



**UNIVERSIDAD
DE ANTIOQUIA**

**Modelo de predicción de precio de energía en el mercado no regulado
usando arquitecturas multimodales basadas en técnicas de Deep Learning**

Autor
Juan Sebastián Peláez Villa

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas
Medellín, Colombia
2021



Modelo de predicción de precio de energía en el mercado no regulado
usando arquitecturas multimodales basadas en técnicas de Deep
Learning

Juan Sebastián Peláez Villa

Trabajo de grado como requisito
para optar al título de
Ingeniero de Sistemas

Asesores:
Julián David Arias Londoño, Ph.D.
Santiago Alejandro Gallón Gómez, Ph.D.

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas
Medellín, Colombia
2021

Agradecimientos

Haber llegado hasta este punto de mi vida es posible gracias a la Universidad de Antioquia quien me extendió la mano en cada una de las diferentes ocasiones que lo solicité. Es por eso que tengo un profundo y grato agradecimiento por cada uno de los departamentos, entidades, servicios y personas que trabajan diariamente por y para los estudiantes.

De la misma manera agradezco a mis asesores Julián David Arias Londoño y Santiago Alejandro Gallón Gómez por la confianza, el acompañamiento, la paciencia, el respeto y la disposición que tuvieron conmigo durante cada etapa del desarrollo de mi proyecto de grados.

Igualmente, es muy importante para mi agradecer a cada uno de los docentes y administrativos del Departamento de Ingeniería de Sistemas quienes constantemente velan por hacer de los estudiantes excelentes profesionales. Personalmente, agradezco a los profesores Carlos Sierra, Diana Margot López, Áldrin Jaramillo, Diego Botia y Julián Arias por haber trascendido más allá de los cursos, me ayudaron a encontrar el tipo de profesional y persona que quiero ser.

De la misma manera, agradezco a mis compañeros y colegas de la Universidad y por supuesto mis amigos más cercanos, quienes han estado ahí permitiendo que lo vivido durante este viaje sea más que maravilloso.

Finalmente, para mi familia tengo mucho más que agradecimiento, nada de esto hubiera sido posible sin ustedes.

Muchas gracias a todos por ayudarme a ser lo que siempre he querido.

Título: Modelo de predicción de precio de energía en el mercado no regulado usando arquitecturas multimodales basadas en técnicas de Deep Learning

Resumen: Establecer el valor del precio de la energía eléctrica es un proceso que ha sufrido grandes cambios a través de los años en cada uno de los países del mundo, especialmente porque estos buscaban garantizar la transparencia en el proceso de definición de los precios en los sistemas eléctricos. El mayor logro de esta serie de cambios fue promover la inclusión de la energía eléctrica en los mercados financieros. Rápidamente los operadores se dieron cuenta de la gran volatilidad de este recurso debido a la gran cantidad de factores que influyen en él. Todos estos factores no están medidos en las mismas frecuencias, ni tampoco son registrados en los mismos medios, lo cual genera una dificultad al momento de ser interpretados. Es en este punto, cuando el interés por entender el comportamiento del precio de la energía y desarrollar modelos predictivos se vuelve una necesidad. El presente proyecto desarrolló arquitecturas multimodales basadas en técnicas de Deep Learning entrenadas con datos diarios asociados a las variables demanda energía SIN, caudal y precio oferta, y datos horarios correspondientes a las variables generación, demanda por comercializador y generación AGC.

Palabras Clave: Deep Learning, precio spot de energía eléctrica, arquitecturas multimodales, mercado no regulado.

Agradecimientos	II
Resumen	III
Lista de Imágenes	VI
Lista de Tablas	VII
1. Introducción	1
2. Objetivos	3
2.1. Objetivo General	3
2.2. Objetivos Específicos	3
3. Antecedentes	4
4. Marco Teórico	6
4.1. Pronósticos del precio de energía eléctrica	6
4.1.1. Pronósticos un día hacia adelante (one day-ahead forecasting)	8
4.2. Aprendizaje profundo	8
4.2.1. Redes neuronales recurrentes	9
4.2.2. Red Recurrente Simple	10
4.2.3. Long Short Term Memory (LSTM)	10
4.2.4. Gated Recurrent Unit (GRU)	12
4.3. Optimización de hiperparámetros	12
4.3.1. Optimización Bayesiana	13
4.4. Métricas de rendimiento	14
4.4.1. MAPE	14
4.4.2. sMAPE	14

5. Metodología	15
5.1. Sección 1: Análisis y procesamiento de datos	15
5.1.1. Etapa 1: Investigación	15
5.1.2. Etapa 2: Definición y selección de variables	16
5.1.3. Etapa 3: Recolección de datos	16
5.1.4. Etapa 4: Preprocesamiento y análisis	17
5.2. Sección 2: Experimentación	25
5.2.1. Etapa 5: Configuración y ejecución de experimentos	25
5.2.2. Etapa 6: Evaluación y selección	26
6. Resultados	27
7. Conclusiones	39
Bibliografía	44

Lista de Imágenes

4.1. Red neuronal profunda de 2 capas ocultas. [19]	9
4.2. Unidades fundamentales en redes recurrentes [2]	11
5.1. Estaciones meteorológicas	17
5.2. Variables climáticas	18
5.3. Generación hidráulica 2018	21
5.4. Generación hidráulica 2019	22
5.5. Generación hidráulica 2020	23
6.1. Arquitecturas diarias y horarias con mejores resultados para tipo de datos.	30
6.2. Arquitecturas con mejores resultados para tipo de datos en variables climáticas.	31
6.3. Arquitectura del mejor modelo completo con imágenes.	33
6.4. Pronósticos realizados por la mejor arquitectura multimodal que tiene en cuenta las variables climáticas.	34
6.5. Pronósticos realizados por la mejor arquitectura multimodal sin tener en cuenta las variables climáticas.	35
6.6. Pronósticos realizados por la mejor arquitectura multimodal sin tener en cuenta las variables climáticas agregando función de activación relu en las capas densas correspondientes a los submodelos horarios y diarios.	37
6.7. Pronosticos realizados por día sobre el conjunto de prueba.	38
7.1. Arquitectura Multimodal con mejor ajuste al comportamiento de los datos.	40

Lista de Tablas

5.1. Conjunto de variables seleccionadas y su respectiva periodicidad. . . .	16
5.2. Energia generada por trimestre por año.	20
5.3. Energia generada por trimestre por año.	23
5.4. Recursos Hídricos asociados a recursos de generación.	24
6.1. Arquitecturas para series diarias y horarias.	28
6.2. Mejores resultados en arquitecturas para series diarias y horarias. . .	29
6.3. Arquitecturas para variables climáticas.	31
6.4. Mejores resultados en arquitecturas para variables climáticas.	32
6.5. Resultados proceso de experimentación modelos completos con imágenes. .	32
6.6. Resultados entrenamiento y predicción sobre modelos completos con imágenes.	32
6.7. Resultados proceso de experimentación modelos completos sin imágenes. .	33
6.8. Resultados entrenamiento y predicción sobre modelos completos sin imágenes.	36
6.9. Resultados de experimentación sobre el conjunto de entrenamiento y prueba iterando las funciones de activación de las capas densas corres- pondientes a los submodelos horarios y diarios.	36

Determinar el precio de la energía es un proceso que ha sufrido grandes cambios a través de los años en cada uno de los países del mundo. Particularmente, en los países latinoamericanos dicha transformación se dió en la década de los 90 con un conjunto de reformas económicas que dieron paso a una nueva construcción del precio de este recurso. Entre las principales razones que influyeron para realizar esta estructuración se encuentran: obtener una mayor eficiencia en la interpretación y el manejo del recurso, fomentar la libre competencia en un sector tradicionalmente estatal, garantizar la transparencia en el proceso de definición de los precios y hacer uso de las nuevas tecnologías de manera que se puedan determinar de forma más exacta las interacciones en los sistemas eléctricos [5].

Para poder alcanzar dichos objetivos se han implementado diversas reformas estructurales, las cuales han logrado que la energía eléctrica sea considerada un commodity, el cual puede ser negociado bajo las condiciones impuestas en los mercados de valores. Poder competir bajo estas reglas permitió desarrollar un gran interés por entender la estructura dinámica del proceso de formación del precio de energía en mercados de corto plazo y a su vez obtener pronósticos confiables en diferentes ventanas de tiempo [12].

La participación de la energía eléctrica en los mercados financieros, permitió considerar este recurso como uno de los commodities con mayor variabilidad, 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 al contado (precio spot) de la energía puede ser de hasta el 50 % [28]. Esta variabilidad afecta directamente la planeación de los ingresos para los agentes y actores del sector eléctrico en Colombia, tales como generadores, despachadores, administradores de recursos, comercializadores, consumidores o grandes industrias.

La variabilidad que se presenta en el cálculo del valor del precio de la energía eléctrica se debe a la influencia de diversos factores externos a su propia generación,

algunos de ellos asociados a variables de mercado, otros a comportamientos climáticos e incluso eventos geopolíticos. Todos estos factores no están medidos en las mismas frecuencias, ni tampoco son observados de la misma manera, lo cual genera una dificultad al momento de ser interpretados por los métodos estadísticos convencionales, especialmente si la intención es construir un modelo con fines de pronóstico del precio.

Es en este punto, cuando el interés por entender el comportamiento del precio de la energía y desarrollar modelos predictivos se vuelve una necesidad, ya que es de vital importancia para los agentes económicos prever el comportamiento en el corto, mediano y largo plazo de este recurso.

Teniendo en cuenta todos los factores directos e indirectos asociados al cálculo del precio de la energía y la diversidad de fuentes de información, este trabajo aplica arquitecturas multimodales basadas en técnicas de aprendizaje profundo para pronosticar el precio spot de energía eléctrica de Colombia para el mercado no regulado.

La construcción del modelo de pronóstico fue producto de un conjunto de etapas, en las cuales inicialmente se definieron los tipos de datos que servirían de entrada al modelo, posteriormente, se precisaron las fuentes de información que contienen dichos datos. Siguiendo este proceso, la siguiente etapa delimitó los datos a un rango específico de fechas. Las últimas etapas estuvieron determinadas por la identificación y selección de arquitecturas para cada una de las frecuencias de datos, así como la exploración de sus respectivos hiperpámetros. Finalmente, con aquellos modelos que presentaban los mejores resultados, se construyó el modelo definitivo.

2.1. Objetivo General

Diseñar un modelo supervisado para pronósticar el precio en bolsa de la energía eléctrica de Colombia en el mercado no regulado, utilizando arquitecturas multimodales de aprendizaje profundo que permitan combinar múltiples tipos de información.

2.2. Objetivos Específicos

- Conocer la evolución, el contexto y el procedimiento por medio del cual la energía eléctrica es transada en el mercado de energía colombiano.
- Diseñar una arquitectura de Deep Learning para la predicción del precio de la energía, capaz de combinar la información de diferente tipo por medio de una estrategia multimodal.
- Realizar un proceso de validación que permita determinar la mejor configuración de parámetros para la cual el modelo de aprendizaje automático propuesto provee los mejores resultados.

El pronóstico del precio de energía eléctrica en el mercado de energía es una tarea que normalmente ha sido tratada con modelos estadísticos de series de tiempo. Teniendo en cuenta la naturaleza del comportamiento del precio spot de la energía, en la literatura se han propuesto diversos métodos estadísticos para el modelamiento y pronóstico del mismo con diferentes objetivos, horizontes de tiempo y en algunas ocasiones incluyen o no variables exógenas fundamentales como precipitación de la superficie de la tierra, caudal de los ríos, precipitación media, entre otros.

Los estudios realizados muestran una amplia gama de métodos estadísticos entre los que están los filtros de suavizamiento exponencial, modelos autorregresivos (AR), modelos autorregresivos de medias móviles (ARMA), modelos autorregresivos de medias móviles integrados (ARIMA) y de larga memoria (ARFIMA), modelos ARIMA estacionales, modelos autorregresivos por umbrales (TAR), modelos de la clase de heterocedasticidad condicional autorregresivos generalizados (GARCH), modelos de parámetros tiempo-variantes, modelos de series de tiempo estructurales, modelos con preprocesamiento de spikes, modelos de difusión con saltos con reversión a la media y modelos de cambios de régimen Markovianos [26] [10] [30] [11] [9] [33] [32], entre otros.

Por otro lado, la disposición de grandes cantidades de información y el gran avance en términos de computación y tratamiento de datos ha permitido usar soluciones asociadas a algoritmos de aprendizaje de máquina como lo pueden ser las redes neuronales artificiales (ANN), máquinas de vectores de soporte (SVM), bosques aleatorios [31] [27] [13], entre otros.

Una implementación particular de estos modelos en el entorno colombiano, es el propuesto por Lira, et al [22] quienes pronostican el precio diario de la energía con modelos ARMAX, PARMAX, ARX y redes neuronales utilizando los precios de la electricidad por hora, la demanda de carga por hora, el precio diario del petróleo, los niveles diarios de los embalses de agua, la contribución fluvial diaria y el precio semestral del gas. Posteriormente, se identifica en el conjunto de datos que la de-

manda del precio de la energía se distribuye en tres grandes grupos: precios bajos, medios y altos; cada uno de ellos determinado por un rango específico de fechas. Una vez identificados estos grupos, el proceso de evaluación tomó cada uno de ellos y los evaluó en los modelos anteriormente mencionados, con el objetivo de predecir los precios en los siguientes días. Los autores reportaron que los modelos ARX presentaron los mejores resultados para los precios medios y para los precios altos respectivamente.

Otro caso aplicado a la predicción del precio de la energía en el mercado colombiano es el desarrollado por Barrientos, et al. [3] quienes hacen uso de modelos de vectores de corrección de errores y redes neuronales. En este trabajo se encontró que el IPC, la tasa de cambio real y el precio internacional de combustibles no afectan el precio de la energía eléctrica en el largo plazo, y por el contrario, el conjunto de variables que tienen el efecto más importante sobre el precio del kilovatio/hora de energía son el PIB, la demanda de energía eléctrica y el nivel de los embalses. Los resultados hallados en este estudio demostraron que el modelo econométrico demuestra una notoria relación inversa entre el precio de la energía y el nivel de los embalses. Los autores concluyeron que, teniendo en cuenta el valor del PIB, el nivel de los embalses y el crecimiento de la demanda, el precio en bolsa en el largo plazo tiende al precio de los contratos de futuros.

Adicionalmente, Bello and Beltrán [4] modelan y pronostican el precio diario de la energía utilizando modelos de volatilidad condicional tipo GARCH concluyendo que es relevante tener en cuenta las variables que son externas al valor de los precios, dado que el mercado está expuesto a épocas de sequía; múltiples épocas de lluvia; terrorismo; regulación estatal; consumo de la población e imposibilidad de almacenamiento, entre otros.

Finalmente, en el estudio desarrollado por Lago et al. [18] el cual no fue desarrollado propiamente para el mercado colombiano, se hace uso de un conjunto de modelos estadísticos y de aprendizaje profundo con el objetivo de comparar su capacidad de pronóstico. Particularmente se seleccionaron cinco estrategias, Deep Neural Network [21]; Recurrent Neural Network [14]; Long Short Term Memory [17]; Gated Recurrent Unit [8] y Convolutional Neural Network [20]. Los resultados de este estudio comparativo mostraron que los modelos DNN, LSTM y GRU demuestran un notable incremento en los resultados de predicción, además, se observó una clara división entre los modelos de aprendizaje profundo y los métodos estadísticos, donde estos primeros presentaron resultados estadísticamente más significativos.

4.1. Pronósticos del precio de energía eléctrica

Teniendo en cuenta la evolución que ha presentado en las últimas décadas el mercado energético, tanto en temas políticos como comerciales, se han desarrollado nuevos retos para los actores que intervienen en él. La excesiva volatilidad de los precios, que al compararla con otros recursos puede llegar a ser de hasta dos órdenes de magnitud, ha obligado tanto a productores, como consumidores y mayoristas a blindarse contra los grandes movimientos de los precios. Por tanto, la predicción del valor de la energía se ha convertido en una tarea fundamental para la toma de decisiones y el desarrollo de estrategias en las empresas del sector energético[32].

La industria ha entendido que impulsar la investigación en la construcción de modelos que generen pronósticos de los precios de la electricidad ha sido el factor con mayor impacto en la interpretación del valor de dicho recurso. La mayor cantidad de soluciones que ha propuesto este desarrollo investigativo están relacionadas en términos de duración del horizonte de tiempos. Por tanto, es normal encontrar métodos orientados al corto plazo (Short Term Price Forecasting - STPF), medio plazo (Medium Term Price Forecasting - MTPF) y métodos de previsión de precios a largo plazo (Long Term Price Forecasting - LTPF).

Cada uno de estos métodos tiene una aplicación particular para cada uno de los contextos del mercado. Por ejemplo, el principal objetivo del LTPF es el análisis y la planificación de la rentabilidad de la inversión; el MTPF suele ser usado para los cálculos del balance, la gestión de riesgos y la fijación de precios de los derivados; por último, el STPF es de vital importancia en las operaciones diarias dado que el mercado al contado suele consistir en 24 subastas horarias que tienen lugar simultáneamente con un día de antelación.

Tal como se puede evidenciar en [18] y [32] se han desarrollado un conjunto de

enfoques para analizar y predecir el comportamiento de los mercados de energía y los precios de la electricidad. Cada una de estas aproximaciones se definen como:

- **Basados en costos de producción:** Estos modelos tienen la capacidad de prever los precios hora a hora, sin embargo, ignoran las prácticas estratégicas de puja asociadas a subastas de precios. Eran apropiados para los mercados regulados con poca incertidumbre en los precios y una estructura estable pero no para los mercados competitivos como los que se ven hoy en día.
- **Basados en teoría de juegos:** Normalmente son tomados como una evolución de los modelos basados en costos de producción ya que si tienen en cuenta las estrategias de pujas que carecen en los anteriores. Estos modelos son especialmente útiles para predecir los precios esperados en mercados sin historial de precios si y solo si se conocen los costes de abastecimiento y concentración de mercado.
- **Basados en modelos fundamentales:** Describen el comportamiento de los precios teniendo en cuenta el impacto de todos aquellos factores físicos y económicos que influyen en el precio de la electricidad. Se definen asociaciones entre cada una de las partes involucradas y posteriormente se modelan los comportamientos, normalmente haciendo uso de técnicas estadísticas, econométricas o no paramétricas.
- **Basados en modelos cuantitativos:** Caracterizan las propiedades estadísticas del precio a través del tiempo. Estos modelos han introducido dos conceptos fundamentales, la prima de riesgo y la distinción entre el pronóstico del precio al contado y el precio a plazo. Su mayor objetivo es evaluar los derivados que influyen en el proceso.
- **Basados en enfoques estadísticos:** Hacen uso de todas aquellas técnicas estadísticas para pronósticos de series de tiempo, en general aquellas que se implementan en los modelos econométricos. Normalmente son comparados con el análisis técnico de los mercados financieros ya que no intentan medir el valor intrínseco o fundamental de un activo, en su lugar, identifican patrones e indicadores en los gráficos de precios que determinarán el rendimiento futuro de un activo.
- **Basados en inteligencia artificial:** Son capaces de modelar el comportamiento de los precios a través de técnicas de machine learning o incluso aprendizaje profundo. Sus mayores características es que estos modelos tienden a ser flexibles y pueden manejar la complejidad y la no linealidad de los datos.

El enfoque basado en técnicas de inteligencia artificial es el principal objeto de estudio en este proyecto, por lo tanto se desarrolló con más profundidad en los demás capítulos del texto.

4.1.1. Pronósticos un día hacia adelante (one day-ahead forecasting)

Para realizar el pronóstico del precio de energía existe un marco de trabajo especializado conocido como *one day-ahead forecasting* en el cual se supone que los pronósticos de precios para el día d se requieren antes de la fecha límite del día $d - 1$.

Aplicar esta metodología cobra sentido ya que dadas las condiciones del mercado, cada agente tiene que presentar 24 ofertas correspondientes a las 24 horas del día. Una vez transcurrido el plazo máximo, el operador del mercado utiliza las ofertas presentadas para calcular el precio oficial en cada una de las correspondientes horas. Posteriormente, a los agentes del mercado se les realiza una asignación de energía que depende del precio oficial del mercado y de las ofertas presentadas por sí mismo.

El uso de este marco de trabajo se puede evidenciar en estudios comparativos de modelos estadísticos contra modelos de inteligencia artificial [18] o incluso en trabajos que hacen uso de funciones de transferencia para modelar el comportamiento del precio de la energía [24].

4.2. Aprendizaje profundo

Tal como se expresa en [14] el aprendizaje profundo es una rama del aprendizaje automático que permite a los sistemas de computación aprender de la experiencia recolectada en el descubrimiento de patrones en los datos por medio de una estructurada jerarquía de conceptos. Teniendo en cuenta que la máquina reconoce los patrones de manera automática no es necesario que una persona especifique explícitamente aquello que debe ser aprendido. Es aquí donde se hace necesario tener una estructura bien definida, porque a partir de construcciones simples es que se pueden identificar los patrones más complicados.

En las últimas décadas el aprendizaje profundo ha venido presentando grandes avances e innovaciones de la mano de las redes neuronales artificiales, tanto así que el término profundo se ha podido asociar al hecho de que ahora se puedan probar diferentes configuraciones de redes neuronales cuya cantidad de capas ocultas no se limite a una sola como lo hace el modelo tradicional de perceptrón multicapa y es gracias a este cambio de paradigma que se ha podido demostrar mejores capacidades de generalización [19].

Teniendo en cuenta esto, una red neuronal profunda como la representada en la imagen 4.1 está definida por $X = [x_1, \dots, x_n]^T \in \mathbb{R}^n$ como entradas de la red, $Y = [y_1, y_2, \dots, y_m]^T \in \mathbb{R}^m$ como salidas de la red, n_k como el número de neuronas en la k -ésima capa oculta y $z_k = [z_{k1}, \dots, z_{kn_k}]^T$ como el vector de estados correspondiente. En este modelo los parámetros están representados por el conjunto de pesos W que establecen los vínculos de conexión entre cada una de las diferentes neuronas de la red [19].

Para realizar la estimación de los parámetros W del modelo se desarrolla un

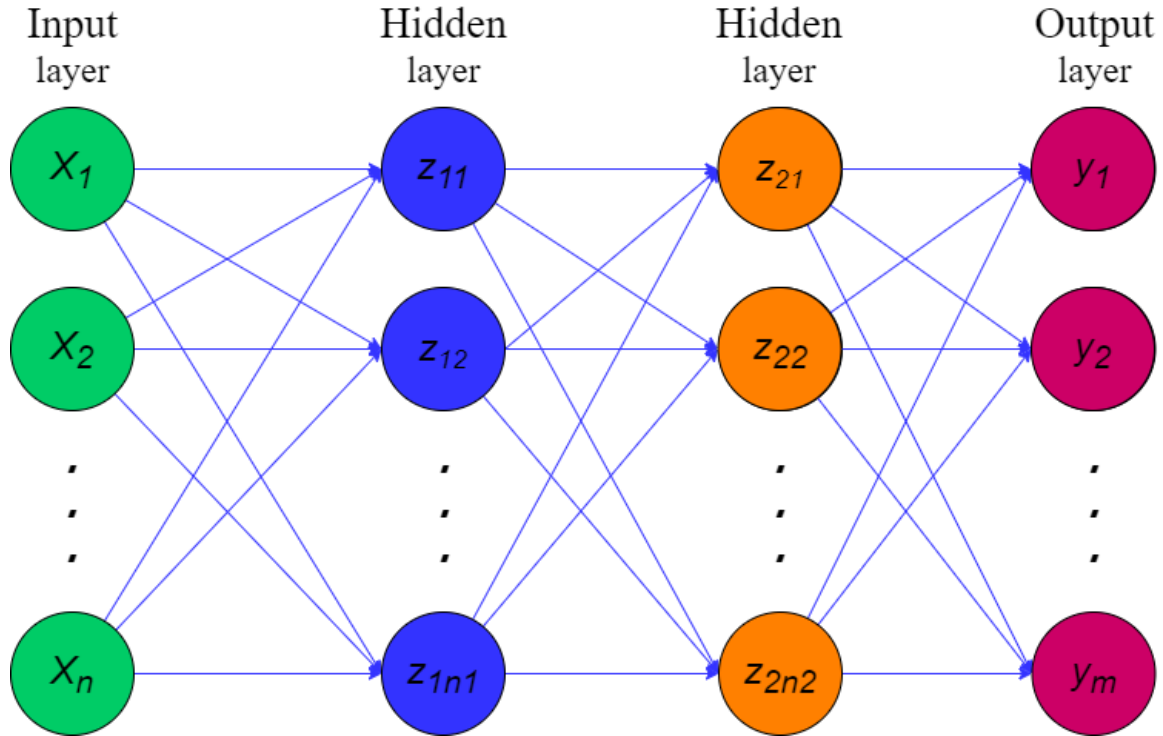


Figura 4.1: Red neuronal profunda de 2 capas ocultas. [19]

proceso conocido como entrenamiento de la red, el cual dado un conjunto de datos $\varphi = \{(X_i, Y_i)\}_{i=1}^N$ con N como la cantidad de registros, se realiza resolviendo el problema de optimización

$$\min_W \sum_{i=1}^N g_i(Y_i, F(X_i, W)) \quad (4.1)$$

donde F es la función que define la red, la cual realiza una transformación $\mathbb{R}^n \rightarrow \mathbb{R}^m$, g_k es una función de costos específica al problema de regresión como lo pueden ser el *error cuadrático medio* o *error medio absoluto*. Tradicionalmente, en los modelos de perceptrón multicapa, el método de *gradiente descendente* es el más usado para resolver problemas de optimización como el de la ecuación 4.1, a pesar de que este también es usado en redes profundas, es más común encontrar diferentes versiones de este mismo método, siendo el *gradiente descendente estocástico* la que más resalta.

4.2.1. Redes neuronales recurrentes

Desarrolladas en la década de los 80s, las redes neuronales recurrentes son catalogadas como uno de los modelos más utilizados en aprendizaje profundo. Este tipo de redes se les asocia la capacidad de tener memoria, puesto que pueden recordar información que ha pasado a través de ellas y por medio de la cual son capaces de predecir la siguiente salida.

Este comportamiento sugiere que las redes recurrentes tengan interiorizados un conjunto de ciclos repetitivos alojados en las capas ocultas que permitan de esta manera guardar el estado de la información por un tiempo determinado hasta que la red sea capaz de predecir los valores futuros. El procedimiento se lleva a cabo teniendo en cuenta que la salida de la capa oculta se retransmite t veces a sí misma generando un proceso recursivo el cual se termina cuando se completa el número de iteraciones establecidos y se pase a la siguiente capa oculta. Finalmente, para acabar con el proceso, los pesos de la red se actualizan teniendo en cuenta los errores generados durante cada iteración.

4.2.2. Red Recurrente Simple

Este tipo de arquitectura es en principio un conjunto de redes neuronales tradicionales acopladas entre sí, cada una de las cuales transmite la información hacia la otra. El tipo de memoria que desarrollan estas redes se conoce como memoria a corto plazo y normalmente no son capaces de interpretar el comportamiento de series de tiempo a largo plazo.

Una red recurrente simple presenta para sus unidades fundamentales la estructura dada por las siguientes ecuaciones 4.2

$$\begin{aligned} h_t &= \sigma(W_i h_{t-1} + V_i x_t + b_i) \\ z_t &= \tanh(W_o h_t + b_o) \end{aligned} \tag{4.2}$$

donde h_t representa el estado oculto de la unidad. x_t y z_t simbolizan la entrada y la salida de la celda en el tiempo t respectivamente. W_i y V_i denotan los pesos de la red, mientras que b_i representa los términos independientes. De la misma manera, W_o y b_o son las matrices de pesos y término independiente de la salida de la unidad respectivamente [16]. En la imagen 4.2c se puede visualizar la estructura de una red recurrente simple.

4.2.3. Long Short Term Memory (LSTM)

Es un modelo de red neuronal que hace uso de otro tipo de unidades diferentes a las usadas en las arquitecturas basadas en perceptrón multicapa. Adicionalmente, este modelo realiza una combinación especial de estas nuevas unidades, así como también productos y sumas entre ellas, las cuales permiten implementar compuertas de control que administran la memoria.

Contrario al modelo tradicional, la característica más importante de las redes LSTM es su capacidad para aprender la dependencia a largo plazo. Esto es posible gracias al uso de las tres compuertas definidas: *input*, *forget* y *output*. La inclusión de la compuerta *forget* en esta estructura está definida con el objetivo de indicar qué tanta información tendrá en cuenta la memoria anterior y qué tanta información ya ha olvidado.

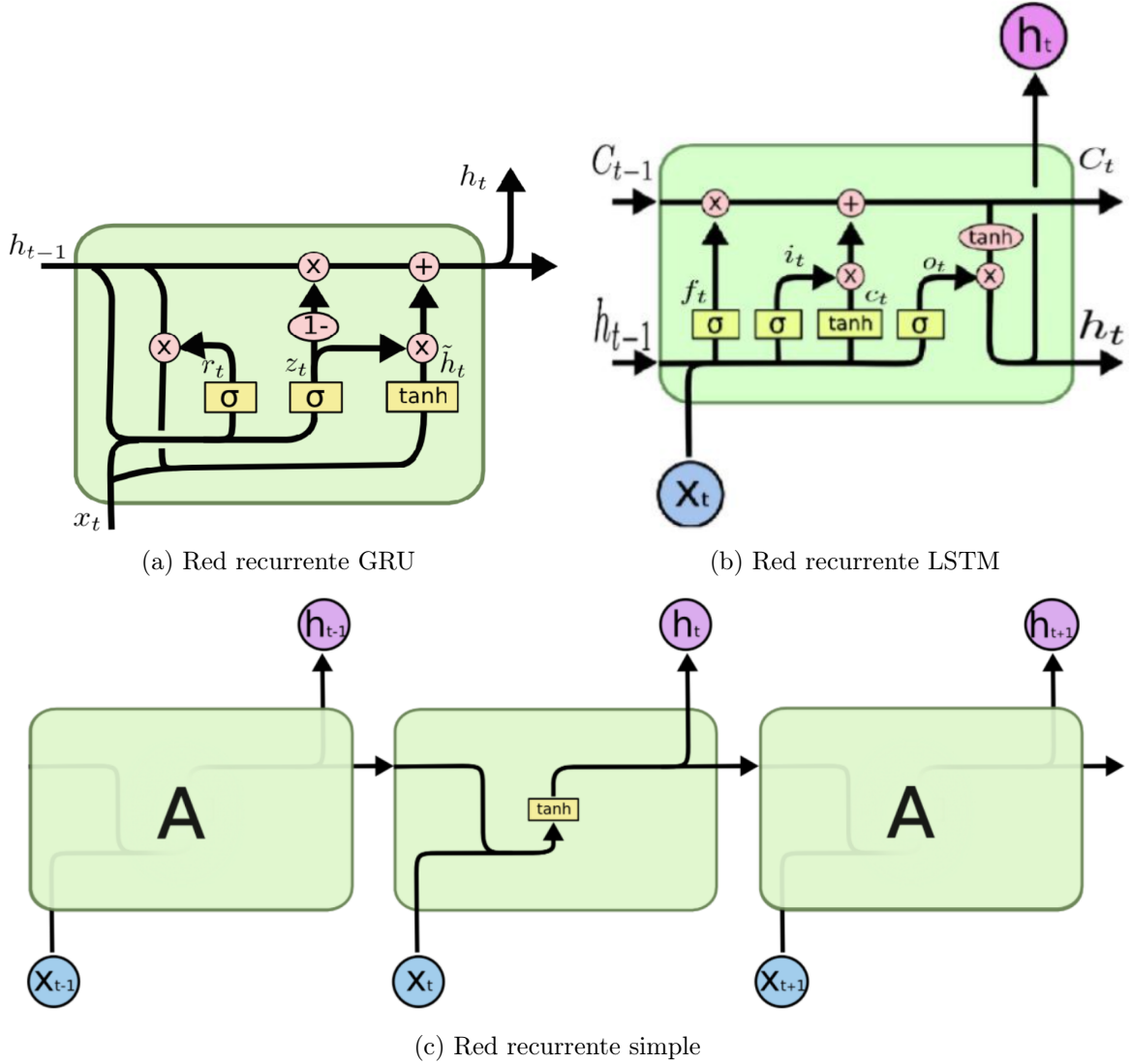


Figura 4.2: Unidades fundamentales en redes recurrentes [2]

En este modelo está definida la salida h_t y C_t las cuales son calculadas bajo el siguiente conjunto de fórmulas (4.3):

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \quad (4.3a)$$

$$f_t = \sigma(W_f X_t + U_f h_{t-1} + b_f) \quad (4.3b)$$

$$O_t = \sigma(W_o X_t + U_o h_{t-1} + b_o) \quad (4.3c)$$

$$\tilde{C}_t = \tanh(W_c X_t + U_c h_{t-1} + b_c) \quad (4.3d)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4.3e)$$

$$h_t = \tanh(C_t) * O_t \quad (4.3f)$$

$$(4.3g)$$

donde i_t , f_t y O_t representan las compuertas *input*, *forget* y *output* en el tiempo t , respectivamente. W_i , W_f , W_o , y W_c son los pesos que asignan la entrada de la capa

oculta a las tres compuertas, mientras que los pesos U_i , U_f , U_o , y U_c asignan la salida de la capa oculta a las mismas compuertas; b_i , b_f , b_o , y b_c son los términos independientes de cada componente. Finalmente, C_t y h_t son los resultados de la célula y la salida de la capa respectivamente [2]. Un ejemplo de esta unidad recurrente se puede visualizar en la imagen 4.2b.

4.2.4. Gated Recurrent Unit (GRU)

Esta arquitectura es una variante de la red recurrente LSTM, donde su mayor diferencia radica en que el modelo GRU mezcla las compuertas *input* y *forget* para definir la compuerta *update*. Este cambio genera que la arquitectura presente muchos menos parámetros y por tanto sea mucho más sencilla de entrenar.

Para el caso particular de la red GRU, la salida del estado h_t se define por el siguiente conjunto de ecuaciones (4.4):

$$z_t = \sigma(W_z X_t + U_z h_{t-1} + b_z) \quad (4.4a)$$

$$r_t = \sigma(W_r X_t + U_r h_{t-1} + b_r) \quad (4.4b)$$

$$\tilde{h}_t = \tanh(W_h X_t + (r_t * h_{t-1})U_h) \quad (4.4c)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4.4d)$$

$$(4.4e)$$

donde r es la compuerta *reset* que indica de qué manera se puede combinar la nueva entrada con la memoria anterior. z es la compuerta *update* la cual propone que porción de los datos almacenados en memoria se mantendrán; si esta compuerta obtiene un valor de 1 entonces la memoria anterior se conserva por completo, pero si el valor es 0, la memoria anterior se olvidará.

Este conjunto de compuertas permite definir otra diferencia con el modelo LSTM, pues mientras la compuerta *forget* en esta red determina automáticamente la cantidad de memoria anterior que se mantiene, en el modelo GRU, toda la memoria previa se mantiene o se olvida por completo [2]. La estructura definida para la unidad GRU se puede visualizar en la imagen 4.2a.

4.3. Optimización de hiperparámetros

Cada uno de los modelos propuestos en el aprendizaje de máquina y por supuesto sus derivados en aprendizaje profundo cuentan con un conjunto de elementos que modifican su comportamiento, generando que el modelo obtenga mejores o peores resultados; para el caso de las redes neuronales se puede mencionar el número de capas ocultas, número de neuronas por capa, tipos de función de activación por neurona, entre otros. Estos elementos se conocen como hiperparámetros y se hace necesario definir un proceso que permite encontrar la combinación de valores más óptimos para que el modelo alcance los mejores resultados.

Realizar este proceso de selección tiene diversas aproximaciones. Una de ellas es la selección manual la cual es viable cuando se desean validar pocos hiperparámetros y en un rango muy corto ya que el proceso consume mucho tiempo en horas-hombre y no garantiza ni siquiera alcanzar un mínimo local. Por otro lado, existe la búsqueda en malla la cual disminuye el tiempo en horas-hombre pero aumenta el tiempo de cómputo y exploración, puesto que dado un conjunto de hiperparámetros el proceso analiza una por una el total de combinaciones. También existe la búsqueda aleatoria la cual toma en cuenta el conjunto de hiperparámetros definidos en el método anterior y explora al azar un subconjunto de esta, lo cual reduce el tiempo de cómputo, pero no tiene en cuenta buenos resultados previos para seguir realizando la exploración en ese mismo sector [6], caso contrario al propuesto por el método de optimización bayesiana, el cual intenta buscar la mejor combinación de hiperparámetros basándose en los mejores resultados obtenidos por experimentos anteriores.

4.3.1. Optimización Bayesiana

Recientemente la optimización Bayesiana ha ido surgiendo como un método de selección de hiperparámetros muy efectivo ya que es capaz de optimizar funciones objetivo de mucha complejidad y con alto coste de evaluación. Este método está desarrollado teniendo en cuenta la inferencia Bayesiana y los procesos Gaussianos, teniendo mayor impacto en expresiones cerradas donde la función objetivo es desconocida.

El proceso de optimización Bayesiana crea un modelo probabilístico para la función objetivo, usando los resultados de experimentos pasados como datos de entrenamiento. Este modelo es mucho menos costoso al momento de calcular y además tiene la característica de entregar información sobre dónde es que se debería realizar la exploración para obtener mejores resultados.

Si se define $P = \{p_1, \dots, p_m\}$ como el conjunto m de hiperparámetros a explorar. Entonces, dado el conjunto de datos de entrenamiento $\{(x_i, y_i)\}_{i=1}^n$ el proceso tiene como objetivo encontrar $P^* = \arg \min_p g(P | \{(x_i, y_i)\}_{i=1}^n)$ donde g es una función de costos cualquiera [29].

Todo el proceso de optimización está dirigido por una función de adquisición, la cual es capaz de definir siguiente conjunto de hiperparámetros a evaluar. En general una función de adquisición debe ser capaz de mantener el equilibrio entre las etapas de exploración y la explotación, donde exploración hace referencia a la búsqueda de regiones en las que la incertidumbre es alta, con el objetivo de encontrar un nuevo conjunto de hiperparámetros que mejoren el rendimiento del modelo y por otro lado explotación se refiere a hacer uso de una región cercana y ya explorada con buenos resultados [29].

4.4. Métricas de rendimiento

Con el objetivo de validar el ajuste de los modelos una tarea fundamental es definir medidas que permitan contrastar los resultados en las diferentes etapas, tanto en la exploración de hiperparámetros como en las fases de entrenamiento e inferencia.

4.4.1. MAPE

Esta métrica está definida como el error porcentual medio absoluto (MAPE) lo que significa que expresa los valores de los errores como un porcentaje de los datos. Teniendo en cuenta que esta métrica es una medida de escala independiente, la hace muy adecuada para comparar diferentes series de tiempo.

La función que define esta métrica está dada por

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{A_i - F_i}{A_i} \right| \quad (4.5)$$

donde A_i corresponde al valor actual para el tiempo i y F_i representa el valor pronosticado para el mismo tiempo i .

A pesar de ser una medida ampliamente usada en todos los ambientes, presenta una serie de dificultades ([23]) entre las cuales la que más resalta es su dificultad por interpretar valores atípicos, ya que cuando el valor del error de porcentaje absoluto (APE) es en extremo pequeño o en extremo grande, el resultado final carece de sentido.

4.4.2. sMAPE

Se conoce como la versión simétrica del MAPE ya que trata de corregir sus mayores dificultades dividiendo el error del pronóstico por el promedio de A_t y F_t [15]. Teniendo en cuenta esta propuesta, no importa si A_t presenta un valor pequeño y F_t presenta uno grande o viceversa. Por lo tanto, el rango de valores que sMAPE puede aceptar va desde 0% a 200%.

El sMAPE está definido por la siguiente ecuación

$$MAPE = \frac{2}{N} \sum_{i=1}^N \left| \frac{A_i - F_i}{A_i + F_i} \right| * 100 \quad (4.6)$$

donde A_i corresponde al valor actual para el tiempo i y F_i representa el valor pronosticado para el mismo tiempo i .

El sMAPE es una métrica que presenta un par de desventajas: inicialmente está el hecho de que es un valor menos intuitivo que el MAPE, además que acarrea el mismo problema de la posible división por cero cuando el valor pronosticado y el valor real sean cero.

Este trabajo se desarrolló teniendo en cuenta un conjunto de 7 etapas divididas entre 2 secciones. La primera sección detalla el proceso de investigación del proyecto, definición y selección de variables, recolección, preprocesamiento y análisis de la información. La segunda sección estuvo enfocada en encontrar el modelo que mejor se ajustara a los datos, para ello fue necesario desarrollar etapas de experimentación, selección y validación.

La metodología presentada propuso un desarrollo lineal acumulativo entre cada una de las secciones y a su vez de las etapas, debido a que los insumos necesarios para implementar una etapa, dependían de lo que se hizo en etapas previas. Sin embargo, una vez se avanzó en las etapas, era posible retornar a cualquiera de las anteriores, con el objetivo de complementar y/o recordar conceptos e información relevante.

5.1. Sección 1: Análisis y procesamiento de datos

5.1.1. Etapa 1: Investigación

Esta etapa se destinó a estudiar los conceptos relacionados con el mercado de valores, especialmente en el sector energético; identificar cuáles son los actores involucrados; analizar el ciclo de vida de la energía eléctrica; comprender el papel que cumplen los factores externos que alteran y modifican el valor del kilovatio/hora; observar la dinámica entre oferta y demanda, y por último, investigar los modelos estadísticos de series de tiempo y de aprendizaje automático que se han desarrollado para realizar pronósticos del precio spot.

5.1.2. Etapa 2: Definición y selección de variables

En esta etapa se definió y estableció que el precio en bolsa de valores de Colombia del kilovatio/hora de energía eléctrica sería la variable objetivo. Adicionalmente, se estudió y determinó las variables predictoras, tanto aquellas que son exógenas como endógenas. En la Tabla 5.1 se muestra el conjunto de variables seleccionadas, así como el rango de fechas disponibles para consulta y la respectiva periodicidad de su información.

CATEGORIA	SUB-CATEGORIA	VARIABLE	AÑO INICIAL	AÑO FINAL	FRECUENCIA
Hidrología	Aportes	Caudal	2000	2020	Diaria
Demanda	Demanda Comercial	Demanda Por Comercializador	2000	2017	Horaria
Demanda	Demanda Nacional	Demanda Energía SIN	2000	2020	Diaria
Oferta	Despacho	Generación AGC	2000	2020	Horaria
Oferta	Generación	Generación	1995	2020	Horaria
Transacciones y Precios	Precios	Precio Oferta	2000	2020	Diario
Clima	-	Temperatura	1990	2020	Diario
Clima	-	Precipitación	1990	2020	Diario
Transacciones y Precios	Precios	Precio Bolsa Nacional	1995	2020	Horaria

Tabla 5.1: Conjunto de variables seleccionadas y su respectiva periodicidad.

Las fuentes oficiales en las cuales se pudo encontrar estas variables fueron National Centers for Environmental Information[25] y XM[35], dicha búsqueda de fuentes también fue producto de esta etapa.

5.1.3. Etapa 3: Recolección de datos

Con las variables y las fuentes de datos definidas, en esta etapa se accedió a cada uno de los sitios web que contienen los archivos de datos, se descargó manualmente cada uno de ellos y se almacenaron en una estructura de carpetas definida por categoría, subcategoría y variable.

Una vez completa la base de datos local, se programó un conjunto de funciones que tenía como objetivo estandarizar cada uno de los archivos de datos, debido a la naturaleza variante de cada una de las características. Una vez aplicadas estas funciones de preprocesamiento, se construyó una carpeta adicional en la cual se guardaron archivos de datos estándar para cada una de las variables.

Esta etapa también involucró un paso adicional de procesamiento, el cual solo se aplicó sobre la información climática. Teniendo en cuenta que estos datos se ubican espacialmente gracias a su latitud y longitud se decidió crear una imagen de tres canales que represente el comportamiento de las variables de forma visual. En la imagen 5.1 se muestran las estaciones meteorológicas de las cuales se extrajeron los valores de las variables y en la imagen 5.2 se muestra la imagen transformada correspondiente

a cada característica. En total se generaron 22160 imágenes que van desde el 1 de Enero de 1990 hasta el 31 de Mayo del 2020.

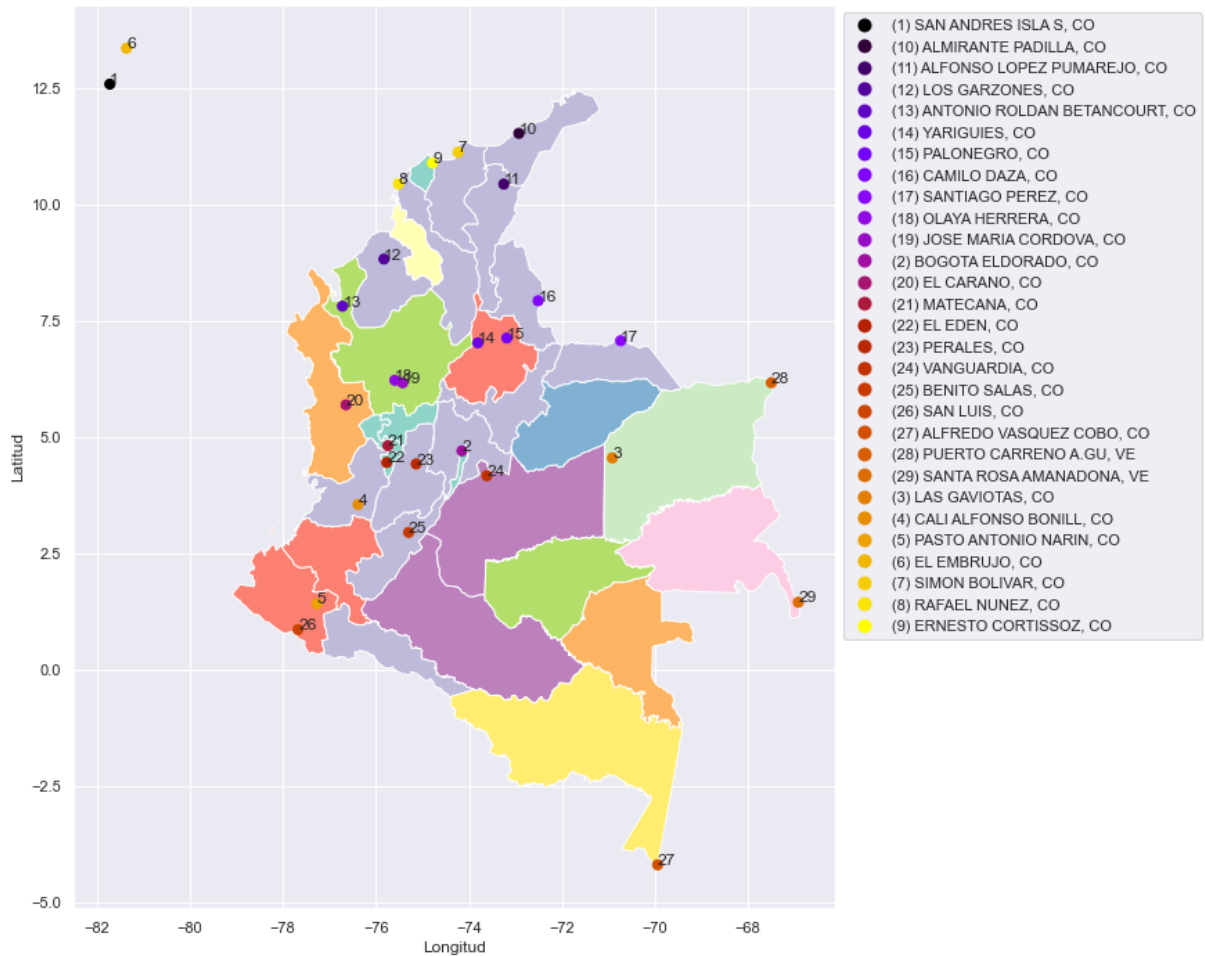


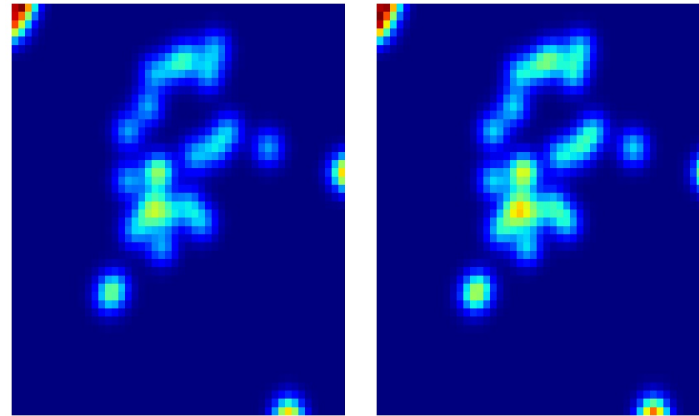
Figura 5.1: Estaciones meteorológicas

5.1.4. Etapa 4: Preprocesamiento y análisis

El objetivo principal de esta etapa fue realizar un análisis de cada una de las variables, el cual permitió definir el conjunto final de características que se entregaron a las etapas posteriores, establecer el rango de fechas a trabajar y consolidar el modo de acceso y almacenamiento.

Para poder realizar este análisis, se construyeron Notebooks de trabajo en los que se revisaba una por una las diferentes subvariables que contenían los datos, posteriormente se graficó la distribución y representación de cada una de ellas. Del procedimiento anterior se pudo construir el siguiente resumen de variables y características:

- Caudal:** Esta variable representa el volumen de agua que atraviesa la superficie de los ríos en un tiempo determinado. Está caracterizada por 5 elementos, nombre del río, región hidrológica, aportes caudal en $\frac{m^3}{s}$, aportes energía en \$kWh y



(a) Precipitación 14/01/2018 (b) Temperatura 14/01/2018

Figura 5.2: Variables climáticas

aporte en %. En total hay registrados 36 ríos diferentes y 6 regiones hidrológicas, los cuales en unas fechas específicas reportan los valores de las tres características adicionales. El rango de fechas asociados a esta variable está entre Enero 1 del 2000 al 30 de Septiembre del 2020.

- **Demanda Por Comercializador:** Esta variable es muy representativa, puesto que define el mercado en el cual se enfocó el presente trabajo. Está caracterizada por 3 elementos, código comercializador, mercado y valor de energía comercializada en cada una de las 24 horas del día. La variable mercado tiene 3 tipos diferentes: consumos, regulado y no regulado. Ya que el interés particular del estudio está asociado al mercado no regulado, se identifican 96 comercializadores únicos que operan en él. Que el mercado sea no regulado, significa que los diferentes actores pueden llegar a un acuerdo de precios y de esta manera pactar un valor fijo por día, mes, semestre o año (depende del acuerdo). El rango de fechas para esta variable está entre el 1 de Enero de 2000 y el 3 de Abril del 2020
- **Demanda Energía SIN:** Esta variable representa la cantidad de energía que es demanda por el Sistema Interconectado Nacional por día. Está caracterizada por 5 elementos: demanda energía SIN, generación, demanda no atendida, exportaciones e importaciones. Los valores registrados para esta variable están en el rango de fechas que oscilan entre Enero 1 del 2000 y 31 de Diciembre del 2020.
- **Generación AGC:** Esta variable indica la cantidad de energía que generan las plantas de reserva que atienden el Sistema Interconectado Nacional siempre que sean requeridas. La variable está caracterizada por 3 elementos: recurso, código agente y el valor de la energía generado en cada una de las 24 horas del día. Se identificó que la variable cuenta con un total de 18 recursos únicos que reportan datos en un rango de fechas entre Enero 1 del 2000 y el Diciembre 31 del 2020.
- **Generación:** Es una de las variables con mayor relevancia dentro del proceso de

definición del precio de la energía, puesto que determina la cantidad del recurso que produce una planta cada hora del día. Está caracterizada por un total de 6 elementos:

- **Recurso:** Representa las plantas de generación, de las cuales hay registradas un total de 258 diferentes.
- **Tipo de generación:** Está asociada a la manera en que se produce la energía y registra 5 tipos diferentes (hidráulica, térmica, cogenerador, eólica y solar).
- **Combustible:** Representa el producto por medio del cual la energía se genera. Hay registrados un total de 14 combustibles únicos.
- **Código agente:** Está asociado a las siglas que representan los agentes que administran los recursos.
- **Tipo Despacho:** Describe la manera por medio de la cual, la energía es despachada. Estos tipos se dividen en 2: despachado centralmente y no despachado centralmente.
- **Valor de la energía despachada** en cada una de las 24 horas del día.

La información asociada a esta variable tiene registros en el rango de fechas entre Julio 20 de 1995 y 31 de Diciembre del 2020.

- **Precio oferta:** Esta variable determina el valor que los agentes asignan al precio que ofertan por kWh de energía. Está caracterizada por 5 elementos: recurso, código agente, precio de oferta ideal, precio de oferta de despacho, precio de oferta declarado. En total hay registrados 62 recursos diferentes en un rango de fechas entre Enero 1 del 2000 al 31 de Diciembre del 2020.
- **Variables meteorológicas:** Estas variables están definidas por la precipitación y la temperatura promedio del país. Como se mencionó anteriormente, están medidas por 29 estaciones meteorológicas entre el 1 de Enero de 1990 hasta el 31 de Mayo del 2020.
- **Precio bolsa nacional:** Según la información oficial, este valor está definido como *"precio de oferta de la última planta flexible para atender la demanda comercial nacional, más delta de incremento para remunerar los costos no cubiertos de las plantas térmicas en el despacho ideal"*. Esta es la variable objetivo del proyecto puesto que representa el valor definitivo con el cual se transa en bolsa el precio de la energía. Los registros de esta variable tienen fechas desde Julio 20 del 1995 hasta 31 de Diciembre del 2020.

Finalizado el proceso de caracterización se pudo identificar el rango de fechas de cada una de las variables, lo cual fue suficiente para determinar que los datos recolectados entre el 1 de Enero del 2000 hasta el 30 de Abril del 2020 son los más adecuados para realizar el proyecto debido a que ocupan la mayor cantidad de información.

Adicionalmente, se encontró en [35] un listado que recoge los códigos de los agentes, para cada uno de ellos asocia el nombre completo, la actividad comercial, el estado y la fecha de registro. Con este mismo documento fue posible conocer que hay un total

de 40 comercializadores inactivos, lo cual garantiza que de los 96 comercializadores registrados en la variable *Demanda Por Comercializador* tan solo 50 de ellos están operativos.

Dado el proceso de caracterización de las variables, también fue posible identificar la gran cantidad de recursos y agentes asociados al proceso de generación de energía, el siguiente paso que se realizó en esta etapa fue desarrollar un análisis de generación por trimestres, con el objetivo de identificar los actores con mayor influencia. El propósito de haber construido el análisis en ese rango de tiempo fue conocer en qué medida eran importantes los recursos, debido a que hay algunos de ellos que tienen mayor participación en una etapa del año mientras que en otras no son relevantes.

Para este análisis, también fue importante tomar una ventana de tiempo adecuada, debido a que si se hubiera evaluado los trimestres de todo el espectro temporal se hubiera podido tener en cuenta actores que antaño fueron muy relevantes pero que a la fecha no lo son, es por esto que solo se tuvo en cuenta los agentes y recursos que aparezcan a partir del año 2018.

Inicialmente se realizó una evaluación para recursos y agentes implicados en la variable generación, como se pudo conocer en el resumen de esta variable, existen diferentes tipos de generación, para el caso particular de esta investigación la única fuente que se tendrá en cuenta será la hidráulica. En la tabla 5.2 se muestra un consolidado de la energía hidráulica generada en kWh por trimestre y por año.

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
2018	4.515'181.684	12.926'931.779	14.816'602.846	14.857'387.841
2019	13.541'862.259	12.538'098.170	15.139'839.603	14.129'667.381
2020	12.650'606.515	7.850'036.536	-	-

Tabla 5.2: Energía generada por trimestre por año.

Teniendo en cuenta la cantidad de recursos que se ven involucrados en el proceso de generación, se desarrolló una regla para el análisis la cual indica que por cada trimestre se tendrán en cuenta el 10% de los recursos que están por encima del percentil 90 de los valores de generación, el 90% restante de los recursos será tratado como Otros.

Dicho análisis nos arrojó los siguientes resultados, los cuales se pueden evidenciar también en las imágenes 5.3, 5.4 y 5.5.

■ **Año 2018:**

- **Trimestre I:** El 25.88% de la generación corresponde al conjunto de Otros recursos que contiene 108 recursos. Los recursos más representativos son San Carlos que general el 12.77% y el Guavio 10.72%.
- **Trimestre II:** En este trimestre repiten los recursos San Carlos y Guavio como más representativos con el 11.95% y 9.62% de energía generada. El

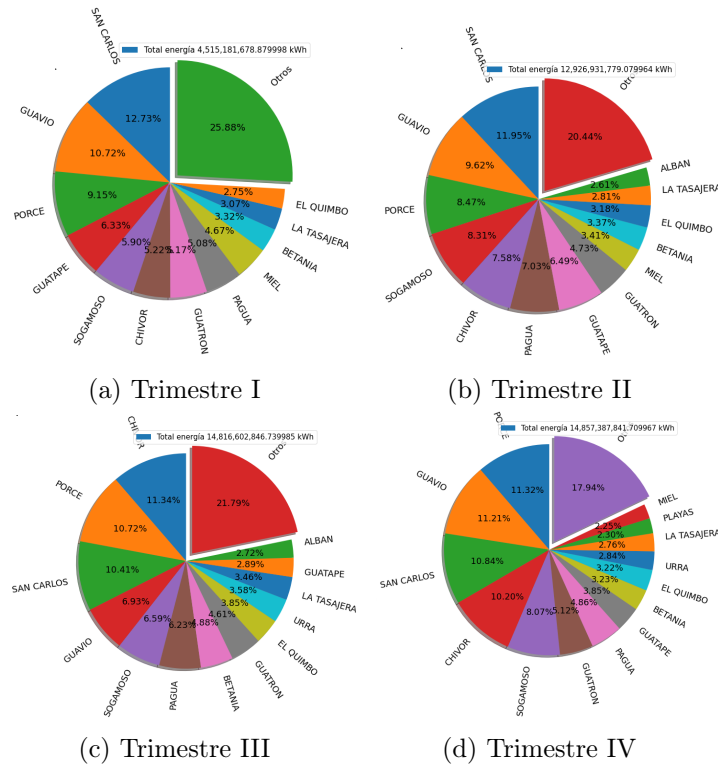


Figura 5.3: Generación hidráulica 2018

conjunto de Otros recursos tiene un total de 113 recursos y generan el 20.44 % de la energía.

- **Trimestre III:** En esta ocasión los recursos más representativos cambian por Chivor con 11.34% y Porce con 10.72% de la energía generada. El recurso San Carlos está muy cerca a estos márgenes con un 10.41%. El conjunto de Otros tiene un total de 114 recursos que generan el 21.79%.
- **Trimestre IV:** Para este trimestre Porce se convierte en el recurso más representativo con 11.32% de la energía, por otro lado están Guavio y San Carlos con 11.21% y 10.48% respectivamente. El conjunto de Otros genera el 17.94% de la energía y está compuesto por 118 recursos.

■ **Año 2019:**

- **Trimestre I:** Este primer trimestre San Carlos genera el 12.28% mientras que Porce el 9.96%, por otro lado el conjunto Otros que en esta ocasión está compuesto por 117 cantidad de recursos genera el 19.77%.
- **Trimestre II:** San Carlos repite como el recurso más representativo con un 12.19% y lo secunda Guavio con 9.38%. El conjunto de 117 recursos produce en total 18.22% de la energía.
- **Trimestre III:** Los recursos más representativos para este trimestre son Porce y Guavio con un total de 10.92% y 10.12% de la energía generada respectivamente. El conjunto de Otros produce un total de 19.47% y tiene incluidos 117 recursos.

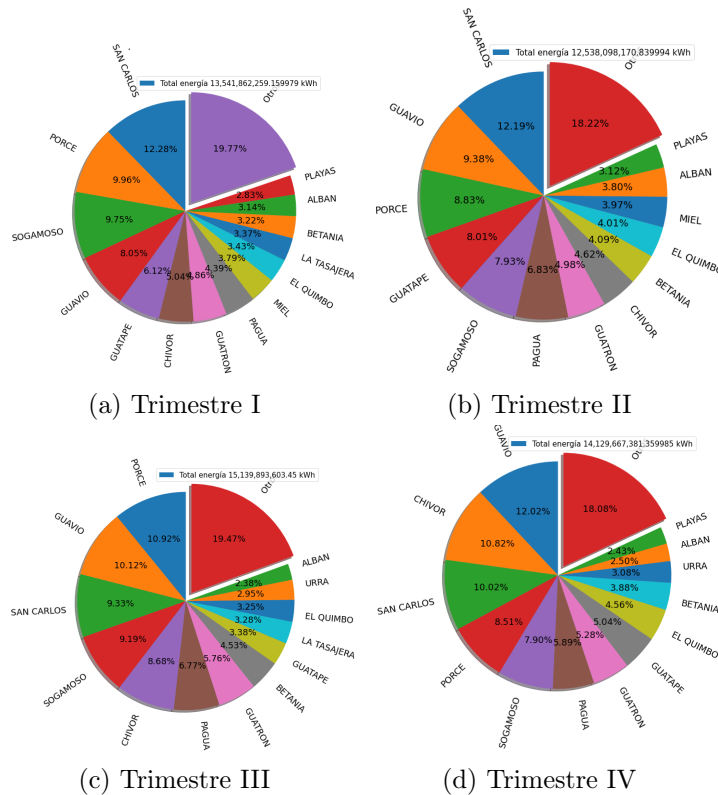


Figura 5.4: Generación hidráulica 2019

- **Trimestre IV:** Guavio pasa a ser el recurso más representativo con un total de 12.02% de la energía generada, seguido por Chivor con un 10.82%, por otro lado el conjunto de 113 de Otros recursos genera el 18.08%.

■ **Año 2020:**

- **Trimestre I:** San Carlos y Porce son los recursos más representativos con 11.74% y 8.57% respectivamente. El conjunto de otros tiene un total de 23.27% de la energía generada.
- **Trimestre II:** En esta ocasión San Carlos repite como más representativo con un 13.27% y lo secunda Guavio con 12.56% del total de energía generada. Para este trimestre el conjunto de 113 de Otros produce 17.40% de la energía generada.

Teniendo en cuenta el análisis previo y con el objetivo de hacer más robusta la selección de recursos, se realizó un conteo de ocurrencias en cada uno de los trimestres para el mismo rango de fechas. El criterio de selección que se tomó fue a partir del cuantil 0.9, todo aquel recurso que tenga por lo menos una ocurrencia en cualquiera de los 10 trimestres será agregado. En la tabla 5.3 se muestra el resultado del conteo ordenado por cuantiles, se pudo confirmar que aquellos recursos identificados en el análisis con mayor participación fueron agregados.

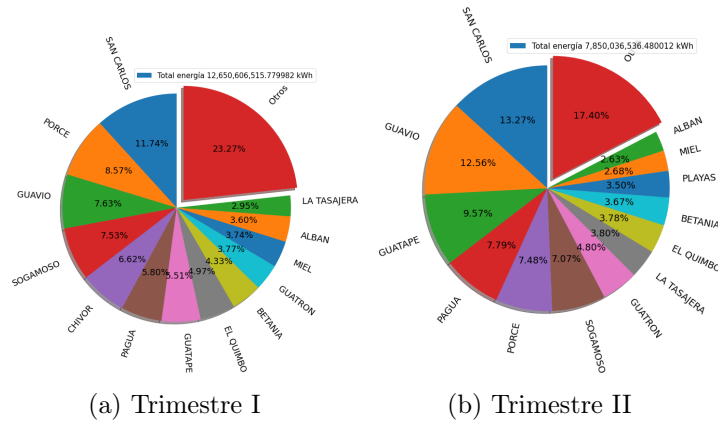


Figura 5.5: Generación hidráulica 2020

con el grupo definitivo de recursos seleccionados se procedió a realizar ajustes sobre aquellas variables que hacen referencia a estos, ya que todas las variables deben coincidir en los actores que interactúan. Un caso particular se presenta en la variable caudal, la cual como se mencionó anteriormente relaciona los recursos hídricos con los recursos generación. En [34] se encontró un documento en el cual se hace la relación de la mayoría de los ríos con su respectivo recurso de generación y además relaciona también la represa a la cual alimentaban. Con esta información se pudo realizar el cruce de recursos y a su vez seleccionar aquellos con los que se va a trabajar. En la tabla 5.4 se muestra la relación entre recursos.

RECURSOS	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
SAN CARLOS	10	10	10	10	10	10	10	10	10	10
GUAVIO	10	10	10	10	10	10	10	10	10	10
PORCE	10	10	10	10	10	10	10	10	10	10
SOGAMOSO	10	10	10	10	10	10	10	10	10	10
PAGUA	10	10	10	10	10	10	10	10	10	8
CHIVOR	10	10	10	10	10	10	10	10	9	8
GUATAPE	10	10	10	10	10	10	10	10	10	6
GUATRON	10	10	10	10	10	10	10	10	10	6
BETANIA	10	10	10	10	10	10	10	10	10	1
EL QUIMBO	10	10	10	10	10	10	10	10	10	0
LA TASAJERA	10	10	10	10	10	10	10	10	8	0
ALBAN	10	10	10	10	10	10	10	10	8	0
MIEL	10	10	10	10	10	10	10	10	7	0
PLAYAS	10	10	10	10	10	10	10	10	5	0
URRA	10	10	10	10	10	10	10	10	4	0

Tabla 5.3: Energía generada por trimestre por año.

Para finalizar esta etapa y a su vez concluir con la sección, el último paso consistió en construir funciones de código que se encargaron de combinar todas las variables de acuerdo a su periodicidad, aplicar los filtros de fechas, seleccionar y extraer los

NOMBRE RIO	RECURSO
Altoanchicaya	Alban
Digua	
Betania Cp	Betania
Magdalena Betania	
Bata	Chivor
El Quimbo	El Quimbo
Nare	Guatapé
Concepción	Guatron
Desv. Eeppm (Nec,Paj,Dol)	
Guadalupe	
Tenche	
Guavio	Guavio
Grande	La Tasajera
Miel	Miel
Desv. Guarino	
Desv. Manso	
Bogota N.R.	Pagua
Chuza	
Guatapé	Playas
Porce	Porce
San Carlos	San Carlos
Sogamoso	Sogamoso
Sinú Urrá	Urra

Tabla 5.4: Recursos Hídricos asociados a recursos de generación.

recursos de generación y agentes comercializadores definidos y por último realizar los cruces entre variables.

Todo este proceso generó 3 conjuntos de datos que serán el insumo correspondiente para las etapas de simulación y entrenamiento. La estructura de estos archivos de datos presenta la siguiente composición.

- **Conjunto de datos horaria:** Este archivo tiene un total de 177480 registros y 85 características de las cuales 15 pertenecen a la variable generación, 57 están asociadas a la variable demanda por comercializador y 13 son de la variable generación AGC.
- **Conjunto de datos diaria:** Tiene un total de 7395 registros con 119 columnas de las cuales 5 son relacionadas con la variable demanda energía SIN, 69 son de la variable caudal y 45 asociadas a la variable precio oferta.
- **Conjunto de datos variable a predecir:** Dado que la variable a predecir está medida en horas por día, este archivo al igual que la sábana de datos horaria, tiene un total de 177480 registros y una sola columna que corresponde al valor del precio en bolsa de la energía eléctrica.

Finalmente, por cada una de estos conjuntos de datos se generó un archivo resumen que contiene la siguiente información por variable: Valor mínimo, valor máximo

diferente de cero, valor máximo, media, mediana, desviación, % valores diferentes de cero, % valores en cero.

5.2. Sección 2: Experimentación

5.2.1. Etapa 5: Configuración y ejecución de experimentos

Esta etapa estuvo dedicada a desarrollar las funciones que fueron necesarias para gestionar y transformar la información en tensores de datos, los cuales posteriormente se usaron para realizar el proceso iterativo de experimentación. Adicionalmente, se generó la distribución de los datos en entrenamiento, validación y pruebas. Este primer conjunto contempla el rango de fechas entre el 1 de Enero del 2000 hasta el 31 de Diciembre del 2019, el segundo conjunto se dividió en el rango de fechas entre el 1 de Enero del 2020 hasta el 31 de Marzo del 2020, finalmente, el último conjunto obtuvo los datos entre el 1 y el 30 de Abril del 2020.

En cuanto al proceso de transformar los datos en tensores, inicialmente se tomó como referencia la documentación oficial de Tensorflow[1], la cual propone un marco de trabajo especial dedicado exclusivamente a datos representados como series de tiempo. Esta metodología se implementó tal como se propone, sin embargo, paralelamente se trabajó otra metodología que consistió en aplicar los mismos pasos con la diferencia que en esta no se hizo uso de ningún framework de desarrollo.

Una vez desarrolladas las rutinas de tratamiento de los tensores, en el siguiente paso se definieron los diferentes tipos de arquitectura que fueron probadas. Las primeras 4 arquitecturas implementadas fueron: redes neuronales recurrentes, Long Short Term Memory, Gated Recurrent Unit y Convolutional Neural Network. Estas redes se probaron con dos composiciones diferentes de capa de salida. En la primera composición se tuvo una sola unidad en la capa de salida, esto con el objetivo que la red predijera el valor siguiente y así tomar este como entrada para la siguiente predicción, este método es conocido como ventana deslizante (*rolling window*). En la segunda composición la capa de salida tuvo tantas unidades como horas se iban a predecir, lo que implica que, si la intención era predecir 24 horas en el futuro, entonces la capa de salida tendría una unidad por hora, en total 24 unidades. A este método se le conoce como *multi-output*.

En la parte final de esta etapa, se procedió a realizar experimentos únicamente con las arquitecturas recurrentes Long Shot Term Memory y Gated Recurrent Unit con la metodología *multi-output*. Una modificación que se realizó en la etapa de experimentación para estas arquitecturas, fue haber implementado la búsqueda de hiperparámetros con el optimizador bayesiano disponible en el framework Keras[7].

Es importante aclarar que cuando se realizó el proceso de experimentación se tuvo en cuenta exclusivamente los datos de entrenamiento y validación. Esto también fue así cuando se realizó la experimentación con la construcción del modelo completo.

El uso del conjunto de pruebas solo fue necesario cuando se requirió comprobar los resultados del mejor modelo.

5.2.2. Etapa 6: Evaluación y selección

En esta etapa, se evaluó el resultado de los experimentos realizados en cada uno de los pasos de la etapa anterior, se tomó cada uno de ellos y se pasaron a formatos tabulares en los cuales pudieron contrastar. Este proceso permitió tener un registro continuo de los resultados obtenidos y de esta manera se hizo el ajuste de aquello que fuera necesario en cualquier etapa del proceso.

Otra de las tareas desarrolladas en esta etapa consistió en consolidar los mejores modelos en una red que permitió la entrada de los diferentes tipos de datos (diarios, horarios, imágenes), con la cual, posteriormente se retornó a la etapa de experimentación y evaluación de hiperparámetros.

El último paso de esta etapa y a su vez del proyecto fue seleccionar el modelo que mejor se adaptó al comportamiento de los datos y con él presentar los resultados definitivos del proyecto.

Tal como fue descrito en la metodología, la búsqueda del modelo que mejor se adaptara a los datos fue un proceso que se dividió en dos partes. Inicialmente, se definieron un conjunto de arquitecturas con las que fueron probadas las series diarias y horarias, posteriormente, se seleccionaron los mejores modelos de cada una de las series y se combinaron en una arquitectura multimodal la cual también pasó por una etapa de experimentación para ajustar los últimos hiperparámetros y de esta manera seleccionar el modelo con mejores resultados. La selección de los modelos más representativos se realizó teniendo en cuenta los resultados obtenidos por las métricas MAPE y sMAPE en el conjunto de validación.

A pesar de que las variables meteorológicas presentan una periodicidad diaria, fueron tratadas con un conjunto de arquitecturas independientes ya que las imágenes necesitan ser procesadas bajo otras configuraciones de red como *ConvLSTM* o *CNN-LSTM*. Esto generó que la arquitectura multimodal inicialmente tuviera 3 tipos de entradas diferentes.

Los resultados del proceso de experimentación también estuvieron determinados por el tratamiento de tiempos que se le da a la serie. Se estableció como condición que los tiempos a predecir serían de 24 horas en el futuro teniendo en cuenta tres tipos diferentes de tiempos en el pasado, estos serían:

- Tipo I: 48 horas y 2 días en el pasado para las series horarias y diarias respectivamente.
- Tipo II: 72 horas y 3 días en el pasado para las series horarias y diarias respectivamente.
- Tipo III: 120 horas y 5 días en el pasado para las series horarias y diarias respectivamente.

La etapa de experimentación sobre las series diarias y horarias se realizó teniendo en cuenta 16 arquitecturas diferentes, cada una de ellas propone diversos tipos de

capas, tanto recurrentes como densas, las cuales presentan un conjunto de hiperparámetros que fueron analizados por el proceso de optimización bayesiana el cual se ejecutó con 20 exploraciones por experimento, minimizando la función de costos *symmetric mean absolute percentage error* (sMAPE) y haciendo uso de *upper confidence bound* (UCB) como función de adquisición. En la Tabla 6.1 se muestran los tipos de capa con su respectivo conjunto de hiperparámetros y el rango sobre el cual se realizó la exploración.

Arquitectura	Frecuencia	Tiempos	MAPE Train	MAPE Val	sMAPE Train	sMAPE Val
5	Diaria	Tipo I	24,222	16,762	22,606	17,109
13	Horaria		18,288	18,899	17,321	18,661
6	Diaria	Tipo II	23,781	17,081	22,175	17,035
13	Horaria		17,657	17,582	16,898	16,651
6	Diaria	Tipo III	23,545	15,701	21,907	15,485
14	Horaria		18,227	18,353	17,174	16,641

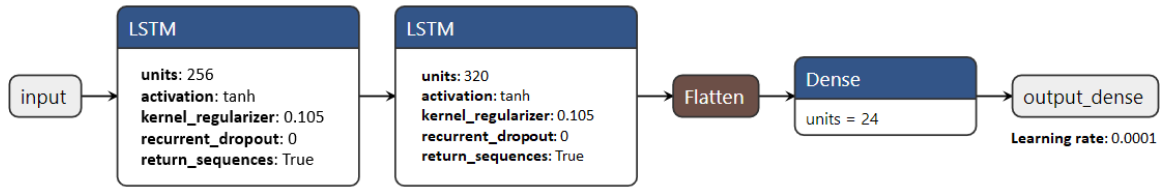
Tabla 6.2: Mejores resultados en arquitecturas para series diarias y horarias.

Teniendo en cuenta los resultados obtenidos (tabla 6.2) se puede concluir que tanto los datos diarios como horarios se adaptan mucho mejor a la estructura de red compuesta por 2 bloques recurrentes y una sola capa densa la cual es la misma capa de salida. Una de las principales diferencias es el tipo de red recurrente, mientras que para la arquitectura 5 y 13 se usa la unidad LSTM, para la arquitectura 6 y 14 se usa la unidad GRU. También se puede identificar que para las series diarias la última capa recurrente retorna toda la secuencia y por consiguiente se debe aplicar una capa de aplanamiento antes de entrar a la capa de salida, ya que de lo contrario, las dimensiones entrada requeridas en la siguiente capa no coincidirían con las dimensiones de salida actuales. Este comportamiento no se ve en las series horarias, puesto que en ellas la última capa recurrente no retorna toda la serie sino un único valor, lo cual presenta una entrada perfecta para la capa densa de salida. Finalmente, tal como se ve en la imagen 6.1 se pueden notar diferencias notables en la selección de hiperparámetros para cada una de las arquitecturas.

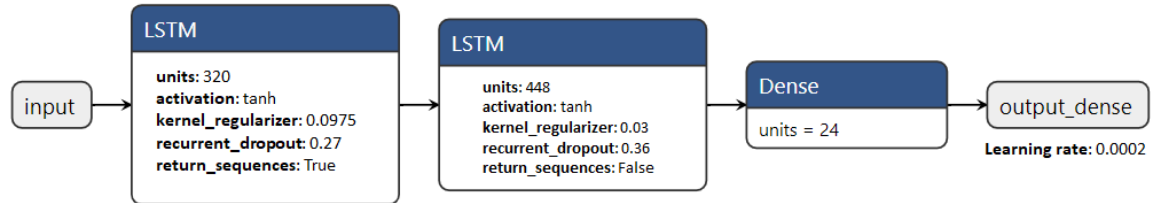
Finalizado el proceso de experimentación con las series diarias y horarias se procedió a ejecutar un procedimiento similar, pero con los datos meteorológicos. En la tabla 6.3 se muestra el conjunto de las 12 arquitecturas que fueron analizadas junto con el rango de hiperparámetros sobre los cuales se realizó el proceso de exploración.

En la tabla 6.4 queda en evidencia como la arquitectura I11 presenta los mejores resultados para los datos del tipo I, por otro lado, la arquitectura I9 es la que tiene la mejor configuración con los datos del tipo II, finalmente, para los datos del tipo III la arquitectura I8 es la que tiene el mejor ajuste. La estructura de cada una de estas redes así como la configuración particular de hiperparámetros se pueden confirmar en la imagen 6.2.

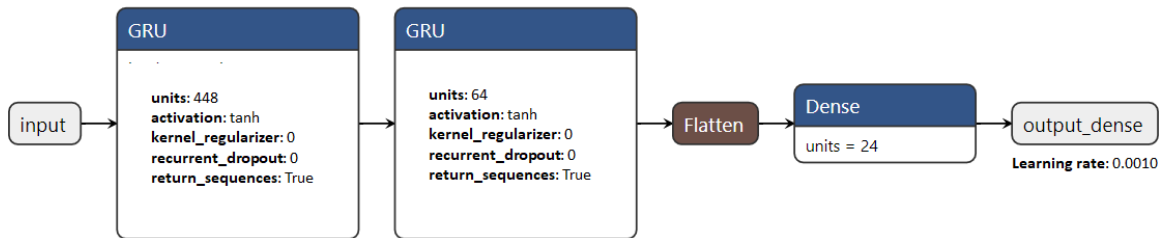
El siguiente paso consistió en tomar los mejores modelos para cada uno de los tipos de datos y combinarlos en una arquitectura multimodal completa la cual tendría 3 diferentes entradas. Adicionalmente, como las arquitecturas creadas ya tenían



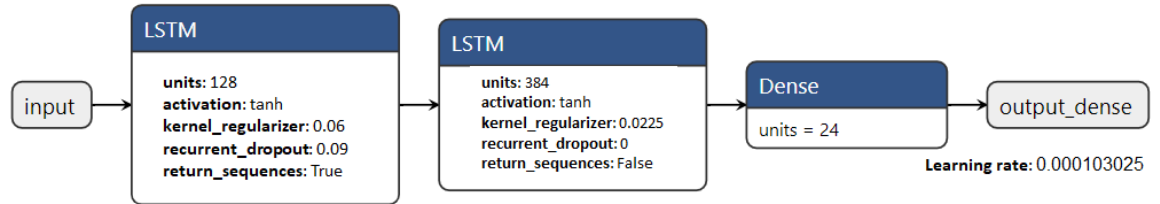
(a) Arquitectura 5 datos diarios Tipo I



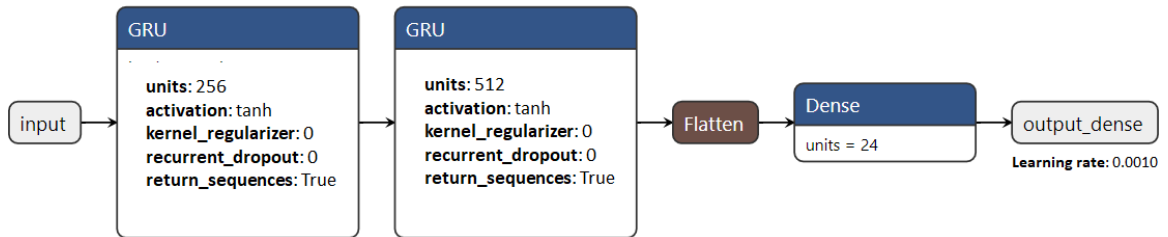
(b) Arquitectura 13 datos horarios Tipo I



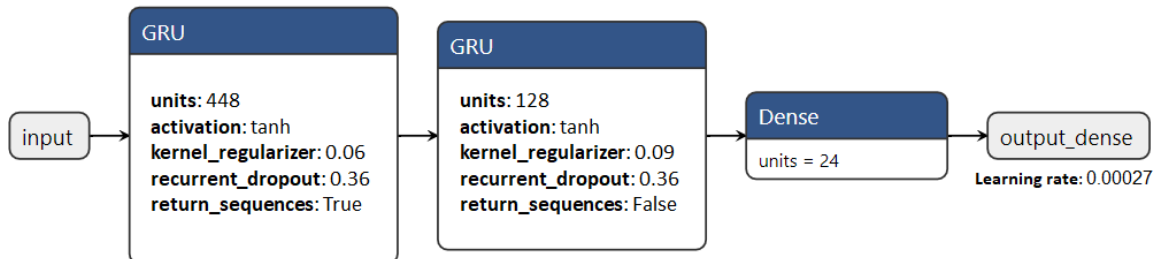
(c) Arquitectura 6 datos diarios Tipo II



(d) Arquitectura 13 datos horarios Tipo II



(e) Arquitectura 6 datos diarios Tipo III

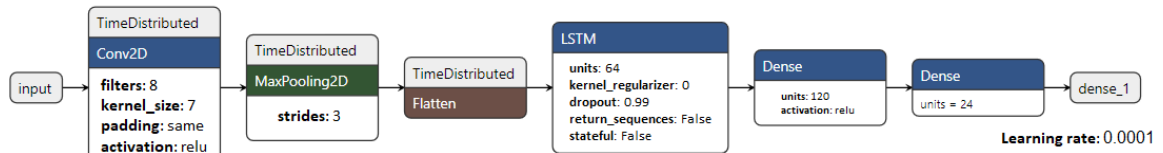


(f) Arquitectura 14 datos horarios Tipo III

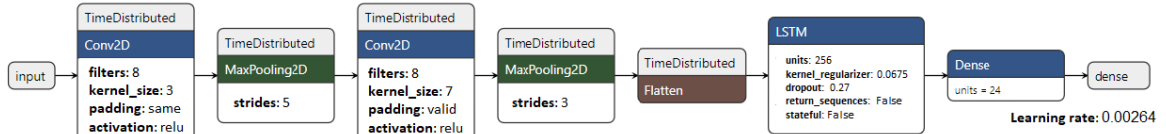
Figura 6.1: Arquitecturas diarias y horarias con mejores resultados para tipo de datos.

Arquitectura	Capa 1	Capa 2	Capa 3	Capa 4	Capa 5	Capa 6	Capa 7	Capa 8	Capa 9	Capa 10	Learning rate
11	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Flatten	tipo: Dense units: [24,20] activation: [relu,tanh,sigmoid]	tipo: Dense units: 24 activation: None	-	[0.001,0.01]
12	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Flatten	tipo: Dense units: [24,20] activation: [relu,tanh,sigmoid]	tipo: Dense units: 24 activation: None	-	-	-	[0.001,0.01]
13	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Flatten	tipo: Dense units: [24,20] activation: [relu,tanh,sigmoid]	tipo: Dense units: 24 activation: None	-	-	-	-	-	[0.001,0.01]
14	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Flatten	tipo: Dense units: 24 activation: None	-	-	[0.001,0.01]
15	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Flatten	tipo: Dense units: 24 activation: None	-	-	-	-	[0.001,0.01]
16	tipo: Conv2D LSTM filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu return_sequences: True	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: Flatten	tipo: Dense units: 24 activation: None	-	-	-	-	-	-	[0.001,0.01]
17	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Flatten)	tipo: LSTM units: [64,512] kernel_regularizer: [0.0,105] dropout: [0.0,99] return_sequences: False stateful: False	tipo: Dense units: [24,20] activation: [relu,tanh,sigmoid]	tipo: Dense units: 24 activation: None	[0.001,0.01]
18	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Flatten)	tipo: LSTM units: [64,512] kernel_regularizer: [0.0,105] dropout: [0.0,99] return_sequences: False stateful: False	tipo: Dense units: [24,20] activation: [relu,tanh,sigmoid]	tipo: Dense units: 24 activation: None	-	-	[0.001,0.01]
19	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Flatten)	tipo: LSTM units: [64,512] kernel_regularizer: [0.0,105] dropout: [0.0,99] return_sequences: False stateful: False	tipo: Dense units: [24,20] activation: [relu,tanh,sigmoid]	tipo: Dense units: 24 activation: None	-	-	-	-	[0.001,0.01]
110	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: 3	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Flatten)	tipo: LSTM units: [64,512] kernel_regularizer: [0.0,105] dropout: [0.0,99] return_sequences: False stateful: False	tipo: Dense units: 24 activation: None	-	[0.001,0.01]
111	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Flatten)	tipo: LSTM units: [64,512] kernel_regularizer: [0.0,105] dropout: [0.0,99] return_sequences: False stateful: False	tipo: Dense units: 24 activation: None	-	-	-	[0.001,0.01]
112	tipo: TimeDistributed(Conv2D) filas: 8 [3,2] kernel_size [3,5] padding [valid,same] activation: relu	tipo: TimeDistributed(MaxPooling2D) pool_size: [3,5]	tipo: TimeDistributed(Flatten)	tipo: LSTM units: [64,512] kernel_regularizer: [0.0,105] dropout: [0.0,99] return_sequences: False stateful: False	tipo: Dense units: 24 activation: None	-	-	-	-	-	[0.001,0.01]

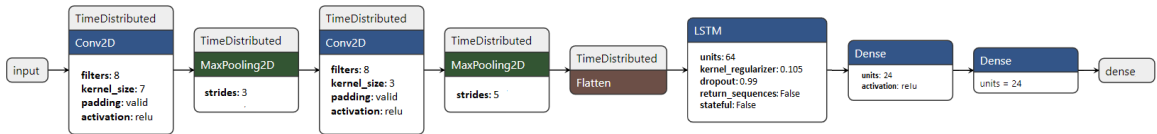
Tabla 6.3: Arquitecturas para variables climáticas.



(a) Arquitectura I9 datos climáticos Tipo I



(b) Arquitectura I11 datos climáticos Tipo II



(c) Arquitectura I8 datos climáticos Tipo III

Figura 6.2: Arquitecturas con mejores resultados para tipo de datos en variables climáticas.

una capa de salida fue necesario agregar dos capas extra, la primera con el objetivo de combinar las ultimas capas de los modelos y la segunda que se convirtiera en la capa de salida del modelo completo. Esta última se creó con tantas unidades de salida como tiempo se iban a predecir.

Para definir la primer capa extra del modelo completo se tuvo en consideración

Arquitectura	Tiempos	MAPE Train	MAPE Val	sMAPE Train	sMAPE Val
I9	Tipo I	131,103	53,123	69,886	77,443
I11	Tipo II	94,746	64,551	59,596	101,506
I8	Tipo III	61,948	68,754	50,65	106,065

Tabla 6.4: Mejores resultados en arquitecturas para variables climáticas.

Arquitectura	Tiempos	Capa Union	Capa Adicional 1	Capa Adicional 2	Capa Final	Learning rate	MAPE Train	MAPE Val	sMAPE Train	sMAPE Val
ModeloCompleto_Concat	Tipo I	Concatenate	tipo: Dense units: 48 activation: relu	tipo: Dense units: 72 activation: sigmoid	tipo: Dense units: 24 activation: None	1,4352E-05	623,118	187,13	62,424	65,077
ModeloCompleto_Suma		Add	-	-	tipo: Dense units: 24 activation: None	0,00026013	16,773	21,664	15,898	18,9
ModeloCompleto_Concat	Tipo II	Concatenate	tipo: Dense units: 48 activation: relu	tipo: Dense units: 48 activation: relu	tipo: Dense units: 24 activation: None	0,00706052	17,805	21,961	17,335	19,968
ModeloCompleto_Suma		Add	-	-	tipo: Dense units: 24 activation: None	0,00029706	17,904	19,001	17,622	18,353
ModeloCompleto_Concat	Tipo III	Concatenate	tipo: Dense units: 72 activation: relu	tipo: Dense units: 24 activation: relu	tipo: Dense units: 24 activation: None	0,00123977	16,661	23,611	16,151	22,296
ModeloCompleto_Suma		Add	-	-	tipo: Dense units: 24 activation: None	4,5828E-05	17,369	17,229	16,682	16,577

Tabla 6.5: Resultados proceso de experimentación modelos completos con imágenes.

2 aproximaciones, una de ellas consistió en concatenar los resultados de las salidas y generar 2 capas densas adicionales (sin tener en cuenta la capa de salida) que se entrenaron a la par del modelo, la otra aproximación consistió en sumar los resultados de las salidas e inmediatamente anexar la última capa densa. Estas dos aproximaciones pasaron por un proceso de exploración de hiperparámetros. En el primer caso para determinar el número de unidades, función de activación y tasa de aprendizaje; en el segundo caso para identificar la mejor tasa de aprendizaje. En la tabla 6.5 se presentan todos los resultados de los procesos de experimentación sobre estos modelos completos y sus respectivas aproximaciones.

Una vez finalizados los procesos de experimentación, el siguiente paso consistió en reentrenar las mejores configuraciones de modelos por tipo de dato, esta vez teniendo en cuenta que los conjuntos de entrenamiento y validación se usaron para entrenar y el conjunto de pruebas se usó para pronosticar. En la tabla 6.6 se relacionan los resultados del proceso de entrenamiento y prueba.

Arquitectura	Tiempos	MAPE Train	MAPE Test	sMAPE Train	sMAPE Test
ModeloCompleto_Concat	Tipo I	5947,67	2002,45	200	200
ModeloCompleto_Suma		16,44	22,92	15,812	19,781
ModeloCompleto_Concat	Tipo II	724,17	289,35	70,558	75,412
ModeloCompleto_Suma		14,712	25,259	14,089	21,407
ModeloCompleto_Concat	Tipo III	49,6	63,446	49,312	93,783
ModeloCompleto_Suma		19,695	68,188	18,719	62,471

Tabla 6.6: Resultados entrenamiento y predicción sobre modelos completos con imágenes.

Los entrenamientos sobre los modelos 2,4 y 6 son el top 3 con mejores resultados en al momento de pronosticar sobre nuevos datos. Es importante notar que dichas

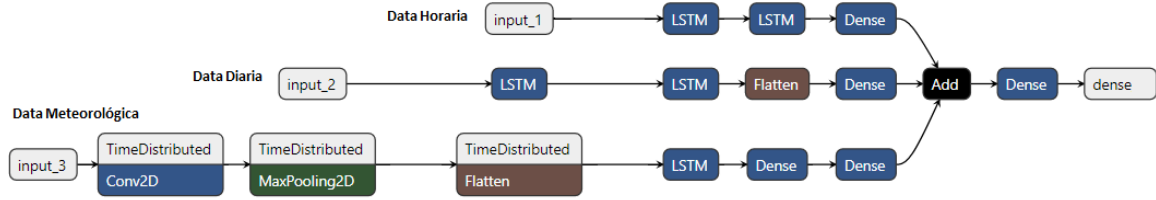


Figura 6.3: Arquitectura del mejor modelo completo con imágenes.

arquitecturas pertenecen a la aproximación número 2, aquella que únicamente suma las salidas de las capas densas. La arquitectura del mejor modelo se puede visualizar en la imagen 6.3, de la misma manera en la imagen 6.4 se pueden visualizar las predicciones sobre el conjunto de entrenamiento y prueba.

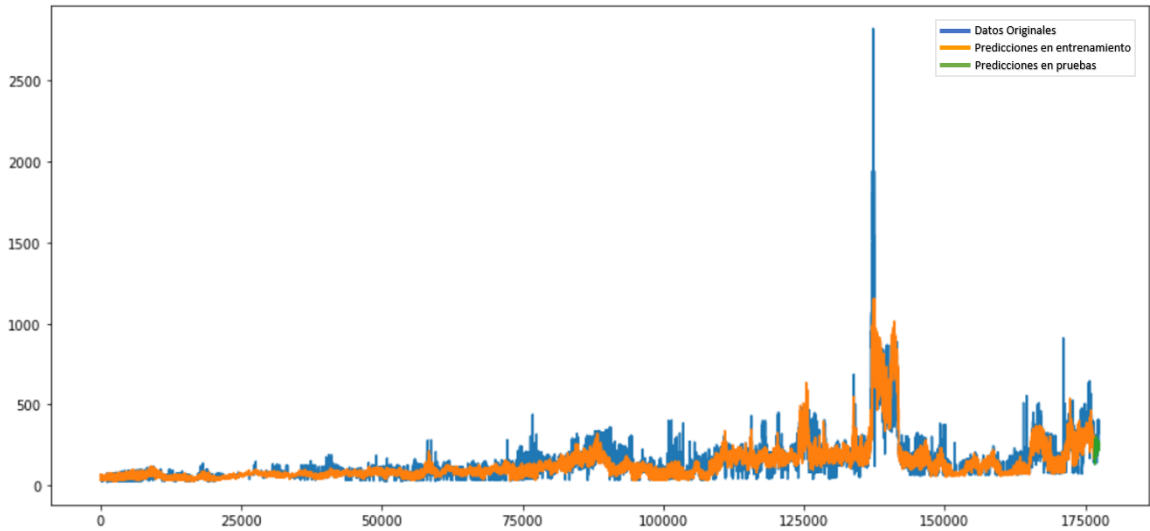
Comparando los resultados de experimentación de las arquitecturas para imágenes con los datos en series se evidenció una gran diferencia en métricas de evaluación de hasta un 95 %. Esto planteó la hipótesis del real benefició que tendrían los datos en imágenes al combinarlos como entradas en el modelo completo. Por tal motivo, se procedió a realizar un paso adicional de experimentación en el cual se tomaron los modelos completos y se quitó la entrada de imágenes, quedando así una arquitectura multimodal de dos entradas, una para datos diarios y otra para datos horarios. A esta configuración de arquitecturas también se anexó las dos aproximaciones de combinación de capas para posteriormente realizar la búsqueda de hiperparámetros (con resultados disponibles en la tabla 6.7) y finalmente realizar pronósticos sobre el conjunto de prueba.

Arquitectura	Tiempos	Capa Union	Capa Adicional 1	Capa Adicional 2	Capa Final	learning rate	MAPE Train	MAPE Val	sMAPE Train	sMAPE Val
ModeloCompleto_SinImagenes_Concat	Tipo I	Concatenate	tipo: Dense units: 72 activation: relu	tipo: Dense units: 24 activation: relu	tipo: Dense units: 24 activation: None	5.6032E-05	15,45	20,182	14,632	17,738
ModeloCompleto_SinImagenes_Suma		Add	-	-	tipo: Dense units: 24 activation: None	0,00096568	17,914	23,91	17,179	21,779
ModeloCompleto_SinImagenes_Concat	Tipo II	Concatenate	tipo: Dense units: 24 activation: relu	tipo: Dense units: 24 activation: sigmoid	tipo: Dense units: 24 activation: None	0,00072357	640,973	221,033	53,516	63,693
ModeloCompleto_SinImagenes_Suma		Add	-	-	tipo: Dense units: 24 activation: None	0,00057459	18,428	21,601	17,697	19,1645
ModeloCompleto_SinImagenes_Concat	Tipo III	Concatenate	tipo: Dense units: 96 activation: relu	tipo: Dense units: 96 activation: relu	tipo: Dense units: 24 activation: None	0,00035861	14,834	20,175	14,948	20,217
ModeloCompleto_SinImagenes_Suma		Add	-	-	tipo: Dense units: 24 activation: None	0,00035861	29,828	20,506	25,56	18,858

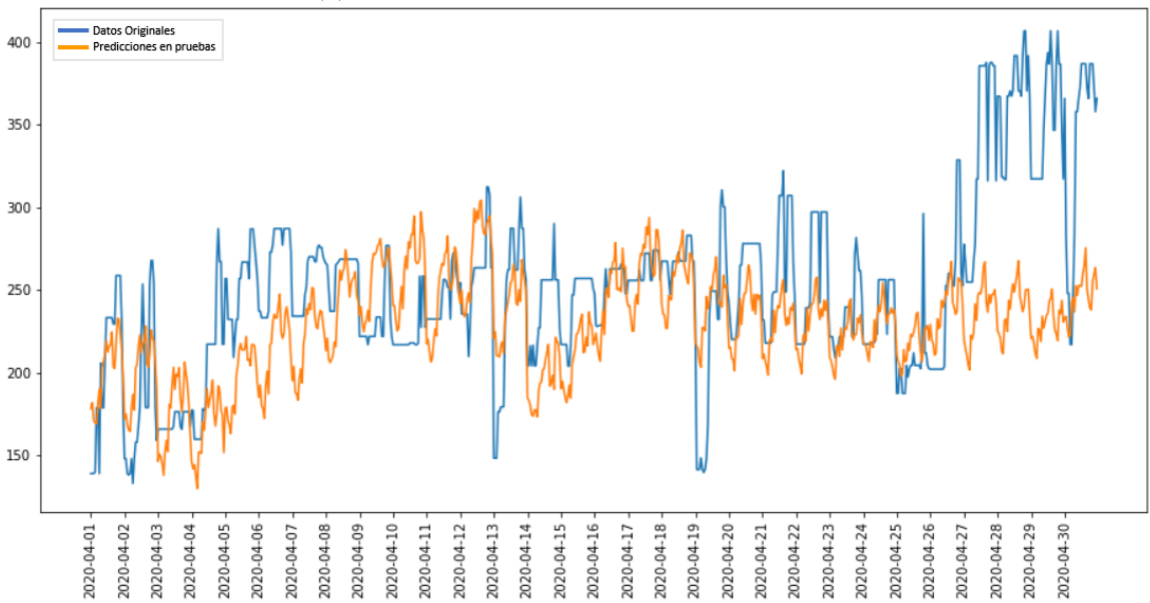
Tabla 6.7: Resultados proceso de experimentación modelos completos sin imágenes.

En la tabla 6.8 se presentan los resultados del proceso de entrenamiento y prueba donde quedó en evidencia que para la mayoría de arquitecturas los resultados mejoraron notablemente. El caso más relevante se presentó en la arquitectura número 6 donde el porcentaje de error para las métricas MAPE y sMAPE en el conjunto de pruebas se redujo en aproximadamente un 52 % generando que dicha arquitectura quede seleccionada como el modelo que mejor se ajusta a los datos. En la imagen 6.5 queda en evidencia las predicciones sobre los diferentes conjuntos.

Luego de seleccionar la arquitectura multimodal *ModeloCompleto SinImagenes*

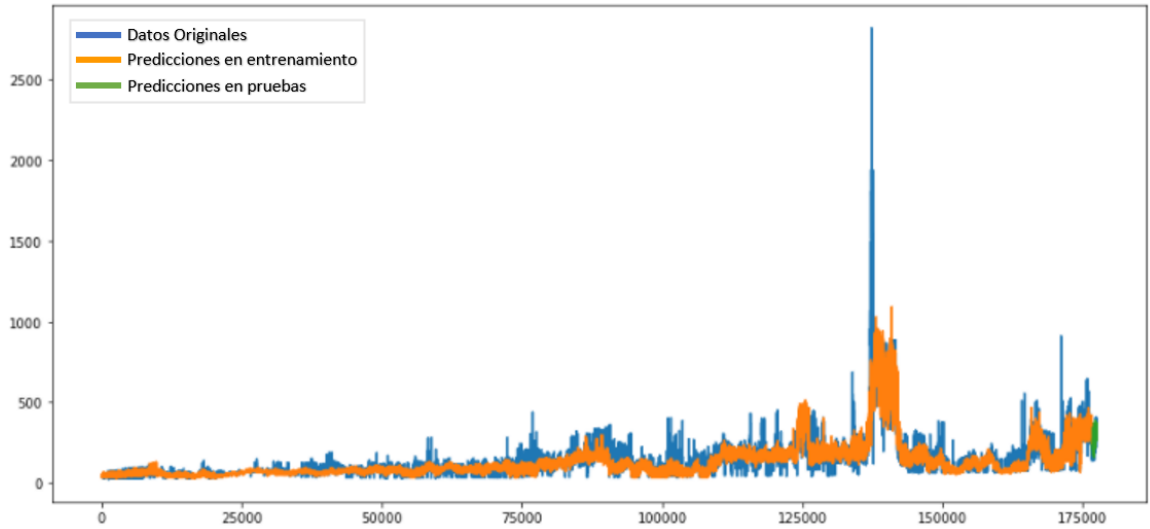


(a) Predicciones sobre la totalidad de datos.

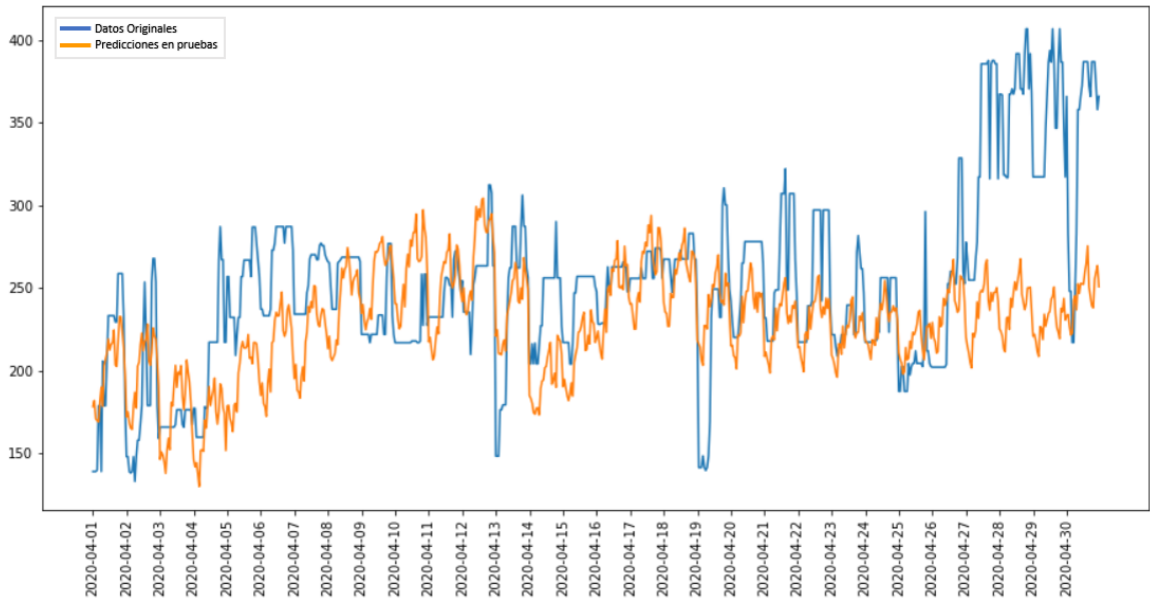


(b) Predicciones sobre el conjunto de pruebas.

Figura 6.4: Pronósticos realizados por la mejor arquitectura multimodal que tiene en cuenta las variables climáticas.



(a) Predicciones sobre la totalidad de datos.



(b) Predicciones sobre el conjunto de pruebas.

Figura 6.5: Pronósticos realizados por la mejor arquitectura multimodal sin tener en cuenta las variables climáticas.

Arquitectura	Tiempos	MAPE Train	MAPE Test	sMAPE Train	sMAPE Test
ModeloCompleto_SinImágenes_Concat	Tipo I	18,203	42,626	17,054	32,698
ModeloCompleto_SinImágenes_Suma		17,169	21,521	16,733	19,231
ModeloCompleto_SinImágenes_Concat	Tipo II	751,272	277,622	61,073	68,05
ModeloCompleto_SinImágenes_Suma		15,37	28,655	14,719	23,607
ModeloCompleto_SinImágenes_Concat	Tipo III	13,256	25,236	12,786	30,198
ModeloCompleto_SinImágenes_Suma		13,531	14,596	12,297	14,992

Tabla 6.8: Resultados entrenamiento y predicción sobre modelos completos sin imágenes.

Suma TipoIII como mejor modelo se identificó que las capas densas que están presentes previas a la suma de las salidas no contaban con función de activación, esto se debe a que la arquitectura para datos diarios y horarios se tomó tal cual había resultado del proceso de experimentación. Por lo tanto, esta particularidad generó una nueva etapa de experimentación en la cual se probó 4 tipos de configuraciones:

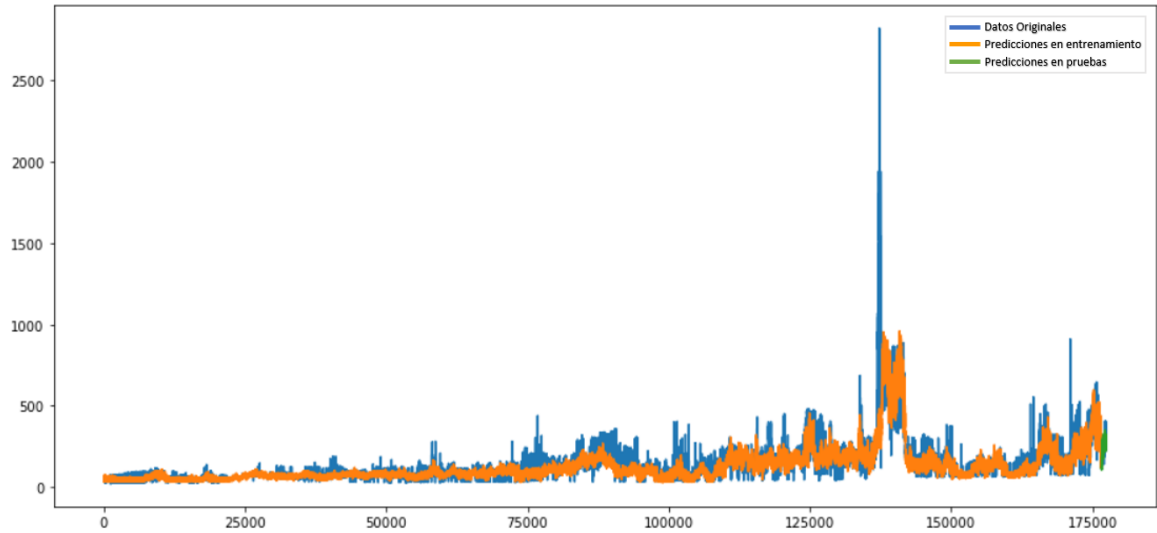
- Ambas capas densas con función de activación relu.
- Ambas capas densas con función de activación tanh.
- La capa densa correspondiente a los datos diarios con función de activación relu y la capa densa para datos horarios con función de activación tanh.
- La capa densa correspondiente a los datos diarios con función de activación tanh y la capa densa para datos horarios con función de activación relu.

Tal como se muestra en la tabla 6.9, el resultado de la configuración que usa ambas capas con función de activación relu presenta las mejores métricas. Al contrastar estos resultados con los obtenidos previamente por el modelo con capas densas sin función de activación, se puede evidenciar que a pesar de que las métricas tengan valores más altos para el conjunto de entrenamiento, esto se ve compensado con la disminución en el porcentaje de error en el conjunto de prueba. Adicionalmente, tal como lo muestra la gráfica de pronósticos para el conjunto de prueba (6.6), la tendencia se ajusta mejor al comportamiento esperado.

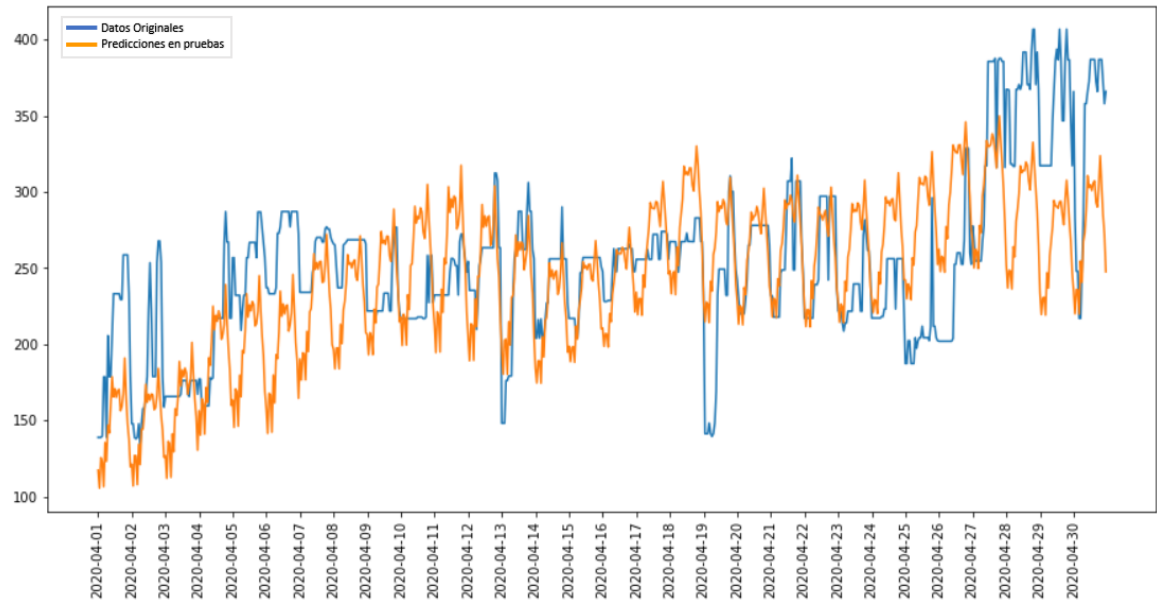
Arquitectura	Tiempos	Activacion Capa Densa Modelo Horario	Activacion Capa Densa Modelo Diario	MAPE Train	MAPE Test	sMAPE Train	sMAPE Test
ModeloCompleto_SinImágenes_Suma	Tipo III	relu	relu	17,653	14,284	16,885	14,739
		relu	tanh	16,944	14,893	16,225	15,255
		tanh	tanh	13,99	17,906	13,579	17,027
		tanh	relu	17,849	15,039	17,194	15,389

Tabla 6.9: Resultados de experimentación sobre el conjunto de entrenamiento y prueba iterando las funciones de activación de las capas densas correspondientes a los submodelos horarios y diarios.

Con el fin de comprobar los resultados obtenidos desde otro punto de vista, se realizó un proceso adicional que consistió en tomar los valores reales y las predicciones que generó el modelo sobre el conjunto de pruebas para dividir las en 7 series, cada una representando un día de la semana. Con estas series filtradas fue posible graficar el comportamiento de las predicciones y de esta manera se verificó cuáles son los días con mejor y peor ajuste. En la imagen 6.7 se pueden visualizar estos resultados.



(a) Predicciones sobre la totalidad de datos.



(b) Predicciones sobre el conjunto de pruebas.

Figura 6.6: Pronósticos realizados por la mejor arquitectura multimodal sin tener en cuenta las variables climáticas agregando función de activación relu en las capas densas correspondientes a los submodelos horarios y diarios.

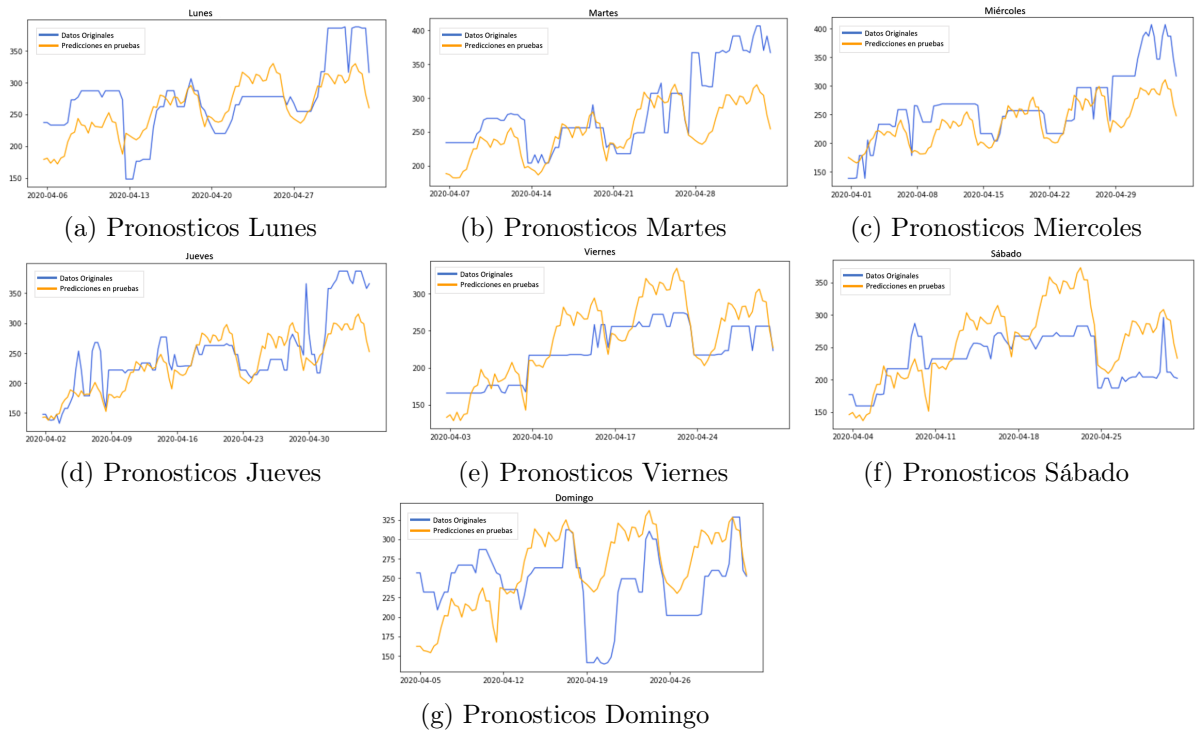


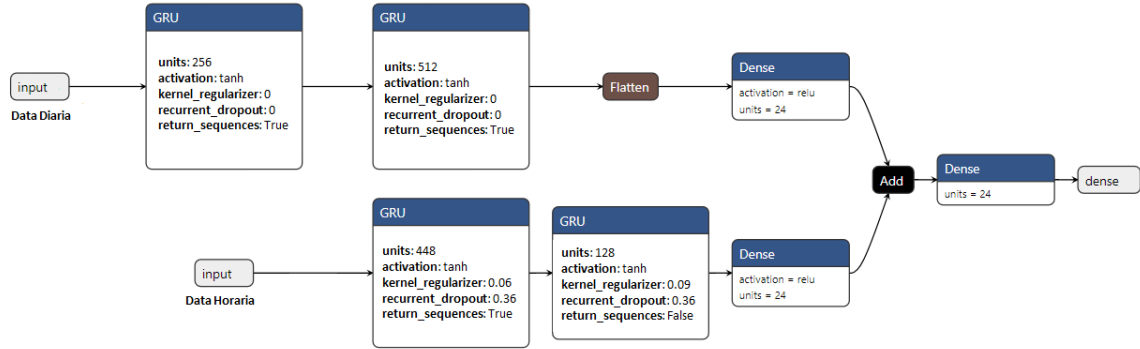
Figura 6.7: Pronosticos realizados por día sobre el conjunto de prueba.

Establecer un valor para el kWh de energía eléctrica es un proceso que ha venido sufriendo diferentes cambios a través de los años, particularmente, a partir de la década de los 90 los diferentes gobiernos latinoamericanos empezaron a implementar reformas administrativas que permitían regular el precio de este recurso. Una gran parte de estos cambios generó que la energía eléctrica fuera tratada como un commodity de gran valor en los mercados financieros, sin embargo, los analistas encontraron que la diversidad de factores internos y externos que se relacionan a él ocasionan que sea uno de los recursos con mayor volatilidad. Por tal motivo, se ha vuelto necesario construir modelos predictivos que interpreten la variabilidad del recurso y pronostiquen de manera adecuada su comportamiento en el corto, mediano y largo plazo.

La metodología propuesta en este proyecto determinó que para el conjunto de variables demanda energía SIN, caudal y precio oferta, expresadas en una periodicidad diaria y por otro lado las variables horarias generación, demanda por comercializador y generación AGC; la arquitectura multimodal representada en la imagen 7.1 es aquella que mejor se ajusta a los datos, alcanzando un 17.653 % y 16.885 % en las métricas MAPE y sMAPE para el conjunto de entrenamiento, mientras que para el conjunto de datos totalmente desconocidos alcanzó 14.284 % y 14.739 % en las mismas métricas.

Se puede notar que el modelo completo no hace uso de las variables meteorológicas, esto es porque al realizar la comparación de resultados en los procesos de experimentación entre las arquitecturas para imágenes contra las arquitecturas de datos en series de tiempo, se encontró hasta un 95 % de diferencia en las métricas de evaluación. Esto plantea la hipótesis del real benefició que tiene la construcción de imágenes a partir de datos meteorológicos.

Gracias a los resultados obtenidos, es posibles confirmar que el uso de una arquitectura multimodal supone una mejoría en el proceso de predicción del precio en bolsa de la energía eléctrica, ya que, si se extraen por aparte cada uno de los submodelos que componen la arquitectura completa y se entrenan de acuerdo a su periodicidad de



(a) Predicciones sobre la totalidad de datos.

Figura 7.1: Arquitectura Multimodal con mejor ajuste al comportamiento de los datos.

tiempos con sus respectivas variables, los resultados obtenidos no van a ser mejores que el alcanzado por el modelo completo (tabla 6.2 y 6.9).

Con el objetivo de mejorar los resultados obtenidos en este proyecto, se propone como trabajo futuro adicionar un conjunto de variables climáticas que permitan interpretar fenómenos meteorológicos de alto impacto en el proceso de generación de energía como lo pueden ser la niña o el niño. Así mismo, se planea hacer uso de otro tipo de arquitecturas basadas en modelos *Transformer* y *Encoder-Decoder* con componente de atención.

- [1] Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu y Xiaoqiang Zheng: *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. https://www.tensorflow.org/tutorials/structured_data/time_series, Software available from tensorflow.org.
- [2] Apaydin, Halit, Hajar Feizi, Mohammad Taghi Sattari, Muslume Sevba Colak, Shahaboddin Shamsirband y Kwok Wing Chau: *Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting*. *Water*, 12(5):1500, 2020.
- [3] Barrientos, Jorge, Edwin Rodas, Esteban Velilla, Mauricio Lopera y Fernando Villada: *Modelo para el pronóstico del precio de la energía eléctrica en Colombia*. *Lecturas de Economía*, (77):91–127, 2012.
- [4] Bello-Rodríguez, Sandra Patricia y Robert Baudilio Beltrán-Ahumada: *Caracterización y pronóstico del precio spot de la energía eléctrica en Colombia*. *Revista de la maestría en Derecho Económico*, 6(6):293–316, 2010.
- [5] Botero, Sergio Botero y Jovan Alfonso Cano Cano: *Análisis de series de tiempo para la predicción de los precios de la energía en la bolsa de Colombia*. *Cuadernos de Economía*, 27(48):173–208, 2008.
- [6] Ceylan, Zeynep: *Estimation of municipal waste generation of Turkey using socio-economic indicators by Bayesian optimization tuned Gaussian process regression*. *Waste Management & Research*, 38(8):840–850, 2020.
- [7] Chollet, François y cols.: *Keras*. <https://keras.io>, 2015.

- [8] Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho y Yoshua Bengio: *Empirical evaluation of gated recurrent neural networks on sequence modeling*. arXiv preprint arXiv:1412.3555, 2014.
- [9] Conejo, Antonio J, Javier Contreras, Rosa Espínola y Miguel A Plazas: *Forecasting electricity prices for a day-ahead pool-based electric energy market*. International journal of forecasting, 21(3):435–462, 2005.
- [10] Contreras, Javier, Rosario Espinola, Francisco J Nogales y Antonio J Conejo: *ARIMA models to predict next-day electricity prices*. IEEE transactions on power systems, 18(3):1014–1020, 2003.
- [11] Cuaresma, Jesús Crespo, Jaroslava Hlouskova, Stephan Kossmeier y Michael Obersteiner: *Forecasting electricity spot-prices using linear univariate time-series models*. Applied Energy, 77(1):87–106, 2004.
- [12] Gallón, Santiago y Jorge Barrientos: *Forecasting the Colombian Electricity Spot Price under a Functional Approach*. International Journal of Energy Economics and Policy, 11(2):67, 2021.
- [13] Gareta, Raquel, Luis M Romeo y Antonia Gil: *Forecasting of electricity prices with neural networks*. Energy conversion and management, 47(13-14):1770–1778, 2006.
- [14] Goodfellow, Ian, Yoshua Bengio, Aaron Courville y Yoshua Bengio: *Deep learning*, volumen 1. MIT press Cambridge, 2016.
- [15] Goodwin, Paul y Richard Lawton: *On the asymmetry of the symmetric MAPE*. International journal of forecasting, 15(4):405–408, 1999.
- [16] Hewamalage, Hansika, Christoph Bergmeir y Kasun Bandara: *Recurrent neural networks for time series forecasting: Current status and future directions*. International Journal of Forecasting, 37(1):388–427, 2021.
- [17] Hochreiter, Sepp y Jürgen Schmidhuber: *Long short-term memory*. Neural computation, 9(8):1735–1780, 1997.
- [18] Lago, Jesus, Fjo De Ridder y Bart De Schutter: *Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms*. Applied Energy, 221:386–405, 2018.
- [19] Lago, Jesus, Fjo De Ridder, Peter Vranckx y Bart De Schutter: *Forecasting day-ahead electricity prices in Europe: The importance of considering market integration*. Applied energy, 211:890–903, 2018.
- [20] LeCun, Yann, Yoshua Bengio y cols.: *Convolutional networks for images, speech, and time series*. The handbook of brain theory and neural networks, 3361(10):1995, 1995.
- [21] LeCun, Yann, Yoshua Bengio y Geoffrey Hinton: *Deep learning*. nature, 521(7553):436–444, 2015.

- [22] Lira, F, C Muñoz, F Nunez y A Cipriano: *Short-term forecasting of electricity prices in the Colombian electricity market*. IET generation, transmission & distribution, 3(11):980–986, 2009.
- [23] Makridakis, Spyros: *Accuracy measures: theoretical and practical concerns*. International journal of forecasting, 9(4):527–529, 1993.
- [24] Nogales, Francisco J y Antonio J Conejo: *Electricity price forecasting through transfer function models*. Journal of the Operational Research Society, 57(4):350–356, 2006.
- [25] Ocean Exploration, NOAA Office of y Research: *Climate Data Online: Dataset Discovery*, 2021. <https://www.ncdc.noaa.gov/cdo-web/datasets>, [Web; accedido el 06-10-2020].
- [26] Ooms, Marius, M Angeles Carnero, Siem Jan Koopman y cols.: *Periodic Heteroskedastic RegARFIMA models for daily electricity spot prices*. En *Econometric Society 2004 Australasian Meetings*, número 158. Econometric Society, 2004.
- [27] Sansom, Damien C, Tom Downs, Tapan K Saha y cols.: *Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants*. Journal of Electrical & Electronics Engineering, Australia, 22(3):227, 2003.
- [28] Santiago, Alberto Muñoz, Jaime Urquijo Vanstrahlengs, Aníbal Castro Otero y Jahir Lombana: *Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH*. Revista de Economía del Rosario, 20(1):127–159, 2017.
- [29] Snoek, Jasper, Hugo Larochelle y Ryan P Adams: *Practical bayesian optimization of machine learning algorithms*. arXiv preprint arXiv:1206.2944, 2012.
- [30] Stevenson, Maxwell J y cols.: *Filtering and forecasting spot electricity prices in the increasingly deregulated Australian electricity market*. School of Finance and Economics, University of Technology, Sydney, 2001.
- [31] Szkuta, BR, L Augusto Sanabria y Tharam S Dillon: *Electricity price short-term forecasting using artificial neural networks*. IEEE transactions on power systems, 14(3):851–857, 1999.
- [32] Weron, Rafal: *Modeling and forecasting electricity loads and prices: A statistical approach*, volumen 403. John Wiley & Sons, 2007.
- [33] Weron, Rafal, Adam Misiorek y cols.: *Forecasting spot electricity prices with time series models*. En *Proceedings of the European electricity market EEM-05 conference*, páginas 133–141, 2005.
- [34] XM: *Descripción sistema hidrológico SIN*, 2021. <http://paratec.xm.com.co/paratec/SitePages/hidrologia.aspx?q=descripcion>, [Web; accedido el 02-01-2021].

- [35] XM: *XM Portal BI, Información Inteligente*, 2021. <http://portalbissrs.xm.com.co/Paginas/Home.aspx>, [Web; accedido el 06-12-2020].