercado de eléctrico colombiano.

Palabras clave

Precio energía, Modelos Analíticos, Pronostico, Mercado Eléctrico

Abstract

Design of an analytical model to forecast the spot price of energy in Colombia

The energy stock price is one of the most volatile commodities in the Colombian market, due to the different factors that can influence its determination, such as the composition of supply and demand, weather and fuel prices, which turns its prediction into a challenge, which generates uncertainty in the agents of the Colombian electricity market to make their transactional decisions in the purchase and sale of energy. Therefore, the objective of this project is to develop an analytical model for forecasting the spot price of energy for the COSENIT organization in order to address the need of its clients to have greater certainty in their strategic and commercial decisions in the Colombian electricity market.

Keywords

Energy Price, Analytical Models, Forecasting, Electricity Market

1. Introducción

La disponibilidad oportuna de información de calidad y confiable es crucial para el desarrollo de cualquier sector económico en un país. Esta información es fundamental para realizar análisis sólidos que respalden la toma de decisiones. El caso del mercado eléctrico no es la excepción, donde diferentes agentes de la cadena productiva deben tomar importantes decisiones de corto, mediano y largo plazo con la información disponible con el fin de mitigar riesgos y/o aprovechar oportunidades.

El sector de electricidad en Colombia es estratégico y transversal para economía nacional, dado que la energía eléctrica es un bien o commodity que impacta a toda la población, desde el sector residencial hasta los comercios y la gran industria, ya que es un insumo fundamental para el desarrollo económico del país y el bienestar de la sociedad. En este contexto, el análisis de los precios de la energía se convierte en una herramienta fundamental para la toma de decisiones y la administración del sector.

Actualmente en el mercado eléctrico colombiano existen dos mecanismos de precios para transar energía, primero, en el corto plazo a través de la bolsa de energía o mercado spot (Barrientos & Martínez, 2016). Segundo, a través de contratos de largo plazo (forward) donde el precio se establece bilateralmente entre las partes. Además, existe un mecanismo adicional que son las subastas de largo plazo, que es activado fundamentalmente por lineamientos del Gobierno Nacional para la promoción de energías renovables no convencionales, donde se establece un contrato financiero de tipo pague lo contratado con un precio de venta para un periodo definido.

A diferencia de los precios de contratos, los precios de bolsa de energía recogen en mayor medida las señales de mercado, dado que está en función principalmente de la oferta, ya que la demanda de energía eléctrica es tomadora de precios en el corto plazo. Adicional, la energía es un bien ligeramente distinto a muchos commodities, en aspectos importantes como su producción, distribución y consumo, por lo tanto, tiende a ser un bien con una alta volatilidad en su precio, puesto que la relación de cambio diaria o mensual para otras materias primas llega a ser tan solo del 5%, mientras que para el precio spot de la energía puede ser de hasta el 50% (Hurtado et al., 2016). Esta variabilidad afecta directamente la planeación de los ingresos y costos para los agentes y usuarios del sector eléctrico en Colombia.

Es así, como toma gran importancia tener información certera de pronósticos que permitan a las participantes del mercado decidir cómo invertir, producir y consumir la energía asumiendo posiciones de compra o venta dependiendo de las estrategias de cada uno (Santiago et al., 2017). Los modelos de pronóstico precisos pueden ayudar a reducir el riesgo y mejorar el rendimiento de la gestión empresarial. Además, la previsión de precios es uno de los campos más importantes de la gestión empresarial, porque uno de los temas más preocupantes y desafiantes para todos los participantes del mercado, es el cambio de precio futuro. Si los cambios futuros de los precios se pronostican con precisión, los responsables de la toma de decisiones pueden proponer una decisión óptima y diseñar un plan razonable para reducir el riesgo de mercado potencial y maximizar los beneficios económicos y sociales (Weron, 2014).

Sin embargo, las herramientas o modelos analíticos de predicción del precio de bolsa de energía tienen una complejidad relevante, dado que el precio tiene características específicas que lo diferencian de otros commodities como se mencionó anteriormente.

Principalmente se debe a diversos factores tales como la demanda inelástica, donde la demanda de electricidad no varía significativamente en respuesta a cambios en los precios. Esto se debe a que la electricidad es un bien necesario y es difícil para los consumidores reducir su consumo incluso si los precios suben. Además, la producción y el consumo de electricidad ocurren simultáneamente debido a las limitaciones en el almacenamiento eficiente de energía. (Mayorga & Sánchez, 2020). Adicionalmente, el precio de electricidad en Colombia tiene un alto impacto por variables exógenas, tales como factores climáticos dada la alta dependencia a la generación hidráulica (Santiago et al., 2017), regulatorios, costos de combustibles para la generación térmica, expansión en la capacidad instalada de las diferentes fuentes de generación (renovables, térmicas), entre otros. Por tanto, contar con un pronóstico preciso del precio de bolsa de energía es un desafío complejo que debe adaptarse a los cambios en los patrones de comportamiento de las diferentes variables que influyen en él mismo (Duque et al., 2016).

Actualmente en el mercado eléctrico. las empresas de consultoría enfrentan el desafío asociado a los mecanismos tecnológicos y de analítica que permitan gestionar el alto volumen de información, donde su consolidación, consistencia de almacenamiento y procesamiento es poco confiable. Esto genera una gestión ineficiente redundando en sobrecostos y pérdidas de los agentes del sector, ya que no se cuentan con modelos de pronóstico del precio de bolsa apropiados, siendo una de las principales variables del mercado eléctrico colombiano, que permitan atender los retos que enfrentan los participantes del mercado.

Es así como desde la organización COSENIT (Corporación de Soluciones Energéticas Integrales) se ha visto la necesidad de complementar sus capacidades analíticas y ampliar la oferta de valor a sus clientes, la cual, a lo largo de su trayectoria, se ha orientado a especialización en el conocimiento del sector energético colombiano, desarrollando productos para resolver problemáticas energéticas propias de los agentes del sector y en particular del sector industrial en Colombia.

Considerando el desafío que es predecir el precio de bolsa y las necesidades de los clientes de COSENIT, se ha determinado trabajar a través de este proyecto empresarial en el desarrollo de un modelo analítico de predicción del precio de bolsa de energía en Colombia, buscando abordar la deficiencia de los modelos actuales al interior de la compañía y aprovechando la oportunidad de las herramientas analíticas y técnicas de análisis de datos (Lucas et al., 2020) que existen actualmente, para elaborar un modelo con un nivel de certeza que permita tomar decisiones de los diferentes agentes del mercado (generadores, comercializadores y usuarios finales) en la planeación de inversiones de nuevos proyectos y para las transacciones operativas y comerciales de compra y venta de energía.

Para el desarrollo de este proyecto empresarial se da uso de la metodología CRISP-DM donde, a través de sus etapas, se busca dar respuesta a la problemática identificada y abordar la necesidad de complementar las capacidades analíticas de la organización COSENIT. En un primer momento se realiza una reseña del contexto sectorial y organizacional, luego se presenta una revisión literaria donde se aterrizan distintas soluciones a problemáticas similares mediante el uso de diferentes modelos analíticos. En segunda instancia se realiza una labor de identificación, y descripción de los datos, con el fin de garantizar unos resultados confiables, luego se desarrollan diez modelos de predicción del

precio de bolsa y se selecciona el que mejores métricas de desempeño arroja. Finalmente, se presenta una aproximación de las bases para la implementación del modelo dentro de la organización.

2. Objetivos

2.1. Objetivo General

Desarrollar un modelo analítico de pronóstico del precio de bolsa de energía en Colombia que facilite la toma de decisiones con certeza de los clientes de COSENIT.

2.2. Objetivos Específicos

- Analizar qué modelos o técnicas estadísticas de pronóstico se utilizan con mayor éxito en el mercado eléctrico, a través de un análisis de revisión bibliográfica.
- Identificar las principales variables que puedan determinar la variabilidad del precio de bolsa en Colombia.
- Desarrollar distintos modelos analíticos de pronóstico del precio de bolsa de energía en Colombia y evaluar su precisión a través de métricas estadísticas de comparación.
- Proponer bases para la implementación del modelo seleccionado dentro de la organización para la toma de decisiones.

3. Alcance

El alcance de este proyecto aplicado es proponer, desarrollar y validar un modelo analítico de pronósticos del precio de bolsa de energía en Colombia para que los clientes de COSENIT que son participantes y actores principales del mercado eléctrico puedan contar con información predictiva para la toma de sus decisiones. Los resultados esperados asociados a este proyecto son los siguientes:

- Marco conceptual, resultado de la revisión de literatura, asociado a modelos y técnicas de pronóstico aplicables a los mercados de energía específicamente a la predicción del precio de bolsa con la identificación de las principales técnicas.
- Variables identificadas a través de criterio experto que puedan afectar el comportamiento del precio de bolsa de energía. Se realiza la selección, recolección, descripción y su análisis de estadística descriptiva de manera visual.
- Selección de las metodologías de analítica y el desarrollo de los modelos bajo herramientas de generación de código abierto (Python) con la finalidad de seleccionar el de mejor desempeño.
- Dar las bases para el despliegue del modelo seleccionado dentro de la organización a través de una recomendación de evaluación de las competencias y la estructura organizacional para su adopción e implementación para la toma de decisiones.

4. Contexto Sectorial y Organizacional

4.1.Contexto Sectorial

El sector energético en Colombia es un sector que corresponde a las actividades primarias de la economía, basadas en el aprovechamiento de los recursos naturales renovables y no renovables a través de diferentes tecnologías para la generación y consumo de energía eléctrica. Esto con el fin de promover el crecimiento económico sostenible y el desarrollo de la sociedad en su conjunto.

Al igual que la mayoría de los países de América Latina Colombia implementó estrategias de apertura y liberalización económica, las cuales alcanzaron su punto máximo a principios de los años noventa. Estas medidas se centraron en fomentar la participación del sector privado con el objetivo de mejorar la eficiencia en la prestación de servicios. Como resultado, se creó un entorno propicio para implementar las transformaciones estructurales necesarias en el sector eléctrico colombiano (Lopez,2015).

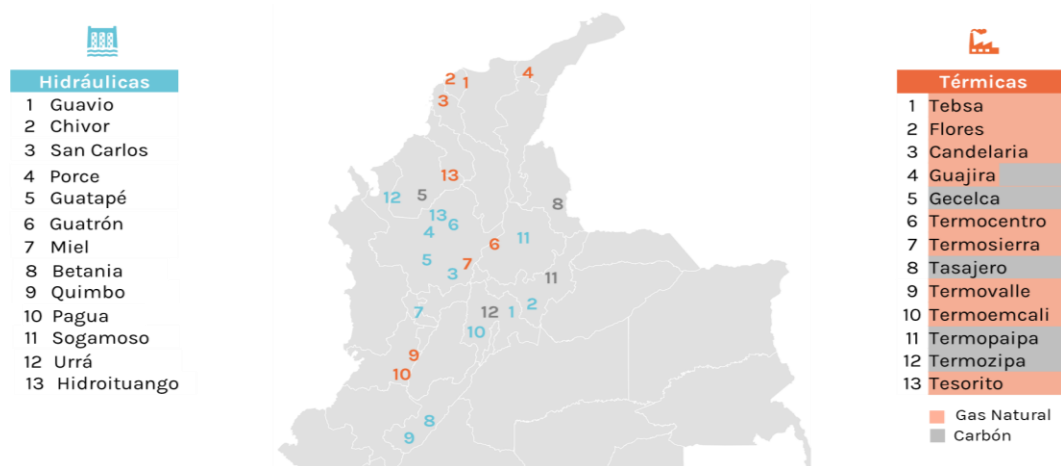
Como resultado, surgió un mercado mayorista altamente competitivo, con el objetivo de promover la eficiencia en la provisión de servicios eléctricos y permitir la entrada libre de agentes interesados en participar en este sector. Este mercado se conoce como el Mercado de Energía Mayorista (MEM), en el cual participan actores involucrados en actividades como generación, transmisión, distribución, comercialización y grandes consumidores de electricidad (Lopez,2015). A continuación, se describen brevemente estas actividades:

Generación: Consiste en la producción de energía eléctrica bajo diferentes tecnologías principalmente las de tipo convencional (plantas hidráulicas y térmicas) y no convencional (eólica, solar, biomas, otras renovables). El Estado debe garantizar que esta actividad se desarrolle por medio de una estructura competitiva, evitando el abuso de poder

de mercado, donde se busca a través de medidas la regulación garantizar el suministro de energía lo más eficientemente posible para todos los usuarios (Lopez,2015).

El sector tiene alrededor de 26 plantas de generación con energía convencional (Ver Figura 1). Sin embargo, desde el año 2016 se ha venido promoviendo por parte de los diferentes gobiernos, la entrada de nuevas tecnologías de tipo renovable no convencional, como la eólica y solar, a través de mecanismos como incentivos tributarios y aspectos regulatorios favorables para incrementar su participación en el mercado eléctrico colombiano.

Figura 1. Plantas de Generación en Colombia

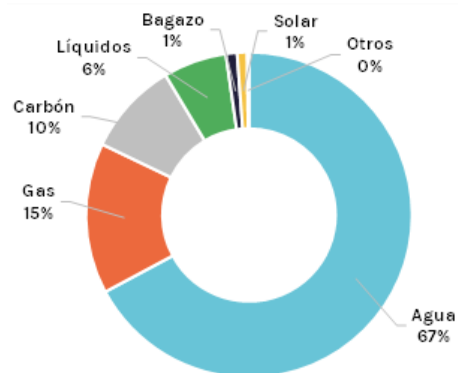


Fuente: Elaboración Propia

El sector eléctrico en Colombia está mayormente dominado por generación de energía hidráulica (67% de la producción) donde las plantas generan electricidad por medio del almacenamiento de agua en embalses o mediante el aprovechamiento del cauce de los ríos (planta filo de agua), es por esto que los cambios climáticos tales como fenómenos de “La Niña” y “El Niño” tienen gran impacto en los precios de la energía dada su alta dependencia a esta tecnología. La generación restante está dada por generación térmica (31%) y las nuevas

tecnologías de energía renovable (principalmente eólica, solar y biomasa) que han ido incursionado en la matriz energética nacional, pero actualmente no representa una parte importante en la estructura de capacidad y generación eléctrica en el país (ver Figura 2).

Figura 2. Generación y Capacidad Instalada



Fuente: Elaboración Propia

Dada la dependencia de la generación hidráulica en el mercado eléctrico colombiano, prever las alteraciones climáticas como los cambios en el nivel de aportes hidráulicos, llevó a la implementación de un mecanismo denominado cargo por confiabilidad, con el fin de garantizar el suministro y evitar posibles racionamientos de energía. El cargo por confiabilidad es un mecanismo de remuneración implementado por la CREG, con el fin de garantizar de manera eficiente la atención de la demanda de energía en condiciones críticas de abastecimiento, es decir, permite hacer viable la inversión en los recursos de generación necesarios garantizando unos ingresos a las plantas de generación y fomentando señales de eficiencia económica para la atención de la demanda. Por lo tanto, la demanda eléctrica debe incorporar en su tarifa el costo de este esquema y esto recibe el nombre de Costo Equivalente Real de Energía (CERE) y en general es el precio piso o mínimo que se establece para la formación de precios en el mercado.

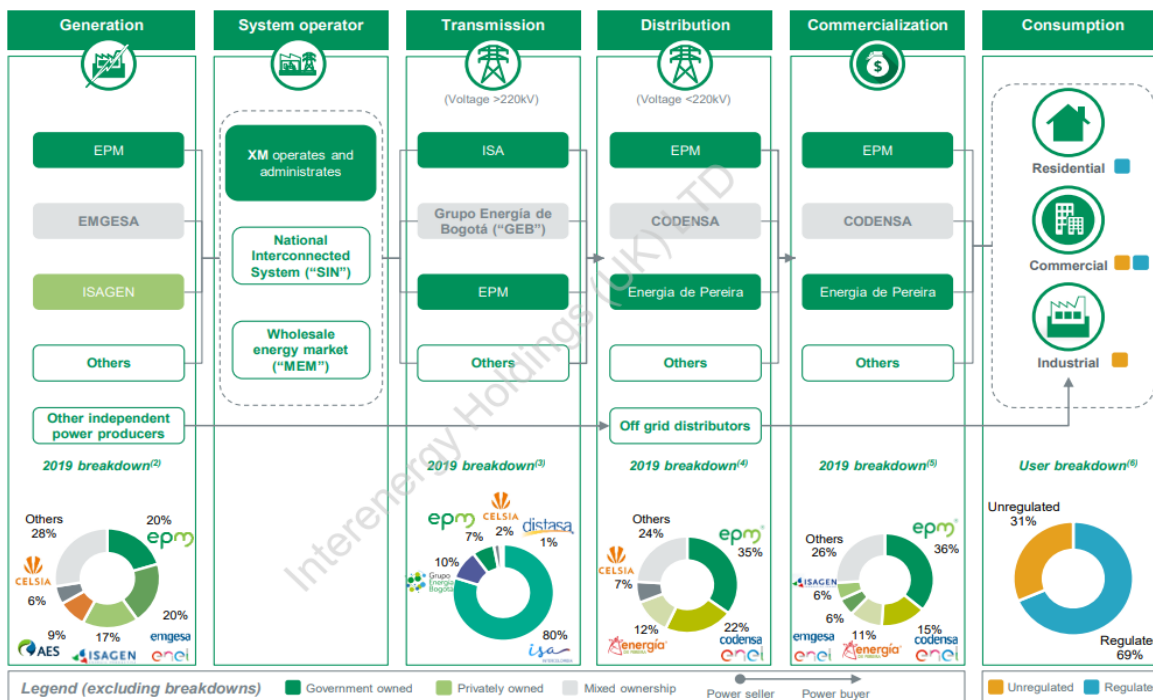
Transmisión: Esta actividad se encarga del transporte de energía desde las centrales o plantas de generación hasta los grandes centros de consumos a altos niveles de tensión a través del Sistema de Transmisión Nacional (STN). Donde el STN es el conjunto de líneas y subestaciones con equipos y transformadores que operan altas tensiones (220-500 kV). La Ley 143 de 1994, indica esta actividad como monopolio natural y por tanto el Estado debe regular el ingreso de cualquier actor (Lopez,2015).

Distribución: El suministro de energía eléctrica antes de llegar al consumidor final implica su transformación a niveles de voltaje intermedios y su transporte a través de redes, subestaciones y transformadores, hasta alcanzar los puntos de consumo. Este proceso de transporte de bloques más pequeños de energía con destino al consumidor final se conoce como distribución. En Colombia, la distribución de la energía se lleva a cabo mediante los Sistemas de Transmisión Regionales (STR) y los Sistemas de Distribución Local (SDL). (Lopez,2015).

Comercialización: Esta actividad implica la adquisición de energía en el mercado mayorista con el propósito de venderla a los usuarios finales. Esta actividad se desarrolla en un entorno competitivo. El rol del Estado se centra en garantizar el abuso de poder de mercado y establecer normas que salvaguarden el bienestar de los usuarios finales.

Demanda: En esta actividad se pueden distinguir dos tipos de usuarios: los usuarios regulados y los usuarios no regulados. Los usuarios regulados no tienen la capacidad de contratar directamente su suministro de energía, sino que son atendidos por un comercializador que actúa en su representación en el mercado. Por otro lado, los usuarios no regulados tienen la libertad de elegir y contratar directamente con un comercializador de su elección (Lopez,2015).

Figura 3. Agentes Sector Eléctrico Colombiano



Fuente: COSENTIT- Interenergy Document

En el mercado energético colombiano, existen tres entidades reguladoras principales. La Comisión de Regulación de Energía y Gas (CREG) es un organismo estatal encargado de regular las actividades relacionadas con la prestación de servicios públicos domiciliarios, incluyendo la energía eléctrica, el gas natural, el gas licuado de petróleo y los combustibles líquidos (CREG,2023). De acuerdo con las leyes 142 y 143 de 1994 la CREG tiene como función regular los monopolios en la prestación de servicios públicos y crear las condiciones para asegurar la disponibilidad de una oferta energética eficiente capaz de abastecer la demanda bajo criterios sociales, económicos, ambientales y de viabilidad financiera, promoviendo y preservando la competencia.

Por otro lado, la Superintendencia de Servicios Públicos Domiciliarios (SSPD) tiene la responsabilidad de controlar y vigilar a las entidades que prestan servicios públicos domiciliarios (SSPD,2023). Esta entidad se encarga de supervisar el cumplimiento de las regulaciones, proteger los derechos de los usuarios y promover la eficiencia en la prestación de servicios. Además, la Unidad de Planeación Minero-Energética (UPME) es una unidad adscrita al Ministerio de Minas y Energía de Colombia. Su función principal es elaborar los planes de expansión del sector eléctrico, teniendo en cuenta aspectos financieros, económicos y ambientales (UPME,2023). La UPME juega un papel clave en la planificación estratégica y el desarrollo sostenible del sector energético en el país.

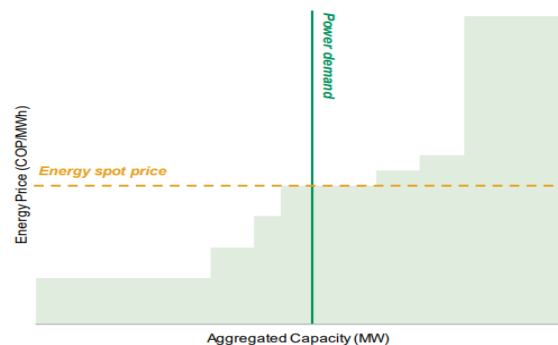
Como se ha mencionado, este sector está caracterizado por ser estratégico y transversal de todas las actividades de la economía nacional, dado su impacto como materia prima y/o comodities para el desarrollo industrial, comercial y residencial. Por tanto, las negociaciones y las maneras de comercializar la energía como servicio son fundamentales para los diferentes actores tanto de la oferta (generadores y comercializadores) como para toda la demanda en sus diferentes categorías de consumo, donde tener información precisa permitiría aprovechar oportunidades o bien mitigar riesgos asociados al proceso de compra y venta de energía.

Uno de los mecanismos utilizados para la transacción de cantidades de energía, según las regulaciones, se centra en las negociaciones bilaterales. Estas negociaciones consisten en establecer contratos a largo plazo entre las partes, acordando un precio para la energía en un momento específico, teniendo en cuenta las condiciones y los riesgos del mercado que se transfieren a través del precio. El objetivo de estos contratos es reducir la volatilidad que puede surgir en los precios del mercado. Estos contratos se respaldan en adquisiciones

necesarias realizadas en la bolsa de energía o mediante acuerdos con otros participantes del Mercado de Energía Mayorista (Gómez et al., 2020).

El segundo mecanismo y objeto de análisis de este proyecto, está enfocado en el mercado spot o precio de bolsa de energía. Su estimación está a cargo del operador del mercado, quien calcula diariamente la demanda de energía agregada del país. Como se observa en la figura, se ordenan los recursos de generación (plantas hidráulicas, térmicas, renovables no convencionales) por mérito, es decir los agentes de la oferta, indican al operador del mercado un único precio al cual están dispuestos a vender su energía, y de acuerdo con estos precios, el operador del mercado los ordena de menor a mayor formando la curva de oferta hasta que se cruce con la curva de demanda. De esta manera, los recursos más económicos son los que entran al despacho de generación, y el precio spot es aquel precio del último recurso de generación que satisface a la demanda.

Figura 4. Formación Precio Spot de Energía

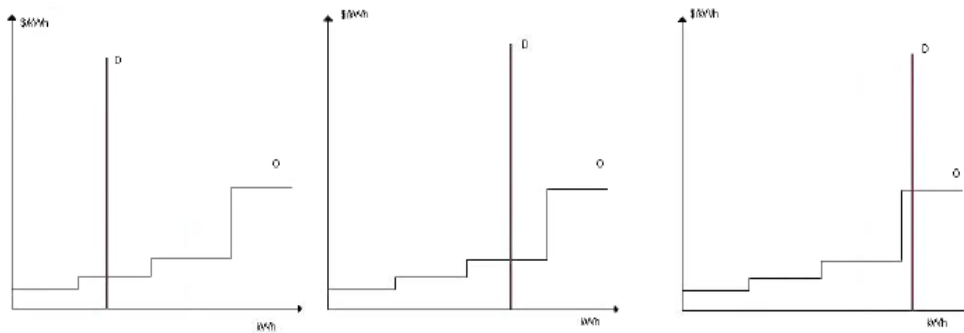


Fuente: COSENIT- Interenergy Document

El incremento de la demanda tiene un impacto positivo en el precio, al igual que en cualquier otro mercado. No obstante, la relación entre la demanda y el precio está condicionada por la forma de la curva de oferta. Para ilustrar este punto, en la siguiente figura

se observa que, bajo una misma curva de oferta (O), los aumentos en la demanda (D) generan un incremento no lineal en el precio. Esta interacción ocurre en el punto de intersección entre ambas curvas, como se mencionó anteriormente. Debido a la forma específica de la curva de oferta, es de esperar que la relación entre la demanda y el precio también dependa de otros factores fundamentales y que no sea lineal.

Figura 5. Curvas de Oferta y Demanda de Energía



Fuente: Tomado de Gómez et al. (2020)

El precio spot de la electricidad, está influenciado por muchos factores interconectados de manera compleja, y es un indicador importante para el sistema de energía eléctrica. Por ejemplo, los resultados de pronósticos precisos pueden guiar a los consumidores, así como el programa de producción de los generadores, ayudando así a los participantes a obtener los máximos beneficios y ayudando a los gerentes a diseñar un plan óptimo de operación del mercado de energía. Además, la previsión también juega un papel vital en la gestión de la inversión.

Desarrollar un modelo de pronóstico de electricidad efectivo y de alta calidad es un tema desafiante para todos los participantes en el mercado de la energía, debido a sus

características inherentes que incluyen la no linealidad, alta volatilidad, alta frecuencia, sensibilidad a la demanda y cambios en las fuentes de generación (Jiang et al.,2023)

La formación del precio spot o de bolsa en Colombia marca un parámetro relevante para los tomadores de decisiones tanto de la oferta como la demanda, dado que poder estimar o pronosticar de manera certera les permite a los participantes del mercado eléctrico colombiano decidir como invertir, producir y consumir energía, estructurando posiciones y portafolios de compra y venta de energía.

4.2.Contexto Organizacional

Es en esta toma decisiones donde la organización eje de este proyecto toma un rol fundamental, debido a su comportamiento de prestación de servicios a los agentes del mercado eléctrico nacional. COSENIT es una organización con más de veinte años de experiencia siendo experta en consultoría y asesoría a todos los agentes de la cadena del mercado (desde la generación hasta la comercialización) y en gran medida contribuyendo a la demanda nacional (principalmente industrial y comercial) a la toma de decisiones informadas para la gestión, control y reducción de sus costos energéticos. Sus principales servicios están agrupados en cuatro líneas de negocio:

- I. Desarrollo y adquisición de proyectos energéticos en Colombia.
- II. Inteligencia de mercados energéticos en Latinoamérica.
- III. Análisis del entorno regulatorio y administración contractual.
- IV. Estrategias de compra y/o venta de energía eléctrica en el mercado spot.

La línea de negocio que este proyecto empresarial busca abordar es la IV, la cual para el cierre financiero del 2022 representó alrededor del 35% de los ingresos de la compañía, y

con una proyección de crecimiento considerable (expectativa de 10% más de ingresos en los próximos 2 años), debido a una alta demanda de pronósticos, elaboración de estrategias y portafolios de compra y venta de energía por diferentes agentes del mercado. Este particular se da, por la promoción por parte del Gobierno, a la entrada de nuevos proyectos de generación, especialmente de tecnologías renovables no convencionales, donde sus sponsors para una debida diligencia deben validar en especial sus ingresos y costos a futuro con las proyecciones basadas en el mecanismo del precio de bolsa.

La compañía cuenta con grandes oportunidades en este rubro de negocio, como se acaba de mencionar existe en el sector una alta demanda por este servicio, la credibilidad y la confianza por parte de los actores del mercado hacia COSENIT por su trayectoria es fuerte. Adicional, cuenta con personal capacitado a nivel académico y empresarial para realizar procesos que generen valor agregado.

Sin embargo, la compañía en su definición estratégica ha visto debilidades en el proceso actual de esta línea de negocio, en especial en todo el tema de competencia a nivel de analítica a la vanguardia de las nuevas tecnologías de la información. El proceso actualmente de pronóstico del precio de energía se maneja de la siguiente manera:

- I. Solicitud de cliente (nuevo o actual) de proyecciones de precios de energía bajo una periodicidad diaria o mensual.
- II. El equipo de COSENIT identifica la información necesaria bajo criterios de experto, principalmente las variables explicativas de la volatilidad del precio de bolsa.
- III. Recopila la información, almacenando como archivos nuevos y particulares de esa solicitud en la herramienta Excel.

- IV. Bajo la herramienta anterior, genera un modelo de regresión lineal múltiple, donde se establece la ecuación lineal para predecir valores futuros.
- V. Se estiman valores futuros y se le comparte al cliente dos o más escenarios posibles de proyecciones para el periodo solicitado.

Del proceso anterior, las principales debilidades identificadas están basadas en que no se tiene trazabilidad en la definición, recolección y almacenamiento de la información. Para cada servicio de consultoría solicitado se elaboran archivos que no guardan relación, ni históricos. La cantidad de información que actualmente suministra el mercado eléctrico colombiano es de gran volumen y no se cuenta con un servicio de almacenamiento on premise o en la nube que permita gestionar dicha cantidad de datos y aprovecharlos al máximo para las solicitudes de los clientes.

Adicionalmente, solo se utiliza un modelo de tipo lineal, el cual no tiene actualmente un desarrollo en un software estadístico robusto, ni cuenta con las pruebas de rigurosidad para que sus pronósticos sean validados, ni comparables con otros modelos estadísticamente y se pueda entregar un nivel de certeza adecuado al servicio del cliente. También se encuentra actualmente a nivel organizacional inmersa en la problemática de la percepción de potenciales clientes de consultoría, respecto a la capacidad de tener herramientas empresariales analíticas que permitan contribuir de manera asertiva a la toma de decisiones.

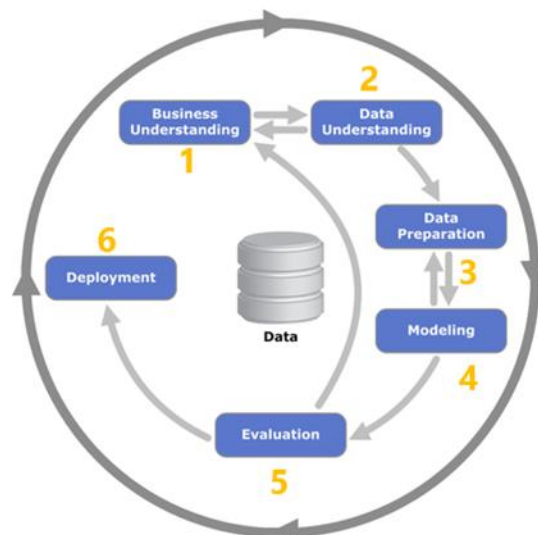
Es así como desde COSENIT se ha visto con urgencia la necesidad de complementar sus capacidades analíticas, aprovechar la alta demanda del mercado por este tipo de servicio potenciando su línea de negocio actual y ampliar su oferta de valor a sus clientes.

5. Estrategia Para la Solución

El proyecto empresarial se aborda desde la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) la cual se ha convertido en un estándar en proyectos de *Data Science* al enfocar sus resultados al entorno del negocio, basada en procesos estándares es una solución de código abierto que describe las actividades comunes que utilizan los expertos en analítica y ciencia de datos (IBM, 2023). En la siguiente Figura la metodología CRISP-DM establece un proyecto de minería de datos como una secuencia de fases:

1. Comprensión del negocio.
2. Compresión de los datos.
3. Preparación de los datos.
4. Modelado.
5. Evaluación.
6. Despliegue.

Figura 6. Metodología CRISP-DM



Fuente: Elaboración Propia a partir de IBM (2021)

5.1. Comprensión del Negocio

Con la problemática y necesidad expuesta en la sección 4.2 por parte de la organización COSENIT, se ha determinado trabajar a través de este proyecto en el desarrollo de un modelo analítico de predicción del precio de bolsa de energía en Colombia, buscando abordar principalmente la deficiencia de los modelos actuales internos, capturar oportunidades de negocio y mejorar la percepción de los clientes respecto al uso de herramientas analíticas que soporten con mayor grado de robustez las recomendaciones para su toma de decisiones en el mercado transaccional spot para estructurar sus compras y ventas de energía.

Se presenta inicialmente como elemento para abordar la estrategia de solución al problema central de desarrollar una herramienta analítica de pronóstico para el mercado eléctrico, la revisión de literatura para determinar qué modelos han sido desarrollados y, por tanto, cuáles podrían ser a nivel teórico los más adecuados para estructurar metodológicamente el modelo de pronóstico a generar al interior de COSENIT.

Se analizaron 25 artículos de investigación tanto a nivel nacional como internacional, la gran mayoría de años recientes. A continuación, se presenta una tabla resumen:

Tabla 1. Marco Conceptual

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
Yang, W., Wang, J., Niu, T., & Du, P. (2020).	A novel system for multi-step electricity price forecasting for electricity market management, Applied Soft Computing	Regresión Lineal Múltiple Regresión Lasso Análisis de Componentes Principales	Precios históricos de la electricidad Precios futuros del gas Demanda de electricidad Temperatura Humedad	El modelo se basa en la regresión lineal múltiple y utiliza técnicas de aprendizaje automático, como el análisis de componentes principales (PCA) y la regresión LASSO (Least Absolute Shrinkage and Selection Operator) para mejorar la precisión de la predicción. Además, se utilizan técnicas de suavizado para predecir los precios de la electricidad a varios pasos. Los resultados del estudio muestran que el modelo propuesto es capaz de predecir con precisión los precios de la electricidad a varios pasos. Se encontró que la incorporación de datos meteorológicos

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
				mejoró significativamente la precisión de la predicción en comparación con los modelos que solo utilizan datos históricos de precios de la electricidad. Además, se demostró que las técnicas de suavizado mejoraron aún más la precisión de la predicción a varios pasos.
Dejamkhooy, A., & Ahmadpour, A. (2022).	Prediction and Evaluation of Electricity Price in Restructured Power Systems Using Gaussian Process Time Series Modeling	Modelo Gaussiano	Precios históricos de electricidad Precio Combustibles Demanda de electricidad Capacidad de generación	En este estudio, se consideran dos sistemas de potencia reestructurados, y los precios de bolsa de estos sistemas se predicen de forma completa y precisa utilizando un modelo de proceso gaussiano (GP) que se adapta para predicciones de series temporales. En este modelado, se utilizan e investigan varios modelos de GP, incluidos dinámicos, estáticos, directos e indirectos, así como sus modelos mixtos. La eficacia y precisión de estos modelos se comparan utilizando indicadores de evaluación apropiados. Los resultados muestran que las combinaciones de los modelos GP tienen errores más bajos que los modelos individuales, y el GP indirecto dinámico fue elegido el mejor modelo.
Jiang, P., Nie, Y., Wang, J., & Huang, X. (2023).	Multivariable short-term electricity price forecasting using artificial intelligence and multi-input multi-output scheme,	Modelo Híbrido de Aprendizaje Automático (Redes Neuronales, Maquinas de Soporte Vectorial)	Precio histórico de electricidad Demanda de electricidad	Formulan un novedoso sistema de multi-bi-pronóstico del precio de la electricidad que utiliza estructuras multivariadas y multientradas y multisalidas. El sistema tiene tres etapas: preprocesamiento de datos, pronóstico combinado y evaluación del desempeño. Los resultados finales se obtienen mediante una estrategia combinada basada en el algoritmo de enjambre. Finalmente, se realizan tres experimentos en el mercado eléctrico australiano para evaluar cuantitativamente el sistema propuesto. Los resultados muestran que el sistema diseñado tiene una capacidad superior para pronosticar el precio de bolsa.
Singh, P., & Kottath, R. (2022)	Influencer-defaulter mutation-based optimization algorithms for predicting electricity prices	Red Neuronal Artificial (ANN)	Precios históricos de la electricidad Demanda de electricidad Temperatura Humedad Precios del combustible	El enfoque propuesto utiliza algoritmos de optimización basados en mutación, denominados influencer-defaulter mutation-based optimization (IDMBO), para optimizar un modelo de redes neuronales artificiales (ANN) utilizado para predecir los precios de la electricidad en un mercado eléctrico. En particular, el modelo ANN se utiliza para predecir los precios de la electricidad en el mercado al día siguiente. Los algoritmos IDMBO se utilizan para optimizar los pesos y sesgos del modelo ANN para mejorar la precisión de la predicción. En general, el estudio demuestra que el enfoque propuesto puede ser útil para mejorar la gestión del mercado de la electricidad y para la toma de decisiones en tiempo real en el mercado.

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
Wang, J., Liu, F., Song, Y., & Zhao, J. (2016).	A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system,	Modelo Híbrido de Redes Neuronales	Precios históricos de la electricidad Factores macroeconómicos Factores meteorológicos (como la temperatura y la humedad)	En este documento se propone un modelo llamado DCANN para pronosticar el precio diario de la electricidad. Este modelo es un sistema híbrido que utiliza una combinación de series de tiempo de precios históricos de la electricidad, factores macroeconómicos y meteorológicos, y factores específicos del mercado de la electricidad para predecir los precios de la electricidad en un mercado de energía. La selección de características se realiza mediante una técnica de selección de características basada en la correlación.
Wang, J., Yang, W., Du, P., & Niu, T. (2020).	Outlier-robust hybrid electricity price forecasting model for electricity market management	Modelo Híbrido (Regresión Lineal PLS y Redes Neuronales ANN)	Precios históricos de la electricidad Factores meteorológicos (como la temperatura y la humedad)	El modelo propuesto en este estudio es un modelo híbrido que combina la regresión por mínimos cuadrados parciales (PLS) y la regresión por redes neuronales artificiales (ANN) para predecir los precios de la electricidad. Para hacer que el modelo sea robusto frente a valores atípicos, se utiliza una técnica de detección de outliers basada en el análisis de componentes principales (PCA) y la distancia de Mahalanobis.
Marcjasz, G., Lago, J., & Weron, R. (2020).	Neural networks in day-ahead electricity price forecasting: single vs. multiple outputs	Redes Neuronales	Precios históricos de la electricidad Demanda y oferta de energía	El estudio utiliza datos históricos de precios de la electricidad y variables relacionadas con el mercado energético, como la demanda y la oferta, para entrenar y evaluar dos enfoques de redes neuronales: redes neuronales de salida única (SNN) y redes neuronales de múltiples salidas (MNN).
Kuo, P.-H., & Huang, C.-J. (2018).	An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks	Redes Neuronales (CNN y LSTM)	Precios históricos de la electricidad Demanda y oferta de energía	Este artículo propone un sistema de pronóstico del precio de la electricidad basado en la combinación de 2 redes neuronales profundas, la Red Neural Convolutiva (CNN) y la Memoria a Largo Corto Plazo (LSTM). Para comparar el rendimiento general de cada algoritmo, se aplicaron medidas de evaluación del error absoluto medio (MAE) y del error cuadrático medio (RMSE) en los experimentos de este documento. Los resultados del experimento muestran que, en comparación con otros métodos tradicionales de aprendizaje automático, el rendimiento de predicción del modelo de estimación propuesto en este documento ha demostrado ser el mejor. Al combinar los modelos CNN y LSTM, la viabilidad y practicidad de la predicción del precio de la electricidad también se confirma en este documento. se ha demostrado que el rendimiento de predicción del modelo de estimación propuesto en este documento es el mejor.
Grothe, O., Kächele, F., & Krüger, F. (2023).	From point forecasts to multivariate probabilistic forecasts: The Schaake shuffle for	Redes Neuronales (ANN) Regresión (SVR, SVR, Gradient Boosting, Decision Tree)	Precios históricos de la electricidad Demanda y oferta de energía Características	El método propuesto utiliza un enfoque basado en ensamblajes para generar predicciones probabilísticas multivariadas de los precios de la electricidad. En particular, se utiliza el método Schaake Shuffle para generar un conjunto de simulaciones de las condiciones meteorológicas futuras. Estas simulaciones se utilizan junto con los datos

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
	day-ahead electricity price forecasting,	Modelos Híbridos	meteorológicas (como la temperatura, la humedad y la velocidad del viento)	históricos para entrenar modelos de redes neuronales artificiales (ANN) y modelos de regresión. Luego, se utiliza un ensamblaje ponderado de estos modelos para generar predicciones probabilísticas multivariadas. El estudio destaca la importancia de considerar la incertidumbre en las predicciones de precios de la electricidad y cómo los enfoques probabilísticos pueden mejorar la toma de decisiones en el mercado de energía.
Weron, R. (2014).	Electricity price forecasting: A review of the state-of-the-art with a look into the future	Modelos Supervisados de Regresión ARIMA, GARCH Redes Neuronales	Datos históricos de precios de la electricidad Demanda y oferta de energía Precios de los combustibles Características meteorológicas	El estudio revisa una amplia variedad de técnicas utilizadas en la predicción de precios de la electricidad, desde métodos estadísticos tradicionales hasta técnicas de inteligencia artificial y aprendizaje automático. También se analizan los desafíos a los que se enfrenta la predicción de precios de la electricidad, como la falta de datos, la volatilidad de los mercados y la complejidad del sistema eléctrico. Algunos de los enfoques revisados incluyen modelos estadísticos como ARIMA y GARCH, técnicas de regresión y redes neuronales artificiales, así como enfoques más nuevos como el aprendizaje profundo y la optimización basada en mutaciones.
Gunduz, S., Ugurlu, U., & Oksuz, I. (2023).	Transfer learning for electricity price forecasting	Modelos de Regresión (SVR, Lineal, SVM) Redes Neuronales	Datos históricos de precios de la electricidad Demanda y oferta de energía	El aprendizaje de transferencia se realiza utilizando modelos de aprendizaje automático, como redes neuronales artificiales (ANN) o modelos de regresión. Se entrenan modelos en un mercado fuente utilizando datos históricos de precios de la electricidad y otras variables relevantes, y luego se transfieren a un mercado objetivo. El modelo se ajusta utilizando datos limitados del mercado objetivo, mejorando la precisión de las predicciones de precios. El método mejora el rendimiento de los algoritmos de última generación en un 7 % para el mercado francés y en un 3 % para el mercado alemán.
Villada, F., Cadavid, D. R., & Molina, J. D. (2014).	Electricity price forecasting using artificial neural networks.	Redes Neuronales GARCH	Precio Histórico Electricidad Niveles de reserva de agua	En este trabajo se propone un modelo para pronosticar el precio de la electricidad en Colombia utilizando redes neuronales artificiales. Se utilizan dos estructuras de redes neuronales que incluyen la serie de precios en la primera y la serie de precios más los niveles de reserva de agua en la segunda. Los resultados se comparan con un modelo de modelo heteroscedástico condicional autorregresivo generalizado (GARCH), que muestra un mejor ajuste dentro del período de entrenamiento, pero las redes neuronales tienen una mejor previsión de rendimiento fuera de la muestra de entrenamiento

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
Marin, J. B., Orozco, E. T., & Velilla, E. (2018).	Forecasting electricity price in Colombia: A comparison between Neural Network, ARMA process and Hybrid Models	ARMAX NARX	Precio Histórico Electricidad	Este estudio tiene como objetivo predecir los precios de la electricidad en el mercado eléctrico colombiano. Para lograr este objetivo, se han implementado análisis econométricos de series temporales convencionales y una técnica alternativa basada en algoritmos de inteligencia artificial. Utilizamos modelos autorregresivos de promedio móvil (ARMAX) y redes neuronales autorregresivas no lineales (NARX). Luego de estimar un modelo híbrido que combina los modelos ARMAX y NARNX, incluyendo insumos exógenos, pronosticamos una serie de tiempo de precios de electricidad en un horizonte de 12 meses (
Oviedo-Gomez, A., Londono-Hernandez, S. M., & Manotas-Duque, D. F. (2021)	Electricity Price Fundamentals in Hydrothermal Power Generation Markets Using Machine Learning and Quantile Regression Analysis	Aprendizaje Automático (SVM, SVR) Regresión	Demanda Capacidad Generación Embalses Consumo Combustibles Precio Electricidad	Este estudio propone un enfoque empírico para identificar los determinantes de los precios y sus efectos en la dinámica de precios. Este documento presenta dos metodologías: un enfoque de aprendizaje automático y un análisis de regresión por cuantiles. El primer método se utiliza para validar los determinantes del precio a través de un proceso de predicción, y el segundo, la regresión por cuantiles, para identificar los efectos no lineales. Los factores más importantes observados son la demanda total del mercado, la capacidad de generación de los embalses de agua y el consumo de combustibles fósiles. Los resultados ofrecen una nueva perspectiva sobre la estructura del mercado y la volatilidad de los precios al contado
Barrientos, J., & Martínez, M. T. (2017).	Análisis de los fundamentales del precio de la energía eléctrica: evidencia empírica para Colombia	Modelo Econométrico VAR	Demanda Hidrología Disponibilidad Generación	En este trabajo estudian los fundamentales del mercado que afectan la formación de los precios de la energía eléctrica en Colombia, así como evalúan el efecto de choques positivos en algunas variables identificadas como responsables de la formación del precio. Para el objetivo se estiman procesos VAR. Adicionalmente se lleva a cabo un ejercicio de pronósticos exploratorios para determinar la trayectoria futura del precio de la energía en los próximos diez años. La conclusión principal del trabajo es que, dadas las condiciones del mercado eléctrico colombiano, las variables que afectan principalmente los precios de la energía son: la demanda, la hidrología y la disponibilidad declarada.
Galón, S., & Barrientos, J. (2021).	Forecasting the Colombian Electricity Spot Price under a Functional Approach	Descomposición de Componentes Principales (PCA)	Precio Electricidad por tipo de día	En este trabajo, apelamos a un punto de vista de Análisis Funcional de Datos (FDA) que permite modelar y pronosticar el precio spot intradiario de la electricidad del Mercado Eléctrico Colombiano. Específicamente, usamos el método Hyndman-Ullah-Shang, que se basa en una descomposición de componentes principales funcionales de las curvas de precios suavizadas no paramétricas, donde los pronósticos a corto plazo se obtienen utilizando los componentes

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
				principales funcionales empíricos y los pronósticos de series temporales univariadas de las puntuaciones estimadas correspondientes.
Lehna, M., Scheller, F., & Herwartz, H. (2022).	Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account	ARIMAX LSTM VAR Modelos Híbridos	Precio Electricidad Carga Consumidor Combustibles Radiación solar Velocidad Viento	Estudio comparativo de cuatro enfoques diferentes para pronosticar el precio al contado de la electricidad para el día siguiente en Alemania. Además del destacado modelo de media móvil autorregresiva integrada estacional ((S)ARIMA(X)) y los modelos de red neuronal de memoria a largo y corto plazo (LSTM), empleamos una red neuronal convolucional LSTM (CNN-LSTM) y un enfoque de modelo autorregresivo vectorial (VAR) multivariante extendido de dos etapas como modelos híbridos. Para un mejor rendimiento, incluyeron influencias externas comunes como la carga del consumidor, el combustible y precios de emisión, radiación solar media y velocidad del viento en el análisis.
Santiago, A. M., Vanstrahlen, J. U., Otero, A. C., & Lombana, J. (2017).	Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH	ARIMA GARCH ARCH	Precio promedio de los treinta días de cada mes	El modelo sugiere que si las variables de estudio presentan características como cambios abruptos en periodos cortos de tiempo (alta volatilidad causada por fenómenos como El Niño y La Niña), distribuciones asimétricas y no cumplen con los supuestos de estacionariedad, entre otros factores, es más apropiado aplicar modelos como ARCH, GARCH y sus derivados. Estos modelos son adecuados para abordar la heterocedasticidad, es decir, la presencia de una varianza no constante en los datos. Al considerar la heterocedasticidad, estos modelos permiten capturar mejor la volatilidad y las características particulares de los datos, mejorando así la precisión de las predicciones.
Galindo Ortiz, A. (2017).	Modelamiento de los precios de la energía en bolsa en Colombia incorporando el efecto del ENSO	Modelo de reversión a la media estimado por máxima verosimilitud, regímenes y probabilidades de transición cambiantes	Índice de Oscilación al Sur (SOI); variación de la temperatura superficial oceánica (SSTA);	El modelo de reversión a la media, cuyos parámetros fueron estimados utilizando el método de máxima verosimilitud, demostró ser efectivo para modelar los procesos subyacentes del precio de la energía. Los resultados obtenidos mostraron un buen ajuste a la serie histórica y lograron incorporar de manera adecuada los impactos causados por la ocurrencia e intensidad del fenómeno El Niño-Southern Oscillation (ENSO). Esto indica que el modelo fue capaz de capturar y reflejar los choques generados por la llegada e intensidad del ENSO en los precios de la energía

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
Agudelo, Adriana P, López-Lezama, Jesús M, & Velilla, Esteban. (2015).	Predicción del Precio de la Electricidad en la Bolsa mediante un modelo Neuronal No-lineal autorregresivo con entradas exógenas	Redes neuronales artificiales	Relación entre generación hidráulica y térmica; Volumen útil diario de energía kWh-mensual; Aporte de energía kWh-mensual; probabilidad del fenómeno de El Niño; Caudal promedio mensual;	Los resultados obtenidos con el modelo NARX-ARIMA seleccionado son coherentes con la serie de precios real, lo que se puede observar a través de un alto coeficiente de determinación del 96% entre los datos reales y los generados por el modelo. Además, se encontró que la correlación muestral (ACF) y la autocorrelación muestral parcial (PACF) del error se sitúan dentro de los intervalos de confianza del 95%. Un hallazgo interesante fue la variable que presenta una mayor correlación con el precio de bolsa, que es la relación entre la generación hidráulica y la generación térmica. También se tuvo en cuenta la demanda de energía eléctrica, que es una variable fundamental en los modelos económicos.
Díez, I. C. (2015).	Proyección de precios de energía eléctrica de mediano plazo en el mercado colombiano mediante la aplicación del índice de Lerner	Índice de Lerner	Costo Marginal	La aplicación de la metodología utilizada permitió realizar estimaciones de los precios futuros de la energía mediante la proyección de los Costos Marginales. Además, se estimó el índice de Lerner, el cual se calculó a partir de la medición de la concentración y la elasticidad. Estos cálculos proporcionaron información importante sobre la relación entre la oferta y la demanda en el mercado energético, y permitieron obtener una visión más precisa de cómo se espera que los precios de la energía se comporten en el futuro.
Uribe Gaviria, E. & Trespalacios Carrasquilla, A. (2014).	Contraste de modelos estocásticos para el precio de la energía en Colombia	Modelo determinístico y estocástico	Caudal Demanda; Generación;	Durante el estudio, se presentaron varios modelos para pronosticar el precio de la energía en Colombia, cada uno con sus respectivos parámetros y ecuaciones. Además de considerar el impacto del fenómeno El Niño, también se tuvieron en cuenta otras variables explicativas, como el nivel de los caudales, la demanda de energía y la generación de energía. Se observó que en los modelos que dependían únicamente de la ocurrencia del fenómeno El Niño, la varianza instantánea parecía depender de las variables explicativas mencionadas. Esto confirma la importancia de considerar estas variables en la construcción de modelos para pronosticar el precio de la energía en Colombia. Estos hallazgos destacan la relevancia de comprender y tener en cuenta múltiples factores que influyen en la formación de los precios de la energía, más allá del impacto aislado del fenómeno El Niño.

AUTOR	REFERENCIA	MÉTODO	VARIABLES	CONCLUSIONES
Hurtado Moreno, L., Quintero Montoya, O. L., & García Rendón, J. J. (2016).	Estimación del precio de oferta de la energía eléctrica en Colombia mediante inteligencia artificial	Redes neuronales + lógica difusa, ARIMA	ENSO Demanda comercial Precio de bolsa Aportes hídricos; Volumen embalse; Precios del gas natural; del petróleo y del carbón Aportes hídricos totales;	El estudio presentó diversas metodologías para estimar el precio ofertado por recurso energético en el Mercado Eléctrico Mayorista de Colombia. Se utilizaron modelos de Inteligencia Artificial y análisis de series de tiempo para abordar esta tarea. Además, se realizó un análisis de variables mediante modelos Logit, evaluando su significatividad en cambios porcentuales del precio ofertado. Estas metodologías permitieron obtener una visión integral y precisa del comportamiento del precio en el mercado energético, así como identificar las variables más relevantes en los cambios significativos.
Ramírez, (2013)	Comparación de la efectividad de diferentes herramientas para modelar el precio del mercado de electricidad en Colombia	ARIMA + ecuaciones diferenciales estocásticas (Wiener y Ornstein-Uhlenbeck)	Precio de la energía	El estudio abordó la estimación del precio ofertado por recurso energético en el Mercado Eléctrico Mayorista de Colombia a través de diversas metodologías. Se emplearon modelos de Inteligencia Artificial y análisis de series de tiempo para este propósito. Además, se realizó un análisis de variables mediante modelos Logit para evaluar su importancia en los cambios porcentuales del precio ofertado. Estas metodologías proporcionaron una visión integral y precisa del comportamiento del precio en el mercado energético, al tiempo que identificaron las variables más relevantes en los cambios significativos. En conjunto, este enfoque diversificado permitió obtener una comprensión sólida y detallada del mercado eléctrico colombiano.

Fuente: Elaboración Propia

Con el análisis de los artículos de investigación se evidenció que en la literatura se han propuesto diversos métodos de pronóstico del precio de bolsa de energía con diferentes objetivos, horizontes de tiempo y en algunas ocasiones incluyen o no variables exógenas. Sin embargo, dichos métodos pueden agruparse en tres grandes grupos:

1. Métodos basados en enfoques estadísticos, estos principalmente hacen uso de todas aquellas técnicas estadísticas para pronósticos de series de tiempo, en general asociadas a modelos de tipo econométrico, donde se busca identificar patrones y tendencias de tipo lineal e histórico de la variable a predecir, en este caso el precio de bolsa de energía y de otros factores exógenos. A menudo son criticados por su capacidad limitada para

modelar el comportamiento (generalmente) no lineal de los precios de la electricidad y las variables fundamentales relacionadas; sin embargo, en aplicaciones prácticas, su rendimiento es comparable al de sus alternativas no lineales

2. Métodos de *machine learning*, estos modelos utilizan algoritmos de aprendizaje automático, como árboles de decisión, bosques aleatorios, redes neuronales, para aprender patrones complejos en los datos históricos del mercado y hacer predicciones futuras. La principal fortaleza es su capacidad para manejar la complejidad de los datos en su varianza y la no linealidad.
3. Métodos híbridos, estos modelos combinan los enfoques estadísticos y de aprendizaje automático para mejorar la precisión de las predicciones. Por ejemplo, se puede utilizar un modelo de regresión lineal para modelar la tendencia a largo plazo y un modelo de redes neuronales para capturar los patrones de corto plazo.

5.2. Comprensión de los Datos

En esta etapa se presentan las fuentes de información que se tuvieron en cuenta para poder llevar a cabo la ejecución del proyecto empresarial a través del desarrollo de la base de datos. Adicional se describe de manera breve las variables identificadas por parte del criterio experto del personal de COSENIT y asociadas a modelos analizados en la literatura de la sección 5.1.

5.2.1. Identificación y Definición de Variables

El equipo de COSENIT debido a su amplia experiencia en el mercado, identificó a priori las principales variables que pueden llegar a influir en el precio de bolsa de energía eléctrica en Colombia. Esta identificación se hizo basada en principalmente en criterios de

experto y criterios económicos y operativos en la composición de la oferta y la demanda de energía a nivel nacional que son las principales variables que determinan el precio de bolsa.

En la siguiente Tabla se muestra el conjunto de las 25 variables (objetivo, exógenas) identificadas:

Tabla 2. Variables Identificadas

Variable	Nombre	Unidad	Descripción
Nivel Embalse Agregado	Embalse Total	%	El nivel de embalse se refiere a la altura del agua almacenada en un embalse o una represa. Es una medida que indica la cantidad de agua disponible en el embalse en un momento dado. El nivel de embalse es un factor importante en la gestión del agua y la generación de energía hidroeléctrica. ya que afecta la cantidad de agua disponible para mover las turbinas y generar electricidad. Un embalse con un nivel bajo puede limitar la capacidad de generación de energía, mientras que un embalse con un nivel alto puede permitir una mayor producción de energía hidroeléctrica.
Aportes Hídricos	Aportes Total	%	Los aportes hídricos de energía en Colombia se refieren a la cantidad de agua disponible en los cuerpos de agua, como ríos y embalses, que se utiliza para la generación de energía hidroeléctrica. Es una medida que indica la disponibilidad de recursos hídricos para la producción de electricidad a partir de plantas hidroeléctricas. Los aportes hídricos de energía varían según las condiciones climáticas y los patrones de precipitación en diferentes regiones del país. La temporada de lluvias tiene un impacto directo en los aportes hídricos, ya que determina la cantidad de agua disponible para la generación de energía hidroeléctrica. En períodos de lluvias intensas, los aportes hídricos suelen ser mayores, lo que aumenta la disponibilidad de agua para la generación de energía.
Generación Hidráulica	Gen Hidro	GWh-día	La generación hidráulica en Colombia se refiere a la producción de energía eléctrica a partir del aprovechamiento de recursos hídricos, como ríos y embalses, mediante la construcción y operación de centrales hidroeléctricas. Es un método de generación de energía renovable y sostenible que aprovecha el flujo del agua para hacer girar turbinas y generar electricidad.

Variable	Nombre	Unidad	Descripción
Generación Gas Natural	Gen Gas	GWh-día	La generación de gas natural en Colombia se refiere a la producción de energía eléctrica a partir de la combustión de gas natural se lleva a cabo a través de centrales térmicas de ciclo combinado, donde el gas natural es quemado en una turbina de gas para producir energía mecánica. Esta energía mecánica se utiliza para hacer girar un generador eléctrico y así generar electricidad. Además, el calor residual de la combustión del gas natural se utiliza para generar vapor de agua, el cual impulsa una turbina de vapor adicional para producir energía adicional.
Generación Combustibles Líquidos	Gen Líquidos	GWh-día	La generación de combustibles líquidos en Colombia se refiere a la producción de energía eléctrica a partir de la combustión de combustibles líquidos, como el petróleo y sus derivados. se utilizan centrales térmicas de ciclo simple o combinado, donde el combustible líquido es quemado en una caldera o una turbina de combustión para producir calor. Este calor se utiliza para generar vapor de agua, el cual impulsa una turbina de vapor
Generación Carbón	Gen Carbón	GWh-día	La generación a carbón en Colombia se refiere a la producción de energía eléctrica mediante la combustión del carbón mineral. Colombia es uno de los principales productores de carbón en el mundo. se utilizan centrales térmicas de ciclo simple o combinado, donde el carbón es quemado en una caldera para producir calor. Este calor se utiliza para generar vapor de agua, el cual impulsa una turbina de vapor y, a su vez, genera electricidad a través de un generador eléctrico.
Generación Eólica	Gen Eólica	GWh-día	La generación eólica en Colombia se refiere a la producción de energía eléctrica a partir del aprovechamiento del viento. Se basa en la utilización de aerogeneradores, también conocidos como turbinas eólicas, que convierten la energía cinética del viento en energía mecánica y luego en energía eléctrica.
Generación Solar	Gen Solar	GWh-día	La generación solar en Colombia se refiere a la producción de energía eléctrica a partir de la radiación solar en el territorio colombiano. Colombia cuenta con un alto potencial para la generación solar debido a su ubicación geográfica cercana al ecuador y a su clima tropical. En Colombia, la generación solar se realiza principalmente a través de la tecnología fotovoltaica. Los paneles solares fotovoltaicos capturan la radiación solar y la convierten directamente en electricidad. Estos paneles están compuestos por células

Variable	Nombre	Unidad	Descripción
			fotovoltaicas que utilizan el efecto fotovoltaico para generar una corriente eléctrica cuando los fotones de la luz solar los golpean.
Generación Biomasa	Gen Biomasa	GWh-día	La generación de biomasa en Colombia se refiere a la producción de energía eléctrica a partir de materiales orgánicos renovables, como residuos agrícolas, forestales, industriales o de origen animal. Estos materiales orgánicos, conocidos como biomasa, son utilizados como combustible para generar calor o electricidad.
Vertimientos	Vertimientos	GWh-día	Los vertimientos de embalses en Colombia se refieren a la liberación controlada de agua almacenada en embalses o represas hacia ríos, afluentes o cuerpos de agua cercanos. Estos vertimientos se llevan a cabo para mantener niveles adecuados de agua en los embalses, regular el caudal de los ríos aguas abajo y satisfacer diferentes usos, como el abastecimiento de agua potable, la generación de energía hidroeléctrica, el riego agrícola, entre otros.
Demanda Eléctrica Nacional	Demanda	GWh-día	La demanda eléctrica nacional en Colombia se refiere a la cantidad total de energía eléctrica requerida por los consumidores y sectores de la economía en todo el país en un determinado período de tiempo. Representa la cantidad de electricidad que se necesita para satisfacer las necesidades de los usuarios y abastecer los diferentes sectores, como residencial, comercial, industrial y público.
Precio de Bolsa Nacional	Precio Bolsa	\$COP/kWh	Se refiere al valor monetario al que se transa la energía eléctrica en el mercado mayorista del país. Es el precio al que los generadores venden la energía y los comercializadores la adquieren para suministrar a los consumidores finales. A nivel operativo es el precio de oferta de la última planta para atender la demanda nacional. Esta es la variable objetivo del proyecto
Precio de Oferta Plantas Hidráulicas	Precio Oferta Hidro	\$COP/kWh	El precio de oferta de las plantas hidráulicas en Colombia se refiere al valor al que los generadores de energía hidroeléctrica están dispuestos a vender la electricidad producida por sus plantas en el mercado mayorista de energía. Este precio de oferta se establece en función de varios factores, como los costos de operación y mantenimiento de la planta, la disponibilidad de agua para generar energía hidroeléctrica, las condiciones del mercado eléctrico y las expectativas de rentabilidad. Es importante destacar que el precio de oferta de las plantas hidráulicas puede variar en función de las condiciones

Variable	Nombre	Unidad	Descripción
			hidrológicas, es decir, la disponibilidad de agua en los embalses. En épocas de lluvia y alto caudal, es posible que las plantas hidroeléctricas ofrezcan su energía a precios más competitivos. Por el contrario, en períodos de escasez de agua, los precios de oferta de las plantas hidráulicas pueden ser más altos debido a la menor disponibilidad de generación.
Precio de Oferta Plantas a Carbón	Precio Oferta Carbón	\$COP/kWh	El precio de oferta de una planta a carbón en Colombia se refiere al valor al que los generadores de energía eléctrica a partir de carbón están dispuestos a vender la electricidad producida por sus plantas en el mercado mayorista de energía. Este precio de oferta se establece en función de diversos factores, como los costos de operación y mantenimiento de la planta, el precio del carbón, las regulaciones ambientales, las condiciones del mercado eléctrico y las expectativas de rentabilidad.
Precio Oferta Plantas a Gas	Precio Oferta Gas	\$COP/kWh	El precio de oferta de una planta a gas en Colombia se refiere al valor al que los generadores de energía eléctrica a partir de gas natural están dispuestos a vender la electricidad producida por sus plantas en el mercado mayorista de energía. Este precio de oferta se establece en función de varios factores, como los costos de operación y mantenimiento de la planta, el precio del gas natural, las regulaciones ambientales, las condiciones del mercado eléctrico y las expectativas de rentabilidad. Es importante tener en cuenta que el precio de oferta de una planta a gas puede estar sujeto a la volatilidad del precio del gas natural, que puede variar según los mercados internacionales, la oferta y demanda de gas y los acuerdos comerciales existentes.
Precio Oferta Plantas a Líquidos	Precio Oferta Líquidos	\$COP/kWh	El precio de oferta de una planta a líquidos se refiere al valor al que los generadores de energía a líquidos están dispuestos a vender la electricidad producida por sus plantas en el mercado mayorista de energía. Este precio de oferta se establece en función de varios factores, como los costos de adquisición del combustible líquido, los costos de operación y mantenimiento de la planta, las regulaciones ambientales, las condiciones del mercado eléctrico y las expectativas de rentabilidad.
Cargo por Confiabilidad	CERE	\$COP/kWh	En Colombia, el cargo por confiabilidad es un componente tarifario establecido por la Comisión de Regulación de Energía y Gas (CREG) que se aplica en el sector eléctrico. Su objetivo es garantizar la disponibilidad y confiabilidad del suministro eléctrico en el país. El cargo por confiabilidad se establece como una tarifa adicional que se

Variable	Nombre	Unidad	Descripción
			cobra a los usuarios finales de energía eléctrica en Colombia. Este cargo se destina a financiar inversiones en infraestructura, mantenimiento y otros recursos necesarios para garantizar la continuidad y calidad del suministro eléctrico.
Precio de Escasez	Precio Escasez	\$COP/kWh	En el contexto del sector eléctrico en Colombia, el "precio de escasez de energía" se refiere al valor monetario al que se negocia o establece el precio de la energía eléctrica cuando existe una insuficiencia en la oferta para satisfacer la demanda de energía en el sistema eléctrico. Cuando la capacidad de generación de energía disponible en el país no es suficiente para cubrir la demanda eléctrica, ya sea debido a condiciones climáticas adversas, problemas en la infraestructura de generación o transmisión, o un aumento inesperado en la demanda, se activan mecanismos para asegurar el suministro eléctrico.
Capacidad Instalada Hidráulica	Cap. Hidro	MW	La capacidad instalada hidroeléctrica en Colombia se refiere a la cantidad total de generación de energía eléctrica que puede ser producida por las plantas hidroeléctricas en el país. Se trata de la capacidad máxima teórica de generación que estas plantas pueden proporcionar en condiciones normales de operación.
Capacidad Instalada Carbón	Cap. Carbón	MW	Es la capacidad máxima teórica de generación que estas plantas pueden proporcionar en condiciones normales de operación.
Capacidad Instalada Gas Natural	Cap. Gas	MW	Es la capacidad máxima teórica de generación que estas plantas pueden proporcionar en condiciones normales de operación.
Capacidad Instalada Solar	Cap. Solar	MW	Es la capacidad máxima teórica de generación que estas plantas pueden proporcionar en condiciones normales de operación.
Tasa Representativa de Mercado	TRM	\$COP/USD	La TRM (Tasa Representativa del Mercado) en Colombia se refiere a la tasa de cambio oficial del peso colombiano (COP) con respecto a otras monedas extranjeras, principalmente el dólar estadounidense (USD). Es el valor de referencia utilizado por el Banco de la República de Colombia para determinar el tipo de cambio en transacciones comerciales y financieras.
ENSO	ENSO	% Variación	ENSO, por sus siglas en inglés, se refiere al fenómeno El Niño-Oscilación del Sur (El Niño-Southern Oscillation en inglés). Es un patrón climático que ocurre de manera irregular y se caracteriza por cambios en las

Variable	Nombre	Unidad	Descripción
			temperaturas de la superficie del océano Pacífico tropical y la atmósfera sobre él.
Henry Hub	Henry Hub	USD/Mbtu	Es utilizado como referencia para la fijación de precios en otros puntos de intercambio de gas natural en los Estados Unidos y también se utiliza como base para los contratos de futuros y opciones de gas natural en los mercados financieros.

Fuente: Elaboración Propia

5.2.2. Base de Datos – Recolección de Variables

Con el proceso de identificación y descripción de variables, se procede a recolectar la información de las variables seleccionadas. Con la finalidad de contar con una base de datos que tuviera información confiable, de calidad y que fuera relevante para el proyecto empresarial, se estimó recolectar información para un rango de fechas de cada una de las variables, entre el 1 de enero del 2010 hasta el 31 de diciembre de 2022, todos los datos con una periodicidad diaria.

Las fuentes oficiales en las cuales se pudo encontrar estas variables para los periodos definido fueron Sinergox. Banco de la República, *National Weather Service* y *Energy Information Administration*. Importante es que a nivel de datos todas las fuentes son públicas y no se incurren en costos para la consecución de la información, adicional que son de alta confiabilidad, principalmente el portal Sinergox (Sinergox, 2023) donde se encuentra el 90% de las variables, dado es una fuente con alta capacidad tecnológica y sujeta a regulación para el suministro de información confiable en tiempo real.

En esta etapa del proceso, se procedió a acceder a los sitios web que contienen los archivos de datos previamente definidos, y de manera manual se descargaron los archivos correspondientes a cada variable. Estos archivos fueron organizados y almacenados en una estructura de carpetas que sigue una jerarquía definida por categoría, subcategoría y variable,

con el objetivo de mantener un orden y facilitar el acceso a los datos durante el análisis posterior. Finalmente se construyó, tanto en la herramienta Excel, como un almacenamiento preliminar dada la disponibilidad de recursos en SQL, la base de datos definitiva conformada por las 26 variables continuas (numéricas) ordenadas cronológicamente, constituyendo un base con 123,448 registros.

5.3.Preparación de los Datos

El objetivo final de esta fase es obtener los datos finales sobre los que aplicarán los modelos de pronóstico del precio de bolsa. Por tanto, es esencial antes de entrenar un modelo predictivo, o incluso antes de realizar cualquier cálculo, es importante realizar una exploración descriptiva de los mismos. Este proceso permite entender mejor qué información contiene cada variable, así como detectar posibles errores.

5.3.1. Número de Observaciones y Valores Ausentes

Es esencial conocer el número de observaciones disponibles y si todas ellas están completas. Los valores ausentes son muy importantes a la hora de crear modelos, la mayoría de los algoritmos no aceptan observaciones incompletas o bien se ven muy influenciados por ellas. Para la base datos definida no se tiene ningún valor nulo (no hay necesidad de realizar técnicas de ajuste de los datos faltantes), y se cuentan con 4,748 registros por cada una de las variables.

Figura 7. Análisis de Observaciones y Nulidad

```

RangeIndex: 4748 entries, 0 to 4747
Data columns (total 26 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Fecha                                     4748 non-null   datetime64[ns]
1   PRECIO BOLSA ($/kWh)                     4748 non-null   float64
2   EMBALSE TOTAL (%)                        4748 non-null   float64
3   APORTES TOTAL (%)                        4748 non-null   float64
4   GEN HIDRO (GWh)                          4748 non-null   float64
5   Gen Carbón (GWh)                         4748 non-null   float64
6   Gen Gas (GWh)                            4748 non-null   float64
7   Gen Líquidos (GWh)                       4748 non-null   float64
8   Gen Eólica (GWh)                         4748 non-null   float64
9   Gen SOLAR (GWh)                          4748 non-null   float64
10  Gen BIOMASA (GWh)                        4748 non-null   float64
11  PRECIO OFERTA HIDRO ($/kWh)              4748 non-null   float64
12  PRECIO OFERTA CARBON ($/kWh)            4748 non-null   float64
13  PRECIO OFERTA GAS ($/kWh)               4748 non-null   float64
14  PRECIO OFERTA LIQUIDOS ($/kWh)          4748 non-null   float64
15  CERE ($/kWh)                             4748 non-null   float64
16  PRECIO ESCASEZ ($/kWh)                   4748 non-null   float64
17  Vertiminetos (GWh)                       4748 non-null   float64
18  CAP. HIDRO (MW)                          4748 non-null   float64
19  CAP. CARBON (MW)                         4748 non-null   float64
20  CAP. GAS (MW)                            4748 non-null   float64
21  CAP. SOLAR (Mw)                          4748 non-null   float64
22  DEMANDA (GWh)                            4748 non-null   float64
23  TRM ($/usd)                              4748 non-null   float64
24  ENSO                                      4748 non-null   float64
25  HENRY HUB (usd/mbtu)                     4748 non-null   float64
dtypes: datetime64[ns](1), float64(25)

```

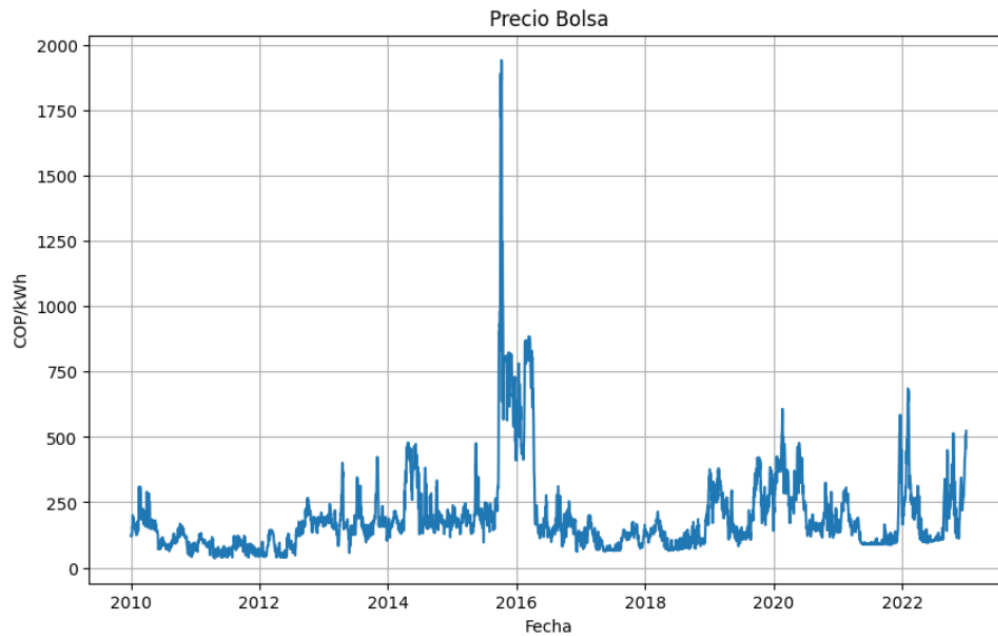
Fuente: Elaboración Propia

5.3.2. *Variable Respuesta*

Al desarrollar un modelo, es crucial examinar la distribución de la variable objetivo, ya que en última instancia es lo que se busca predecir. En el caso de la variable "precio de bolsa", se observa una distribución asimétrica con una cola positiva. Esto se debe a que algunas observaciones en el período analizado presentan valores de precio significativamente superiores a la media. Estos picos de precios superiores están asociados a eventos como el Fenómeno del Niño, que genera una alta volatilidad en los precios, como se puede apreciar

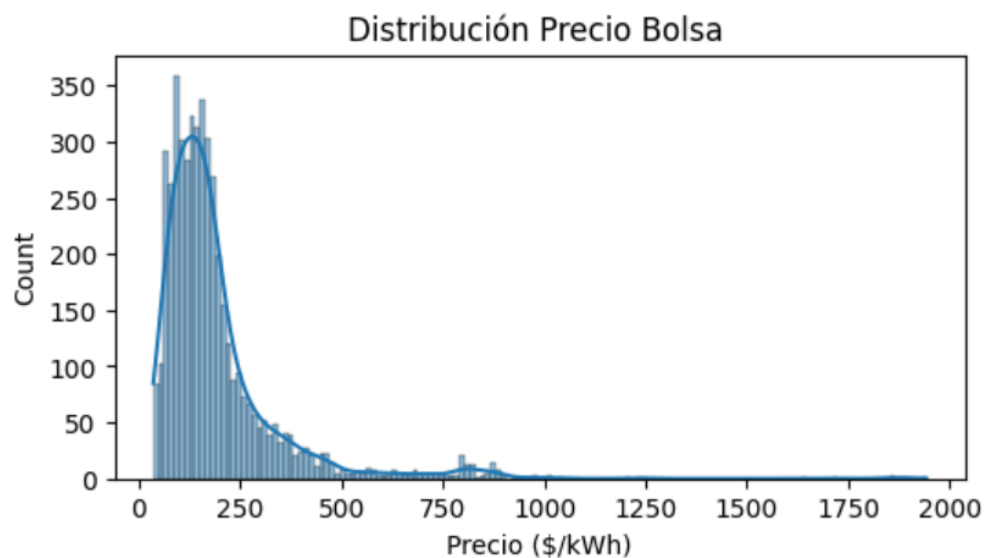
en los picos ocurridos en el año 2016. Estos eventos excepcionales contribuyen a la asimetría en la distribución de la variable precio de bolsa.

Figura 8. Serie de Precio Spot de Energía en Colombia (2010-2022)



Fuente: Elaboración Propia

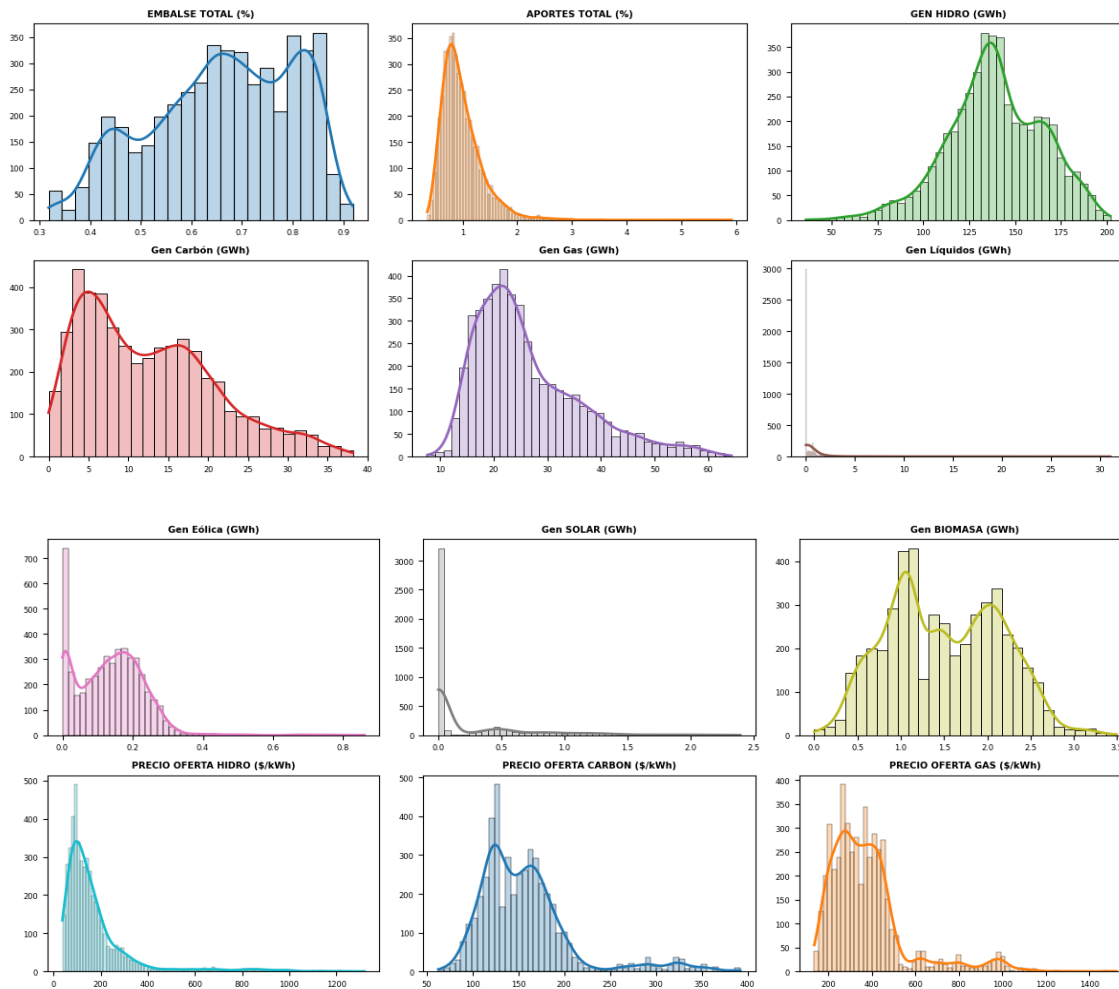
Figura 9. Distribución de Datos – Variable Objetivo

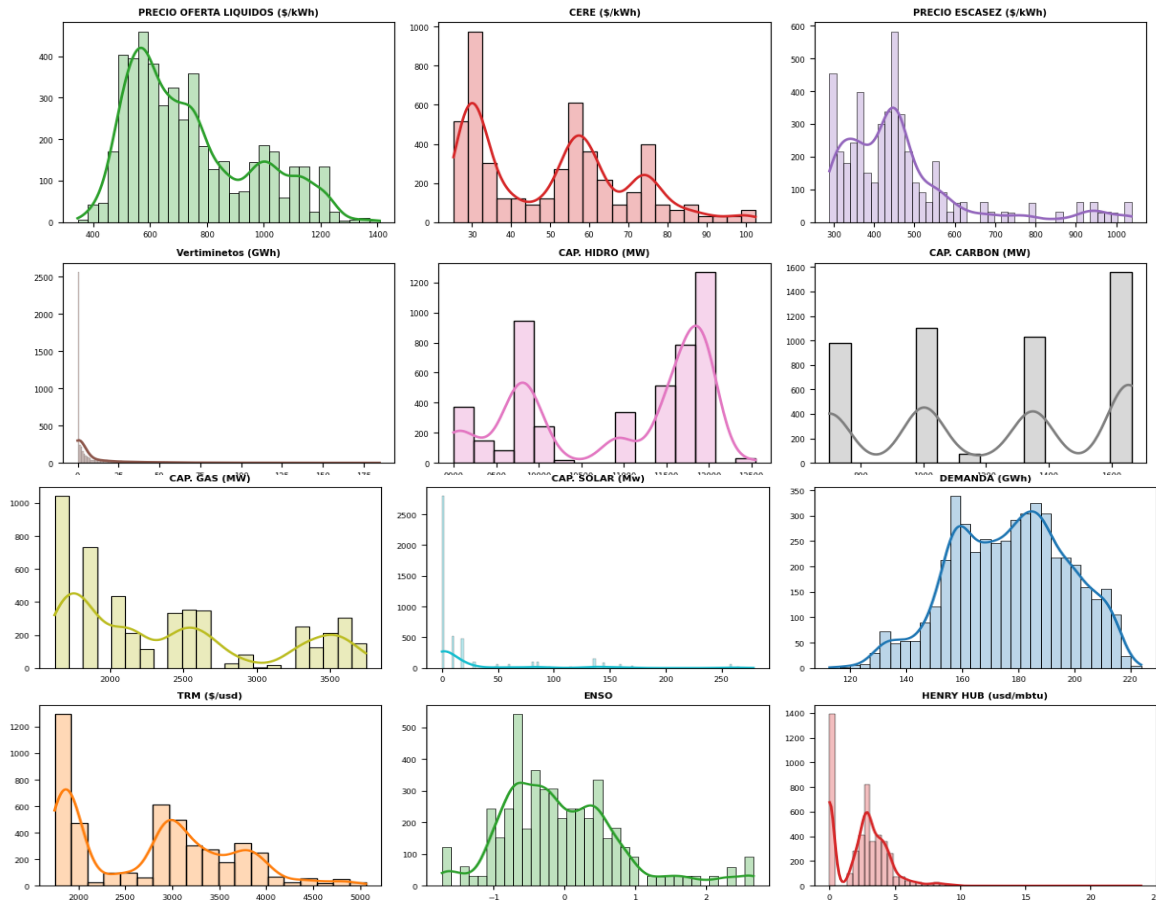


Fuente: Elaboración Propia

5.3.3. Distribución Variables Exógenas

Figura 10. Distribución de Datos – Variable Exógenas



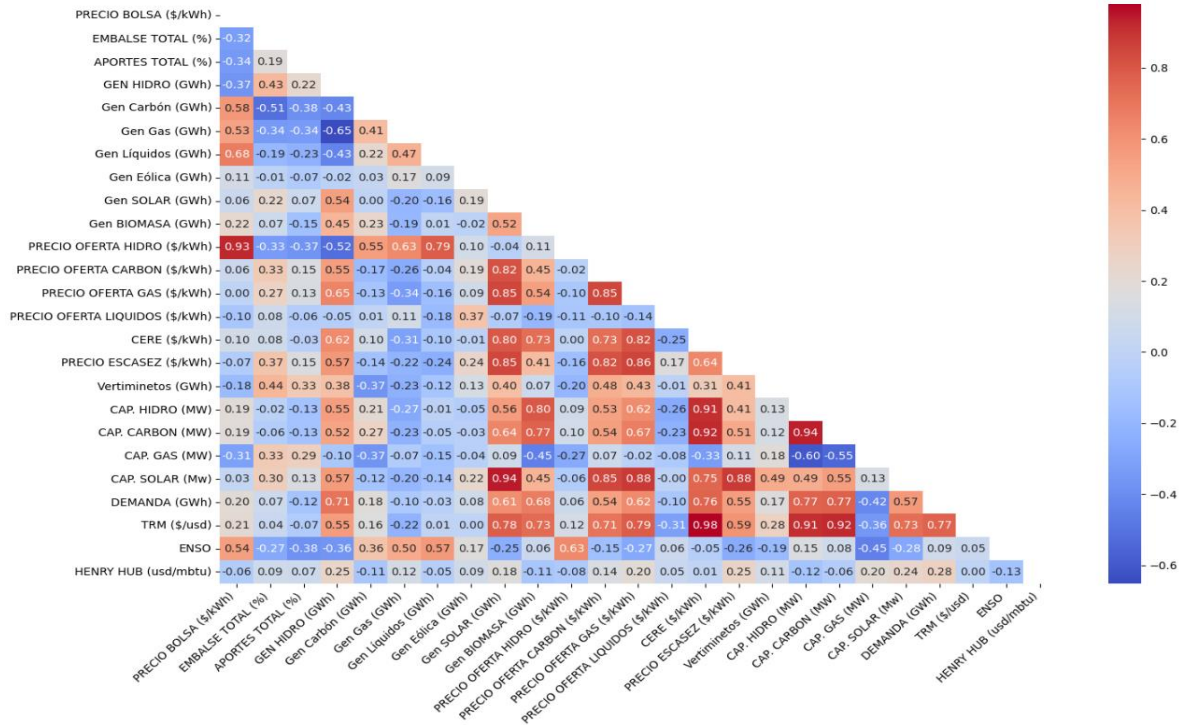


Fuente: Elaboración Propia

5.3.4. Análisis de Correlación

Se llevó a cabo un análisis de correlaciones del cual se pudo identificar las relaciones existentes entre las diferentes variables y la dependencia con la variable respuesta. Donde encontró claramente lo que a nivel teórico se había expuesto anteriormente y es la alta dependencia del precio de bolsa a las variaciones del precio de oferta de las plantas hidráulicas, seguido de las variaciones en los niveles de generación de energía de las plantas térmicas (gas, carbón y líquidos) y los cambios en la variable ENSO.

Figura 11. Análisis de Correlación Variables



Fuente: Elaboración Propia

5.3.5. Estandarización de Variables

La escala y la varianza de los predictores numéricos pueden tener un impacto significativo en el desarrollo y la ejecución de modelos de machine learning. Algunos algoritmos, como SVR, redes neuronales y regresión, son sensibles a estas diferencias. Si no se igualan de alguna manera los predictores, aquellos que se midan en una escala mayor o tengan una mayor varianza pueden dominar el modelo, incluso si no tienen una relación más fuerte con la variable respuesta. Por lo tanto, existen diferentes técnicas para mitigar este efecto, y en este proyecto se utilizó la normalización como técnica de referencia (Rodrigo, 2023).

Esta consiste en transformar los datos de forma que todos los predictores estén aproximadamente en la misma escala. Se usó para toda la base de datos, a través de la

herramienta Python (StandardScaler), que permite dividir cada predictor entre su desviación típica después de haber sido centrado a la media, de esta forma, los datos pasan a tener una distribución normal y su uso para el desarrollo de los modelos da garantía de que los predictores serán variables fiables y validadas estadísticamente.

5.4.Modelado

El objetivo de esta etapa es desarrollar el conjunto de modelos con el fin de que puedan ser validados y comparados entre sí, esto para cumplir con el objetivo general del proyecto empresarial de desarrollar un modelo de pronóstico del precio de bolsa y así abordar la necesidad de COSENIT como organización en ampliar su oferta de valor con mayores capacidades analíticas.

Las principales técnicas de modelado que tomaron como base dada la revisión de literatura en la sección 5.1 son:

5.4.1. *K-Nearest Neighbor (KNN)*

El algoritmo K-Nearest Neighbors (KNN) es un método de aprendizaje automático supervisado utilizado para clasificar o predecir nuevas instancias en función de la similitud con las instancias de entrenamiento más cercanas. KNN es un enfoque intuitivo y simple que se basa en el supuesto de que objetos similares tienden a estar cerca unos de otros en el espacio de características. El algoritmo KNN determina la clase o el valor objetivo de una nueva instancia calculando la distancia entre esa instancia y los puntos de entrenamiento más cercanos. El "K" en KNN se refiere al número de vecinos más cercanos que se toman en consideración para tomar una decisión. Dada su sencillez, suele dar peores resultados que otros algoritmos, pero es un buen referente como línea base (Rodrigo, 2023).

5.4.2. Regresión Lineal

La regresión lineal es un método estadístico utilizado para modelar la relación lineal entre una variable continua, conocida como variable dependiente, y una o más variables independientes o predictores. Consiste en ajustar una ecuación lineal que mejor se ajuste a los datos observados. Sin embargo, la regresión lineal no se limita solo a modelos sin regularización. También se pueden aplicar técnicas de regularización, como la regresión ridge y lasso durante el ajuste del modelo. Estas variantes de regresión lineal tienen como objetivo evitar el sobreajuste, reducir la varianza y mitigar el efecto de la correlación entre los predictores. Al aplicar la regularización, se obtienen modelos con un mayor poder predictivo al encontrar un equilibrio entre el ajuste a los datos y la complejidad del modelo. Estas técnicas son útiles cuando se trabaja con conjuntos de datos complejos y con múltiples variables predictoras (Rodrigo, 2023).

5.4.3. Árboles de Decisión

Los árboles de regresión en términos generales, durante el entrenamiento las observaciones se dividen en distintos nodos a medida que se construye la estructura del árbol. Cada nodo representa una bifurcación basada en los valores de los predictores. Al predecir el valor para una nueva observación, se sigue el camino correspondiente en el árbol según los valores de sus predictores hasta llegar a un nodo terminal. La predicción del árbol se obtiene como la media de los valores de la variable objetivo de las observaciones de entrenamiento que se encuentran en ese nodo terminal específico. (Rodrigo, 2023).

Algunas ventajas de este tipo de modelos son que, al tratarse de métodos no paramétricos, no es necesario que se cumpla ningún tipo de distribución específica las variables, no se ven muy influenciados por outliers (valores atípicos).

5.4.4. *Random Forest*

Un modelo Random Forest, o Bosque Aleatorio, está compuesto por un conjunto de árboles de decisión individuales. Cada árbol se entrena con una muestra ligeramente diferente de los datos de entrenamiento, generada mediante el método de bootstrapping, que implica tomar muestras aleatorias con reemplazo del conjunto de datos original. Cada árbol se construye tomando en cuenta solo un subconjunto de las características disponibles, lo que ayuda a mejorar la diversidad y reducir la correlación entre los árboles.

La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que conforman el modelo. En problemas de regresión, se toma el promedio de las predicciones individuales como resultado final.

Los métodos basados en árboles se han convertido en uno de los referentes dentro del ámbito predictivo debido a los buenos resultados que generan en problemas muy diversos. Entre sus principales ventajas están que son muy útiles en la exploración de datos, permiten identificar de forma rápida y eficiente las variables (predictores) más importantes. Tienen buena escalabilidad, pueden aplicarse a conjuntos de datos con un elevado número de observaciones (Rodrigo, 2023).

5.4.5. *Gradient Boosting Trees*

Un modelo Gradient Boosting Trees se compone de un conjunto de árboles de decisión individuales que se entrenan de forma secuencial. Cada árbol subsiguiente utiliza la información del árbol anterior para aprender de sus errores y mejorar en cada iteración. Durante la construcción de cada árbol individual, las observaciones se distribuyen en diferentes nodos, formando la estructura del árbol hasta llegar a nodos terminales. Para predecir el valor de una nueva observación, se suman las predicciones de todos los árboles

individuales que componen el modelo. De esta manera, el modelo aprovecha la capacidad de cada árbol para mejorar la precisión general del ensamblaje. (Rodrigo, 2023).

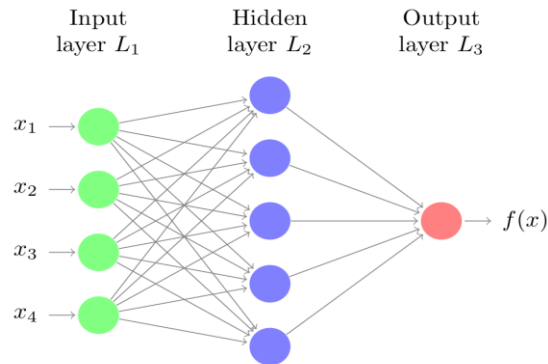
5.4.6. Máquina Soporte Vectorial (SVR)

La SVR se basa en el concepto de maximizar el margen entre la función de regresión y los datos, mientras se permite un cierto grado de error o "tolerancia" en el ajuste. A diferencia de los modelos de regresión tradicionales, la SVR no se enfoca en ajustar todos los puntos de datos, sino en encontrar una función que esté dentro de una banda definida por la tolerancia.

La optimización en la SVR se lleva a cabo mediante técnicas de programación cuadrática, donde se busca minimizar la función de pérdida y maximizar el margen bajo las restricciones establecidas por la tolerancia. Estas restricciones son determinadas por los vectores de soporte, que son los puntos de datos más cercanos a la función de regresión. El número de vectores de soporte es generalmente mucho menor que el número total de puntos de datos, lo que hace que la SVR sea computacionalmente eficiente (Rodrigo, 2023).

5.4.7. Redes Neuronales

Las redes neuronales son modelos que se componen de capas y neuronas interconectadas mediante pesos. Cada neurona realiza una operación simple y transmite información a través de conexiones ponderadas. Las capas ocultas realizan transformaciones complejas en los datos, mientras que la capa de salida produce los resultados finales. Los pesos de las conexiones se ajustan mediante algoritmos de optimización para minimizar una función de pérdida. Las redes neuronales son capaces de modelar relaciones no lineales y capturar patrones complejos en los datos, lo que las hace efectivas en problemas de aprendizaje automático.

Figura 12. Modelo Teórico Redes Neuronales

Fuente: Tomado de (Rodrigo, 2023).

En una red neuronal, la capa inicial (representada en verde) se denomina capa de entrada o input layer y recibe los datos originales, es decir, los valores de los predictores. La capa intermedia (representada en azul), conocida como capa oculta o hidden layer, recibe los valores provenientes de la capa de entrada, los cuales son ponderados por los pesos correspondientes (representados por las flechas grises). La última capa, llamada output layer, combina los valores provenientes de la capa oculta para generar la predicción final del modelo.

5.4.8. Modelos Híbridos

Los modelos híbridos en el contexto del aprendizaje automático se refieren a la combinación de diferentes técnicas o algoritmos con el objetivo de construir un modelo que mejore las predicciones. Estos modelos aprovechan las fortalezas de diferentes enfoques para mejorar el rendimiento y abordar las limitaciones de un solo algoritmo. Los ejemplos comunes de modelos híbridos son:

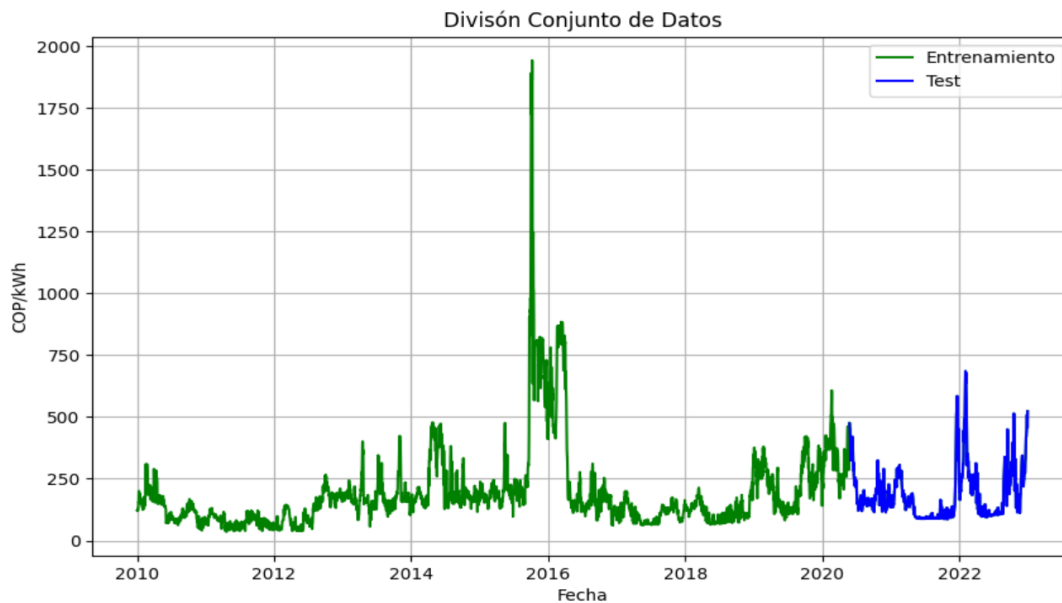
Ensamble por promedio: En este enfoque, se promedian las predicciones de varios modelos para obtener una predicción final.

Ensamble por votación: Aquí, cada modelo emite su propia predicción y la clase final se determina por mayoría de votos. Ejemplos populares incluyen el método de votación mayoritaria en Bagging o el método de votación ponderada en Boosting.

5.4.9. Datos de Entrenamiento y Test

La evaluación de la capacidad predictiva de un modelo implica verificar qué tan cercanas son sus predicciones a los valores reales de la variable objetivo. Para realizar una evaluación precisa, se requiere tener un conjunto de observaciones cuyos valores de la variable objetivo sean conocidos pero que no hayan sido utilizados en el ajuste del modelo. Con este fin, los datos disponibles se dividen en un conjunto de entrenamiento y un conjunto de prueba. El tamaño adecuado de estas divisiones depende de la cantidad de datos disponibles y del nivel de confianza necesario en la estimación del error. Al utilizar el conjunto de prueba, se puede evaluar qué tan bien generaliza el modelo y si es capaz de hacer predicciones precisas en datos no vistos durante el proceso de entrenamiento.

Para la división de la base de datos se realizó de manera determinística y no aleatoria, esto con el fin de entrenar el modelo con la mayor cantidad de posibles precios de bolsa resultantes dadas las condiciones de las variables exógenas. Por tanto, se tomó para el entrenamiento el 80% de manera cronológica desde el 1 de enero de 2010, y el 20% restante como test para probar y validar el ajuste del modelo. Adicional se procura en esta división, contar con el criterio experto donde se validó que en los datos de entrenamiento se tenían diferentes choques o eventos, como el Fenómeno del Niño en el año 2016 o la Niña para el periodo 2011, donde las variables responden a estos cambios climatológicos. En la siguiente Figura se puede observar la variable objetivo precio de bolsa en sus conjuntos de entrenamiento y test.

Figura 13. División Conjuntos de Datos (Entrenamiento – Test)

Fuente: Elaboración Propia

5.5. Evaluación

Existe una gran variedad de métricas que permiten evaluar qué tan bien se ajusta el modelo frente a los datos y su capacidad predictiva. Además, estas métricas permiten comparar los diferentes algoritmos o modelos desarrollados para determinar cuál es el mejor a nivel estadístico. A continuación, se describen las métricas que se utilizarán para evaluar los modelos mencionados en la sección 5.4.

5.5.1. MSE

El Mean Squared Error, o Error Cuadrático Medio, es una medida que calcula el promedio de los errores al cuadrado entre los valores predichos y los valores reales en un conjunto de datos. El MSE se calcula sumando los errores al cuadrado y dividiendo el resultado por el número de muestras en el conjunto de datos. Matemáticamente, se puede expresar de la siguiente manera:

$$\text{MSE} = (1/n) * \sum (y_i - \hat{y}_i)$$

Donde:

- n es el número de muestras en el conjunto de datos.
- y_i es el valor real.
- \hat{y}_i es el valor predicho por el modelo.

El MSE es una medida de la dispersión de los errores, donde valores más bajos indican un mejor ajuste del modelo a los datos. Sin embargo, el MSE no está en la misma escala que los valores originales, lo que puede dificultar su interpretación en una unidad de medida, pero es una métrica importante utilizada como una medida relativa de rendimiento para comparar modelos y determinar su nivel de certeza.

5.5.2. *MAPE*

El Error Porcentual Absoluto Medio (MAPE, por sus siglas en inglés: Mean Absolute Percentage Error) es una métrica utilizada para evaluar el error relativo entre las predicciones y los valores reales en un problema de pronóstico o regresión. Se calcula como el promedio de los errores porcentuales absolutos entre las predicciones y los valores reales. La fórmula para calcular el MAPE es la siguiente:

$$\text{MAPE} = (1/n) * \sum (| (y_true - y_pred) / y_true |) * 100$$

Donde:

- n es el número de muestras en el conjunto de datos.
- y_true son los valores reales.
- y_pred son las predicciones del modelo.

El MAPE se expresa como un porcentaje, lo que permite evaluar el error relativo en términos porcentuales. Un MAPE más bajo indica un mejor rendimiento del modelo, ya que implica un menor error relativo en las predicciones.

5.5.3. *R-Squared*

El Coeficiente de Determinación R², es una medida que indica la proporción de la varianza de los valores de respuesta que se puede predecir a partir de las variables predictoras utilizadas en el modelo. La métrica R² toma valores entre 0 y 1, donde 0 indica que el modelo no explica la variabilidad de los datos y 1 indica que el modelo explica perfectamente la variabilidad de los datos. R² se calcula mediante la fórmula:

$$R^2 = 1 - (SSR / SST)$$

Donde:

- SSR es la Suma de los Cuadrados de la Regresión, que representa la variabilidad explicada por el modelo.
- SST es la Suma Total de los Cuadrados, que representa la variabilidad total de los datos

6. Resultados Solución

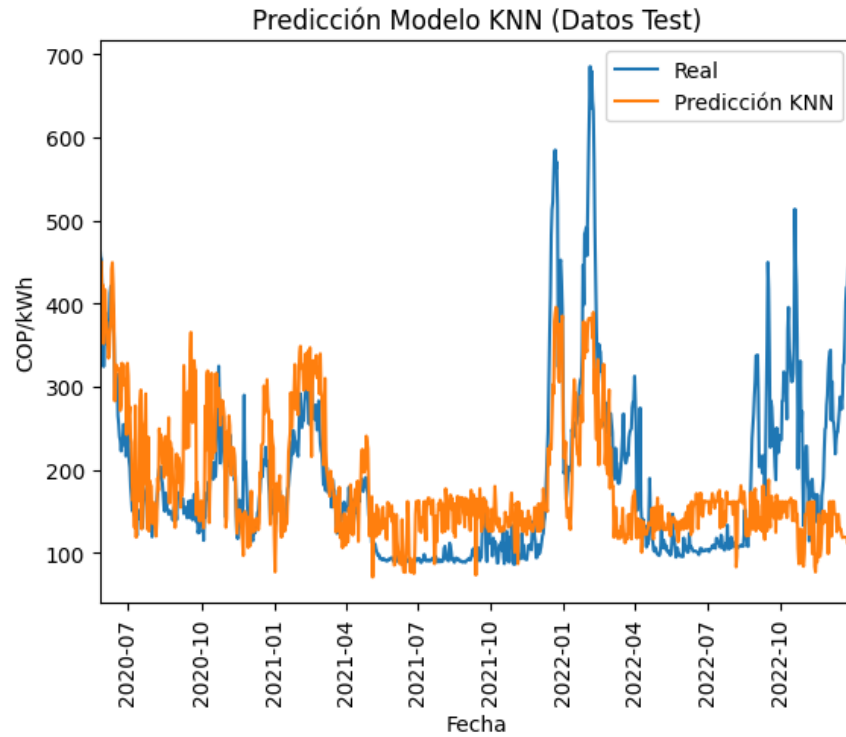
Tal como fue descrito en la metodología el objetivo de esta etapa es desarrollar el conjunto de modelos bajo código abierto en este caso Python, en el cual se elaboraron, ejecutaron, validaron y compararon entre sí los modelos, esto para cumplir con el objetivo general del proyecto empresarial de desarrollar un modelo de pronóstico del precio de bolsa y así abordar la necesidad de COSENIT como organización en ampliar su oferta de valor con mayores capacidades analíticas.

Para el desarrollo de los modelos se tomó todo el conjunto de variables de la base de datos, se procedió a estandarizar bajo la técnica de normalización, luego se dividieron los datos en entrenamiento y test (ver Sección 5.4.9) tanto para las variables exógenas como para la variable objetivo. Con esto se entrenó cada uno de los modelos seleccionados y se comprobó su nivel de certeza frente a los datos de test, donde se estimaron las métricas elegidas. Para ver el código de programación ver Anexo 1.

A continuación, se presentan de manera grafica cada uno de los resultados de los modelos de predicción, sus métricas de desempeño y finalmente una comparación entre todos para elegir el que mejor se adapta al comportamiento de los datos.

6.1.1. *K-Nearest Neighbor (KNN)*

Figura 14. Predicción Modelo KNN



Fuente: Elaboración Propia

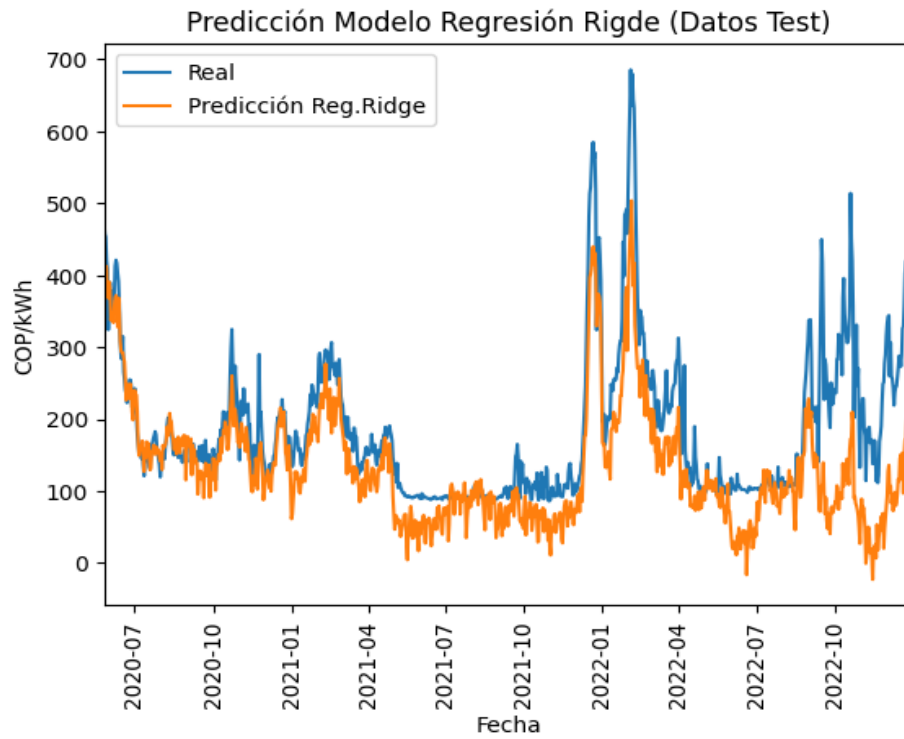
Tabla 3. Métricas de Desempeño Modelo KNN

Modelo	MSE	MAPE	R-Squared
KNN	7,045	0.33	0.36

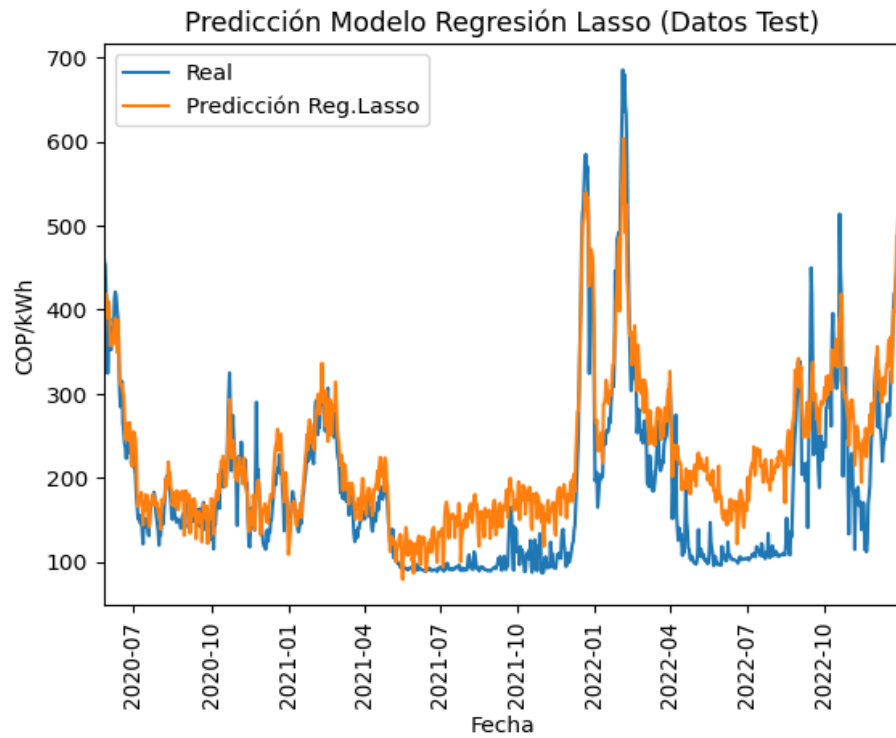
Fuente: Elaboración Propia

6.1.2. Regresión Lineal (Ridge y Lasso)

Figura 15. Predicción Modelo Regresión Ridge



Fuente: Elaboración Propia

Figura 16. Predicción Modelo Regresión Lasso

Fuente: Elaboración Propia

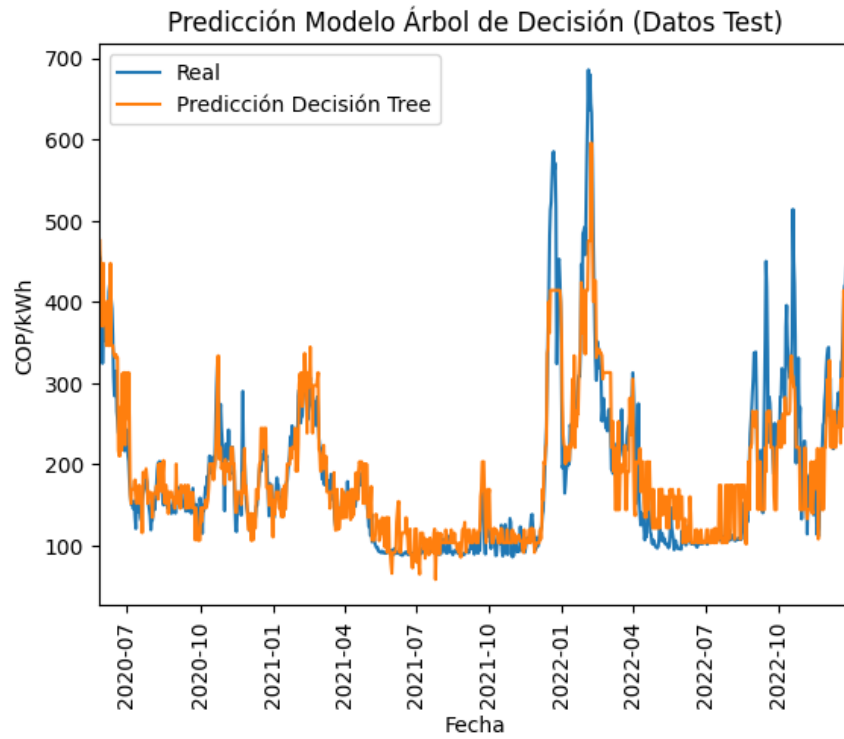
Tabla 4. Métricas de Desempeño Modelos Regresión

Modelo	MSE	MAPE	R-Squared
Regresión Ridge	6,565	0.29	0.40
Regresión Lasso	2,939	0.32	0.73

Fuente: Elaboración Propia

6.1.3. Árboles de Decisión

Figura 17. Predicción Modelo Decisión Tree



Fuente: Elaboración Propia

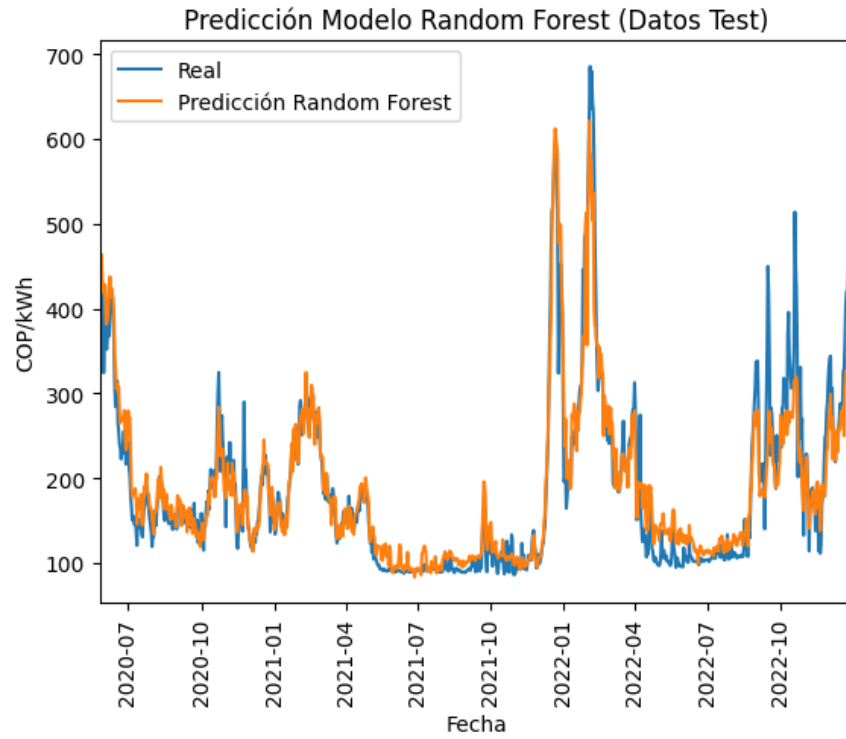
Tabla 5. Métricas de Desempeño Modelo Decision Tree

Modelo	MSE	MAPE	R-Squared
Árbol de Decisión	1,874	0.16	0.83

Fuente: Elaboración Propia

6.1.4. Random Forest

Figura 18. Predicción Modelo Random Forest



Fuente: Elaboración Propia

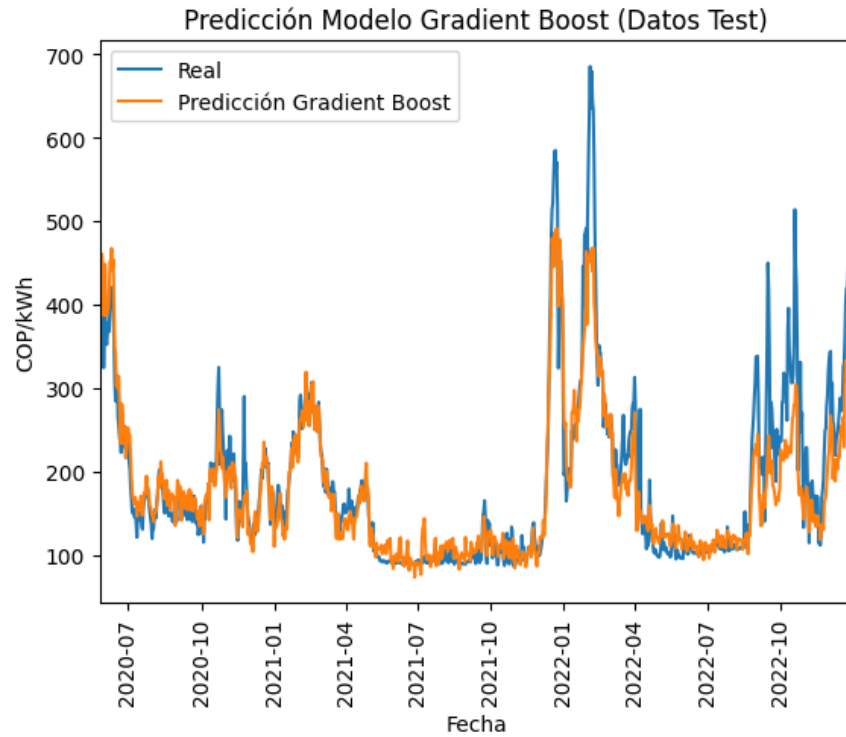
Tabla 6. Métricas de Desempeño Modelo Random Forest

Modelo	MSE	MAPE	R-Squared
Random Forest	945	0.11	0.91

Fuente: Elaboración Propia

6.1.5. Gradient Boosting Trees

Figura 19. Predicción Modelo Gradient Boost



Fuente: Elaboración Propia

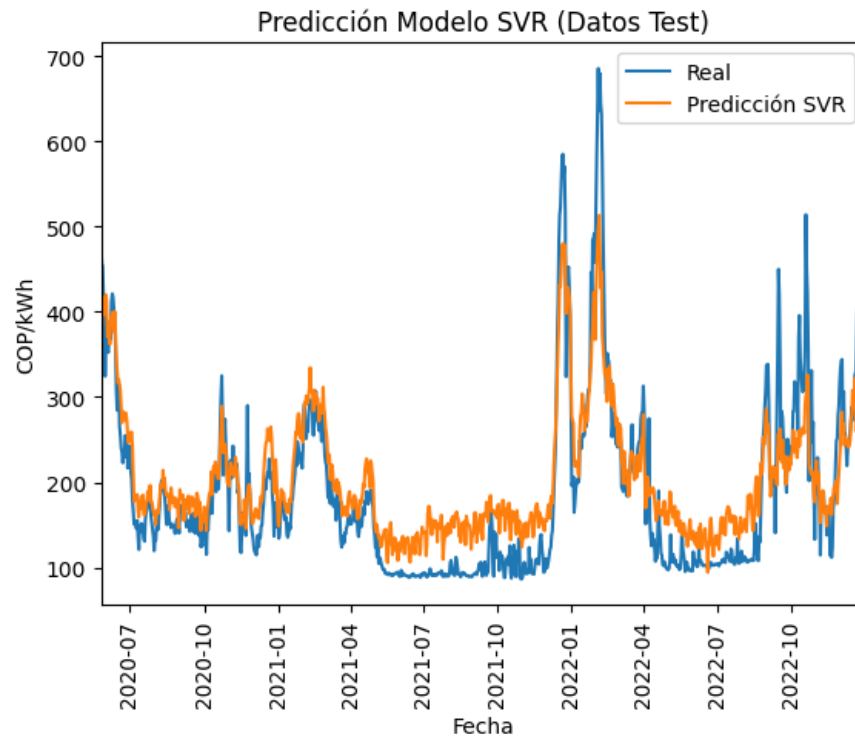
Tabla 7. Métricas de Desempeño Modelo Gradient Boost

Modelo	MSE	MAPE	R-Squared
Gradient Boost	1,555	0.12	0.86

Fuente: Elaboración Propia

6.1.6. Máquina Soporte Vectorial (SVR)

Figura 20. Predicción Modelo SVR



Fuente: Elaboración Propia

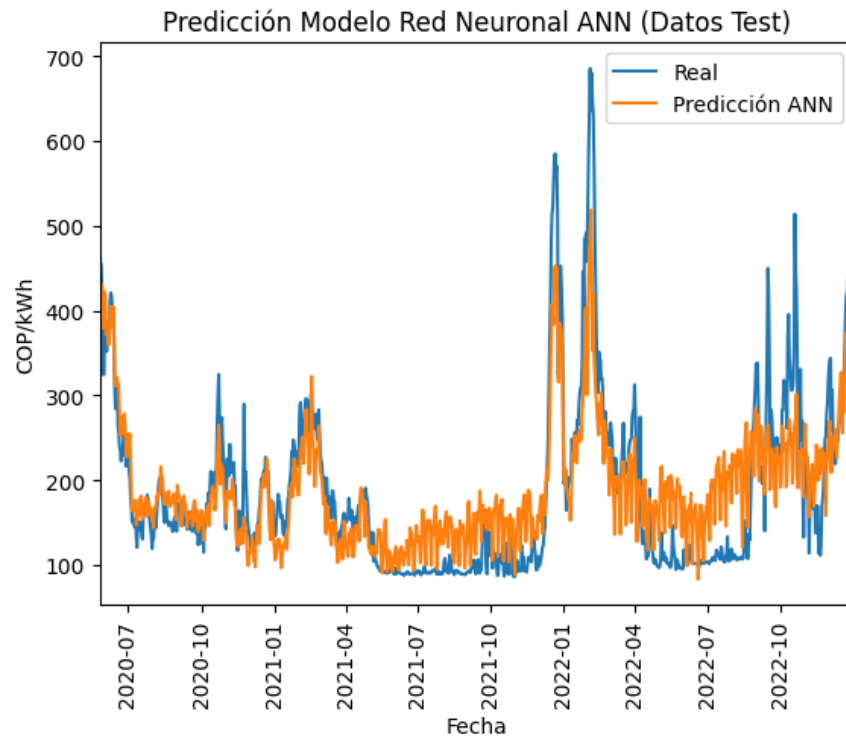
Tabla 8. Métricas de Desempeño Modelo SVR

Modelo	MSE	MAPE	R-Squared
SVR	2,184	0.26	0.80

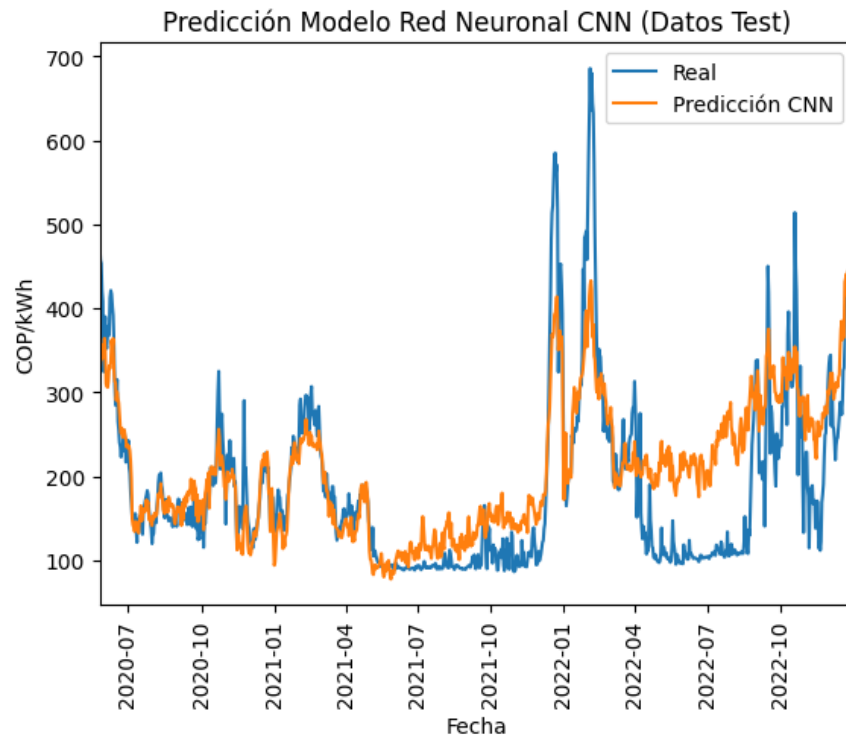
Fuente: Elaboración Propia

6.1.7. Redes Neuronales (ANN y CNN)

Figura 21. Predicción Modelo Red Neuronal ANN



Fuente: Elaboración Propia

Figura 22. Predicción Modelo Red Neuronal CNN

Fuente: Elaboración Propia

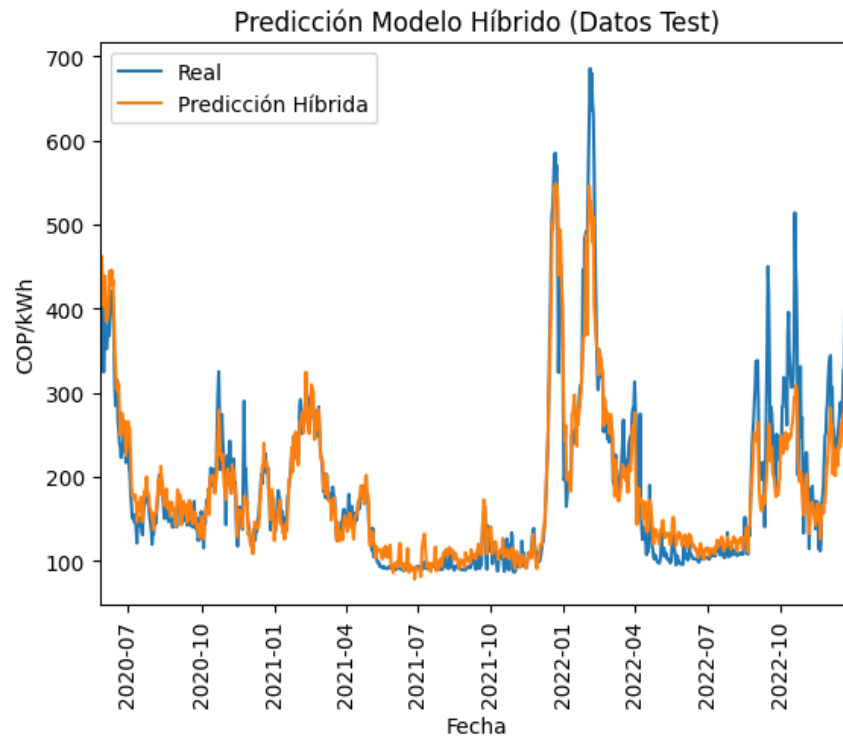
Tabla 9. Métricas de Desempeño Modelos Redes Neuronales

Modelo	MSE	MAPE	R-Squared
Red ANN	3,083	0.27	0.72
Red CNN	4.145	0.31	0.62

Fuente: Elaboración Propia

6.1.8. Modelos Híbridos (Voting Regressor)

Figura 23. Predicción Modelo Híbrido



Fuente: Elaboración Propia

Tabla 10. Métricas de Desempeño Modelo Híbrido

Modelo	MSE	MAPE	R-Squared
Voting Regressor	1,111	0.11	0.90

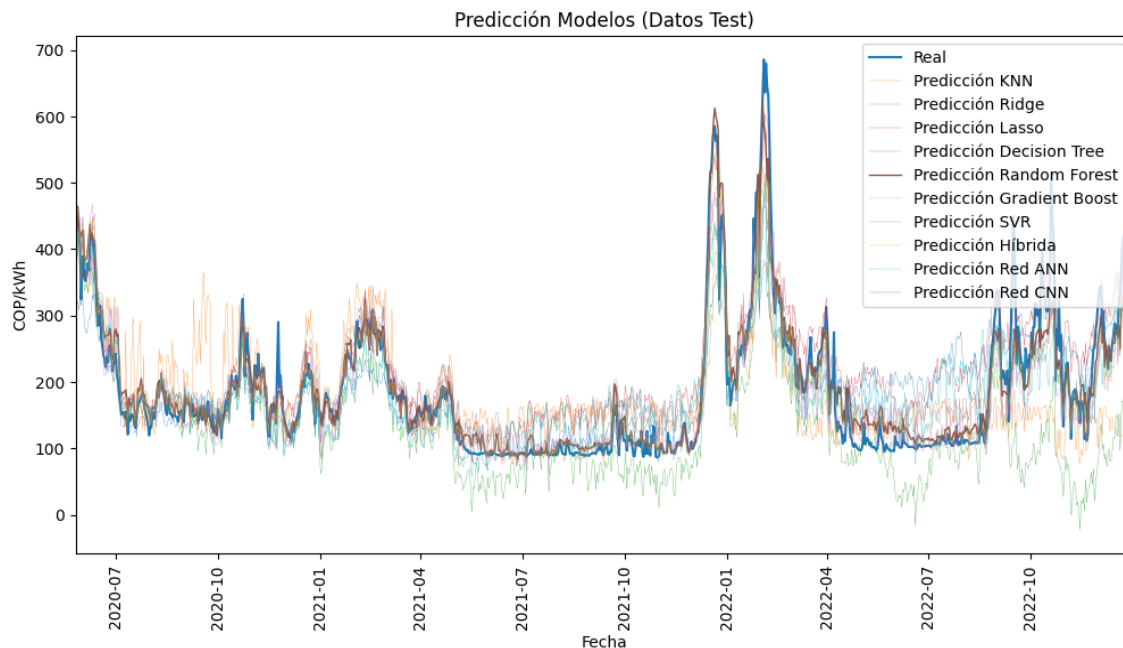
Fuente: Elaboración Propia

6.1.9. Comparación de Modelos

A continuación, se presenta la comparación grafica de los valores que cada modelo predice frente a la serie de datos de test real del precio de bolsa de energía. En esta se observa como los modelos de Random Forest, Gradient Boost, Vooting Regressor, siguen en gran medida la tendencia de la serie de test, aún en los momentos donde se presentan picos por variaciones en los precios de bolsa diarios, lo cual es un indicador de que dichos modelos a través de las técnicas de aprendizaje y los parámetros iniciales recogen el gran desafío que

es capturar la volatilidad del comportamiento del precio spot. Adicional, se establece una tabla resumen de las métricas para su comparación y una gráfica específicamente para la métrica MSE, la cual se toma como eje referencia para la elección del modelo de referencia para COSENT, esto debido a que es una métrica de uso común y simple para el desarrollo y comparación de modelos de machine learning, pero no excluye las otras métricas que también priorizan al modelo Random Forest como el óptimo.

Figura 24. Comparación Predicciones Modelos



Fuente: Elaboración Propia

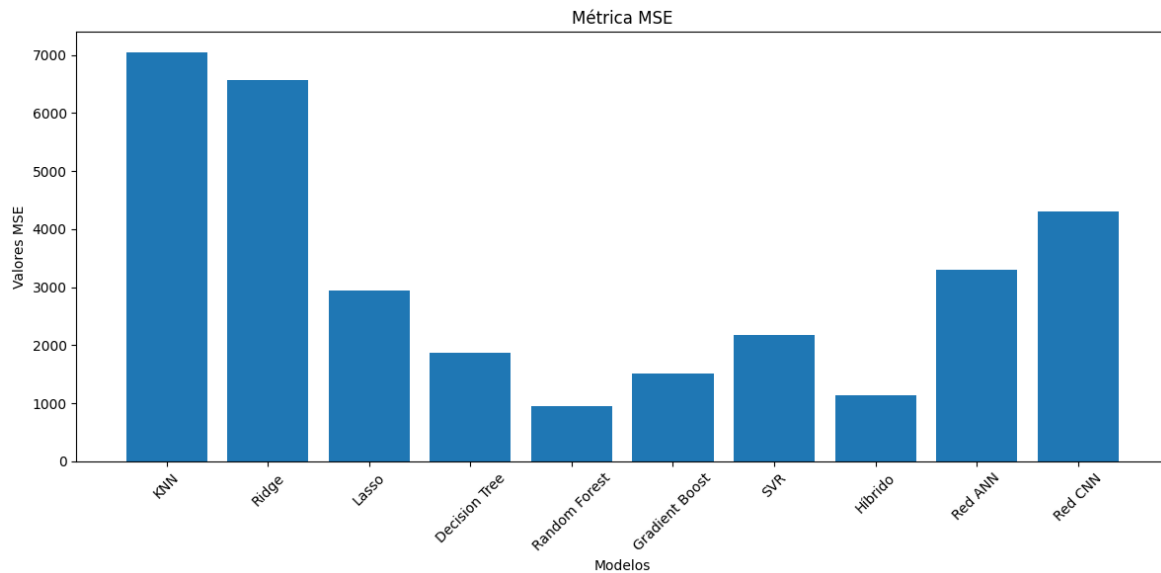
Tabla 11. Resumen Métricas de Desempeño por Modelo

Modelo	MSE	MAPE	R-Squared
KNN	7,045	0.33	0.36
Regresión Ridge	6,565	0.29	0.40
Regresión Lasso	2,939	0.32	0.73
Árbol de Decisión	1,874	0.16	0.83
Random Forest	945	0.11	0.91
Gradient Boost	1,555	0.12	0.86
SVR	2,184	0.26	0.80

Red ANN	3,083	0.27	0.72
Red CNN	4.145	0.31	0.62
Voting Regressor	1,111	0.11	0.90

Fuente: Elaboración Propia

Figura 25. Comparación Modelos por Métrica Desempeño MSE



Fuente: Elaboración Propia

Después de realizar la validación de todos los modelos entrenados, se determinó que el mejor modelo para predecir el precio de bolsa de energía en Colombia para la línea de negocio de la organización COSENIT es el modelo de Bosque Aleatorio (Random Forest). Este modelo se adapta de manera favorable a la variabilidad de los datos y logra capturar la tendencia real de la serie de manera precisa. Además, es capaz de anticipar correctamente los picos que indican la volatilidad del mercado y muestra un nivel de certeza estadística confiable en comparación con los otros modelos evaluados.

Adicionalmente, se llevó a cabo un análisis para determinar qué variables tienen una mayor influencia en el modelo seleccionado. Esto se realizó tomando la contribución relativa de cada característica en la predicción del resultado objetivo. Estos valores son

proporcionados por el modelo y se calculan en función de la reducción del error o la ganancia de información obtenida al utilizar cada característica.

Los valores de importancia de las características generalmente oscilan entre 0 y 1. Cuanto más cercano a 1 sea el valor de importancia de una característica, se considera que tiene una mayor influencia en la predicción del resultado. Por tanto, se encontró que la variable "Precio Oferta Hidro" es la más relevante, con una contribución de 0,91, lo cual respalda la relación mencionada con anterioridad tanto teóricamente como en el contexto del mercado. Este hallazgo indica que la dependencia hidráulica desempeña un papel significativo en la predicción del modelo, ya que los cambios en esta variable tienen un impacto importante en los resultados.

7. Plan y Recomendaciones de Implementación

El plan de despliegue o implementación de los modelos desarrollados y en específico el de mejor desempeño debe procurar estar alineado con las capacidades organizacionales dentro de la compañía. Por tanto, se plantean unos lineamientos base que implicarían tareas y validaciones en COSENIT, con el fin de incrementar sus capacidades analíticas para generar una oferta de mayor valor a sus clientes actuales y potenciales.

- I. Evaluación de capacidades operativas: Es fundamental evaluar las capacidades operativas de COSENIT para asegurarse de que cuente con los recursos y la logística necesaria para implementar los modelos desarrollados. Por tanto, se debe considerar aspectos como la disponibilidad de datos para entrenamiento de los modelos y para la predicción, el acceso a sistemas y herramientas, y la capacidad de gestionar el flujo de trabajo necesario para aplicar los modelos en la operación diaria.

De esta manera dada la actualidad de la organización es necesario, implementar un sistema de base de datos en la nube (Azure, Google Cloud, Oracle) para almacenar la información necesaria. Adicional una licencia de tener Python empresarial y tener soporte del sistema para ejecutar y/o modificar los modelos. Y algo relevante es con el equipo construir una API de descarga de las fuentes de información para disminuir el trabajo operativo y el grado de error en el almacenamiento de datos.

- II. Evaluación de capacidades económicas: Realizar un análisis financiero para determinar la viabilidad económica de la implementación de los modelos. Evaluar los costos asociados, como la adquisición de tecnología (licencias SQL, servicio en la nube, licencia Python/Anaconda, equipos de cómputo con una capacidad eficiente de procesamiento), la contratación de personal especializado y los posibles gastos

operativos adicionales. Esto con el fin de asegurar de que la inversión sea sostenible y genere un retorno adecuado para COSENIT.

- III. Evaluación de capacidades de conocimiento: Evaluar el nivel de conocimiento existente dentro de la compañía en relación con los modelos desarrollados. Identificar las brechas de conocimiento y las habilidades necesarias para implementar y utilizar eficazmente los modelos. Considerar el costo de tiempo y económico de la capacitación interna o externa para cerrar estas brechas.
- IV. Planificación de recursos: Desarrollar un plan detallado para asignar los recursos necesarios a la implementación de los modelos. Esto incluye recursos humanos, financieros y tecnológicos. Definir roles y responsabilidades claras para el equipo encargado de la implementación y asignar los recursos de manera eficiente.
- V. Implementación gradual: Considerar la implementación gradual de los modelos, comenzando con un piloto a nivel de predicciones para análisis interno y para casos de consultoría de corto alcance, antes de expandir su aplicación a toda la línea de negocio de la organización. Esto permitirá obtener retroalimentación temprana, identificar y abordar posibles problemas y ajustar el enfoque según sea necesario.
- VI. Monitoreo y evaluación continua: Establecer mecanismos para monitorear y evaluar el desempeño de los modelos implementados. Esto incluye la recolección de datos relevantes, el análisis de resultados y la iteración continua para mejorar los modelos y optimizar su rendimiento.
- VII. Comunicación y capacitación: Comunica de manera efectiva los avances, beneficios y resultados obtenidos a lo largo del proceso de implementación a los clientes, socios y empleados de COSENIT.

8. Conclusiones y Recomendaciones

Este proyecto empresarial tuvo como objetivo principal desarrollar un modelo analítico para la predicción del precio spot en Colombia, esto con el fin de ser implementado en la organización COSENIT para ampliar sus capacidades analíticas, entrar en la era de la transformación digital y generar valor a sus clientes en el mercado eléctrico colombiano.

Este objetivo se logró cumplir a través de un proceso elaborado a través de la metodología CRISP-DM, en la cual se lograron identificar los principales modelos y técnicas desarrolladas a nivel nacional e internacional, se seleccionaron las variables exógenas para explicar el comportamiento y la variabilidad del precio de energía y finalmente se desarrollaron diez modelos de machine learning. Al analizar sus métricas de desempeño el modelo que mejor se comportó para la base de datos analizada es el modelo Random Forest (MSE 945) el cual es una técnica basada en arboles de decisión con un alto potencial y robustez para predicción de variables.

Este modelo seleccionado permite a la organización COSENIT abordar su necesidad actual de contar con una herramienta potente que genere un mayor grado de certeza para satisfacer la demanda de los clientes en términos de predicciones que respalden de manera sólida la toma de decisiones empresariales de compra y venta de energía eléctrica en el mercado spot. El desarrollo del modelo está fundamentado y estructurado de manera que sea replicable y adaptable en el tiempo. Por tanto, se recomienda implementarlo dentro de la compañía, pero es importante tener en cuenta aspectos organizacionales relevantes para asegurar su despliegue, mantenimiento y generación de valor de manera efectiva.

El alcance de este proyecto empresarial se limita hasta el desarrollo y selección del modelo, pero se recomienda un proceso adaptativo donde se implemente aspectos que

permitan no solo abordar la necesidad expuesta, si no que se dé una transformación digital y analítica integral al interior de la compañía, esto mediante la adquisición de servicios en la nube para almacenar, generar y analizar información necesaria para la operatividad del negocio que permita tomar decisiones informadas.

Adicionalmente, se propone para futuros trabajos vislumbrar la posibilidad de mejorar los modelos expuestos con técnicas de optimización, dado que su desarrollo estuvo enfocado en técnicas determinísticas, esto podría conllevar que con hiperparámetros (ajustes de fondo a los parámetros de los modelos analizados para mejorar su desempeño), mayor cantidad de variables exógenas y análisis constante de la evolución del mercado tener modelos certeros que den un alto valor a los actores del mercado eléctrico colombiano. Así mismo validar qué modelos se comportan mejor para predicciones en diferentes periodicidades y puedan aplicar a estrategias diferentes como compra/venta en tiempo real de energía, inversiones de largo plazo o planeación de mediano plazo, esto permitirá posicionar a COSENIT como la consultoría referente a nivel nacional, líder en el uso de análisis de datos para impulsar el éxito empresarial.

9. Referencias Bibliográficas

- Agudelo, Adriana P, López-Lezama, Jesús M, & Velilla, Esteban. (2015). Predicción del Precio de la Electricidad en la Bolsa mediante un Modelo Neuronal No-Lineal Autorregresivo con Entradas Exógenas. *Información tecnológica*, 26(6), 99-108. <https://dx.doi.org/10.4067/S0718-07642015000600012>
- Barrientos Marín, J., & Toro Martínez, M. (2016). La hidrología como predictor del comportamiento del precio de energía en bolsa. *Perfil de Coyuntura Económica*, 25. <https://doi.org/10.17533/udea.pece.n25a07>
- Barrientos, J., & Martínez, M. T. (2017). Análisis de los fundamentales del precio de la energía eléctrica: Evidencia empírica para Colombia. *Revista de Economía del Caribe*, 19, 33-59.
- CREG. (2023). Comisión de Regulación de Energía y Gas. <https://www.creg.gov.co/>
- Dejamkhooy, A., & Ahmadpour, A. (2022). Prediction and Evaluation of Electricity Price in Restructured Power Systems Using Gaussian Process Time Series Modeling. *Smart Cities*, 5(3), 889-923. <https://doi.org/10.3390/smartcities5030045>
- Díez, I. C. (2015). Proyección de precios de energía eléctrica de mediano plazo en el mercado colombiano mediante la aplicación del índice de Lerner. [Trabajo de grado]. Universidad EAFIT, Medellín, Colombia
- Duque, J. P. B., García, J. J., & Velásquez, H. (2016). Efectos del cargo por confiabilidad sobre el precio spot de la energía eléctrica en Colombia. *Cuadernos de Economía*, XXXV (68), 491-519.

- Galindo, A. F. (2017). Modelamiento de los precios de la energía en bolsa en Colombia incorporando el efecto del ENSO. [Trabajo de grado]. Universidad de Los Andes, Bogotá D.C., Colombia.
- Galón, S., & Barrientos, J. (2021). Proyección del Precio Spot de la Electricidad Colombiana bajo un Enfoque Funcional. *Revista internacional de economía y política energética*, 11 (2), 67–74. Obtenido de <https://www.econjournals.com/index.php/ijeeep/article/view/10607>
- Gómez-Cano, L., Catalina-Cuellar, S., & Méndez-Vargas, R. (2020). Modelo de pronóstico para estimar el comportamiento del precio en bolsa de la energía en Colombia. *Pensamiento y Acción*. Recuperado de: <https://www.semanticscholar.org/paper/Modelo-de-pron%C3%B3stico-para-estimar-el-comportamiento-G%C3%B3mez-Cano-Catalina-Cuellar/a1a0b9166d83cb7e01c214d50ce97c9662bbe1dc>
- Grothe, O., Kächele, F., & Krüger, F. (2023). From point forecasts to multivariate probabilistic forecasts: The Schaaque shuffle for day-ahead electricity price forecasting. *Energy Economics*, 120, 106602. <https://doi.org/10.1016/j.eneco.2023.106602>
- Gunduz, S., Ugurlu, U., & Oksuz, I. (2023). Transfer learning for electricity price forecasting. *Sustainable Energy, Grids and Networks*, 34, 100996. <https://doi.org/10.1016/j.segan.2023.100996>
- Hurtado Moreno, L., Quintero Montoya, O. L., & García Rendón, J. J. (2016). Estimación del precio de oferta de la energía eléctrica en Colombia mediante inteligencia

- artificial. *Revista De Métodos Cuantitativos Para La Economía Y La Empresa*, 18, Páginas 54 a 87. <https://doi.org/10.46661/revmetodoscuanteconomia.2205>
- IBM. (2023). Introducción al CRISP-DM. Recuperado de: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=guide-introduction-crisp-dm>
- Jiang, P., Nie, Y., Wang, J., & Huang, X. (2023). Multivariable short-term electricity Price forecasting using artificial intelligence and multi-input multi-output scheme. *Energy Economics*, 117, 106471. <https://doi.org/10.1016/j.eneco.2022.106471>
- Kuo, P., & Huang, C. (2018). An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. *Sustainability*, 10(4), 1280. <https://doi.org/10.3390/su10041280>
- Lehna, M., Scheller, F., & Herwartz, H. (2022). Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. *Energy Economics*, 106, 105742. <https://doi.org/10.1016/j.eneco.2021.105742>
- López, A. (2015). Comportamiento del mercado eléctrico colombiano por los intercambios de electricidad regional con la comunidad andina de naciones Can y caso de estudio con Ecuador. Recuperado de: <http://hdl.handle.net/20.500.11912/2369>.
- Lucas, A., Pegios, K., Kotsakis, E., & Clarke, D. (2020). Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression. *Energies*, 13(20), 5420. <https://doi.org/10.3390/en13205420>
- Marcjasz, G., Lago, J., & Weron, R. (2020). Neural networks in day-ahead electricity price forecasting: Single vs. multiple outputs. ArXiv, abs/2008.08006.

- Marin, J. B., Orozco, E. T., & Velilla, E. (2018). Forecasting electricity price in Colombia: A comparison between Neural Network, ARMA process and Hybrid Models. *International Journal of Energy Economics and Policy*, 8(3), 97–106. Retrieved from <https://www.econjournals.com/index.php/ijeep/article/view/6181>
- Martín Mayorga, R. A., & Píneros Sánchez, L. F. (2020). Propuesta de un modelo con redes neuronales y metodología Box & Jenkins para el pronóstico del precio de bolsa de la energía en Colombia. <https://repository.libertadores.edu.co/handle/11371/2654>
- Oviedo-Gomez, A., Londono-Hernandez, S. M., & Manotas-Duque, D. F. (2021). Electricity Price Fundamentals in Hydrothermal Power Generation Markets Using Machine Learning and Quantile Regression Analysis. *International Journal of Energy Economics and Policy*, 11(5), 66–77. Retrieved from <https://www.econjournals.com/index.php/ijeep/article/view/11346>
- Rodrigo, J. (2023). Ciencia de Datos. <https://www.cienciadedatos.net/>
- Santiago, A. M., Vanstrahlengs, J. U., Otero, A. C., & Lombana, J. (2017). Pronóstico del precio de la energía en Colombia utilizando modelos arima con igarch. *Revista de economía del Rosario*, 20(1), 127-159
- Singh, P., & Kottath, R. (2022). Influencer-defaulter mutation-based optimization algorithms for predicting electricity prices. *Utilities Policy*, 79, 101444. <https://doi.org/10.1016/j.jup.2022.101444>
- SSPD. (2023). Superintendencia de Servicios Públicos Domiciliarios. <http://www.superservicios.gov.co/>
- UPME. (2023). Unidad de Planeación Minero Energética. <https://www1.upme.gov.co/>

- Uribe Gaviria, E. & Trespalacios Carrasquilla, A. (2014). Contraste de modelos estocásticos para el precio de la energía en Colombia. *Revista de la Facultad de Ciencias*, 3(1), 41–55. <https://revistas.unal.edu.co/index.php/rfc/article/view/49524>
- Villada, F., Cadavid, D. R., & Molina, J. D. (2014). Electricity price forecasting using artificial neural networks. *Revista Facultad De Ingeniería Universidad De Antioquia*, (44), 111–118. <https://doi.org/10.17533/udea.redin.18508>
- Wang, J., Liu, F., Song, Y., & Zhao, J. (2016). A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system. *Applied Soft Computing*, 48, 281-297. <https://doi.org/10.1016/j.asoc.2016.07.011>
- Wang, J., Yang, W., Du, P., & Niu, T. (2020). Outlier-robust hybrid electricity price forecasting model for electricity market management. *Journal of Cleaner Production*, 249, 119318. <https://doi.org/10.1016/j.jclepro.2019.119318>
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030-1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>
- Yang, W., Wang, J., Niu, T., & Du, P. (2020). A novel system for multi-step electricity price forecasting for electricity market management. *Applied Soft Computing*, 88, 106029. <https://doi.org/10.1016/j.asoc.2019.106029>

Anexos Técnico No.1

```

#Bibliotecas necesarias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.neighbors import KNeighborsRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.metrics import mean_absolute_percentage_error

from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso, ElasticNet

from sklearn.tree import DecisionTreeRegressor

from sklearn.ensemble import RandomForestRegressor

from sklearn.ensemble import GradientBoostingRegressor

from sklearn.svm import SVR

from sklearn.ensemble import VotingRegressor

from tensorflow import keras
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Dense, Dropout
from tensorflow.keras.callbacks import EarlyStopping

from tensorflow.keras import layers
import tensorflow as tf

#Cargar base de datos
data = pd.read_excel("/content/Prueba_V2.xlsx", header=0)

#Convertir la fecha y agregarla a la base de datos
data['Fecha'] = pd.to_datetime(data['Fecha'])
data = data.set_index('Fecha').sort_index()

# Separar los datos en características y etiquetas
X = data.drop(['PRECIO BOLSA ($/kWh)'], axis=1)
y = data['PRECIO BOLSA ($/kWh)']

#Normalizar las variables independientes
scaler = StandardScaler()
X_norm = scaler.fit_transform(X)

# Dividir los datos en conjuntos de entrenamiento y prueba
train_size = int(len(data) * 0.8)
X_train = X_norm[:train_size]

```

```

y_train = y[:train_size]
X_test = X_norm[train_size:]
y_test = y[train_size:]

# Crear el modelo KNN y ajustarlo a los datos de entrenamiento
knn_model = KNeighborsRegressor()
knn_model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_knn = knn_model.predict(X_test)

# Calcular metricas
mse_knn_test = mean_squared_error(y_test, y_pred_knn)
print('Error cuadrático medio K-NN:', mse_knn_test)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_knn))
mape_knn = mean_absolute_percentage_error(y_test, y_pred_knn)
print("MAPE:", mape_knn)

# Crear el modelo Regresión Ridge y ajustarlo a los datos de entrenamiento
ridge_model = Ridge()
ridge_model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_ridge = ridge_model.predict(X_test)

# Calcular el error cuadrático medio
mse_ridge = mean_squared_error(y_test, y_pred_ridge)
print('Error cuadrático medio Ridge:', mse_ridge)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_ridge))
mape_ridge = mean_absolute_percentage_error(y_test, y_pred_ridge)
print("MAPE:", mape_ridge)

# Crear modelo Regresión Lasso
lasso = Lasso(alpha=0.1)
lasso.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_lasso = lasso.predict(X_test)

# Calcular métricas de rendimiento
mse_lasso = mean_squared_error(y_test, y_pred_lasso)
r2_lasso = r2_score(y_test, y_pred_lasso)
print("Lasso - MSE:", mse_lasso)
print("Lasso - R2 Score:", r2_lasso)
mape_lasso = mean_absolute_percentage_error(y_test, y_pred_lasso)
print("MAPE:", mape_lasso)

# Entrenar el modelo de regresión en Decision Tree
dt_model = DecisionTreeRegressor(random_state=42)
dt_model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_dt = dt_model.predict(X_test)

```

```

# Calcular el error cuadrático medio
mse_dt = mean_squared_error(y_test, y_pred_dt)
print('Error cuadrático medio Decision Tree:', mse_dt)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_dt))
mape_dt = mean_absolute_percentage_error(y_test, y_pred_dt)
print("MAPE:", mape_dt)

# Entrenar el modelo de regresión en Random Forest
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_rf = rf_model.predict(X_test)

# Calcular el error cuadrático medio
mse_rf = mean_squared_error(y_test, y_pred_rf)
print('Error cuadrático medio Random Forest:', mse_rf)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_rf))
mape_rf = mean_absolute_percentage_error(y_test, y_pred_rf)
print("MAPE:", mape_rf)

# Crear el modelo Gradient Boosting y ajustarlo a los datos de entrenamiento
gb_model = GradientBoostingRegressor()
gb_model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_gb = gb_model.predict(X_test)

# Calcular el error cuadrático medio
mse_gb = mean_squared_error(y_test, y_pred_gb)
print('Error cuadrático medio Gradient Boosting:', mse_gb)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_gb))
mape_gb = mean_absolute_percentage_error(y_test, y_pred_gb)
print("MAPE:", mape_gb)

# Entrenar el modelo de regresión en SVR
model = SVR(kernel='linear')
model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred_svr = model.predict(X_test)

# Calcular el error cuadrático medio
mse_SVR = mean_squared_error(y_test, y_pred_svr)
print('Error cuadrático medio:', mse_SVR)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_svr))
mape_SVR = mean_absolute_percentage_error(y_test, y_pred_svr)
print("MAPE:", mape_SVR)

# Crear el modelo de ANN
model_ann = Sequential()
model_ann.add(Dense(64, input_dim=X_train.shape[1], activation='relu'))

```

```

model_ann.add(Dropout(0.2)) # Agregar dropout para regularización
model_ann.add(Dense(32, activation='relu'))
model_ann.add(Dense(1))

# Compilar el modelo de ANN
model_ann.compile(optimizer='adam', loss='mean_squared_error')

# Utilizar EarlyStopping para detener el entrenamiento si no hay mejoras
early_stopping = EarlyStopping(monitor='val_loss', patience=3, restore_best_weights=True)

# Entrenar el modelo de ANN
history = model_ann.fit(X_train, y_train, epochs=50, batch_size=32, validation_data=(X_test, y_test), callbacks=[early_stopping])

# Realizar predicciones con el modelo de ANN
y_pred_ann = model_ann.predict(X_test)

# Calcular las métricas de rendimiento
mse_ann = mean_squared_error(y_test, y_pred_ann)
r2_ann = r2_score(y_test, y_pred_ann)
print("MSE:", mse_ann)
print("R2 Score:", r2_ann)
mape_ann = mean_absolute_percentage_error(y_test, y_pred_ann)
print("MAPE:", mape_ann)

# Crear el modelo de red neuronal convolucional
model = keras.Sequential()
model.add(layers.Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(X_train.shape[1], 1)))
model.add(layers.Conv1D(filters=64, kernel_size=3, activation='relu'))
model.add(layers.Dropout(0.5))
model.add(layers.MaxPooling1D(pool_size=2))
model.add(layers.Flatten())
model.add(layers.Dense(100, activation='relu'))
model.add(layers.Dense(1))

# Compilar el modelo
model.compile(loss='mse', optimizer='adam')

# Entrenar el modelo
X_train_cnn = X_train.reshape((X_train.shape[0], X_train.shape[1], 1))
model.fit(X_train_cnn, y_train, epochs=50, batch_size=16, verbose=2)

# Predecir con el modelo en el conjunto de prueba
X_test_cnn = X_test.reshape((X_test.shape[0], X_test.shape[1], 1))
y_pred_cnn = model.predict(X_test_cnn)

# Agregar predicciones al conjunto de prueba
y_test_pred_cnn = pd.Series(y_pred_cnn.flatten(), index=y_test.index)

# Evaluar el modelo en el conjunto de prueba
mse_cnn = model.evaluate(X_test_cnn, y_test, verbose=0)
print(f'MSE: {mse_cnn:.4f}')
print(f'Coefficient of determination: %.2f' % r2_score(y_test, y_test_pred_cnn))

# Calcular MAPE

```

```

mape_cnn = mean_absolute_percentage_error(y_test, y_test_pred_cnn)
print("MAPE:", mape_cnn)

# Crear el modelo de ensamblaje de promedio
ensemble_model = VotingRegressor([('rf', rf_model), ('gb', gb_model)])
ensemble_model.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba con el modelo de ensamblaje
y_pred_voting = ensemble_model.predict(X_test)

# Calcular el error cuadrático medio y el coeficiente de determinación para el modelo de Random Forest
mse_voting = mean_squared_error(y_test, y_pred_voting)
print('Error cuadrático medio Random Forest:', mse_voting)
print('Coefficient of determination: %.2f' % r2_score(y_test, y_pred_voting))
mape_voting = mean_absolute_percentage_error(y_test, y_pred_rf)
print("MAPE:", mape_voting)

# Graficar los datos

plt.figure(figsize=(12, 6))

plt.plot(y_test.index, y_test, label='Real', linewidth=1.5)
plt.plot(y_test.index, y_pred_knn, label='Predicción KNN', linewidth=0.2)
plt.plot(y_test.index, y_pred_ridge, label='Predicción Ridge', linewidth=0.2)
plt.plot(y_test.index, y_pred_lasso, label='Predicción Lasso', linewidth=0.2)
plt.plot(y_test.index, y_pred_dt, label='Predicción Decision Tree', linewidth=0.2)
plt.plot(y_test.index, y_pred_rf, label='Predicción Random Forest', linewidth=1)
plt.plot(y_test.index, y_pred_gb, label='Predicción Gradient Boost', linewidth=0.2)
plt.plot(y_test.index, y_pred_svr, label='Predicción SVR', linewidth=0.2)
plt.plot(y_test.index, y_pred_voting, label='Predicción Híbrida', linewidth=0.2)
plt.plot(y_test.index, y_pred_ann, label='Predicción Red ANN', linewidth=0.2)
plt.plot(y_test.index, y_test_pred_cnn, label='Predicción Red CNN', linewidth=0.2)
plt.legend()

# Ajustar el espaciado entre las marcas de tiempo en el eje x
plt.xticks(rotation=90)

# Establecer los límites del eje x
plt.xlim(y_test.index[0], y_test.index[-1])

plt.xlabel('Fecha')
plt.ylabel('COP/kWh')
plt.title('Predicción Modelos (Datos Test)')

# Mostrar el gráfico
plt.show()

#Graficar Metricas
plt.figure(figsize=(12, 6))

# Datos
variables = ['KNN', 'Ridge', 'Lasso', 'Decision Tree', 'Random Forest', 'Gradient Boost', 'SVR', 'Híbrido', 'Red ANN', 'Red CNN']
valores = [mse_knn_test, mse_ridge, mse_lasso, mse_dt, mse_rf, mse_gb, mse_SVR, mse_voting, mse_ann, mse_cnn ]

# Crear gráfico de barras
plt.bar(variables, valores)

# Personalizar el gráfico
plt.xlabel('Modelos')
plt.ylabel('Valores MSE')
plt.title('Métrica MSE')

# Mostrar el gráfico
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

```