



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**PREDICCIÓN ORIENTADA A COSTOS PARA LA DETECCIÓN DE LAS K  
HORAS PUNTA EN MERCADOS ELÉCTRICOS**

TESIS PARA OPTAR AL GRADO DE  
MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL ELÉCTRICO

**MATÍAS IGNACIO MEDINA MEDINA**

PROFESOR GUÍA:  
RODRIGO MORENO VIEYRA

PROFESOR CO-GUÍA:  
VÍCTOR BUCAREY LÓPEZ

MIEMBROS DE LA COMISIÓN:  
PATRICIO MENDOZA ARAYA  
JUAN VELÁSQUEZ SILVA

SANTIAGO DE CHILE

2026

RESUMEN DE LA TESIS PARA OPTAR  
AL GRADO DE MAGÍSTER EN CIENCIAS  
DE LA INGENIERÍA, MENCIÓN ELÉCTRICA  
Y MEMORIA PARA OPTAR AL TÍTULO  
DE INGENIERO CIVIL ELÉCTRICO  
POR: MATÍAS IGNACIO MEDINA MEDINA  
FECHA: 2026  
PROF. GUÍA: RODRIGO MORENO VIEYRA

## **PREDICCIÓN ORIENTADA A COSTOS PARA LA DETECCIÓN DE LAS K HORAS PUNTA EN MERCADOS ELÉCTRICOS**

Los consumidores eléctricos sujetos a cargos por capacidad deben anticipar las horas de máxima demanda del sistema para decidir cuándo recurrir a un medio de generación propia y reducir así su costo de operación. Este problema presenta una escasez estructural de datos, ya que las horas punta se definen una sola vez por periodo de facturación. Para abordarlo, esta tesis reformula la identificación de las k horas punta como una clasificación binaria a nivel horario, lo que multiplica las instancias de entrenamiento, e incorpora una función de pérdida asimétrica basada en el arrepentimiento que pondera cada error según su costo económico. Se desarrollaron dos modelos de redes neuronales con esta pérdida y un modelo de referencia convencional, evaluados sobre el subsistema de distribución de la Región Metropolitana. Los modelos propuestos superaron de forma significativa al de referencia, con ahorros de hasta 791 USD/MW al año, y redujeron el costo de operación hasta en un 18.65 % frente a no utilizar un predictor. Además, dos análisis de sensibilidad mostraron que la ventaja es robusta y transferible a otros mercados, caracterizados por su asimetría de costos y su número de horas penalizadas.

*Medios a mi futuro,  
historias a mi pasado.*

# Agradecimientos

Luego de recorrer este largo viaje, lleno de penas y alegrías, de controles y tareas, de mate y de café, me doy cuenta de que toda esta travesía nunca podría haber sido llevada cabo sin el apoyo ni la compañía de mucha gente a mi alrededor.

En primer lugar, quiero agradecer a mi polola Evelyn, mi musa, mi fuente de inspiración. Agradezco tu amor intenso y genuino, con el que has sido un apoyo fundamental en estos últimos años de carrera. Agradezco aquel día de Septiembre en el que llegaste a mi vida. Y, agradezco que, desde que estamos juntos, has teñido mi mundo con una paleta de colores que solo tú puedes generar.

A mi familia, a mi Tata Carlos y a mi Tía Margarita, a mi Padrino Cesar, a mi Carlitos Daniel, a mi Toto Carol y a mis dos hermanitos chicos Gustavito y Cristóbal. Todos ustedes forman parte de lo que soy y los llevo a cada uno conmigo, siempre, donde quiera que vaya.

A mis amigos que conocí en el liceo, que en este punto ya son parte de mi familia: Sabato, Krostó y Gobo. ¿Cuántas juntas habremos tenido? ¿Cuántas veces habremos hablado? ¿Cuántas veces me han acompañado, contenido, escuchado y aconsejado? Imposible recordarlas todas, así como me es imposible agradecerles por tanto que me han entregado.

A mis amigos que formé en la u. A los Panapatines. A Montserrat, Adrián, Pepe, que los conocí antes de ser patinadmins. A Pablo Snacky, cuya amistad guarda óperas notables. A mi eterno compañero de eléctrica Maldopipe y a mi socio Byron, que espero que nuestro culto a los completos sea eterno.

A mi compadre Nikolito, que en paz descansa en las heladas tundras de Siberia. Me tendiste la mano durante todos los viernes de estos últimos 5 años. Viernes en los que nuestro podcast nos ha llevado a conversar de todo. Desde las maravillas de la ciencia hasta las trivialidades del día a día, pasando por unos sofisticados módulos de apreciación al arte. Nada más que desearte un ¡Feliz Jueves!, amigo mío.

A mis profesores guías, Rodrigo Moreno y Víctor Bucarey, por darme toda clase de oportunidades y consejos, por resolver todas mis dudas y por ayudarme a definir el camino que transité a lo largo de estos años de tesis y universidad.

Y finalmente, y a la más importante, quiero agradecer a mi madre, Marisol, por todos estos años siendo mi motivación y mi compañía, mi ayuda y mi consuelo. Por ser el pilar fundamental de mi vida y de mi persona. Si logré lo que logré, ha sido gracias a ti mamá.

# Tabla de Contenido

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>1</b>  |
| 1.1. Motivación . . . . .   | 1         |
| 1.2. Hipótesis . . . . .  | 2         |
| 1.3. Objetivos . . . . .  | 3         |
| 1.3.1. Objetivos generales . . . . .  | 3         |
| 1.3.2. Objetivos específicos . . . . .  | 3         |
| 1.3.3. Estructura de la tesis . . . . .   | 3         |
| <b>2. Estado del arte</b>   | <b>4</b>  |
| 2.1. Modelos de predicción de demanda eléctrica . . . . .                               | 4         |
| 2.1.1. Modelos de predicción de horas punta de demanda eléctrica . . . . .              | 5         |
| 2.1.2. Modelos de predicción de horas punta coincidentes de demanda eléctrica . . . . . | 7         |
| 2.2. Modelos con funciones de pérdida asimétricas . . . . .                             | 8         |
| 2.3. Resumen del estado del arte . . . . .  | 9         |
| 2.4. Contribuciones . . . . .   | 11        |
| <b>3. Metodología</b>   | <b>12</b> |
| 3.1. Nomenclatura . . . . .   | 12        |
| 3.1.1. Conjuntos e índices . . . . .  | 12        |
| 3.1.2. Parámetros . . . . .   | 12        |
| 3.1.3. Variables del problema . . . . .   | 13        |
| 3.1.4. Variables de entrada . . . . .   | 13        |
| 3.1.5. Variables de salida . . . . .  | 13        |
| 3.1.6. Funciones . . . . .  | 13        |
| 3.2. Descripción general . . . . .  | 14        |
| 3.2.1. Implementación de los modelos de predicción . . . . .                            | 15        |
| 3.3. Selección de variables y procesamiento de datos de entrada . . . . .               | 15        |
| 3.3.1. Selección de variables del modelo . . . . .                                      | 15        |
| 3.3.2. Procesamiento de datos . . . . .   | 17        |
| 3.4. Diseño de la función de pérdida . . . . .  | 18        |
| 3.4.1. Formulación del problema . . . . .   | 18        |
| 3.4.2. Función objetivo y restricciones . . . . .                                       | 19        |
| 3.4.3. Matriz de decisión . . . . .   | 20        |
| 3.4.4. Definición de las funciones de pérdida basada en el arrepentimiento . . . . .    | 21        |
| 3.5. Diseño, entrenamiento y validación del modelo de aprendizaje . . . . .             | 23        |
| 3.5.1. Arquitectura de la red neuronal . . . . .  | 24        |
| 3.5.2. Entrenamiento del modelo . . . . .   | 26        |

|           |   |           |
|-----------|---|-----------|
| 3.5.3.    | Selección y ajuste de hiperparámetros . . . . .   | 27        |
| 3.5.4.    | Entrenamiento final del modelo y métricas de evaluación . . . . .                             | 29        |
| <b>4.</b> | <b>Caso de estudio</b>  | <b>31</b> |
| 4.1.      | Descripción del caso de estudio . . . . .   | 31        |
| 4.1.1.    | El mercado eléctrico de potencia en Chile . . . . .   | 32        |
| 4.1.2.    | Descripción del subsistema de distribución . . . . .  | 32        |
| 4.2.      | Datos de entrada . . . . .  | 32        |
| 4.2.1.    | Consumo eléctrico del subsistema . . . . .  | 33        |
| 4.2.2.    | Variables meteorológicas . . . . .  | 33        |
| 4.2.3.    | Costos económicos . . . . .   | 36        |
| 4.2.4.    | Horas punta oficiales . . . . .   | 37        |
| 4.3.      | Configuración del experimento . . . . .   | 37        |
| 4.3.1.    | Selección del periodo de análisis . . . . .   | 37        |
| 4.3.2.    | División en conjuntos de entrenamiento, validación y prueba . . . . .                         | 38        |
| 4.3.3.    | Conjunto de hiperparámetros evaluados . . . . .   | 38        |
| 4.4.      | Configuración de los análisis de sensibilidad . . . . .                                       | 39        |
| 4.4.1.    | Configuración del análisis de sensibilidad de los costos económicos . . . . .                 | 39        |
| 4.4.2.    | Configuración del análisis de sensibilidad de las horas punta penalizadas . . . . .           | 40        |
| <b>5.</b> | <b>Resultados y análisis</b>  | <b>41</b> |
| 5.1.      | Desempeño de los modelos . . . . .  | 41        |
| 5.1.1.    | Avisos de los modelos . . . . .   | 41        |
| 5.1.2.    | Análisis conjunto de curvas de demanda, variables climáticas y avisos de predicción . . . . . | 43        |
| 5.1.3.    | Evaluación económica . . . . .  | 45        |
| 5.2.      | Análisis de sensibilidad . . . . .  | 49        |
| 5.2.1.    | Análisis de sensibilidad de los costos . . . . .  | 49        |
| 5.2.2.    | Análisis de sensibilidad de las horas punta penalizadas . . . . .                             | 51        |
| 5.2.3.    | Conclusión conjunta de los análisis de sensibilidad. . . . .                                  | 53        |
| <b>6.</b> | <b>Conclusiones</b>   | <b>55</b> |
| 6.1.      | Conclusiones . . . . .  | 55        |
| 6.2.      | Trabajo futuro . . . . .  | 57        |
|           | <b>Bibliografía</b>   | <b>58</b> |
|           | <b>Anexos</b>   | <b>62</b> |
| A.        | Cálculo de arrepentimientos . . . . .   | 62        |
| A.1.      | Consideraciones iniciales . . . . .   | 62        |
| A.2.      | Cálculo de utilidades . . . . .   | 63        |
| A.3.      | Cálculo de arrepentimientos . . . . .   | 63        |
| B.        | Cálculo de costos variables de generación propia . . . . .                                    | 66        |
| C.        | Resultados . . . . .  | 67        |
| C.1.      | Gráficos de avisos y variables climáticas . . . . .   | 67        |

# Índice de Tablas

|      |   |    |
|------|---|----|
| 2.1. | Tabla resumen del estado del arte. . . . .  | 10 |
| 3.1. | Matriz de decisiones del problema que enfrenta el consumidor sujeto a cobros por demanda punta. . . . .   | 22 |
| 3.2. | Matriz de arrepentimiento para el modelo de predicción. . . . .   | 23 |
| 3.3. | Matriz de arrepentimiento para la variante del modelo de predicción. . . . .  | 23 |
| 4.1. | Costos específicos utilizados por los modelos de predicción. . . . .  | 36 |
| 4.2. | Hiperparámetros elegidos para los modelos de predicción. . . . .  | 39 |
| 5.1. | Matriz de Confusión del modelo enfocado en predicción. . . . .  | 41 |
| 5.2. | Matriz de Confusión del modelo enfocado en costos. . . . .  | 42 |
| 5.3. | Matriz de Confusión del modelo enfocado en costos modificado. . . . .   | 42 |
| 5.4. | Desempeño del modelo en términos de métricas de calidad promedio en la predicción. . . . .  | 42 |
| 5.5. | Desempeño de los modelos en términos de métricas económicas promedio. . . . .   | 46 |
| 5.6. | Métricas económicas de las políticas comparativas. . . . .  | 46 |
| 5.7. | Comparación del costo total de operación frente al modelo de referencia (enfocado en predicción), mediante test t de Welch sobre las 10 réplicas. $\Delta$ es la reducción media de costo, expresada en USD/MW. . . . . | 49 |

# Índice de Ilustraciones

|      |   |    |
|------|---|----|
| 3.1. | Esquema general de la metodología empleada para el desarrollo de los modelos de predicción de horas punta basados en funciones de pérdida asimétricas. . .  | 15 |
| 3.2. | Esquema del problema que enfrenta el consumidor sujeto a cobros por las $k$ horas punta de demanda eléctrica del sistema eléctrico al que está conectado, y que cuenta con un medio de generación local para desplazar su demanda. . . .  | 19 |
| 3.3. | Arquitectura de la red neuronal utilizada para la predicción de horas punta. . .  | 26 |
| 3.4. | Validación cruzada desarrollada para el ajuste de hiperparámetros. . . . .  | 28 |
| 4.1. | Curvas de demanda del subsistema de distribución de la Región Metropolitana de Santiago de Chile entre los años 2018 y 2023, cuyos datos fueron extraídos de la Plataforma de Recepción de Medidas de Transferencias Económicas (PRMTE) del Coordinador Eléctrico Nacional. . . . . | 33 |
| 4.2. | Curvas de las variables meteorológicas en la Región Metropolitana de Santiago de Chile, cuyos datos fueron extraídos desde dos estaciones meteorológicas publicadas por la Dirección Meteorológica de Chile, ubicadas en Quinta Normal y Cerro Colorado. . . . .                    | 34 |
| 5.1. | Curvas de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos modificado. . . . .   | 43 |
| 5.2. | Desglose de los costos de operación promedio de los modelos desarrollados durante todo el periodo de control de horas punta. . . . .  | 47 |
| 5.3. | Análisis de sensibilidad del arrepentimiento como porcentaje del costo total de operación, en función del factor de sensibilidad. La línea punteada indica el caso base. . . . .  | 50 |
| 5.4. | Análisis de sensibilidad del arrepentimiento como porcentaje del costo total de operación, en función del número de horas punta penalizadas $k$ . La línea punteada indica el caso base ( $k = 52$ ). . . . .   | 51 |
| C.1. | Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en predicción. . . . .   | 67 |
| C.2. | Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos. . . . .   | 68 |
| C.2. | Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos (Continuación). . . . .  | 69 |

# Lista de Acrónimos

| <b>Acrónimo</b> | <b>Inglés</b>                            | <b>Español</b>                                 |
|-----------------|--|--|
| LTLF            | Long-Term Load Forecasting               | Pronóstico de demanda a largo plazo            |
| MTLF            | Mid-Term Load Forecasting                | Pronóstico de demanda a mediano plazo          |
| STLF            | Short-Term Load Forecasting              | Pronóstico de demanda a corto plazo            |
| VSTLF           | Very Short-Term Load Forecasting         | Pronóstico de demanda a muy corto plazo        |
| DRNN            | Deep Recurrent Neural Network            | Red neuronal recurrente profunda               |
| TCN             | Temporal Convolutional Network           | Red convolucional temporal                     |
| ARIMA           | Autoregressive Integrated Moving Average | Modelo autorregresivo integrado de media móvil |
| SARIMA          | Seasonal ARIMA                           | ARIMA estacional                               |
| GAM             | Generalized Additive Model               | Modelo aditivo generalizado                    |
| SVM             | Support Vector Machine                   | Máquina de vectores de soporte                 |
| RF              | Random Forest                            | Bosque aleatorio                               |
| GBM             | Gradient Boosting Machine                | Máquina de potenciación por gradiente          |
| SVR             | Support Vector Regression                | Regresión por vectores de soporte              |
| LSTM            | Long Short-Term Memory                   | Memoria de largo corto plazo                   |
| GRU             | Gated Recurrent Unit                     | Unidad recurrente con compuertas               |
| RNN             | Recurrent Neural Network                 | Red neuronal recurrente                        |
| CNN             | Convolutional Neural Network             | Red neuronal convolucional                     |
| DTW             | Dynamic Time Warping                     | Alineamiento temporal dinámico                 |
| FC-DenseNet     | Fully Convolutional Dense Network        | Red densa completamente convolucional          |

| <b>Acrónimo</b> | <b>Inglés</b>                                    | <b>Español</b>  |
|-----------------|--|---|
| ReLU            | Rectified Linear Unit                            | Unidad lineal rectificada                                       |
| Adam            | Adaptive Moment Estimation                       | Estimación del momento adaptativo                               |
| RMSE            | Root Mean Squared Error                          | Raíz del error cuadrático medio                                 |
| MAPE            | Mean Absolute Percentage Error                   | Error porcentual absoluto medio                                 |
| MSE             | Mean Squared Error                               | Error cuadrático medio  |
| CDF             | Cumulative Distribution Function                 | Función de distribución acumulada                               |
| BCE             | Binary Cross Entropy                             | Entropía cruzada binaria  |
| PIB             | Gross Domestic Product (GDP)                     | Producto interno bruto  |
| ESS             | Energy Storage System                            | Sistema de almacenamiento de energía                            |
| BESS            | Battery Energy Storage System                    | Sistema de almacenamiento en baterías                           |
| EV              | Electric Vehicle                                 | Vehículo eléctrico  |
| PV              | Photovoltaic                                     | Fotovoltaico  |
| CP              | Coincident Peak                                  | Pico coincidente  |
| CEN             | National Electric Coordinator (Chile)            | Coordinador Eléctrico Nacional                                  |
| CNE             | National Energy Commission (Chile)               | Comisión Nacional de Energía                                    |
| SEN             | National Electric System (Chile)                 | Sistema Eléctrico Nacional                                      |
| PRMTE           | Transfer Payments Measurement Reception Platform | Plataforma de Recepción de Medidas de Transferencias Económicas |
| PNEP            | Average Node Energy Price                        | Precio Nudo de Energía Promedio                                 |
| USD             | United States Dollar                             | Dólar estadounidense  |
| SFC             | Specific Fuel Consumption                        | Consumo específico de combustible                               |
| ENAP            | National Petroleum Company (Chile)               | Empresa Nacional del Petróleo                                   |
| VP              | True Positive (TP)                               | Verdadero positivo  |
| VN              | True Negative (TN)                               | Verdadero negativo  |
| FP              | False Positive (FP)                              | Falso positivo  |
| FN              | False Negative (FN)                              | Falso negativo  |

# Capítulo 1

## Introducción

### 1.1. Motivación

En varios países del mundo existen mercados de potencia cuyo objetivo es remunerar la infraestructura eléctrica necesaria para garantizar la suficiencia de capacidad y transmisión, especialmente durante las horas de mayor estrés del sistema eléctrico [1]. Para ello, estos mercados establecen que los usuarios paguen por la infraestructura que les permite acceder a la energía, exponiéndolos a un cobro por demanda punta, el cual consiste en un pago adicional en función de los consumos realizados en las horas de mayor demanda del sistema al que se encuentran conectados.

Estos cargos se aplican principalmente a clientes medianos y grandes, y operan mediante la aplicación de un precio adicional por el consumo en las horas punta, a menudo varias veces superior a la tarifa base, llegando a representar, en algunos países o jurisdicciones, más del 20% de la tarifa eléctrica anual de un usuario del sistema [2]. Esto ha llevado a que los clientes busquen estrategias para controlar y reducir su consumo durante las horas sujetas al cobro por demanda punta, requiriendo herramientas que les permitan tomar decisiones orientadas a disminuir sus costos eléctricos a través de una gestión eficiente de su consumo.

Este desafío se intensifica en un mercado cada vez más complejo, debido a la creciente participación de energías renovables, la incertidumbre en la demanda y otros factores. En este contexto, la identificación de las horas punta de demanda eléctrica se presenta como una tarea compleja, lo que hace que los modelos de predicción constituyan una herramienta valiosa al permitir anticipar con antelación las horas con mayor probabilidad de cobro y, en consecuencia, ajustar el consumo de forma estratégica.

A esta complejidad se suma una dificultad de carácter estructural. Las horas punta se definen una única vez por periodo de facturación sobre la demanda agregada del sistema, por lo que la cantidad de eventos etiquetados disponibles para construir un modelo es intrínsecamente escasa. Esta condición no es propia de un mercado en particular, sino que es común a cualquier mecanismo de cargo por capacidad, y limita severamente la aplicación directa de técnicas de aprendizaje supervisado que requieren un volumen elevado de observaciones. Por este motivo, anticipar las horas punta no puede entenderse únicamente como una tarea de

predicción, sino que exige además una formulación que haga viable el aprendizaje a partir de un número reducido de eventos de interés.

Ahora bien, anticipar con precisión las horas punta y decidir con el menor costo posible no constituyen necesariamente el mismo objetivo. Las metodologías habitualmente empleadas en estos modelos recurren a funciones de pérdida simétricas, tales como RMSE, MAE o BCE, que no distinguen entre el costo asociado a un falso positivo y a un falso negativo, penalizando ambos errores con el mismo peso. Este enfoque resulta adecuado para fines puramente predictivos, pero no refleja los costos diferenciales ni la calidad de las decisiones que se toman a partir de estas estimaciones [3].

No obstante, cuando las predicciones sirven como base para la toma de decisiones, los costos asociados a ambos tipos de error son inherentemente asimétricos, ya que en este contexto los falsos negativos generan cobros adicionales por contribuir a la demanda máxima durante horas punta, mientras que los falsos positivos pueden implicar usos innecesarios de recursos para desplazar o reducir demanda, lo que generalmente son costos relativamente menores.

Frente a un *peak*, y dado que su carga máxima no necesariamente coincide con las horas punta del sistema, el cliente dispone de dos estrategias principales: reducir su carga, apagando o disminuyendo el uso de equipos no esenciales, o desplazar su consumo mediante generación o almacenamiento local, recurriendo a grupos electrógenos, paneles fotovoltaicos o sistemas de almacenamiento de energía (ESS) [2]. La elección entre estas alternativas, en cada hora, es precisamente la decisión que el modelo busca informar.

En este contexto, se propone desarrollar un modelo de predicción de horas punta orientado a la toma de decisiones económicas. Para abordar la escasez estructural de datos, el problema de seleccionar las  $k$  horas de máxima demanda se reformula como una clasificación binaria a nivel horario, lo que multiplica el número de instancias de entrenamiento disponibles. Sobre esta base, el modelo incorpora explícitamente los costos asociados a las decisiones del cliente mediante una función de pérdida asimétrica basada en la minimización del arrepentimiento, integrando la asimetría de los errores tanto en el entrenamiento como en la predicción.

De este modo, se espera desarrollar una herramienta que no solo entregue predicciones precisas, sino que también facilite la toma de decisiones económicamente eficientes. Con ello, se busca demostrar que es posible formular modelos cuyo enfoque trascienda a aquellos puramente predictivos y que, en cambio, conduzcan a mejores decisiones desde el punto de vista económico.

## 1.2. Hipótesis

En el contexto de mercados de potencia, un modelo de predicción de horas punta entrenado con un objetivo orientado a costos conduce a decisiones económicamente más eficientes que las derivadas de un modelo de predicción convencional, entrenado con una función de pérdida simétrica.

## 1.3. Objetivos

A continuación, se presentan los objetivos generales y específicos que se buscan desarrollar en este trabajo.

### 1.3.1. Objetivos generales

Desarrollar un modelo de predicción de horas punta de demanda eléctrica, basado en redes neuronales y entrenado mediante una función de pérdida asimétrica orientada a minimizar el arrepentimiento acumulado asociado a las decisiones de un consumidor de electricidad, con el fin de evaluar su desempeño económico en comparación con un modelo de predicción convencional.

### 1.3.2. Objetivos específicos

- Modelar el problema de decisión al que se enfrenta un usuario conectado a un subsistema eléctrico sujeto a cobros por demanda punta, considerando los elementos técnicos y económicos asociados.
- Diseñar e implementar una función de pérdida asimétrica basada en la minimización del arrepentimiento, que refleje los costos diferenciales asociados a las decisiones del cliente.
- Desarrollar un modelo de predicción de horas punta de demanda eléctrica mediante redes neuronales entrenado con la función de pérdida propuesta.
- Evaluar y comparar el desempeño del modelo propuesto frente a un modelo de predicción convencional, utilizando métricas económicas sobre un subsistema eléctrico representativo.

### 1.3.3. Estructura de la tesis

El documento se divide en seis capítulos. El primero presenta la motivación, los objetivos de esta investigación y la hipótesis de trabajo. En el Capítulo 2 se expone el estado del arte sobre los modelos de predicción de demanda eléctrica, con énfasis en la detección de horas punta, y sobre el uso de funciones de pérdida asimétricas en problemas de clasificación, y se sintetizan las contribuciones de este trabajo respecto de dicho estado del arte. El Capítulo 3 describe la metodología implementada, abordando la selección y el procesamiento de las variables de entrada, el diseño de la función de pérdida basada en el arrepentimiento y el diseño, entrenamiento y validación del modelo de aprendizaje. En el Capítulo 4 se presenta el caso de estudio, detallando el mercado eléctrico de potencia chileno, los datos de entrada utilizados y la configuración del experimento. El Capítulo 5 expone los resultados obtenidos junto con su respectivo análisis, incluyendo la evaluación económica de las decisiones generadas por cada modelo, el análisis de sensibilidad frente a variaciones en los costos del caso de estudio y el análisis de sensibilidad respecto al número de horas punta penalizadas ( $k$ ). Finalmente, el Capítulo 6 indica las conclusiones principales, acompañadas de las directrices que pueden orientar futuras investigaciones en el área.

# Capítulo 2

## Estado del arte

### 2.1. Modelos de predicción de demanda eléctrica

La previsión de demanda eléctrica siempre ha sido un tópico muy importante en la gestión energética de los sistemas eléctricos modernos. Los operadores y participantes del mercado eléctrico utilizan estas herramientas para la toma de decisiones relacionadas al despacho o para la reserva de giro de las unidades de generación [4][5]. Mientras que los planificadores hacen uso de previsiones a largo plazo para identificar la necesidad de planificar e invertir en generación según cómo aumente la tendencia de consumo eléctrico de una región o de un país [5]. Debido al carácter de las decisiones anteriores, el aumento de pequeños puntos en la precisión de las predicciones puede traducirse en beneficios significativos, de millones de dólares, tanto en la operación como en la planificación del sistema eléctrico [6].

Dependiendo de su aplicación, los modelos de predicción de demanda eléctrica se clasifican en metodologías de largo plazo (LTLF, superiores a un año), mediano plazo (MTLF, entre varios meses y un año), corto plazo (STLF, entre varias horas y días) y muy corto plazo (VSTLF, entre varios minutos y horas) [7].

En [8] se propone un modelo de predicción de carga energética a corto plazo utilizando redes neuronales recurrentes profundas (DRNN) y un proceso de filtrado de muestras. El filtrado de muestras se logra mediante el uso de agrupamiento *K-Means*, que determina el número de clústeres a utilizar en el modelo. Este enfoque muestra una mejora significativa en la precisión de la predicción.

En [9] presenta una evaluación comparativa de cinco técnicas de pronóstico de carga a corto plazo: Regresión Lineal Múltiple, Series Temporales Estocásticas, Suavizado Exponencial Generalizado, Método de Espacio de Estados, Enfoque Basado en Conocimiento. Estas técnicas se aplicaron para obtener pronósticos horarios de carga eléctrica (hasta 24 horas) durante las temporadas de máxima demanda en invierno y verano. El estudio demuestra que el método de espacio de estados ofrece una precisión superior en comparación con los otros enfoques evaluados.

En [10] propone un modelo de predicción de carga eléctrica a corto plazo que combina una red neuronal convolucional temporal (TCN) con mecanismos de atención espacial y temporal.

Este enfoque permite al modelo enfocarse dinámicamente en las características más relevantes de los datos de entrada, mejorando así la precisión de las predicciones. El modelo propuesto se evaluó utilizando datos de consumo eléctrico y meteorológicos, mostrando un rendimiento superior en comparación con modelos tradicionales.

### 2.1.1. Modelos de predicción de horas punta de demanda eléctrica

La predicción de horas punta de demanda eléctrica (horas punta o *peaks*) comparte varias similitudes con la predicción de demanda general; sin embargo, la diferencia sustancial radica en las variables de salida. En este ámbito, se distinguen dos enfoques principales para la detección de *peaks*: el primero busca estimar la magnitud de la máxima demanda, ya sea diaria, mensual, anual o estacional, mientras que el segundo se orienta a identificar las horas específicas que conforman el bloque de mayor demanda [11].

Por otro lado, las variables de entrada de estos modelos suelen agruparse en dos grandes categorías: endógenas y exógenas. Las primeras corresponden a aquellas que se derivan directamente de la propia serie temporal de demanda, como los valores históricos, las tendencias y los patrones estacionales [11]. Las segundas, en cambio, representan factores externos que influyen en la demanda eléctrica, pero que no forman parte intrínseca de la serie temporal [11].

Dentro de las variables exógenas más utilizadas destacan las condiciones climáticas (temperatura, humedad relativa, velocidad del viento, presión atmosférica, irradiación solar, horas de luz), las variables de calendario (hora del día, día de la semana, mes, feriados), los eventos masivos (deportivos, culturales) y los factores económicos (PIB, crecimiento poblacional, precio de la electricidad). En términos generales, los pronósticos de corto plazo se apoyan principalmente en variables climáticas, de calendario y en la ocurrencia de eventos masivos; mientras que en los de mediano y largo plazo cobran mayor relevancia los factores económicos, orientados a capturar las tendencias estructurales del crecimiento económico y demográfico [11].

Entre los enfoques más utilizados para la predicción de horas punta destacan los métodos estadísticos, como ARIMA, SARIMA y los modelos de suavizamiento exponencial, así como los modelos de regresión aditiva generalizada (GAM) y sus variantes cuantílicas [11]. También se emplean algoritmos de aprendizaje automático, tales como redes neuronales artificiales, máquinas de soporte vectorial (SVM), regresión de vectores de soporte (SVR), bosques aleatorios (RF) y potenciación del gradiente (GBM). Más recientemente, se han incorporado arquitecturas de *deep learning*, incluyendo memoria larga a corto plazo (LSTM), unidades recurrentes cerradas (GRU) y redes convolucionales temporales (TCN). Algunos estudios emplean modelos híbridos, que combinan métodos tradicionales de series de tiempo con técnicas de aprendizaje automático.

En particular, en [12] se propone una comparación exhaustiva de varios modelos para la predicción de la demanda eléctrica máxima. Los modelos evaluados incluyen ARIMA, LSTM, GRU, así como combinaciones híbridas ARIMA-LSTM y ARIMA-GRU. El estudio se centra en la evaluación de la precisión de estos modelos en la predicción de la carga eléctrica máxima, proporcionando una visión integral de su rendimiento en este contexto.

Luego, en [13] propone un modelo híbrido entre redes neuronales recurrentes (RNN) y unidades de puerta (GRU) para la predicción de la carga máxima diaria en sistemas eléctricos. Se aplica la técnica *Dynamic Time Warping* (DTW) para mejorar la precisión de las predicciones. El enfoque se evalúa con un conjunto de datos de carga de la *European Network on Intelligent Technologies* (EUNITE), mostrando resultados satisfactorios en comparación con otros métodos existentes.

En [14] se propone un algoritmo híbrido para la predicción de la demanda máxima de electricidad. Este algoritmo combina el método de SVR con técnicas de lógica difusa, permitiendo una adaptación en línea a los cambios en los patrones de carga. La metodología busca mejorar la precisión de las predicciones en entornos dinámicos y con datos ruidosos. El estudio demuestra que el enfoque propuesto puede ser eficaz para la predicción de la carga máxima en sistemas de energía eléctrica.

En [15] se propone un enfoque de aprendizaje automático supervisado utilizando RF, GBM y regresión logística para predecir la probabilidad de que el próximo día operativo contenga la hora punta del mes y la probabilidad de que una hora específica sea la hora punta del día. El estudio se aplica al sistema *Duke Energy Progress* y logra capturar 69 días punta de 72 meses de prueba con un umbral de probabilidad de excedencia del 3%. En el 90% de los días punta, la hora real de máxima demanda se encuentra entre las dos horas con las probabilidades más altas.

En [16] se propone un marco único para apoyar a los tomadores de decisiones en la gestión diaria de las actividades eléctricas, prestando especial atención al seguimiento de la demanda máxima mensual. Este enfoque integra avances en aprendizaje automático y dinámica de sistemas, con el potencial de contribuir a pronósticos más precisos que ayuden a desarrollar estrategias para abordar las variaciones entre la oferta y la demanda.

En [17] se propone un modelo híbrido para la predicción de la carga eléctrica a corto plazo. Este modelo combina redes neuronales convolucionales, unidades recurrentes con puertas y un mecanismo de atención para mejorar la precisión de las predicciones. El modelo se aplica a un conjunto de datos que incluye información de carga eléctrica regional y datos meteorológicos, mostrando mejoras significativas en comparación con otros modelos comunes.

En [18] se diseña un sistema en tiempo real que coordina *Battery Energy Storage Systems* (BESS), vehículos eléctricos (EVs) y paneles solares fotovoltaicos (PV) para gestionar la demanda pico. Se modela usando predicción del día siguiente mediante modelos ARMA y redes neuronales, y se analiza qué tan perjudiciales son los errores en esas predicciones para la gestión energética. Se observa que los errores pueden exceder el 300%, generando ciclos de carga-descarga innecesarios, aumentando los costos y reduciendo la vida útil de las baterías. También se explora una técnica de detección de incertidumbre utilizando variables ambientales como irradiación solar, temperatura y humedad.

Finalmente, en [19] se propone una metodología para pronosticar la densidad de la demanda máxima de electricidad a largo plazo (hasta diez años) mediante modelos aditivos semiparamétricos. Estos modelos son utilizados para estimar las relaciones entre la demanda y las variables de interés, incluidas las temperaturas, los efectos de calendario y algunas variables demográficas y económicas. Se pronostican las distribuciones de la demanda mediante una combinación de simulación de temperatura, supuestos escenarios económicos futuros y

*bootstrap* residual. Este enfoque se ha aplicado específicamente al caso de Australia del Sur.

### 2.1.2. Modelos de predicción de horas punta coincidentes de demanda eléctrica

En la literatura relacionada a la predicción de horas punta de demanda eléctrica, un subcampo específico y menos explorado corresponde a la predicción de horas punta coincidentes (*Coincident Peak*, CP). Estas se definen como las horas en las que la demanda eléctrica alcanza su punto máximo a nivel del sistema, y resultan de especial relevancia ya que son utilizadas para la asignación de cargos por capacidad y transmisión a los usuarios finales, en función de su contribución a dicha demanda máxima.

Para la predicción de horas punta coincidentes, en [20] se presenta un modelo híbrido para predecir los días de máxima demanda eléctrica durante un periodo de facturación. El modelo propuesto combina ARIMA, regresión logística y redes neuronales artificiales para evaluar los resultados individuales de estos modelos estadísticos y de aprendizaje automático, con el fin de predecir si un día determinado será un día de carga máxima para el periodo de facturación. La implementación del modelo mostró una precisión del 70 % en la predicción de los días de máxima carga real, lo que permitió identificar posibles ahorros de aproximadamente 80.000 USD para una universidad estadounidense durante un periodo de prueba de un año.

En [21] se presenta un modelo de red neuronal *feed-forward* para predecir eventos de CP en una región o sistema con 24 horas de antelación. La red neuronal se entrena para predecir la función de distribución acumulativa (CDF) estimada de la demanda del sistema basándose en datos históricos y meteorológicos. Los resultados muestran que la red neuronal ofrece una mejora significativa en la precisión de la predicción de eventos CP con respecto a modelos tradicionales, lo cual permite generar decisiones de reducción de consumo lo que puede llevar a una reducción efectiva de costos asociados a estos picos de demanda. Un hallazgo importante en este estudio es que se declara formalmente el problema del consumidor sujeto a cargos por CP, planteado como un problema de maximización de las utilidades del cliente.

En [22] se propone un modelo de series temporales para la previsión de la carga eléctrica horaria a corto plazo, con capacidad para estimar la carga máxima diaria. Este enfoque integra la experiencia de operadores expertos para mejorar la precisión de las predicciones, superando a técnicas convencionales como redes neuronales artificiales y modelos Box-Jenkins.

En [1] se propone un enfoque basado en estimadores de Monte Carlo para predecir eventos de CP a nivel diario y horario, utilizando datos reales de carga y pronósticos de diferentes jurisdicciones. Además, se presenta un análisis retrospectivo de estrategias con umbrales adaptativos para la predicción, y se derivan implicaciones prácticas para la reducción de carga mediante sistemas de almacenamiento de energía en baterías (BESS).

En [23] se presenta una comparación entre varios modelos de clasificación, entre los que se incluye *Naïve Bayes*, SVM, *Random Forest*, *AdaBoost*, CNN, LSTM y *Stacked Autoencoder*, con el objetivo de determinar cuál es el mejor predictor de la hora exacta en que se alcanza el pico de demanda eléctrica coincidente diaria con 24 horas de antelación en una red de distribución. Para el estudio, se analizaron datos históricos de demanda y clima, los cuales se dividieron en dos conjuntos según si correspondían al periodo de invierno o de verano.

Los resultados indicaron que el modelo SVM fue el segundo con mejor rendimiento, siendo superado únicamente por LSTM.

Por último, en [24] se propone un modelo de predicción de la demanda máxima de energía basado en inteligencia artificial, incorporando características económicas y climáticas. El estudio destaca la importancia de una previsión precisa de la demanda máxima mensual para gestionar eficientemente la programación de combustible, el mantenimiento de plantas de energía y la expansión de la infraestructura de transmisión. Se utiliza un conjunto de datos integral que incluye características económicas, climáticas y el número de consumidores, lo que mejora la precisión de la previsión en comparación con modelos existentes. El modelo de regresión de soporte vectorial (SVR) demostró ser el más preciso, con un R-cuadrado de 99 %, un RMSE de 28 y un MAPE de 0.1355.

## 2.2. Modelos con funciones de pérdida asimétricas

A diferencia de las funciones de pérdida simétricas, que penalizan de manera uniforme los errores positivos y negativos, las funciones de pérdida asimétricas permiten asignar diferentes pesos a estos errores, reflejando así la realidad de muchos problemas prácticos donde ciertos errores pueden tener consecuencias más graves que otros.

En la literatura existen varios trabajos que han explorado el uso de funciones de pérdida asimétricas en diversos contextos. En [25] se propone el uso de funciones de pérdida asimétricas en modelos de predicción de demanda eléctrica para optimizar la capacidad eléctrica contratada. Dado que los consumidores de alta tensión enfrentan costos eléctricos que dependen no solo de la energía consumida, sino también de la capacidad contratada, es crucial predecir con precisión la demanda futura para evitar costos innecesarios. El estudio demuestra que las funciones de pérdida asimétricas, al penalizar de manera diferenciada los errores de sobreestimación y subestimación, mejoran la precisión de las predicciones y, por ende, la optimización de la capacidad contratada.

En [26] propone un enfoque innovador para la segmentación de imágenes médicas altamente desequilibradas, como la detección de lesiones en la esclerosis múltiple. El estudio presenta una red neuronal convolucional completamente conectada (FC-DenseNet) que utiliza una función de pérdida asimétrica basada en el índice de Tversky para mitigar el problema del desequilibrio de clases. Este enfoque permite una segmentación más precisa de las lesiones, mejorando la tasa de verdaderos positivos y reduciendo la tasa de falsos negativos. Los resultados obtenidos muestran un rendimiento superior en comparación con métodos tradicionales, destacando la eficacia de la función de pérdida asimétrica en escenarios de segmentación con clases desbalanceadas.

En [27] propone una función de pérdida personalizada que mejora significativamente la predicción de los rendimientos de activos financieros. Se destaca que los errores en la predicción de rendimientos no son iguales en términos de impacto financiero, y por lo tanto, una función de pérdida estándar como el error cuadrático medio (MSE) puede no ser la más adecuada. La función de pérdida propuesta tiene en cuenta la asimetría de los errores, penalizando más fuertemente los errores que podrían inducir decisiones de inversión desfavorables. Los resultados experimentales muestran que esta función de pérdida mejora la eficiencia de

los algoritmos de predicción de rendimientos, ofreciendo una mejor relación riesgo-retorno.

Existen otros trabajos que desarrollan funciones de pérdida asimétrica para abordar el problema de la toma de decisiones bajo incertidumbre mediante la minimización del arrepentimiento (*regret minimization*). En [3] se propone un enfoque innovador para la prevención de la deserción de clientes en empresas, utilizando un método predictivo y de optimización. El objetivo principal es maximizar el beneficio neto al identificar y retener a los clientes más valiosos, en lugar de centrarse únicamente en las probabilidades de deserción. El modelo desarrollado se basa en la minimización del arrepentimiento y puede resolverse eficientemente mediante métodos de descenso de gradiente estocástico. Los resultados obtenidos de 12 conjuntos de datos de predicción de deserción destacan la efectividad del enfoque propuesto, superando a otras estrategias bien establecidas en términos de beneficio promedio.

En [28] se presenta una nueva función de pérdida denominada *Empirical Soft Regret* (ESR), diseñada para mejorar significativamente el rendimiento en tareas de toma de decisiones contextualizadas. A diferencia de las métricas tradicionales como el error cuadrático medio, que se centran en la precisión de las predicciones, ESR se enfoca en minimizar el arrepentimiento asociado a decisiones subóptimas. Dado que el arrepentimiento no es diferenciable, los autores proponen una versión diferenciable de este término, permitiendo su integración en modelos de aprendizaje automático basados en gradientes. El enfoque propuesto supera significativamente a otros métodos de vanguardia en problemas reales de toma de decisiones, como la recomendación de noticias y la atención médica personalizada.

## 2.3. Resumen del estado del arte

En la Tabla 2.1 se presentan, a modo de resumen, las principales contribuciones de los trabajos más relevantes, destacando los aspectos distintivos de sus modelos y alcances.

Tabla 2.1: Tabla resumen del estado del arte.

| Referencia | ¿Predice una hora punta coincidente? | Modelo utilizado  | ¿Utiliza una función de pérdida asimétrica? | Contribuciones   |
|------------|--------------------------------------|---|---|--|
| [1]        | Sí                                   | Generación de escenarios y estimadores de Monte Carlo                         | No  | Genera implicaciones prácticas en base a las predicciones del modelo mediante soluciones de almacenamiento en baterías (BESS).           |
| [17]       | No                                   | Red neuronal convolucional (CNN) y unidades recurrentes con puertas (GRU)     | No  | Utiliza variables climáticas y de calendario como características del modelo, además de implementar validación cruzada y early stopping. |
| [20]       | Sí                                   | Modelos ARIMA, regresión logística (RL) y redes neuronales artificiales (ANN) | No  | Genera implicaciones prácticas en base a las predicciones del modelo mediante acciones de respuesta a la demanda.                        |
| [21]       | Sí                                   | Red neuronal feedforward (FNN)  | No  | Enuncia formalmente el problema del consumidor sujeto a cargos por hora punta, sin medios de generación propia.                          |
| [26]       | No                                   | Redes neuronales convolucionales densamente conectadas (DenseNet)             | Sí  | Incorpora funciones de pérdida asimétricas para reflejar diferencias en las decisiones.  |
| [28]       | No                                   | Red neuronal densamente conectada (FCNN)                                      | Sí  | Incorpora funciones de pérdida asimétricas basadas en arrepentimiento para reflejar diferencias en las decisiones.                       |

## 2.4. Contribuciones

Considerando los objetivos planteados, las contribuciones de este trabajo al estado del arte son las siguientes:

- Se formula el problema de decisión de un consumidor sujeto a cargos por demanda punta como un problema de optimización de utilidades, en el que el cliente decide el uso de un medio de generación propia para desplazar su consumo en las horas pronosticadas como punta y así evitar los costos de facturación asociados. Esta formulación integra de manera explícita los elementos técnicos y económicos del cargo por capacidad en la representación del problema.
- Se propone una reformulación del problema de identificación de las  $k$  horas punta que, en lugar de abordarlo como una selección estructurada sobre la totalidad del periodo de control, se plantea como un problema de clasificación binaria puntual a nivel horario. Esta decisión responde a que, en los mercados de cargo por capacidad, las horas punta se definen sobre todo el periodo de facturación, de modo que el problema ofrece muy pocos casos para entrenar, lo que es en esencia uno por año. Esto constituye una limitación estructural del mecanismo y no propia del caso chileno estudiado. La reformulación sacrifica parte de la estructura de competencia global entre horas, pero incrementa en varios órdenes de magnitud el número de observaciones disponibles, habilitando el entrenamiento de redes neuronales pese a la escasez de años de registro.
- Se proponen dos funciones de pérdida asimétrica, basadas en la minimización del arrepentimiento acumulado de las decisiones tomadas a partir de las predicciones, que penalizan de forma diferenciada los errores de falsos positivos y falsos negativos según sus costos económicos reales.
- Se implementa y valida una herramienta de predicción de horas punta de demanda eléctrica que integra explícitamente los costos asociados a los errores de predicción mediante las funciones de pérdida asimétricas diseñadas, evaluando su desempeño económico frente a un modelo de predicción convencional en el mercado de potencia chileno mediante métricas de arrepentimiento y costo operacional.

# Capítulo 3

## Metodología

En este capítulo se presenta la metodología empleada para desarrollar el modelo de predicción de horas punta de demanda eléctrica propuesto. En primera instancia se enuncia la nomenclatura, donde se describen brevemente los conjuntos, variables y parámetros utilizados. Posteriormente, se realiza una descripción general del modelo y de todos sus módulos, y, finalmente, se describen detalladamente los módulos del modelo, incluyendo la selección de variables y el procesamiento de datos de entrada, el diseño de la función de pérdida y el diseño del modelo de predicción.

### 3.1. Nomenclatura

#### 3.1.1. Conjuntos e índices

| Símbolo                  | Descripción  | Unidad |
|--------------------------|--|--------|
| $\tau$                   | Conjunto de horas del periodo de control.                          | -      |
| $\{S_1, \dots, S_\tau\}$ | Conjunto de demandas totales del sistema en el periodo de control. | -      |

#### 3.1.2. Parámetros

| Parámetro    | Descripción  | Unidad  |
|--------------|--|---------|
| $p^{max}$    | Potencia máxima consumida por el cliente.                    | MW      |
| $p^{max-cs}$ | Potencia máxima extraída desde la red.                       | MW      |
| $p^{max-gp}$ | Potencia máxima aportada por el sistema de generación local. | MW      |
| $\pi^{cp}$   | Cargo por consumo en horas puntas.                           | USD/MW  |
| $\pi^{cs}$   | Cargo por consumo de energía desde la red.                   | USD/MWh |
| $\pi^{gp}$   | Costos por uso del sistema de generación local.              | USD/MWh |

### 3.1.3. Variables del problema

| Variable   | Descripción   | Unidad |
|------------|---|--------|
| $p_t$      | Potencia consumida por el cliente en la hora $t$ .                    | MWh    |
| $p_t^{cs}$ | Potencia extraída desde la red por el cliente en la hora $t$ .        | MWh    |
| $p_t^{gp}$ | Potencia aportada por el sistema de generación local en la hora $t$ . | MWh    |

### 3.1.4. Variables de entrada

| Variable      | Descripción   | Unidad |
|---------------|---|--------|
| $X_t^{Dem}$   | Demanda de potencia total del subsistema en la hora $t$ . | MW     |
| $X_t^{Tem}$   | Temperatura del subsistema en la hora $t$ .               | °C     |
| $X_t^{Hum}$   | Humedad relativa del subsistema en la hora $t$ .          | %      |
| $X_t^{Pres}$  | Presión atmosférica del subsistema en la hora $t$ .       | hPa    |
| $X_t^{Sens}$  | Sensación térmica del subsistema en la hora $t$ .         | °C     |
| $X_t^{Hora}$  | Hora del día en la hora $t$ .                             | -      |
| $X_t^{Día}$   | Día del mes en la hora $t$ .                              | -      |
| $X_t^{Mes}$   | Mes del año en la hora $t$ .                              | -      |
| $X_t^{Pand}$  | Indicador de si es pandemia en la hora $t$ .              | -      |
| $X_t^{Viern}$ | Indicador de si es viernes en la hora $t$ .               | -      |

### 3.1.5. Variables de salida

| Variable              | Descripción   | Unidad |
|-----------------------|---|--------|
| $y_t^{peak}$          | Variable binaria que indica si la hora $t$ es <i>peak</i> .                             | {0,1}  |
| $\hat{y}_t^{peak-p}$  | Predicción de hora <i>peak</i> en la hora $t$ del modelo enfocado en predicción.        | {0,1}  |
| $\hat{y}_t^{peak-c}$  | Predicción de hora <i>peak</i> en la hora $t$ del modelo enfocado en costos.            | {0,1}  |
| $\hat{y}_t^{peak-cm}$ | Predicción de hora <i>peak</i> en la hora $t$ del modelo enfocado en costos modificado. | {0,1}  |

### 3.1.6. Funciones

| Variable     | Descripción   | Unidad |
|--------------|---|--------|
| $r_n(\{S\})$ | Función de clasificación que produce el subconjunto de índices $t$ de los $n$ valores más grandes dentro de un conjunto $\{S\}$ | -      |

## 3.2. Descripción general

Se propone un modelo de predicción de horas punta de demanda eléctrica, cuyo objetivo consiste en identificar de manera *ex-ante* las  $k$  horas de máxima demanda eléctrica dentro de un sistema o subsistema eléctrico específico, incorporando en las predicciones los costos asociados a las decisiones de los usuarios eléctricos en base a estas estimaciones. El modelo se basa en redes neuronales y es entrenado mediante una función de pérdida asimétrica, diseñada para minimizar el arrepentimiento acumulado asociado a las decisiones de un usuario que enfrenta cargos por demanda punta. Para abordar este problema, se desarrolla un enfoque estructurado en tres módulos principales: (i) la selección de variables y el procesamiento de datos de entrada, (ii) el diseño de las funciones de pérdida asimétricas centradas en la minimización del arrepentimiento, y (iii) el diseño, entrenamiento y validación del modelo de predicción.

El primer módulo tiene como objetivo preparar la base de datos adecuada para el entrenamiento del modelo de predicción. Para ello, se seleccionan las variables de entrada que influyen en los máximos de demanda eléctrica y se define la variable de salida del modelo. A continuación, se recopilan los datos correspondientes a estas variables y se aplican procesos de normalización, generación de ventanas deslizantes y filtrado de datos con el objetivo de construir una base de datos apta para el entrenamiento del modelo.

El segundo módulo se enfoca en el diseño de la función de pérdida asimétrica basada en el arrepentimiento. Este proceso comienza con la formulación del problema que enfrenta un consumidor sujeto a cobros por demanda punta. Para ello, se definen sus utilidades por cada hora durante el periodo de control de horas punta, considerando para ello los costos de generación local, las tarifas por consumo desde la red y los cargos por consumo en hora punta. A partir de esta formulación, se construyen las matrices de decisión y de arrepentimiento del problema, y finalmente se diseñan las funciones de pérdida que incorporan explícitamente estas consideraciones económicas y con la que se entrenará al modelo de predicción.

Por último, el tercer módulo aborda el diseño, entrenamiento y validación del modelo de predicción. Este modelo, basado en algoritmos de aprendizaje de máquinas, emplea la arquitectura de una red neuronal que utiliza las variables seleccionadas y los datos procesados en el primer módulo para predecir las  $k$  horas punta de demanda eléctrica. El modelo se entrena utilizando la función de pérdida asimétrica desarrollada en el segundo módulo, y se ajusta hiperparamétricamente mediante técnicas de validación cruzada. Finalmente, su desempeño se evalúa utilizando métricas adecuadas como la precisión y el arrepentimiento acumulado, lo que permite medir tanto la calidad predictiva como el impacto económico de las decisiones generadas por el modelo.

En la Figura 3.1 se muestra un esquema general de la metodología empleada para el desarrollo de los modelos de predicción de horas punta basados en funciones de pérdida asimétricas.

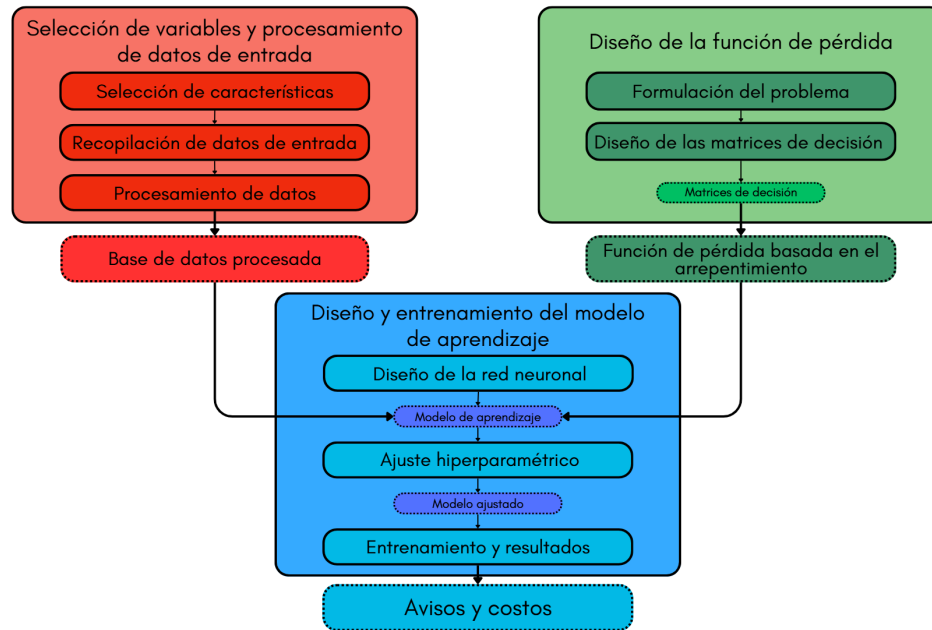


Figura 3.1: Esquema general de la metodología empleada para el desarrollo de los modelos de predicción de horas punta basados en funciones de pérdida asimétricas.

### 3.2.1. Implementación de los modelos de predicción

La implementación de los modelos de predicción se llevó a cabo utilizando el lenguaje de programación Python (versión 3.9.2), empleando las bibliotecas TensorFlow (2.18.0) y Keras (3.8.0) para la construcción y entrenamiento de las redes neuronales. Además, se utilizó la biblioteca Pandas para la manipulación y análisis de datos, y Matplotlib para la visualización de resultados. Por último, se utilizó un equipo con un procesador Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 32 GB de RAM, 6 núcleos físicos y 12 procesadores lógicos para llevar a cabo el entrenamiento y la evaluación de los modelos.

## 3.3. Selección de variables y procesamiento de datos de entrada

A continuación, se describe el proceso de selección de las variables de entrada y salida del modelo de predicción, junto con el procesamiento de los datos necesarios para entrenar el modelo.

### 3.3.1. Selección de variables del modelo

La selección adecuada de variables explicativas constituye un componente central en el diseño de modelos de predicción de los máximos de demanda eléctrica, ya que determina su capacidad para capturar patrones temporales, meteorológicos y de comportamiento asociados

al consumo energético [29][30]. La literatura en STLF y VSTLF ha demostrado consistentemente que tanto las variables numéricas, como la demanda histórica o la temperatura, así como las variables categóricas, relativas a información temporal o eventos específicos, contribuyen significativamente a mejorar el desempeño predictivo, especialmente en la detección de picos de demanda [31][32]. Asimismo, la representación de variables categóricas mediante técnicas de codificación como *One-Hot Encoding* se ha consolidado como una práctica estándar para evitar la introducción de relaciones ordinales artificiales entre categorías [33].

En primer lugar, se selecciona como variable numérica del modelo a la demanda eléctrica total del subsistema eléctrico ( $X_t^{Dem}$ ), que representa la demanda total de potencia en el subsistema eléctrico elegido en la hora  $t \in \{1, \dots, \tau\}$ . Este atributo se considera fundamental para la predicción de picos de demanda, ya que la demanda total del sistema está directamente relacionada con los picos de demanda eléctrica, debido a que los picos de demanda se producen cuando la demanda total del sistema alcanza su máximo.

En segundo lugar, se incorporan variables meteorológicas como segunda fuente de variables numéricas, las cuales fueron seleccionadas por su capacidad para influir significativamente en el comportamiento de los usuarios y, por consiguiente, en el consumo de energía. Además, son utilizadas ampliamente en la literatura como variables predictivas de picos de demanda eléctrica, por lo que se consideran variables relevantes para el modelo [7] [18]. Estas incluyen la temperatura ( $X_t^{Tem}$ ), presión atmosférica ( $X_t^{Pres}$ ), humedad relativa ( $X_t^{Hum}$ ) y sensación térmica ( $X_t^{Sens}$ )

Por otro lado, variables categóricas incorporadas representan aspectos temporales que pueden influir en la demanda eléctrica. Estas variables se representan mediante codificación binaria mediante *One-Hot Encoding*, lo que permite representar cada categoría de forma independiente y evitar problemas de orden o jerarquía entre ellas. Dentro de las variables elegidas se incluye la hora del día ( $X_t^{Hora}$ ), el día del mes ( $X_t^{Día}$ ) y el mes del año ( $X_t^{Mes}$ ). Además, y de forma adicional, se consideran variables categóricas ad hoc para representar situaciones específicas que pueden influir en la demanda eléctrica o capturar comportamientos particulares, como si es pandemia ( $X_t^{Pand}$ ) o si es viernes ( $X_t^{Viern}$ ). Estas últimas, se codifican como binarias, donde 1 indica que la condición se cumple y 0 indica que no.

Por último, la variable de salida u objetivo del modelo de predicción se define como una variable binaria que indica qué tan probable es que la hora  $t \in \{1, \dots, \tau\}$  sea *peak* (1) o *no peak* (0). Esta formulación permite clasificar las horas en función de su probabilidad de ser picos de demanda eléctrica, de modo que se puedan identificar las  $k$  horas de máxima demanda dentro del subsistema eléctrico, que corresponderán a las horas utilizadas para el cobro de la tarifa por hora punta de demanda.

Cabe enfatizar que clasificar cada hora de manera independiente constituye una decisión de modelación deliberada y no una simple convención. Las  $k$  horas punta se definen sobre la totalidad del periodo de control, por lo que un planteamiento que abordara el periodo completo como una única instancia dispondría de muy pocos casos para entrenar, en esencia uno por año. Esto constituye una limitación estructural del mecanismo de cargo por capacidad y no una particularidad del caso de estudio. Para superarla, se reformula el problema como una clasificación binaria a nivel horario, de modo que cada hora de cada día se convierte en una observación de entrenamiento. Si bien esta reformulación sacrifica la representación explícita de que las horas punta se determinan por comparación entre sí, incrementa en varios

órdenes de magnitud el tamaño efectivo de la muestra y habilita el entrenamiento de redes neuronales pese a la escasez de años de registro.

### 3.3.2. Procesamiento de datos

Una vez seleccionadas las variables explicativas del modelo, se procede al procesamiento y preparación de las series temporales con el objetivo de construir una base adecuada para el entrenamiento supervisado. En tareas de predicción de demanda eléctrica, este proceso es fundamental para garantizar la calidad de los datos, la coherencia temporal y la correcta representación de las dependencias entre observaciones [29][30]. El procesamiento realizado comprende tres etapas principales: normalización, construcción de ventanas temporales y filtrado según criterios regulatorios.

Las variables numéricas fueron normalizadas utilizando la técnica de *Min-Max Scaling*, que transforma cada atributo a un rango acotado entre 0 y 1. Esta técnica es ampliamente utilizada en modelos basados en redes neuronales, ya que evita que el modelo ajuste desproporcionadamente sus parámetros en torno a variables con escalas mayores, ignorando o reduciendo la influencia de variables de menor escala [34]. La normalización se aplicó de manera independiente para cada año del periodo analizado, con el fin de evitar que variaciones estructurales o cambios de escala entre años introduzcan sesgos en el modelo y afecten su capacidad para aprender patrones característicos de cada periodo.

Posteriormente, para organizar los datos de las series temporales en un formato adecuado para modelos de predicción supervisada, se utilizó una matriz de Hankel, la cual permite estructurar el conjunto de observaciones en ventanas deslizantes que preservan la secuencia temporal de la señal. Una matriz de Hankel se construye disponiendo vectores consecutivos de la serie en columnas cuyos elementos se solapan de manera ordenada, generando una representación que captura la dinámica subyacente y facilita la extracción de patrones temporales. Este enfoque es ampliamente utilizado en técnicas de identificación de sistemas y filtrado, donde la estructura Hankel constituye la base para organizar mediciones en bloques temporales coherentes [35]. Más recientemente, la literatura en pronóstico de series de tiempo ha mostrado que las descomposiciones y factorizaciones de matrices tipo Hankel ofrecen una forma efectiva de modelar dependencias temporales y mejorar el rendimiento predictivo, lo que justifica su uso como mecanismo sistemático para la creación de ventanas deslizantes [36]. La matriz se construyó considerando una ventana de 24 horas hacia el pasado, para capturar dependencias secuenciales y patrones diarios previos, y una ventana de 24 horas hacia el futuro, para permitir que el modelo aprenda a predecir eventos basados en pronósticos a corto plazo.

Finalmente, tanto las variables numéricas procesadas como las variables categóricas codificadas fueron filtradas para conservar únicamente las horas pertenecientes al periodo de control de horas punta, siguiendo la regulación aplicable al subsistema eléctrico estudiado. Este paso garantiza que el entrenamiento se focalice en los intervalos relevantes para la determinación tarifaria y evita introducir ruido proveniente de horas que no intervienen en el cálculo de la tarifa punta.

## 3.4. Diseño de la función de pérdida

A continuación, se muestra la metodología utilizada para diseñar las funciones de pérdida del modelo de predicción, las cuales se basan en el arrepentimiento asociado a las decisiones tomadas por el cliente.

### 3.4.1. Formulación del problema

Se considera un consumidor sujeto a cargos por hora punta de demanda, el cual presentará cargos adicionales si consume dentro de las  $k$  horas de máxima demanda dentro del sistema o subsistema eléctrico en el que está conectado. El consumidor, por lo tanto, está incentivado a reducir su consumo desde la red o desplazarlo mediante un medio de generación local durante las  $k$  horas punta de su sistema/subsistema eléctrico. Se asume que el consumidor posee una función de utilidad genérica  $g$  que depende de la potencia consumida en el tiempo  $t$ ,  $p_t$ , a lo largo de todo un periodo de control  $\tau$ . También, se asume que el consumidor posee métodos de generación local, por lo que su consumo en la hora  $t$  puede provenir tanto desde el sistema  $p_t^{cs}$  como de forma local  $p_t^{gp}$ .

Se establece que  $\pi^{cp}$  corresponderá al cargo por consumo dentro de una de las  $k$  horas punta,  $\pi^{gp}$  el cobro por generación local,  $\pi^{cs}$  al cargo por consumo desde la red y que  $X_t^{Dem}$  es el total de demanda de potencia del sistema en  $t$ . Además, se da por asumido que los costos y cargos guardarán la relación mostrada en (3.1), ya que de esta forma se impone que el cliente utilice los medios de generación local únicamente para evitar los *peaks* del sistema mas no en horas fuera de punta.

$$\pi^{cs} < \pi^{gp} < \pi^{cp} + \pi^{cs} \quad (3.1)$$

Finalmente, se asume que el consumo  $p_t$  del cliente libre de distribución es lo suficientemente pequeño como para no afectar a una hora punta dentro de su subsistema. Esto implica que el cliente no influye en la demanda total del sistema, lo que le permite tomar decisiones sin preocuparse por alterar los picos de demanda del sistema eléctrico.

A partir de lo anterior, el desafío de este problema radica en la dificultad de predecir si una determinada hora corresponde a uno de los  $k$  *peaks*, ya que el cliente no tiene certeza sobre si una hora  $t \in \{1, \dots, \tau - 1\}$ , cuya demanda total de potencia es  $X_t^{Dem}$ , será clasificada dentro de los  $k$  *peaks* hasta que se haya observado toda la demanda futura dentro del periodo  $\tau$ . Esta incertidumbre se ve agravada por el hecho de que el cliente no puede anticipar cuándo ocurrirá uno de los  $k$  *peaks* coincidentes, dado que este depende del comportamiento agregado de todos los usuarios del sistema o subsistema.

En la Figura 3.2 se muestra un esquema que busca representar el problema que enfrenta el consumidor sujeto a cobros por las  $k$  horas punta de demanda eléctrica del sistema eléctrico al que está conectado.

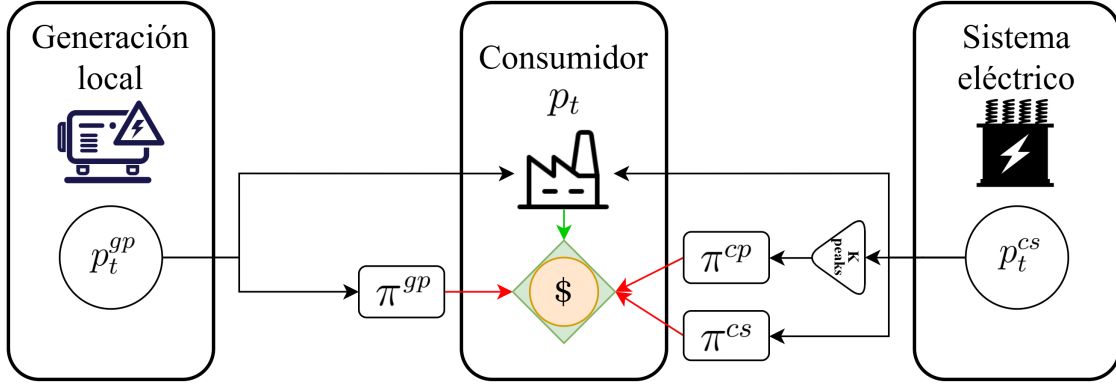


Figura 3.2: Esquema del problema que enfrenta el consumidor sujeto a cobros por las  $k$  horas punta de demanda eléctrica del sistema eléctrico al que está conectado, y que cuenta con un medio de generación local para desplazar su demanda.

### 3.4.2. Función objetivo y restricciones

En base a lo expuesto anteriormente y considerando la formulación planteada en [21] como referencia, se establece que, para maximizar las utilidades en un tiempo  $t$  sujeto a las potenciales  $k$  horas *peak* de un subsistema en específico, el cliente debe resolver el siguiente problema de optimización, el cual se muestra en ecuación (3.2). En este problema, se considera la función de utilidad del cliente, los cargos tarifarios por su consumo desde la red, los cargos por consumir durante una hora punta del subsistema y los costos relacionados a la operación de un sistema de generación local.

$$\text{maximizar } \begin{pmatrix} \text{Función de utilidad} + \text{Cargos por consumo de la red} + \\ \text{Cargos por hora punta} + \text{Costos por generación local} \end{pmatrix} \quad (3.2)$$

La función de utilidad definida en (3.3) se presenta como una función genérica  $g$  que depende de la potencia consumida por el cliente en una hora específica a lo largo del periodo de control de horas punta. Para simplificar el análisis, se asume que  $g$  es cóncava, monótona creciente y continua.

$$\text{Función de utilidad} = g(p_t) \quad (3.3)$$

Los cargos por consumo desde la red, definidos en (3.4), dependen de la potencia consumida desde la red en la hora  $t$  y del cargo por consumo de energía en dicha hora.

$$\text{Cargos por consumo de la red} = -\pi^{cs} \cdot p_t^{cs} \quad (3.4)$$

Los cargos por hora punta, definidos en (3.5), corresponden a los cobros aplicados por el uso de la red eléctrica durante las  $k$  horas de mayor demanda del subsistema al que se encuentra conectado el cliente. Estos cargos dependen del consumo eléctrico del cliente en cada una de dichas horas *peak*, así como de la tarifa por hora punta correspondiente a su subsistema.

Para formalizar esta definición, se considera el conjunto de demandas del subsistema a lo largo del periodo de control, denotado por  $\{S\} := \{S_1, \dots, S_\tau\}$ . Además, y se define la función de clasificación  $r_n(\{S\})$ , que produce el conjunto de índices  $t$  de los  $n$  valores más grandes en  $\{S\}$ . Por último, se introduce la función indicatriz  $\mathbf{1}[t \in r_k(\{S\})]$ , que toma el valor de 1 si la hora  $t$  corresponde a una de las  $k$  horas de máxima demanda del subsistema y 0 en caso contrario.

$$\text{Cargos por hora punta} = -\pi^{cp} \cdot p_t^{cs} \cdot \mathbf{1}[t \in r_k(\{S\})] \quad (3.5)$$

Los costos por generación local, definidos en (3.6), dependen de la potencia utilizada al operar el sistema de generación local y de los costos asociados a dicha operación.

$$\text{Costos por generación local} = -\pi^{gp} \cdot p_t^{gp} \quad (3.6)$$

Finalmente, la función objetivo de la ecuación (3.2) se puede expresar a partir de las definiciones anteriores (3.3), (3.4), (3.5) y (3.6), dando como resultado la ecuación (3.7).

$$\text{maximizar}_{p_t} \sum_{t \in \tau} \left[ g(p_t) - \pi^{cs} \cdot p_t^{cs} - \pi^{cp} \cdot p_t^{cs} \cdot \mathbf{1}[t \in r_k(\{S\})] - \pi^{gp} \cdot p_t^{gp} \right] \quad (3.7)$$

Con respecto a las restricciones del problema, en primer lugar, se establece la restricción de balance de potencia, que asegura que la potencia total consumida por el cliente en una hora  $t$  es igual a la suma de la potencia extraída desde la red y la potencia aportada por el sistema de generación local. Esta restricción se presenta en la ecuación (3.8).

$$p_t = p_t^{cs} + p_t^{gp} \quad (3.8)$$

Por otro lado, las restricciones mostradas en (3.9), (3.10) y (3.11) corresponden, respectivamente, a los límites de la potencia total consumida por el cliente, la potencia consumida desde la red y la potencia suministrada por el sistema de generación local. Estas restricciones aseguran que el consumo del cliente no exceda los límites establecidos para cada tipo de potencia.

$$0 \leq p_t \leq p^{max} \quad (3.9)$$

$$0 \leq p_t^{cs} \leq p^{max-cs} \quad (3.10)$$

$$0 \leq p_t^{gp} \leq p^{max-gp} \quad (3.11)$$

### 3.4.3. Matriz de decisión

A partir de la formulación anterior, para una hora  $t \in \tau$ , las utilidades del cliente vendrán dadas por la función objetivo descrita en 3.7 evaluada al tomar la decisión de consumo  $d$  en base a la predicción de hora punta  $\alpha_t \in \{Peak(1), No Peak(0)\}$  y a la observación

$\beta_t \in \{Peak (1), No Peak (0)\}$ .

$$Utilidades_t = FO(\beta_t, d(\alpha_t)) \quad (3.12)$$

$$Utilidades_t = g(p_t[d(\alpha_t)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t)] - (\pi^{cs} + \mathbf{1}_{\beta_t} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t)] \quad (3.13)$$

Luego, para construir las matrices de decisión del problema, se asume que el consumo  $p_t$  del cliente es constante durante las horas de control, y se considera que los costos y cargos tarifarios siguen la relación mostrada en (3.1). Con esto, se puede definir una matriz de decisiones que represente los costos asociados a las decisiones tomadas por el cliente sujeto a los costos por hora punta. Esta matriz se construye considerando las posibles combinaciones entre los resultados reales y las predicciones realizadas por el modelo.

Teniendo esto en mente, para una hora  $t$  específica, existen 4 posibles resultados en los que el modelo de predicción puede incurrir. Estos son:

- Verdadero negativo (VN): se predice “*No peak*” y ocurre “*No Peak*”. Se consume  $p_t$  desde la red y se cobra el cargo por consumo de energía desde la red,  $\pi^{cs}$ .
- Falso negativo (FN): se predice “*No peak*” y ocurre “*Peak*”. Se consume  $p_t$  desde la red y se paga el cargo por consumo de energía desde la red,  $\pi^{cs}$ , más el cargo por consumir dentro de una hora punta,  $\pi^{cp}$ .
- Verdadero positivo (VP): se predice “*Peak*” y ocurre “*Peak*”. Se desplaza la demanda  $p_t$  a través del medio de generación propia, por lo que consume desde este medio y se paga por su uso correcto a un costo  $\pi^{gp}$ .
- Falso positivo (FP): se predice “*Peak*” y ocurre “*No peak*”. Como se predice un *peak*, se trata de desplazar la demanda  $p_t$  mediante el medio de generación propia y se paga por su uso incorrecto a un costo  $\pi^{gp}$ .

Para representar estos resultados, se construyó una matriz de decisiones, la cual se muestra en la Tabla 3.1 (los cálculos detallados de cada celda se encuentran en el Apéndice A.2). En esta matriz, las filas representan los resultados reales y las columnas representan las predicciones del modelo. Los valores dentro de la matriz corresponden a los costos asociados a cada combinación de predicción y resultado real.

#### 3.4.4. Definición de las funciones de pérdida basada en el arrepentimiento

El modelo de predicción propuesto busca estimar las horas de máxima demanda de un subsistema eléctrico arbitrario, considerando los costos asociados a las decisiones tomadas por el cliente según las predicciones realizadas por el modelo de aprendizaje. Para ello, se propone un entrenamiento utilizando una función de pérdida que minimice el arrepentimiento

Tabla 3.1: Matriz de decisiones del problema que enfrenta el consumidor sujeto a cobros por demanda punta.

|      |         | Predicción                                 |                               |
|------|---------|--|-------------------------------|
|      |         | No peak                                    | Peak                          |
| Real | No peak | $g(p_t) - \pi^{cs} \cdot p_t$              | $g(p_t) - \pi^{gp} \cdot p_t$ |
|      | Peak    | $g(p_t) - (\pi^{cs} + \pi^{cp}) \cdot p_t$ | $g(p_t) - \pi^{gp} \cdot p_t$ |

acumulado a lo largo del periodo de control de horas punta. Esta función de pérdida se define en primer lugar como la media del arrepentimiento en cada hora del periodo de control, como se muestra en la ecuación (3.14).

$$L_{regret} = \frac{1}{\tau} \sum_{t \in \tau} f(regret_t) \quad (3.14)$$

Donde  $f$  es una función que transforma el arrepentimiento en una métrica de pérdida adecuada para el entrenamiento del modelo.

Este contexto, el arrepentimiento en una hora  $t \in \tau$  se define como la diferencia entre las utilidades obtenidas al tomar la decisión óptima en una hora específica y las utilidades resultantes de la decisión adoptada a partir de las predicciones del modelo para esa misma hora. Para ello, se entiende como decisión óptima a aquella que se tomaría bajo un escenario de información perfecta, es decir, cuando se conoce con certeza si ocurrirá o no una hora punta.

Considerando lo anterior, el arrepentimiento se calcula como la diferencia entre las utilidades planteadas en la ecuación (3.12) cuando se decide  $d$  en función de la predicción  $\alpha_t \in \{Peak (1), No Peak (0)\}$  y ocurre el resultado real  $\beta_t \in \{Peak (1), No Peak (0)\}$ , y las utilidades correspondientes a la decisión óptima  $d^*$  la cual se asume con conocimiento perfecto de  $\beta_t$ , como se muestra en la ecuación (3.15).

$$regret_t = FO(\beta_t, d^*(\beta_t)) - FO(\beta_t, d(\alpha_t)) \quad (3.15)$$

Luego, tomando en cuenta la matriz de decisiones mostrada en la Tabla 3.1 y la ecuación (3.15), se puede calcular el arrepentimiento asociado a cada combinación de predicción y resultado real, dando como resultado la matriz de arrepentimiento mostrada en la Tabla 3.2 (los cálculos para cada combinación se muestran en la sección A.3 del anexo).

En esta matriz, el arrepentimiento es cero cuando se predice correctamente, mientras que al realizar una predicción incorrecta se incurre en un valor asociado a qué tanto se aleja el costo de la decisión al costo de la decisión óptima. De este modo, el arrepentimiento representa los sobrecostos incurridos por el cliente al tomar una decisión alejada del óptimo.

Luego, a partir de la matriz anterior, se define la función de pérdida para el modelo de predicción que busca minimizar el arrepentimiento acumulado a lo largo de todas las horas

Tabla 3.2: Matriz de arrepentimiento para el modelo de predicción.

|      |         | Predicción                         |                       |
|------|---------|------------------------------------|-----------------------|
|      |         | No peak                            | Peak                  |
| Real | No peak | 0                                  | $\pi^{gp} - \pi^{cs}$ |
|      | Peak    | $(\pi^{cs} + \pi^{cp}) - \pi^{gp}$ | 0                     |

del periodo de control  $\tau$ , la cual se muestra en la ecuación (3.16).

$$\mathcal{L}_{regret} = \frac{1}{\tau} \sum_{t \in \tau} \left[ (\pi^{gp} - \pi^{cs}) \cdot (1 - y_t) \cdot \hat{y}_t^{peak-c} + ((\pi^{cp} + \pi^{cs}) - \pi^{gp}) \cdot y_t \cdot (1 - \hat{y}_t^{peak-c}) \right] \quad (3.16)$$

Además, se desarrolla una variante de la función de pérdida anterior, la cual busca incorporar un incentivo para predecir verdaderos positivos. Esta variante incluye un término adicional que premia los verdaderos positivos predichos, incentivando al modelo a predecir correctamente las horas *peak* y a entregar más avisos de horas *peak*. En la Tabla 3.3 se muestra la matriz de arrepentimiento para esta variante, donde se puede observar que el arrepentimiento asociado a un verdadero positivo es  $-reward$ , lo que significa que el modelo recibe un premio por predecir correctamente un *peak*.

Tabla 3.3: Matriz de arrepentimiento para la variante del modelo de predicción.

|      |         | Predicción                         |                       |
|------|---------|------------------------------------|-----------------------|
|      |         | No peak                            | Peak                  |
| Real | No peak | 0                                  | $\pi^{gp} - \pi^{cs}$ |
|      | Peak    | $(\pi^{cs} + \pi^{cp}) - \pi^{gp}$ | $-reward$             |

Finalmente, a partir de la matriz de arrepentimiento anterior, se define la función de pérdida para esta variante en la ecuación (3.17).

$$\mathcal{L}_{regret\ modificada} = \frac{1}{\tau} \sum_{t \in \tau} \left[ (\pi^{gp} - \pi^{cs}) \cdot (1 - y_t) \cdot \hat{y}_t^{peak-cm} + ((\pi^{cp} + \pi^{cs}) - \pi^{gp}) \cdot y_t \cdot (1 - \hat{y}_t^{peak-cm}) - reward \cdot y_t \cdot \hat{y}_t^{peak-cm} \right] \quad (3.17)$$

### 3.5. Diseño, entrenamiento y validación del modelo de aprendizaje

A continuación, se presenta una descripción detallada de la arquitectura de la red neuronal utilizada en los modelos de predicción de horas punta, en donde se explicita el número de

capas, la cantidad de neuronas por capa, las funciones de activación y el esquema de regularización aplicado. Asimismo, se detallan los procedimientos de entrenamiento y validación, especificando el optimizador utilizado, los criterios de parada y el proceso de selección y ajuste de hiperparámetros. Además, se expone la función de pérdida utilizada como referencia para comparar los resultados obtenidos con aquellos generados por las funciones de pérdida propuestas. Finalmente, se describen las métricas de evaluación utilizadas para medir el rendimiento del modelo, tanto en términos de calidad de la predicción como de su impacto económico en el cliente.

### 3.5.1. Arquitectura de la red neuronal

Los modelos de predicción propuestos abordan un problema de clasificación binaria, cuyo objetivo es identificar las  $k$  horas de mayor probabilidad de máximos de demanda eléctrica dentro de un subsistema eléctrico, a lo largo de un periodo de control determinado. Para este propósito, se implementó una red neuronal densa (*Dense Neural Network*), una arquitectura ampliamente utilizada en tareas de clasificación cuando se dispone de un conjunto de variables previamente transformadas y estructuradas como vectores de características [34]. Su principal atributo es que cada neurona de una capa está completamente conectada con todas las neuronas de la capa anterior, lo que le permite modelar relaciones no lineales y patrones complejos en los datos de entrada.

En una red neuronal artificial, cada neurona opera como una unidad computacional elemental que transforma un conjunto de entradas en una salida escalar. Para ello, la neurona recibe los valores de entrada  $x_i$  y los combina mediante una suma ponderada  $\sum_i w_i x_i$ , donde  $w_i$  representan los pesos que determinan la influencia relativa de cada entrada. A esta combinación lineal se le añade un término adicional denominado *bias*  $b$ , que desplaza la activación y permite una mayor flexibilidad en el ajuste del modelo. El resultado de esta operación lineal se introduce posteriormente en una función de activación no lineal  $\phi(\cdot)$ , tal como ReLU, sigmoide o *tanh*, cuya finalidad es incorporar no linealidad y permitir que la red neuronal represente relaciones complejas entre las variables de entrada [34]. En consecuencia, la salida de cada neurona viene dada por la expresión mostrada en la ecuación (3.18), constituyendo así el componente fundamental a partir del cual se construyen arquitecturas más profundas y expresivas [37][38].

$$y = \phi \left( \sum_i w_i x_i + b \right) \quad (3.18)$$

Con lo anterior en mente, se decidió utilizar una red compuesta por una capa de entrada con  $n$  neuronas, donde cada una representa una variable de entrada previamente normalizada y codificada, tal como se describe en la sección 3.3. Posteriormente, se incluyen dos capas ocultas de 64 y 32 neuronas, respectivamente, ambas utilizando la función de activación *Rectified Linear Unit* (ReLU). Esta función es ampliamente utilizada en redes neuronales debido a su simplicidad computacional y su capacidad para mitigar el problema del desvanecimiento del gradiente durante el entrenamiento de redes profundas [34]. La función ReLU se define como se muestra en la ecuación (3.19).

$$\text{ReLU}(x) = \max(0, x) \tag{3.19}$$

La capa de salida está compuesta por una única neurona con activación *Sigmoid*, la cual produce un valor continuo entre 0 y 1 que representa la probabilidad estimada de que la hora  $t$  corresponda a un *peak*. Esta formulación probabilística es estándar para problemas de clasificación binaria. La función *Sigmoid* se define como se muestra en la ecuación (3.20).

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{3.20}$$

Por otro lado, el sobreajuste ocurre cuando un modelo de aprendizaje automático logra representar con gran precisión los patrones presentes en los datos de entrenamiento, pero a costa de perder capacidad de generalización frente a datos no observados. En esta situación, el modelo no solo aprende las relaciones subyacentes relevantes, sino que también memoriza ruido, fluctuaciones aleatorias o particularidades específicas del conjunto de entrenamiento, lo que se traduce en un rendimiento deficiente al momento de predecir nuevos ejemplos [34].

Una técnica ampliamente utilizada para mitigar el fenómeno del sobreajuste es *dropout*, la cual consiste en desactivar aleatoriamente una proporción de neuronas durante el entrenamiento. Este mecanismo obliga a la red a no depender excesivamente de unidades o rutas particulares dentro de la arquitectura, promoviendo la redundancia y la diversificación de los patrones aprendidos, así como mejorando su capacidad de generalización [39].

Por último, en este problema de clasificación supervisada es fácil identificar problemas relacionados con el desbalance de clases en el conjunto de datos, ya que las horas punta son mucho menos frecuentes que las horas no punta. Este desequilibrio puede inducir al modelo a favorecer las clases mayoritarias, reduciendo su capacidad para reconocer adecuadamente las clases minoritarias y deteriorando el desempeño global [40].

Una estrategia habitual para mitigar el efecto del desbalance de clases consiste en incorporar ponderaciones que penalicen más fuertemente los errores cometidos sobre clases subrepresentadas [38]. En Keras, la función `compute_class_weight` permite calcular automáticamente estos pesos a partir de la frecuencia relativa de cada clase en el conjunto de entrenamiento [41]. Estas ponderaciones se entregan posteriormente al modelo mediante el parámetro `class_weight` durante el proceso de entrenamiento, incentivando que la red neuronal otorgue mayor importancia a las clases minoritarias y mejorando la capacidad de generalización en escenarios con desbalance.

Finalmente, con el objetivo de ilustrar la arquitectura de la red neuronal utilizada para los modelos de predicción de horas punta, se presenta un esquema en la Figura 3.3. En este esquema, se puede observar la disposición de las capas y las neuronas, así como las funciones de activación utilizadas en cada una de ellas.

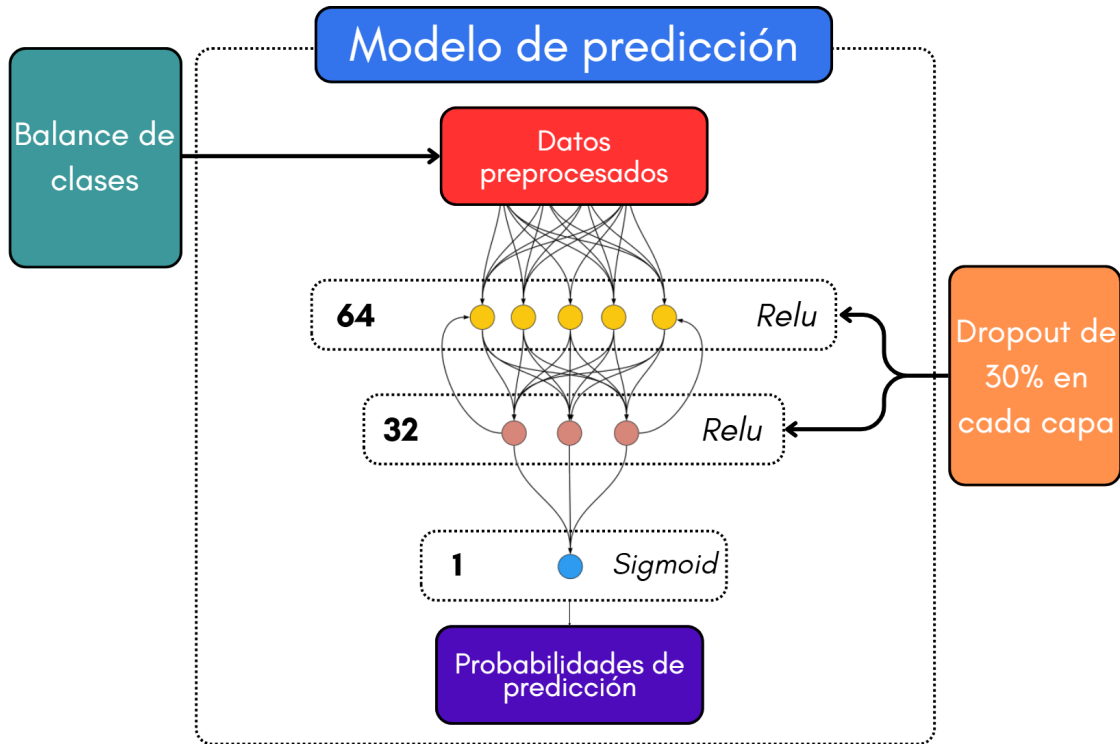


Figura 3.3: Arquitectura de la red neuronal utilizada para la predicción de horas punta.

### 3.5.2. Entrenamiento del modelo

El entrenamiento de un modelo de red neuronal consiste en un proceso iterativo mediante el cual el algoritmo ajusta sus parámetros internos (pesos y sesgos) con el fin de aprender patrones presentes en los datos históricos y así para minimizar una función de pérdida que cuantifica el error entre las predicciones del modelo y los valores reales. Este proceso se lleva a cabo mediante técnicas de optimización basadas en gradientes, donde el modelo aprende patrones presentes en los datos a través de múltiples pasadas o épocas sobre el conjunto de entrenamiento. Durante el entrenamiento, el modelo evalúa su desempeño sobre un conjunto de validación, lo que permite monitorear su capacidad de generalización y ajustar hiperparámetros claves. El objetivo final es obtener un modelo capaz de reproducir adecuadamente la estructura subyacente de los datos y realizar predicciones precisas sobre información no vista previamente [34][38].

Para el entrenamiento de los modelos de predicción de máximos de demanda, el conjunto de datos se divide en dos partes: un conjunto de entrenamiento y validación, equivalente aproximadamente al 80 % de las observaciones, y un conjunto de prueba, correspondiente al 20 % restante, siguiendo recomendaciones ampliamente aceptadas en la literatura de aprendizaje supervisado [34]. Esta separación se realiza respetando el orden temporal de los datos, dividiendo la información por años completos. De esta forma, se logra preservar la estructura secuencial de los datos, evitar la mezcla de información entre distintos periodos y capturar la variabilidad anual y estacional [42].

Una vez realizada la partición, el conjunto de entrenamiento y validación se utiliza para

ajustar los hiperparámetros y entrenar la red neuronal. En particular, con el fin de comparar los modelos propuestos con las distintas funciones de pérdida diseñadas en la sección 3.4, se implementa un modelo de referencia utilizando una función de pérdida convencional: la entropía cruzada binaria (*binary cross-entropy*, BCE), que es comúnmente utilizada en problemas de clasificación binaria por su capacidad de medir la discrepancia entre las predicciones del modelo y las etiquetas reales [34][38]. En la ecuación 3.21 se muestra la función de pérdida BCE dada una etiqueta real  $y \in \{0, 1\}$  y una probabilidad predicha  $\hat{y} \in [0, 1]$ .

$$\mathcal{L}_{\text{BCE}}(y, \hat{y}) = -\left(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right) \quad (3.21)$$

De esta forma, se construyen tres modelos de predicción de horas punta de demanda eléctrica: un modelo de referencia enfocado en predicción entrenado con una función BCE; un modelo enfocado en costos, entrenado con la función de pérdida basada en el arrepentimiento económico; y un modelo enfocado en costos modificado, entrenado con la variante de la función de pérdida basada en el arrepentimiento económico.

Por último, para el proceso de optimización de los modelos de aprendizaje se emplea el algoritmo *Adam*, el cual corresponde a uno de los algoritmos de optimización más utilizados y que combina las ventajas de los métodos basados en gradientes con momentum y del escalamiento adaptativo de las tasas de aprendizaje [43]. Complementariamente, se implementa la técnica de *Early Stopping*, la cual es una técnica ampliamente empleada para evitar el sobreajuste y que consiste en monitorizar el error en el conjunto de validación durante el entrenamiento y detener el proceso cuando dicho error deja de mejorar de manera sostenida. De esta forma, se impide que el modelo continúe ajustándose a patrones específicos del conjunto de entrenamiento, preservando su capacidad de generalización y mejorando su desempeño en datos no observados [44].

Finalmente, una vez completado el entrenamiento, el desempeño final de cada modelo se evalúa utilizando exclusivamente el conjunto de prueba, que contiene datos completamente nuevos para el modelo y permite obtener una estimación realista de su capacidad predictiva.

### 3.5.3. Selección y ajuste de hiperparámetros

En modelos de aprendizaje automático, los hiperparámetros corresponden a configuraciones externas del modelo que se fijan antes del proceso de entrenamiento y que influyen directamente en su rendimiento y capacidad de generalización. A diferencia de los parámetros internos (pesos y sesgos), los hiperparámetros no son aprendidos durante el entrenamiento, sino que deben ser determinados mediante un procedimiento de antemano [38][34]. Entre los hiperparámetros más relevantes en redes neuronales se encuentran el número de épocas de entrenamiento (*epochs*), la tasa de aprendizaje (*learning rate*) y el tamaño del lote (*batch size*).

El objetivo del proceso de selección y ajuste de hiperparámetros es identificar la combinación que permita obtener el mejor desempeño para cada uno de los modelos considerados. Para ello, se implementó una búsqueda en malla (*grid search*), la cual es una estrategia que consiste en definir un conjunto discreto y determinístico de valores posibles para cada hiper-

parámetro y evaluar todas las combinaciones [45]. Cada combinación es entrenada y evaluada mediante un procedimiento consistente y replicable, capturando los efectos conjuntos de los distintos hiperparámetros.

Para evaluar cada configuración, se empleó un esquema de validación cruzada de 5 pliegues adaptado a datos temporales. El conjunto de entrenamiento y validación, definido en la sección anterior, se divide en cinco segmentos conservando el orden cronológico. En cada iteración, el modelo se entrena con el 80 % del pliegue y se valida con el 20 % restante, asegurando que no se mezcle información futura en el entrenamiento y preservando la estructura secuencial de la serie temporal, lo que es fundamental en tareas de predicción basadas en datos dependientes del tiempo [45].

Las métricas obtenidas en los cinco pliegues se promedian para cada combinación de hiperparámetros, y se selecciona aquella que optimiza las métricas de interés previamente definidas. En el caso del modelo enfocado en predicción, se priorizan combinaciones que maximizan recuperación (*recall*), precisión (*precision*), puntaje F1 (*F1-score*) y exactitud (*accuracy*), dado que el objetivo consiste en identificar de manera correcta las horas *peak*. Por otro lado, para los modelos enfocados en costos, se busca minimizar la función de pérdida asociada al arrepentimiento y, simultáneamente, maximizar recuperación, precisión y F1-score, considerando que su propósito es reducir los costos derivados de las decisiones tomadas por el cliente a partir de las predicciones generadas.

En la Figura 3.4 se presenta un diagrama que ilustra el proceso de validación cruzada desarrollado para el ajuste paramétrico de los modelos de aprendizaje, destacando la división del conjunto de datos en pliegues y la evaluación de las métricas de rendimiento en cada iteración.

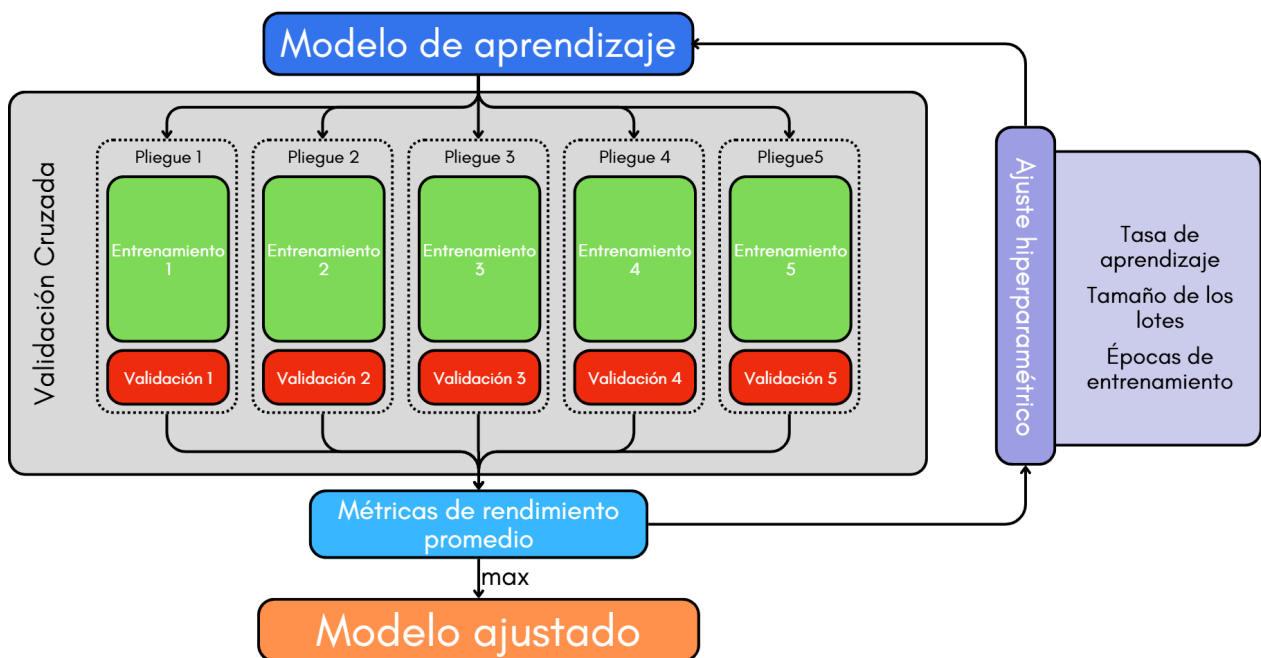


Figura 3.4: Validación cruzada desarrollada para el ajuste de hiperparámetros.

### 3.5.4. Entrenamiento final del modelo y métricas de evaluación

Una vez seleccionados los hiperparámetros óptimos mediante el procedimiento descrito en la sección anterior, se entrena cada uno de los modelos utilizando la totalidad del conjunto de entrenamiento y validación. Dado que el entrenamiento de una red neuronal incorpora componentes estocásticos tales como la inicialización aleatoria de los pesos, el barajado de los mini-lotes y las máscaras de dropout, una única ejecución no permite estimar de forma robusta el desempeño del modelo ni descartar que los resultados dependan de una inicialización particularmente favorable o desfavorable. Por este motivo, cada modelo se entrena de forma independiente 10 veces, utilizando una semilla aleatoria distinta en cada réplica, de modo que cada una parta desde una inicialización de pesos diferente.

Cada una de estas 10 réplicas se evalúa sobre el conjunto de prueba, clasificando cada hora como peak o no peak según la probabilidad entregada por la red neuronal. A partir de estas predicciones se construye una matriz de confusión, la cual resume el número de verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN) obtenidos, y se calculan tanto métricas de calidad como métricas económicas, con el objetivo de evaluar simultáneamente la capacidad predictiva y el impacto económico de las decisiones derivadas del modelo.

Las métricas de calidad y económicas reportadas en el Capítulo 5 corresponden al promedio sobre estas 10 réplicas, acompañado de su desviación estándar. Este procedimiento permite caracterizar tanto el desempeño esperado de cada modelo como su variabilidad frente a la aleatoriedad del entrenamiento, entregando una estimación más confiable y reproducible y permitiendo evaluar la robustez de las diferencias observadas entre los modelos.

En primer lugar, las métricas de calidad consideradas son la precisión, la recuperación y la exactitud, las cuales son ampliamente utilizadas para evaluar modelos de clasificación binaria [45]. Sus definiciones se muestran en las ecuaciones (3.22)–(3.25).

- Precisión: mide la proporción de verdaderos positivos entre todas las predicciones positivas realizadas por el modelo. Se muestra en la ecuación (3.22).

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (3.22)$$

- Recuperación: mide la proporción de verdaderos positivos entre todos los casos positivos reales. Se define en la ecuación (3.23).

$$\text{Recuperación} = \frac{VP}{VP + FN} \quad (3.23)$$

- Exactitud: evalúa la proporción de predicciones correctas (tanto positivas como negativas) sobre el total de predicciones realizadas. Se define en la ecuación (3.24).

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.24)$$

- Puntaje F1: es la media armónica de la precisión y el *recall*, lo cual permite tener una

métrica balanceada que combina ambas. Se define en la ecuación (3.25).

$$F1\text{-score} = 2 \cdot \frac{\textit{Precisión} \cdot \textit{Recuperación}}{\textit{Precisión} + \textit{Recuperación}} \quad (3.25)$$

Además de las métricas de calidad, se consideran métricas económicas que permiten cuantificar el impacto monetario asociado a las decisiones del cliente. La primera de ellas es el arrepentimiento económico, que mide la diferencia entre los costos incurridos al seguir las predicciones del modelo y los costos que se habrían obtenido bajo decisiones óptimas. Esta métrica incorpora explícitamente los costos asociados a falsos positivos y falsos negativos, y se formaliza en la ecuación (3.26):

$$\textit{Arrepentimiento} = \sum_{t \in \tau} \left[ FP \cdot (\pi^{gp} - \pi^{cs}) + FN \cdot ((\pi^{cp} + \pi^{cs}) - \pi^{gp}) \right] \quad (3.26)$$

Por último, se define el costo total de operación, el cual considera todos los costos asociados a las decisiones tomadas durante el periodo de control, incluyendo el consumo de energía desde la red, el uso del medio de generación propia y los cargos por consumo en horas punta. Este costo se calcula sumando los costos de las decisiones tomadas en cada hora del periodo de control, como se muestra en la ecuación (3.27).

$$\textit{Costo total de operación} = \sum_{t \in \tau} \left[ VN \cdot \pi^{cs} + FP \cdot \pi^{gp} + FN \cdot (\pi^{cp} + \pi^{cs}) + VP \cdot \pi^{gp} \right] \quad (3.27)$$

Finalmente, para evaluar si las diferencias de desempeño entre modelos son estadísticamente significativas y no atribuibles a la variabilidad entre semillas, se comparan las métricas obtenidas en las 10 réplicas mediante un test t de Welch para muestras independientes. Como la dispersión entre réplicas difiere de forma apreciable, se consideró apropiado utilizar el test t de Welch debido a que no requiere que los grupos comparados tengan igual varianza, mantiene un control adecuado del error de tipo I bajo heterocedasticidad y conserva su potencia cuando las varianzas son similares [46]. Para cada comparación se reportan el estadístico  $t$ , los grados de libertad, el valor- $p$  y el intervalo de confianza al 95 % de la diferencia de medias.

# Capítulo 4

## Caso de estudio

En este capítulo se presenta una descripción del caso de estudio desarrollado en este trabajo, junto con los datos de entrada utilizados para la predicción de las horas punta. Posteriormente, se detallan los hiperparámetros utilizados en los modelos de predicción junto con los detalles de la configuración del experimento.

### 4.1. Descripción del caso de estudio

El objetivo de este estudio consiste en evaluar y comparar económicamente las decisiones tomadas a partir de las predicciones generadas por los modelos desarrollados al ser aplicadas en el mercado de potencia chileno, el cual busca remunerar la suficiencia de capacidad del sistema eléctrico mediante un cargo adicional a los clientes que consumen durante las 52 horas de máxima demanda del subsistema al que están conectados. Para ello, el caso de estudio planteado consiste en la implementación de las herramientas de predicción de horas punta, descritas en los capítulos anteriores, en el subsistema de distribución de la Región Metropolitana de Santiago de Chile.

Con este propósito, se obtienen las clasificaciones de todas las horas del periodo de control de horas punta analizado, generado por cada modelo de predicción entrenado y ajustado. A partir de dichas clasificaciones, se calculan tanto las métricas de calidad como las métricas económicas definidas anteriormente. Posteriormente, se realiza una comparación entre el enfoque propuesto y los enfoques convencionales, se realiza un análisis conjunto de los avisos obtenidos y de las variables climáticas, y se desarrolla una evaluación económica de las decisiones generadas.

Finalmente, se realizan dos análisis de sensibilidad complementarios sobre los modelos desarrollados. El primero evalúa su desempeño económico frente a variaciones en la relación de costos entre los falsos positivos y los falsos negativos, mientras que el segundo examina su robustez frente a cambios en el número de horas punta penalizadas ( $k$ ). La configuración de ambos análisis se detalla en la Sección 4.4.

### 4.1.1. El mercado eléctrico de potencia en Chile

En Chile, el mercado de potencia del sistema eléctrico tiene por objetivo remunerar la suficiencia de capacidad del sistema, especialmente durante los momentos de máxima demanda o mayor estrés del sistema eléctrico. Esto se asegura mediante un cargo adicional a los clientes que consumen durante las horas de máxima demanda de los subsistemas en los que están conectados.

Para ello, el Coordinador Eléctrico Nacional (CEN) toma las 52 horas de mayor demanda de potencia para cada subsistema, leídas en un periodo de control comprendido entre las 18:00 hasta las 22:00 horas, en días hábiles, y entre los meses de Abril y Septiembre, y realiza un cobro a aquellos usuarios en función de cuánto aportan a la demanda punta en sus respectivos subsistemas en dichas horas de control [47].

Este proceso se realiza de forma *ex post*, lo que significa que, durante el año, los clientes desconocen cuáles serán los horarios que el CEN clasificará como horas punta [48]. Además, para este cálculo, el CEN separa a los clientes libres que cuentan con su propio alimentador y a aquellos que se conectan a un alimentador de distribución. A estos últimos se les realiza el cobro de demanda punta según su contribución a las 52 horas de mayor del consumo total de dicha zona de concesión [47].

Debido a lo anterior, si los clientes libres conectados a un sistema de distribución logran predecir qué horas serán *peak*, existe una oportunidad para reducir sus costos por concepto de hora punta, desplazando sus consumos hacia horas fuera del periodo de control o cambiando el suministro de energía durante esas horas por un método alternativo, como por ejemplo generadores diésel o bancos de baterías.

### 4.1.2. Descripción del subsistema de distribución

El subsistema eléctrico elegido abarca 32 comunas de la Región Metropolitana de Chile y suministra energía a más de 2 millones de clientes, de los cuales el 89.9% corresponde a clientes regulados y el resto a clientes libres. Durante el año 2024, el subsistema distribuyó un total de 14.8 TWh de energía eléctrica, alcanzando una demanda máxima de 2814 MW. lo que lo convierte en uno de los principales subsistemas de distribución del país.

## 4.2. Datos de entrada

Los datos de entrada recopilados para el modelo de predicción abarcan el periodo comprendido entre 2018 y 2023, y se agrupan en tres categorías principales: consumo eléctrico total del subsistema de distribución, variables meteorológicas correspondientes a la zona geográfica de interés, y costos económicos asociados al consumo de energía desde la red, la generación propia y a los costos por potencia durante las 52 horas de máxima demanda.

Por otro lado, la variable objetivo del modelo corresponde a las 52 horas de máxima demanda del subsistema de distribución seleccionado, las cuales son calculadas y publicadas anualmente por el Coordinador Eléctrico Nacional.

A continuación, se describen en profundidad los datos de entrada utilizados en el modelo de predicción.

#### 4.2.1. Consumo eléctrico del subsistema

El consumo eléctrico total del subsistema de distribución de la Región Metropolitana de Santiago de Chile se obtiene a partir de los datos históricos de consumo e inyección eléctrica registrados por los medidores de cabecera de las subestaciones del Sistema Eléctrico Nacional (SEN), publicados en la Plataforma de Recepción de Medidas de Transferencias Económicas del Coordinador Eléctrico Nacional (PRMTE). Estos medidores registran, de forma horaria, las transferencias de energía de todos los coordinados eléctricos del sistema, incluyendo aquellas correspondientes al subsistema seleccionado.

Los datos de consumo recopilados abarcan el periodo comprendido entre 2018 y 2023, y fueron procesados para filtrar únicamente aquellos medidores correspondientes a la zona de concesión del subsistema en estudio. Para ello, se utilizaron las claves identificadoras proporcionadas por el CEN en su Informe de Retiro de Potencia, las cuales son empleadas para el cálculo oficial de las 52 horas de máxima demanda.

Finalmente, los datos de consumo son procesados para obtener las curvas de demanda total del subsistema para todos los años de análisis. Dichas curvas se presentan en la Figura 4.1 y muestran la evolución del consumo eléctrico del subsistema escogido a lo largo del periodo de estudio.



Figura 4.1: Curvas de demanda del subsistema de distribución de la Región Metropolitana de Santiago de Chile entre los años 2018 y 2023, cuyos datos fueron extraídos de la Plataforma de Recepción de Medidas de Transferencias Económicas (PRMTE) del Coordinador Eléctrico Nacional.

#### 4.2.2. Variables meteorológicas

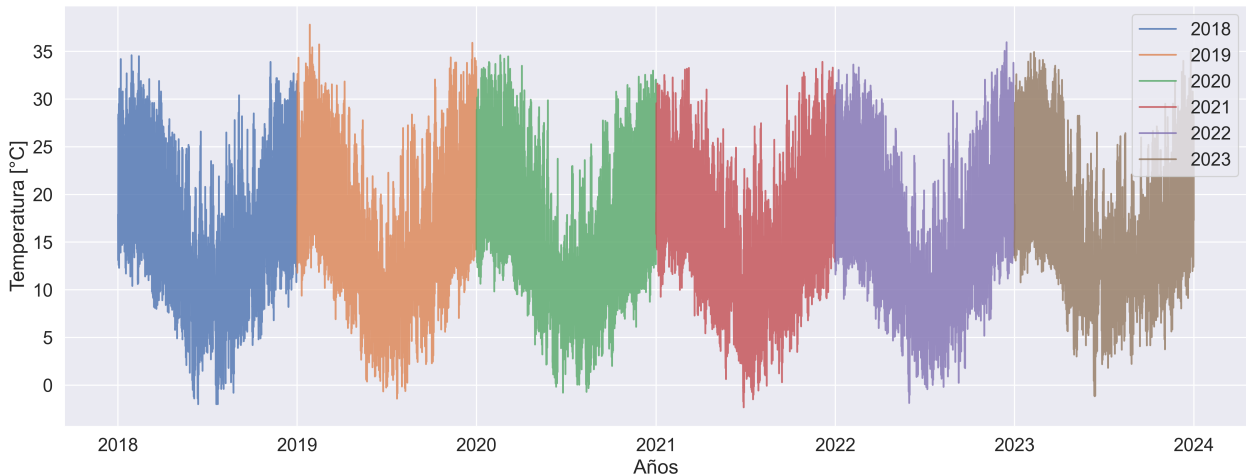
Las variables meteorológicas utilizadas en el modelo de predicción incluyen los siguientes factores:

- Temperatura medida en grados Celsius ( $^{\circ}\text{C}$ ).
- Humedad relativa medida en porcentaje (%).
- Presión atmosférica medida en hectopascales (hPa).
- Sensación térmica medida en grados Celsius ( $^{\circ}\text{C}$ ) y calculada utilizando la fórmula de sensación térmica descrita en la ecuación (4.1), donde  $T$  es la temperatura en grados Celsius y  $V$  es la velocidad del viento en kilómetros por hora (km/h) [49].

$$\text{Sensación térmica} = 13.12 + 0.6215 \cdot T - 11.37 \cdot V^{0.16} + 0.3965 \cdot T \cdot V^{0.16} \quad (4.1)$$

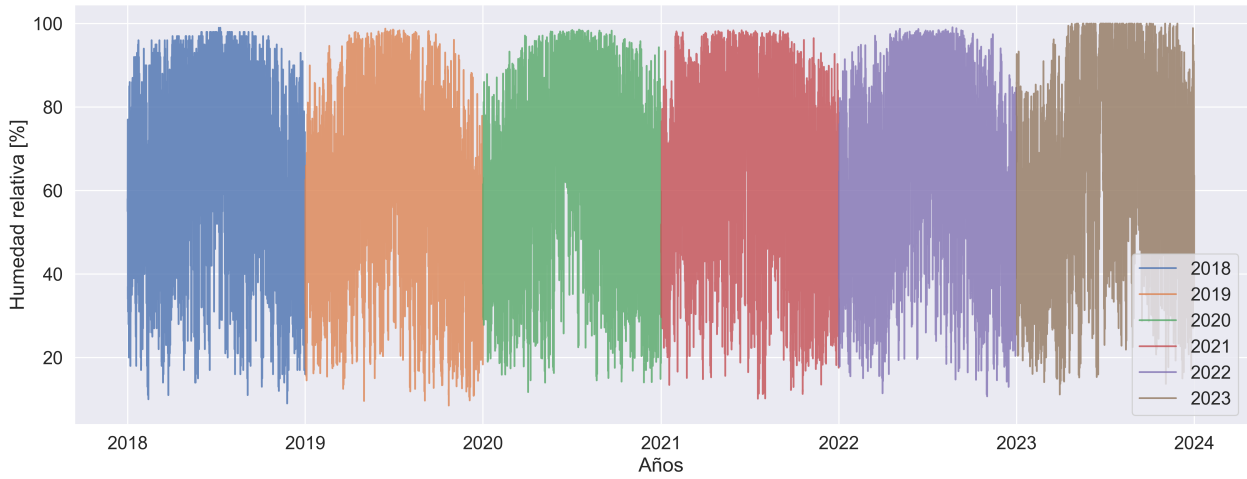
Estas variables son recopiladas desde dos estaciones meteorológicas publicadas por la Dirección Meteorológica de Chile, ubicadas en Quinta Normal y Cerro Colorado, ambas en la Región Metropolitana. La elección de estas ubicaciones radica en la proximidad geográfica a la zona de concesión del subsistema elegido, lo que garantiza que las condiciones climáticas reflejadas en los datos sean representativas de la región y guarden relación con la demanda eléctrica del subsistema.

Finalmente, las variables meteorológicas se procesaron para obtener las curvas de cada variable a lo largo del tiempo, las cuales se presentan en la Figura 4.2. Estas curvas muestran la evolución de las variables meteorológicas en la región metropolitana entre 2018 y 2023, mostrando las variaciones climáticas a lo largo de este periodo.

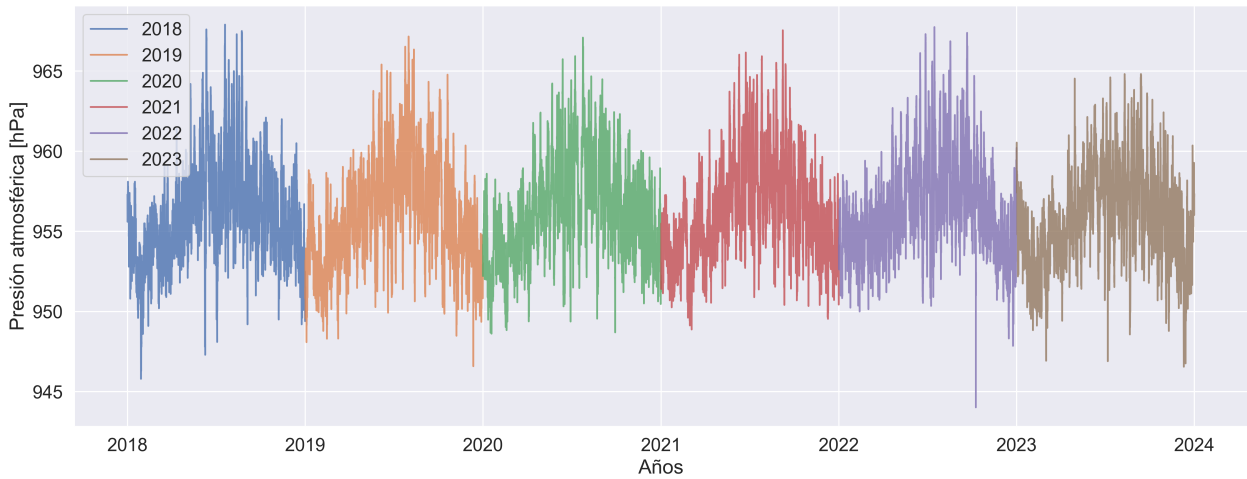


(a) Temperatura.

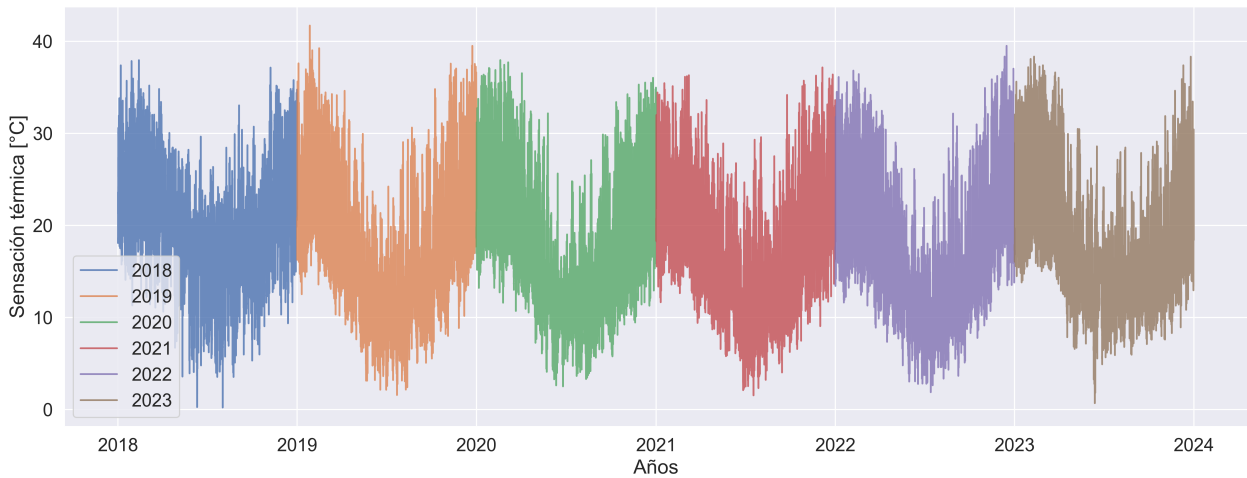
Figura 4.2: Curvas de las variables meteorológicas en la Región Metropolitana de Santiago de Chile, cuyos datos fueron extraídos desde dos estaciones meteorológicas publicadas por la Dirección Meteorológica de Chile, ubicadas en Quinta Normal y Cerro Colorado.



(b) Humedad relativa.



(c) Presión atmosférica.



(d) Sensación térmica.

Figura 4.2: Curvas de las variables meteorológicas en la Región Metropolitana de Santiago de Chile, cuyos datos fueron extraídos desde dos estaciones meteorológicas publicadas por la Dirección Meteorológica de Chile, ubicadas en Quinta Normal y Cerro Colorado (continuación).

### 4.2.3. Costos económicos

Los costos utilizados en el modelo de predicción se dividen en tres: costos de energía consumida desde la red, costos de potencia en hora punta y costos de generación propia. Estos datos fueron fundamentales para el entrenamiento y la evaluación de los modelos enfocados en costos. Además, resultaron imprescindibles para calcular el arrepentimiento y el costo total de operación de todos los modelos de predicción, permitiendo evaluar su desempeño económico en función de las decisiones tomadas respecto a las horas punta pronosticadas.

En primer lugar, para los costos de energía consumida desde la red se utiliza como referencia el Precio Nudo de Energía Promedio (PNEP) para el subsistema elegido publicado por la Comisión Nacional de Energía (CNE) en el informe técnico preliminar de fijación de precios de nudo promedio del sistema eléctrico nacional para el año 2023, y se expresa en dólares por megavatio-hora (USD/MWh) [50].

En segundo lugar, para los costos de potencia en hora punta se utilizó el Precio Nudo Promedio de Potencia (PNPP) para el subsistema elegido publicado por la CNE en el mismo informe técnico preliminar mencionado anteriormente, y se expresa en dólares por megavatio por mes (USD/MW/mes) [50].

Cabe destacar que el costo de la potencia en hora punta utilizado en el modelo se calculó dividiendo el valor publicado en el informe de precio nudo promedio (11.34 [USD/kW/mes]) por 52, obteniendo así un costo por hora de potencia de 218.077 [USD/MW], necesario para poder trabajar de forma horaria con los avisos y predicciones, y poder calcular el arrepentimiento y los costos totales de la operación de los modelos de predicción. Lo anterior se desarrolló así debido a que el cálculo del cobro por potencia se realiza considerando el promedio de las 52 horas de máxima demanda del año multiplicadas por el precio de la potencia considerado en el informe anteriormente mencionado.

Por último, se consideró un grupo electrógeno diésel como medio de generación propia para desplazar consumo en horas catalogadas como punta. Para ello, el costo variable de combustible diésel se calculó como  $\pi^{gp} = SFC \cdot P$  [USD/MWh], donde  $SFC$  es el consumo específico de un generador estándar expresado en [L/kWh] y  $P$  corresponde al precio del diésel expresado en [USD/m<sup>3</sup>], en donde se toma como referencia los valores reportados por ENAP en su Tabla de Precios de Paridad (en el Anexo B se muestran los cálculos de los costos variables de generación propia).

Finalmente, los valores específicos de los costos utilizados en los modelos de predicción se detallan en la Tabla 4.1.

Tabla 4.1: Costos específicos utilizados por los modelos de predicción.

| <b>Costo de la energía desde la red</b><br>[USD/MWh] | <b>Costo de la potencia en hora punta</b><br>[USD/MW] | <b>Costo de generación propia</b><br>[USD/MWh] |
|--|---|--|
| 93.655   | 218.077   | 126.4  |

#### 4.2.4. Horas punta oficiales

Las horas punta oficiales utilizadas en el modelo de predicción corresponden a las 52 horas de máxima demanda del subsistema estudiado, calculadas y publicadas anualmente por el CEN.

Los datos fueron extraídos de los antecedentes para el cálculo del informe de retiro de potencia elaborado por el CEN, los cuales se encuentran disponibles en su sitio web oficial [51]. Estos datos incluyen las 52 horas punta oficiales correspondientes a cada año, desde 2018 hasta 2023, y sus respectivos consumos leídos.

Este conjunto de datos fue utilizado como variable objetivo del modelo de predicción, y se empleó para entrenar y evaluar el desempeño del modelo mediante el uso de métricas de calidad y económicas.

### 4.3. Configuración del experimento

A continuación, se describen los aspectos relevantes de la configuración del experimento realizado para evaluar el desempeño de los modelos de predicción de horas punta. Estos aspectos incluyen la selección del periodo de análisis, la división en conjuntos de entrenamiento, validación y prueba, y el conjunto de hiperparámetros evaluados.

#### 4.3.1. Selección del periodo de análisis

El periodo de análisis seleccionado para el experimento abarca los años comprendidos entre el 2018 y el 2023. La elección de este intervalo se basa principalmente en la disponibilidad de datos históricos del consumo eléctrico del subsistema de distribución elegido en la página web del Coordinador Eléctrico Nacional (CEN).

Esta disponibilidad acotada de mediciones oficiales, sumada a que el cargo por capacidad define las horas punta sobre todo el periodo de control y aporta por tanto una sola instancia por año, es lo que motiva la reformulación del problema como una clasificación a nivel horario descrita en la Sección 3.3.1. De esta forma, se dispone de un volumen de observaciones suficiente para entrenar los modelos.

El conjunto de datos se filtró para incluir en el análisis únicamente las horas correspondientes al periodo de control de horas punta. Este periodo considera únicamente los días hábiles, excluyendo fines de semana y festivos. Además, el estudio tuvo en cuenta los cambios en la normativa que define dicho periodo a lo largo de los años. Por lo que, de acuerdo con esta normativa, para los distintos años analizados se consideraron las siguientes horas de control para el cálculo de las horas punta:

- 2018 a 2019: Abril a Septiembre, entre las 18:00 y las 23:00 horas.
- 2020 a 2021: Junio a Julio, entre las 18:00 y las 22:00 horas.

- 2022 a 2023: Abril a Septiembre, entre las 18:00 y las 22:00 horas.

Cabe destacar que el periodo de análisis incluye datos de los años 2020 y 2021, que corresponden a un periodo marcado por la pandemia de COVID-19. Este evento pudo haber influido en los patrones de consumo eléctrico, ya que durante este tiempo se produjeron cambios significativos en el comportamiento de los consumidores y en la actividad económica. Sin embargo, se consideró que la inclusión de estos datos era relevante para el análisis, ya que permite evaluar cómo los modelos se adaptan a situaciones excepcionales y cómo estas pueden afectar las predicciones de horas punta. Con esto en mente, y con el fin de incorporar el efecto de la pandemia en el modelo, se utilizó una variable categórica que indica si el año corresponde a un periodo de pandemia o no, lo que permite al modelo asociar patrones específicos de esos años a dicho evento.

Finalmente, para el estudio se consideraron datos horarios, lo que implica que cada observación corresponde a un registro de consumo eléctrico y variables meteorológicas por hora. Esto permite una granularidad adecuada en el análisis según la literatura, y una mejor comprensión de los patrones de consumo a lo largo del tiempo.

#### 4.3.2. División en conjuntos de entrenamiento, validación y prueba

El conjunto de datos se dividió siguiendo una asignación temporal en cada subconjunto. Para el entrenamiento y validación del modelo mediante validación cruzada, se utilizaron los datos correspondientes a los años 2018 a 2022, lo que representa un total de cinco años de información, es decir, un 83.3% del total de datos. Este subconjunto se empleó en su totalidad para el entrenamiento final del modelo. Por otro lado, los datos del año 2023 se reservaron exclusivamente para la evaluación final, representando un 16.7% del total.

#### 4.3.3. Conjunto de hiperparámetros evaluados

Para definir los hiperparámetros óptimos de los modelos de predicción, se realizó una búsqueda en malla (*grid search*), en donde se eligió un conjunto determinístico de valores para cada hiperparámetro y se probó el desempeño de cada posible combinación de estos valores mediante una validación cruzada. El conjunto de hiperparámetros que se consideró se muestra a continuación:

- Épocas (*epochs*)  $\in \{10, 50, 100, 150\}$ .
- Tamaño del lote (*batch size*)  $\in \{12, 16, 32, 64\}$ .
- Tasa de aprendizaje (*learning rate*)  $\in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ .

A partir de cada combinación del conjunto de hiperparámetros previamente definido, se aplicó una validación cruzada de 5 pliegues, en la cual cada pliegue correspondió a un año calendario entre 2018 y 2022. Este enfoque implicó dividir el conjunto de datos en cinco subconjuntos anuales, entrenar el modelo utilizando cuatro de ellos y evaluar su rendimiento con

el subconjunto restante. El proceso se repitió cinco veces por combinación, rotando el pliegue utilizado para la validación, de manera que cada subconjunto fuera utilizado exactamente una vez como conjunto de evaluación. Luego, a partir de las métricas obtenidas en cada pliegue, se calcularon los valores promedio correspondientes a cada combinación de hiperparámetros.

Finalmente, para el modelo enfocado en predicción se seleccionó la combinación que maximizó la recuperación, la precisión y el *F1 score*, mientras que para los modelos enfocados en costos se seleccionó la combinación que minimizó la función de pérdida relacionada con el arrepentimiento del modelo, y se maximizó la recuperación y la precisión. En resumen, se buscó maximizar la calidad predictiva o económica del modelo, dependiendo de su enfoque específico.

Finalmente, los hiperparámetros seleccionados para cada modelo se muestran en la Tabla 4.2.

Tabla 4.2: Hiperparámetros elegidos para los modelos de predicción.

| Modelo                               | Épocas | Tasa de aprendizaje | Tamaño del lote |
|--------------------------------------|--------|---------------------|-----------------|
| <b>Enfocado en predicción</b>        | 10     | 0.001               | 16              |
| <b>Enfocado en costos</b>            | 150    | 0.01                | 12              |
| <b>Enfocado en costos Modificado</b> | 150    | 0.01                | 64              |

## 4.4. Configuración de los análisis de sensibilidad

Con el propósito de evaluar la robustez de los modelos desarrollados, se plantean dos análisis de sensibilidad que operan sobre ejes distintos del caso de estudio. El primero mantiene fijo el número de horas punta penalizadas y varía la relación entre los costos asociados a los falsos positivos y los falsos negativos. El segundo mantiene fija la estructura de costos del caso de estudio y varía el número de horas punta penalizadas. De este modo, cada análisis aísla el efecto de una de las dos dimensiones que determinan el desempeño económico de los modelos y permiten caracterizar las condiciones bajo las cuales el enfoque propuesto mantiene su ventaja económica, entregando evidencia sobre la robustez del método frente al número de horas penalizadas y sobre su transferibilidad a regímenes de costos distintos del caso base.

### 4.4.1. Configuración del análisis de sensibilidad de los costos económicos

El primer análisis evalúa el comportamiento de las funciones de pérdida desarrolladas tanto en el escenario en que los falsos positivos tienen el mismo costo que los falsos negativos, como en los casos en que uno de ellos resulta más costoso que el otro. Para ello, se define el factor de sensibilidad presentado en la ecuación (4.2), el cual establece la relación entre los costos asociados a los falsos positivos y los falsos negativos generados por los avisos de los modelos.

$$Factor\ de\ sensibilidad = \frac{c(falsos\ positivos)}{c(falsos\ negativos)} = \frac{\pi^{gp} - \pi^{cs}}{(\pi^{cp} + \pi^{cs}) - \pi^{gp}} \quad (4.2)$$

El factor de sensibilidad se consideró dentro del intervalo  $[0.1, 10]$ , abarcando valores tanto menores como mayores a uno y garantizando la reciprocidad entre ellos, es decir, que cada valor sea el inverso multiplicativo de otro dentro del rango analizado. Una vez definidos los costos económicos para cada valor del factor de sensibilidad, se procede a entrenar y validar los modelos aplicando la misma metodología descrita en el Capítulo 3. Cabe mencionar que, para este análisis, únicamente se reentrenan y validan los modelos enfocados en costos, ya que el modelo enfocado en predicción no cuenta con una función de pérdida asimétrica que permita ajustar su entrenamiento según los costos económicos del caso de estudio.

Finalmente, con el fin de capturar la variabilidad asociada a la inicialización y al dropout, el experimento se repite cinco veces para cada valor del factor de sensibilidad, reportándose la media y la desviación estándar de las métricas obtenidas.

#### 4.4.2. Configuración del análisis de sensibilidad de las horas punta penalizadas

El segundo análisis evalúa la sensibilidad de los modelos frente al número de horas punta penalizadas  $k$ . A diferencia del factor de sensibilidad, el parámetro  $k$  es un parámetro estructural definido por el regulador, ya que corresponde a la cantidad de horas de máxima demanda sobre las que se aplica el cargo por capacidad del sistema. El propósito de este análisis es evaluar la robustez de los modelos desarrollados frente a variaciones en la cantidad de horas punta penalizadas, parámetro que difiere entre distintos sistemas y mercados eléctricos y que, por lo tanto, condiciona la transferibilidad de la metodología propuesta a otros contextos regulatorios.

En el caso de estudio chileno este valor se fija en 52 horas, pero dicha elección responde a una definición regulatoria particular y puede diferir en otros mercados de cargo por capacidad. Por ello, se evalúa el desempeño de los modelos considerando  $k \in \{28, 40, 52, 64, 76\}$ , donde 52 corresponde al caso base del estudio y los valores restantes representan desviaciones de  $\pm 12$  y  $\pm 24$  horas en torno a él.

Para este análisis se mantienen fijos los parámetros tarifarios del caso de estudio ( $\pi^{cp}$ ,  $\pi^{cs}$  y  $\pi^{gp}$ ), de modo que la variación observada en el desempeño se atribuya exclusivamente al número de horas penalizadas y no a un cambio simultáneo en la estructura de costos, dimensión ya abordada por el análisis anterior.

Dado que el número de horas punta forma parte de la definición del problema de clasificación, cada valor de  $k$  exige realizar un reentrenamiento a todos los modelos. En consecuencia, para cada  $k$  se realiza una nueva búsqueda en malla de hiperparámetros, siguiendo el mismo procedimiento de validación cruzada descrito en la Sección 4.3.3, y se selecciona la combinación correspondiente. Del mismo modo, el experimento se repite cinco veces para cada valor de  $k$ , reportándose la media y la desviación estándar de las métricas obtenidas.

Finalmente, dado que el costo total de operación crece con el número de horas punta penalizadas, las métricas económicas no resultan directamente comparables entre distintos valores de  $k$  en términos absolutos. Por esta razón, el arrepentimiento se normaliza como porcentaje del costo total de operación, lo que permite una comparación consistente del desempeño económico de los modelos a lo largo de los distintos valores de  $k$ .

# Capítulo 5

## Resultados y análisis

En este capítulo se presentan los resultados obtenidos a partir de la implementación de los modelos de predicción de horas punta en el subsistema de distribución de la Región Metropolitana de Santiago de Chile, utilizando la metodología desarrollada en el Capítulo 3.

### 5.1. Desempeño de los modelos

#### 5.1.1. Avisos de los modelos

A continuación, se presentan las matrices de confusión correspondientes a los modelos enfocado en predicción, enfocado en costos y enfocado en costos modificado. Estas se muestran en las Tablas 5.1, 5.2 y 5.3, respectivamente, y resumen los resultados obtenidos, indicando la cantidad de horas clasificadas correcta e incorrectamente como punta o no punta.

Los resultados se presentan en términos de promedios sobre las 10 iteraciones junto con la desviación estándar, y se interpretan de la siguiente manera: los verdaderos positivos (VP) son las horas que fueron correctamente clasificadas como punta; los falsos positivos (FP) son las horas que fueron incorrectamente clasificadas como punta; los verdaderos negativos (VN) son las horas que fueron correctamente clasificadas como no punta; y los falsos negativos (FN) son las horas que fueron incorrectamente clasificadas como no punta.

Tabla 5.1: Matriz de Confusión del modelo enfocado en predicción.

|            |         | Clase predicha |            |
|------------|---------|----------------|------------|
|            |         | No peak        | Peak       |
| Clase real | No peak | 458.4 ± 2.2    | 9.6 ± 2.2  |
|            | Peak    | 20.4 ± 2.6     | 31.6 ± 2.6 |

Tabla 5.2: Matriz de Confusión del modelo enfocado en costos.

|            |         | Clase predicha  |                |
|------------|---------|-----------------|----------------|
|            |         | No peak         | Peak           |
| Clase real | No peak | 460.4 $\pm$ 3.2 | 7.6 $\pm$ 3.2  |
|            | Peak    | 18.8 $\pm$ 0.8  | 33.2 $\pm$ 0.8 |

Tabla 5.3: Matriz de Confusión del modelo enfocado en costos modificado.

|            |         | Clase predicha  |                |
|------------|---------|-----------------|----------------|
|            |         | No peak         | Peak           |
| Clase real | No peak | 448.6 $\pm$ 3.1 | 19.4 $\pm$ 3.1 |
|            | Peak    | 14.4 $\pm$ 2.9  | 37.6 $\pm$ 2.9 |

Las métricas de calidad se calcularon a partir de los resultados de las matrices de confusión y se presentan en la Tabla 5.4. Estas métricas incluyen la exactitud, la precisión, la recuperación y el puntaje F1.

Tabla 5.4: Desempeño del modelo en términos de métricas de calidad promedio en la predicción.

| Modelo                               | Exactitud promedio | Precisión promedio | Recuperación promedio | Puntaje F1 prom.  |
|--------------------------------------|--------------------|--------------------|-----------------------|-------------------|
| <b>Enfocado en predicción</b>        | 0.9423 $\pm$ 0.01  | 0.7692 $\pm$ 0.02  | 0.6077 $\pm$ 0.05     | 0.6773 $\pm$ 0.02 |
| <b>Enfocado en costos</b>            | 0.9492 $\pm$ 0.01  | 0.8172 $\pm$ 0.06  | 0.6385 $\pm$ 0.01     | 0.7161 $\pm$ 0.02 |
| <b>Enfocado en costos Modificado</b> | 0.9350 $\pm$ 0.01  | 0.6610 $\pm$ 0.02  | 0.7231 $\pm$ 0.06     | 0.6893 $\pm$ 0.02 |

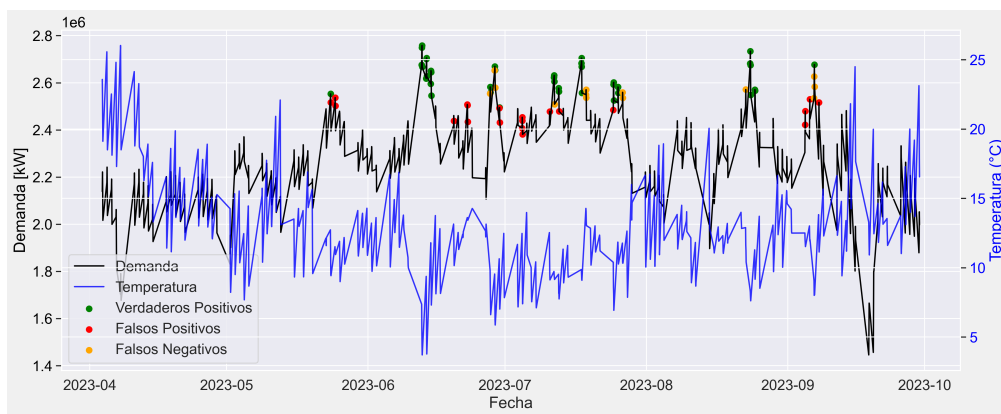
En este contexto, las métricas de calidad seleccionadas se interpretan en función de la relación entre la cantidad de avisos de horas punta generados, la cantidad de horas punta correctamente predichas y las 52 horas punta reales observadas en el periodo de control. Bajo este criterio, la métrica de precisión indica el porcentaje de avisos que resultaron acertados respecto de la totalidad de avisos emitidos. En el caso del modelo enfocado en predicción, en promedio se emitieron 41.2 avisos, de los cuales 31.6 coincidieron con las horas punta reales, alcanzando una precisión promedio del 76.92 %. El modelo enfocado en costos emitió en promedio 40.8 avisos, de los cuales 33.2 fueron correctos, obteniendo una precisión del 81.72 % y posicionándose como el mejor modelo en esta métrica. Finalmente, el modelo enfocado en costos modificado acertó en promedio 37.6 de 57.0 avisos emitidos, lo que representa una precisión del 66.10 %.

La métrica de recuperación indica el porcentaje de las 52 horas punta reales que fueron correctamente identificadas por el modelo. Bajo este criterio, el modelo enfocado en predicción identificó en promedio 31.6 de las 52 horas punta, resultando en una recuperación promedio del 60.77%. Por su parte, el modelo enfocado en costos predijo correctamente en promedio 33.2 de las 52 horas punta, alcanzando una recuperación del 63.85%. Finalmente, el modelo enfocado en costos modificado identificó en promedio 37.6 de las 52 horas punta, alcanzando una recuperación del 72.31% y posicionándose como el mejor modelo en esta métrica.

La métrica de exactitud representa el porcentaje de aciertos, considerando tanto de horas punta como de horas no punta, dentro del total de 520 horas del periodo de control. Bajo este criterio, el modelo enfocado en predicción logró una exactitud promedio del 94.23%, mientras que el modelo enfocado en costos obtuvo un 94.92% y el modelo enfocado en costos modificado registró 93.50%. Por último, el puntaje F1 (*F1-score*), definido como la media armónica entre precisión y recuperación, refleja un balance entre la cantidad de avisos de horas punta emitidos y los aciertos obtenidos. En este caso, el modelo enfocado en predicción logró un puntaje F1 promedio del 67.73%, el modelo enfocado en costos alcanzó un 71.61% y el modelo enfocado en costos modificado obtuvo un 68.93%.

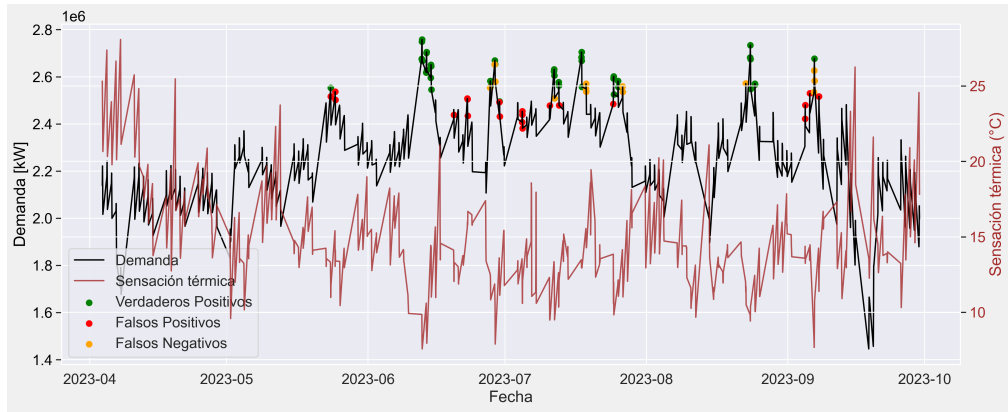
### 5.1.2. Análisis conjunto de curvas de demanda, variables climáticas y avisos de predicción

En la Figura 5.1 se presentan de manera conjunta la curva de demanda eléctrica del subsistema de distribución de la Región Metropolitana y cada una de las cuatro variables climáticas consideradas (temperatura, sensación térmica, humedad relativa y presión atmosférica, respectivamente), junto con los avisos generados por el modelo. Con el fin de permitir una comparación directa entre variables, la figura fija un único modelo, el enfocado en costos modificado, por ser el de mejor recuperación de horas punta. De esta forma los avisos permanecen constantes entre paneles y la única diferencia entre ellos es la variable climática superpuesta, lo que permite aislar la relación entre cada variable y las horas de máxima demanda. Las gráficas equivalentes para los modelos enfocado en predicción y enfocado en costos se presentan, para cada variable, en el Anexo C.1.

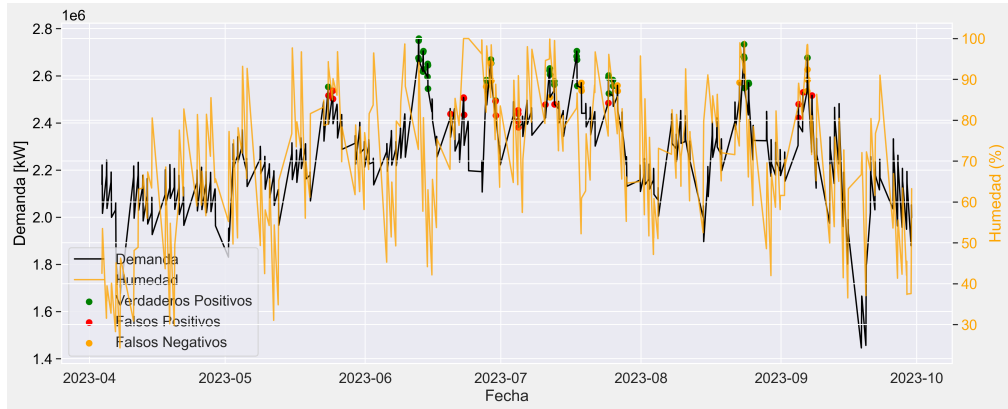


(a) Temperatura.

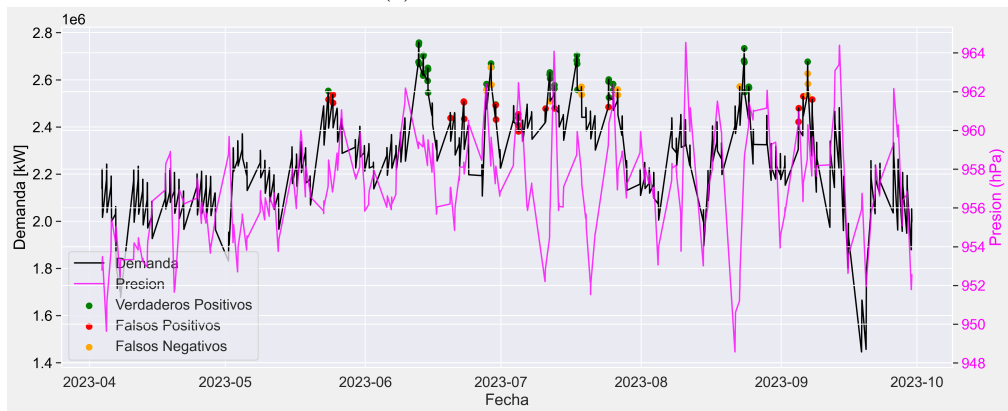
Figura 5.1: Curvas de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos modificado.



(b) Sensación térmica.



(c) Humedad relativa.



(d) Presión atmosférica.

Figura 5.1: Curvas de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos modificado (continuación).

A partir de la figura se observa que los avisos de horas punta no se distribuyen de manera homogénea respecto de todas las variables climáticas. En el caso de la temperatura y de la sensación térmica, los avisos tienden a concentrarse en los tramos en que ambas variables alcanzan sus valores más bajos, correspondientes al periodo invernal.

Lo anterior puede considerarse una característica positiva de los modelos, en tanto resulta coherente que, a medida que disminuye la temperatura, la demanda eléctrica tienda a au-

mentar debido al mayor uso de sistemas de calefacción y otros dispositivos eléctricos. Esta relación es especialmente relevante en el subsistema de distribución analizado, dado que la mayoría de los clientes corresponde al segmento residencial, lo que incrementa la probabilidad de que la demanda responda a patrones de consumo agregados y relativamente homogéneos. No obstante, esta misma característica puede constituir una limitación si la temperatura no se comporta como un predictor robusto de la demanda en todas las circunstancias, o si el modelo no logra incorporar adecuadamente otros factores que también influyen en su variación.

Esta dualidad se refleja en distintos intervalos de las figuras analizadas. En particular, los modelos predicen correctamente las horas punta de mediados de junio, periodo que registra la temperatura más baja del horizonte temporal analizado. Sin embargo, también se observan periodos con temperaturas igualmente bajas en los que no se registran horas punta, como ocurre durante la primera semana de julio.

A partir de lo anterior se concluye que los modelos desarrollados son sensibles a la temperatura y a la sensación térmica. Ahora bien, si bien la temperatura mínima del periodo se muestra como un buen indicador de horas punta, existe un umbral a partir del cual esta variable deja de comportarse como un predictor confiable. Esto sugiere que, en determinados escenarios, otros factores adquieren mayor relevancia en la explicación de los máximos de demanda eléctrica.

Cabe precisar que la temperatura y la sensación térmica comparten parte de su señal por construcción, dado que esta última se calcula a partir de la temperatura y de la velocidad del viento, por lo que es esperable que su comportamiento frente a los avisos sea similar.

En contraste, para la humedad relativa y la presión atmosférica los avisos se reparten de manera más difusa a lo largo del rango de ambas variables, sin concentrarse de forma evidente en torno a valores extremos. Esto sugiere que, en el periodo analizado, la coocurrencia entre las horas punta y estas dos variables es menos marcada que la observada para la temperatura y la sensación térmica.

### **5.1.3. Evaluación económica**

A partir de los resultados obtenidos de los modelos, se calcularon las métricas de desempeño económico, las cuales se presentan en la Tabla 5.5. Estas métricas consideran, en primer lugar, el arrepentimiento, definido como la diferencia entre los costos asociados a las decisiones tomadas por cada modelo y aquellos asociados a las decisiones óptimas. Y en segundo lugar, se incluye el costo total de operación, que representa el el gasto acumulado asociado a las decisiones tomadas por cada modelo durante todo el periodo de control de horas punta.

Tabla 5.5: Desempeño de los modelos en términos de métricas económicas promedio.

| Modelo                               | Arrepentimiento promedio [USD/MW] | Costo total promedio de la operación [USD/MW] |
|--------------------------------------|-----------------------------------|---|
| <b>Enfocado en predicción</b>        | 4095.12 $\pm$ 417.59              | 54498.46 $\pm$ 417.59                         |
| <b>Enfocado en costos</b>            | 3733.10 $\pm$ 207.30              | 54136.44 $\pm$ 207.30                         |
| <b>Enfocado en costos Modificado</b> | 3304.03 $\pm$ 486.81              | 53707.37 $\pm$ 486.81                         |

Además, con el objetivo de comparar los resultados obtenidos por los modelos, se consideraron tres políticas adicionales: (i) una política que consiste en operar siempre el medio de generación propia durante las horas de control, (ii) una política que utiliza exclusivamente la red en dichas horas, y (iii) una política óptima que asume información perfecta respecto a si cada hora corresponde o no a un *peak*, acertando por tanto en todos los casos. En la Tabla 5.6 se presentan las métricas económicas correspondientes a estas políticas comparativas.

Tabla 5.6: Métricas económicas de las políticas comparativas.

| Modelo                        | Arrepentimiento [USD/MW] | Costo total de la operación [USD/MW] |
|-------------------------------|--------------------------|--------------------------------------|
| <b>Uso solo del generador</b> | 15324.660                | 65728.000                            |
| <b>Uso solo de la red</b>     | 9637.264                 | 60040.604                            |
| <b>Operación óptima</b>       | -                        | 50403.340                            |

Luego, en la Figura 5.2 se presenta un desglose detallado de los distintos componentes que integran el costo total de operación del cliente durante todo el periodo de control de horas punta, tanto para los modelos desarrollados como para las políticas comparativas. Estos costos incluyen los costos totales por la energía consumida desde la red, los costos asociados al consumo durante horas punta y los costos derivados del uso del generador eléctrico estándar, tanto cuando su operación es correcta como cuando se utiliza erróneamente.

Al analizar el desempeño del modelo enfocado en costos, se observa que en promedio identificó cerca de dos horas punta adicionales respecto al modelo de referencia, pasando de 31.6 a 33.2 verdaderos positivos. Cada una de estas horas correctamente anticipada permite desplazar el consumo hacia el medio de generación propia en lugar de abastecerse desde la red durante una hora punta, evitando así el cargo por capacidad, lo que representa un ahorro de 185.332 [USD/MW] por hora. Al mismo tiempo, el modelo redujo en promedio en dos la cantidad de falsos avisos, de 9.6 a 7.6 falsos positivos. Cada falso positivo evitado equivale a un encendido innecesario menos del generador, lo que significa que en lugar de operar la generación propia sin que ocurra una hora punta, el cliente se abastece correctamente desde la red, con un ahorro de 32.745 [USD/MW] en cada caso. La combinación de ambos efectos se traduce en una reducción promedio del costo total de operación de 0.66 %, equivalente a 362.02 [USD/MW], respecto al modelo de referencia.

Esta mejora es relevante porque se produce en ambos tipos de error a la vez, dado que el modelo enfocado en costos disminuye en promedio tanto los falsos positivos, de 9.6 a 7.6, como los falsos negativos, de 20.4 a 18.8, respecto a la referencia. Un simple desplazamiento

del umbral de decisión sobre un mismo clasificador intercambia un tipo de error por otro y no puede reducir ambos de forma simultánea. El hecho de que el modelo enfocado en costos sí lo logre indica que la pérdida basada en el arrepentimiento no se limita a reponderar las clases, sino que induce una frontera de decisión distinta, mejor alineada con la estructura de costos del problema.

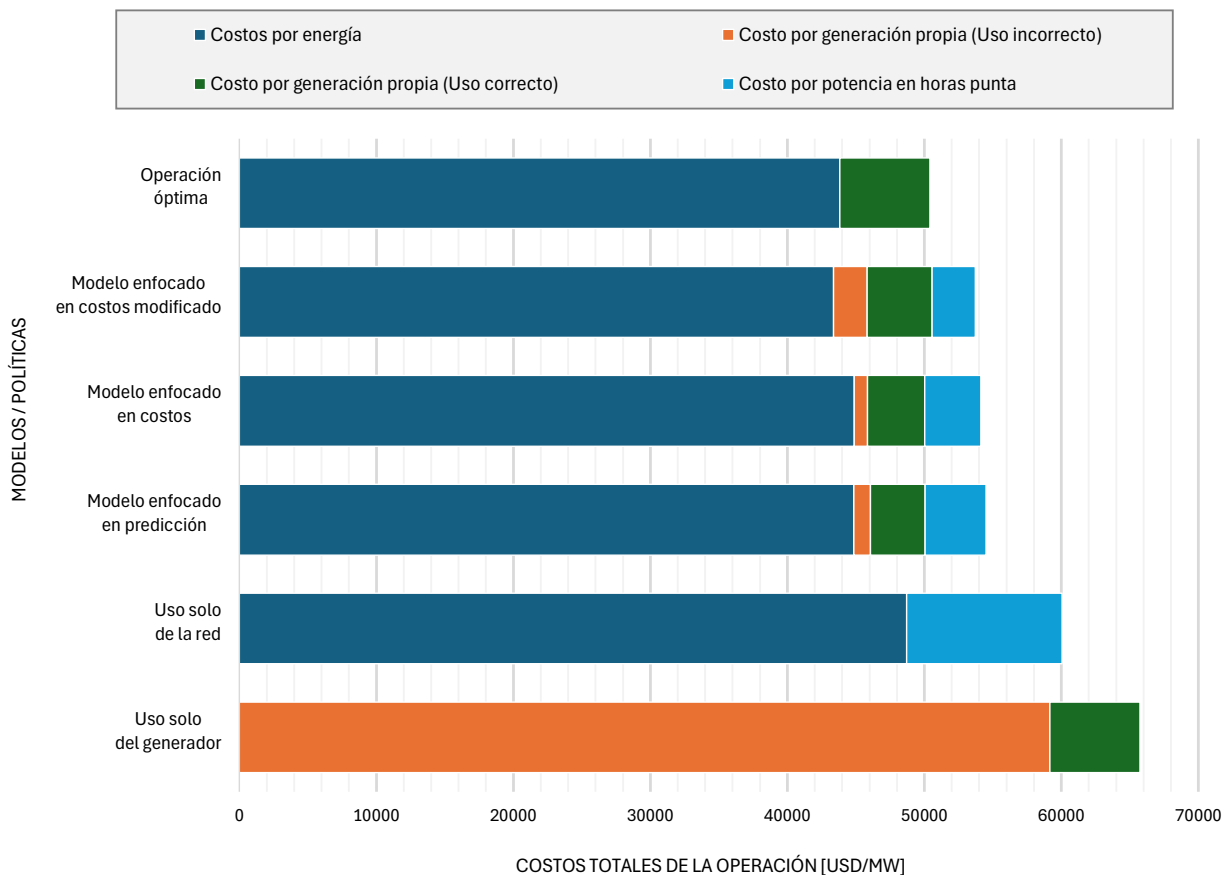


Figura 5.2: Desglose de los costos de operación promedio de los modelos desarrollados durante todo el periodo de control de horas punta.

En cuanto al modelo enfocado en costos modificado, en promedio identificó cerca de seis horas punta más que el modelo de referencia, cada una con el mismo ahorro de 185.332 [USD/MW] por desplazar el consumo mediante generación propia. Para lograrlo, sin embargo, incurrió en promedio en cerca de diez encendidos innecesarios adicionales del generador, es decir, diez falsos positivos más, cada uno con un sobre costo de 32.745 [USD/MW] respecto a haberse abastecido desde la red. Pese a este mayor número de operaciones incorrectas, el modelo obtuvo un arrepentimiento menor que el de la referencia y redujo el costo total de operación en un 1.45 %, equivalente a 791.09 [USD/MW].

Este resultado vuelve explícita una aparente tensión entre las métricas de calidad y el desempeño económico. El modelo enfocado en costos modificado registra la peor precisión, 66.10 %, y la peor exactitud, 93.50 %, de los tres modelos y, sin embargo, alcanza el menor arrepentimiento. Lejos de ser contradictorio, este comportamiento es el objetivo buscado por la formulación. La clave está en la asimetría de los costos unitarios: anticipar correctamente

una hora punta aporta un beneficio de 185.3 [USD/MW], cerca de 5.7 veces el sobre costo de 32.7 [USD/MW] de cada encendido innecesario del generador. Bajo esta relación, el punto de operación económicamente óptimo no es el que equilibra ambos errores, sino uno de alta recuperación y baja precisión, en el que conviene tolerar varios falsos positivos con tal de no omitir una hora punta. Las métricas de clasificación simétricas penalizan precisamente ese régimen, lo que explica que el modelo más eficiente en términos económicos sea, a la vez, el peor evaluado por ellas. La función de pérdida asimétrica desplaza la frontera de decisión hacia ese régimen, algo que la entropía cruzada binaria no puede lograr por construcción, dado que pondera por igual ambos errores.

Esta diferencia de estrategias se refleja en el desglose de costos de la Figura 5.2. El modelo enfocado en costos, con 40.8 avisos en promedio frente a los 41.2 de la referencia y los 57.0 del modelo modificado, es el más conservador y el que presenta la menor cantidad de horas de uso incorrecto del generador. El modelo enfocado en costos modificado, en cambio, concentra un mayor uso de la generación propia y un menor uso de la red, lo que resulta coherente con su disposición a emitir más avisos para no omitir horas punta.

Luego, al comparar los resultados obtenidos por los modelos con las políticas comparativas mostradas en la Tabla 5.6 y en la Figura 5.2, se observa que ambos modelos lograron reducir significativamente el costo total de operación frente a la política de utilizar siempre el generador y a la de utilizar siempre la red durante el periodo de control. En particular, el modelo enfocado en costos logró una reducción del 17.64 % respecto a la política de utilizar siempre el generador, equivalente a 11591.56 [USD/MW], mientras que el modelo enfocado en costos modificado alcanzó una reducción del 18.29 %, correspondiente a 12020.63 [USD/MW]. De manera análoga, frente a la política de utilizar siempre la red, el modelo enfocado en costos logró una reducción del 9.83 %, equivalente a 5904.16 [USD/MW], y el modelo enfocado en costos modificado una del 10.55 %, equivalente a 6333.23 [USD/MW].

En suma, ambos modelos enfocados en costos mejoran el desempeño económico de la operación, tanto frente al modelo de referencia enfocado en predicción como frente a las políticas consideradas, generando ahorros relevantes respecto de los modelos convencionales basados en funciones de pérdida simétricas y, en especial, respecto de no utilizar ningún mecanismo de predicción.

Una vez caracterizado el desempeño económico promedio de cada modelo, resta determinar si las diferencias de costo observadas frente al modelo de referencia son atribuibles al desempeño de los modelos o simplemente a la variabilidad entre semillas. Con este fin, se aplicó un test  $t$  de Welch sobre el costo total de operación de las 10 réplicas de cada modelo, tomando como referencia el modelo enfocado en predicción. La Tabla 5.7 resume, para cada comparación, la reducción media de costo ( $\Delta$ ), su intervalo de confianza al 95 %, el estadístico  $t$ , los grados de libertad y el valor- $p$  asociado.

La Tabla 5.7 muestra que ambos modelos con pérdida asimétrica reducen el costo de operación respecto a la referencia de forma estadísticamente significativa ( $p < 0.05$ ). La reducción del modelo enfocado en costos modificado es la mayor y la más robusta, de 791.1 [USD/MW], con un intervalo de confianza al 95 % que se mantiene holgadamente por sobre cero. La del modelo enfocado en costos, si bien significativa, es de menor magnitud y su intervalo de confianza llega a valores cercanos a cero, con un límite inferior de 44 [USD/MW], por lo que debe interpretarse como una mejora real pero acotada. En conjunto, el test confirma

Tabla 5.7: Comparación del costo total de operación frente al modelo de referencia (enfocado en predicción), mediante test t de Welch sobre las 10 réplicas.  $\Delta$  es la reducción media de costo, expresada en USD/MW.

| Comparación         | $\Delta$ | IC 95 %     | $t$  | gl   | $p$   |
|---------------------|----------|-------------|------|------|-------|
| Modificado vs. ref. | 791.1    | [364, 1218] | 3.90 | 17.6 | 0.001 |
| Costos vs. ref.     | 362.0    | [44, 680]   | 2.46 | 13.2 | 0.029 |

que la ventaja económica de ambos modelos no proviene del azar entre semillas, sino que constituye un efecto atribuible a la formulación basada en el arrepentimiento.

## 5.2. Análisis de sensibilidad

### 5.2.1. Análisis de sensibilidad de los costos

En la Figura 5.3 se presenta el análisis de sensibilidad de los modelos frente a distintos valores del factor de sensibilidad, definido en la ecuación 4.2 como la razón entre el costo de un falso positivo y el de un falso negativo. Su propósito es evaluar la robustez de los modelos ante variaciones en dichos costos y determinar cómo afectan su desempeño económico. Para aligerar la lectura, se denota por  $s$  al logaritmo en base 10 de ese factor,  $s = \log_{10}(\text{Factor de sensibilidad})$ . De este modo,  $s = 0$  corresponde a costos de falsos positivos y negativos iguales, y dos valores opuestos  $\pm s$  representan asimetrías de igual magnitud pero sentido contrario. El eje horizontal de la figura corresponde a  $s$ , y el arrepentimiento se expresa como porcentaje del costo total de operación, de modo que las diferencias en la magnitud absoluta de los costos no afecten la comparación.

El signo de  $s$  indica cuál de los dos errores es más costoso y, con ello, la estrategia que conviene adoptar. Cuando  $s > 0$ , activar la generación propia en una hora que no es punta (un falso positivo) resulta más costoso que omitir una hora punta (un falso negativo), por lo que conviene emitir menos avisos y tolerar falsos negativos. Cuando  $s < 0$  la situación se invierte: omitir una hora punta pasa a ser lo más costoso y conviene emitir más avisos. En  $s = 0$  ninguna decisión es claramente preferible y el problema se aproxima a una clasificación convencional.

A partir de esta lectura, la figura distingue dos regímenes. Hacia los extremos, donde la asimetría de costos es marcada ( $|s| \geq 0.3$ , esto es, un factor de sensibilidad mayor que dos o menor que un medio), los modelos enfocados en costos superan al modelo enfocado en predicción. Una diferencia considerable entre el costo de ambos errores vuelve claramente más conveniente una de las dos decisiones, lo que permite que durante el entrenamiento el modelo desarrolle una inclinación definida hacia la opción de menor costo y la aproveche al predecir, reduciendo el arrepentimiento total. Esta inclinación se traduce en estrategias opuestas según el signo de  $s$ . Cuando  $s < 0$ , los modelos enfocados en costos emiten un mayor número de avisos, dado que omitir una hora punta resulta muy costoso frente al ahorro de desplazarla con generación propia. Cuando  $s > 0$  emiten menos avisos, porque el alto costo de utilizar

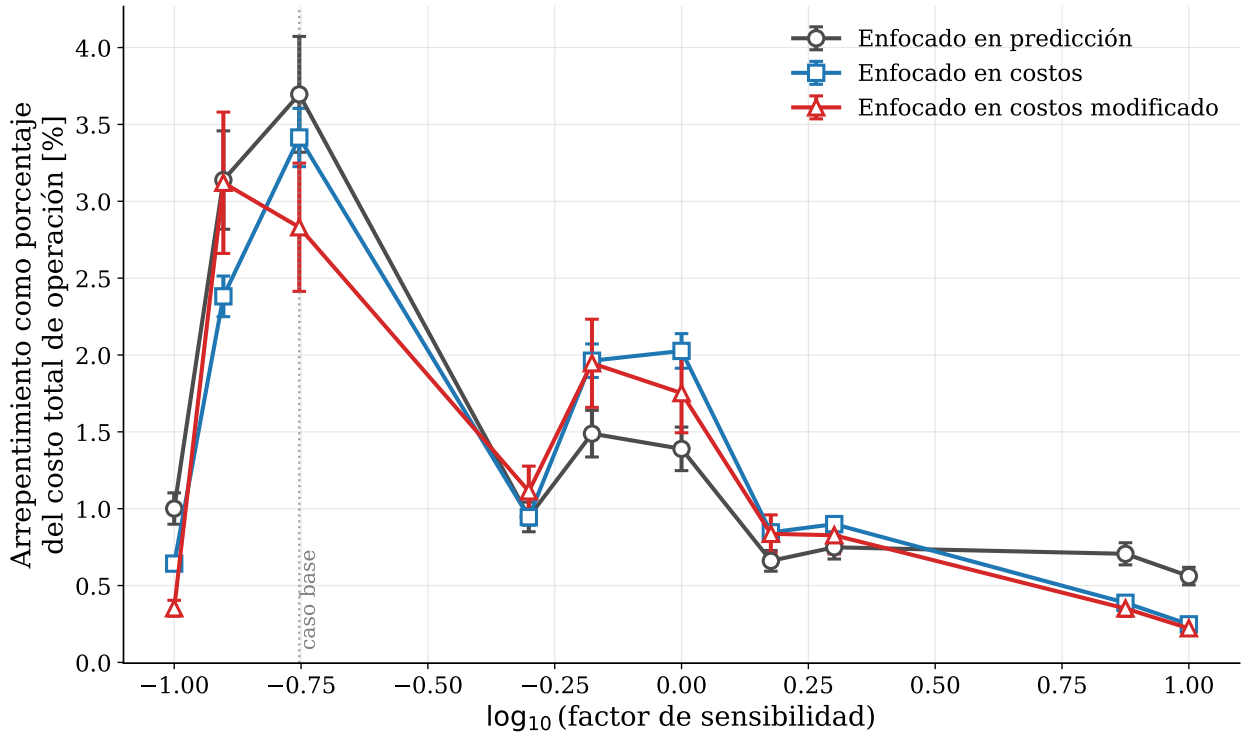


Figura 5.3: Análisis de sensibilidad del arrepentimiento como porcentaje del costo total de operación, en función del factor de sensibilidad. La línea punteada indica el caso base.

la generación propia no se compensa con el ahorro de evitar el cargo por capacidad, lo que aumenta los falsos negativos y la frecuencia con que se consume desde la red en horas punta.

Hacia el centro de la figura esta jerarquía se invierte. En el entorno del punto simétrico, aproximadamente para  $|s| \leq 0.3$  (esto es, cuando ninguno de los dos errores cuesta más del doble que el otro), el modelo enfocado en predicción iguala o supera a ambos modelos enfocados en costos. La diferencia es máxima en  $s = 0$ , donde el arrepentimiento del modelo enfocado en predicción alcanza un 1.39% del costo total, frente a un 1.75% del modelo enfocado en costos modificado y un 2.03% del modelo enfocado en costos. Esta ventaja, del orden de medio punto porcentual, se atenúa de forma gradual hacia los bordes del intervalo hasta desaparecer en el cruce con las curvas de los modelos enfocados en costos.

Este comportamiento admite una explicación en dos niveles. Por una parte, en el régimen casi simétrico la función de pérdida desarrollada pierde su componente direccional, ya que la diferencia entre el costo de un falso positivo y el de un falso negativo es demasiado pequeña para inducir una inclinación clara hacia un tipo de decisión, de modo que el mecanismo que sustenta su ventaja en los extremos deja de operar. Por otra parte, la pérdida no degenera exactamente en la entropía cruzada simétrica del modelo de referencia, pues conserva la ponderación de cada hora según la magnitud económica de su error. En ausencia de asimetría direccional, esa ponderación concentra la señal de entrenamiento en un número reducido de horas de alto costo en lugar de distribuirla sobre el conjunto completo de observaciones, lo que, bajo la escasez estructural de datos propia del problema, produce un clasificador menos robusto que el del modelo enfocado en predicción, que aprovecha de manera uniforme todas

las horas disponibles. Como en el régimen simétrico una buena clasificación coincide con un buen desempeño económico, la mejor generalización del modelo de referencia se traduce indirectamente en un menor arrepentimiento.

Lejos de constituir una debilidad, este resultado delimita con precisión el dominio de aplicabilidad de la metodología propuesta. La superioridad de los modelos enfocados en costos no es universal sino condicional, pues se manifiesta cuando existe una asimetría de costos suficiente que explotar y se desvanece cuando dicha asimetría es débil. Esta frontera puede expresarse de forma operativa mediante el propio valor de  $s$ . Para un practicante de otro mercado, el enfoque asimétrico resulta preferible cuando  $|s|$  supera aproximadamente 0.3, es decir, cuando la razón entre los costos de falsos positivos y falsos negativos se aleja de la unidad más allá de un factor cercano a dos, mientras que en regímenes de costos aproximadamente simétricos un clasificador convencional bien ajustado es la opción más conveniente. Esta caracterización convierte a  $s$  en un criterio explícito sobre cuándo conviene adoptar el método propuesto, lo que constituye en sí mismo un aporte para su transferencia a otros esquemas de cargo por capacidad.

### 5.2.2. Análisis de sensibilidad de las horas punta penalizadas

En la Figura 5.4 se muestra la evolución del arrepentimiento, expresado como porcentaje del costo total de operación, en función del número de horas punta penalizadas.

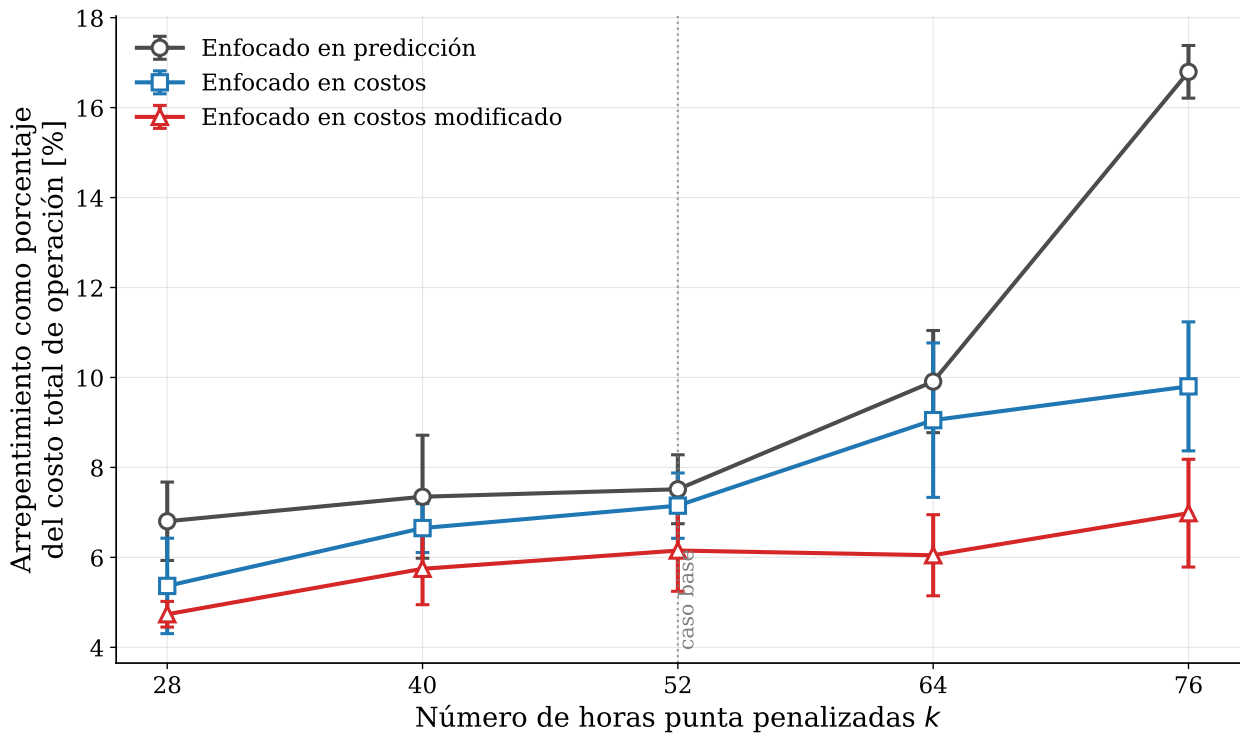


Figura 5.4: Análisis de sensibilidad del arrepentimiento como porcentaje del costo total de operación, en función del número de horas punta penalizadas  $k$ . La línea punteada indica el caso base ( $k = 52$ ).

En primer lugar, se observa que el arrepentimiento aumenta con  $k$  para los tres modelos.

Esto resulta esperable, ya que un mayor número de horas penalizadas implica, por un lado, una mayor cantidad de horas punta que el modelo debe identificar correctamente y, por otro, una mayor exposición acumulada al cargo por capacidad, de modo que cada error de clasificación tiene un impacto económico potencialmente mayor.

En segundo lugar, el modelo enfocado en costos modificado obtiene el menor arrepentimiento en todos los valores de  $k$  evaluados, manteniéndose como el modelo de mejor desempeño económico a lo largo de todo el rango analizado. Esta consistencia indica que la ventaja de la función de pérdida asimétrica no depende de la configuración particular del caso de estudio, sino que se sostiene tanto en esquemas de penalización menos exigentes como más exigentes que el chileno.

En tercer lugar, el modelo de referencia enfocado en predicción es el que más se deteriora a medida que aumenta  $k$ . Mientras que para valores bajos del parámetro su arrepentimiento es comparable al de los modelos enfocados en costos (6.80% para  $k = 28$ ), para el escenario más exigente este se dispara hasta un 16.80% del costo total para  $k = 76$ . Este comportamiento se explica por la caída en la recuperación del modelo enfocado en predicción, la cual desciende hasta 0.28 para  $k = 76$ : al ser entrenado con una función de pérdida simétrica que no incorpora la magnitud de los costos, el modelo adopta una postura excesivamente conservadora, emitiendo pocos avisos de hora punta y dejando sin identificar la mayoría de las horas penalizadas, justamente en el escenario donde no acertarlas es más costoso.

Por el contrario, el modelo enfocado en costos modificado conserva una recuperación elevada incluso para valores altos de  $k$  (0.80 para  $k = 64$  y 0.77 para  $k = 76$ ), lo que refleja su tendencia a emitir un mayor número de avisos. Esta estrategia, más agresiva en la identificación de horas punta, resulta económicamente ventajosa precisamente cuando la cantidad de horas penalizadas es grande, dado que el costo de omitir una hora punta se acumula sobre un número mayor de instancias. Como resultado, la ventaja económica de la función de pérdida asimétrica respecto al modelo de referencia se amplía a medida que aumenta  $k$ : la diferencia entre el modelo enfocado en costos modificado y el modelo enfocado en predicción crece desde aproximadamente 1 a 2 puntos porcentuales en el rango bajo y central, hasta cerca de 10 puntos porcentuales para  $k = 76$ .

En el otro extremo, para valores bajos de  $k$  los tres modelos exhiben un desempeño más cercano entre sí. Este resultado es coherente con la lógica observada en el análisis de sensibilidad del factor de sensibilidad: cuando la penalización es menos extensa, la presión económica que induce la asimetría de la función de pérdida es menor, por lo que los modelos enfocados en costos no logran desarrollar una inclinación tan marcada hacia un tipo de decisión, y su ventaja respecto al modelo simétrico se reduce.

Cabe señalar, por último, el comportamiento del modelo enfocado en costos para  $k = 64$ , donde su arrepentimiento se eleva hasta un 9.05%, acercándose al del modelo de referencia y separándose del modelo enfocado en costos modificado. Este punto presenta, además, la mayor desviación estándar registrada en el análisis ( $\pm 958.8$  [USD/MW]), lo que evidencia una menor estabilidad del modelo enfocado en costos en este régimen intermedio. Pese a ello, dicho modelo continúa superando en promedio al modelo de referencia enfocado en predicción, mientras que el modelo enfocado en costos modificado mantiene un desempeño claramente superior y estable.

Finalmente, a partir del análisis de sensibilidad de las horas punta penalizadas se concluye que la ventaja económica de los modelos basados en funciones de pérdida asimétricas es robusta frente a variaciones en el número de horas penalizadas  $k$ : en todo el rango evaluado, los modelos enfocados en costos obtienen un mejor desempeño económico que el modelo de referencia —siendo el modelo enfocado en costos modificado el de mejor desempeño en cada caso— y dicha ventaja se vuelve más pronunciada en los escenarios de penalización más extensa. Este resultado complementa el análisis del factor de sensibilidad, dado que muestra que el beneficio de la metodología propuesta no se restringe a la configuración particular del caso de estudio chileno, sino que se mantiene —e incluso se acentúa— ante esquemas regulatorios que penalicen un número distinto de horas punta.

### 5.2.3. Conclusión conjunta de los análisis de sensibilidad.

Conviene precisar cómo se relacionan estas dos dimensiones con la estructura tarifaria específica del caso de estudio. Los tres parámetros de costo del problema ( $\pi^{cs}$ ,  $\pi^{cp}$  y  $\pi^{gp}$ ) no inciden de manera independiente sobre la conveniencia del enfoque propuesto. Por una parte, la función de pérdida asimétrica distingue entre los dos tipos de error únicamente a través de la razón entre sus costos, esto es, el factor de sensibilidad definido en la ecuación 4.2, del cual  $s$  es su logaritmo en base 10. De este modo, la influencia de la estructura tarifaria sobre lo que los modelos aprenden queda mediada por dicho factor, de manera que distintas tarifas que comparten una misma razón de costos conducen al mismo comportamiento. Por otra parte, el nivel absoluto de los cargos no altera el desempeño relativo de los modelos, pues al expresar el arrepentimiento como porcentaje del costo total de operación la métrica se vuelve invariante frente a un reescalamiento proporcional de las tres tarifas. La estructura tarifaria incide entonces sobre el problema a través de dos magnitudes que los análisis de sensibilidad recorren de forma explícita: el factor de sensibilidad, resumido por  $s$ , y el número de horas punta penalizadas  $k$ . Por ello, los resultados no dependen de los valores tarifarios particulares del caso chileno, sino de su posición en el plano  $(s, k)$ , lo que permite caracterizar la dependencia respecto de la estructura tarifaria de cualquier mercado de cargo por capacidad a partir de las dos dimensiones estudiadas.

Los dos análisis de sensibilidad desarrollados, considerados en conjunto, permiten delimitar una región de operación dentro de la cual los modelos enfocados en costos resultan más eficientes que el modelo convencional de referencia. Esta región queda caracterizada por dos dimensiones. La primera es la relación de costos entre los falsos positivos y los falsos negativos, estudiada en el análisis del factor de sensibilidad, que mostró que la ventaja de la metodología propuesta crece a medida que aumenta la asimetría entre ambos costos. La segunda es la frecuencia del evento, es decir, el número de horas punta penalizadas  $k$ , estudiada en el análisis de las horas punta, que mostró que dicha ventaja se mantiene al aumentar o disminuir la cantidad de horas penalizadas. De este modo, ambos análisis acotan en conjunto una zona del espacio de operación en la que el desempeño superior de los modelos desarrollados se encuentra respaldado por la evidencia obtenida.

A partir de lo anterior, resulta verosímil extender este comportamiento hacia zonas más amplias del espacio considerado. Esto se debe a que la ventaja de los modelos enfocados en costos no proviene de los valores numéricos particulares del caso de estudio, sino del hecho de alinear el entrenamiento con la estructura económica de las decisiones del cliente. Dado

que dicho mecanismo opera con independencia de la configuración específica de costos y de la cantidad de horas penalizadas, es esperable que la región de eficiencia identificada se extienda más allá de los puntos efectivamente evaluados, abarcando combinaciones intermedias de ambas dimensiones.

Con todo, esta extensión descansa sobre el supuesto de que el comportamiento observado de manera independiente en cada dimensión se conserva al considerarlas de forma conjunta. Los resultados obtenidos hacen razonable dicho supuesto, y su confirmación sobre la totalidad del espacio de operación queda como una verificación que podría abordarse en trabajos futuros, en particular en las regiones más alejadas del caso base. Esta precisión acota el alcance de la lectura conjunta sin restar validez a la región de eficiencia ya establecida.

En síntesis, los análisis de sensibilidad realizados, junto con esta lectura conjunta, permiten concluir que la metodología propuesta es extrapolable a otros mercados de cargo por capacidad. Su desempeño no se encuentra atado a las condiciones particulares del mercado de potencia chileno, sino que se sostiene ante distintas relaciones de costos y distintos números de horas penalizadas, características que varían precisamente de un mercado a otro. De esta forma, el aporte de este trabajo trasciende el caso de estudio analizado y resulta aplicable a una clase más amplia de esquemas regulatorios que remuneran la suficiencia de capacidad mediante cargos a la demanda en horas punta.

# Capítulo 6

## Conclusiones

### 6.1. Conclusiones

El presente trabajo abordó el problema de un consumidor eléctrico sujeto a cargos por capacidad, quien debe anticipar las horas de máxima demanda del sistema para decidir cuándo recurrir a un medio de generación propia y reducir así su costo de operación. La dificultad central de este problema no es únicamente predictiva, sino estructural, ya que las horas punta se definen una vez por periodo de facturación sobre la demanda agregada del sistema, de modo que la cantidad de instancias de entrenamiento disponibles es intrínsecamente escasa, condición común a cualquier mecanismo de cargo por capacidad y no una limitación particular del caso chileno estudiado. Para sortear esta escasez, la metodología propuesta reformuló la selección de las  $k$  horas punta como un problema de clasificación binaria a nivel horario. Esta decisión implica un compromiso explícito, en donde se sacrifica la estructura combinatoria y competitiva del problema original, en el que las horas compiten entre sí por pertenecer al conjunto de las  $k$  de mayor demanda, pero a cambio se multiplica el número de observaciones de entrenamiento, lo que vuelve viable el entrenamiento y ajuste de redes neuronales.

Sobre esta base se desarrollaron dos modelos de clasificación con funciones de pérdida asimétricas, el modelo enfocado en costos y el modelo enfocado en costos modificado, junto a un modelo de referencia entrenado con la entropía cruzada binaria convencional. Los tres se evaluaron sobre el caso de estudio del subsistema de distribución de la Región Metropolitana, utilizando datos históricos de demanda, variables climatológicas y la estructura tarifaria vigente del mercado de potencia chileno en dicho subsistema.

En el caso base estudiado, en el que se penalizan las 52 horas de máxima demanda, ambos modelos enfocados en costos superaron al modelo de referencia en el rendimiento económico de sus decisiones. El modelo enfocado en costos modificado alcanzó el mejor desempeño, con un arrepentimiento promedio de 3304.03 [USD/MW] al año frente a los 4095.12 [USD/MW] del modelo de referencia, equivalente a un ahorro de 791.1 [USD/MW] al año, mientras que el modelo enfocado en costos obtuvo un ahorro de 362.0 [USD/MW] al año. Un test  $t$  de Welch sobre diez réplicas confirmó que ambas mejoras son estadísticamente significativas y no atribuibles a la variabilidad entre semillas, con valores  $p$  de 0.001 y 0.029 respectivamente,

lo que consolida al modelo enfocado en costos modificado como el más robusto, mientras que el modelo enfocado en costos, si bien significativo, alcanza una mejora de menor magnitud. Al comparar contra las políticas de referencia, ambos modelos redujeron de forma significativa el costo de operación respecto de no utilizar ningún mecanismo de predicción, con reducciones de hasta un 17.66 % para el modelo enfocado en costos y de hasta un 18.65 % para el modelo enfocado en costos modificado.

Los dos análisis de sensibilidad realizados permitieron caracterizar las condiciones bajo las cuales el enfoque propuesto conserva su ventaja. El análisis sobre la relación de costos mostró que la superioridad de los modelos enfocados en costos no es universal sino condicional ya que se manifiesta cuando existe una asimetría suficiente entre el costo de un falso positivo y el de un falso negativo, aproximadamente cuando la razón entre ambos costos se aleja de la unidad más allá de un factor cercano a dos, y se desvanece cuando ambos costos son comparables, régimen en el que un clasificador convencional bien ajustado resulta preferible.

El análisis sobre el número de horas penalizadas mostró, por su parte, que la ventaja del enfoque es robusta frente a variaciones en  $k$  e incluso se acentúa a medida que  $k$  crece. Tomados en conjunto, ambos análisis identifican las dos magnitudes a través de las cuales la estructura tarifaria de cualquier mercado incide sobre el problema, la asimetría de costos resumida por  $s$  y el número de horas penalizadas  $k$ , de modo que un mercado distinto del chileno queda caracterizado por su posición en el plano  $(s, k)$ . Esto otorga a la metodología un carácter transferible, ya que un practicante de otro mercado de cargo por capacidad puede determinar de antemano, a partir de su propio valor de  $s$ , si el enfoque asimétrico le conviene, sin necesidad de reevaluar el método desde cero.

Estos resultados admiten una lectura con implicancias prácticas concretas. Desde el punto de vista regulatorio, los parámetros con que se diseña el cargo por capacidad, el número de horas penalizadas y el monto del cargo por hora, determinan la posición del mercado en el plano  $(s, k)$  y, con ello, el valor que un consumidor puede capturar mediante una predicción consciente de los costos. Un esquema que concentre el cargo en menos horas eleva tanto la asimetría como el costo de no anticipar una hora punta, lo que fortalece el incentivo a la gestión activa de la demanda. De este modo, al fijar estos parámetros, el regulador define de manera implícita cuán valiosa resulta la anticipación para el consumidor.

En el plano operacional, los modelos se traducen directamente en una política de operación del medio de generación propia, ya que el modelo enfocado en costos modificado adopta una postura más agresiva, emitiendo más avisos y privilegiando no omitir horas punta costosas, mientras que el modelo enfocado en costos opera de forma más parsimoniosa en el uso del generador. La elección entre ambos no es puramente técnica, sino que depende del perfil de riesgo del consumidor y de los costos asociados al uso del generador, y el factor de asimetría ofrece una perilla explícita para ubicar ese punto de operación.

En cuanto a su adopción práctica, la metodología requiere únicamente información habitualmente disponible para un consumidor, esto es, el historial de demanda del subsistema al cual está conectado y los parámetros tarifarios. Además, estos modelos pueden reentrenarse con cada ciclo de facturación y entrega una recomendación operativa simple. Dado que el ahorro se expresa por unidad de potencia, su magnitud absoluta escala con la capacidad involucrada, por lo que el beneficio resulta especialmente relevante para consumidores industriales y comerciales de gran tamaño que dispongan de un medio de generación despachable.

A partir de lo anterior se valida la hipótesis planteada en este trabajo: es posible desarrollar un modelo de predicción de horas punta de demanda eléctrica que reduzca los costos asociados a las decisiones de un usuario sujeto a cargos por demanda punta, superando a los modelos basados en metodologías convencionales. El resultado central es que esta mejora se obtiene precisamente porque las predicciones no se orientan a minimizar el error estadístico promedio, sino el costo económico de las decisiones que de ellas se derivan. En consecuencia, este trabajo muestra que existen predicciones que, aun apartándose de la media estadística, conducen a decisiones más eficientes, y que alinear el objetivo de entrenamiento con la estructura económica del problema constituye una vía concreta para que la predicción de demanda genere valor real para el consumidor eléctrico.

## 6.2. Trabajo futuro

Como trabajo futuro y mejora a la metodología desarrollada, se propone eliminar el supuesto que considera constante el consumo del cliente, incorporándolo como una variable del problema. Esto permitiría desarrollar un modelo que integre la respuesta de la demanda del cliente, ampliando así el abanico de decisiones disponibles y reflejando de forma más realista su comportamiento ante los cargos por demanda punta.

Por otro lado, se sugiere abordar el segundo supuesto considerado en este trabajo, el cual limita la aplicación del modelo únicamente a aquellos clientes cuyo consumo es lo suficientemente bajo como para no tener la capacidad de desplazar una hora punta en su subsistema a partir de los avisos de predicción. Con esto en mente, deja como puesto analizar cuánta potencia es efectivamente necesaria para lograr dicho desplazamiento, según las particularidades del subsistema en el que se implemente. Esto permitiría definir un umbral de potencia que distinga entre aquellos clientes que pueden beneficiarse del modelo y aquellos que no.

Se sugiere incorporar otras variables de entrada que reflejen aspectos adicionales que puedan influir en la demanda total del subsistema, como eventos deportivos, festividades masivas o comportamientos simultáneos de los consumidores. Esto permitiría al modelo considerar un espectro más amplio de observaciones que faciliten la identificación de nuevos patrones de predicción, lo que podría mejorar su capacidad para detectar horas punta y optimizar las decisiones de los clientes sometidos a cargos por demanda punta.

Además, si bien los modelos proporcionan una holística de los resultados posibles, sería interesante abordar el problema desde una perspectiva de optimización, determinando el número óptimo de avisos que podrían utilizarse.

Una línea natural de investigación futura consiste en evaluar empíricamente el efecto de la función de pérdida propuesta sobre arquitecturas alternativas, tales como redes recurrentes, convolucionales o modelos más livianos como la regresión logística.

Por último, se propone explorar otros medios de generación propia que el cliente podría utilizar para desplazar su consumo o abastecerse de manera alternativa, tales como baterías u otros sistemas de almacenamiento. Esto con el fin de ampliar las posibilidades de respuesta del cliente ante horas pronosticadas como punta, lo que podría mejorar aún más su eficiencia económica frente a los cargos por demanda punta.

# Bibliografía

- [1] Carmona, R., Yang, X., y Zeng, C., “Coincident peak prediction for capacity and transmission charge reduction”, 2024, <https://arxiv.org/abs/2407.04081>. Preprint arXiv:2407.04081. Accedido: 13-oct-2025.
- [2] Liu, Z., Wierman, A., Chen, Y., Razon, B., y Chen, N., “Data center demand response: Avoiding the coincident peak via workload shifting and local generation”, *Performance Evaluation*, vol. 70, no. 10, pp. 770–791, 2013, [doi:10.1016/j.peva.2013.08.014](https://doi.org/10.1016/j.peva.2013.08.014).
- [3] Gómez-Vargas, N., Maldonado, S., y Vairetti, C., “A predict-and-optimize approach to profit-driven churn prevention”, *Eur. J. Oper. Res.*, vol. 324, no. 2, pp. 555–566, 2025, [doi:10.1016/j.ejor.2025.02.008](https://doi.org/10.1016/j.ejor.2025.02.008).
- [4] Afshin, M. y Sadeghian, A., “Pca-based least squares support vector machines in week-ahead load forecasting”, en *Proceedings of the IEEE/IAS Industrial and Commercial Power Systems Technical Conference (ICPS)*, pp. 1–6, 2007, [doi:10.1109/ICPS.2007.4292100](https://doi.org/10.1109/ICPS.2007.4292100).
- [5] Setiawan, A., Koprinska, I., y Agelidis, V., “Very short-term electricity load demand forecasting using support vector regression”, en *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, vol. 1, pp. 2888–2894, 2009, [doi:10.1109/IJCNN.2009.5179063](https://doi.org/10.1109/IJCNN.2009.5179063).
- [6] Feinberg, E. A. y Genethliou, D., *Load Forecasting*, pp. 269–285. Springer, 2005, [doi:10.1007/0-387-23471-3\\_12](https://doi.org/10.1007/0-387-23471-3_12).
- [7] Ávila, F., “Diseño de un sistema de gestión de demanda basado en lógica difusa para micro-redes”, Tesis M.Sc., Universidad de Chile, Santiago, Chile, 2013.
- [8] Kateule, R., Sakala, L. A., y Tunga, M., “Short-term energy load forecasting model with sample filtering”, *Tanzania Journal of Science*, vol. 49, no. 1, pp. 130–139, 2023, [doi:10.4314/tjs.v49i1.12](https://doi.org/10.4314/tjs.v49i1.12).
- [9] Moghram, I. y Rahman, S., “Analysis and evaluation of five short-term load forecasting techniques”, *IEEE Power Engineering Review*, vol. 9, no. 11, pp. 42–43, 1989, [doi:10.1109/MPER.1989.4310383](https://doi.org/10.1109/MPER.1989.4310383).
- [10] Tan, Y., Li, Y., Wang, Z., Li, Y., y Zhang, X., “Short-term load forecasting using channel and temporal attention mechanism based temporal convolutional network”, *Electric Power Systems Research*, vol. 205, p. 107761, 2022, [doi:10.1016/j.epsr.2021.107761](https://doi.org/10.1016/j.epsr.2021.107761).
- [11] Dai, S., Meng, F., Dai, H., Wang, Q., y Chen, X., “Electrical peak demand forecasting: A review”, 2021, <https://arxiv.org/abs/2108.01393>. Preprint arXiv:2108.01393. Accedido: 13-oct-2025.

- [12] Pierre, A. A., Akim, S. A., Semenyó, A. K., y Babiga, B., “Peak electrical energy consumption prediction by arima, lstm, gru, arima-lstm and arima-gru approaches”, *Energies*, vol. 16, no. 12, p. 4739, 2023, [doi:10.3390/en16124739](https://doi.org/10.3390/en16124739).
- [13] Yu, Z., Niu, Z., Tang, W., y Wu, Q., “Deep learning for daily peak load forecasting: A novel gated recurrent neural network combining dynamic time warping”, *IEEE Access*, vol. 7, pp. 17184–17194, 2019, [doi:10.1109/ACCESS.2019.2895604](https://doi.org/10.1109/ACCESS.2019.2895604).
- [14] Dhillon, J. y Rahman, S. A., “Peak electricity load forecasting using online support vector regression”, en *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–4, 2016, [doi:10.1109/CCECE.2016.7726784](https://doi.org/10.1109/CCECE.2016.7726784).
- [15] Bovornkeeratiroj, P. *et al.*, “Predicting peak day and peak hour of electricity demand with ensemble machine learning”, *Frontiers in Energy Research*, vol. 10, p. 944804, 2022, [doi:10.3389/fenrg.2022.944804](https://doi.org/10.3389/fenrg.2022.944804).
- [16] Ibrahim, B., Rabelo, L., Sarmiento, A. T., y Gutierrez-Franco, E., “A holistic approach to power systems using innovative machine learning and system dynamics”, *Energies*, vol. 16, no. 13, p. 5225, 2023, [doi:10.3390/en16135225](https://doi.org/10.3390/en16135225).
- [17] Hua, Q., Fan, Z., Mu, W., Cui, J., Xing, R., Liu, H., y Gao, J., “A short-term power load forecasting method using cnn-gru with an attention mechanism”, *Energies*, vol. 18, no. 1, p. 106, 2025, [doi:10.3390/en18010106](https://doi.org/10.3390/en18010106).
- [18] Mahmud, K., Ravishankar, J., Hossain, M. J., y Dong, Z. Y., “The impact of prediction errors in the domestic peak power demand management”, *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4567–4579, 2020, [doi:10.1109/TII.2019.2946292](https://doi.org/10.1109/TII.2019.2946292).
- [19] Hyndman, R. J. y Fan, S., “Density forecasting for long-term peak electricity demand”, *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, 2010, [doi:10.1109/TPWRS.2009.2036017](https://doi.org/10.1109/TPWRS.2009.2036017).
- [20] Saxena, H., Aponte, O., y McConky, K. T., “A hybrid machine learning model for forecasting a billing period’s peak electric load days”, *International Journal of Forecasting*, vol. 35, no. 4, pp. 1288–1303, 2019, [doi:10.1016/j.ijforecast.2019.03.025](https://doi.org/10.1016/j.ijforecast.2019.03.025).
- [21] Dowling, C. P., Kirschen, D. S., y Zhang, B., “Coincident peak prediction using a feed-forward neural network”, en *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 912–916, 2018, [doi:10.1109/GlobalSIP.2018.8646654](https://doi.org/10.1109/GlobalSIP.2018.8646654).
- [22] Amjady, N., “Short-term hourly load forecasting using time-series modeling with peak load estimation capability”, *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 498–505, 2001, [doi:10.1109/59.932287](https://doi.org/10.1109/59.932287).
- [23] Liu, J. y Brown, L. E., “Prediction of hour of coincident daily peak load”, en *Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, (Washington, DC), 2019, [doi:10.1109/ISGT.2019.8791587](https://doi.org/10.1109/ISGT.2019.8791587).
- [24] Aziz, A., Mahmood, D., Qureshi, M. S., Qureshi, M. B., y Kim, K., “Ai-based peak power demand forecasting model focusing on economic and climate features”, *Frontiers in Energy Research*, vol. 12, p. 1328891, 2024, [doi:10.3389/fenrg.2024.1328891](https://doi.org/10.3389/fenrg.2024.1328891).
- [25] Lin, J.-L., Zhang, Y., Zhu, K., Chen, B., y Zhang, F., “Asymmetric loss functions for contract capacity optimization”, *Energies*, vol. 13, no. 12, p. 3123, 2020, [doi:10.3390/en13123123](https://doi.org/10.3390/en13123123).

- [26] Hashemi, S. R., Salehi, S. S. M., Erdogmus, D., Prabhu, S. P., Warfield, S. K., y Gholipour, A., “Asymmetric loss functions and deep densely connected networks for highly imbalanced medical image segmentation: Application to multiple sclerosis lesion detection”, *IEEE Access*, vol. 7, pp. 138–149, 2019, [doi:10.1109/ACCESS.2018.2886371](https://doi.org/10.1109/ACCESS.2018.2886371).
- [27] Dessain, J., “Improving the prediction of asset returns with machine learning by using a custom loss function”, *SSRN Electronic Journal*, 2021, [doi:10.2139/ssrn.3973086](https://doi.org/10.2139/ssrn.3973086).
- [28] Tan, S. y Frazier, P. I., “Asymptotically optimal regret for black-box predict-then-optimize”, 2024, <https://arxiv.org/abs/2406.07866>. Preprint arXiv:2406.07866. Accedido: 13-oct-2025.
- [29] Hippert, H., Pedreira, C., y Souza, R., “Neural networks for short-term load forecasting: a review and evaluation”, *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001, [doi:10.1109/59.910780](https://doi.org/10.1109/59.910780).
- [30] Hong, T. y Fan, S., “Probabilistic electric load forecasting: A tutorial review”, *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016, [doi:https://doi.org/10.1016/j.ijforecast.2015.11.011](https://doi.org/10.1016/j.ijforecast.2015.11.011).
- [31] *Modeling and Forecasting Electricity Loads*, cap. 3, pp. 67–100. John Wiley & Sons, Ltd, 2006, [doi:https://doi.org/10.1002/9781118673362.ch3](https://doi.org/10.1002/9781118673362.ch3).
- [32] Hahn, H., Meyer-Nieberg, S., y Pickl, S., “Electric load forecasting methods: Tools for decision making”, *European Journal of Operational Research*, vol. 199, no. 3, pp. 902–907, 2009, [doi:https://doi.org/10.1016/j.ejor.2009.01.062](https://doi.org/10.1016/j.ejor.2009.01.062).
- [33] Poslavskaya, E. y Korolev, A., “Encoding categorical data: Is there yet anything ‘hotter’ than one-hot encoding?”, 2023, <https://arxiv.org/abs/2312.16930>.
- [34] Goodfellow, I., Bengio, Y., y Courville, A., *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>. [En línea]. Disponible en: <http://www.deeplearningbook.org>.
- [35] Verhaegen, M. y Verdult, V., *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, 2007.
- [36] Khanna, R., Lane, C., Mattei, N., y Rushing, J. W., “Hankel matrix factorization for time series forecasting”, arXiv preprint arXiv:2006.03134, 2020, <https://arxiv.org/abs/2006.03134>.
- [37] Haykin, S., *Neural Networks and Learning Machines*. Pearson, 3rd ed., 2009.
- [38] Bishop, C. M., *Pattern Recognition and Machine Learning*. Springer, 2006.
- [39] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., y Salakhutdinov, R., “Dropout: A simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [40] He, H. y Garcia, E. A., “Learning from imbalanced data”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [41] Keras Team, “Class weight utility: compute\_class\_weight”, 2025, [https://www.tensorflow.org/api\\_docs/python/tf/keras/utils/class\\_weight](https://www.tensorflow.org/api_docs/python/tf/keras/utils/class_weight). Accedido: 12-oct-2025.
- [42] Moon, J., Park, S., Rho, S., y Hwang, E., “Interpretable short-term electrical load forecasting scheme using cubist”, *Computational Intelligence and Neuroscience*, vol. 2022, p. 6892995, 2022, [doi:10.1155/2022/6892995](https://doi.org/10.1155/2022/6892995).

- [43] Kingma, D. P. y Ba, J., “Adam: A method for stochastic optimization”, International Conference on Learning Representations (ICLR), 2015, <https://arxiv.org/abs/1412.6980>.
- [44] Prechelt, L., “Early stopping—but when?”, en *Neural Networks: Tricks of the Trade*, pp. 55–69, Springer, 1998.
- [45] Hastie, T., Tibshirani, R., y Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd ed., 2009.
- [46] Welch, B. L., “The generalization of ‘Student’s’ problem when several different population variances are involved”, *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947, doi: [10.1093/biomet/34.1-2.28](https://doi.org/10.1093/biomet/34.1-2.28).
- [47] Coordinador Eléctrico Nacional (CEN), “Informe retiros de potencia para balance de potencia de suficiencia 2023”. Gerencia de Mercado, Departamento de Transferencias de Potencia y Cargos de Transmisión, jul 2024, <https://www.coordinador.cl/mercados/documentos/potencia-de-suficiencia/balances-anuales-de-potencia-de-suficiencia/2023-balances-anuales-de-potencia-de-suficiencia/>. Accedido: 08-dic-2024.
- [48] Ministerio de Economía, Fomento y Reconstrucción, Chile, “Decreto n.º 62, aprueba reglamento de transferencias de potencia entre empresas generadoras establecidas en la ley general de servicios eléctricos”. Diario Oficial, n.º 38490, jun 2006, <https://www.leychile.cl/Navegar?idNorma=250604>. Accedido: 08-dic-2024.
- [49] Oszcewski, R. J. y Bluestein, M., “The new wind chill equivalent temperature chart”, *Bulletin of the American Meteorological Society*, vol. 86, no. 10, pp. 1453–1458, 2005, doi: [10.1175/BAMS-86-10-1453](https://doi.org/10.1175/BAMS-86-10-1453).
- [50] Comisión Nacional de Energía, “Fijación de precios de nudo promedio del sistema eléctrico nacional: Informe técnico preliminar”, 2024, <https://www.cne.cl/tarificacion/electrica/precio-nudo-promedio/fijaciones-2023/>. Accedido: 08-dic-2024.
- [51] Coordinador Eléctrico Nacional (CEN), “Antecedentes de cálculo para las transferencias económicas”. Gerencia de Mercado, Departamento de Transferencias de Potencia y Cargos de Transmisión, 2023, <https://www.coordinador.cl/mercados/documentos/transferencias-economicas/antecedentes-de-calculo-para-las-transferencias-economicas/>. Accedido: 12-oct-2025.
- [52] Wärtsilä Corporation, “Wärtsilä generating sets”, 2025, <https://www.wartsila.com/marine/products/engines-and-generating-sets/generating-sets/wartsila-gensets>. Accedido: 12-oct-2025.
- [53] Empresa Nacional del Petróleo (ENAP), “Tabla de precios de paridad”, 2025, <https://www.enap.cl/informacion-comercial/tabla-de-precios-de-paridad>. Accedido: 12-oct-2025.

# Anexos

## Anexo A. Cálculo de arrepentimientos

### A.1. Consideraciones iniciales

Al asumir que el cliente posee un consumo constante  $p_t$ , se tendrá que al considerar la predicción del modelo  $\alpha_t \in \{Peak (1), No Peak (0)\}$  en la hora  $t \in \tau$  y la observación real  $\beta_t \in \{Peak (1), No Peak (0)\}$  en la misma hora, se tendrán las siguientes definiciones:

$$p_t[d(\beta_t) = 1] = p_t[d(\alpha_t) = 1] = p_t = cte \quad (A.1)$$

$$p_t[d(\beta_t) = 0] = p_t[d(\alpha_t) = 0] = p_t = cte \quad (A.2)$$

Las cuales indican que el consumo de potencia  $p_t$  del cliente es el mismo independientemente de la predicción de  $\alpha_t$  y la observación de  $\beta_t$ , debido a que se considera que el cliente buscará satisfacer su consumo ya sea por el medio de generación distribuida o por la red.

Luego, se definen las potencias de generación propia  $p_t^{gp}$  y de consumo desde la red  $p_t^{cs}$  como:

$$p_t^{cs}[d(\beta_t = 1)] = p_t^{cs}[d(\alpha_t) = 1] = 0 \quad (A.3)$$

$$p_t^{cs}[d(\beta_t) = 0] = p_t^{cs}[d(\alpha_t) = 0] = p_t \quad (A.4)$$

$$p_t^{gp}[d(\beta_t) = 1] = p_t^{gp}[d(\alpha_t) = 1] = p_t \quad (A.5)$$

$$p_t^{gp}[d(\beta_t) = 0] = p_t^{gp}[d(\alpha_t) = 0] = 0 \quad (A.6)$$

Las cuales indican que el cliente consumirá toda su potencia  $p_t$  desde la red en caso de que la hora  $t$  no sea o no se prediga como punta ( $\beta_t = 0$  o  $\alpha_t = 0$ ), y que el cliente generará

toda su potencia  $p_t$  por medio de generación propia en caso de que la hora  $t$  sea o se prediga como punta ( $\beta_t = 1$  o  $\alpha_t = 1$ ).

## A.2. Cálculo de utilidades

Para las posibles combinaciones de las predicciones  $\alpha_t \in \{Peak (1), No Peak (0)\}$  y las observaciones  $\beta_t \in \{Peak (1), No Peak (0)\}$ , las utilidades en una hora  $t \in \tau$  específica considerando para ello las ecuaciones 3.12 y 3.13 descritas en el la Sección 3.4.3 serán:

- $\alpha = 0 \wedge \beta = 0$ :

$$\begin{aligned}
 Utilidades_t &= FO(\beta_t = 0, d(\alpha_t = 0)) \\
 &= g(p_t[d(\alpha_t = 0)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 0)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=0]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 0)] \\
 &= g(p_t) - \pi^{gp} \cdot 0 - (\pi^{cs} + 0 \cdot \pi^{cp}) \cdot p_t \\
 &= g(p_t) - \pi^{cs} \cdot p_t
 \end{aligned} \tag{A.7}$$

- $\alpha = 1 \wedge \beta = 1$ :

$$\begin{aligned}
 Utilidades_t &= FO(\beta_t = 1, d(\alpha_t = 1)) \\
 &= g(p_t[d(\alpha_t = 1)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 1)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=1]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 1)] \\
 &= g(p_t) - \pi^{gp} \cdot p_t - (\pi^{cs} + 1 \cdot \pi^{cp}) \cdot 0 \\
 &= g(p_t) - \pi^{gp} \cdot p_t
 \end{aligned} \tag{A.8}$$

- $\alpha = 1 \wedge \beta = 0$ :

$$\begin{aligned}
 Utilidades_t &= FO(\beta_t = 0, d(\alpha_t = 1)) \\
 &= g(p_t[d(\alpha_t = 1)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 1)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=0]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 1)] \\
 &= g(p_t) - \pi^{gp} \cdot p_t - (\pi^{cs} + 0 \cdot \pi^{cp}) \cdot 0 \\
 &= g(p_t) - \pi^{gp} \cdot p_t
 \end{aligned} \tag{A.9}$$

- $\alpha = 0 \wedge \beta = 1$ :

$$\begin{aligned}
 Utilidades_t &= FO(\beta_t = 1, d(\alpha_t = 0)) \\
 &= g(p_t[d(\alpha_t = 0)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 0)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=1]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 0)] \\
 &= g(p_t) - \pi^{gp} \cdot 0 - (\pi^{cs} + 1 \cdot \pi^{cp}) \cdot p_t \\
 &= g(p_t) - (\pi^{cs} + \pi^{cp}) \cdot p_t
 \end{aligned} \tag{A.10}$$

## A.3. Cálculo de arrepentimientos

Para las posibles combinaciones de las predicciones  $\alpha_t \in \{Peak (1), No Peak (0)\}$  y las observaciones  $\beta_t \in \{Peak (1), No Peak (0)\}$ , las utilidades en una hora  $t \in \tau$  específica considerando para ello la ecuación 3.15 descritas en el la Sección 3.4.4 serán:

- $\alpha = 0 \wedge \beta = 0$ :

$$\begin{aligned}
\text{regret}_t &= FO(\beta_t = 0, d^*(\beta_t = 0)) - FO(\beta_t = 0, d(\alpha_t = 0)) \\
&= g(p_t[d^*(\beta_t = 0)]) - \pi^{gp} \cdot p_t^{gp}[d^*(\beta_t = 0)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=0]} \cdot \pi^{cp}) \cdot p_t^{cs}[d^*(\beta_t = 0)] \\
&\quad - \left( g(p_t[d(\alpha_t = 0)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 0)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=0]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 0)] \right) \\
&= g(p_t) - \pi^{gp} \cdot 0 - (\pi^{cs} + 0 \cdot \pi^{cp}) \cdot p_t \\
&\quad - \left( g(p_t) - \pi^{gp} \cdot 0 - (\pi^{cs} + 0 \cdot \pi^{cp}) \cdot p_t \right) \\
&= g(p_t) - \pi^{cs} \cdot p_t - (g(p_t) - \pi^{cs} \cdot p_t) \\
&= 0
\end{aligned} \tag{A.11}$$

- $\alpha = 1 \wedge \beta = 1$ :

$$\begin{aligned}
\text{regret}_t &= FO(\beta_t = 1, d^*(\beta_t = 1)) - FO(\beta_t = 1, d(\alpha_t = 1)) \\
&= g(p_t[d^*(\beta_t = 1)]) - \pi^{gp} \cdot p_t^{gp}[d^*(\beta_t = 1)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=1]} \cdot \pi^{cp}) \cdot p_t^{cs}[d^*(\beta_t = 1)] \\
&\quad - \left( g(p_t[d(\alpha_t = 1)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 1)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=1]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 1)] \right) \\
&= g(p_t) - \pi^{gp} \cdot p_t - (\pi^{cs} + 1 \cdot \pi^{cp}) \cdot 0 \\
&\quad - \left( g(p_t) - \pi^{gp} \cdot p_t - (\pi^{cs} + 1 \cdot \pi^{cp}) \cdot 0 \right) \\
&= g(p_t) - \pi^{gp} \cdot p_t - (g(p_t) - \pi^{gp} \cdot p_t) \\
&= 0
\end{aligned} \tag{A.12}$$

- $\alpha = 1 \wedge \beta = 0$ :

$$\begin{aligned}
\text{regret}_t &= FO(\beta_t = 0, d^*(\beta_t = 0)) - FO(\beta_t = 0, d(\alpha_t = 1)) \\
&= g(p_t[d^*(\beta_t = 0)]) - \pi^{gp} \cdot p_t^{gp}[d^*(\beta_t = 0)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=0]} \cdot \pi^{cp}) \cdot p_t^{cs}[d^*(\beta_t = 0)] \\
&\quad - \left( g(p_t[d(\alpha_t = 1)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 1)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=0]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 1)] \right) \\
&= g(p_t) - \pi^{gp} \cdot 0 - (\pi^{cs} + 0 \cdot \pi^{cp}) \cdot p_t \\
&\quad - \left( g(p_t) - \pi^{gp} \cdot p_t - (\pi^{cs} + 0 \cdot \pi^{cp}) \cdot 0 \right) \\
&= g(p_t) - \pi^{cs} \cdot p_t - (g(p_t) - \pi^{gp} \cdot p_t) \\
&= \pi^{gp} \cdot p_t - \pi^{cs} \cdot p_t \\
&= (\pi^{gp} - \pi^{cs}) \cdot p_t
\end{aligned} \tag{A.13}$$

- $\alpha = 0 \wedge \beta = 1$ :

$$\begin{aligned}
regret_t &= FO(\beta_t = 1, d^*(\beta_t = 1)) - FO(\beta_t = 1, d(\alpha_t = 0)) \\
&= g(p_t[d^*(\beta_t = 1)]) - \pi^{gp} \cdot p_t^{gp}[d^*(\beta_t = 1)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=1]} \cdot \pi^{cp}) \cdot p_t^{cs}[d^*(\beta_t = 1)] \\
&\quad - \left( g(p_t[d(\alpha_t = 0)]) - \pi^{gp} \cdot p_t^{gp}[d(\alpha_t = 0)] - (\pi^{cs} + \mathbf{1}_{[\beta_t=1]} \cdot \pi^{cp}) \cdot p_t^{cs}[d(\alpha_t = 0)] \right) \\
&= g(p_t) - \pi^{gp} \cdot p_t - (\pi^{cs} + 1 \cdot \pi^{cp}) \cdot 0 \\
&\quad - \left( g(p_t) - \pi^{gp} \cdot 0 - (\pi^{cs} + 1 \cdot \pi^{cp}) \cdot p_t \right) \\
&= g(p_t) - \pi^{gp} \cdot p_t - (g(p_t) - (\pi^{cs} + \pi^{cp}) \cdot p_t) \\
&= -\pi^{gp} \cdot p_t + (\pi^{cs} + \pi^{cp}) \cdot p_t \\
&= (\pi^{cs} + \pi^{cp} - \pi^{gp}) \cdot p_t
\end{aligned} \tag{A.14}$$

Finalmente, considerando un consumo de potencia unitario  $p_t = 1 \text{ MW}$ , las expresiones de arrepentimiento para cada combinación de predicción y observación quedan como se muestra en la Tabla 3.2.

## Anexo B. Cálculo de costos variables de generación propia

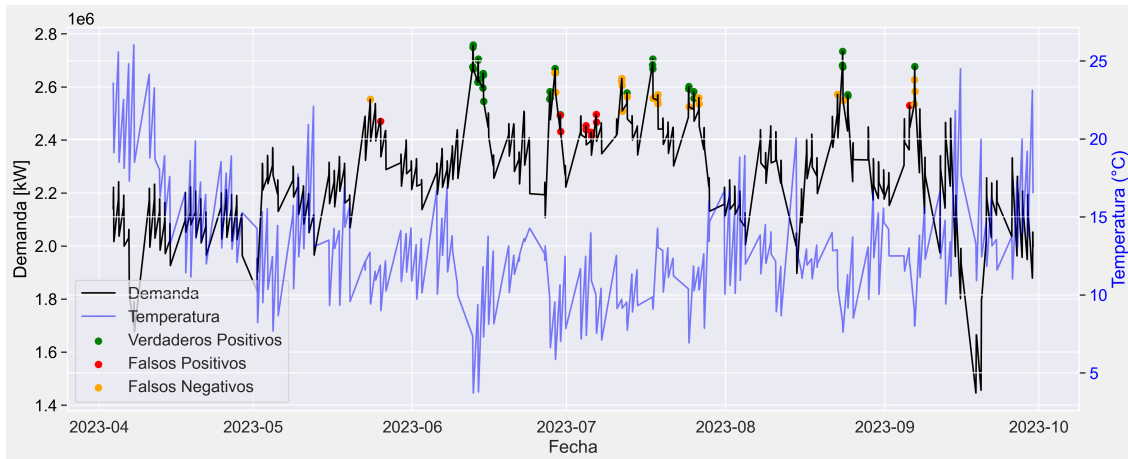
Como referencia para el medio de generación propia se tomó a los modelos del grupo electrógeno diésel Wärtsilä 32, los cuales reportan un consumo específico de combustible (SFC, por sus siglas en inglés) de aproximadamente  $174 [g/kWh]$  (equivalente a  $0.205 [L/kWh]$  asumiendo una densidad del diésel de  $0.84 [kg/L]$ ) [52]. Además, se toma como referencia el precio del diésel según los valores reportados por ENAP en su Tabla de Precios de Paridad [53] el cual fluctúa entre 630 y 600  $[USD/m^3]$ . Finalmente, a partir de estos dos valores, el costo variable de generación propia ( $\pi^{gp}$ ) se puede calcular como sigue:

$$\begin{aligned}\pi^{gp} &= SFC \times P \quad [\text{USD/MWh}] \\ &= 0.205 \frac{\text{L}}{\text{kWh}} \times 615 \frac{\text{USD}}{\text{m}^3} \times 1000 \frac{\text{kWh}}{\text{MWh}} \times 0.001 \frac{\text{m}^3}{\text{L}} \\ &= 126.4 \text{ USD/MWh}\end{aligned} \tag{B.1}$$

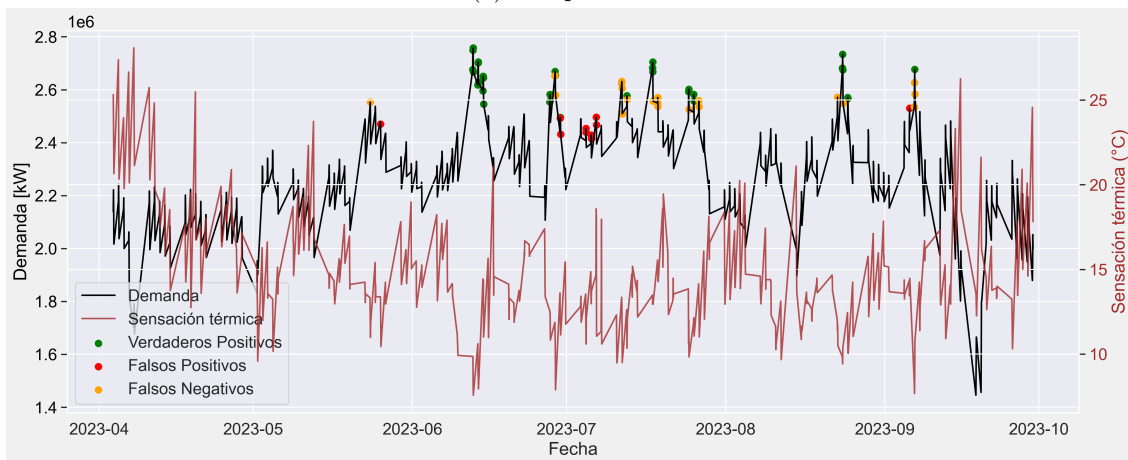
# Anexo C. Resultados

## C.1. Gráficos de avisos y variables climáticas

En las Figuras C.1 y C.2 se presentan las curvas de demanda junto con las variables climáticas para el modelo enfocado en predicción y enfocado en costos, respectivamente. En cada una de estas figuras se muestran de forma separada las variables climáticas analizadas: temperatura (subfiguras (a)), sensación térmica (subfiguras (b)), humedad (subfiguras (c)) y presión (subfiguras (d)).

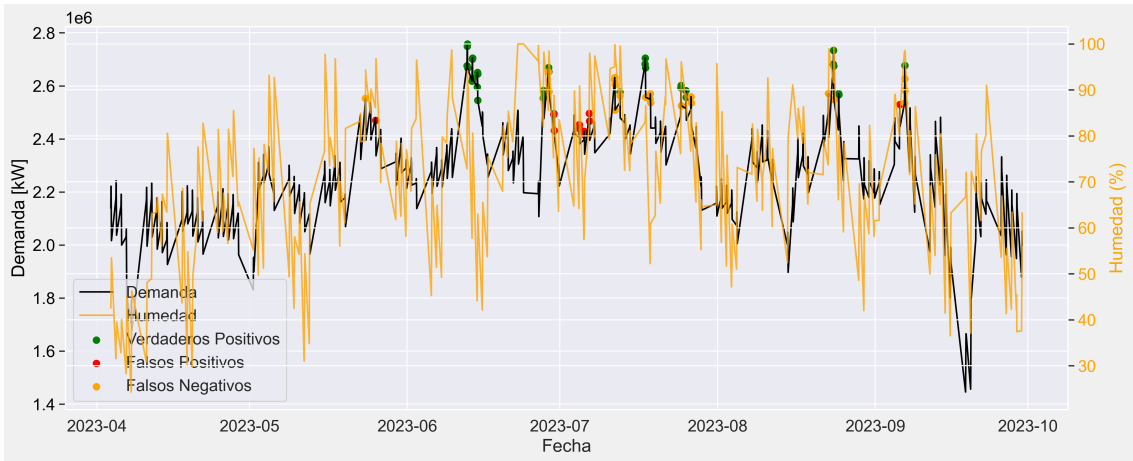


(a) Temperatura.

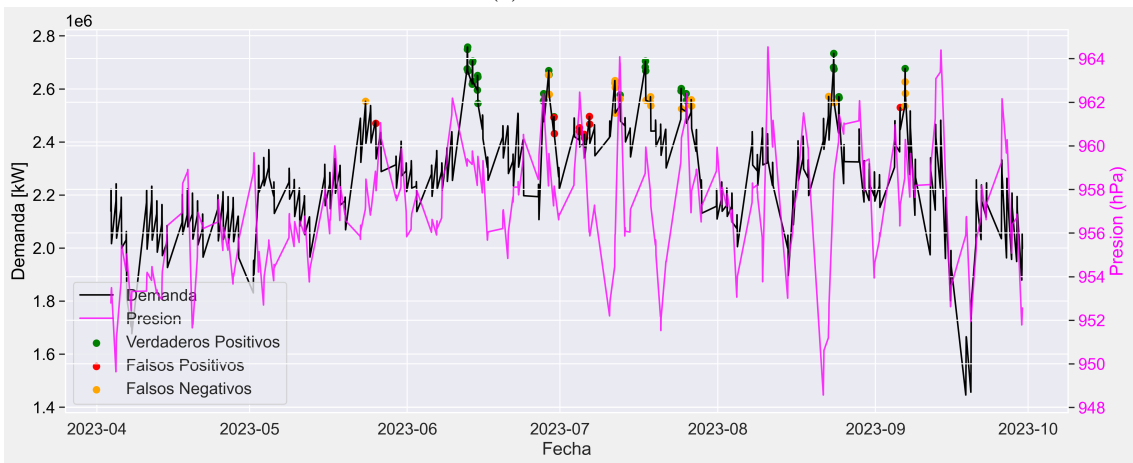


(b) Sensación térmica.

Figura C.1: Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en predicción.

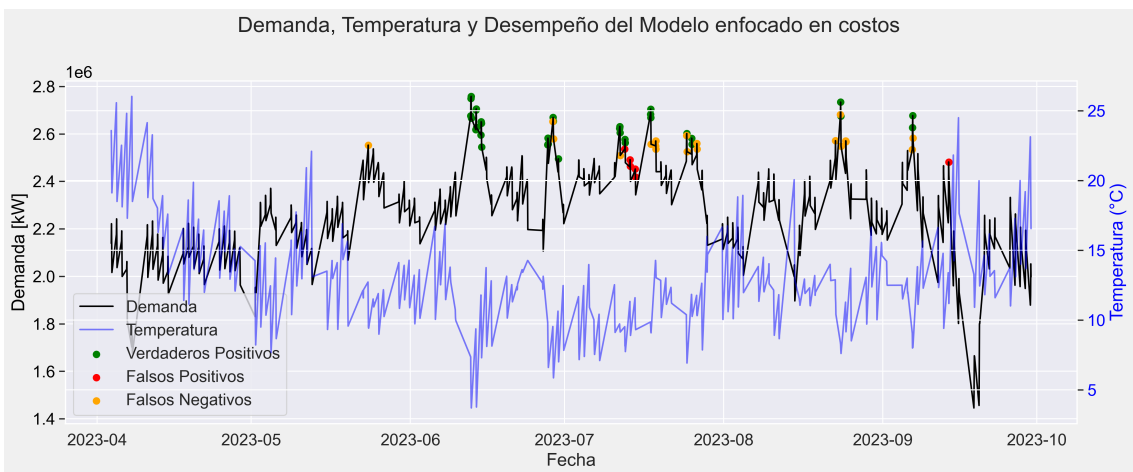


(c) Humedad.



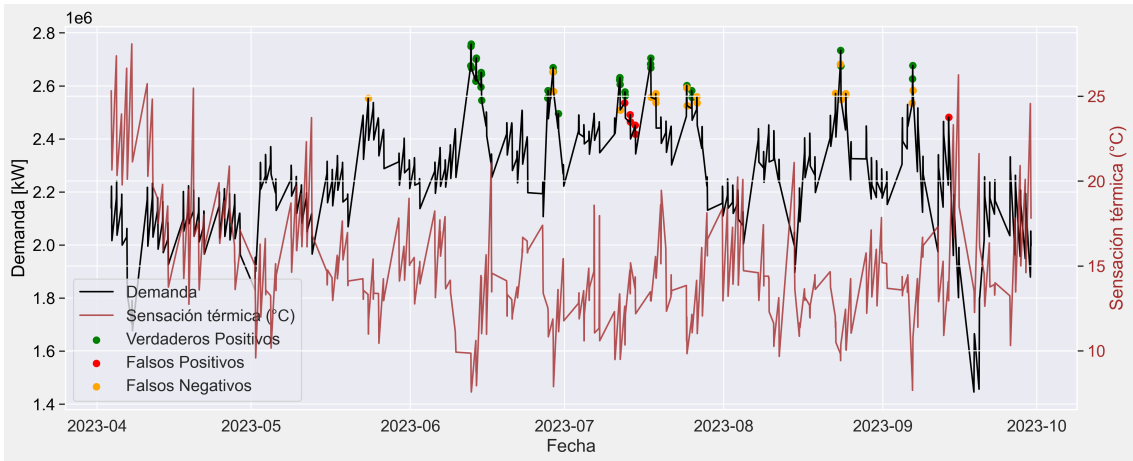
(d) Presión.

Figura C.1: Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en predicción (Continuación).

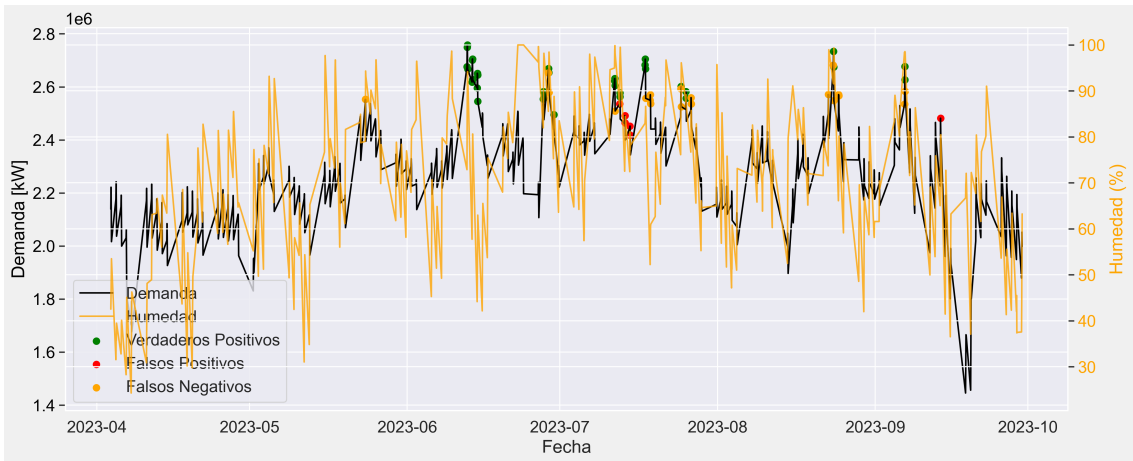


(a) Temperatura.

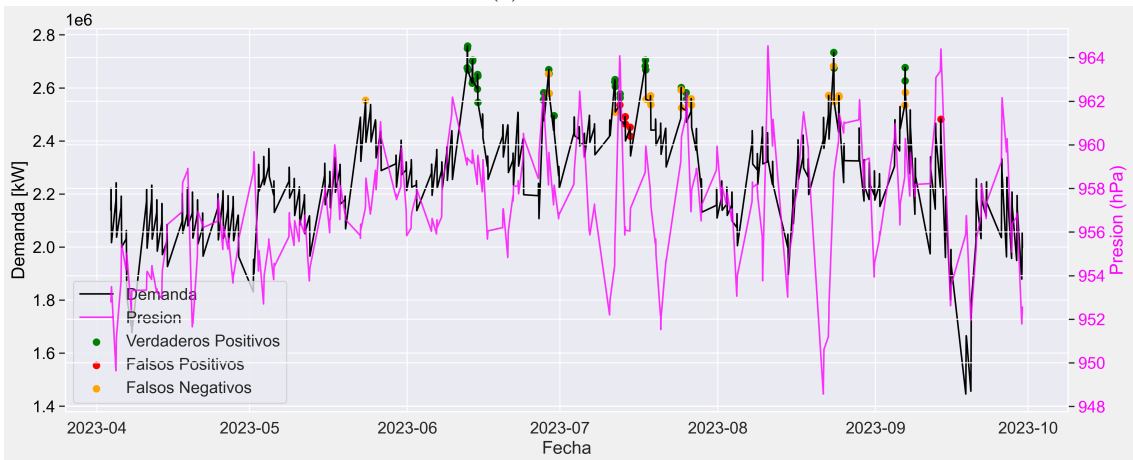
Figura C.2: Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos.



(b) Sensación térmica.



(c) Humedad.



(d) Presión.

Figura C.2: Curva de variables climáticas y de demanda, junto con los avisos del modelo enfocado en costos (Continuación).