

**DEMANDA DE LA ENERGÍA ELÉCTRICA EN COLOMBIA:
UN MODELO DE PRONÓSTICO**

**JONATHAN GUTIÉRREZ ROMERO
JULIO CESAR MORENO CAÑON**

Taller de aplicación económica, especialización en economía para ingenieros

Director: ALEX SMITH ARAQUE

**ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO
FACULTAD DE ECONOMÍA
ESPECIALIZACIÓN EN ECONOMÍA PARA INGENIEROS
BOGOTÁ
2019**

CONTENIDO

TABLAS.....	4
INTRODUCCIÓN	5
OBJETIVOS	6
MERCADO ELÉCTRICO COLOMBIANO	7
Tarifa de Energía.....	10
Modelo Actual De Pronóstico De Demanda.....	13
ANÁLISIS DE VARIABLES QUE INTERVIENEN EN LA DEMANDA DE ENERGÍA ELECTRICA.....	18
Histórico de demanda de energía eléctrica.....	18
Temperatura promedio en la región.....	21
Precipitaciones (lluvia) en la región.....	23
Producto Interno Bruto (PIB) en la región.....	25
MODELO MACHINE LEARNING PARA LA PREDICCIÓN DE LA DEMANDA.....	26
Definición de variables y periodos de la data	26
<i>Histórico de demanda energética</i>	27
<i>Temperatura promedio en la región</i>	29
<i>Precipitaciones totales en la región</i>	29
<i>Producto Interno Bruto (PIB)</i>	29
Desarrollo del modelo.....	30
<i>Herramienta de programación</i>	30
<i>Creación del DataFrame con las variables de entrada y salida</i>	30
<i>Separación (Train – Test) y normalización de los datos</i>	31
<i>Compilación del modelo y obtención de las predicciones discretizadas por hora</i>	31
RESULTADOS DEL MODELO	33
CONCLUSIONES.....	35
BIBLIOGRAFÍA	37

ILUSTRACIONES

Ilustración 1 Diagrama unifilar Compañía Energética de Occidente.....	15
Ilustración 2 Demanda de energía eléctrica en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores	18
Ilustración 3 Demanda de energía eléctrica en 2017 en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores	19
Ilustración 4 Demanda de energía eléctrica en los meses de Junio y Julio de 2017 en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores	19
Ilustración 5 Demanda de energía eléctrica en la primer semana del mes de mayo de 2017 en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores	20
Ilustración 6 Demanda de energía eléctrica típica para domingos y festivos en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores	20
Ilustración 7 Demanda de energía eléctrica típica para los sábados y sábados antes de un lunes feriado en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores	21
Ilustración 8 Temperatura promedio en estaciones meteorológicas de la región. Fuente: Corporación autónoma regional (CAR). Elaboración: Autores	22
Ilustración 9 Temperatura promedio de la región. Fuente: Corporación autónoma regional (CAR). Elaboración: Autores	22
Ilustración 10 Demanda energética y temperatura promedio mensual para una región del país. Elaboración: Autores	23
Ilustración 11 Precipitaciones totales en estaciones meteorológicas de la región. Fuente: Corporación autónoma regional (CAR). Elaboración: Autores	24
Ilustración 12 Comparación de los valores de precipitaciones totales mensuales y demanda energética mensual. Elaboración: Autores	24
Ilustración 13 Demanda energética total anual y PIB de una región central del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país y DANE Elaboración: Autores	25
Ilustración 14 Correlación entre los promedios mensuales de demanda energética, temperatura promedio y precipitaciones en la región.....	29
Ilustración 15 Correlación entre la demanda energética total por año y el PIB anual de la región.	29
Ilustración 16 Características generales de los datos cargados en COLAB	30
Ilustración 17 Características estadísticas de la demanda eléctrica horaria de 2012 a 2017.	31
Ilustración 18 Separación de datos de prueba y entrenamiento y normalización de los datos.	31
Ilustración 19 Compilación e iteración del modelo de entrenamiento y predicción de la demanda energética.....	32
Ilustración 20 Demanda real y predicción de energía para el día 06 de Abril de 2018.....	33
Ilustración 21 Desviación por tramo de la predicción de demanda para 763 registros de prueba.....	33
Ilustración 22 Demanda real y predicción de energía para el día 04 de Julio de 2014.	34

TABLAS

Tabla 1 Fronteras de importación Compañía Energética de Occidente Fuente: Compañía Energética de Occidente.....	14
Tabla 2 Fronteras de importación Compañía Energética de Occidente Fuente: Compañía Energética de Occidente.....	14
Tabla 3 Resumen cálculo de desviación pronóstico-propuesta XM datos 2018. Fuente: Autores.....	16
Tabla 4 Cálculo penalizaciones por desviaciones pronóstico (datos 2018) Fuente: autores.....	17
Tabla 5 Características de las cuatro variables a evaluar en el desarrollo del modelo.....	27
Tabla 6 Detalle variables del histórico de demanda a incluir en el modelo.....	28
Tabla 7 Detalle variables de temperatura promedio en la región a incluir en el modelo.....	29
Tabla 8 Detalle variables del PIB en la región a incluir en el modelo.....	30
Tabla 9 Calculo de desviaciones del pronóstico de demanda según el modelo propuesto.....	34

INTRODUCCIÓN

La ley de servicios públicos domiciliarios, (Ministerio de Minas y Energía, 1994) , estipula las bases legales para la negociación directa y la creación de contratos no estandarizados en el mercado de corto plazo. (Amador & Ramírez, 2011). La Ley promueve la competencia al abrir la participación de los agentes privados y la eliminación del monopolio del Estado en la prestación de los servicios públicos (Amador & Ramírez, 2011). La Ley desarrolla un modelo tarifario y un sistema de subsidios; por otro lado, se separaron los roles de vigilancia y control de manera que el desarrollo del sector eléctrico en Colombia está incentivado al crecimiento y a la modernización de la infraestructura en busca de mejorar el servicio de energía eléctrica en el país.

En Colombia el mercado eléctrico opera libremente de acuerdo con las condiciones de oferta y demanda, al crear competencia los agentes del mercado públicos y privados y los comercializadores celebran contratos de energía eléctrica con los generadores, los precios del kilovatio hora, kWh, se establecen en común acuerdo con las partes que intervienen en dicho mercado.

La administración y la operación del mercado se realiza por medio de XM el cual tiene a su cargo las funciones de Centro Nacional de Despacho -CND-, Administrador del Sistema de Intercambios Comerciales -ASIC- y Liquidador y Administrador de Cuentas de cargos por Uso de las Redes del SIN –LAC.

Ahora bien, en virtud de la asignación de los recursos de generación eléctrica y las restricciones técnicas y operativas del sistema y el precio en bolsa del kWh se requiere de la proyección de la demanda eléctrica como elemento central de la planeación y expansión de la infraestructura eléctrica.

El CND utiliza los pronósticos diarios de demanda realizados por cada uno de los agentes del mercado para determinar, además de la demanda, las políticas de inversión. Para regular los pronósticos la Comisión de Regulación de Energía y Gas (CREG) propone un pago por desviaciones en las predicciones de demanda horaria castigando monetariamente a los operadores de red causando un impacto financiero, (Comisión de Regulación de Energía y Gas, 2019). En este sentido resulta importante desarrollar un modelo de pronósticos que tenga un grado de precisión que permita a las empresas minimizar los pagos por desajustes de las proyecciones.

OBJETIVOS

Objetivo Principal

Diseñar y estimar un modelo de pronóstico de demanda diaria de energía para un operador de red, sobre el cual se evalúe el impacto de la resolución CREG 034 de 2019 con el modelo actual y el modelo estimado en este trabajo.

Objetivos Específicos

1. Describir el mercado eléctrico colombiano, su operación, su regulación y los agentes que intervienen en él como marco institucional que configura el mercado de intercambios eléctricos tanto de la bolsa de energía como de los componentes de la tarifa.
2. Evaluar los modelos que actualmente utiliza el operador de red, Compañía Energética de Occidente, para el pronóstico de demanda en Colombia.
3. Realizar el análisis de las variables que tienen las bases de consumos por parte de los clientes de la Compañía Energética de Occidente.
4. Elaborar un algoritmo con la información anterior que permita realizar pronóstico de demanda horaria y con base en ellas realizar el modelo de pronóstico utilizando Machine Learning como una técnica que supera los pronósticos y cumpla con lo estipulado en la resolución CREG 034 de 2019, (menor a 4%)

MERCADO ELÉCTRICO COLOMBIANO

En la bolsa de energía los agentes declararan disponibilidad y ofertan precios en lugar de declarar costos (Vélez Álvarez, 2019), pero con este esquema es muy factible que los cambios de precio en la bolsa se trasladen al consumidor final debido a que la energía es un servicio “in time”, para evitar este problema de variación de precio se iniciaron los contratos a largo plazo entre usuarios no regulados, generadores y empresas comercializadoras (mercado regulado). Sin embargo, el modelo de bolsa de mercado trae consigo un nuevo problema, las Generadoras de punta, en su mayoría térmicas, tienen un costo de producción más elevado al no realizar su generación de manera convencional, por lo que su actividad no está garantizada y su precio no hace que sea atractivo un contrato a largo plazo, por lo que, en el año 2006 se crea el cargo por confiabilidad, lo que garantiza que la generación térmica se remunere de acuerdo con su disponibilidad y capacidad de generación tanto en puntas de demanda como en periodos de escasez.

El modelo de Cargo por Capacidad consistía en una remuneración administrada de la capacidad de generación, que garantizaba un ingreso fijo anual por megavatio instalado, a un precio definido por el regulador. No existía una obligación concreta de los generadores, asociada a esta remuneración. El nuevo esquema de Cargo por Confiabilidad permite a los generadores contar con un ingreso de dinero fijo, independientemente de su participación diaria en el mercado mayorista, asignado para períodos de hasta 20 años, reduciendo el riesgo de su inversión. Como contraprestación, los generadores deben mantenerse disponibles y cumplir con sus Obligaciones de Energía Firme cuando se presenten las condiciones de escasez en el sistema. Para la asignación y la determinación del precio se utiliza un mecanismo de mercado que propende por la eficiencia: las subastas de obligaciones de energía firme.¹

Uno de los componentes esenciales del nuevo esquema es la existencia de las Obligaciones de Energía Firme (OEF), que corresponden a un compromiso de los generadores respaldado por activos de generación capaces de producir energía firme durante condiciones críticas de abastecimiento. Para estos propósitos, se subastan entre los generadores las Obligaciones de Energía Firme (OEF) que se requieren para cubrir la demanda del Sistema. El generador al que se le asigna una OEF recibe una remuneración conocida y estable durante un plazo determinado, y se compromete a entregar determinada cantidad de energía cuando el precio de bolsa supera un umbral previamente establecido por la Comisión de Regulación de Energía y Gas (CREG) y denominado Precio de Escasez. Dicha remuneración es liquidada y recaudada por el ASIC y pagada por los usuarios del Sistema Interconectado Nacional (SIN), a través de las tarifas que cobran los comercializadores. Las OEF del nuevo Cargo por Confiabilidad establecen un vínculo jurídico entre la demanda del MEM y los generadores, que permite, tanto a generadores como a usuarios del sistema, obtener los beneficios derivados de un mecanismo estable en el largo plazo y que da señales e incentivos para la inversión en nuevos recursos de generación, garantizando de esta forma el suministro de energía eléctrica necesario para el crecimiento del país.

El anterior modelo de Cargo por Capacidad consistía en un mecanismo de remuneración administrado de la capacidad de generación, que garantizaba un ingreso fijo anual por megavatio instalado, a un precio definido por el regulador. No existía una obligación concreta de los generadores, asociada a esta remuneración. El nuevo esquema de Cargo por Confiabilidad permite a los generadores contar con un ingreso de dinero fijo, independientemente de su participación diaria en el mercado mayorista, asignado para períodos de hasta 20 años, reduciendo el riesgo de su inversión. Como contraprestación, los

¹ <https://www.xm.com.co/Promocin%20Primera%20Subasta%20de%20Energia%20Firme/abc2.pdf>

generadores deben mantenerse disponibles y cumplir con sus Obligaciones de Energía Firme cuando se presenten las condiciones de escasez en el sistema. Para la asignación y la determinación del precio se utiliza un mecanismo de mercado que propende por la eficiencia: las subastas de obligaciones de energía firme.

El mercado eléctrico colombiano se constituye de agentes y usuarios, los agentes son aquellos que se encargan de intermediar y controlar el mercado.

- CNO: Consejo Nacional de Operación Garantizar que la operación del sistema interconectado nacional sea segura, confiable y económica y ser el ejecutor del Reglamento de Operación
- ASIC: Administrador del Sistema de Intercambios Comerciales y registro de fronteras comerciales, de los contratos de energía a largo plazo.
- LAC: Es el liquidador y administrador de cuentas del Sistema de Transmisión Nacional, encargado de facturar, cobrar y distribuir los cargos por uso del Sistema de Transmisión Nacional (STN).
- CND (XM): El Centro Nacional de despacho. Planea, supervisa y controla la operación de los recursos de generación, transmisión e interconexión para garantizar una operación segura, confiable y económica.

Generación es el inicio de la cadena productiva, para los agentes generadores ponen en el mercado la energía disponible, por lo tanto de acuerdo al pronóstico de la demanda, se hace uso de la disponibilidad de cada uno de ellos en el corto plazo ordenándolos por precio. La transmisión tiene un monopolio y pertenece a Interconexión S.A. (ISA), la transmisión se encarga de llevar grandes cantidades de energía del generador al usuario final, si la energía que se pronosticó consumir es menor a la real, la transmisión puede ser sobrecargada.

La distribución de energía como su nombre lo indica, consiste en llevar la energía a los usuarios finales, este agente es el que menos se ve afectado por los pronósticos puesto que simplemente se encargan del transporte local y su infraestructura está diseñada con base a la capacidad máxima instalada en el usuario final. La comercialización es la serie de actividades que se realizan de cara al cliente final, facturación, atención al cliente, medición de consumos, el comercializador es el que realiza las negociaciones a largo plazo directamente con el agente generador o a corto plazo en bolsa, sus negociaciones dependen directamente del pronóstico de demanda, pues de allí define que tanta energía compra en el largo plazo.

Operación y Regulación del Mercado Eléctrico Colombiano y bolsa de energía.

No se puede describir la operación del mercado eléctrico colombiano sin antes hablar de los Expertos de Mercados (en adelante XM), compañía nacional que opera y administra el mercado a través del Centro Nacional de Despacho y administra las transacciones del mercado de energía mayorista. En el artículo 33 de la ley 143 de 1994 se cita:

La operación del sistema interconectado se hará procurando atender la demanda en forma confiable, segura y con calidad del servicio mediante la utilización de los recursos disponibles en forma económica y conveniente para el país.

Y en el artículo 34 se definen como funciones específicas del *Centro Nacional de Despacho*:

- a) Planear la operación de los recursos de generación, interconexión y transmisión del sistema nacional, teniendo como objetivo una operación segura, confiable y económica;*
- b) Ejercer la coordinación, supervisión, control y análisis de la operación de los recursos de generación, interconexión y transmisión incluyendo las interconexiones internacionales;*
- c) Determinar el valor de los intercambios resultantes de la operación de los recursos energéticos del sistema interconectado nacional;*
- d) Coordinar la programación del mantenimiento de las centrales de generación y de las líneas de interconexión y transmisión de la red eléctrica nacional;*
- e) Informar periódicamente al Consejo Nacional de Operación acerca de la operación real y esperada de los recursos del sistema interconectado nacional y de los riesgos para atender confiablemente la demanda;*
- f) Informar las violaciones o conductas contrarias al Reglamento de Operaciones;*
- g) Las demás atribuciones que le confiera la presente Ley.*

De acuerdo con los artículos de la ley anteriormente citada, XM planea, coordina, supervisa y controla la operación de los recursos de generación y transmisión de energía incluyendo la interconexión Venezuela, o mejor las interconexiones internacionales; cumpliendo con los reglamentos expedidos por la Comisión Nacional de Energía y Gas (en adelante CREG) y los acuerdos técnicos realizados y aprobados por el Consejo Nacional de Operación (en adelante CNO), en el marco de la planeación de la operación se encuentran lo análisis respectivos de la demanda del sistema asegurando la operación confiable del mismo.

La bolsa de energía es el sistema de intercambio de la energía en Colombia, que sirve para comprar y vender contratos de energía en el sistema interconectado nacional, mediante la suscripción de contratos de largo y corto plazo y sobre precios definidos por el mercado y la oferta de los generadores. XM (XM S.A. ESP, 2017) describe las operaciones en el mercado mayorista como el mercado de grandes bloques de energía eléctrica en el que comercializadores y generadores venden y compran energía tanto en los contratos de corto plazo como en los contratos de largo plazo.

Las transacciones en los contratos de largo plazo es responsabilidad de las partes, quienes pactan precio y cantidad, el ASIC (Administrador de Intercambios Comerciales) registra los contratos una vez se hayan cumplido los requisitos regulatorios. Considerando que debe establecerse un orden de mérito de precio, el ASIC debe aplicar primero las condiciones de precio de todos los contratos registrados para luego asignar las cantidades horariamente.

Por otra parte, el mercado a corto plazo es un mercado para el día siguiente, en el que los generadores de manera confidencial registran en la bolsa su disponibilidad de energía a nivel horario y la oferta de precio, de acuerdo con esta información el CND asigna los recursos más económicos para cubrir la demanda hora a hora, en este mercado todos los agentes están obligados a participar. Por último, la energía se transporta a través del SIN lo que facilita que generadores y comercializadores ubicados en cualquier parte del país puedan negociar energía.

Tarifa de Energía.

Conociendo el funcionamiento del mercado eléctrico colombiano se hace más sencillo conocer la naturaleza de la tarifa. La fórmula tarifaria para los usuarios regulados se propuso en el documento CREG 043 de 2007, se publicó en la resolución 119 de 2007 y se complementó con la resolución 156 de 2009, a continuación, se realizará un breve recorrido por dichos documentos con el fin de entender los componentes tarifarios de la energía eléctrica. La fórmula tarifaria definida en el documento CREG 043 de 2007 (Comisión de Regulación de Energía y Gas, 2007) se describe en la ecuación 1:

$$CUv_{n,m} = G_m + T_m + D_{n,m} + Cv_m + PR_{n,m} + R_m$$

Ecuación 1 Ecuación que rige la tarifa energética del país.

En donde:

CUv_{n,m}: Componente variable del Costo Unitario de Prestación del Servicio (\$/kWh) para los usuarios conectados al nivel de tensión n, correspondiente al mes m, del Comercializador Minorista i, en el Mercado de Comercialización j.

G_m: Costo de compra de energía (\$/kWh) para el mes m, del Comercializador Minorista i, en el Mercado de Comercialización j.

T_m: Costo por uso del Sistema Nacional de Transmisión (\$/kWh) para el mes m.

D_{n,m}: Costo por uso de Sistemas de Distribución (\$/kWh) correspondiente al nivel de tensión n para el mes m.

CV_i: Margen de Comercialización correspondiente al mes m, del Comercializador Minorista i, en el Mercado de Comercialización j que incluye los costos variables de la actividad de comercialización, expresado en (\$/kWh).

R_m: Costo de Restricciones y de Servicios asociados con generación en \$/kWh asignados al Comercializador Minorista i en el mes m, conforme al Capítulo VI de la presente Resolución.

PR_{n,m}: Costo de compra, transporte y reducción de pérdidas de energía (\$/kWh) acumuladas hasta el nivel de tensión n, para el mes m, del Comercializador Minorista i, en el Mercado de Comercialización j.

n: Nivel de tensión de conexión del usuario.

m: Es el mes para el cual se calcula el Costo Unitario de Prestación del Servicio.

La componente G, para los generadores la exactitud en los pronósticos de demanda hace que los contratos a largo plazo sean más confiables y así mismo les permita declarar sus reservas con mayor precisión, la tarifa de dicha componente depende de los contratos bilaterales firmados, la demanda comercial regulada, el costo promedio ponderado de compras de energía de cada comercializador y el precio de la energía comprada en bolsa para cada comercializador, en resumen, el precio depende de las transacciones realizadas en bolsa tanto a largo como a corto plazo. La generación en Colombia es principalmente hidráulica, por lo que los precios de corto plazo dependen de la capacidad disponible de las plantas y ellas dependen directamente de los cambios climáticos, para disminuir el impacto por cambios climáticos abruptos se implementó el llamado cargo por confiabilidad, lo que permitió ejecutar grandes proyectos que se encuentran dentro del plan de expansión nacional, con este cargo por confiabilidad, por ejemplo, el país logró enfrentar a dos fenómenos del niño sin tener que acudir al racionamiento de energía.

La componente T, la transmisión de energía: El precio en el STN (Sistema de Trasmisión Nacional) son costos calculados por el XM (publicados por el Liquidador de Cuentas) se calcula de acuerdo con el cálculo de costo unitario de cada activo en el sistema, su vida útil, su reposición y su expansión, siendo las dos últimas consideradas como inversión, la inversión no se renumera si la aportó el estado; de acuerdo con ese costo unitario se estima una anualidad a la que se agregan algunas tasas e ingresos por otros conceptos. La tasa de retorno que se fijó de acuerdo a la metodología publicada en la resolución CREG 083 de 2008 (Comisión de Regulación de Energía y Gas, 2008) en el que se indica que dicha tasa es del 11,50% anual, adicional a esta tasa se reconoce un costo por administración, operación y mantenimiento (AOM), recordemos que de acuerdo con lo dicho anteriormente la transmisión es un monopolio natural de ISA.

La componente D, distribución de energía: La distribución de energía se realiza por niveles de tensión, 1 hasta 1.000 V, 2 hasta 30.000V, 3 hasta 57.500 V y 4 hasta 220.000 V. El nivel 4 de tensión en la distribución de energía es denominado STR (Sistema de Trasmisión Regional) y su remuneración se calcula a partir de los ingresos de los OR que tienen mercados de

comercialización en el mismo STR, el LAC se encarga de estimar este cálculo de acuerdo con la metodología publicada en la resolución CREG 015 de 2018 con base en una relación entre el ingreso mensual y la demanda del mercado de comercialización pertenecientes a ese STR. En los niveles de tensión 3, 2 y 1, se calcula como una relación entre el ingreso del OR (Operador de Red), las conexiones realizadas en dicho nivel y el índice de pérdidas por la energía de entrada para el nivel tensión.

La componente R, Restricciones, como su nombre lo indica son las restricciones que se tienen en el momento de realizar el transporte de energía eléctrica, debido a estas restricciones de transporte la energía debe comprarse a un generador que oferta un precio más alto, algunas causas de estas restricciones son capacidad de transporte de las líneas, porque la línea se encuentra indisponible ya sea por falla o en su defecto por actos terroristas.

La componente Pr, las pérdidas desconocidas, partiendo del hecho de que ningún sistema es 100% eficiente, la cadena de producción de energía eléctrica tampoco lo es, ya que por el calentamiento de conductores o de cualquier otro elemento del sistema se transforma en calor, a estas se le llaman pérdidas técnicas, por otro lado las conexiones fraudulentas, la defraudación de fluidos, el subregistro de energía, la alteración de conexiones son las pérdidas no técnicas y también se reconocen dentro de la tarifa. Cada operador de red tiene un porcentaje de pérdidas reconocidas, lo que incentiva a estos a mantener sus indicadores por debajo del tope reconocido.

MODELO ACTUAL DE PRONÓSTICO DE DEMANDA

Para el pronóstico de demanda, XM publica los resultados de un pronóstico propio, el operador de red está en la libertad de aceptar o no dicho pronóstico (XM comparte los resultados de dicho modelo, más no el modelo en sí), si el operador de red no acepta el pronóstico propuesto por XM, debe subir su propio pronóstico. El modelo propuesto por XM se basa en modelos de pronóstico Sarima (Seasonal AutoRegresive Integrated Moving Average Model), el principio básico de este modelo es la estimación de las relaciones de los valores pasados con los valores actuales, la incorporación de los shocks en el mercado al comportamiento de los agentes y la agregación de componentes estacionales. La ventaja de estos modelos es sencillez de la estimación de parámetros, pero su debilidad yace en la adaptabilidad. En general el pronóstico realizado con este tipo de modelos es conservador, porque tienden a la media de largo plazo y evitan los valores extremos, que son en realidad los que se requiere predecir.

Se trata de un modelo que utiliza los datos históricos y se busca que las predicciones futuras se den por medio de la asociación de los valores pasados. Los datos reales del modelo propuesto por XM toma la información de la demanda por hora durante un período de tiempo, para replicar con el modelo la demanda de años anteriores al momento actual. Si bien es un análisis univariado se asume que en el dato de un momento del tiempo, se resume toda la información relevante del sector sobre la cual se toman decisiones por parte de los actores.

Por otro lado, la estacionalidad del modelo hace que el comportamiento de este frente a fenómenos atmosféricos únicos en cada año no se tenga en cuenta, por ejemplo, en las temporadas de lluvia el comportamiento de la demanda tiende a bajar, puesto que las fallas en las redes cortan el suministro y un porcentaje de ellas se debe a fenómenos atmosféricos tales como descargas, vendavales, tormentas, etc. Por otro lado, las temporadas soleadas conllevan a un aumento en el consumo de los clientes, puesto que se consume más en aires acondicionados, ventiladores, refrigeradores, etc.

Para el caso del pronóstico de demanda de la compañía energética de Occidente, se utiliza el modelo propuesto por XM y se realiza un aumento en la curva para que esté más de acuerdo con el crecimiento de la demanda o en su defecto con el fenómeno climático. Este ajuste se realiza de manera manual y observando el comportamiento de la semana inmediatamente anterior.

Para el modelo propuesto inicialmente por XM se tienen errores entre el 5 y el 7%, por lo tanto, siempre es necesario acudir al cálculo descrito en el párrafo anterior y para el caso de este pronóstico, se tiene que los valores finales de error son entre el 4 y el 6 %. Para conocer la demanda de la compañía energética de occidente es necesario conocer la ecuación de demanda, en la *Tabla 1* y *Tabla 2* se observan las fronteras registradas con importación y exportación de energía.

Importación				
Código	Nodo	Estado	Nivel Tensión	Observación
Frt00224	1	Activa	N4	Línea Pance-Jamundí (Sub Pance)
Frt00288	7	Activa	N3	Planta Rio Palo Principal
Frt00295	1	Activa	N2	Planta Mondomo Principal
Frt00348	9	Activa	N3	Planta Sajandi
Frt00350	6	Activa	N2	Planta Silvia
Frt10320	4	Activa	N4	San Berbardino Principal
Frt10321	8	Activa	N3	Sub_Centro Cto Coconuco Principal
Frt10322	3	Activa	N4	Línea Rio Mayo – Principal
Frt32029	5	Activa	N4	Línea San Martin - Zaque (M-Principal)
Frt10324	3	Activa	N4	Planta Florida II
Frt10325	2	Activa	N4	Cabaña - Paez 230 Kv Principal
Frt14354	1	Activa	N4	Línea Pance- Santander (M-Principal)
Frt22639	2	Activa	N3	Proenca 2
Frt28373	2	Activa	N3	Proenca 2
Frt29362	4	Activa	N4	Línea Guapi*-San Bernardino

Tabla 1 Fronteras de importación Compañía Energética de Occidente Fuente: Compañía Energética de Occidente.

Exportación				
Código	Nodo	Estado	Nivel Tensión	Observación
Frt00038	3	Activa	N4	Linea Principal-Riomayo
Frt00039	5	Activa	N4	Zaque-Principal
Frt00036	4	Activa	N4	San Berbardino Principal
Frt10568	1	Activa	N4	Linea Pance-Jamundi (Sub Pance)
Frt00035	8	Activa	N3	Sub_Centro Cto Coconuco Principal
Frt00041	2	Activa	N4	Cabaña - Paez 230 Kv Principal
Frt11244	1	Activa	N4	Subestacion Jamundi Principal (Tp2)
Frt11672	1	Activa	N4	Subestacion Jamundi Trafo 1 Principal
Frt26429	1	Activa	N4	Jamundi Transf 3
Frt14380	1	Activa	N4	Linea Pance- Santander (M-Principal)
Frt29361	4	Activa	N4	Linea Sambernardino-Guapi

Tabla 2 Fronteras de importación Compañía Energética de Occidente Fuente: Compañía Energética de Occidente.

Una vez conocidas cuales son las fronteras, en la *Ilustración 5* se observa la ubicación de cada una de estas fronteras y que se reportan a la ASIC.

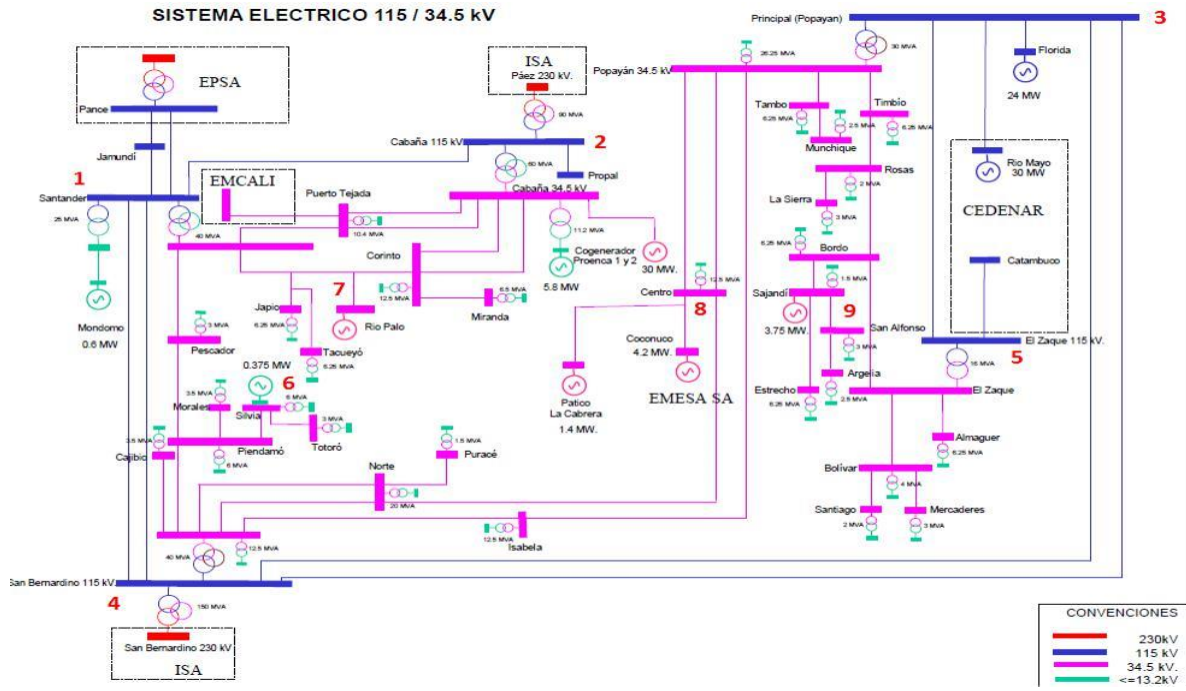


Ilustración 1 Diagrama unifilar Compañía Energética de Occidente.

La ecuación de demanda es la sumatoria de entradas y salidas de energía al sistema (importación y exportación). XM basa sus datos del modelo en la energía reportada en cada una de las fronteras comerciales por medio del ASIC, estas fronteras son reportadas por cada uno de los comercializadores con una periodicidad mensual y se presentan como el consumo de energía hora a hora por cada uno de ellos, para el caso del departamento del Cauca se presentan 9 fronteras comerciales. Al realizar un seguimiento al pronóstico propuesto con XM se encontró la siguiente desviación, que se presenta como un resumen mensual y se realiza la estimación de acuerdo con la ecuación propuesta en la resolución CREG 100 de 2019 (Comisión de Regulación de Energía y Gas, 2019):

$$Desviación_{i,m} = \max. \left(0, \frac{|D_{p,i,m} - D_{r,i,m}|}{D_{r,i,m}} \times 100 - 4\% \right)$$

Ecuación 2 Cálculo de desviación en el pronóstico de la energía.

Donde:

$|D_{p,i,m} - D_{r,i,m}|$ Diferencia absoluta

$D_{p,i,m}$ Predicción de demanda en kWh del comercializador i para el mercado m para el tramo

$D_{r,i,m}$ Demanda real en kWh del comercializador i para el mercado m para el tramo.

En dónde los tramos se definen como los periodos horarios 1 y 2 y contemplan la demanda de las 00:00 a las 12:00 horas y de las 12:01 a las 24:00 horas En la *Tabla 3* se presenta un resumen del promedio de desviaciones presentadas mensualmente., esta información es el resultado de contrastar la demanda real vs la demanda pronosticada para cada periodo de aplicación. En dónde los tramos se definen como los periodos horarios 1 y 2 y contemplan la demanda de las 00:00 a las 12:00 horas y de las 12:01 a las 24:00 horas.

Mes	Promedio de Desviación tramo 1	Promedio de Desviación tramo 2
ene	0%	0%
feb	1%	1%
mar	7%	5%
abr	2%	2%
may	1%	1%
jun	1%	0%
jul	1%	1%
ago	1%	1%
sep	7%	4%
oct	0%	0%
nov	1%	1%
dic	5%	3%

Tabla 3 Resumen cálculo de desviación pronóstico-propuesta XM datos 2018. Fuente: Autores.

Y de acuerdo con la resolución, todo pronóstico con valores de desviación mayores a 0 (teniendo en cuenta que la ecuación tolera hasta un 4% de desviación y que este ya está contemplado en la *Tabla 3*) serán penalizados. Al aplicar la penalización descrita en la resolución CREG 034 de 2019 (en la resolución CREG 100 no está definida) se obtienen las siguientes penalizaciones, ver *Tabla 4*.

MES	COSTO
Ene	\$ 2 360 062
Feb	\$ 9 544 358
Mar	\$ 39 685 619
Abr	\$ 14 448 927
May	\$ 8 374 755
Jun	\$ 6 525 252
Jul	\$ 8 601 936
Ago	\$ 10 377 992
Sep	\$ 65 055 163
Oct	\$ 4 626 946
Nov	\$ 7 512 236
Dic	\$ 19 397 035
Total general	\$ 196 510 280

Tabla 4 Cálculo penalizaciones por desviaciones pronóstico (datos 2018) Fuente: autores.

Para disminuir el error en los pronósticos propuestos por XM, actualmente se realiza el pronóstico comparando la serie histórica con la propuesta por XM se eliminan picos atípicos en la demanda y se suaviza de acuerdo con el histórico de demanda para días ordinarios, fines de semana y festivos, obteniendo los resultados descritos en las tablas

ANÁLISIS DE VARIABLES QUE INTERVIENEN EN LA DEMANDA DE ENERGÍA ELÉCTRICA

Buscaremos ahora, plantear un modelo de predicción de demanda energética diaria utilizando Machine Learning (ML) que mejore los porcentajes de desviación en las predicciones actuales, para dicho modelo analizaremos la pertinencia de incluir como entradas las siguientes cuatro variables:

- Histórico de demanda de energía eléctrica (Utilizado en modelos actuales)
- Temperatura promedio de la región de estudio
- Precipitaciones mensuales de la región de estudio
- PIB agregado para la región de estudio.

De las cuatro variables a analizar únicamente el histórico de demanda de energía eléctrica es incluido en los modelos de predicción actuales, así mismo discretizaremos el modelo para cada una de las 24 horas del día de manera tal que para cada una de las horas del día se tendrá un modelo que se entrenará y se probará de manera independiente.

Ahora bien, es necesario describir las relaciones entre las variables de entrada propuestas y la demanda de energía eléctrica por lo que a continuación realizamos un estudio general de las series de datos que tenemos y de la variable a predecir para entender las relaciones y analizar su influencia en el modelo de predicción final.

Histórico de demanda de energía eléctrica

Se tienen datos de la demanda de energía eléctrica con resolución horaria de 2012 a 2017 para una región del país.

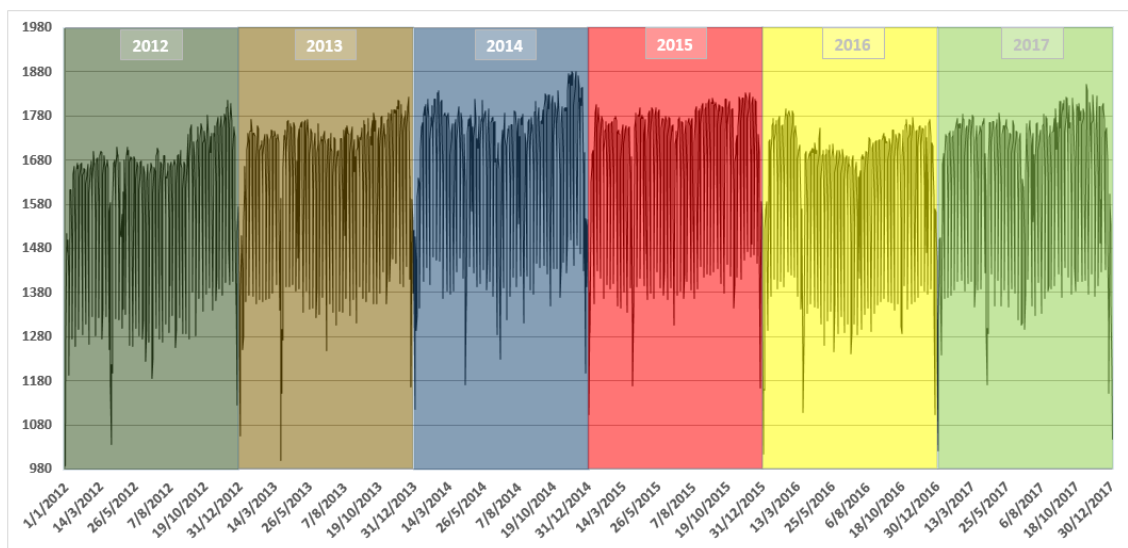


Ilustración 2 Demanda de energía eléctrica en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores

Se observa que la serie de demanda energética tiene un patrón año a año, identificando una estacionalidad de la serie, se identifica una relación con el tipo de día y época del año, por ejemplo, se observan picos mínimos de demanda energética en las festividades de fin de año, 25 de diciembre y 01 de enero, así como en los días feriados de semana santa, del mismo modo se identifica un aumento en el consumo de energía eléctrica en el cuarto trimestre del año y una disminución promedio en la época de vacaciones de junio. En la siguiente ilustración se muestra el consumo de energía eléctrica durante 2017.

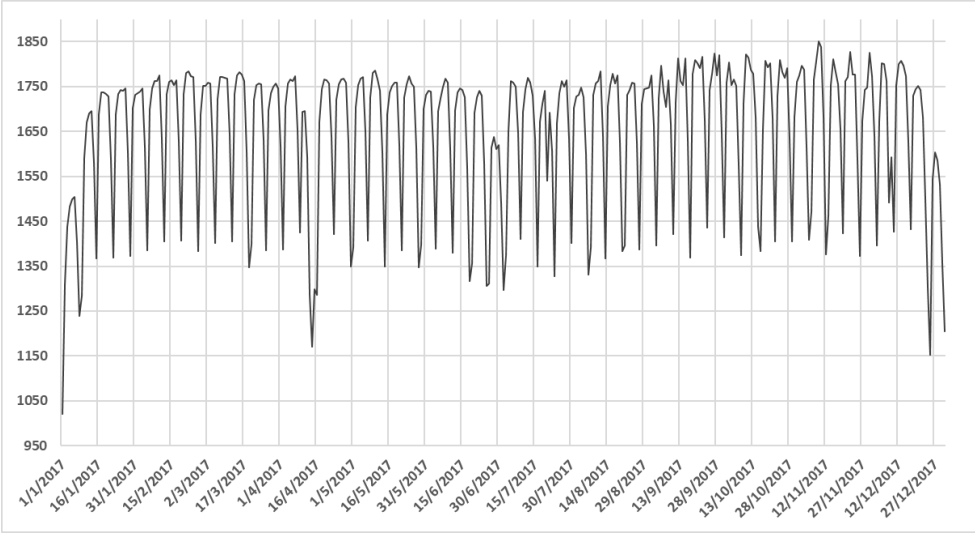


Ilustración 3 Demanda de energía eléctrica en 2017 en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores

Como hemos mencionado, se evidencia una relación entre el tipo de día y época del año con el consumo de energía que es atribuible a la actividad industrial de la región, entendiendo que durante las festividades de fin de año y semana santa este sector económico disminuye su producción afectando así el consumo de energía. Ahora bien, en las siguientes ilustraciones detallamos la demanda energética durante los meses de Junio y Julio.

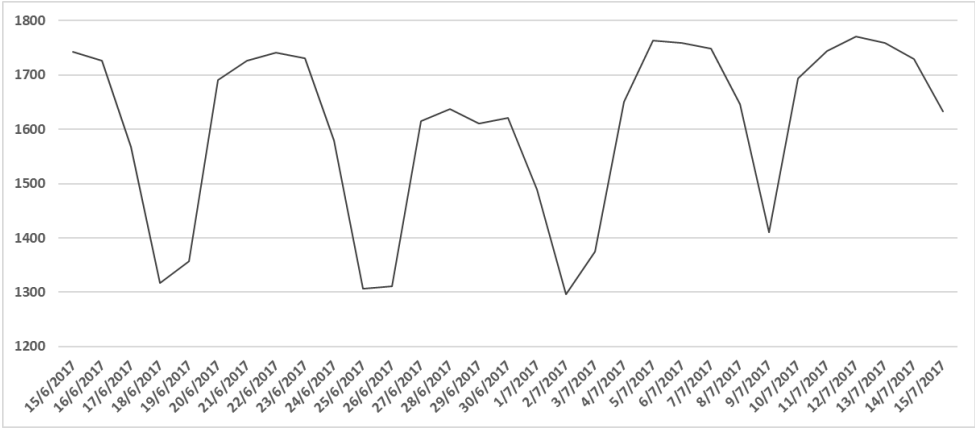


Ilustración 4 Demanda de energía eléctrica en los meses de Junio y Julio de 2017 en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores

Para esta época del año podemos evidenciar la disminución del consumo energético durante los fines de semana, así como durante la semana de vacaciones de mitad de año atribuible al receso escolar masivo de esta época.

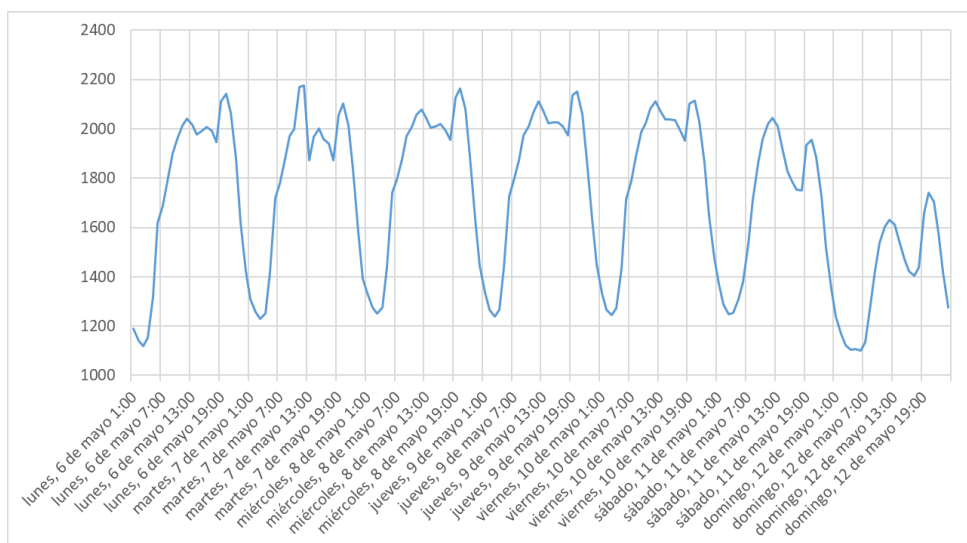


Ilustración 5 Demanda de energía eléctrica en la primer semana del mes de mayo de 2017 en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores

En la Ilustración 5 se muestra el comportamiento horario de la demanda eléctrica durante una semana típica del año 2017, evidenciando la disminución de la energía demandada los fines de semana y en las noches, así como picos máximos de demanda cerca al medio día y las 19:00, reflejando en más detalle el comportamiento de consumo de los sectores residenciales y comerciales, claro sin olvidar el sector industrial que no consume de la misma manera los fines de semana.

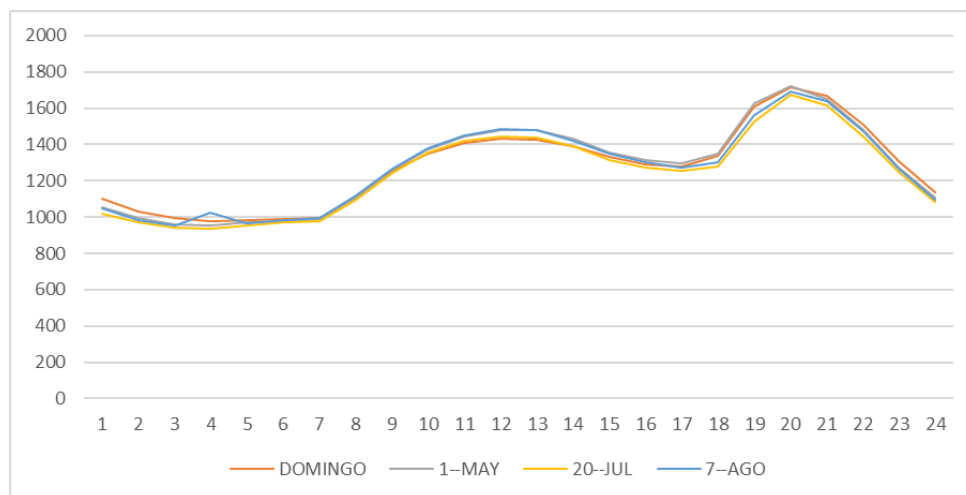


Ilustración 6 Demanda de energía eléctrica típica para domingos y festivos en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores

De acuerdo con la data seleccionada y como se observa en la Ilustración 6, la demanda energética de la región para los días feriados diferentes a festividades de fin de año y semana santa es similar a la demanda energética de los domingos, por otro lado, en la Ilustración 7 se logra evidenciar que la curva demanda promedio de los días sábados previos a un lunes festivo aunque mantiene la misma tendencia es menor en su magnitud.

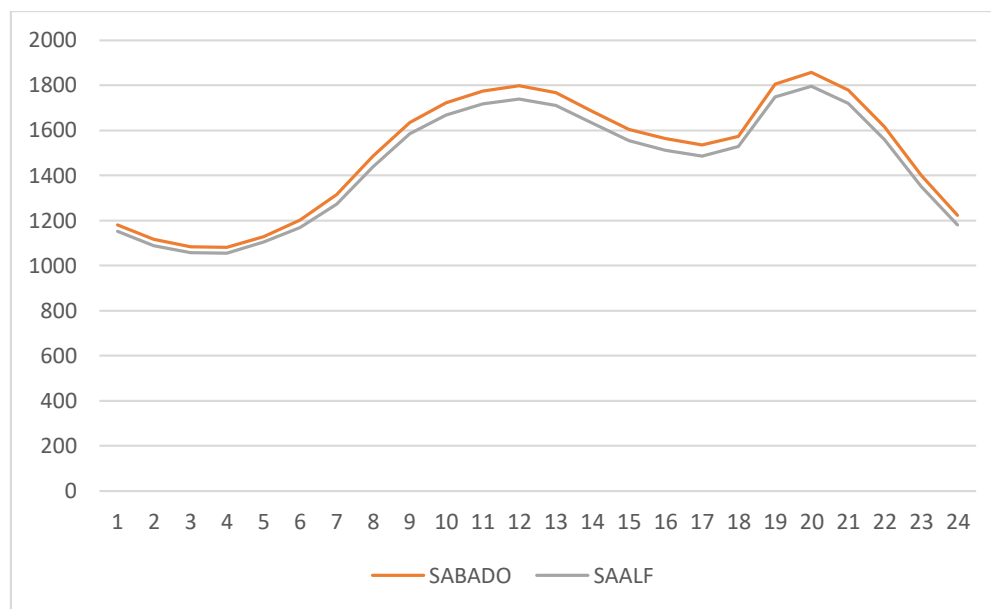


Ilustración 7 Demanda de energía eléctrica típica para los sábados y sábados antes de un lunes feriado en una región del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país. Elaboración: Autores

Como evidenciamos, el papel del histórico de demanda energética es de gran relevancia teniendo en cuenta que contiene todas las tendencias de consumo a lo largo del tiempo, reflejando los impactos de los diferentes actores y la cotidianidad en el consumo.

Temperatura promedio en la región

Luego de observar la relación entre la demanda y los tipos de día, la siguiente variable a tener en cuenta que podría influir en la demanda de energía eléctrica es la temperatura entendiendo, claro que no debe estar tan marcada como en otros países que viven el fenómeno de las cuatro estaciones meteorológicas en el año (Cabezón Manchado, 2018), analizándolo desde nuestro punto de vista, como país tenemos épocas del año en donde la temperatura promedio aumenta por la ausencia de lluvias y nubosidad, así como otras épocas marcadas por altos niveles de precipitación (lluvia) las cuales pueden variar en periodo e intensidad de acuerdo a la región del país. Para las épocas con temperaturas elevadas se espera el aumento del consumo energético de la región ante la evidente utilización en masa de sistemas de refrigeración y ventilación. En las siguientes ilustraciones buscaremos exponer las características en cuanto a temperatura de la región estudiada.

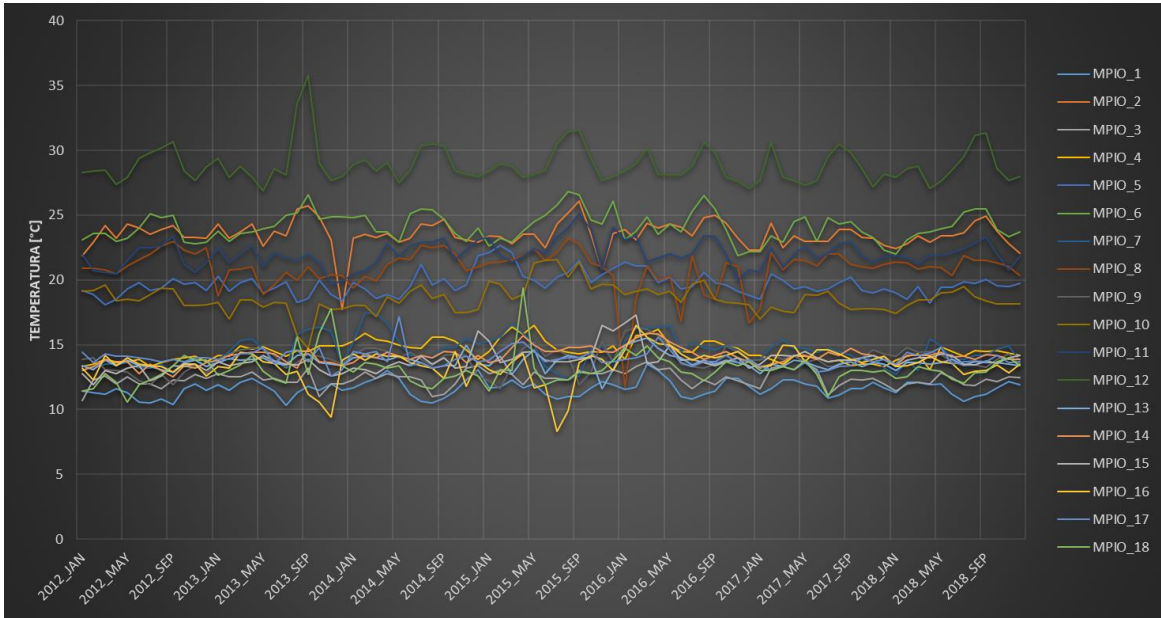


Ilustración 8 Temperatura promedio en estaciones meteorológicas de la región. Fuente: Corporación autónoma regional (CAR). Elaboración: Autores

En la Ilustración 8 se observa la temperatura promedio con resolución mensual en las 18 estaciones de medición instaladas en diferentes municipios a lo largo y ancho de la región de estudio, luego, en la Ilustración 9 se observa la temperatura promedio en la región, cabe anotar que se logra obtener datos para el periodo comprendido entre enero de 2012 a diciembre de 2018.

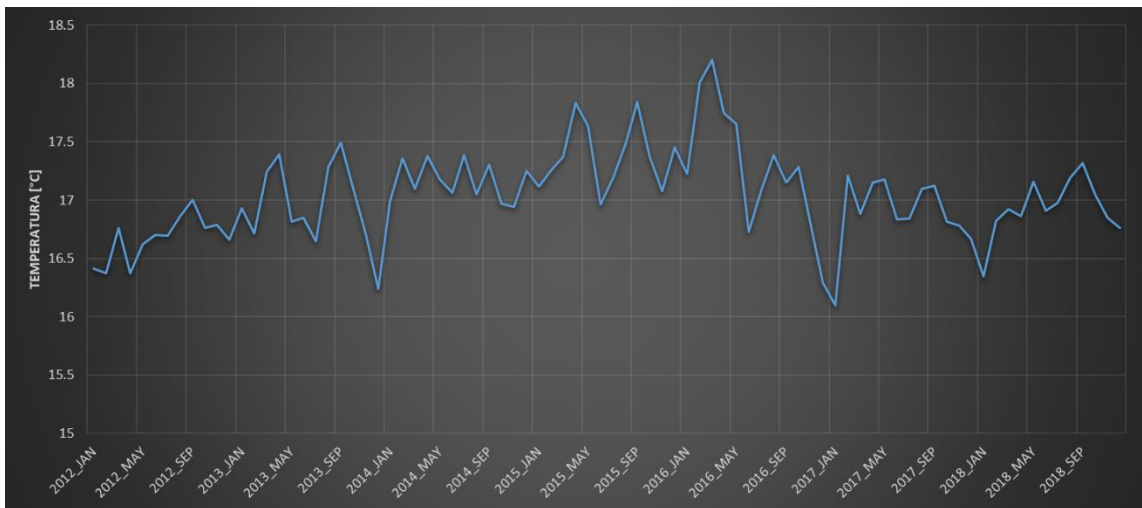
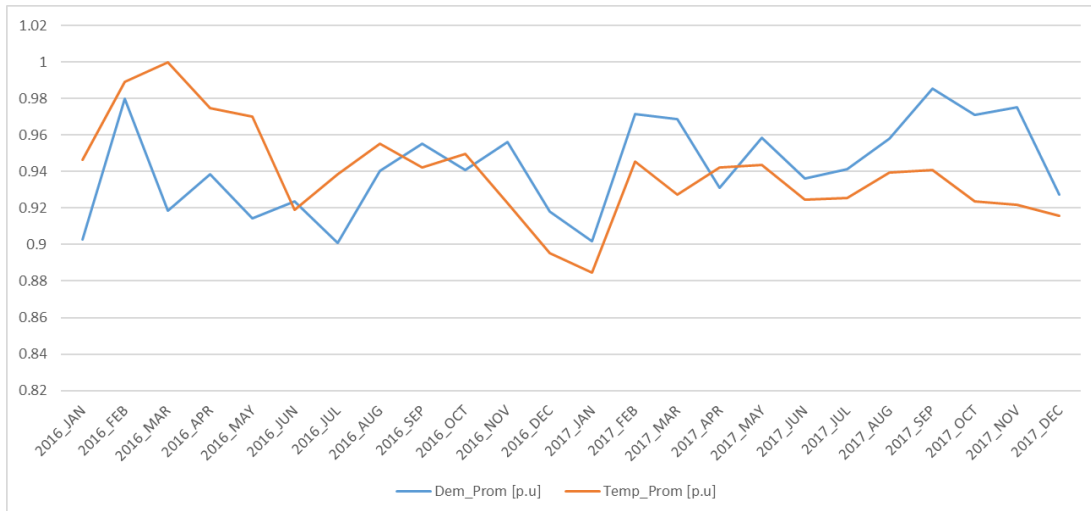


Ilustración 9 Temperatura promedio de la región. Fuente: Corporación autónoma regional (CAR). Elaboración: Autores

Adicionalmente en la Ilustración 10 se muestra el detalle de la tendencia entre la temperatura promedio y la demanda energética promedio mensual entre 2016 y 2017 encontrando que existe alguna correlación entre estas dos variables.



*Ilustración 10 Demanda energética y temperatura promedio mensual para una región del país.
Elaboración: Autores*

Precipitaciones (lluvia) en la región.

Para los momentos con aumento en los niveles de precipitación se esperan disminuciones puntuales de la energía demandada por la imposibilidad de suministrarla ante fallas en la infraestructura eléctrica ocasionadas por la lluvia, los vientos y las descargas atmosféricas.

Al igual que los datos de temperatura promedio, se logró recopilar la información de precipitaciones en la región obtenida para 64 estaciones de monitoreo de lluvias con resolución mensual para el periodo comprendido entre enero de 2012 a diciembre de 2017. Se identifican dos épocas durante el año con aumento en las lluvias de la región comprendidas en los meses de abril-mayo y octubre-noviembre. Se evaluará en la etapa de entrenamiento y prueba la inclusión de esta variable en el modelo a construir de acuerdo con la correlación existente entre las épocas de lluvia y la energía dejada de suministrar por fallas en la infraestructura eléctrica en la región de estudio.

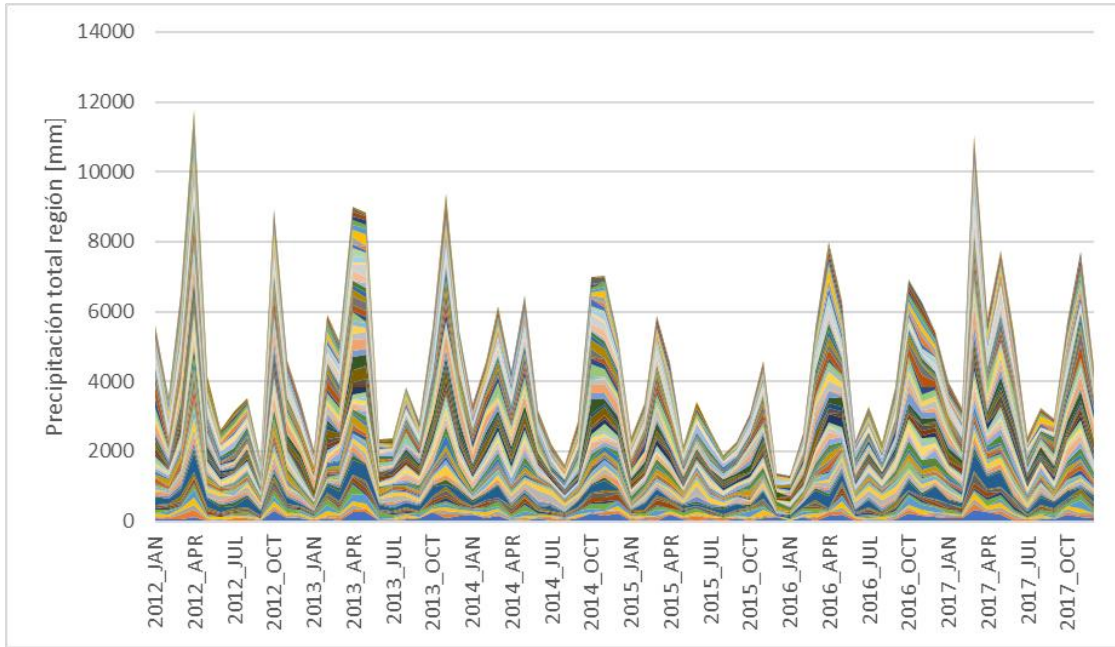


Ilustración 11 Precipitaciones totales en estaciones meteorológicas de la región. Fuente: Corporación autónoma regional (CAR). Elaboración: Autores

Con los datos obtenidos y realizando una transformación en por unidad (p.u) en la Ilustración 12 se muestra la tendencia del histórico de demanda energética y las precipitaciones totales para la región de estudio, en donde se observa gráficamente que no existe correlación entre estas dos variables.

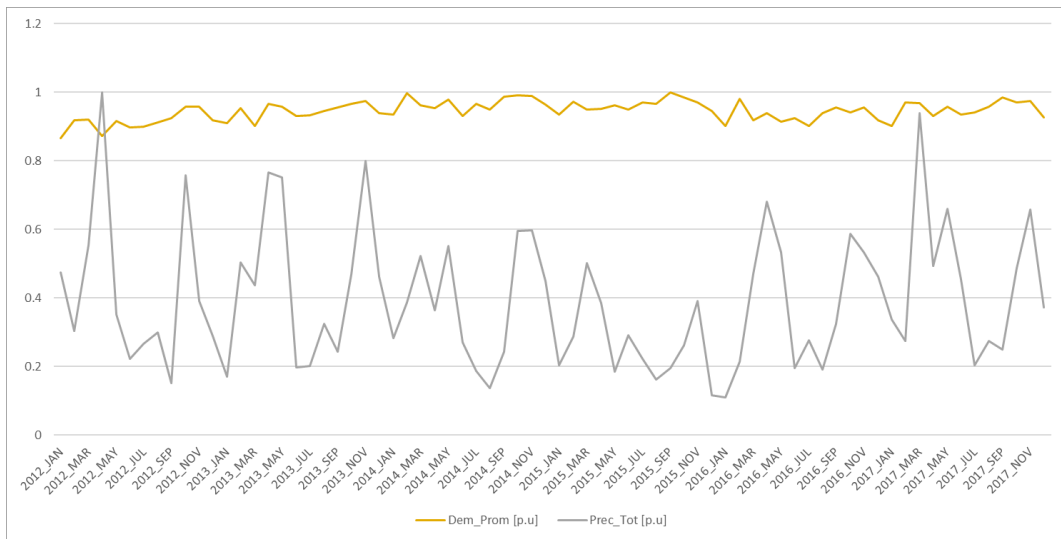
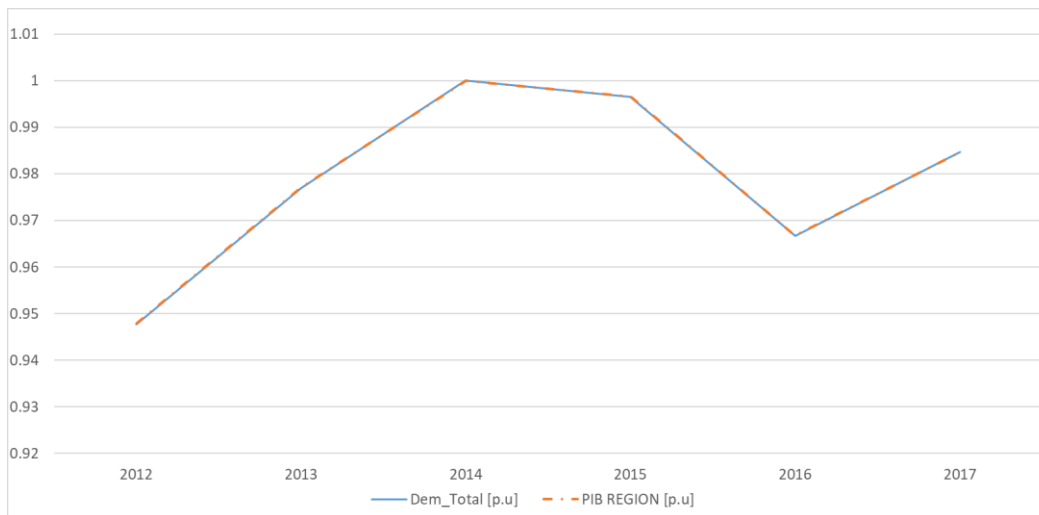


Ilustración 12 Comparación de los valores de precipitaciones totales mensuales y demanda energética mensual. Elaboración: Autores

Producto Interno Bruto (PIB) en la región.

Por último y ante la evidente y conocida correlación entre el producto interno bruto (PIB) y el consumo energético, se logra evidenciar una correlación total entre dichas variables para la región en estudio.



*Ilustración 13 Demanda energética total anual y PIB de una región central del país. Fuente: Reporte de demanda real al CND por una distribuidora de la región central del país y DANE
Elaboración: Autores*

MODELO MACHINE LEARNING PARA LA PREDICCIÓN DE LA DEMANDA.

De acuerdo con los datos recopilados y el tipo de predicción requerida se selecciona el método Random Forest Regressor para el entrenamiento y testeo del modelo, dicho modelo hace parte de los modelos Machine Learning de tipo Ensemble, que comparado con los métodos convencionales de regresión lineal multivariable mejoran significativamente la precisión de las predicciones (Takefuji & Shoji, n.d.)

El método Random Forest Regressor es una combinación de predictores de árboles, de modo que cada árbol depende de los valores de un vector aleatorio muestreado independientemente y con la misma distribución para todos los árboles en el bosque cuyo error converge a un límite a medida que aumenta el número de árboles en el bosque. La precisión de un bosque de clasificadores de árboles depende de la fuerza de los árboles individuales en el bosque y de la correlación entre ellos.(Pavlov, 2019)

El uso de una selección aleatoria de características para dividir cada nodo produce tasas de error favorables y hace más robusto el modelo con respecto al ruido. El monitor interno de error, de fuerza y correlación, se utilizan para mostrar la respuesta del modelo al aumento de la cantidad de características utilizadas en la división. Las estimaciones internas también se utilizan para medir la importancia de cada una de las variables independientes.(Pavlov, 2019)

Para entenderlo un poco mejor, supongamos que hacemos una pregunta compleja a miles de personas al azar, luego agregamos sus respuestas. En muchos casos, se encontrará que esta respuesta agregada es mejor que la respuesta de un experto. Esto se llama la sabiduría de la multitud. Del mismo modo, si se agregan las predicciones de un grupo de predictores (como clasificadores o regresores), a menudo se obtendrán mejores predicciones que con el mejor predictor individual. Un grupo de predictores se llama conjunto; así, esta la técnica se llama Ensemble Learning, y un algoritmo de Ensemble Learning es llamado un método Ensemble. Por ejemplo, se puede entrenar a un grupo de clasificadores de árbol de decisión, cada uno en un subconjunto aleatorio diferente del conjunto de entrenamiento. Para hacer predicciones, solo se obtienen las predicciones de todos los árboles individuales, luego predice la clase que obtiene más votos. Tal conjunto de árboles de decisión es llamado Random Forest, y a pesar de su simplicidad, este es uno de los más Potentes algoritmos de aprendizaje automático disponibles en la actualidad.(Géron, 2017)

Definición de variables y periodos de la data

De acuerdo con la data recopilada se tienen cuatro grupos de variables definidas como se muestra en la Tabla 5, luego de la recopilación y la evaluación de disponibilidad de datos de cada variable se define como periodo para el entrenamiento y testeo del modelo el comprendido entre enero de 2012 hasta abril de 2018, periodo sobre el cual se tiene información disponible en la plataforma <https://www.datos.gov.co/> y proporcionados por la corporación autónoma regional (CAR) para las variables de temperatura promedio y precipitaciones en la región.

Data	Data_Inicio	Data_Fin	Resolución	Tipo_Variables	Cant_Datos
Histórico de demanda energética	1-jun-04	30-jun-19	Horaria	Demanda por hora	132168
Precipitación (lluvia)	1-ene-12	30-abr-18	Mensual	64 Estaciones metereologicas	5364
Temperatura Promedio	1-ene-12	30-abr-18	Mensual	18 Estaciones metereologicas	1512
PIB_Región	1-ene-05	31-dic-18	Anual	Capital + Departamento	28

Tabla 5 Características de las cuatro variables a evaluar en el desarrollo del modelo

Histórico de demanda energética

En el histórico de demanda energética se incluyeron variables tipo Dummy para cada tipo de día recomendado por la comisión de regulación de energía y gas (CREG) y se agregan variables que describan el comportamiento de cada una de las horas en el histórico, en la Tabla 6 se detalla cada una de las 56 variables de entrada para el modelo respecto al histórico de demanda.

Variable	Descripcion	Tipo dato
YEAR	Año	Entero
MONTH	Mes	Entero
DAY	Día	Entero
WEEK	Semana del año	Entero
DAY_WEEK	Día de la semana	Entero
DTOTAL	Demanda Total del día	Float
DMAX	Demanda maxima del día	Float
LUNES	Identificador día de la semana	Dummy (1 - 0)
MARTES	Identificador día de la semana	Dummy (1 - 0)
MIERCOLES	Identificador día de la semana	Dummy (1 - 0)
JUEVES	Identificador día de la semana	Dummy (1 - 0)
VIERNES	Identificador día de la semana	Dummy (1 - 0)
SABADO	Identificador día de la semana	Dummy (1 - 0)
DOMINGO	Identificador día de la semana	Dummy (1 - 0)
VIALF	Identificador Viernes antes de Lunes Festivo	Dummy (1 - 0)
SAALF	Identificador Sabado antes de Lunes Festivo	Dummy (1 - 0)
DOALF	Identificador Domingo antes de Lunes Festivo	Dummy (1 - 0)
LF	Identificador Lunes Festivo	Dummy (1 - 0)
DSS	Identificador Domingo de Ramos	Dummy (1 - 0)
LSS	Identificador Lunes Semana Santa	Dummy (1 - 0)
MSS	Identificador Martes Semana Santa	Dummy (1 - 0)
MMSS	Identificador Miercoles Semana Santa	Dummy (1 - 0)
JSS	Identificador Jueves Semana Santa	Dummy (1 - 0)
VSS	Identificador Viernes Semana Santa	Dummy (1 - 0)
SSS	Identificador Sabado Semana Santa	Dummy (1 - 0)
DDSS	Identificador Domingo de resurrección	Dummy (1 - 0)
20_JUL	Identificador Fiesta Patria	Dummy (1 - 0)
7_AGO	Identificador Batalla de Boyacá	Dummy (1 - 0)

Variable	Descripcion	Tipo dato
8_DIC	Identificador Día de la Virgen	Dummy (1 - 0)
24_DIC	Identificador 24/Diciembre	Dummy (1 - 0)
25_DIC	Identificador 25/Diciembre	Dummy (1 - 0)
31_DIC	Identificador 31/Diciembre	Dummy (1 - 0)
1_ENE	Identificador 01/Enero	Dummy (1 - 0)
2_ENE	Identificador 02/Enero	Dummy (1 - 0)
1_MAY	Identificador Día del Trabajo	Dummy (1 - 0)
LUNESV	Identificador Lunes semana vacaciones (Junio)	Dummy (1 - 0)
MARTESV	Identificador Martes semana vacaciones (Junio)	Dummy (1 - 0)
MIERCOLESV	Identificador Miercoles semana vacaciones (Junio)	Dummy (1 - 0)
JUEVESV	Identificador Jueves semana vacaciones (Junio)	Dummy (1 - 0)
VIERNESV	Identificador Viernes semana vacaciones (Junio)	Dummy (1 - 0)
SABADOV	Identificador Sabado semana vacaciones (Junio)	Dummy (1 - 0)
DOMINGOV	Identificador Domingo semana vacaciones (Junio)	Dummy (1 - 0)
SAALFRY	Identificador Sabado antes de Fiesta de Reyes	Dummy (1 - 0)
DOALFRY	Identificador Domingo antes de Fiesta de Reyes	Dummy (1 - 0)
LFYR	Identificador Lunes Fiesta de Reyes	Dummy (1 - 0)
LUNESD	Identificador Lunes semana vacaciones (Diciembre)	Dummy (1 - 0)
MARTESD	Identificador Martes semana vacaciones (Diciembre)	Dummy (1 - 0)
MIERCOLESD	Identificador Miercoles semana vacaciones (Diciembre)	Dummy (1 - 0)
JUEVESD	Identificador Jueves semana vacaciones (Diciembre)	Dummy (1 - 0)
VIERNESD	Identificador Viernes semana vacaciones (Diciembre)	Dummy (1 - 0)
SABADOD	Identificador Sabado semana vacaciones (Diciembre)	Dummy (1 - 0)
DOMINGOD	Identificador Domingo semana vacaciones (Diciembre)	Dummy (1 - 0)
DDM	Identificador día de la madre	Dummy (1 - 0)
DIAPAPA	Identificador día de la padre	Dummy (1 - 0)
AH1-24	Demanda de las 24 horas del día anterior a la predicción	Float
OH1-24	Demanda de las 24 horas de una semana antes de la predicción	Float

Tabla 6 Detalle variables del histórico de demanda a incluir en el modelo.

Temperatura promedio en la región

Teniendo en cuenta que la correlación entre la temperatura y la demanda de energía eléctrica es de 0,41 (Tabla 7), se definen como variables de entrada del modelo el promedio mensual de la temperatura medida en cada una de las 18 estaciones meteorológicas de la región en estudio; se repite el dato mensual para cada uno de los días de cada mes y para los meses en que no se tiene registro de la medición se asigna el promedio histórico de dicho mes en el periodo seleccionado.

Variable	Descripción	Tipo dato
Tp_Estación 1-18	Temperatura promedio mensual para cada una de las 18 estaciones de medición disponibles en la región	Float

Tabla 7 Detalle variables de temperatura promedio en la región a incluir en el modelo.

Precipitaciones totales en la región

De acuerdo con lo observado gráficamente en la Ilustración 12, se evalúa la correlación entre la demanda de energía eléctrica y las precipitaciones totales mensuales (Ilustración 14) arrojando como resultado 0,062, con lo cual descartamos la inclusión de las mediciones de precipitación mensuales para el modelo.

```
[ ] corr_matrixm = dftpm.corr()
    corr_matrixm["Dem_Prom"].sort_values(ascending=False)

Dem_Prom      1.000000
Temp_Prom     0.410784
Prec_Tot      0.062828
Name: Dem_Prom, dtype: float64
```

Ilustración 14 Correlación entre los promedios mensuales de demanda energética, temperatura promedio y precipitaciones en la región.

Producto Interno Bruto (PIB)

Se corrobora lo identificado gráficamente en la Ilustración 13 calculando la correlación entre el PIB y la demanda de energía eléctrica obteniendo como resultado una correlación total (Ilustración 15) por lo que se establece como variable de entrada del modelo el dato agregado de producto interno bruto anual de la región, repitiendo el dato anual para cada uno de los 365 días del año correspondiente.

```
[ ] corr_matrixy = dftpy.corr()
    corr_matrixy["Tot_Energy_Dem"].sort_values(ascending=False)

Tot_Energy_Dem  1.000000
PIB REGION     1.000000
Year           0.411087
Name: Tot_Energy_Dem, dtype: float64
```

Ilustración 15 Correlación entre la demanda energética total por año y el PIB anual de la región.

Variable	Descripcion	Tipo dato
PIB_Capital	PIB anual para la capital de la región de estudio	Float
PIB_Departamento	PIB anual para el departamento de la región de estudio	Float

Tabla 8 Detalle variables del PIB en la región a incluir en el modelo.

Desarrollo del modelo

A continuación, se describe el paso a paso en la programación y compilación del modelo, haciendo un paneo desde la plataforma seleccionada para la programación, hasta la obtención de las predicciones.

Herramienta de programación.

Para la programación se utilizó la plataforma libre Colaboratory de google, que es un entorno de notebook de Jupyter gratuito que no requiere ninguna configuración, descarga ni instalación y se ejecuta completamente en la nube.

Colaboratory permite escribir y ejecutar código, guardar y compartir análisis, y acceder a recursos informáticos potentes, todo de manera gratuita y desde el navegador en una página con entorno interactivo utilizando lenguaje Python.

Creación del DataFrame con las variables de entrada y salida

Luego de la limpieza de datos y definición de variables tipo Dummy, se realiza el cargue de datos en la plataforma COLAB, teniendo como información general la siguiente:

```
[ ] dft.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2312 entries, 0 to 2311
Columns: 162 entries, H_1 to Pp_Villa
dtypes: float64(110), int64(52)
memory usage: 2.9 MB
```

Ilustración 16 Características generales de los datos cargados en COLAB

Así mismo las características estadísticas de los datos cargados son los siguientes:

	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_10	H_11	H_12
count	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000
mean	1221.591095	1162.342717	1134.045881	1140.273280	1237.309623	1427.104402	1511.533402	1627.744485	1752.063272	1821.548377	1880.829393	1913.269053
std	75.362875	73.416337	73.619560	79.794867	119.267047	231.616601	253.101555	240.950127	228.764340	210.146253	205.750413	204.933298
min	855.000151	809.960487	794.172208	795.162440	805.975758	804.723388	734.592201	749.585318	814.967123	887.469136	954.492299	993.884148
25%	1173.054740	1115.341665	1087.054345	1087.472890	1162.636620	1265.948165	1389.074729	1555.456033	1709.928019	1794.465108	1852.400233	1883.445327
50%	1232.924546	1175.985583	1148.202753	1156.151147	1260.935283	1500.232560	1621.058993	1726.333278	1842.467199	1903.857750	1962.009172	1992.476743
75%	1277.591115	1218.843815	1191.760400	1203.488610	1330.225249	1622.180987	1704.555024	1801.768233	1908.523382	1958.803680	2015.230740	2049.683196
max	1405.583310	1331.568030	1316.524200	1291.231640	1456.174440	1773.163735	1862.907770	1944.128880	2056.694450	2110.674380	2166.479160	2208.054110

8 rows x 162 columns

	H_13	H_14	H_15	H_16	H_17	H_18	H_19	H_20	H_21	H_22	H_23	H_24
count	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000	2312.000000
mean	1883.844183	1839.444262	1819.481001	1803.919628	1782.186202	1787.335770	1978.735058	2023.831157	1938.462022	1758.685169	1525.305988	1340.044046
std	190.382582	190.357558	217.441898	232.239825	230.810839	211.698490	190.785330	173.850368	157.023379	134.139057	103.291455	85.202597
min	1000.000000	996.928029	962.155272	933.196167	932.149815	974.362892	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	937.488841
25%	1849.659383	1781.345020	1704.206861	1666.385818	1638.284784	1653.506570	1856.786130	1899.956698	1827.694227	1669.612216	1463.828862	1291.614506
50%	1955.000560	1911.483745	1918.463688	1913.496795	1892.135905	1864.784760	2040.117992	2090.081930	1993.741535	1798.252202	1548.746170	1355.320842
75%	2009.885352	1972.273565	1979.250860	1975.894140	1952.807881	1935.124035	2113.658847	2155.788343	2057.098498	1857.287580	1599.751022	1401.382847
max	2184.321940	2128.584440	2141.327360	2145.191700	2155.487900	2199.443820	2343.149720	2336.272740	2241.099160	2037.481390	1761.061360	1552.662550

Ilustración 17 Características estadísticas de la demanda eléctrica horaria de 2012 a 2017.

Separación (Train – Test) y normalización de los datos.

La separación de la data se realiza por medio de la librería `train_test_split` y definiendo como parámetros de separación que el 33% del total de registros serán reservados y ocultos al modelo de entrenamiento para probar las predicciones finales.

De acuerdo a la naturaleza de los datos recopilados se hace necesario realizar proceso de normalización para ajustar los valores medidos y recopilados en diferentes escalas a una escala común, para lo anterior se realiza la estandarización de los datos por medio de la librería `StandardScaler` desde el módulo `sklearn.preprocessing`, esta estandarización transforma los datos eliminando la media y escalando a la varianza de la unidad. (Müller & Guido, 2016)

```
#Separación Train and Test
import numpy as np
from sklearn.model_selection import train_test_split
hora=1
while hora<25:
    Y=dft["H_"+str(hora)]
    X=dft.drop(["H_1", "H_2", "H_3", "H_4", "H_5", "H_6", "H_7", "H_8", "H_9", "H_10", "H_11",
    X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.33, random_state=42)

#Estandarizacion de la data con StandardScaler
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

X_train_esc = scaler.fit_transform(X_train)
X_test_esc = scaler.fit_transform(X_test)
```

Ilustración 18 Separación de datos de prueba y entrenamiento y normalización de los datos.

Compilación del modelo y obtención de las predicciones discretizadas por hora.

Con la librería `sklearn.ensemble` se importa el módulo `RandomForestRegressor` con la siguiente configuración de parámetros:

- `n_estimators = 50` (Cantidad de árboles entrelazados)
- `Random_state = 50` (Semilla de origen del modelo, con este parámetro se garantiza homogeneidad para cada iteración del modelo)

```
[ ] #Entrenamiento Modelo
from sklearn.ensemble import RandomForestRegressor
from sklearn.datasets import make_regression

mpd = RandomForestRegressor(n_estimators = 50, random_state = 50)
mpd.fit(X_train_esc, y_train.ravel())

#Predicción luego del entrenamiento
y_pred = mpd.predict(X_test_esc)

if hora==1:
    dtp=X_test
    exec('H_{0} = y_test'.format(hora))

#y_pred_inv = scaler.inverse_transform(y_pred)
exec('PH_{0} = y_pred'.format(hora))

#Calculo del error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
e_sqr=mean_squared_error(y_test, y_pred)
e_abs=mean_absolute_error(y_test, y_pred)
if hora==1:
    Error = {'hora':[hora], 'eabs':[e_abs], 'esqr':[e_sqr]}
    E=pd.DataFrame(data=Error)
else:
    ne = {'hora':[hora], 'eabs':[e_abs], 'esqr':[e_sqr]}
    E = E.append(ne, ignore_index=True)
```

Ilustración 19 Compilación e iteración del modelo de entrenamiento y predicción de la demanda energética.

Por medio de los operadores *if* y *while* se realiza la iteración de modelos independientes pero homogéneos para cada una de las horas de cada día.

RESULTADOS DEL MODELO

Luego del entrenamiento y predicción de los datos por medio del método *Random Forest Regressor* e iterando el modelo para cada una de las 24 horas del día, con 50 árboles entrelazados se obtienen los resultados de la predicción para los 763 días separados como prueba, en la Ilustración 20 se observa la comparación grafica entre la demanda real y la predicción para el 06 de abril de 2018 escogido aleatoriamente.

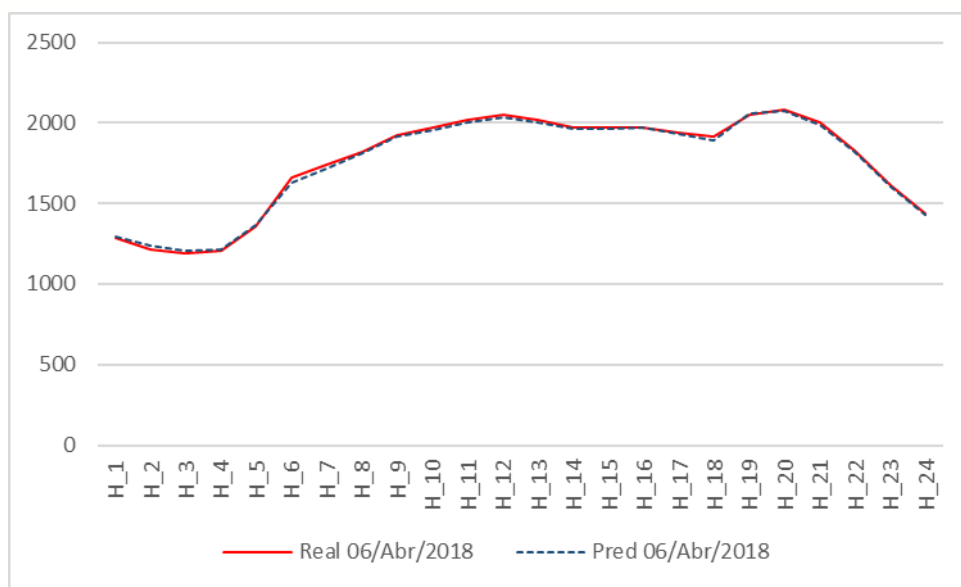


Ilustración 20 Demanda real y predicción de energía para el día 06 de Abril de 2018.

De acuerdo con lo anterior y según la separación realizada para la prueba del modelo, se realizó la predicción de la demanda de energía para cada una de las 24 horas de 763 días, se realiza el cálculo de desviación de acuerdo con lo estipulado en la resolución CREG 100 de 2019 y descrita en la Ecuación 2 del presente documento.

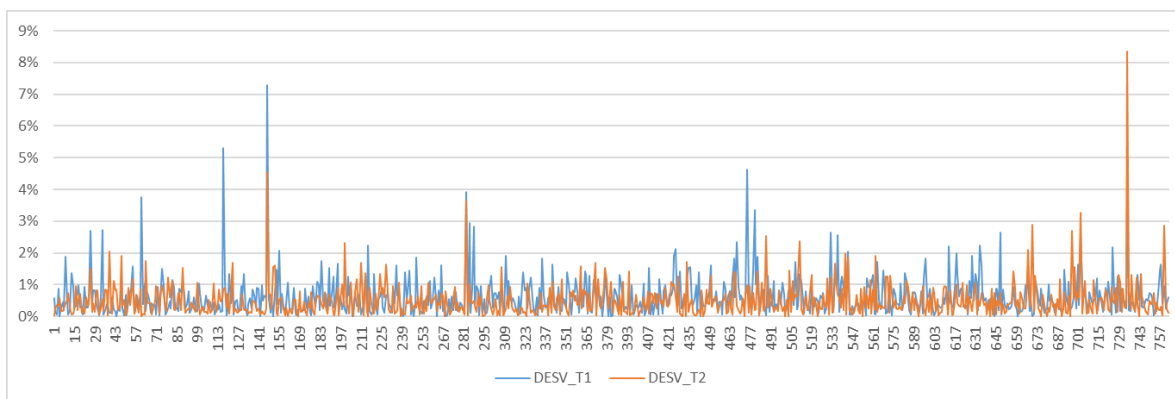


Ilustración 21 Desviación por tramo de la predicción de demanda para 763 registros de prueba

En la Ilustración 22 se observa la predicción realizada para el día 04 de julio de 2014, en el cual se tiene una desviación en el tramo 2 de 8.33%, la más alta de todo el modelo.

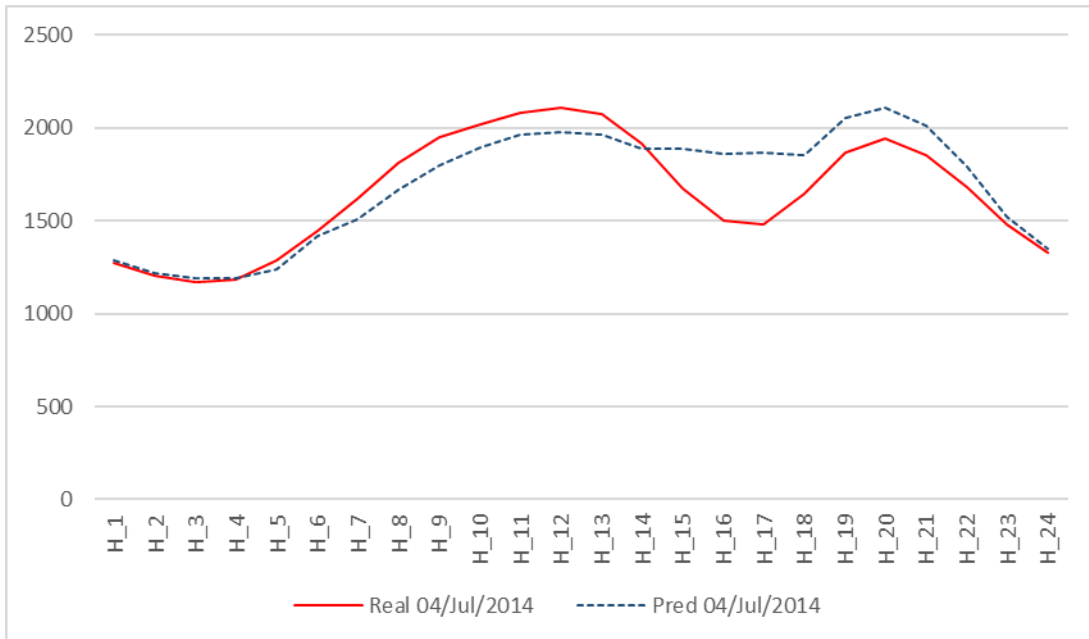


Ilustración 22 Demanda real y predicción de energía para el día 04 de Julio de 2014.

Según el cálculo realizado se tiene un promedio de desviación neto inferior al 1 %, para efectos del cálculo de penalizaciones por desviación (restando la tolerancia permitida de 4% en la desviación), con el modelo propuesto se obtiene un rango de ventaja a favor de 3.37% en los pronósticos, lo que conlleva a asegurar que se evitaran las sanciones por parte del ente regulador.

	Desv_Prom_T1	Desv_Prom_T2
Neta	0.6210%	0.4972%
Restando Tolerancia del 4%	-3.3790%	-3.5028%

Tabla 9 Calculo de desviaciones del pronóstico de demanda según el modelo propuesto.

CONCLUSIONES

De acuerdo con el análisis de datos realizados y el desarrollo del modelo propuesto se pueden llegar a las siguientes conclusiones:

- De acuerdo con la Ilustración 2, se identifica que la demanda de energía eléctrica tiene un patrón que se repite año tras año, por lo cual se hace oportuno incluir los datos de mes y día en las variables independientes del modelo.
- En la Ilustración 3 podemos concluir que existe una correlación entre la época del año y la demanda de energía eléctrica.
- Existe una relación directa entre la operación rutinaria de la industria y el consumo de energía eléctrica.
- Las semanas típicas a lo largo del año siguen un comportamiento similar de acuerdo a la hora del día, teniendo horas pico y horas valle claramente plasmados en la Ilustración 4 y en la Ilustración 5.
- Por medio de la Ilustración 6 y la Ilustración 7 se evidencia la particularidad de los días festivos y su impacto en el consumo de energía incluso en los días previos y posteriores, lo que hace necesario tener variables dentro del modelo que representen de manera adecuada el tipo de día como variable independiente que puede aportar a mejorar la predicción de la demanda energética.
- La disponibilidad de datos de temperatura y precipitaciones en la región, permitió evaluar en el modelo la pertinencia de su inclusión como variables independientes para la predicción de la demanda de energía eléctrica.
- De acuerdo con la Ilustración 12 no se evidencia relación gráfica entre las precipitaciones en la región y la demanda de energía eléctrica, lo cual se evidencia con el resultado de correlación de 0,0628 entre estas dos variables expuesto en la Ilustración 14.
- Los niveles de precipitación en la región no impactan en la demanda de energía eléctrica y su eliminación del modelo de entrenamiento y pruebas permitió obtener mejores resultados de la predicción.
- Gráficamente se observa en la Ilustración 10 una correlación entre la temperatura promedio de la región y el consumo de energía eléctrica, dicho aumento en el consumo de energía eléctrica durante los meses con mayor temperatura promedio se puede atribuir al aumento en el uso y la potencia de equipos de refrigeración y ventilación. Estadísticamente se realiza el cálculo de dicha correlación dando como resultado 0,410, que nos permite incluir como variable dentro del modelo los datos de temperatura promedio por estación meteorológica.
- En la Ilustración 13 se evidencia gráficamente una correlación perfecta entre el PIB de la región y el consumo de energía eléctrica, lo cual sugiere que el cálculo del PIB de la región se basa casi de manera exclusiva en la energía consumida por el conjunto de industria, comercio y residencias. Lo anterior se ratifica en el cálculo de la correlación expuesto en la Ilustración 15.
- La herramienta libre de programación y código COLAB de GOOGLE, seleccionada como plataforma de programación, brinda versatilidad, accesibilidad y dinámica

únicas para el desarrollo de modelos matemáticos y estadísticos basados en lenguaje Python.

- La estandarización de los datos de nuestro modelo es relevante para obtener resultados adecuados debido al tipo de datos incluidos en el modelo y sus diferencias de escala marcadas drásticamente.
- El modelo propuesto disminuye la desviación promedio de la mayoría de modelos de predicción de demanda del país de las diferentes distribuidoras de energía eléctrica.
- Existen oportunidades de mejora dentro del modelo propuesto mediante el análisis y selección de los mejores parámetros en la separación de datos de entrenamiento y prueba, así como la optimización de parámetros del modelo de entrenamiento (Random Forest Regressor).
- Para garantizar la mejor predicción de demanda de energía eléctrica se deben recopilar datos y buscar las mejores predicciones diarias de temperatura promedio de la región, garantizando así mantener las desviaciones estimadas del modelo.
- Para la compañía energética de occidente el uso del modelo propuesto representaría un ahorro anual de \$ 196'510,280 al no incurrir en penalizaciones por desviaciones fuera de lo permitido en las regulaciones propuestas.

BIBLIOGRAFÍA

- Amador, L., & Ramírez, R. (2011). La Ley 142 de 1994 desde una perspectiva de análisis económico del derecho. *Con-Texto • Revista de Derecho y Economía*, 34.
- Bello Rodríguez, S. P., & Beltrán Ahumada, R. B. (2010). *Caracterización y pronóstico del precio spot de la energía eléctrica en Colombia*. 6, 293–316.
- Cabezón Manchado, M. (2018). *Predicción demanda eléctrica española. Implementación de redes neuronales recurrentes en Python*.
- CDEC-SIC. (2013). Sistema Interconectado Central. *Mapa SIC*, 1.
- Comisión de Regulación de Energía y Gas. (2007). *DEFINICIÓN DE LA FÓRMULA TARIFARIA DE PERIODO TARIFARIO DOCUMENTO CREG-*.
- Comisión de Regulación de Energía y Gas. (2019). *CREG 034 de 2019: Por la cual se ordena hacer público un proyecto de resolución de consulta, “Por medio de la cual se hacen modificaciones a las Resoluciones CREG 025 de 1995, CREG 121 de 1998, CREG 062 de 2000 y CREG 034 de 2001 y se establecen otras dis.*
- Comisión de Regulación de Energía y Gas. (2008). *Creg083-2008.pdf*.
- Comisión de Regulación de Energía y Gas. (2019). *Modificaciones a las Resoluciones CREG 025 de 1995 y CREG 063 de 2000 y se establecen otras disposiciones.*
- Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O’Reilly Media.
- Ministerio de Minas y Energía. (1994). Ley 142 de 1994. *Diario Oficial*.
- Ministerio de Minas y Energía. (1994). Ley 143 de 1994. *Diario Oficial*.
- Ministerio de Minas y Energía. (2017). *MinMinas*. Ministerio de Minas y Energía.
- Müller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python* (Dawn Schanafelt (ed.); First Edit). O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. O’Reilly.
- Pavlov, Y. L. (2019). Random forests. *Random Forests*, 1–122.
https://doi.org/10.1007/978-3-662-56776-0_10
- Redacción El Tiempo. (1995). Una bolsa de energía para los colombianos. *Periódico El Tiempo*. <https://www.eltiempo.com/archivo/documento/MAM-372460>

- Rodríguez Gómez, J. C., Acosta Peñaloza, C. E., Ramírez, H. H., & Villamizar, N. L. (1999). *Historia de la empresa de energía de Bogotá 1896-1927: Vol. I*.
- Takefuji, Y., & Shoji, K. (n.d.). *Effectiveness of ensemble machine learning over the conventional multivariable linear regression models*. 4.
- Tudela, G. N. S. (2011). Análisis y pronóstico de la demanda de potencia eléctrica en Bolivia: una aplicación de redes neuronales. *Revista Latinoamericana de Desarrollo Económico*.
- Vélez Álvarez, L. G. (2019). *Génesis del mercado eléctrico Colombiano*. Alponiente. <https://alponiente.com/genesis-del-mercado-electrico-colombiano/>
- XM S.A. ESP. (2017). *2-5-contratos-de-largo-plazo-registrados*. Xm.Com.Co. <http://informesanuales.xm.com.co/2017/SitePages/operacion/2-5-contratos-de-largo-plazo-registrados.aspx>