

Predictive Recommendation Model for the Energy Dispatch of the Paute Hydropower Complex

Modelo Predictivo de Recomendación para el Despacho Energético del Complejo Hidroeléctrico Paute

R.M. Monteros¹

¹Corporación Eléctrica del Ecuador CELEC EP - CELEC SUR, Cuenca, Ecuador
E-mail: marcelo.monteros@celec.gob.ec

Abstract

This work proposes to make the most of the water resource used for the generation of electricity in Ecuador. Three models based on artificial intelligence have been made for the Mazar, Molino and Sopladora hydroelectric plants that belong to the Paute-Integral hydroelectric complex. For the implementation of the predictive recommendation algorithms, the behavior of the Mazar, Molino and Sopladora plants was first modeled, after which optimization was carried out to maximize electricity generation according to the capacity of the hydroelectric plants and hydrology. Finally, with the results obtained, it is observed that the maximization of electricity generation is achieved for the Mazar and Molino plants. Regarding the Sopladora plant, whose energy dispatch depends directly on the electricity generation of the Molino plant, the evaluation point remains to measure the impact produced by the optimization of the Molino plant.

Index terms— Hydroelectric dispatch, Energy dispatch optimization, predictive model.

Resumen

El presente trabajo propone aprovechar al máximo el recurso hídrico utilizado para la generación de energía eléctrica en el Ecuador. Se ha realizado tres modelos basados en inteligencia artificial para las centrales hidroeléctricas Mazar, Molino y Sopladora que pertenecen al complejo hidroeléctrico Paute-Integral. Para la implementación de los algoritmos predictivos de recomendación, primero se modeló el comportamiento de las centrales Mazar, Molino y Sopladora, posterior a lo cual se procedió a la optimización para maximizar la generación eléctrica acorde a la capacidad de las centrales hidroeléctricas y la hidrología. Finalmente, con los resultados obtenidos, se logra la maximización de la generación eléctrica para las centrales Mazar y Molino. Respecto a la central Sopladora, cuyo despacho energético depende directamente de la generación eléctrica de la central Molino, queda como punto de evaluación medir el impacto producido por la optimización de la central Molino.

Palabras clave— Despacho Hidroeléctrico, Optimización despacho energético, modelo predictivo.

Recibido: 18-10-2021, Aprobado tras revisión: 18-01-2022

Forma sugerida de citación: Monteros, M. (2022). "Modelo Predictivo de Recomendación para el Despacho Energético del Complejo Hidroeléctrico Paute". Revista Técnica "energía". No. 18, Issue II, Pp. 104-112

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

© 2022 Operador Nacional de Electricidad, CENACE

1. INTRODUCCIÓN

El propósito del presente estudio es diseñar un algoritmo predictivo de recomendación que permita predecir la producción de energía eléctrica de tres centrales hidroeléctricas Mazar, Molino y Sopladora, aprovechando el contexto de la capacidad de producción en ese preciso momento, la hidrología y la filosofía de operación.

Para lograrlo se establecen las siguientes fases:

- **Estado del Arte.** En esta fase se realiza el estudio e investigación de lo que se conoce acerca de la predicción de la demanda energética. Esto permite enfocar las acciones para la implementación del algoritmo de predicción de generación eléctrica.
- **Objetivos y Metodología.** En esta fase se describen los objetivos general y específico del presente estudio, así como la metodología aplicada para la construcción y evaluación del modelo predictivo de recomendación.
- **Contribución.** En esta fase se procede a construir tres modelos predictivos uno para cada una de las centrales Mazar, Molino y Sopladora. El primer objetivo es determinar el comportamiento normal de las tres centrales de generación en base al contexto de la hidrología y la filosofía de operación. Y el segundo objetivo es la construcción de los modelos predictivos que aprovechen la capacidad instalada y los recursos hídricos para recomendar la producción más acertada.
- **Descripción de resultados.** La descripción de resultados está dispuesta en dos etapas, la primera es un análisis acerca del impacto de la optimización en la operación de las centrales de generación. Este análisis se realiza comparando los datos actuales con los optimizados. La segunda etapa corresponde al análisis del impacto observable en el escenario de vertimientos lo que constituye el aporte significativo de este trabajo.

2. ESTADO DEL ARTE

El despacho energético depende de la demanda y de la oferta existente en un Sistema Eléctrico de Potencia (SEP). Sin embargo, el modelo predictivo de recomendación que se propone está enfocado únicamente a maximizar la producción de energía de las centrales de generación de menor costo de producción de energía eléctrica, que son las centrales Molino, Mazar y Sopladora. Esta decisión está basada en la obligación que tiene el CENACE de garantizar el

abastecimiento de energía al mínimo costo, preservando la eficiencia y transparencia [1].

El emplazamiento de las centrales de generación antes mencionadas se muestra en la Fig. 1, en la cual, el modelo predictivo de recomendación debe tener la capacidad de predecir la generación total del complejo Paute Integral. Para lograrlo debe considerar: la estación del año, los caudales de entrada, el nivel de los embalses, restricciones técnicas como equipos en mantenimiento, equipos disponibles, al igual que feriados largos y el clima.

Para diseñar el modelo predictivo de recomendación, se han investigado algunas alternativas; sin embargo, la mayoría de los algoritmos se centran en el pronóstico de la demanda de la energía eléctrica requerida a corto plazo. Ahora, el objetivo del presente trabajo radica en determinar la máxima cantidad de producción de energía eléctrica que se puede generar en base a caudales y niveles de embalse de las centrales de generación Molino, Mazar y Sopladora. Por tanto, se analizan los métodos basados en inteligencia artificial que han dado resultado en el campo de despacho energético.



Figura 1: Centrales hidroeléctricas de generación en estudio. Fuente: [2]

2.1. Redes Neuronales Artificiales

Las redes neuronales artificiales han sido usadas para el escenario tradicional, en que los SEP son diseñados y operados como una sola entidad, y se requiere un pronóstico de demanda de corto plazo, que permita garantizar que el suministro sea confiable. El uso de este método de aprendizaje automático ha reportado muchos éxitos en la predicción a corto plazo en ámbitos de STLF (Short Term Load Forecast), principalmente enfocado en el reconocimiento y clasificación de patrones [3].

Considerando que la predicción de la oferta de energía está basada en hechos observados, se identifican dos razones por las cuales se considera a las redes neuronales aptas para la tarea de predicción y estas son [3]: (1) Las redes neuronales artificiales tienen la capacidad de aproximar numéricamente cualquier función continua con una precisión arbitraria. (2) Las redes neuronales artificiales es un método donde los resultados dependen de los datos proporcionados; por lo

tanto, no se requiere postular modelos y estimar parámetros.

La arquitectura de red neuronal más utilizada para la demanda de corto plazo es la Perceptrón Multicapa y las metodologías que se han aplicado son las siguientes:

En un caso, se utilizaron datos de la demanda de energía de cada hora del día como series temporales distintas, correspondientes a los años 2011 y 2012; es decir, se tiene 24 series y, por tanto, se desarrollaron 24 redes neuronales. Con este método se ha logrado que la predicción de cada hora sea independiente. Adicional, se ha considerado que, al realizarlo de la forma expuesta, cada red neuronal resultante tiene menor cantidad de parámetros con relación a una sola red neuronal con 24 nodos de salida [3]. En el otro caso, se utilizaron datos de la serie de tiempo recolectadas cada 15 minutos del periodo que comprende desde el 2007 al 2016. Los datos que contiene son: día, hora, minutos y demanda [4].

Para el pre-procesamiento de datos, se ha establecido que para determinar el perfil de la demanda el factor más importante es el calendario, diferenciando los días laborables, de los fines de semana y días festivos, además del día de la semana y hora del día. Con estas consideraciones, se ha realizado la siguiente clasificación, basado en tres tipos de días que son:

- a) Días laborables.
- b) Fines de semana.
- c) Días Feriados.
- d) Otro factor importante es el clima; por lo cual, se utilizó la temperatura diaria máxima ponderada, obtenida de estaciones meteorológicas.
- e) Finalmente, con el objeto de caracterizar la estacionalidad anual de la demanda, cada mes del año es representado por la humedad relativa media mensual, velocidad del viento media mensual y precipitación media mensual [4] [3].

Los diseños de los trabajos revisados están basados en el Perceptrón Multicapa (MLP) feedForward full conected. Este diseño está conformado por una capa oculta cuya función de activación corresponde a la Tangente Hiperbólica Sigmoidea; mientras que, la función de activación de la única neurona de la capa de salida utiliza una función lineal.

En lo referente a la cantidad de neuronas, el trabajo [3] no indica la cantidad de neuronas en la capa oculta; sin embargo, indica que se han utilizado 13 entradas para predecir la demanda del día siguiente y determinar el despacho. Las entradas son:

- La demanda de la misma hora del día anterior y de las cinco horas antes de ese día.
- La demanda de la misma hora del mismo día de la semana anterior.
- La clasificación del tipo de día: laborable, fin de semana o feriado, que ocupa dos entradas.

- La temperatura máxima ponderada del día y las características del mes que son: humedad relativa, velocidad del viento y la precipitación media mensual ocupan otras cuatro entradas.

Otro trabajo [4] indica el uso de ocho entradas para predecir la demanda, una capa oculta con 27 neuronas, una capa de salida que corresponde a la demanda de energía eléctrica. Las entradas utilizadas son: a) temperatura, b) humedad, c) hora del día, d) día de la semana, e) una variable que indica si es un día festivo / fin de semana, f) carga media del día anterior, g) carga de la misma hora del día anterior, y h) Carga de la misma hora y el mismo día de la semana anterior.

Finalmente, para el entrenamiento de las redes neuronales artificiales se ha utilizado el método backpropagation implementado en MatLab con el algoritmo Levenberg Marquardt. [4].

2.2. Algoritmos Random Forest

El pronóstico de la demanda a corto plazo hace referencia a la predicción con una hora, un día, o una semana de antelación de la demanda y su precisión está directamente relacionada con la seguridad, la estabilidad y la economía en el funcionamiento del SEP [5].

El algoritmo Random Forest tiene la ventaja de ser tolerante al ruido, resistente al sobreajuste y sólo se necesita examinar pocos parámetros en comparación con otros métodos [5].

Con lo descrito anteriormente, se analiza el trabajo realizado por Huang et al. [5], en el que se propone un método predictivo STLF de estimación de la demanda energética basado en random forest. Para lo cual, se realiza la siguiente metodología:

Los datos utilizados corresponden a la carga histórica del año 2012 de una ciudad al noreste de China. El conjunto total de datos contiene 366 días, el 9% (33 días) se utilizan para el conjunto de pruebas y el 91% (333 días) con conjunto de entrenamiento. El muestreo es de cada hora.

Con el objeto de aumentar la fiabilidad del experimento, el conjunto de prueba se distribuye al azar y en cuatro trimestres.

Basados en los datos anteriores se establecen dos tipos de predicciones: con una hora de anticipación y con un día de anticipación; para lo cual, se utilizan las siguientes características:

- El valor de la carga historia de tiempo (t-i) (i=1,2,3...;240)
- Día laborable (1->Sí,2-> No)
- Día actual.
- El momento (de 0 a 23, 24 horas del día)

En este estudio se ignoran las covariables como la temperatura.



Un punto importante en esta metodología es que se realiza una etapa de preselección para obtener las variables más importantes para la clase objetivo. Únicamente las variables más importantes antes obtenidas sirven para entrenar el modelo de Random Forest.

3. OBJETIVOS Y METODOLOGÍA

3.1. Objetivo general

Crear un modelo predictivo de recomendación para el proceso del pre-despacho de energía que permita maximizar la producción de energía eléctrica en las centrales Mazar, Molino y Sopladora en base a los generadores en operación, los caudales de entrada de agua y cuotas de los embalses de Mazar y Amaluza.

3.2. Objetivos específicos

- Identificar los algoritmos predictivos utilizados para despacho energético a corto plazo y sus ventajas.
- Identificar las variables predictivas que se pueden considerar para la creación del modelo predictivo.
- Analizar la información disponible acerca del despacho energético en las centrales de generación Mazar, Molino y Sopladora.
- Crear un modelo predictivo para cada central de generación con las dependencias existentes por tratarse de un complejo hidroeléctrico en cascada.
- Pruebas del algoritmo con días puntuales en los meses de enero, febrero y marzo del 2020.

3.3. Metodología de trabajo

Con el objeto de alcanzar el objetivo general y los objetivos específicos, la metodología del trabajo conlleva los siguientes pasos:

1. Obtención de la información referente al despacho energético de las centrales de generación Mazar, Molino y Sopladora por separado.
2. Análisis de datos de cada una de las centrales de generación y determinación del comportamiento.
3. Definición del conjunto de información que se utilizará para el aprendizaje y evaluación del modelo predictivo de recomendación del despacho energético. Para este proceso se utiliza información de 5 años para las Centrales Mazar y Molino (46.297 registros), y para la Central Sopladora se utiliza información de tres años (28.752 registros).
4. Entrenamiento del modelo predictivo de recomendación utilizando el algoritmo Random Forest. Para el entrenamiento se utilizará el 80%

de los datos y el 20% restante para la evaluación del modelo.

5. Evaluación del modelo predictivo de recomendación con métricas de regresión error cuadrático medio (RMSE) y el coeficiente de determinación (R^2 score).
6. Verificación de las variables predictivas más importantes con Random Forest.
7. Ajuste de datos de entrenamiento hasta conseguir un error adecuado para la predicción del modelo.
8. Prueba del modelo predictivo de recomendación con la información real.

4. CONTRIBUCIÓN

Con el objeto de corroborar los objetivos planteados, se realiza tres modelos predictivos, uno para cada central de generación Mazar, Molino y Sopladora. La tecnología que se utiliza para la construcción de los modelos antes descritos es Random Forest. Las principales ventajas y las razones por las que se escogió esta tecnología son las siguientes:

- Es un modelo de propósito general que sirve tanto para clasificación como para regresión. Esto quiere decir que funciona correctamente para el presente caso de estudio.
- Este modelo es apto para trabajar con grandes volúmenes de datos, dado que tiene la capacidad de combinar los principios de bagging (algoritmos ensamblados) con selección de la variable aleatoria, añadiendo así diversidad y reduciendo el coste computacional. Esta característica es de gran utilidad ya que la información es horaria por 5 años y representa una gran cantidad de datos.

4.1. Organización del piloto

Para la organización del piloto se procedió a construir tres modelos predictivos uno para cada una de las centrales: Mazar, Molino y Sopladora. El objetivo es determinar el comportamiento óptimo de las tres centrales de generación basado en la filosofía de operación, las condiciones ambientales y la capacidad de producción disponible al momento.

Para el caso de la Central Sopladora, por ser una central que no posee embalse, la generación de energía depende directa y totalmente de la generación de la Central Molino.

4.2. Fase 1. Comportamiento de las tres centrales

Para determinar el comportamiento de las tres centrales se realiza el modelo predictivo de las tres centrales de generación y se evalúan los resultados. Los datos por considerar en los tres modelos corresponden a:

- El caudal de entrada, medido en m^3/s .
- La cota de los embalses para el caso de las centrales Mazar y Molino, para el caso de la

central Sopladora es el nivel de la cámara de interconexión. Estos datos son medidos en metros sobre el nivel del mar (m.s.n.m.).

- Las unidades de generación disponibles, medido en unidades. Cada central de generación cuenta con las siguientes unidades de generación:
 - La central Mazar: 2 unidades de generación.
 - La central Molino: 10 unidades de generación.
 - La central Sopladora: 3 unidades de generación.

Respecto al central Molino, se incluye la información de la generación de la Central Mazar y para el caso de la Central Sopladora se incluye la generación de la Central Molino.

4.2.1 Modelo predictivo para la central Mazar

En este acápite se explica el proceso involucrado en el entrenamiento del modelo predictivo de recomendación para la central Mazar.

Obtención de datos. Los datos fueron obtenidos del sistema SARDOM (sistema desarrollado internamente en CELEC SUR). Este sistema es utilizado para recopilar la información de la generación de las centrales de generación.

Los datos obtenidos para el análisis del comportamiento de la central Mazar se basaron en los campos expuestos en la Tabla 1.

Tabla 1: Campos Central Mazar

| Campos | Descripción |
|-------------------|---|
| Hora día | Constituyen las 24 horas del día |
| Mes | Constituyen los 12 meses del año |
| Día semana | Determina el día de la semana del lunes al domingo (1 al 7) |
| Energía real | Constituye el target (objetivo), y es la cantidad de energía producida en una hora, día y mes específico. |
| Cota | Constituye el nivel del agua en el embalse. |
| Caudal de entrada | Medido en metros cúbicos por segundo, constituye la cantidad de agua que ingresa al embalse. |
| Unidades en línea | Constituye la cantidad de máquinas operando en la hora específica. En este caso son 2 unidades de generación. |

La información obtenida acorde a la Tabla 1, está basada en la historia de 5 años.

Análisis de datos: La dispersión de los datos se puede revisar en la Tabla 2.

En la Tabla 2 se puede observar que el caudal más elevado se encuentra en el último cuartil y sus valores superan los 109 m³/s. De igual manera, respecto a la cota el último cuartil supera los 2.153 metros sobre el nivel del mar, lo cual coincide con la presencia de varios vertimientos.

Tabla 2: Dispersión de datos cota y caudal

| | Nivel Mazar m.s.n.m. | Caudal entrada m ³ /s |
|-----|----------------------|----------------------------------|
| Min | 2.121,39 | 0,00 |
| 25% | 2.143,23 | 37,83 |
| 50% | 2.150,22 | 65,59 |
| 75% | 2.153,14 | 109,28 |
| Max | 2.156,20 | 1.174,52 |

Los datos en percentiles de energía se visualizan en la Tabla 3.

Tabla 3: Distribución percentil

| | P MW |
|-----|--------|
| Min | 0,00 |
| 25% | 0,00 |
| 50% | 84,82 |
| 75% | 154,65 |
| Max | 173,84 |

Considerando que cada unidad de generación tiene la capacidad máxima de 85 MW, en la Tabla 3 se puede corroborar que el 25% del tiempo no se produce energía eléctrica, del 25% al 50% se genera energía eléctrica equivalente a una máquina y sólo del 50% del tiempo en adelante se utilizan las dos máquinas para generar energía.

Definición de datos de entrenamiento: El conjunto de datos disponible para el entrenamiento consiste en 5 años desde el 2015 hasta el 2019, en total 46.297 registros. De estos registros se obtienen el 20% para la evaluación del modelo y se utilizan los datos correspondientes a los tres primeros meses posteriores para la evaluación del modelo como test en vivo, esto es: enero, febrero y marzo del 2020.

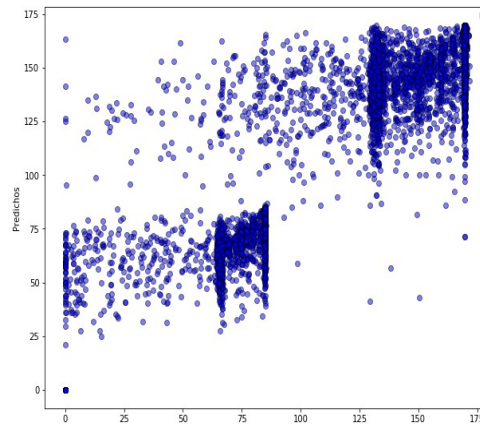


Figura 2: Evaluación del Modelo Predictivo
Fuente: Elaborado por el autor

Entrenamiento del modelo predictivo: El entrenamiento del modelo predictivo se realiza con 250 árboles, utilizando el 80% de los datos para el entrenamiento y el 20% para la evaluación.

Evaluación del modelo: Los resultados de la evaluación se observan en la Fig. 2. Las métricas utilizadas para evaluar el modelo de predicción se observan en la Tabla 4:



Tabla 4: Métricas de evaluación

| Modelo | RMSE | RMSLE | R ² score |
|-----------------------|-------|-------|----------------------|
| RandomForestRegressor | 17,97 | 0,42 | 0,92 |

La Tabla 4 indica que el error en la predicción es de 17,97; sin embargo, se tiene una precisión del 92% en la predicción de los datos de evaluación. El valor de RMSLE de 0,42 indica que no existe sobreajuste en la predicción, lo que implica que el modelo tiene una capacidad aceptable para generalizar.

4.2.2 Modelo predictivo para la central Molino

En este punto se explica el proceso involucrado en el entrenamiento del modelo predictivo de recomendación para la central Molino.

Obtención de datos: Los datos también fueron obtenidos del sistema SARDOM. Los datos obtenidos para el análisis del comportamiento de la central Molino se basaron en los campos: hora día, mes, día semana, energía real, cota, caudal entrada, unidades en línea, como se puede observar en la Tabla 5.

Tabla 5: Campos Central Molino

| Campos | Descripción |
|---------------------|--|
| Hora día | Constituyen las 24 horas del día |
| Mes | Constituyen los 12 meses del año |
| Día semana | Determina el día de la semana del lunes al domingo (1 al 7) |
| Energía real Molino | Constituye el target (objetivo), y es la cantidad de energía producida en una hora, día y mes específico. |
| Cota | Constituye el nivel del agua en el embalse Amalusa. |
| Caudal entrada | Medido en metros cúbicos por segundo, constituye la cantidad de agua que ingresa al embalse. |
| Unidades en línea | Constituye la cantidad de máquinas operando en la hora específica. En este caso son 10 unidades de generación. |
| Energía real Mazar | Constituye la energía que produjo la central Mazar |

La información obtenida acorde a la Tabla 5, está basada en la historia de 5 años, recopilada en forma horaria.

Análisis de datos: La dispersión de los datos se puede revisar en la Tabla 6, en la cual se puede observar que el caudal más elevado se encuentra en el último cuartil y sus valores superan los 168 m³/s. De igual manera, respecto a la cota el último cuartil supera los 1.990 metros sobre el nivel del mar, lo cual coincide con la presencia de varios vertimientos.

Tabla 6: Dispersión de datos cota y caudal Molino

| Campos | Cota m.s.n.m. | Caudal m ³ /s | Energía Mazar MWh |
|--------|---------------|--------------------------|-------------------|
| Min | 1.971,65 | 0 | 0 |
| 25% | 1.983,33 | 67,77 | 0 |
| 50% | 1.985,92 | 125,40 | 84,82 |
| 75% | 1.990,34 | 168,20 | 154,65 |
| Max | 1.992,06 | 1.250,40 | 173,84 |

Distribución de datos de la variable target: La distribución de datos de la variable target durante los cinco años analizados mantiene la forma reflejada en la Tabla 7.

Tabla 7: Distribución de datos target Molino

| Campos | Energía Mw |
|--------|------------|
| Min | 0 |
| 25% | 436,81 |
| 50% | 604,01 |
| 75% | 778,71 |
| Max | 1.098,28 |

En la Tabla 7 se puede observar el comportamiento de la generación de energía de la Central Molino. Este comportamiento refleja que el 50% del tiempo de producción de energía eléctrica representa valores superiores a 604 MWh de un total de 1.100 MWh.

Definición de datos de entrenamiento: El conjunto de datos disponible para el entrenamiento consiste en 5 años desde el 2015 hasta el 2020, en total 46.297 registros. De estos registros se obtiene el 20% para la evaluación del modelo y se utilizan los datos correspondientes a los tres primeros meses posteriores para la evaluación del modelo como test en vivo, esto es: enero, febrero y marzo del 2020.

Entrenamiento del modelo predictivo: El entrenamiento del modelo predictivo se realiza con 550 árboles, utilizando el 80% de los datos para el entrenamiento y el 20% para la evaluación.

Evaluación del modelo: Los resultados de la evaluación se observan en la Fig. 3.

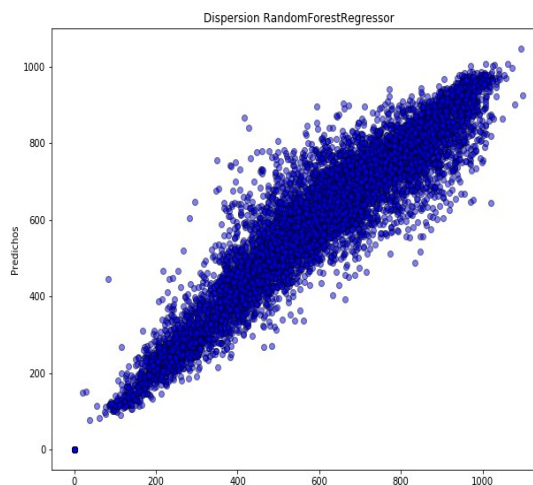


Figura 3: Evaluación del Modelo Predictivo
Fuente: Elaborado por el autor

Las métricas utilizadas para evaluar el modelo de predicción se observan en la Tabla 8:

Tabla 8: Métricas de evaluación.

| Modelo | RMSE | RMSLE | R ² score |
|-----------------------|-------|-------|----------------------|
| RandomForestRegressor | 75,28 | 0,14 | 0,89 |



La Tabla 8 indica que el error en la predicción es de 75,28, el cual compara el valor predicho respecto al valor conocido; sin embargo, se tiene una precisión del 89% en la predicción de los datos de evaluación. Por tanto, el valor RMSE se le considera tolerable dentro de los parámetros indicados. El valor de RMSLE de 0,14 nos indica que no existe sobreajuste en la predicción, lo que implica que el modelo tiene una capacidad aceptable para generalizar.

4.2.3 Modelo predictivo para la central Sopladora

En este acápite se explica el proceso involucrado en el entrenamiento del modelo predictivo de recomendación para la central Sopladora.

Obtención de datos: Los datos obtenidos para el análisis del comportamiento de la central Sopladora, obtenidos del sistema SARDOM, se basaron en los campos: hora, día, mes, día de la semana, energía real, cota, caudal entrada, y unidades en línea; expuestos anteriormente en Mazar, adicionando los campos expuestos en la Tabla 9.

Tabla 9: Campos Central Sopladora

| Campos | Descripción |
|--------------------------------|---|
| Hora día | Constituyen las 24 horas del día |
| Mes | Constituyen los 12 meses del año |
| Día semana | Determina el día de la semana del lunes al domingo (1 al 7) |
| Energía real Sopladora | Constituye el target (objetivo), y es la cantidad de energía producida en una hora, día y mes específico. |
| Cota | Constituye el nivel del agua en el embalse Amalusa. |
| Caudal entrada | Medido en metros cúbicos por segundo, constituye la cantidad de agua que ingresa al embalse. |
| Unidades en línea | Constituye la cantidad de máquinas operando en la hora específica. En este caso son 3 unidades de generación. |
| Energía real Molino | Constituye la energía que produjo la central Molino |
| Cámara Interconexión Sopladora | Constituye una cámara que almacena el agua temporalmente. |

La información obtenida acorde a la Tabla 9, está basada en la historia de 3 años, recopilada en forma horaria.

Análisis de datos: La dispersión de los datos se puede revisar en la Tabla 10, en la que se puede observar que el caudal más elevado se encuentra en el último cuartil y sus valores superan los 123,17 m³/segundos. De igual manera, respecto al nivel de la cámara de interconexión en el último cuartil supera los 1.316,42 metros sobre el nivel del mar.

Tabla 10: Dispersión de datos cota y caudal.

| | Cámara Interconexión m.s.n.m | Caudal m ³ /s | Energía MolinoM Wh |
|-----|------------------------------|--------------------------|--------------------|
| Min | 1.311,99 | 0 | 0 |
| 25% | 1.315,79 | 74,86 | 444,63 |
| 50% | 1.316,07 | 97,76 | 573,23 |
| 75% | 1.316,42 | 123,17 | 719,94 |
| Max | 1.318,00 | 185,59 | 1.098,28 |

Distribución de datos de la variable target: La distribución de datos de la variable target durante los tres años analizados mantiene la forma reflejada en la Tabla 11, en la que se puede observar el comportamiento de la generación de energía de la Central Sopladora. Este comportamiento refleja que el 50% del tiempo de producción de energía eléctrica representa valores superiores a 260 MWh de un total de posible de 487 MWh.

Definición de datos de entrenamiento: El conjunto de datos disponible para el entrenamiento consiste en 3 años desde el 2017 hasta el 2020, en total 28.752 registros. De estos registros se obtiene el 20% para la evaluación del modelo y se utilizan los datos correspondientes a los tres primeros meses posteriores para la evaluación del modelo como test en vivo, esto es: enero, febrero y marzo del 2020. El piloto se lo realiza con datos del mes de abril de 2020.

Tabla 11: Distribución de datos target Central Sopladora

| Energía Mw | |
|------------|--------|
| Count | 28.752 |
| Mean | 260,56 |
| Std | 96,28 |
| Min | 0 |
| 25% | 204,62 |
| 50% | 260,41 |
| 75% | 324,30 |
| Max | 484,50 |

Entrenamiento del modelo predictivo: El entrenamiento del modelo predictivo se realiza con 400 árboles, utilizando el 80% de los datos para el entrenamiento y el 20% para la evaluación.

Evaluación del modelo: Los resultados de la evaluación se observan en la Fig. 4.

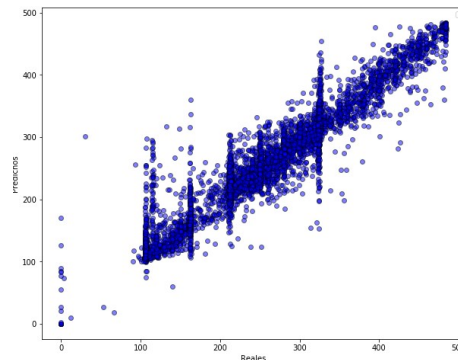


Figura 4: Evaluación del Modelo Predictivo Central Sopladora
Fuente: Elaborado por el autor



Las métricas utilizadas para evaluar el modelo de predicción se observan en la Tabla 12:

Tabla 12: Métricas de evaluación

| Modelo | RMSE | RMSLE | R ² score |
|-----------------------|-------|-------|----------------------|
| RandomForestRegressor | 28,07 | 0,22 | 0,91 |

La Tabla 12 indica que el error en la predicción es de 28,07; sin embargo, se tiene una precisión del 91% en la predicción de los datos de evaluación. El valor de RMSLE de 0,22 nos indica que no existe sobreajuste en la predicción, lo que implica que el modelo tiene una capacidad aceptable para generalizar.

4.3. Fase 2. Optimización de los modelos predictivos de recomendación

Para la realización de la optimización, se determina, tanto para la Central Mazar como para la Central Molino un límite de cota que representa la presencia de vertimiento inminente. En caso de llegar a dicho límite, se debe proceder con el despacho de la energía máxima, o en su defecto, con el despacho del 90% del volumen de agua ingresado en el embalse el día anterior, esto acorde al siguiente artículo [6].

Acorde a lo indicado, se determina que el límite de cota que indica vertimiento inminente para el embalse de la Central Mazar es de: 2.151,5 m.s.n.m. y para la cota del embalse de la Central Molino es de: 1.989 m.s.n.m. [6].

5. RESULTADOS

La descripción de resultados está dispuesta en una primera etapa, que consiste en un análisis acerca del impacto de la optimización en la operación de las centrales de generación. Este análisis se realiza comparando los datos actuales con los optimizados.

5.1. Primera etapa

Par el caso de la Central Mazar, la comparativa de los dos escenarios actual y optimizado se puede observar en las Fig. 5 y 6 Escenarios Central Mazar.

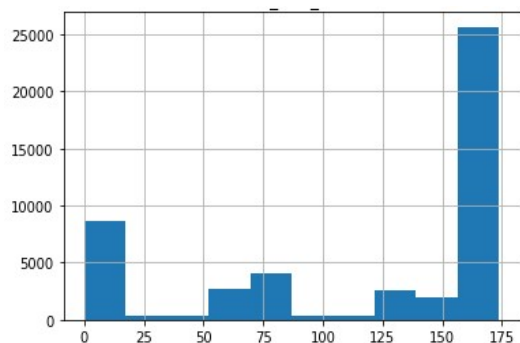


Figura 5: Escenario Optimizado Mazar. Fuente: Realizado por el autor

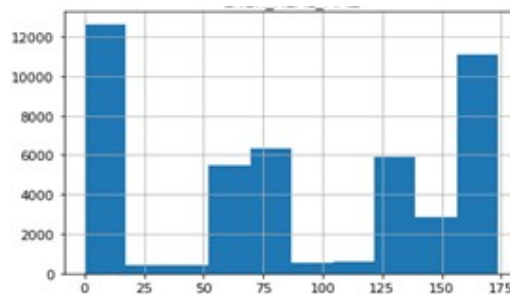


Figura 6: Escenario Actual Mazar. Fuente: Realizado por el autor

Par el caso de la Central Molino, la comparativa de los dos escenarios actual y optimizado se puede observar en las Figs. 7 y 8 Escenarios Central Molino.

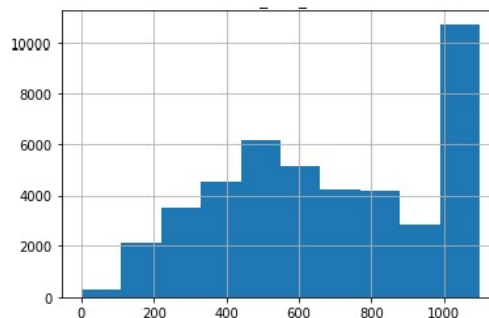


Figura 7: Escenario Optimizado Molino. Fuente: Realizado por el autor

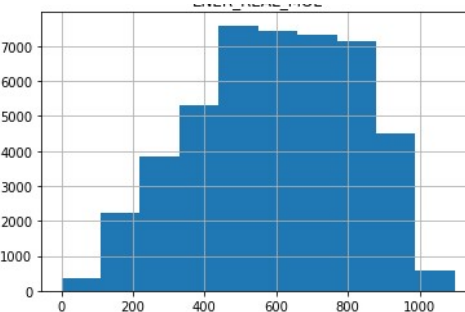


Figura 8: Escenario actual Molino. Fuente: Realizado por el autor

De este análisis se concluye que podría existir un incremento de generación de energía aplicando el modelo predictivo de recomendación con una seguridad del 94% para el caso de la central Molino y 96% para el caso de la Central Mazar.

Para el caso de la central Sopladora, depende directamente de la producción de energía de la Central Molino y por ende de los niveles de la cámara de interconexión; por tanto, no se modela una optimización para esta central, simplemente se modela el estado actual.

6. CONCLUSIONES Y RECOMENDACIONES

Durante la ejecución del piloto se realizó una comparación entre los valores predichos, los valores reales y los valores estimados por el organismo



encargado de regular el despacho energético del Ecuador (CENACE). Cabe recalcar que los valores predichos del algoritmo de recomendación cumplen con la optimización esperada y denotan un incremento en la generación de energía durante el periodo de invierno.

Para el caso de la central Mazar, cuando la cota llega al valor de vertimiento inminente que corresponde a 2.151,5 m.s.n.m. el modelo predictivo recomienda el valor máximo de generación, esto es potencias cercanas a los 170 MW.

Así mismo, cuando no existen unidades disponibles para la generación, el valor predicho es 0 y cuando la cota está llegando a su valor mínimo de operación 2.105 m.s.n.m y tiene poco caudal de entrada al embalse, y existe una máquina disponible, el algoritmo respeta la generación mínima factible de 65 MWh acorde a lo esperado.

Para el caso de la Central Molino, cuando la cota llega al valor de vertimiento inminente que corresponde a 1.989 m.s.n.m. el modelo predictivo recomienda el valor máximo de generación; esto es, el máximo valor de generación 1.100 MWh menos el 5% de la potencia como reserva rodante, requerido por AGC. Así mismo, cuando no existen unidades disponibles para la generación, el valor predicho es 0 y cuando la cota del embalse está llegando a su valor mínimo de operación 1.975 m.s.n.m y tiene poco caudal de entrada, y existe una máquina disponible, el algoritmo respeta la generación mínima factible de 25 MWh acorde a lo esperado.

Para la central Sopladora no se generó ninguna optimización. Para el test, se consideraron cinco escenarios reales en los que se probó el algoritmo planteado vs el algoritmo del CENACE. Los datos predichos del algoritmo planteado en tres escenarios tuvieron menor diferencia respecto al despacho real en comparación con la predicción del algoritmo del CENACE.

Finalmente, es importante recalcar que los factores determinantes para la predicción adecuada son la cota, el caudal de entrada y las unidades de generación.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Asamblea Nacional de la República del Ecuador, LEY ORGANICA DEL SERVICIO PÚBLICO DE ENERGÍA ELECTRICA, Quito, Pichincha: LEY 0 REGISTRO OFICIAL SUPLEMENTO 418, 2015.
- [2] CELEC, «www.celec.gob.ec.» 31 12 2015. [En línea]. Available: www.celec.gob.ec.
- [3] O. Barboza, «Automatización de previsión de demanda horaria de potencia,» Revista Científica de la UCSA, pp. 4-14, 2014.
- [4] I. F. Sinaluisa Lozano, A. F. Morocho Caiza y C. Marquez Zurita, «Predicción de demanda de

energía eléctrica mediante redes neuronales artificiales,» Risti, pp. 505-519, 2019.

- [5] N. Huang, L. Guobo y X. Dianguo, «A Permutation Importance-Based Feature Selection Method for Short-Term Electricity Load Forecasting Using Random Forest,» Energies, 2016.
- [6] J. Zalamea, «Despacho Hidrotérmico de Mediano Plazo aplicado al Complejo Hidroeléctrico Paute Integral,» Revista Técnica “energía”, n° 18, pp. 95-105, 2021.
- [7] G. T. Doran, «There's a S.M.A.R.T. way to write management's goals and objectives,» Management Review (AMA FORUM), vol. 70, pp. 35-36, 1981.
- [8] CONELEC, Estudio y Gestion de la Demanda Electrica, Quito, Provincia, 2013.
- [9] ARCONEL, Estadística Anual y Multianual del Sector Eléctrico Ecuatoriano, Quito, Pichincha, 2018.
- [10] CELEC EP, «Plan Estratégico 2017-2021,» Cuenca, 2019.
- [11] G. Argüello, INFORME OPERATIVO ANUAL, 2019.



Rubén Marcelo Monteros Guerrero.- Nació en Quito en 1977. Recibió su título de Ingeniero en sistemas e Informática de la Escuela Politécnica del Ejército en 2000; y en el 2008 su título de MBA en la Escuela Politécnica de Ejército (ESPE), Master en Inteligencia Artificial en el 2020. Sus campos de investigación están relacionados a la implementación de sistemas basados en modelos predictivos para la optimización del despacho energético y sistemas SCADA para la operación de las centrales de generación eléctrica.