



Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

Estudiante:

Carlos Alberto Jusquini Tinoco

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesores

Álvaro Jaramillo Duque, Doctor (PhD)

Walter Mauricio Villa Acevedo, Doctor (PhD)

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

Cita	(Jusquini Tinoco, 2024)
Referencia	Jusquini Tinoco, C.A. (2024). <i>Predicción del precio en bolsa de la energía eléctrica en Colombia</i> [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	Colombia.



Especialización en Analítica y Ciencia de Datos, Cohorte VI.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de contenido

Tabla de contenido	3
Lista de tablas	5
Lista de figuras	6
Siglas, acrónimos y abreviaturas	8
Resumen	9
Abstract	10
1. Descripción del problema	11
1.1 Problema de negocio	12
1.2 Aproximación desde la analítica de datos	12
1.3 Origen de los datos	12
1.4 Métricas de desempeño	13
2. Objetivos	14
2.1 Objetivo general	14
2.2 Objetivos específicos.....	14
3. Datos.....	15
3.1 Datos originales.....	15
3.2 Datasets	16
3.3 Analítica descriptiva.....	16
3.4 Análisis de Componentes Principales	21
4. Proceso de analítica	22
4.1 Pipeline principal.....	22
4.2 Preprocesamiento	23
4.3 Modelos	23
4.3.1 DNN	24

Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

4.3.2 RNN	26
4.3.3 LSTM.....	27
4.3.3 GRU	29
4.3.4 SARIMAX	30
4.4 Métricas	31
5. Metodología	32
5.1 Baseline	32
5.2 Validación	34
5.3 Iteraciones y evolución.....	34
5.4 Herramientas	37
6. Resultados y discusión	37
6.1 Métricas.....	37
6.2 Evaluación cualitativa	38
6.2 Consideraciones de producción.....	40
7. Conclusiones	41
Referencias	42

Lista de tablas

Tabla 1. Parámetros utilizados en los modelos de Deep Learning.....	24
Tabla 2. Resultados baseline Modelos Deep Learning	32
Tabla 3. Resultados baseline SARIMAX.....	32
Tabla 4. Desempeño Modelo SARIMAX ajustado.....	35
Tabla 5. Top 10 modelos Deep Learning	38

Lista de figuras

Figura 1. Precio en Bolsa de la Energía - serie	17
Figura 2. Diagrama de Cajas y Bigotes.....	18
Figura 3. Autocorrelación de la serie	18
Figura 4. Diagrama de Cajas y Bigotes Mensual.....	19
Figura 5. Correlación de variables	20
Figura 6. Estructura final de los datos.....	21
Figura 7. Análisis de Componentes Principales del Dataset.....	21
Figura 8. Diagrama de flujo	23
Figura 9. Red Neuronal Profunda (DNN).....	25
Figura 10. Arquitectura Modelo DNN.....	25
Figura 11. Red Neuronal Recurrente (RNN)	26
Figura 12. Arquitectura Modelo RNN	27
Figura 13. Red Neuronal LSTM	28
Figura 14. Arquitectura Modelo LSTM.....	28
Figura 15. Red Neuronal Recurrente Cerrada GRU	29
Figura 16. Arquitectura Modelo GRU	30
Figura 17. Arquitectura Modelo SARIMAX	31
Figura 18. Predicción Modelo baseline GRU	33
Figura 19. Desempeño vs. épocas Modelo baseline GRU	33
Figura 20. Predicción Modelo SARIMAX	34
Figura 21. Evolución del MSE en el entrenamiento modelo DNN.....	35
Figura 22. Evolución del MSE en el entrenamiento modelo RNN.....	36
Figura 23. Evolución del MSE en el entrenamiento modelo LSTM.....	36

Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

Figura 24. Evolución del MSE en el entrenamiento modelo GRU	36
Figura 25. Predicción con mejor modelo DNN	38
Figura 26. Predicción con mejor modelo RNN.....	39
Figura 27. Predicción con mejor modelo LSTM	39
Figura 28. Predicción con mejor modelo GRU.....	39

Siglas, acrónimos y abreviaturas

MEM	Mercado de Energía Mayorista
XM	XM Compañía de Expertos en Mercados
SIN	Sistema Interconectado Nacional
CREG	Comisión de Regulación de Energía y Gas
UPME	Unidad de Planeación Minero-Energética
RNN	Recurrent Neural Network
LSTM	Long short-term memory
GRU	Gated Recurrent Unit
VBA	Visual Basic for Applications
ONI	Ocean Niño Index

Resumen

El precio de la energía eléctrica es uno de los productos con más volatilidad en los mercados mundiales, y el Mercado de Energía Mayorista Colombiano no es la excepción. Por lo cual se hace necesario, desde diferentes sectores del mercado y especialmente para los actores que participan en él conocer la tendencia futura en su comportamiento con la finalidad de minimizar el riesgo de exposición a la bolsa en el caso de los agentes comercializadores y maximizar los beneficios en el caso de los agentes generadores. Además, para el caso de los Usuarios No Regulados permite reducir sus costos a la hora de realizar compra de energía en contratos.

Este trabajo pretende implementar modelos de series temporales, como el SARIMAX y modelos de Deep Learning como: DNN, RNN, LSTM y GRU con la finalidad de predecir el precio en bolsa de la energía en Colombia en un horizonte de 30 días y considerando las siguientes variables: Precio Promedio Ponderado de la Energía en Bolsa, Demanda de Energía, Disponibilidad Declarada de los Recursos Hidráulicos, Disponibilidad Declarada de los Recursos Térmicos, Generación de los Recursos Hidráulicos, Generación de los Recursos Térmicos, Aportes Hídricos del Sistema, Precio de Oferta de los Recursos, Volumen útil de los embalses, índice ONI (Ocean Niño Index) y Vertimientos del Sistema. Las métricas de desempeño utilizadas para la evaluación de los modelos fueron: MSE, RMSE, MAE y MAPE; con base en las cuales el modelo que presentó un mejor desempeño en el horizonte establecido fue el RNN con un error del 9.84%.

Palabras clave: precio de bolsa, energía eléctrica, series temporales, Deep Learning, predicción, pronóstico

Repositorio del proyecto: <https://github.com/cjusquini/MonografiaUdeA>

Abstract

The price of electricity is one of the most volatile products in world markets, and the Colombian Wholesale Energy Market is no exception. Therefore, it is necessary, from different sectors of the market and especially for the actors that participate in it, to know the future trend in its behavior to minimize the risk of exposure to the stock market in the case of the trading agents and maximize the benefits in the case of the generating agents. In addition, in the case of Non-Regulated Users, it allows reducing their costs when purchasing energy in contracts.

This research work intends to implement time series models, such as SARIMAX and Deep Learning models such as: DNN, RNN, LSTM and GRU in order to predict the energy stock market price in Colombia in a 30-day horizon and considering the following variables: Weighted Average Energy Exchange Price, Energy Demand, Declared Availability of Hydraulic Resources, Declared Availability of Thermal Resources, Generation of Hydraulic Resources, Generation of Thermal Resources, Water Contributions of the System, Offer Price of Resources, Useful Volume of Reservoirs, ONI (Ocean Niño Index) and Discharges of the System. The performance metrics used for the evaluation of the models were: MSE, RMSE, MAE and MAPE; based on which the model that presented the best performance in the established horizon was the RNN with an error of 9.84%.

Keywords: stock market price, electric power, time series, Deep Learning, prediction, forecasting.

Project repository: <https://github.com/cjusquini/MonografiaUdeA>

1. Descripción del problema

El Mercado de Energía Mayorista en Colombia (MEM) es un componente fundamental del sector eléctrico del país administrado por XM, donde se negocia la compra y venta de energía entre generadores y comercializadores. Este mercado puede definirse como un conjunto de sistemas de intercambio de información entre generadores y comercializadores de grandes bloques de energía eléctrica en el Sistema Interconectado Nacional (SIN) para suscribir contratos de energía a largo plazo y comprar energía en bolsa sobre cantidades y precios definidos con sujeción a los reglamentos y normas establecidas (Zapata & Ochoa, 2008). Asimismo, desempeña un papel crucial en el suministro de energía confiable y económica para los colombianos, promueve la competencia y la eficiencia en toda la cadena de valor del sector energético colombiano. Las transacciones en el MEM se realizan a través de contratos bilaterales de energía y transacciones en la bolsa de energía.

Este mercado se rige por reglas y normativas establecidas por la Comisión de Regulación de Energía y Gas (CREG), con el objetivo de garantizar la competencia, transparencia y eficiencia en las operaciones (XM Administradores del mercado eléctrico, s. f.). En estas operaciones se prioriza la oferta de energía más económica para cubrir la proyección de demanda suministrada por la Unidad de Planeación Minero-Energética (UPME); esto con el fin de minimizar los costos de operación y garantizar una operación segura del SIN (Sede electrónica Comisión de Regulación de Energía y Gas (CREG), s. f.).

En este orden de ideas, es oportuno aclarar que el precio de bolsa, también conocido como precio spot, es un indicador clave en el MEM. Este precio en condiciones normales de operación corresponde al mayor precio de oferta de las unidades con despacho centralizado que han sido programadas por orden de mérito para generar en el despacho ideal, y que no presentan inflexibilidad. Representa un precio único para el sistema interconectado en cada periodo horario. El precio de bolsa refleja las condiciones de mercado, la disponibilidad de recursos energéticos y otros factores como la climatología y el estado y disponibilidad de los activos del SIN (XM Administradores del mercado eléctrico, s. f.).

1.1 Problema de negocio

El precio de la energía eléctrica es uno de los productos con más volatilidad en los mercados mundiales. El cambio diario o mensual del precio spot de energía puede ser de hasta el 50%, mientras que, al mismo tiempo, para otras materias primas es de solo el 5%. La planeación de los ingresos para los agentes o actores del sector eléctrico en Colombia, tales como generadores, comercializadores, grandes usuarios o clientes finales (grandes industrias), dependen en gran medida de la evolución del precio spot de la energía. De acá surge la importancia para estos agentes prever o desarrollar modelos de pronósticos de precios de la energía en el corto y mediano plazo (Muñoz-Santiago et al., 2017). Un precio alto de la energía en bolsa puede incentivar la inversión en nuevas tecnologías, proyectos de generación y la contratación de energía a largo plazo con el fin de mitigar el riesgo asociado a la exposición a la bolsa de energía (Asociación Nacional de Empresas Generadoras – ANDEG, s. f.), especialmente en épocas de baja hidrología en el país, esto debido a que la capacidad instalada de la matriz de generación eléctrica colombiana es un 65% Hidráulica y por lo cual tiene un alto impacto en la formación del precio de la Energía en Bolsa.

1.2 Aproximación desde la analítica de datos

En este trabajo se pretende estudiar el comportamiento del precio de la energía en bolsa en Colombia, su relación con otras variables del mercado eléctrico colombiano a través de análisis estadísticos como la correlación y distribuciones de probabilidad. Además, realizar el pronóstico en un horizonte de mediano plazo utilizando modelos de aprendizaje profundo basados en redes neuronales como lo son: DNN, RNN, LSTM y GRU; y modelos ARIMA analizando cada una de las componentes de la serie temporal.

1.3 Origen de los datos

Los datos para este trabajo fueron obtenidos a través de dos fuentes, la API de XM (EquipoAnaliticaXM/API_XM, s. f.) y de la National Oceanic and Atmospheric Administration (Climate Prediction Center - ONI, s. f.). En la API XM se tiene disponible la información pública asociada al Mercado de Energía Eléctrico Colombiano y puede ser obtenida desde diferentes herramientas, entre ellas el lenguaje de programación Python, Excel (VBA), entre otras. facilitando así la extracción de la información.

En este trabajo, la extracción de los datos se realizó a través del lenguaje de programación Python. Por otra parte, el índice ONI se tomó del centro de predicción climática de la NOAA. La cual es responsable de describir y predecir los cambios en el medio ambiente, investigando los océanos, la atmósfera, el espacio y el sol.

Para resolver el problema de la predicción del precio de bolsa de la energía eléctrica, se consideran las siguientes variables: el Precio Promedio Ponderado de Bolsa, Demanda de Energía, Disponibilidad Declarada de los Recursos Hidráulicos, Disponibilidad Declarada de los Recursos Térmicos, Generación de los Recursos Hidráulicos, Generación de los Recursos Térmicos, Aportes Hídricos del Sistema, Precio de Oferta de los Recursos, Volumen útil de los embalses, índice ONI (Ocean Niño Index) y Vertimientos del Sistema.

1.4 Métricas de desempeño

En el ámbito del Machine Learning, evaluar el desempeño de un modelo es fundamental para comprender su efectividad y su capacidad para generalizar sobre datos no vistos. Las métricas de desempeño proporcionan una forma cuantitativa de medir el grado de ajuste entre las predicciones del modelo y los valores reales del conjunto de datos. Cuatro de las métricas más comunes utilizadas para este tipo de problema son:

- **Error Absoluto Medio (MAE):** Esta métrica calcula el promedio de las diferencias absolutas entre las predicciones del modelo y los valores observados. Se expresa como la suma de las diferencias absolutas dividida por el número total de observaciones, lo que proporciona una medida de la magnitud promedio de los errores del modelo (Steurer & Hill, 2020).
- **Error Cuadrático Medio (MSE):** Este indicador mide el promedio de los cuadrados de las diferencias entre las predicciones del modelo y los valores reales. Es una métrica útil, ya que penaliza de manera más significativa los errores grandes. Esto se debe al cuadrado en el cálculo, lo que puede ayudar a identificar discrepancias importantes entre las predicciones y los valores observados (Steurer & Hill, 2020).
- **Raíz del Error Cuadrático Medio (RMSE):** Esta métrica es simplemente la raíz cuadrada del MSE y proporciona una medida del error promedio en la misma escala que la variable objetivo. Es particularmente útil para interpretar el error en términos del

dominio del problema, ya que está en las mismas unidades que los datos originales (Steurer & Hill, 2020).

- **Error Porcentual Absoluto Medio (MAPE):** En este caso se calcula el error porcentual promedio entre las predicciones del modelo y los valores observados. Es especialmente útil cuando se necesita evaluar el desempeño del modelo en términos de precisión porcentual, lo que lo hace adecuado para aplicaciones donde la interpretación de los errores en términos relativos es crucial (Steurer & Hill, 2020).

2. Objetivos

2.1 Objetivo general

Implementar modelos de pronóstico utilizando análisis de series temporales y modelos de Deep Learning para pronosticar de manera diaria el precio en bolsa de la energía en Colombia con un horizonte de un mes, esto permitirá a los agentes de mercado disminuir riesgos, cubrirse ante la volatilidad del precio en bolsa y maximizar beneficios al momento de transar energía en el mercado.

2.2 Objetivos específicos

- Identificar aspectos claves en el tratamiento de series temporales y modelos de Deep Learning para pronóstico, incluyendo diferentes arquitecturas y técnicas para la adecuación de la base de datos.
- Desarrollar módulos para el pre-procesamiento y análisis exploratorio de la base de datos originales para el entrenamiento de los modelos a implementar.
- Implementar diferentes modelos para el análisis de series temporales como ARIMA y de Deep Learning basados en Redes neuronales como RNN, GRU y LSTM, para el pronóstico del Precio de la Energía en Bolsa.
- Evaluar el desempeño de los modelos usando métricas adecuadas al problema del pronóstico de series temporales como el MAE, MSE, RMSE y MAPE.

3. Datos

3.1 Datos originales

Los datos originales fueron obtenidos a través de dos fuentes, la API de XM (EquipoAnaliticaXM/API_XM, s. f.) y de la National Oceanic and Atmospheric Administration (Climate Prediction Center - ONI, s. f.). En la API XM se tiene disponible la información pública asociada al Mercado de Energía Eléctrico Colombiano y puede ser usada desde diferentes herramientas, entre ellas el lenguaje de programación Python, Excel (VBA), etc. facilitando la extracción de la información. En este caso, se realizó a través del lenguaje de programación Python. Por otra parte, el índice ONI se tomó del centro de predicción climática de la NOAA.

La tabla se compone de 3652 filas y 16 columnas equivalentes a un registro único por día de la información desde el año 2010 hasta el 2019.

Formato: los datos están en formato .CSV, con un peso de 611 KB.

Distribución: Los datos están en granularidad diaria y representan un valor totalizado a nivel del sistema. Las variables se encuentran en kWh y el Precio de Bolsa se encuentra en COP/kWh y su estructura es la siguiente:

- **fecha:** Fecha de operación
- **precio_bolsa:** Precio en bolsa de la energía en la última versión de la liquidación realizada por XM, en \$COP/kWh.
- **demanda:** Demanda Real de Energía (no incluye las pérdidas técnicas y no técnicas), en kWh.
- **disponibilidad_declarada:** Disponibilidad Declarada por los recursos para la operación, en kWh.
- **generacion:** Generación de todos los recursos, en kWh.
- **aportes_energia:** Aportes Hídricos del sistema, en kWh.
- **precio_oferta:** Precio de Oferta de los recursos Despachados Centralmente, en \$COP/kWh.
- **vol_util:** Volumen útil de los embalses del SIN, en kWh.
- **indice_ONI:** Índice Niño Oceánico

3.2 Datasets

Para predecir el Precio en Bolsa de la Energía en un horizonte de 30 días, previamente se realizó la limpieza e imputación de datos y se dividieron los datos en conjuntos de entrenamiento, validación y test. Para el caso del modelo ARIMA, se utilizó únicamente la serie del Precio de Bolsa, tomando el primer 80% de los datos para entrenamiento y el 20% restante para pruebas, esto conservando el orden para no afectar su dependencia temporal.

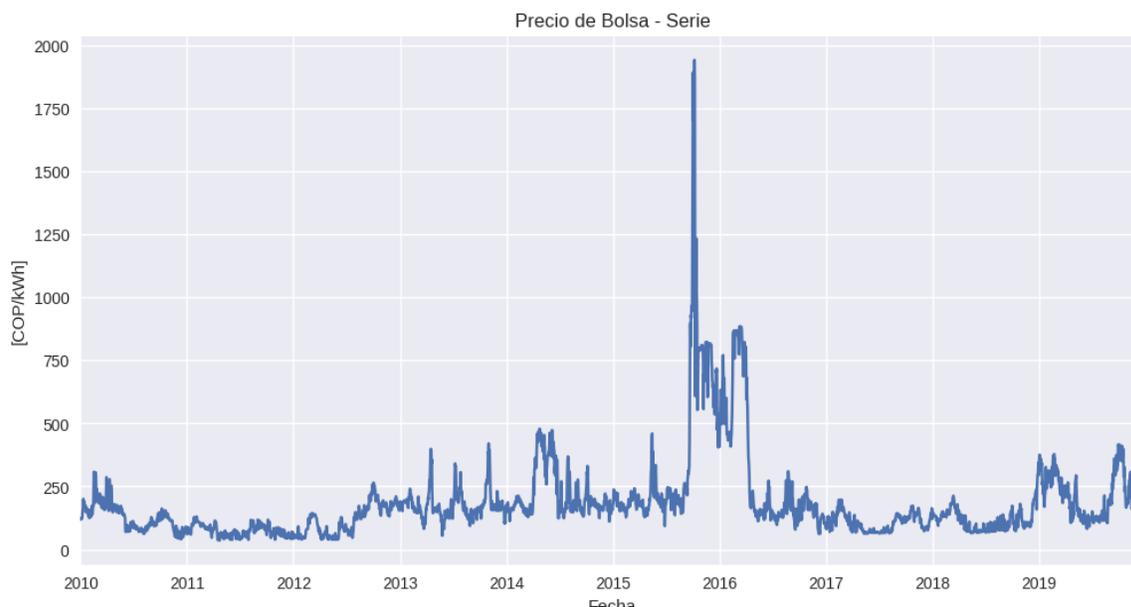
Para los modelos de Deep Learning los datos se transformaron a tensores (3 dimensiones) para ser ingresados a los modelos; se tomó el primer 80% de los datos para entrenamiento y el 20% restante para validación y pruebas, conservando el orden.

3.3 Analítica descriptiva

Se realizó un análisis univariado de la serie del Precio en Bolsa de la Energía, incluyendo gráfico de líneas, diagrama de cajas y bigotes, autocorrelación de la serie y correlación de las variables con la variable de interés.

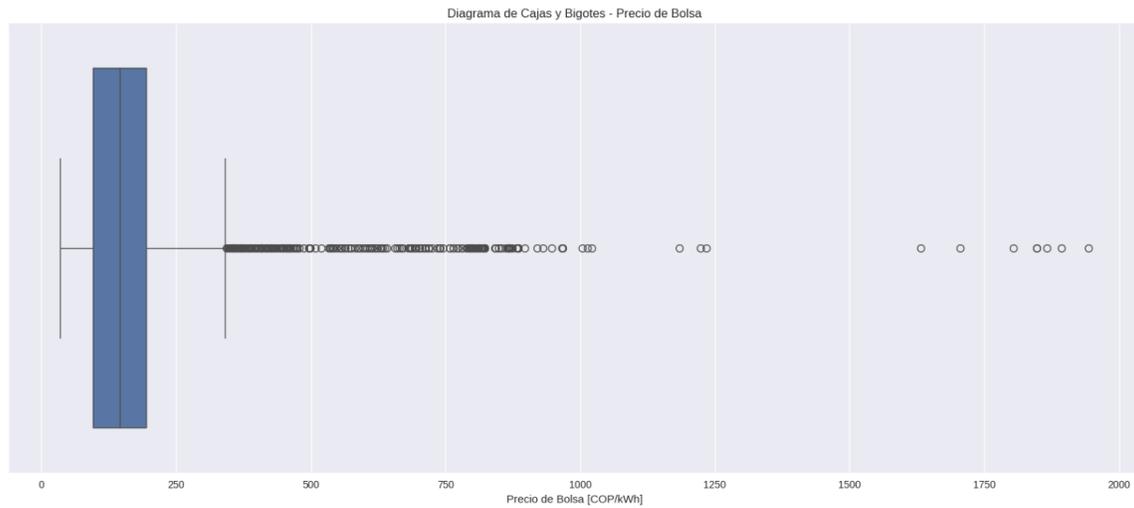
La serie del Precio en Bolsa de la Energía se observa en la **Figura 1.**, en la cual se evidencia su estacionalidad y que para algunos meses del periodo 2015-2016 se presentan datos atípicos asociados al fenómeno del niño ocurrido en esos años. Este aumento en el precio se presentó debido a la escasez hídrica y la necesidad de realizar el despacho de recursos térmicos en el sistema para atender la demanda nacional, representando una diferencia importante respecto a los recursos de generación con fuentes renovables. Luego de este periodo, el precio comenzó a regularse debido a la recuperación de los embalses del SIN.

Figura 1. Precio en Bolsa de la Energía - serie



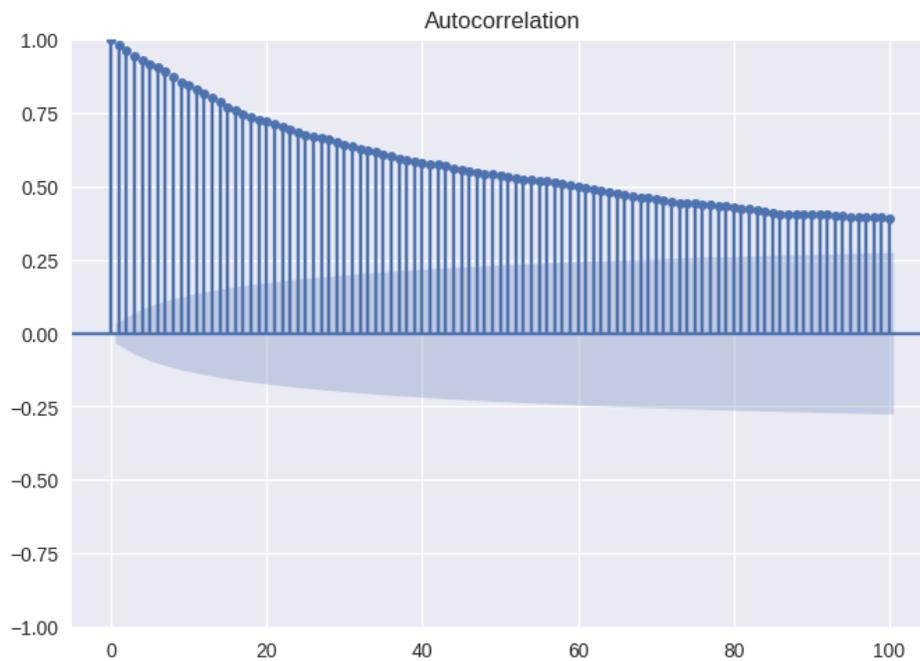
En la **Figura 2.** se presenta un diagrama de cajas y bigotes dónde es posible observar los cuartiles y datos atípicos. El valor mínimo de la serie es de 35.35 \$COP/kWh, el primer cuartil tiene un valor de 95.6 \$COP/kWh, el segundo, equivalente a la mediana tiene un valor de 146.2 \$COP/kWh y el tercero tiene un valor de 194.49 \$COP/kWh. La serie presenta un valor máximo de 1942.69 \$COP/kWh y algunos datos atípicos entre el tercer cuartil y el máximo debido a la escasez hídrica mencionada previamente. Este valor máximo se alcanzó en el periodo 2015-2016 por la escasez de lluvias y, por ende, una mejor Disponibilidad Declarada de los recursos hidráulicos. Asimismo, es notable que el precio no oscila en valores tan altos, su valor promedio es de 185 \$COP/kWh.

Figura 2. Diagrama de Cajas y Bigotes



En la **Figura 3.** se observa la correlación vs. los retardos, es decir, cómo se relaciona un elemento con los anteriores. Es notorio que la serie se encuentra autocorrelacionada, pues se conserva hasta un 50% de relación lineal de los elementos considerando hasta 50 retardos. Es posible asociar esta alta cantidad de retardos con la volatilidad de la serie, si bien el Precio en Bolsa de la Energía tiende a ser muy volátil generalmente no se dan cambios bruscos, por lo cual es apropiado plantear un pronóstico de la serie con base en sus valores anteriores.

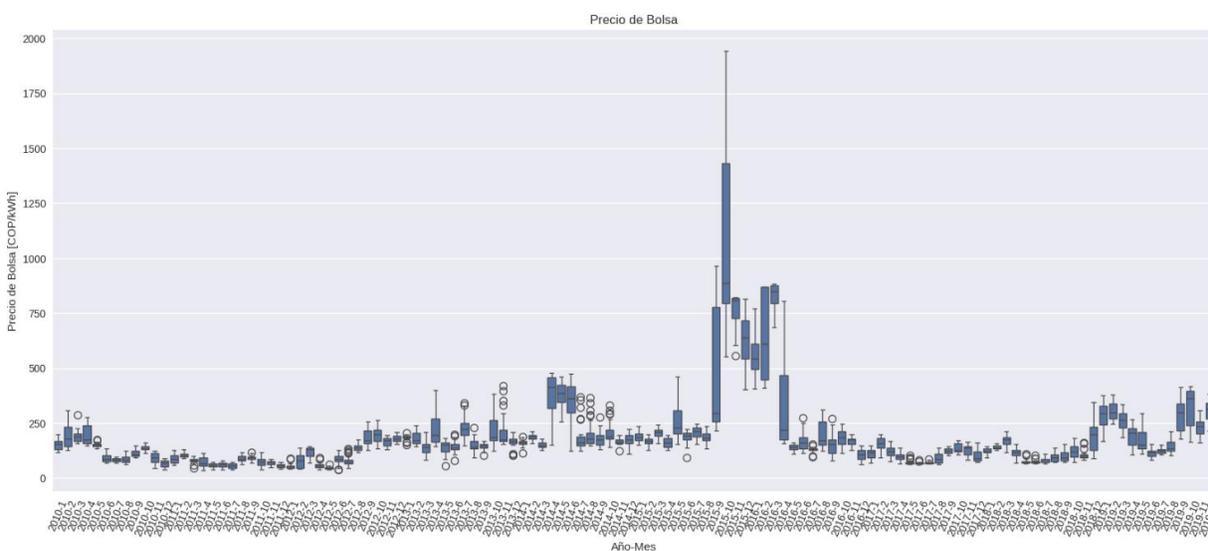
Figura 3. Autocorrelación de la serie



Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

En la **Figura 4**, es posible observar que en verano (meses: dic., ene., feb., mar., abr.), debido a las altas temperaturas, alta Demanda de Energía y el bajo nivel de los embalses del sistema los precios tienden a ser más volátiles y con una tendencia al alza. Especialmente en los años 2015 y 2016, en los que se presentó el fenómeno de El Niño, el cual fue el más severo y uno de los más largos desde 1950. La anterior anomalía se tradujo en una sensación térmica extrema en todo el país. En 54 municipios se registraron temperaturas superiores a los 40°C a la sombra. Los departamentos más afectados por las altas temperaturas fueron Antioquia, Bolívar, Boyacá, Caldas, Cesar, Chocó, Cundinamarca, Huila, La Guajira, Magdalena y Nariño (IDEAM, 2016). Así mismo, durante el 2015 se registró una reducción en el volumen de lluvias de hasta un 30%, principalmente en las áreas andina y caribe (Contreras, 2016).

Figura 4. *Diagrama de Cajas y Bigotes Mensual*



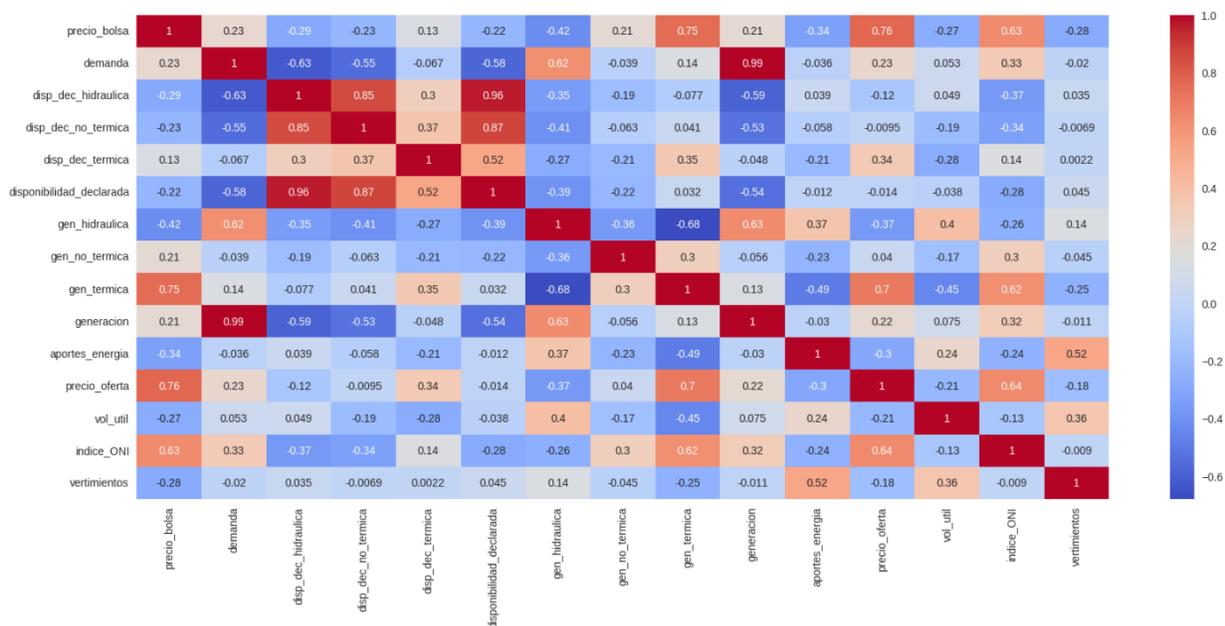
Finalmente, se realizó un análisis de correlación para eliminar aquellas variables linealmente dependientes que, comparadas con el Precio en Bolsa de la Energía, aportan el mismo valor al modelo de datos. Se fijó un umbral del 80% en la dependencia lineal como criterio para conservar una de las dos variables linealmente correlacionadas.

En la **Figura 5** se observa la matriz de correlación y de la cual es posible eliminar las siguientes variables y algunas agregaciones de acuerdo con el criterio establecido: Generación Real, Disponibilidad Declarada y Disponibilidad Declarada de Recursos no térmicos para finalmente obtener la estructura final a trabajar de los datos en la **Figura 6** y la cual tiene las

Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

siguientes variables: Precio en Bolsa de la Energía, Demanda Real Nacional, Disponibilidad Declarada de los recursos hidráulicos, Disponibilidad Declarada de los recursos térmicos, Generación Real de los recursos hidráulicos, Generación Real de los recursos térmicos, Aportes Hídricos de los ríos del SIN, Precio de Oferta de los recursos, Volúmen Útil de los embalses del SIN, Índice ONI y Vertimientos del SIN.

Figura 5. Correlación de variables



De la matriz de correlación de variables en la **Figura 5** es posible establecer ciertas conclusiones:

- El precio en Bolsa de la Energía se encuentra altamente correlacionado de manera positiva con el Precio de Oferta de los recursos, la Generación Térmica y el Índice ONI, es decir, al aumentar una variable la otra tiene la misma tendencia.
- El precio en Bolsa de la Energía se encuentra correlacionado de manera negativa con la Disponibilidad Declarada de los recursos hidráulicos, la Generación de los recursos hidráulicos, Aportes Hídricos de los ríos y Volumen Útil de los embalses. Indicando que a medida que alguna de estas variables aumenta, el precio tiende a disminuir, lo cual es consistente con las variables y la regulación del MEM ya que las energías renovables tienden a ser más económicas.

Figura 6. Estructura final de los datos

fecha	precio_bolsa	demanda	disp_dec_hidraulica	disp_dec_termica	gen_hidraulica	gen_no_termica	gen_termica	aportes_energia	precio_oferta	vol_util	indice_ONI	vertimientos
2019-12-27	331.918402	1.982598e+08	215490.925926	103867.906746	1.250668e+08	4106171.06	64368474.66	100303900.0	307.435903	1.117787e+10	0.5	0.000000e+00
2019-12-28	279.387930	1.921742e+08	215309.518519	104209.728836	1.304055e+08	4890928.09	52769349.45	80942900.0	311.201968	1.112934e+10	0.5	0.000000e+00
2019-12-29	246.472593	1.796031e+08	218024.148148	106788.868564	1.203988e+08	4056935.56	50588485.89	78816500.0	326.929484	1.108964e+10	0.5	0.000000e+00
2019-12-30	323.892900	1.902380e+08	212496.222222	109175.364583	1.290546e+08	5109396.88	52338556.02	110690400.0	314.140355	1.107393e+10	0.5	0.000000e+00
2019-12-31	274.955105	1.747399e+08	217699.370370	106451.575203	1.138374e+08	4528824.75	51810625.29	167617400.0	324.844129	1.112610e+10	0.5	8.970006e+06

3.4 Análisis de Componentes Principales

El Análisis de Componentes Principales (PCA por sus siglas en inglés) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un conjunto de datos con muchas variables, el objetivo será reducirlas a un menor número explicando la mayor variabilidad posible en la información a través de su representación en componentes principales (Gurrea, 2000).

En este caso, se realizó una estandarización de los datos y se aplicó la técnica PCA para explicar una variabilidad del 95% de los datos, pasando de 14 variables a 7 componentes principales conservando la variable de análisis, como se puede observar en la **Figura 7**.

Figura 7. Análisis de Componentes Principales del Dataset

fecha	PC0	PC1	PC2	PC3	PC4	PC5	PC6	PC7	precio_bolsa
2010-01-01	4.203538	3.565771	4.510770	1.345584	-0.627948	1.078003	-1.006162	1.445185	120.501228
2010-01-02	4.160252	2.920335	4.233935	1.235097	-0.609312	1.287115	-1.051695	1.014595	118.495340
2010-01-03	4.349066	3.062140	4.175426	1.408133	-0.733180	0.869889	-0.888278	0.989648	121.160145
2010-01-04	4.076867	1.784141	3.933963	1.067841	-0.635992	1.248735	-0.909209	0.293162	123.962937
2010-01-05	4.227628	1.394444	4.084858	1.172538	-0.644346	1.478602	-1.240434	-0.174062	120.666437

Dependiendo el objetivo del proyecto, la complejidad de los modelos y qué tan importante sea conservar, o no las variables originales, puede ser más beneficioso trabajar con las componentes principales del conjunto de datos.

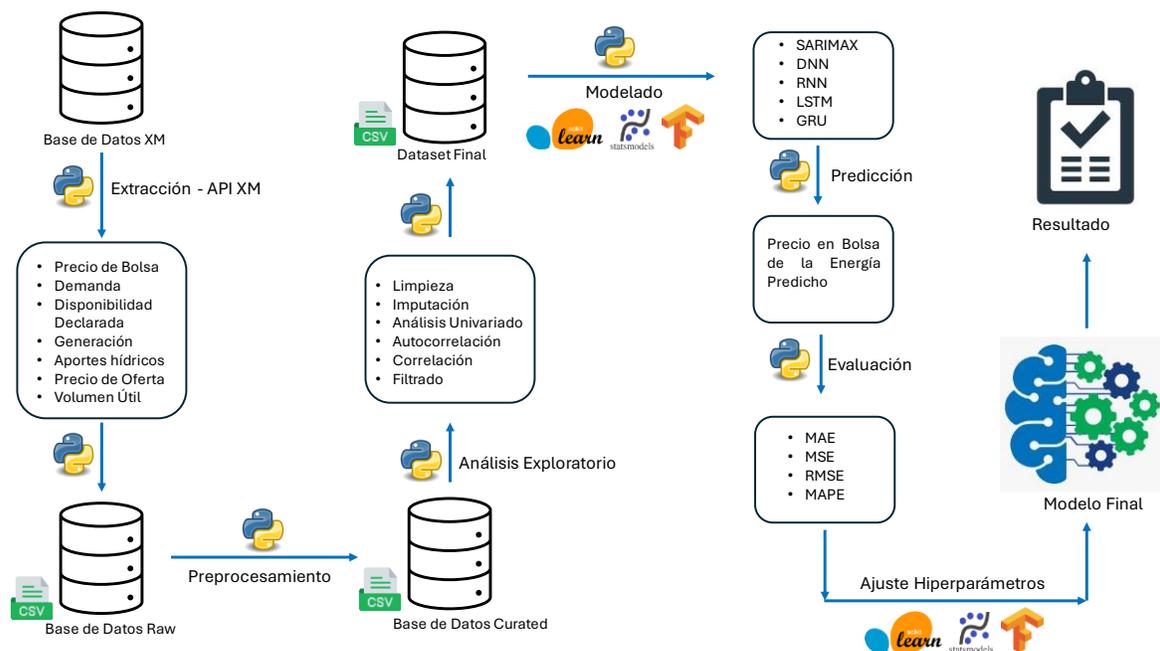
4. Proceso de analítica

4.1 Pipeline principal

A continuación, en la **Figura 8** se muestra el diagrama de flujo del trabajo realizado. En primer lugar, se realiza la extracción de la información de la base de datos de XM a través de la API XM la cual es consultada utilizando el lenguaje de programación Python y se almacena en formato .csv en su estructura original. Luego, se realiza el preprocesamiento de la información, el cual incluye limpieza y estructura de datos para luego ser almacenados en formato .csv bajo la etiqueta de “procesado”. Posteriormente se realiza el análisis exploratorio de datos (EDA) el cual incluye limpieza, imputación, análisis univariado, análisis de autocorrelación, correlación de variables y filtrado de las mismas para eliminar la redundancia en el conjunto de datos. En este paso se realizó la unión de todas las variables en una sola estructura denominada “Dataset Final” sobre el cual se realizará la implementación de modelos.

Finalmente, utilizando algunas librerías de Python como: SciKit Learn, TensorFlow, StatModels, Pandas, Numpy, entre otras. Se realiza la separación del conjunto de datos, es decir, la división de los datos en datos de entrenamiento, validación y pruebas. Estos conjuntos se usarán para implementar los modelos, evaluarlos y realizar el ajuste de hiperparámetros, que permitan minimizar el error de acuerdo con las métricas de desempeño establecidas en la predicción del Precio en Bolsa de la Energía.

Figura 8. Diagrama de flujo



4.2 Preprocesamiento

En el preprocesamiento de datos se realizaron las agregaciones necesarias a cada una de las variables para darles estructura de serie temporal multivariada. También, se aplicaron técnicas para la limpieza e imputación de los datos, seleccionar las variables relevantes y estandarizar el conjunto de datos con la finalidad de obtener un resultado adecuado en la implementación de modelos.

4.3 Modelos

Para realizar el pronóstico del Precio en Bolsa de la Energía en Colombia, se utilizaron 5 modelos, uno de ellos basado en técnicas estadísticas para la predicción de series de tiempo, SARIMAX y otros cuatro modelos basados en aprendizaje profundo a través de Redes Neuronales, estos son: DNN, RNN, LSTM y GRU. Finalmente, se pretende comparar los diferentes modelos considerando las métricas de desempeño en el pronóstico de la variable de interés

En cada uno de los modelos se realizaron distintas iteraciones variando distintos parámetros, en el caso del modelo SARIMAX, se variaron los parámetros p , d , q , P , D , Q y s . Para los modelos de Deep Learning se variaron los parámetros: número de neuronas del modelo, número de retardos, optimizador, función de pérdida, épocas de entrenamiento y el tamaño del lote en entrenamiento. Además, se utilizó la misma semilla, se fijaron dos capas de Dropout para eliminar

el sobreajuste y una capa Densa de 20 Neuronas antes de la salida con la finalidad de mejorar el desempeño.

En la **Tabla 1** se observan los parámetros utilizados en las diferentes iteraciones.

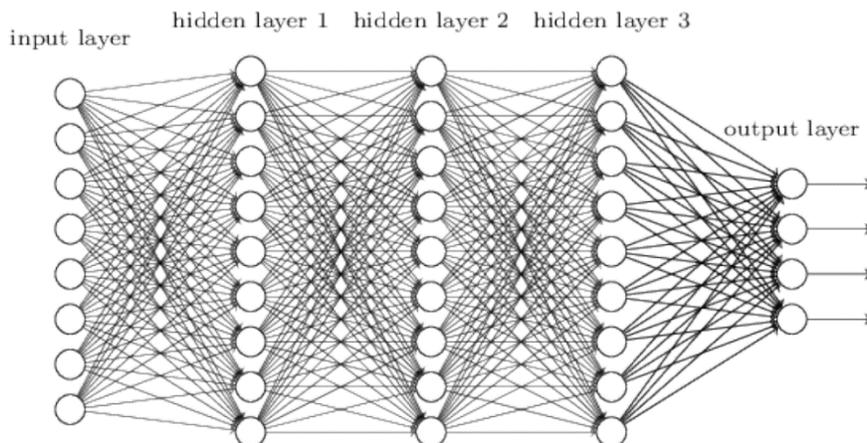
Tabla 1. *Parámetros utilizados en los modelos de Deep Learning*

Parámetro	Valor
Retardos	[60]
Neuronas	[30, 40]
Modelos	['DNN', 'RNN', 'LSTM', 'GRU']
Métricas	['MAE', 'MAPE', 'MSE', 'RMSE']
Horizonte	[31]
Optimizadores	Adam(learning_rate = 1e-6)
Función de pérdida	'MSE'
Épocas	[100]
Lote	[64, 32]

4.3.1 DNN

Una red neuronal es una serie de capas de neuronas, cada una calcula una suma ponderada de la salida de las neuronas de la capa anterior y aplica una función de activación. Las DNN son redes neuronales con varias capas ocultas. En general, cada capa consigue extraer un cierto grado de abstracción, de modo que una DNN permite aprender patrones complejos y genéricos (Arana, 2021). En la **Figura 9** se presenta una arquitectura de Red Neuronal Profunda, donde se puede apreciar las diferentes capas de neuronas y cómo cada neurona de la capa siguiente se conecta con la salida de cada neurona de la anterior, así capa a capa, hasta la salida.

Figura 9. Red Neuronal Profunda (DNN).



Nota. Fuente: (Nielsen, 2015)

Con base en los parámetros iterados anteriormente y bajo la misma semilla utilizando la API Keras de TensorFlow para su implementación, la arquitectura del mejor modelo de tipo DNN se puede observar en la **Figura 10**.

Figura 10. Arquitectura Modelo DNN

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 60, 30)	390
flatten (Flatten)	(None, 1800)	0
dropout (Dropout)	(None, 1800)	0
dense_1 (Dense)	(None, 20)	36020
dropout_1 (Dropout)	(None, 20)	0
dense_2 (Dense)	(None, 1)	21

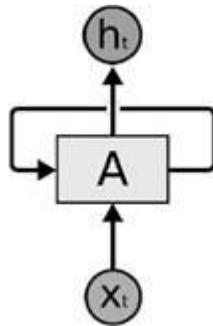
```
=====  
Total params: 36431 (142.31 KB)  
Trainable params: 36431 (142.31 KB)  
Non-trainable params: 0 (0.00 Byte)  
=====
```

4.3.2 RNN

Las redes neuronales recurrentes son una clase de aprendizaje profundo basada en los trabajos de David Rumelhart en 1986. Las RNN son conocidas por su capacidad para procesar y obtener información de datos secuenciales. Por lo tanto, el análisis de vídeo, la subtítulos de imágenes, el Procesamiento del Lenguaje Natural (PLN) y el análisis de la música, dependen de las capacidades de las redes neuronales recurrente, pues las RNN capturan activamente sus dependencias secuenciales y temporales (Arana, 2021).

En la **Figura 11** se puede observar la arquitectura básica de una Red Neuronal Recurrente.

Figura 11. Red Neuronal Recurrente (RNN)



Nota. Fuente: (Nielsen, 2015), (Chung et al., 2014)

Con base en los parámetros iterados anteriormente y bajo la misma semilla utilizando la API Keras de TensorFlow para su implementación, la arquitectura del mejor modelo de tipo RNN se puede observar en la **Figura 12**.

Figura 12. *Arquitectura Modelo RNN*

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
simple_rnn (SimpleRNN)	(None, 40)	2120
flatten (Flatten)	(None, 40)	0
dropout (Dropout)	(None, 40)	0
dense (Dense)	(None, 20)	820
dropout_1 (Dropout)	(None, 20)	0
dense_1 (Dense)	(None, 1)	21

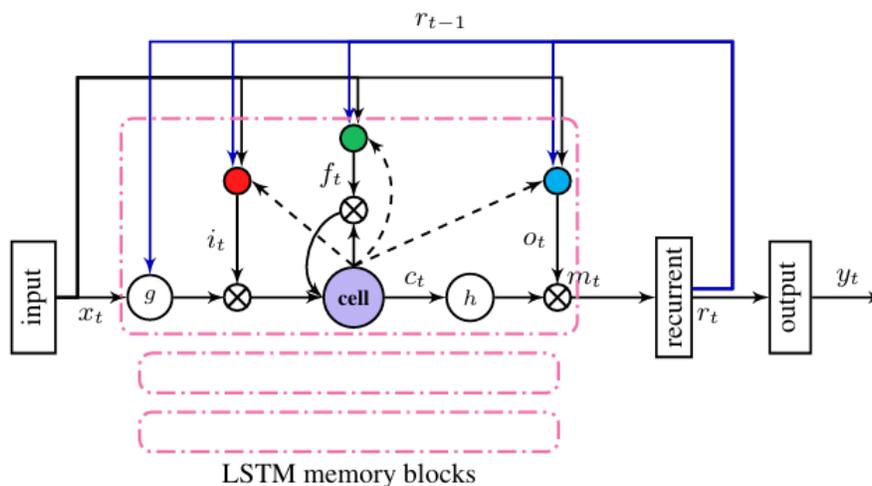
```
=====  
Total params: 2961 (11.57 KB)  
Trainable params: 2961 (11.57 KB)  
Non-trainable params: 0 (0.00 Byte)  
=====
```

4.3.3 LSTM

Una red neuronal recurrente LSTM tiene como objetivo aprender dependencias a largo plazo; es decir, aprender las dependencias de valores futuros de una secuencia en función a los valores anteriores. Como toda red neuronal recurrente, una LSTM tiene la forma de una cadena con unidades (o celdas) repetidas, pero en vez de tener una estructura simple dentro de cada unidad, en una LSTM se tiene 4 componentes que interactúan entre sí.

En la **Figura 13** se puede observar la arquitectura básica de una Red Neuronal LSTM.

Figura 13. Red Neuronal LSTM



Nota. Fuente: (Sak, Senior & Beaufays, 2014)

Con base en los parámetros iterados anteriormente y bajo la misma semilla utilizando la API Keras de TensorFlow para su implementación, la arquitectura del mejor modelo de tipo LSTM se puede observar en la **Figura 14**.

Figura 14. Arquitectura Modelo LSTM

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30)	5160
flatten (Flatten)	(None, 30)	0
dropout (Dropout)	(None, 30)	0
dense (Dense)	(None, 20)	620
dropout_1 (Dropout)	(None, 20)	0
dense_1 (Dense)	(None, 1)	21

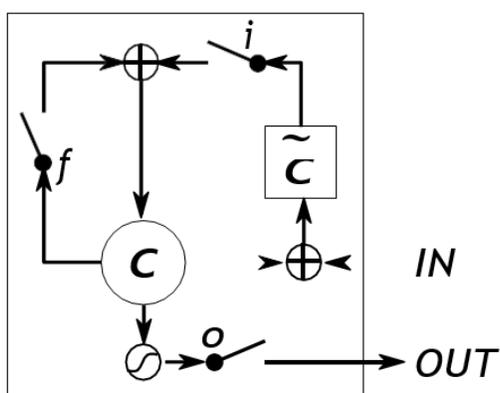
=====
 Total params: 5801 (22.66 KB)
 Trainable params: 5801 (22.66 KB)
 Non-trainable params: 0 (0.00 Byte)
 =====

4.3.3 GRU

Es una unidad recurrente controlada (GRU por sus siglas en inglés), para que cada unidad recurrente capture de forma adaptativa dependencias de diferentes escalas temporales. De forma similar a la unidad LSTM, la GRU tiene unidades de compuerta que modulan el flujo de información dentro de la unidad, sin tener células de memoria separadas (Chung et al., 2014).

En la **Figura 15** se puede observar la arquitectura básica de una red tipo GRU.

Figura 15. Red Neuronal Recurrente Cerrada GRU



Nota. Fuente: (Chung et al., 2014)

Con base en los parámetros iterados anteriormente y bajo la misma semilla utilizando la API Keras de TensorFlow para su implementación, la arquitectura del mejor modelo de tipo GRU se puede observar en la **Figura 16**.

Figura 16. Arquitectura Modelo GRU

```
Model: "sequential"
-----
```

Layer (type)	Output Shape	Param #
gru (GRU)	(None, 30)	3960
flatten (Flatten)	(None, 30)	0
dropout (Dropout)	(None, 30)	0
dense (Dense)	(None, 20)	620
dropout_1 (Dropout)	(None, 20)	0
dense_1 (Dense)	(None, 1)	21

```
-----
Total params: 4601 (17.97 KB)
Trainable params: 4601 (17.97 KB)
Non-trainable params: 0 (0.00 Byte)
-----
```

4.3.4 SARIMAX

La palabra ARIMA significa Modelos Autorregresivos Integrados de Medias Móviles. Definimos un modelo como **autorregresivo** si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error considerado innovación (De Arce y Mahia, 2003).

Teniendo en cuenta la definición anterior, la estacionariedad de la serie de acuerdo con la Prueba de Dickey-Fuller aumentada (Alonso, 2010) y que el Precio en Bolsa de la Energía se ve afectado por variables exógenas se realizó el ajuste del modelo de tipo SARIMAX (Modelos autorregresivos integrados de medias móviles estacional con variables exógenas) utilizando la librería statsmodels de Python y teniendo en cuenta el Precio de Oferta de los Recursos como variable exógena debido a su alta correlación con la variable de interés.

Con base en los parámetros iterados anteriormente y bajo la misma semilla utilizando el paquete tsa (Time Series Analysis) de la librería statmodels, la arquitectura del mejor modelo de tipo SARIMAX iterando los parámetros mencionados previamente entre los valores del 1 al 6 se puede observar en la *Figura 17*.

Figura 17. Arquitectura Modelo SARIMAX

SARIMAX Results						
Dep. Variable:	precio_bolsa		No. Observations: 2922			
Model:	SARIMAX(2, 0, 2)x(2, 0, 2, 12)		Log Likelihood		-14404.186	
Date:	Thu, 20 Jun 2024		AIC		28826.372	
Time:	15:27:13		BIC		28880.192	
Sample:	01-01-2010 - 12-31-2017		HQIC		28845.757	
Covariance Type: opg						
	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	1.4280	0.040	35.272	0.000	1.349	1.507
ar.L2	-0.4310	0.040	-10.798	0.000	-0.509	-0.353
ma.L1	-0.2942	0.041	-7.190	0.000	-0.374	-0.214
ma.L2	-0.1916	0.006	-30.179	0.000	-0.204	-0.179
ar.S.L12	0.9942	0.133	7.471	0.000	0.733	1.255
ar.S.L24	-0.6032	0.099	-6.101	0.000	-0.797	-0.409
ma.S.L12	-1.0144	0.138	-7.370	0.000	-1.284	-0.745
ma.S.L24	0.5649	0.109	5.200	0.000	0.352	0.778
sigma2	1110.6586	5.283	210.216	0.000	1100.303	1121.014
Ljung-Box (L1) (Q):	0.03	Jarque-Bera (JB):		3436331.97		
Prob(Q):	0.86	Prob(JB):		0.00		
Heteroskedasticity (H):	13.50	Skew:		-4.86		
Prob(H) (two-sided):	0.00	Kurtosis:		170.72		

4.4 Métricas

Las métricas utilizadas para evaluar el desempeño de la predicción del Precio en Bolsa de la Energía de los modelos implementados miden que tan precisa es la predicción respecto a los valores reales. Las métricas son: MAE, MSE, RMSE y MAPE.

Remitirse a la sección *1.4 Métricas de desempeño* para complementar su definición.

5. Metodología

5.1 Baseline

Se trabajó con los 5 modelos mencionados en el capítulo anterior. La primera iteración para los modelos de Deep Learning fue utilizando en todos los modelos un número de retardos de 60, 25 neuronas, 30 épocas y un tamaño de lotes de 32 para predecir en un horizonte de 30 días. Cabe aclarar que se fijó la semilla aleatoria utilizada de la misma manera en todos los modelos.

Para el modelo SARIMAX la iteración base fue utilizando los parámetros $p = 1$, $d = 0$, $q = 1$; $P = 1$, $Q = 1$, $D = 0$ y $s = 12$. Es oportuno mencionar, que debido a que la serie es estacionaria y no requiere diferenciación es posible dejar los parámetros de diferenciación d y D en 0 sin ver cambios significativos en el desempeño.

Cabe aclarar que el cálculo de las métricas de desempeño en esta y próximas iteraciones se hace en un horizonte de 30 días. Gráficamente se dejan todos los datos de prueba con la finalidad de observar el desempeño del modelo en el mediano y largo plazo.

Los resultados de los modelos de Deep Learning se observan en la **Tabla 2**.

Tabla 2. Resultados baseline Modelos Deep Learning

Modelo	Neuronas	LookBack	Epocas	BatchSize	MAE [\$COP/kWh]	MSE [\$COP/kWh]	RMSE [\$COP/kWh]	MAPE [%]
GRU	25	60	31	32	24.91	919.12	30.32	15.21
RNN	25	60	31	32	25.32	1,047.96	32.37	15.35
DNN	25	60	31	32	30.41	1,335.89	36.55	19.30
LSTM	25	60	31	32	47.57	3,747.00	61.21	28.73

Los resultados del modelo SARIMAX se observan en la **Tabla 3**.

Tabla 3. Resultados baseline SARIMAX

Modelo	p	d	q	P	D	Q	s	MAE [\$COP/kWh]	MSE [\$COP/kWh]	RMSE [\$COP/kWh]	MAPE [%]
SARIMAX	1	-	1	1	-	1	12	6.002	75.931	8.714	5.04

Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

Asimismo, tomando el modelo de Deep Learning con un mejor desempeño, en este caso el GRU es posible observar en la **Figura 18** los valores reales vs. los valores predichos del Precio en Bolsa de la Energía. Adicionalmente en la **Figura 19** es posible observar el desempeño del modelo durante el entrenamiento a lo largo de las épocas, en el cual a pesar de ser un modelo sencillo se aprecia una buena predicción y un MSE bajo que se ajusta de buena forma en el horizonte deseado.

Figura 18. Predicción Modelo baseline GRU

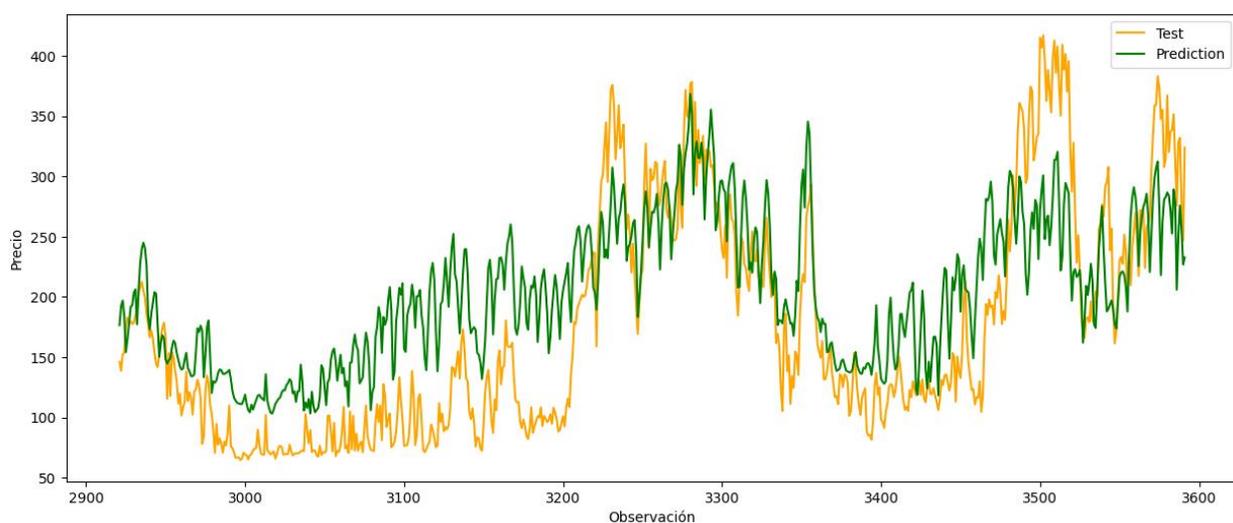
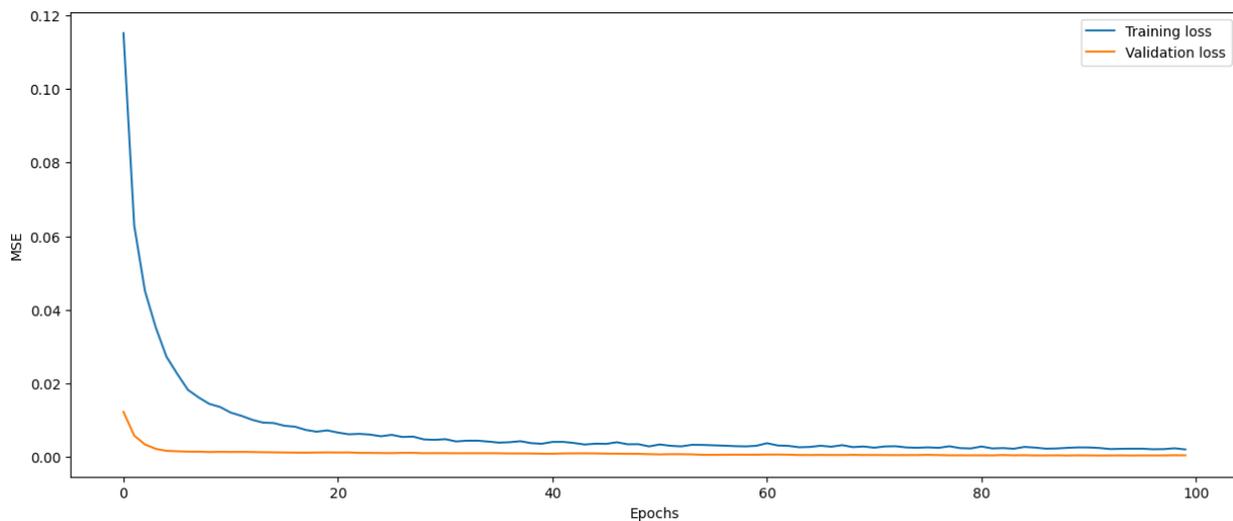
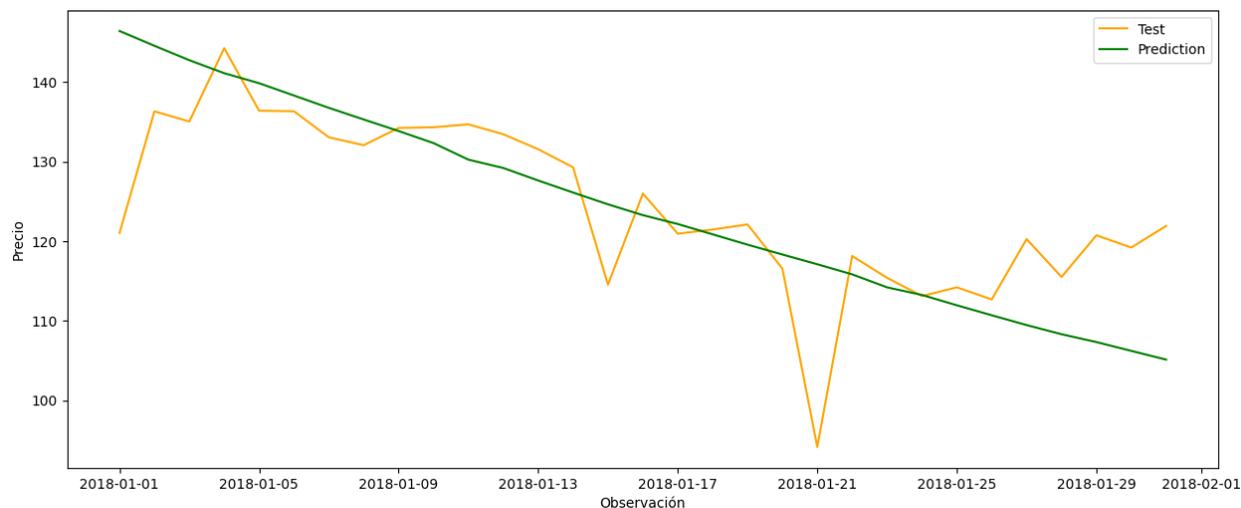


Figura 19. Desempeño vs. épocas Modelo baseline GRU



Por otra parte, en la **Figura 20** se observa la predicción con la primera iteración del modelo SARIMAX, sin considerar variables exógenas y con un valor limitado de los valores de sus componentes obteniendo un desempeño regular.

Figura 20. Predicción Modelo SARIMAX



5.2 Validación

Para predecir el Precio en Bolsa de la Energía en un horizonte de 30 días se dividieron los datos en conjuntos de entrenamiento, prueba y validación. Para el caso del modelo SARIMAX, se utilizó la serie del Precio de Bolsa y el Precio de Oferta como variable exógena, tomando el primer 80% de los datos para entrenamiento y el 20% restante para pruebas y validación, esto conservando el orden con el fin de no afectar su dependencia temporal. De igual manera se realizó para los modelos de Deep Learning con la diferencia que estos últimos, reciben una estructura tensorial (3 o más dimensiones).

5.3 Iteraciones y evolución

Tal como se mencionó en el capítulo **4.3 Modelos**, se realizaron iteraciones variando los parámetros allí mencionados. Se realizó el ajuste de hiperparámetros de los modelos utilizados a lo largo de los diferentes ensayos, esto basado en las métricas de desempeño presentadas en la sección **1.4 Métricas de desempeño**.

Para el modelo SARIMAX se variaron los parámetros realizando combinaciones de estos en un rango entre 0 y 6, presentando un mejor desempeño en el horizonte establecido el modelo con orden (2, 0, 1) y estacionalidad (1, 0, 1, 12) tal como se observa en la **Tabla 4**.

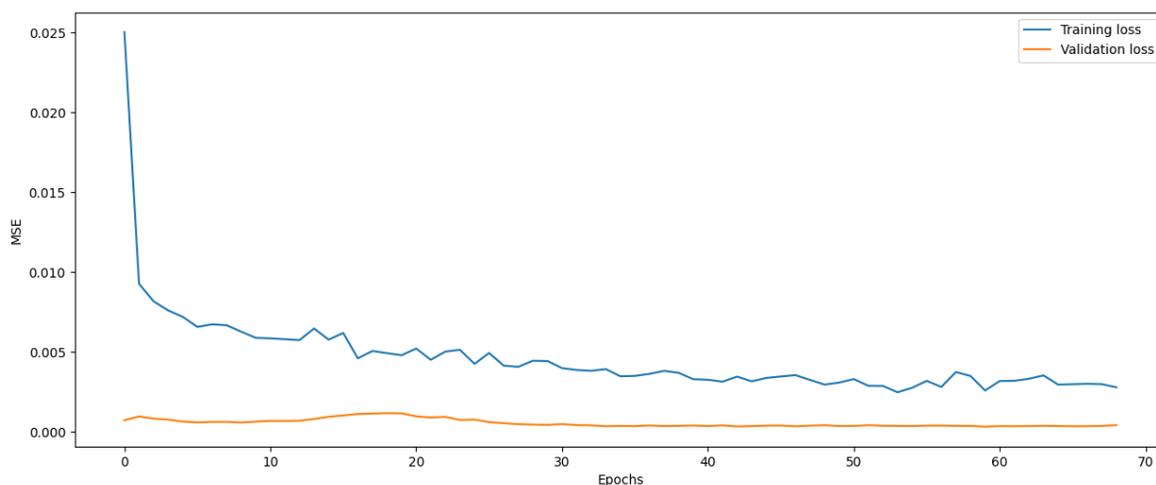
Tabla 4. Desempeño Modelo SARIMAX ajustado

Modelo	p	d	q	P	D	Q	s	MAE [\$COP/kWh]	MSE [\$COP/kWh]	RMSE [\$COP/kWh]	MAPE [%]
SARIMAX	2	-	1	1	-	1	12	5.5	70.107	8.37	4.64

Por otra parte, en la *Figura 21*, *Figura 22*, *Figura 23* y *Figura 24* se observan la evolución de la métrica de desempeño MSE para los modelos de Deep Learning mencionados en la sección *4.3 Modelos*.

De manera general tienen un buen desempeño durante el entrenamiento y en la predicción, destacando el modelo RNN el cual por su arquitectura es capaz de “recordar” los datos relevantes en distintos instantes de tiempo, en el corto y largo plazo. Si bien en los parámetros iniciales se estableció un número mayor de las épocas para el entrenamiento, debido al EarlyStopping no se observan en su totalidad, pues con esto se busca evitar un entrenamiento innecesario del modelo y se ahorra costo computacional.

Figura 21. Evolución del MSE en el entrenamiento modelo DNN



Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

Figura 22. *Evolución del MSE en el entrenamiento modelo RNN*

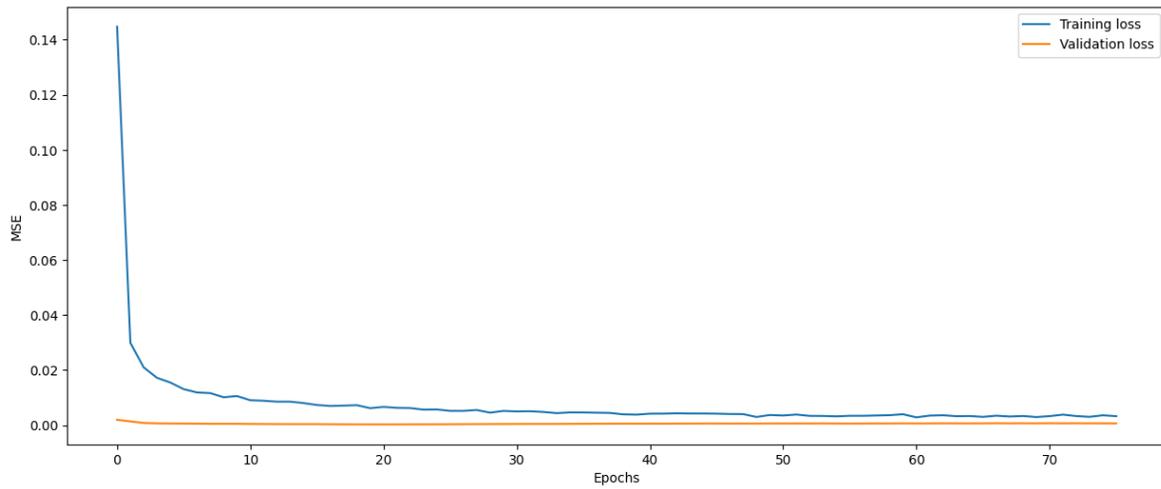


Figura 23. *Evolución del MSE en el entrenamiento modelo LSTM*

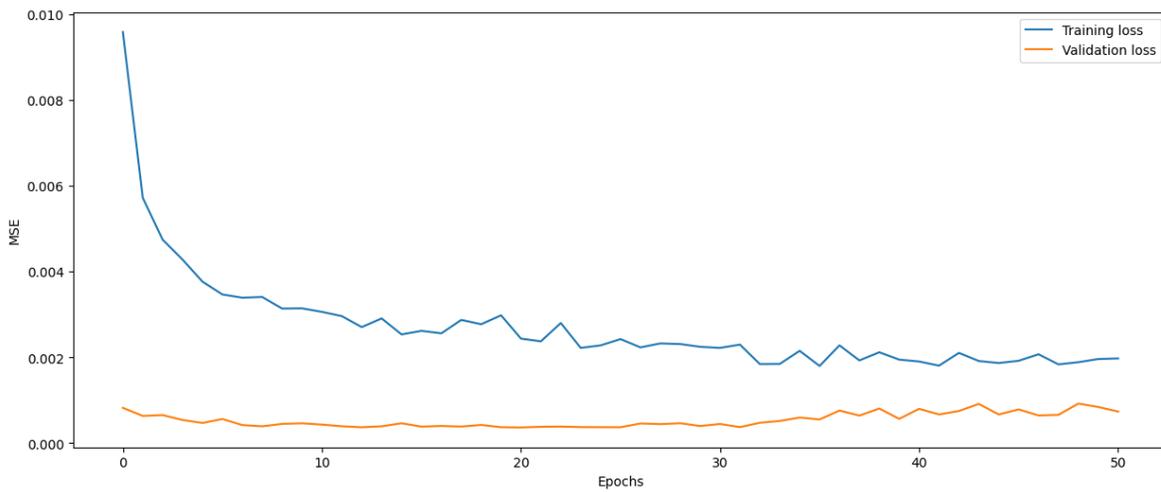
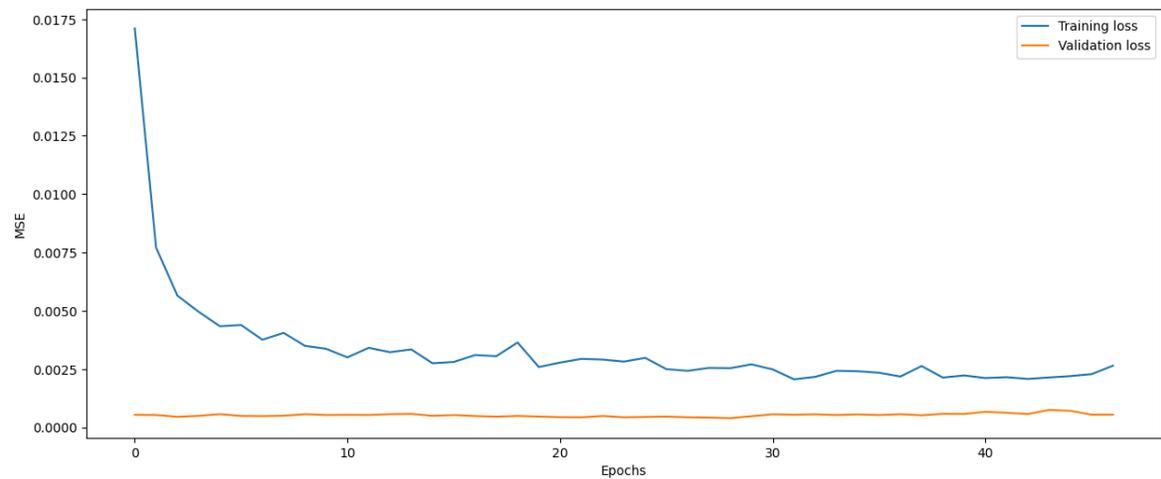


Figura 24. *Evolución del MSE en el entrenamiento modelo GRU*



5.4 Herramientas

Las herramientas utilizadas en el desarrollo de este trabajo fueron:

- Python como lenguaje de programación
- Google Colab para ejecución de Python
- Visual Studio Code como IDE
- Microsoft Excel para construcción de tablas y resultados
- Git y GitHub como herramientas de control de versiones
- Google Drive y OneDrive para trabajo colaborativo del documento y repositorio.

6. Resultados y discusión

La visualización de resultados se hizo a través de tablas dónde se puede observar las métricas de desempeño para las iteraciones realizadas en la implementación de los modelos, el tipo de cada modelo y los parámetros relevantes que se utilizaron para obtener dicho desempeño. Estas métricas son el: MAE [$\$/\text{COP/kWh}$], MSE [$\$/\text{COP/kWh}$], RMSE [$\$/\text{COP/kWh}$] Y MAPE [%]. Adicionalmente, se presentan gráficos con los resultados comparando los valores reales vs. los predichos.

A continuación, se muestran los resultados más relevantes obtenidos en las iteraciones realizadas y los resultados del mejor modelo con el ajuste de hiperparámetros.

6.1 Métricas

En la **Tabla 5** se muestra un top 10 de los resultados obtenidos. Como se mencionó previamente el modelo con mejor desempeño, dados los parámetros anteriores, es de tipo RNN con 30 neuronas, un retardo de 60 elementos, un tamaño de lote de 32 y entrenado en 100 épocas.

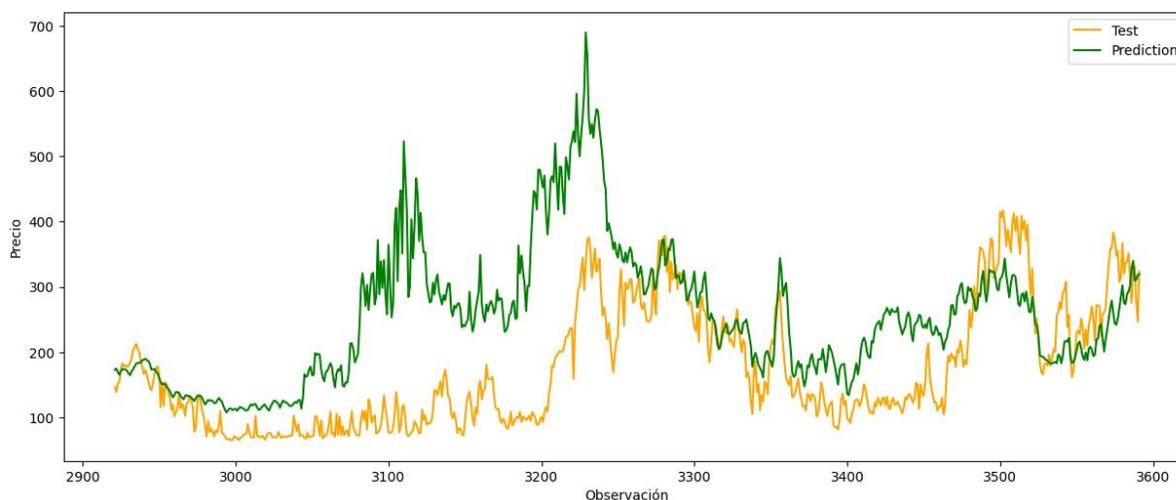
Tabla 5. Top 10 modelos Deep Learning

Modelo	Neuronas	LookBack	Epocas	BatchSize	MAE [\$COP/kWh]	MSE [\$COP/kWh]	RMSE [\$COP/kWh]	MAPE [%]
RNN	40	60	31	32	15.59	421.30	20.53	9.84
RNN	30	60	31	32	17.52	463.98	21.54	10.57
DNN	30	60	31	64	18.39	529.19	23.00	11.56
DNN	40	60	31	32	18.61	575.40	23.99	11.88
DNN	30	60	31	32	20.37	663.54	25.76	12.96
LSTM	30	60	31	32	20.47	816.55	28.58	12.89
DNN	40	60	31	64	24.52	974.09	31.21	15.10

6.2 Evaluación cualitativa

En la *Figura 25*, *Figura 26*, *Figura 27* y *Figura 28* se observan los resultados de la predicción del conjunto de prueba. Cabe aclarar que, aunque gráficamente se presenta todo el conjunto de prueba, las métricas de error se calcularon en el horizonte establecido (30 días). Por lo cual, es posible que, si se toma un horizonte mayor o menor, un modelo con diferentes características, se ajuste de mejor manera.

Figura 25. Predicción con mejor modelo DNN



Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

Figura 26. *Predicción con mejor modelo RNN*

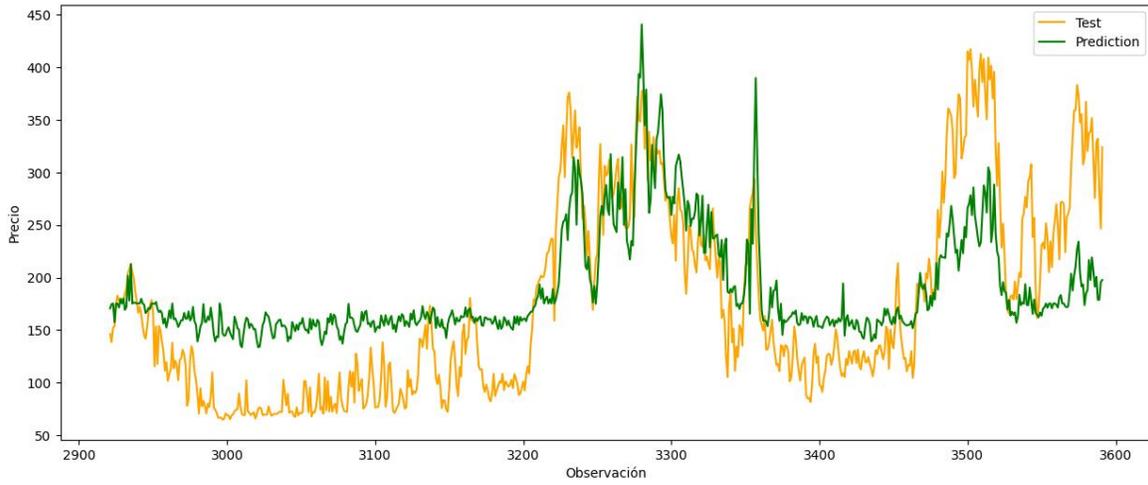


Figura 27. *Predicción con mejor modelo LSTM*

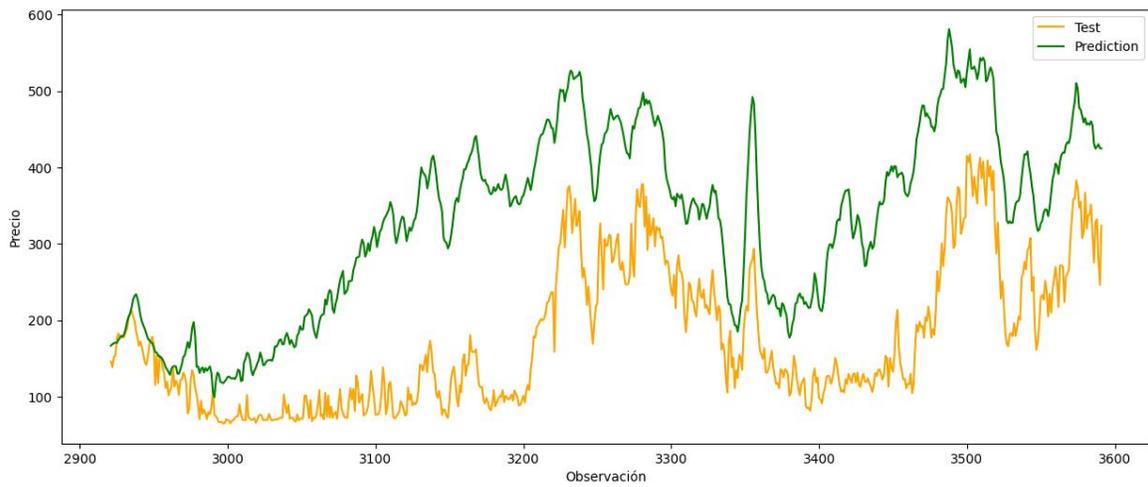
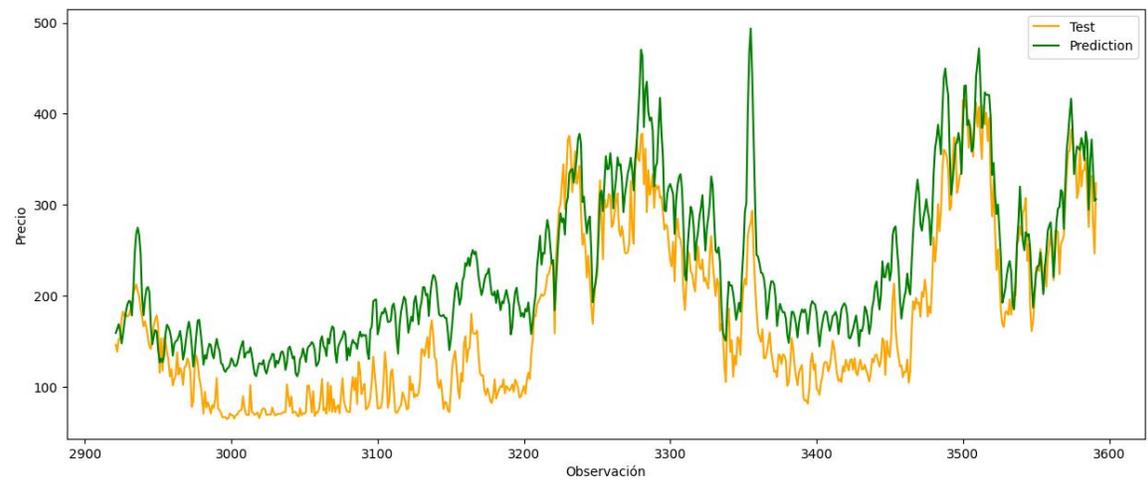


Figura 28. *Predicción con mejor modelo GRU*



Se observa que se manejó de manera correcta el sobreentrenamiento (OverFitting) utilizando capas de DropOut, si bien la predicción puede mejorar los modelos tienden a generalizar de buena manera el comportamiento de la serie, lo cual es deseado por la volatilidad que presenta la misma dependiendo de las variables de la operación del SIN y algunas otras variables del MEM.

6.2 Consideraciones de producción

Se recomienda para realizar la predicción del Precio en Bolsa de la Energía considerar, en primer lugar, una mayor cantidad de parámetros y dependiendo la necesidad o el horizonte que se desee predecir, establecer los parámetros adecuados. En este orden de ideas y con base en los resultados, el modelo RNN sería una excelente alternativa para la predicción captando la variabilidad y dependencia temporal de la serie en el corto y largo plazo; lo cual lo hace ideal para conjuntos de datos de gran dimensión. Adicionalmente, reentrenar periódicamente el modelo debido a cambios constantes en las variables consideradas para que el desempeño no se vea afectado.

7. Conclusiones

El precio en bolsa de la energía eléctrica en Colombia es una variable con alta volatilidad, sin embargo, con base en las pruebas realizadas también es estacionaria y presenta una alta autocorrelación; siendo una variable ideal para realizar su pronóstico teniendo en cuenta características propias y adicional, variables exógenas. Entre las variables con las que se correlaciona destacan el Precio de Oferta de los recursos y la Disponibilidad Declarada de los recursos hidráulicos y recursos térmicos. Esto tiene mucho sentido porque la generación de energía de los recursos térmicos tiende a ser más costosa debido al alto Costo de Suministro de Combustibles, afectando directamente la formación del Precio en Bolsa de la Energía.

El Mercado de Energía Mayorista Colombiano es un mercado con pocos participantes en la oferta de generación y muchos participantes en la demanda, por lo cual se hace necesario hacer un seguimiento constante de la regulación vigente con la finalidad de que sea un mercado competitivo y que fomente la participación de nuevos agentes, pequeñas empresas y el usuario final.

Finalmente, en la implementación de los modelos, adicional a la arquitectura y parámetros utilizados en los modelos es importante el horizonte a predecir, ya que como se observó gráficamente en la sección **6. Resultados y discusión** puede haber arquitecturas diferentes que se ajusten mejor si la predicción se evalúa en un horizonte mayor.

El modelo SARIMAX implementado, aunque es un modelo sencillo de implementar tiene un gran potencial y requiere de un amplio conocimiento estadístico y del componente teórico para realizar un correcto ajuste de hiperparámetros y mejorar su desempeño.

Los modelos de Deep Learning utilizados, de manera general, presentaron un buen desempeño, especialmente el modelo RNN debido a que su arquitectura se adapta bien a la dependencia temporal del precio de bolsa sin ser muy complejo su entrenamiento, obteniendo un error del 9.84% respecto a la serie original en un horizonte de 30 días. Cabe aclarar que es de gran importancia realizar un buen preprocesamiento de datos y análisis exploratorio para conocer la serie, y de ser necesario, aplicar técnicas estadísticas como transformación o diferenciación de esta para facilitar su análisis y pronóstico.

Referencias

- Alonso, J. C. (2010). *Tutorial para pruebas de raíces unitarias: Dickey-Fuller aumentado y Phillips-Perron en easyreg* (No. 9100). Universidad Icesi.
- Arana, C. (2021). *Redes Neuronales Recurrentes: Análisis de los Modelos Especializados en Datos Secuenciales*. CEMA Working Papers: Serie Documentos de Trabajo., Article 797. <https://ideas.repec.org/p/cem/doctra/797.html>.
- Asociación Nacional de Empresas Generadoras – ANDEG. (s. f.). Recuperado 17 de junio de 2024, de <https://www.andeg.org/>.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling* (arXiv:1412.3555). arXiv. <https://doi.org/10.48550/arXiv.1412.3555>.
- Climate Prediction Center—ONI. (s. f.). Recuperado 17 de junio de 2024, de https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.
- Contreras, D. (2016). *El impacto de El Niño en Colombia*. Revista Fasecolda, 163, Article 163.
- De Arce, R., & Mahía, R. (2003). Modelos Arima. Programa CITUS: Técnicas de Variables Financieras, 5-6.
- EquipoAnaliticaXM/API_XM (s. f.). Recuperado 17 de junio de 2024, de https://github.com/EquipoAnaliticaXM/API_XM.
- Gurrea, M. (2000). *Análisis de componentes principales*. Proyecto e-Math financiado por la Secretaría de Estado de Educación y Universidades (MECD).
- Jhana, N. (2020). *Multi-Variate Time Series Forecasting Tensorflow*. Kaggle. <https://www.kaggle.com/code/nicholasjhana/multi-variate-time-series-forecasting-tensorflow>.
- Muñoz-Santiago, A., Urquijo-Vanstrahlengs, J., Castro-Otero, A., & Lombana, J. (2017). *Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH*. Revista de Economía del Rosario, 20(1), Article 1. <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.6152>.
- Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. <http://neuralnetworksanddeeplearning.com>.

Predicción del precio en bolsa de la energía eléctrica en Colombia en un horizonte de mediano plazo

- Peralta, D., Herrera-Poyatos, A., & Herrera, F. (2016, septiembre 14). *Un Estudio sobre el Preprocesamiento para Redes Neuronales Profundas y Aplicación sobre Reconocimiento de Dígitos Manuscritos*.
- Regal, A., Morzán, J., Fabbri, C., Herrera Medina, G., Yaulli, G., Palomino, A., & Gil, C. (2019). *Proyección del precio de criptomonedas basado en Tweets empleando LSTM*. *Ingeniare. Revista chilena de ingeniería*, 27, 696-706. <https://doi.org/10.4067/S0718-33052019000400696>.
- Sak, H., Senior, A. W., & Beaufays, F. (2014). *Long short-term memory recurrent neural network architectures for large scale acoustic modeling*.
- Sede electrónica Comisión de Regulación de Energía y Gas (CREG). (s. f.). Recuperado 17 de junio de 2024, de <https://creg.gov.co/>.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-52452-8>.
- Steurer, M., & Hill, R. J. (2020). *Metrics for measuring the performance of machine learning prediction models: An application to the housing market*. Department of Economics. University of Graz.
- Torres Sánchez, A. S. (2023). *Predicción de la demanda de energía eléctrica usando modelos de inteligencia artificial para series temporales*. <https://bibliotecadigital.udea.edu.co/handle/10495/37564>.
- XM Administradores del mercado eléctrico. (s. f.). Recuperado 17 de junio de 2024, de <https://www.xm.com.co/>.
- Zapata, M. M. G., & Ochoa, C. M. (2008). *Modelación de la volatilidad de los precios de la energía eléctrica en Colombia*. *Revista Ingenierías Universidad de Medellín*, 7(12), Article 12.