

# Data Mining and Short-Term Power Demand Forecasting in the Ecuadorian Electric System

## Minería de Datos y Proyección a Corto Plazo de la Demanda de Potencia en el Sistema Eléctrico Ecuatoriano

A.O. Gallo<sup>1</sup>F.E. Pérez<sup>1</sup>D.A. Salinas<sup>2</sup><sup>1</sup>Escuela Politécnica Nacional, Ecuador

E-mail: angel.gallo@epn.edu.ec; fabian.perez@epn.edu.ec

<sup>2</sup>Agencia de Regulación y Control de Recursos Naturales No Renovables, Ecuador

E-mail: diego.salinas@controlrecursosyenergia.gob.ec

### Abstract

This article presents a computational tool developed in the Python programming language for data mining and short-term load forecasting of the National Interconnected System (SNI), using the predictive approach of the Light Gradient Boosting Machine (LGBM), Random Forest, Decision Tree machine learning algorithms, and the statistical algorithm for time series ARIMA. The implementation of the Hyperopt and Grid Search functions to define the main hyperparameters of the machine learning algorithms and the parameters  $p$ ,  $d$ ,  $q$  of the ARIMA algorithm respectively, together with the application of feature engineering allow the algorithms to be properly adjusted and modeled for the data series. The procedure for data mining and load forecasting shows the reliability and versatility of using the computational tool, obtaining relevant results, such as the reduction of anomalies in the data series to improve the precision in the projected electrical demand curves.

**Index terms**— Machine learning, Data mining, Electrical Power, Short-term load forecasting, National Interconnected System.

### Resumen

En este artículo se presenta una herramienta computacional elaborada en lenguaje de programación Python para minería de datos y proyección a corto plazo de la demanda de potencia eléctrica del Sistema Nacional Interconectado (SNI), utilizando el enfoque predictivo de los algoritmos de aprendizaje automático Light Gradient Boosting Machine (LGBM), Random Forest, Decision Tree, y el algoritmo estadístico para series de tiempo ARIMA. La implementación de las funciones Hyperopt y Búsqueda de Cuadrícula para definir los principales hiperparámetros de los algoritmos de aprendizaje automático y los parámetros  $p$ ,  $d$ ,  $q$  del algoritmo ARIMA respectivamente, junto con la aplicación de la ingeniería de características permiten ajustar y modelar adecuadamente los algoritmos para las series de datos. El procedimiento para minería de datos y proyección de demanda muestra la confiabilidad y versatilidad de utilizar la herramienta computacional obteniendo resultados relevantes, como la disminución de anomalías en las series de datos para mejorar la precisión en las curvas de demanda eléctrica proyectadas.

**Palabras clave**— Aprendizaje Automático, Minería de datos, Potencia Eléctrica, Proyección de demanda a corto plazo, Sistema Nacional Interconectado.

Recibido: 04-04-2021, Aprobado tras revisión: 23-07-2021

Forma sugerida de citación: Gallo, A.; Pérez, F.; Salinas, D. (2021). "Minería de Datos y Proyección a Corto Plazo de la Demanda de Potencia en el Sistema Eléctrico Ecuatoriano". Revista Técnica "energía". No. 18, Issue I, Pp. 72-85

ISSN On-line: 2602-8492 - ISSN Impreso: 1390-5074

© 2021 Operador Nacional de Electricidad, CENACE

## 1. INTRODUCCIÓN

El pronóstico de demanda eléctrica es el insumo principal para una adecuada operación y expansión del sector eléctrico ecuatoriano; además, es el primer paso crucial para cualquier estudio de planificación. Algunas aplicaciones del pronóstico de la demanda son: abastecer el balance de generación-demanda, planificar la operación del sistema, elaborar planes de expansión. Por tales razones, es de sumo interés que los resultados de los estudios de proyección de demanda sean los más precisos posibles.

Se entiende que una proyección de demanda a corto plazo, utilizada para estudios de operación, es significativamente diferente del largo plazo utilizada en estudios de planificación. En la proyección de demanda a corto plazo, por ejemplo, de la próxima semana, se predice la demanda para cada hora de la próxima semana. En este contexto, la proyección de demanda a corto plazo es un estudio eléctrico importante para la planificación y operación del Sistema Nacional Interconectado (SNI). Una mala planificación puede traer consigo una mala programación del abastecimiento de la demanda horaria de potencia eléctrica; provocando pérdidas al requerir unidades de generación más costosas para suplir la demanda.

Sin embargo, en bases de datos históricas robustas (demanda de potencia horaria) la presencia de datos faltantes y la presencia de desviaciones atípicas “outliers” en las series de datos, es un factor clave de análisis previo a la elaboración de cualquier proyección, ya que afectan directamente a la calidad y precisión de los diferentes métodos de predicción. Los procedimientos más frecuentes para tratar problemas de datos faltantes y valores atípicos han sido la eliminación o la sustitución/imputación de estos por valores como la media de la variable, siendo procedimientos clásicos que en general presentan inconvenientes y carencias [1].

El aprendizaje automático como un subcampo de la inteligencia artificial ha introducido una gran variedad de técnicas aplicadas en el campo de estudios de proyección a corto plazo del consumo de electricidad, mostrando un mejor rendimiento que las técnicas tradicionales como el algoritmo ARIMA para series de tiempo [2]. Por ejemplo, en estudios como [2], [3] se utiliza el algoritmo ARIMA para validar la aplicación de algoritmos de inteligencia artificial para la proyección de demanda a corto plazo, así mismo en [4], [5] se propone el uso de varios algoritmos computacionales para realizar tareas de proyección de demanda y al comparar sus resultados mediante indicadores de error se infiere que algoritmo tiene un mejor rendimiento. No obstante, no se presenta a detalle el ajuste y calibración de los algoritmos de aprendizaje automático para las series de datos, el ajuste adecuado de un algoritmo predictivo es uno de los principales objetivos y desafíos en el proceso de aprendizaje automático, ya que permite mejorar el rendimiento de las proyecciones realizadas. Además, se

presentan metodologías para la detección de valores atípicos que no indican un procedimiento claro y eficaz para tratar con anomalías presentes en las series de datos.

Sobre la base de lo mencionado, se ha propuesto un procedimiento para minería de datos y proyección a corto plazo de demanda horaria de potencia eléctrica incorporado en dos módulos ejecutables computacionales desarrollados en lenguaje de programación Python. El módulo de minería de datos tiene como objetivo identificar y mitigar desviaciones y comportamientos anómalos en los datos de estudio. Esclarecer dichos comportamientos permite contar con bases de datos completas y confiables, que permiten la adecuada calibración de cada algoritmo mediante la implementación de funciones de autoajuste para disminuir el error en los resultados del módulo de proyección de demanda eléctrica.

## 2. MARCO TEÓRICO

### 2.1. Modelos de proyección

Los modelos de proyección se clasifican en dos grandes grupos [6] que se describen en las secciones 2.1.1 y 2.1.2.

#### 2.1.1 Técnicas basadas en inteligencia artificial (IA)

Las técnicas basadas en inteligencia artificial incluyen métodos de aprendizaje automático, como redes neuronales artificiales, árboles de decisión, bosques aleatorios, máquina de vector soporte, y entre otros que han tenido un éxito notable cuando se trata de sistemas de potencia.

#### 2.1.2 Enfoques estadísticos

Son modelos paramétricos convencionales para análisis de series de tiempo estocásticas, tales como regresión lineal, suavizado exponencial, y el modelo Autorregresivo Integrado de Media Móvil (ARIMA) [6], [7].

#### Modelo Autorregresivo Integrado de Media Móvil

El modelo está compuesto por tres partes, un modelo de auto regresión (AR) donde hay una combinación de valores pasados; un componente de promedio móvil (MA), que utiliza errores de proyecciones pasadas en un modelo similar a la regresión; y una integración (I) [2].

Un modelo ARIMA no estacional se lo denomina un modelo “ARIMA ( $p, d, q$ )”, donde:

- $p$ : Es el número de términos autorregresivos.
- $d$ : Es el número de diferencias no estacionales necesarias para que la serie se vuelva estacional.
- $q$ : Es el número de errores de pronóstico en retraso en la ecuación de predicción.

Con los parámetros  $p, d, q$  establecidos y si se considera una serie temporal estacional para simplificar

las cosas, con  $d=0$ , la ecuación se puede escribir como:

$$y_t = \phi_0 + \phi_1 \times y_{t1} + \phi_2 \times y_{t2} + \dots + \phi_p \times y_{tp} + \theta_0 + \theta_1 \times e_{t1} + \theta_2 \times e_{t2} + \dots + \theta_q \times e_{tq} \quad (1)$$

Donde  $\phi_0, \phi_1, \dots, \phi_p$ , son los coeficientes autorregresivos. Los retrasos asociados son  $y_{t1}, y_{t2}$  hasta el grado de polinomio autorregresivo  $p$ . Y  $\theta_0, \theta_1, \dots, \theta_q$  son los coeficientes de promedio móvil. Los retrasos asociados son  $e_{t1}, e_{t2}, \dots$  hasta el grado del polinomio de promedio móvil  $q$ .

El algoritmo ARIMA se encuentra disponible en la biblioteca "statsmodels" de Python, y puede ser modelado y ajustado para problemas de predicción de demanda de potencia eléctrica.

### 2.1.3 Evaluación de la precisión en la proyección

A medida que hay más algoritmos de pronóstico disponibles, es cada vez más importante evaluar que tan cerca están los resultados de proyección de los valores reales. El Porcentaje de Error Medio Absoluto (MAPE, por sus siglas en inglés), es un indicador de precisión y rendimiento de la proyección de demanda que mide el tamaño del error absoluto en términos porcentuales. Dicha métrica es implementada para medir la calidad de los resultados de proyección del presente estudio debido a que es un indicador fácil de interpretar y evaluar.

#### Error porcentual medio absoluto (MAPE)

El Error Porcentual Medio Absoluto (*Mean Absolute Percentage Error*) es el porcentaje de error promedio en un conjunto de valores proyectados. Varía desde cero en adelante, y un valor más bajo indica un mejor modelo [8]. Se describe matemáticamente como:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

Donde:

$y_i$ : Es el valor real por pronosticar

$\hat{y}_i$ : Es el valor pronosticado

$n$ : E el número de datos pronosticados o evaluados

### 2.2. Anomalías en series de tiempo de demanda eléctrica

En el análisis de series de tiempo, es un requerimiento fundamental detectar valores faltantes y valores atípicos. Esto debido a que, mientras más escasos sean los datos, más difícil será crear un pronóstico preciso [6].

A continuación, se presentan tres alternativas para tratar anomalías en las series de tiempo.

- Sustituir valores atípicos con un promedio de valores vecinos.
- Pronosticar los valores con un método predictivo.
- Eliminar las observaciones.

Debido al enfoque predictivo de los algoritmos de aprendizaje automático se determinó que, el reemplazo de anomalías por un valor proyectado más probable es la mejor opción para el desarrollo de este estudio.

### 2.3. Aprendizaje automático

La tarea principal del aprendizaje automático es modelar algoritmos que sean capaces de aprender de datos históricos y realizar predicciones sobre nuevos datos de entrada [9].

De manera general en la Fig. 1 se presenta un esquema de las etapas que conlleva un problema de aprendizaje automático.

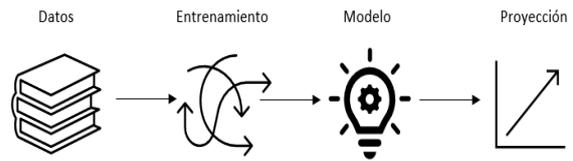


Figura 1: Esquema general de aprendizaje automático [10]

#### 2.3.1 Aprendizaje supervisado

##### Regresión

El resultado a predecir es un valor continuo. Existe una gran gama de algoritmos de aprendizaje supervisado que están disponibles en la biblioteca scikit-learn de Python. Entre ellos se encuentran disponibles algoritmos tipo ensemble como el algoritmo Random Forest que ha tenido resultados muy buenos en su aplicación como algoritmo de regresión de aprendizaje supervisado [2].

- **Algoritmo Decision Tree Regressor:** Es un modelo de decisiones similares a un árbol que representa un algoritmo que solo contiene sentencias de control condicional [12].
- **Algoritmo Random Forest Regressor:** Un bosque aleatorio ajusta varios árboles de decisión de submuestras aleatorias del conjunto de datos, en donde el propósito es disminuir la varianza del estimador, ya que los árboles de decisión individuales suelen presentar una gran varianza [12], [13].
- **Algoritmo Light Gradient Boosting Machine Regressor (LGBM):** Se constituye a partir de modelos de árboles de decisión. Los árboles se agregan uno a la vez y se ajustan para corregir los errores de predicción cometidos por modelos anteriores.

#### 2.3.2 Ingeniería de las características

La Ingeniería de las Características (Feature Engineering) es el proceso de crear o mejorar características o atributos del conjunto de datos. Las características se crean en función del sentido común, experiencia, o dominio del conocimiento. El resultado de este proceso es un conjunto significativo de

características que los algoritmos pueden procesar para identificar patrones en el conjunto de datos de entrada y construir un mejor modelo de aprendizaje automático [9], [12].

### 2.3.3 Ajuste de hiperparámetros

Los algoritmos de aprendizaje automático tienen un conjunto específico de parámetros (denominados también hiperparámetros) que deben estimarse a partir del conjunto de datos de entrada. Los principales hiperparámetros de los algoritmos utilizados en el presente estudio constituidos a partir de árboles de decisión son: número de estimadores, profundidad máxima, características máximas, y criterio [12].

Se han desarrollado funciones que permiten ajustar automáticamente los diferentes hiperparámetros [14].

#### Búsqueda de Cuadrícula:

En la función Búsqueda de Cuadrícula los hiperparámetros son proporcionados por el analista, y la mejor combinación es elegida según un indicador de puntuación de error. La combinación de hiperparámetros forma una cuadrícula de búsqueda.

#### Hyperopt

Función que proporciona la configuración automática de algoritmos de la biblioteca de aprendizaje automático scikit-learn [15]. Hyperopt adiciona un algoritmo de búsqueda para optimización de hiperparámetros denominado Estimador de Parzen Estructurado en Árbol. El algoritmo TPE, por sus siglas en inglés (*Tree-structured Parzen Estimator*) busca elegir los parámetros con mejor rendimiento para el siguiente paso, dejando atrás valores deficientes.

## 3. PROCEDIMIENTO PARA ANÁLISIS PREDICTIVO Y DESARROLLO DE LA HERRAMIENTA COMPUTACIONAL

### 3.1. Demanda del Sistema Nacional Interconectado

La Fig. 2 representa la demanda horaria de potencia eléctrica total del Sistema Nacional Interconectado por día y registrada en periodos de 30 minutos desde el 1 de enero de 2014 hasta el 30 de junio de 2020 [16].

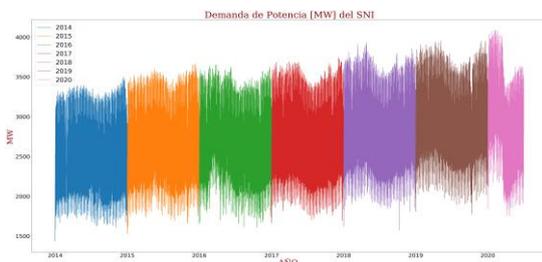


Figura 2: Demanda de potencia eléctrica SNI enero 2014-junio 2020 [16]

Los datos registrados en MW con un periodo de 30 minutos son recopilados como se indica en la Fig. 3,

representando una serie de tiempo en donde existen 48 datos por día y 17 520 datos recopilados por año.

Fecha	Hora	Total
30/6/2014	00:00:00	2 088,6
30/6/2014	00:30:00	1 996,0
30/6/2014	01:00:00	1 985,2
.	.	.
.	.	.
30/6/2014	23:30:00	2 451,3

Figura 3: Formato de registro de mediciones de la demanda horaria de potencia eléctrica del SIN [Elaboración Propia]

### 3.2. Declaración de un problema de aprendizaje automático supervisado

Un problema de aprendizaje supervisado donde los atributos o características son variables de entrada relacionadas con la variable objetivo, que se considera como dependiente de las entradas, consiste en modelar un problema de entrada/salida, donde la entrada es del tipo matricial y la salida es un vector como se indica en la Fig. 4.

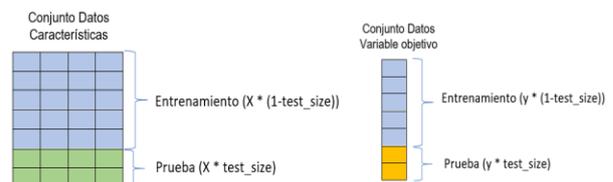


Figura 4: Estructura del conjunto de datos de variables características y variable objetivo

Para definir un conjunto de datos como problema de aprendizaje supervisado en primer lugar se define las variables características “X” (entrada) y la variable de salida “y” (objetivo), después se divide dichos conjuntos en subconjuntos de prueba y entrenamiento como se observa en la Fig. 4. El tamaño (*test size*) puede variar entre 10%, 20% o 30% del tamaño del conjunto de datos. Al realizar un análisis de sensibilidad para el conjunto de datos de demanda de potencia cambiando el tamaño del conjunto de prueba desde el 30% al 10% para las series de datos de demanda de potencia, el rendimiento del algoritmo evaluado por la métrica de precisión MAPE con valores desde 0,96% al 0,94% presentó la tendencia de mejorar con la disminución del tamaño del conjunto de entrenamiento, o visto de otra manera, con el aumento del tamaño del conjunto de datos de entrenamiento. Sin embargo, con un conjunto de datos de prueba del 10% se obtuvo un rendimiento decreciente al volver a un MAPE del 0,95%. Para el desarrollo de este trabajo se utilizó un tamaño del 11% para el conjunto de datos de prueba debido a que dicho porcentaje provocó que el entrenamiento del algoritmo de proyección se lleve de manera más adecuada, y dado que, los algoritmos de aprendizaje automático no requieren de una muestra exagerada de conjunto de datos de entrada para inferir predicciones de gran calidad, y también, por los esfuerzos computacionales que esto involucra; para cada tarea de proyección en las series de datos con históricos anuales de demanda de potencia eléctrica, se ajustaron como

conjunto de entrada, las mediciones correspondientes a los últimos dos meses, tomando como referencia la fecha inicial de proyección.

Además, se debe señalar que, debido al horizonte de proyección a corto plazo los datos históricos de demanda de potencia deben corresponder a los registros de mediciones más recientes. Esto debido a que, los resultados de los estudios de proyección en el corto plazo son utilizados principalmente en la operación y control del sistema de potencia.

En segundo lugar, se modela el algoritmo de aprendizaje automático ajustando sus hiperparámetros. Luego, el conjunto de entrenamiento (entradas y salidas) proporciona un modelo ajustado a los datos. La evaluación del modelo se realiza con el conjunto de prueba de variables características y variable de salida.

Finalmente, el modelo se encuentra ajustado y es validado con una métrica de rendimiento, de tal manera que si se proporciona un nuevo conjunto de variables características como entrada; el modelo es capaz de inferir valores proyectados. La métrica para evaluar las predicciones en el presente estudio es el indicador de error MAPE presentado en la sección 2.1.3.

### 3.3. Enfoques de aprendizaje supervisado para series de tiempo y aplicación de la ingeniería de características

Para acoplar series de tiempo (Fig. 3) en un problema de aprendizaje supervisado es necesario representar una serie de tiempo como un modelo autorregresivo (AR), donde el modelo entrega la proyección de  $y_{t+1}$  en función de los  $n$  valores previos, como se observa en la Fig. 5. Para ello, en Python se retrasa la serie en un número específico de pasos, en este estudio el número de pasos de retraso corresponde a 48, debido al número de registros diarios de demanda de potencia recopilados cada 30 minutos.

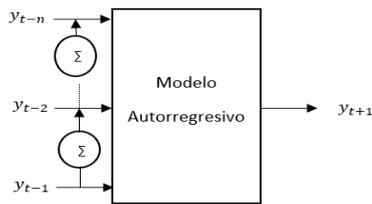


Figura 5: Estructura de un modelo autorregresivo [17]

Además de inferir retrasos y estructurar un modelo de regresión en las series de tiempo, el procedimiento de ingeniería de características permite inferir nuevas características en las series de datos. Esta reestructuración del conjunto de datos permite definir un modelo matricial multivariable no lineal (ver Fig. 6) que toma en cuenta el comportamiento de las diferentes variables intervinientes (hora, tipo de día, etc.) como los modelos implementados en [2], [3].

Fecha	y	lag_1	...	lag_48	hour	weekday	is_weekend	d_prom	d_max	d_min
2014-01-02 00:00:00	1.882,2	1910,35	..	2140,04	0	3	0	1885,84021	2536,64	1436,05
2014-01-02 00:30:00	1.750,5	1882,19	..	2121,6	0	3	0	1880,46833	2536,64	1436,05
2014-01-02 01:00:00	1.697,4	1750,52	..	2090,73	1	3	0	1872,7375	2536,64	1436,05
2014-01-02 01:30:00	1.657,0	1697,43	..	2005,27	1	3	0	1864,54375	2536,64	1436,05
2014-01-02 23:30:00	2.463,61	2621,76	..	2553,13	23	2	0	2712,2325	3362,87	2049,41

Valor actual depende de su valor en retraso

Figura 6: Reestructuración final de series de tiempo

### 3.4. Identificación y tratamiento de anomalías en las series de datos

#### 3.4.1 Datos Faltantes

Funciones de la biblioteca pandas de Python permiten crear arreglos de fecha y tiempo en un intervalo con frecuencia definida por el usuario. Como se ilustra en la Fig. 7 el primer arreglo se compara fecha por fecha con las fechas de los datos históricos de demanda horaria de potencia eléctrica existentes en un periodo anual. En el caso de detectar una comparación nula con una fecha se almacena la fecha faltante.



Figura 7: Detección de valores faltantes

#### 3.4.2 Valores atípicos

El procedimiento de detección de valores atípicos, a través del crecimiento anual de potencia eléctrica, consiste en determinar el crecimiento entre un dato de potencia horaria registrada en el periodo actual y el mismo periodo del siguiente año como se indica en (3).

$$Crecimiento \% = \left( \frac{MW \text{ Actual}}{MW \text{ Anterior}} - 1 \right) \times 100 \quad (3)$$

Para determinar la presencia de valores atípicos se utilizó el gráfico de caja, el cual es un gráfico que se construye a partir de los cuartiles y la media del conjunto de datos. En la Fig. 8 se presentan los gráficos de caja para el crecimiento horario anual entre los periodos 2017-2018 y 2018-2019.

Se puede observar que los crecimientos por encima del 30% o por debajo del -20% indican posibles límites umbrales de crecimiento que pueden ser ingresados por usuario para detectar comportamientos anómalos en las series de datos.



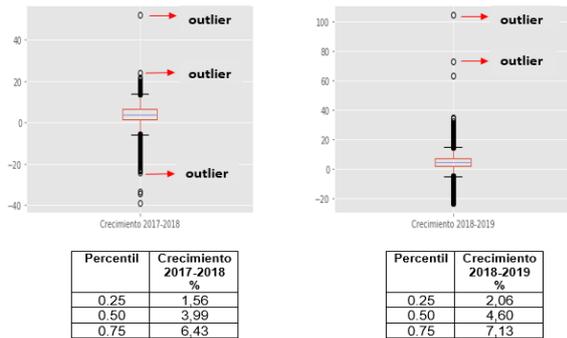


Figura 8: Gráfico de caja crecimiento anual

### 3.4.3 Procedimiento para tratar con anomalías en las series de tiempo

Al tener presente el enfoque predictivo del aprendizaje automático se determinó que, el reemplazo de anomalías por un valor pronosticado más probable es la mejor opción para el desarrollo del presente estudio.

Un algoritmo de aprendizaje automático requiere de un conjunto de datos de entrada para poder realizar nuevas predicciones en su salida. Dichas entradas pueden corresponder a valores en intervalos anteriores según el horizonte de tiempo que se desee proyectar, por ejemplo, si se desea proyectar la medición en la hora siguiente, la entrada es la medición de la hora anterior; de esta manera, proyectar el día siguiente correspondería a las mediciones del día anterior.

Para calibración de hiperparámetros se utilizó la función Hyperopt, y a manera de ejemplo ilustrativo del procedimiento de predicción, a continuación, se describe el procedimiento para la proyección de 48 mediciones de demanda de potencia que corresponden a la demanda del SNI el día 2014-04-24. Del procedimiento presentado en la sección 3.3 se obtiene el conjunto de datos *data* que se ilustra en la Fig. 9, compuesto por conjuntos de variables características “X” y variables objetivo “y” para el entrenamiento y prueba del algoritmo de predicción. Para determinar el conjunto de datos de entrada requerido por el algoritmo Random Forest para la proyección de un día de demanda de potencia, se divide el conjunto *data* en dos partes como se observa en la Fig.9; la primera parte corresponde a todos los datos hasta dos días antes del día a proyectar (conjunto de entrenamiento), y la segunda parte corresponde a los datos de demanda MW del día anterior (conjunto de entrada).

Finalmente se elimina la variable objetivo “y” y con este conjunto de datos de entrada el algoritmo puede realizar 48 predicciones futuras de demanda de potencia. Para evaluación de los diferentes algoritmos implementados en las tareas de predicción sobre las series históricas de demanda de potencia eléctrica en análisis se eliminaron algunos días correspondientes a 48 datos registrados de demanda de potencia eléctrica en los datos históricos del SNI para proceder a realizar las predicciones correspondientes. El resultado del análisis mostró que el algoritmo que mejor se acopla a las series

históricas de demanda horaria de potencia eléctrica del SNI, es el algoritmo Random Forest. Debido a esto, el algoritmo Random Forest fue implementado en el módulo de minería de datos.

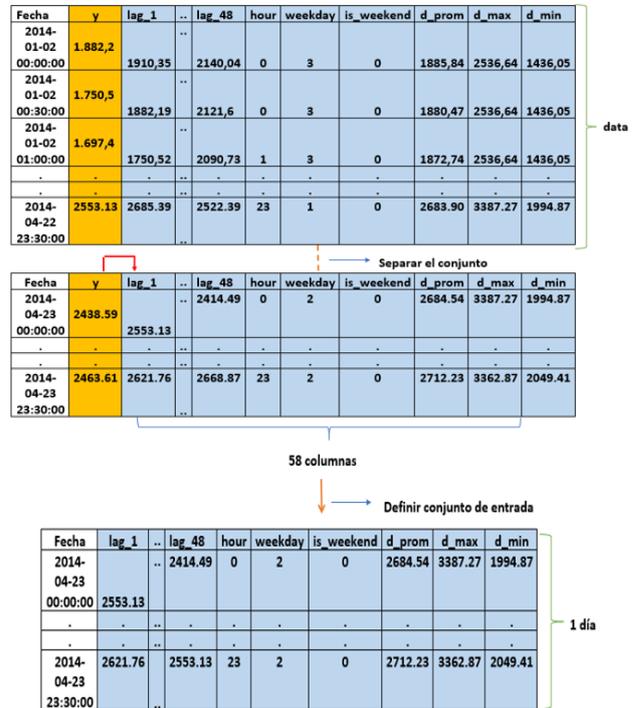


Figura 9: Conjunto de datos de entrada para proyectar un día de demanda de potencia MW

## 3.5. Proyección a corto plazo de la demanda horaria de potencia eléctrica

### 3.5.1 Procedimiento para la proyección a corto plazo de la demanda horaria de potencia eléctrica

El procedimiento llevado a cabo para la proyección de demanda horaria de potencia eléctrica del SNI con un horizonte de dos semanas se definió a partir de los conceptos de la sección 3.4.3; no obstante, el conjunto de datos de entrada requerido por los algoritmos de proyección debe corresponder con un intervalo de 7 días, y no de un solo día. Al realizar la proyección semanal con el conjunto de entrada especificado, se procede a estructurar un nuevo conjunto de entrada semanal (7 días) con dichos datos proyectados, para la predicción de la siguiente semana. Por ejemplo, en la Fig. 10 se representa la forma en que se realiza la proyección de demanda horaria desde el 2014-04-30 al 2014-05-13.

Es importante mencionar que, si se debe proyectar un día festivo, el conjunto de entrada se estructura con el mismo día feriado del año anterior; de esta manera, el comportamiento de proyección de demanda asimilará la tendencia de consumo que se presentó el mismo día festivo del año anterior, lo que permite reducir el margen de error en la proyección de días especiales.

Conjunto de Entrada	Proyección
2014-04-23	→ 2014-04-30
2014-04-24	→ 2014-05-01
.	→ .
.	→ .
2014-04-29	→ 2014-05-06

Figura 10: Conjunto de datos de entrada para proyección semanal

### 3.6. Herramienta Computacional

La programación de la herramienta computacional en Python se estructuró de forma modular con la aplicación de subrutinas. El diagrama de flujo de la Fig.11 esquematiza todos los procesos ordenados para llevar a cabo la ejecución de todas las tareas que brinda la herramienta computacional.

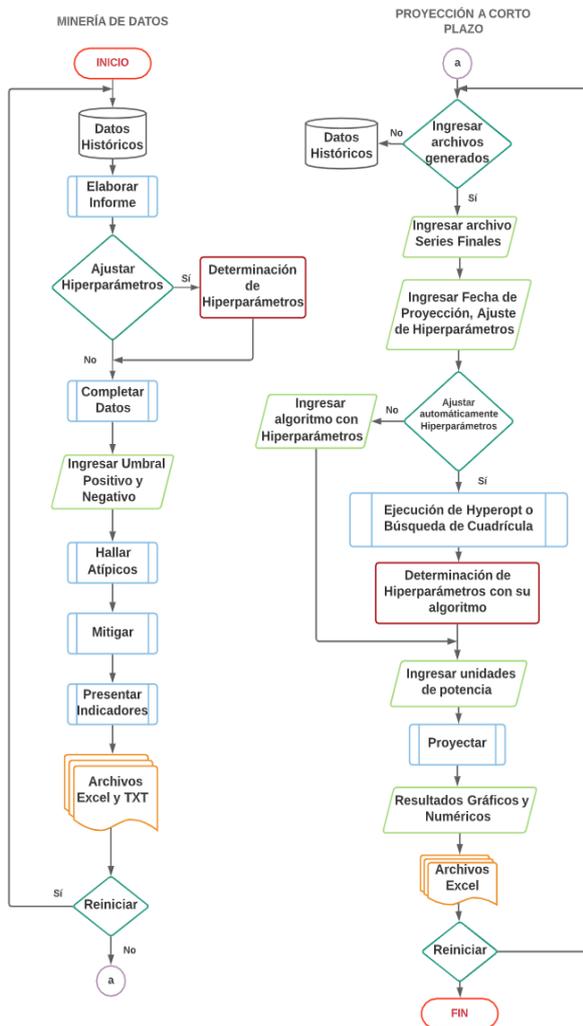


Figura 11: Diagrama de flujo de la herramienta computacional

A pesar de que el módulo de minería de datos y el módulo de proyección a corto plazo constituyen una misma interfaz gráfica, cada módulo puede ejecutarse de manera independiente con el fin de realizar análisis por separado de minería de datos o proyección de demanda. El usuario puede pasar de un módulo a otro dependiendo de las necesidades de estudio que requiera; sin embargo,

como se verá más adelante, mitigar anomalías en las series de datos permite en general un mejor entrenamiento y ajuste de un algoritmo de proyección.

## 4. APLICACIÓN DE LA HERRAMIENTA COMPUTACIONAL Y ANÁLISIS DE RESULTADOS

### 4.1. Análisis de minería de datos

Las tareas de minería de datos que se realizan a continuación permiten estructurar series completas y libres de datos atípicos para poder utilizarlas en las tareas de proyección de demanda. Las tareas de detección, imputación, y mitigación de valores faltantes y valores atípicos se llevan sobre los registros históricos de la demanda de potencia eléctrica del SNI, con el fin de resaltar la importancia de eliminar anomalías en los registros de mediciones de demanda previos al periodo de proyección, ya que dichos registros afectan directamente la calidad de proyección a corto plazo.

#### 4.1.1 Análisis y detección de datos faltantes

El análisis exploratorio de datos presenta información característica estructurando un informe que es mostrado en pantalla por la interfaz gráfica y, a la vez, esta información es exportada a un archivo de texto como se presenta en la Fig. 12.

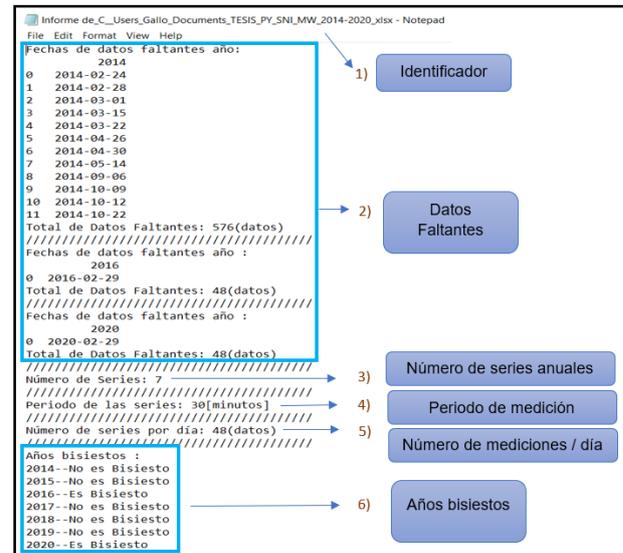


Figura 12: Informe preliminar

La información presentada permite ubicar los datos faltantes por fecha y para cada año de las series históricas en análisis; datos faltantes correspondientes a un total de 672 mediciones. El informe también realiza un análisis exploratorio sobre las series de datos, indicando el número de series anuales de los datos históricos ingresados, calcula el periodo de registro de medición y el número de mediciones por día. También indica si es que un año corresponde o no a un año bisiesto.

### 4.1.2 Mitigación de valores faltantes

Según se indicó en la sección 3.6 la herramienta computacional cuenta con la subrutina Completar Datos que permite la imputación de datos faltantes por medio del procedimiento descrito en la sección 3.4.3.

En la Tabla 1 se presenta el formato correspondiente al archivo Excel generado con los resultados obtenidos; por ejemplo, se puede observar que el día 2014-04-30, detectado en el procedimiento anterior como un día con datos faltantes, es proyectado e imputado en la serie de datos originales. De la misma manera, todos los datos faltantes de cada fecha presentada en el informe preliminar son reemplazados por un valor proyectado.

Tabla 1: Formato en Excel de Series Completas

Fecha	Hora	MW
2014-04-29	23:00:00	2666,39
2014-04-29	23:30:00	2552,93
2014-04-30	0:00:00	2305,565
2014-04-30	0:30:00	2212,006
2014-04-30	1:00:00	2143,655
2014-04-30	1:30:00	2087,64
2014-04-30	2:00:00	2071,442
2014-04-30	2:30:00	2032,114
2014-04-30	3:00:00	2046,558
2014-04-30	3:30:00	2002,75
2014-04-30	4:00:00	2007,757
2014-04-30	4:30:00	2022,757
2014-04-30	5:00:00	2042,243
2014-04-30	5:30:00	2187,772
2014-04-30	6:00:00	2207,476
2014-04-30	6:30:00	2210,748

Mediciones faltantes proyectadas

### 4.1.3 Detección de valores atípicos

Los umbrales de crecimiento anual ingresados por usuario, como se mencionó en la sección 3.4.2, para análisis de las series históricas de demanda de potencia eléctrica del SNI fueron +35% y -25%. Los resultados de la tarea de detección de valores atípicos (outliers) son exportados a un archivo Excel como se muestra en la Tabla 2. Los valores atípicos se presentan en diferentes hojas de cálculo, indicando los outliers detectados en todos los periodos.

En la Tabla 3 se presenta un resumen del número total de valores atípicos encontrados en cada periodo de crecimiento anual. Puede observarse que los periodos con presencia de un mayor número de datos atípicos en los históricos de demanda horaria de potencia nacional corresponden a los años 2015-2016, 2016-2017, y 2019-2020.

Tabla 2: Formato en Excel con outliers detectados

2015	Total 2015	2016	Total 2016	Crec.
2015-04-18 19:30	3150,02	2016-04-16 19:30	1789,72	-43,2
2015-04-18 20:00	3103,88	2016-04-16 20:00	1896,31	-38,9
2015-04-18 20:30	3077,59	2016-04-16 20:30	2070,13	-32,7
2015-04-18 21:00	2988,48	2016-04-16 21:00	2085,56	-30,2
2015-04-18 21:30	2921,84	2016-04-16 21:30	2128,61	-27,1
2015-07-03 10:30	2999,08	2016-07-01 10:30	1984,7	-33,8
2015-07-03 11:00	3061,80	2016-07-01 11:00	2098,22	-31,5
2015-10-22 13:30	2993,69	2016-10-20 13:30	1612,48	-46,1
2015-10-22 14:00	3008,18	2016-10-20 14:00	1854,82	-38,3
2015-11-19 11:30	3073,85	2016-11-17 11:30	1679,3	-45,4
2015-12-02 11:30	3048,14	2016-11-30 11:30	1976	-35,1

Tabla 3: Datos atípicos identificados en históricos de demanda

Periodo	Número de datos atípicos encontrados	
	Crecimiento Anual > 35 %	Crecimiento Anual < -25 %
2015-2016	0	11
2016-2017	11	1
2017-2018	1	2
2018-2019	3	0
2019-2020	0	30
Total	15	44

### 4.1.4 Mitigación de valores atípicos

La subrutina Mitigar permite reemplazar los valores atípicos identificados en el apartado anterior. Los reemplazos se realizan eliminando dichos valores y aplicando el procedimiento para tratar anomalías en las series de datos descrito en la sección 3.4.3.

Como ejemplo, en la Tabla 4 y en la Tabla 5 se presentan los valores atípicos identificados y reemplazados el sábado 16 de abril de 2016; atípicos reemplazados por valores más próximos a un comportamiento normal de demanda de potencia eléctrica en las respectivas horas.



Tabla 4: Valores atípicos detectados el 2016-04-16

2015	MW 2015	2016	MW 2016	Crec.
2015-04-18 19:30	3150,02	2016-04-16 19:30	1789,72	-43,2
2015-04-18 20:00	3103,88	2016-04-16 20:00	1896,31	-38,9
2015-04-18 20:30	3077,59	2016-04-16 20:30	2070,13	-32,7
2015-04-18 21:00	2988,48	2016-04-16 21:00	2085,56	-30,2
2015-04-18 21:30	2921,84	2016-04-16 21:30	2128,61	-27,1
2015-07-03 10:30	2999,08	2016-07-01 10:30	1984,7	-33,8
2015-07-03 11:00	3061,80	2016-07-01 11:00	2098,22	-31,5
2015-10-22 13:30	2993,69	2016-10-20 13:30	1612,48	-46,1
2015-10-22 14:00	3008,18	2016-10-20 14:00	1854,82	-38,3
2015-11-19 11:30	3073,85	2016-11-17 11:30	1679,3	-45,4
2015-12-02 11:30	3048,14	2016-11-30 11:30	1976	-35,1

Tabla 5: Valores atípicos reemplazados en las series de datos

2016	MW 2016
2016-04-16 19:00	2603,44
2016-04-16 19:30	3387,178
2016-04-16 20:00	3355,938
2016-04-16 20:30	3300,941
2016-04-16 21:00	3196,891
2016-07-01 21:30	3089,602

#### 4.1.5 Indicadores

El módulo de minería de datos permite calcular el factor de carga mensual para las series históricas. Se presenta al factor de carga como un indicador de la mejora de la curva de demanda diaria; debido a su relación directa con la demanda promedio, y al hecho de que al eliminar tendencias atípicas en las series de datos dicha magnitud de demanda tiende a aumentar.

### 4.2. Análisis de proyección de demanda a corto plazo

#### 4.2.1 Ajuste de Hiperparámetros

El módulo de proyección de demanda de potencia eléctrica a corto plazo permite autoajustar los principales hiperparámetros del algoritmo de aprendizaje automático Random Forest con la función Hyperopt, para obtener un modelo adecuado de predicción para los datos de demanda de potencia en análisis.

#### 4.2.2 Importancia de mitigar valores faltantes y valores atípicos

Según el procedimiento descrito en la sección 3.5.1 para la proyección a corto plazo de la demanda horaria de

potencia eléctrica, la primera semana de proyección depende directamente de la semana inmediatamente anterior, como se observa en la Fig. 13, en donde la proyección de demanda va desde el 2016-04-20 hasta el 2016-05-03 (2 semanas). En el primer recuadro se observa que el día de proyección 23 de abril de 2016 depende netamente del día 16 de abril de 2016, razón por la cual, si el día 16 de abril presenta algún comportamiento anormal, y lo cual, si ocurrió debido al terremoto suscitado en Ecuador en dicha fecha, repercute en la calidad de predicción a llevarse a cabo el día 23 de abril de 2016.



Figura 13: Proyección de la demanda de potencia del 2016-04-20 al 2016-05-03

Por lo mencionado, en la Fig. 14 se presenta una parte de la proyección realizada por el algoritmo Random Forest para el periodo indicado en el párrafo anterior, específicamente la proyección del día 23 de abril de 2016, donde se puede observar que los comportamientos anómalos detectados el 16 de abril de 2016 provocan errores porcentuales muy notables en los resultados de proyección, al contrario de trabajar con series libres de tendencias atípicas que permiten reducir el error porcentual. La disminución de los errores porcentuales permite mantener un margen aceptable como en los resultados del estudio realizado para proyección a corto plazo en [3], [4].

Es importante mencionar que, tratar las tendencias atípicas que afectan a los datos de entrada para la tarea de proyección de demanda de potencia del algoritmo predictivo depende del criterio del analista, ya que, si la tendencia de consumo de demanda de potencia es afectada a mediano o largo plazo por factores no controlables como pandemias o fenómenos naturales a gran escala, provocará que el comportamiento de la demanda se mantenga por debajo de un consumo normal en un periodo de tiempo extendido. De esta manera, el módulo de proyección de demanda puede trabajar con las series de datos que provienen del módulo de minería de datos o cualquier otra serie de datos de demanda de potencia eléctrica en las que se requiera realizar tareas de proyección.



Proyección utilizando las series históricas originales de demanda horaria del SNI					Proyección utilizando las series históricas de demanda horaria de potencia del SNI resultantes del módulo de minería de datos				
Fecha	Hora	Proyección	MW Real	Error %	Fecha	Hora	Proyección	MW Real	Error %
2016-04-23	17:30	2584,2	2528,84	0,76	2016-04-23	17:30	2584,2	2528,84	0,76
2016-04-23	18:00	2612,4	2558,45	2,06	2016-04-23	18:00	2603,2	2558,45	1,72
2016-04-23	18:30	2894,1	2914,21	0,7	2016-04-23	18:30	2923,1	2914,21	0,3
2016-04-23	19:00	3104,2	3064,59	1,28	2016-04-23	19:00	3090,1	3064,59	0,83
2016-04-23	19:30	2965,8	3109,85	4,85	2016-04-23	19:30	3000,8	3109,85	3,63
2016-04-23	20:00	2583,4	3030,73	17,31	2016-04-23	20:00	3181,53	3030,73	4,74
2016-04-23	20:30	2316,5	3032,73	30,92	2016-04-23	20:30	3187,9	3032,73	4,87
2016-04-23	21:00	2267,5	2960,63	30,57	2016-04-23	21:00	3151,3	2960,63	6,05
2016-04-23	21:30	2297,9	2881,2	25,39	2016-04-23	21:30	3088,7	2881,2	6,72
2016-04-23	22:00	2322,1	2776,96	19,59	2016-04-23	22:00	3012,2	2776,96	7,81
2016-04-23	22:30	2305,6	2704,19	17,29	2016-04-23	22:30	2511,9	2704,19	7,65
2016-04-23	23:00	2198,6	2575,59	17,15	2016-04-23	23:00	2250,5	2575,59	14,44
2016-04-23	23:30	2089,8	2460,85	17,76	2016-04-23	23:30	2098,8	2460,85	17,25

Figura 14: Proyección del día 23 de abril de 2016

#### 4.2.3 Resultados de proyección a corto plazo de la demanda de potencia eléctrica del SNI

Dos de los escenarios definidos para realizar proyecciones correspondieron a temporada seca en julio de 2019 y temporada lluviosa en marzo de 2020, otro escenario definido correspondió a semanas con presencia de días feriados en el mes de febrero de 2020.

En la Fig. 15 se presenta las proyecciones de demanda horaria de potencia eléctrica con sus respectivas curvas de comparación, para la temporada seca que va desde el 2019-07-15 hasta el 2019-07-28. Los resultados de proyección en temporada seca exponen un error MAPE menor al 3%.

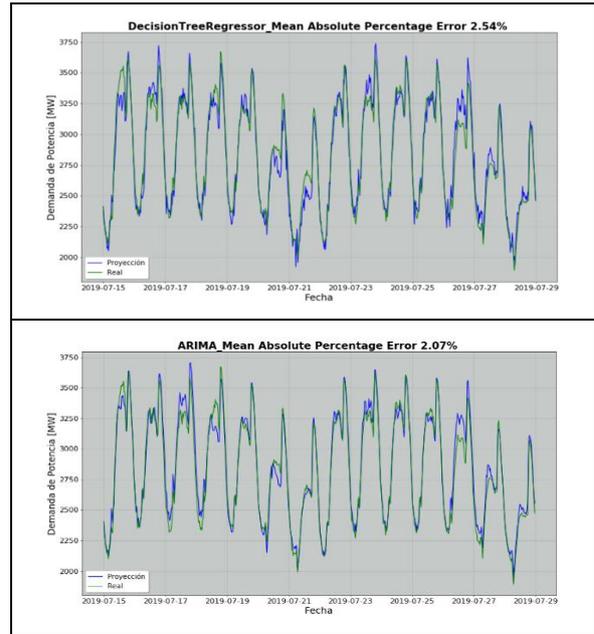
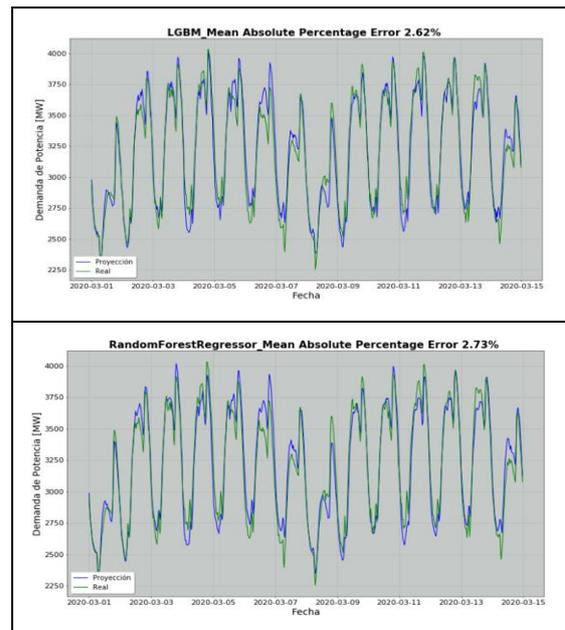
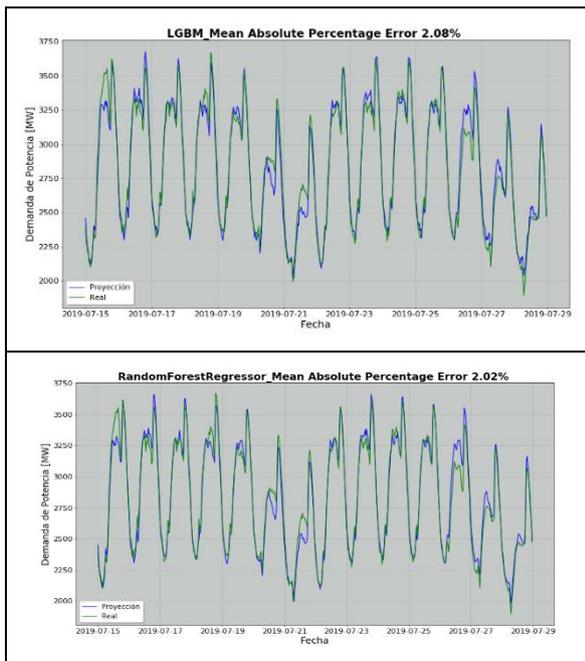


Figura 15: Temporada seca Desde 2019-07-15 hasta 2019-07-28

En la Fig.16 se presenta las curvas de proyección obtenidas para un escenario de temporada lluviosa desde el 2020-03-01 hasta el 2020-03-14.

Los resultados para las proyecciones de demanda de potencia eléctrica en temporada lluviosa se presentan con porcentajes de error medio absoluto menores al 4% para los algoritmos LGBM, Random Forest y Decision Tree; no obstante, el algoritmo estadístico ARIMA presenta un MAPE mayor al 5%, indicando problemas de predicción en los días laborales lunes 9 y martes 10 de marzo de 2020.



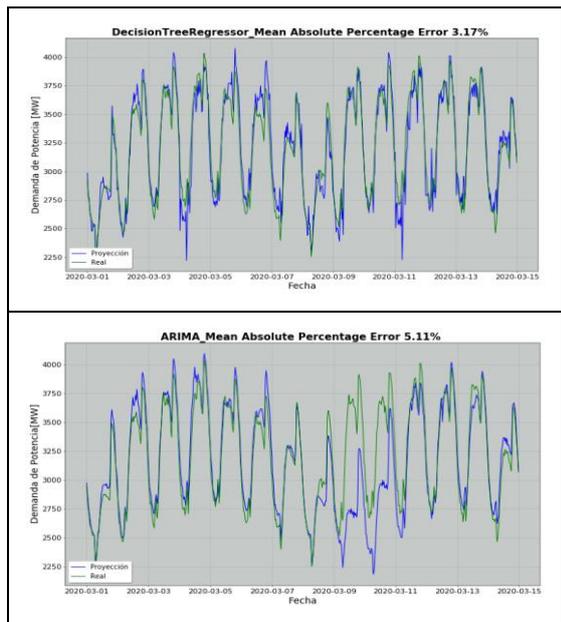


Figura 16: Temporada lluviosa Desde 2020-03-01 hasta 2020-03-14

En la Fig. 17 se presentan las proyecciones semanales (2 semanas) de la demanda horaria de potencia eléctrica con presencia de días feriados que corresponden al feriado nacional de carnaval en febrero de 2020.

Las curvas proyectadas por la herramienta computacional presentan un buen resultado al ajustarse a comportamientos de demanda críticos, como los que se presentan en los días feriados de carnaval lunes 24 y martes 25 de febrero de 2020. Sin embargo, el algoritmo de predicción ARIMA no logra un buen ajuste.

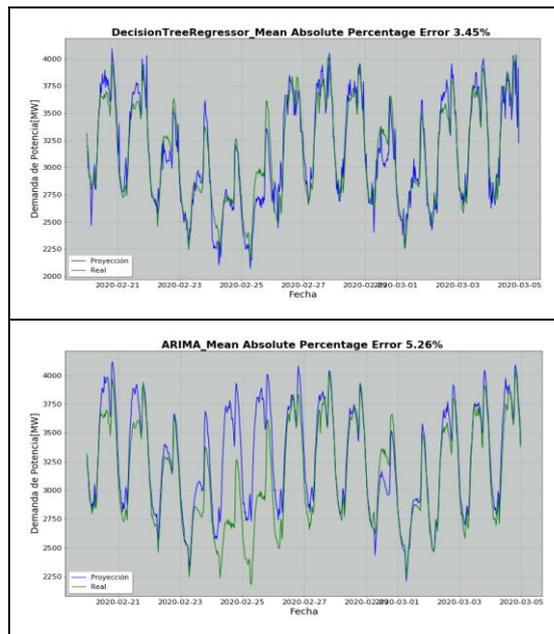
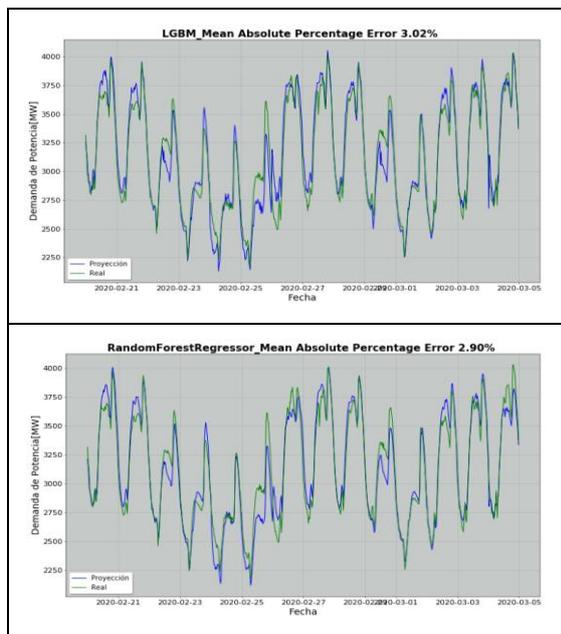


Figura 17: Semana con días feriados Desde 2020-02-20 hasta 2020-03-05

#### 4.2.4 Errores

En la Tabla 6 se presentan los porcentajes de error medio absoluto de los resultados de proyección.

Tabla 6: Errores de proyección

Temporada seca Desde 2019-07-15 hasta 2019-07-28				
Error	LGBM	RANDOM FOREST	DECISION TREE	ARIMA
MAPE [%]	2,08	2,02	2,54	2,07
Temporada lluviosa Desde 2020-03-01 hasta 2020-03-14				
Error	LGBM	RANDOM FOREST	DECISION TREE	ARIMA
MAPE [%]	2,62	2,73	3,17	5,11
Semana con presencia de días feriados Desde 2020-02-20 hasta 2020-03-05				
Error	LGBM	RANDOM FOREST	DECISION TREE	ARIMA
MAPE [%]	3,02	2,90	3,45	5,26

El error porcentual medio absoluto para los algoritmos predictivos Light Gradient Boosting Machine, Random Forest, y Decision Tree, se mantienen por debajo de un error del 4%; no obstante, el algoritmo ARIMA presenta un error porcentual medio absoluto de hasta el 5,26% en el escenario de proyección con presencia de días feriados.

#### 4.2.5 Tiempos de simulación

En la Tabla 7 se presenta como referencia los tiempos de simulación de las diferentes tareas realizadas por la herramienta computacional para las mediciones horarias de potencia eléctrica del SNI, con un total de 113 856 datos históricos.



Tabla 7: Tiempos de simulación herramienta computacional

Minería de datos	
Tarea	Tiempo [min]
Informe	0,6
Completar Series (672 Datos)	1,6
Hallar atípicos (59 Datos)	2,24
Mitigar Atípicos (59 Datos)	6,76
Indicadores	0,3
Proyección de demanda a corto plazo	
Tarea	Tiempo [min]
Hyperopt	3,27
Búsqueda de Cuadrícula	21
LGBM	0,17
Random Forest	0,54
Decision Tree	0,075
ARIMA	0,71

Se puede observar que la herramienta desarrollada no requiere grandes esfuerzos computacionales para realizar sus funciones. En el análisis de minería de datos, la tarea de mitigación de valores típicos es la que presenta un mayor tiempo de simulación debido a que la herramienta debe buscar cada medición específica en todo el conjunto de datos y reemplazarla por un valor pronosticado. Por otra parte, en el módulo de proyección de demanda la tarea que presenta un mayor tiempo de simulación es el autoajuste de parámetros  $p$ ,  $d$ ,  $q$  del algoritmo ARIMA, como se presentó en la sección 2.3.3 dicha función evalúa todas las combinaciones posibles de los parámetros dentro de rangos definidos, al contrario de la función hyperopt que busca solo evaluar las mejores combinaciones de hiperparámetros para los algoritmos de scikit-learn.

#### 4.2.5 Análisis de resultados

Los análisis realizados con la herramienta computacional permitieron evaluar el rendimiento de varios algoritmos predictivos en tareas de minería de datos y tareas de proyección de demanda de potencia eléctrica.

El análisis de minería de datos permitió identificar valores faltantes en las series de demanda de potencia eléctrica del SNI, los valores faltantes encontrados se presentaron en su mayoría en diferentes meses del año 2014, y en los días 29 de febrero de los años bisiestos 2016 y 2020, con un total de 672 datos faltantes, la imputación de dichos datos permitió generar series anuales históricas completas.

Se determinó la presencia de valores atípicos en las series, los valores atípicos se presentaron con mayor incidencia en los años 2016 y 2020, específicamente 11 comportamientos anómalos en el mes de abril de 2016 y 30 comportamientos anómalos en el mes de abril de 2020. Dichos valores fueron eliminados y reemplazados por valores más probables a un comportamiento normal de demanda de potencia eléctrica. Al realizar un análisis de los valores atípicos identificados se puede afirmar que, los comportamientos anómalos presentados el sábado 16 de abril de 2016 se debieron al terremoto que se presentó

en Manabí y que tuvo un gran impacto en todo el país, reduciendo la demanda de potencia con un decrecimiento anual mayor al 40%. Por otra parte, los valores atípicos identificados en el mes de abril de 2020 indican la disminución del consumo de potencia eléctrica debido al estado de emergencia en el que se declaró a todo el país debido a la pandemia de Coronavirus.

Las curvas de proyección a corto plazo presentaron en general un porcentaje de error medio absoluto menor al 4% en los algoritmos de aprendizaje automático Light Gradient Boosting Machine, Random Forest y Decision Tree, al contrario del algoritmo ARIMA que presenta un error MAPE de hasta 5,26% para los diferentes escenarios planteados, también los tiempos de simulación para los algoritmos de scikit-learn se presentaron por debajo del tiempo de simulación del algoritmo ARIMA, indicando menores esfuerzos computacionales. En contraste con los resultados de [2] y [3], se determina que los algoritmos de aprendizaje automático presentan una mejor interpretación del comportamiento no lineal de la demanda de potencia.

## 5. CONCLUSIONES Y RECOMENDACIONES

La aplicación del enfoque predictivo del aprendizaje automático permitió mitigar valores faltantes y valores atípicos que se encuentran inmersos en las series de datos. Los valores faltantes y valores atípicos han sido identificados en la literatura técnica como los causantes directos de errores en estudios de proyección. De esta manera, la herramienta computacional implementada brinda la posibilidad de realizar análisis de proyección de demanda eléctrica a corto plazo con un mayor grado de confiabilidad.

El presente estudio permitió corroborar el rendimiento y calidad de varios algoritmos de aprendizaje automático, así como el comportamiento de predicción del algoritmo estadístico ARIMA mediante la comparación de error en diferentes escenarios. De esta manera, se determinó que el algoritmo Random Forest obtuvo un mejor rendimiento de proyección. Dicho análisis comparativo mostró que los algoritmos de aprendizaje automático tienden a representar de mejor manera el comportamiento no lineal de factores que afectan el consumo de demanda eléctrica como el clima, o la época del año.

El procedimiento de ingeniería de características para caracterizar e identificar los diferentes comportamientos de la demanda de potencia eléctrica, determinó que los algoritmos de aprendizaje automático pueden asimilar de mejor manera los diferentes patrones del conjunto de datos analizado por la herramienta computacional. Además, el adecuado ajuste de los diferentes hiperparámetros de los algoritmos de proyección permitió construir modelos de aprendizaje automático que presentaron en general un buen desempeño para diferentes escenarios de proyección.

El problema de proyección de demanda descrito a detalle en este estudio, que define la proyección de la demanda de potencia eléctrica en un horizonte de dos semanas basándose en el consumo de potencia de días previos, determinó que, el análisis de minería de datos de los registros horarios de demanda de potencia en la semana previa al pronóstico es un requerimiento primordial para disminuir errores de proyección.

El ajuste adecuado de los diferentes hiperparámetros de los algoritmos de aprendizaje automático es un requerimiento muy importante para que el rendimiento de proyección sea el mejor posible. Es por esto, que se recomienda un estudio minucioso del aporte que realiza cada hiperparámetro en la construcción de un modelo de aprendizaje automático para ajustar algoritmos más confiables y eficientes.

### REFERENCIAS BIBLIOGRÁFICAS

[1] R. Medina and E. Montoya, "Estimación estadística de valores faltantes en series históricas de lluvia," trabajo de fin de máster, Universidad Tecnológica de Pereira, Pereira, 2008.

[2] S. Jurado, A. Nebot, F. Mugica, and N. Avellana, "Hybrid methodologies for electricity load forecasting: Entropy-Based Feature Selection with Machine Learning and Soft Computing Techniques," *ENERGY*, vol. 86, pp. 276–291, Jun. 2015.

[3] V. Hinojosa, "Pronóstico de Demanda de Corto Plazo en Sistemas de Suministro de Energía Eléctrica utilizando Inteligencia Artificial," Universidad Nacional de San Juan, Argentina, 2007.

[4] M. Khan and A. Abraham, "Short Term Load Forecasting Models in Czech Republic Using Soft Computing Paradigms," *International Journal of Knowledge-Based Intelligent Engineering Systems*, vol. 7, pp. 172–179, 2004.

[5] A. El khantach, M. Hamlich, and N. Belbounaguia, "Short-term load forecasting using machine learning and periodicity decomposition," *Energy*, vol. 7, no. 3, pp. 382–394, Jun. 2019.

[6] K. Berk, *Modeling and Forecasting Electricity Demand A Risk Management Perspective*. Germany: Springer, 2015.

[7] V. Hinojosa, "Pronóstico de Demanda de Corto Plazo en Sistemas de Suministro de Energía Eléctrica utilizando Inteligencia Artificial," Universidad Nacional de San Juan, Argentina, 2007.

[8] M. Jacob, C. Neves, and D. Vukadinović Greetham, "Short Term Load Forecasting," in *Forecasting and Assessing Risk of Individual Electricity Peaks. Mathematics of Planet Earth*, Springer, 2019, pp. 15–37.

[9] Y. Liu, *Python Machine Learning by Example*, 2nd ed. Birmingham: Packt, 2017.

[10] L. Igual and S. Seguí, *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications*. Switzerland: Springer, 2017.

[11] L. BREIMAN, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[12] Scikit Learn, "scikit-learn Machine Learning in Python." Jun. 28, 2020, Accessed: Jun. 28, 2020. [Online]. Available: <https://scikit-learn.org/stable/index.html>.

[13] Medium, "A Trip to Random Forest." Mar. 2018, Accessed: Jun. 28, 2020. [Online]. Available: <https://medium.com/greyatom/a-trip-to-random-forest-5c30d8250d6a>.

[14] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps: A practical Implementation Guide to Predictive Data Analytics Using Python*. Bangalore: Apress, 2017.

[15] K. Brent, J. Bergstra, and C. Eliasmith, "Hyperopt-Sklearn," in *Automated Machine Learning: Methods, Systems, Challenges*, Waterloo: Springer, 2019.

[16] A. O. Gallo, "Análisis predictivo para minería de datos y proyección de demanda a corto plazo de la demanda de potencia en el sistema eléctrico ecuatoriano", Escuela Politécnica Nacional, Quito, Ecuador, Oct. 2020.

[17] G. Bontempi, S. Ben Taieb, and Y. Le borgne, "Machine learning strategies for time series forecasting," Berlin Germany, Jul. 2012, vol. 138, pp. 62–77.



**Angel Gallo Cruz.-** Nació en Quito-Ecuador. Realizó sus estudios de tercer nivel en la carrera de Ingeniería Eléctrica de la Escuela Politécnica Nacional. Ha colaborado en estudios y elaboración de informes técnicos de minería de datos, proyección de demanda, y estabilidad de voltaje en la ex Agencia de Regulación y Control de Electricidad. Sus áreas de interés incluyen: Aprendizaje automático para análisis de datos, Operación de Sistemas Eléctricos de Potencia, y Sistemas de Control aplicado al SEP.



**Fabián Pérez Yauli.-** Nació en Ambato-Ecuador. Obtuvo el título de Ingeniero Eléctrico en la Escuela Politécnica Nacional, Quito-Ecuador en 2004. En 2012 obtuvo el grado de Doctor en Ingeniería Eléctrica en la Universidad Nacional de San Juan,

Argentina, con una beca otorgada por el Servicio Alemán de Intercambio Académico (DAAD). Actualmente, es profesor titular a tiempo completo en la Escuela Politécnica Nacional. Su rama de investigación incluye protecciones de sistemas de potencia y procesamiento de señales.



**Diego Salinas Herrera.-** Nació en Catamayo-Loja. En 2011, obtuvo su título de Ingeniero Eléctrico en la Escuela Politécnica Nacional. Tiene más de 12 años de experiencia en la realización de estudios especializados para la planificación, operación,

regulación y control del sector eléctrico ecuatoriano. Ha formado parte de equipos interinstitucionales para la Integración Eléctrica de la Región Andina. Actualmente, forma parte del equipo de profesionales de la Dirección de Estudios e Información del Sector Eléctrico de la Agencia de Regulación y Control de Energía y Recursos Naturales No Renovables.