

II Jornadas Internacionales de Estadística Aplicada 5 y 6 de Diciembre de 2019

Técnicas de Minería de Datos aplicadas al Análisis de Curvas de Demanda de Potencia de San Salvador de Jujuy y el Gran Buenos Aires

Autores: Pablo Nicolás Ramos, Melisa Rocío Valdiviezo

Facultad de Ingeniería, Universidad Nacional de Jujuy,
San Salvador de Jujuy

Email: pablonicolasr777@gmail.com, melisan8012@gmail.com
Teléfono: (0388-154207007), (0388-154586212)

RESUMEN.

Uno de los principales problemas para modelar el consumo de energía eléctrica en un lugar determinado, consiste en la extracción del conocimiento cuando éste se encuentra almacenado en grandes volúmenes de información como, por ejemplo, registros históricos. De acuerdo con esta representación, cada hecho ocurrido y registrado está compuesto por tres de componentes (t, P, T) en donde t representa el tiempo en el que se registró la muestra y P representa la potencia eléctrica consumida en ese instante y T es la temperatura registrada en el mismo momento. El registro diario cuenta con N casos que representa cada una de las parejas de estímulo-respuesta conocidas. Los objetivos de este trabajo consisten en el estudio cualitativo de las curvas de demanda de energía en la provincia de Jujuy y el gran Buenos Aires; y hallar una función que permita mapear el vector de variables de entrada T al vector de variables de salida P ; donde F es una función cualquiera, en este caso el consumo de energía eléctrica. Se utiliza la Regresión Polinomial para obtener el modelo representativo para cada conjunto de datos.

Palabras Claves: Curva de demanda, Curva de Potencia, Minería de Datos, Predicción, RMSE.

1. Introducción

La calidad de servicio de un sistema eléctrico (Grainger, 2002) puede cuantificarse a través de varios parámetros relacionados con: la continuidad del servicio, las fluctuaciones de voltajes, el contenido armónico de las formas de onda de voltaje y de corriente, variaciones de frecuencia, y la regulación. El concepto de calidad del servicio es bastante amplio, de manera que no es posible sintetizarlo en un solo parámetro o índice. Un parámetro comúnmente utilizado para evaluar la calidad del servicio, consiste en monitorear periódicamente los registros de consumo de energía eléctrica. El estudio de su comportamiento sirve para tomar medidas que garanticen el rendimiento óptimo del sistema en su conjunto.

En este trabajo de investigación se analiza el efecto que tienen las variaciones de demanda de energía en la provincia de Jujuy y en el gran Buenos Aires. El modelo se realizó mediante el software Rapidminer®, donde se representa la curva de demanda del consumo diario para dos días determinados. Después de evaluar la curva de demanda de potencia para las dos provincias, se

observa que hay cambios significativos en el conjunto de datos de la provincia de Jujuy debido al mal aprovechamiento de las instalaciones eléctricas.

2. Metodología

El estudio del problema puede ser definido del tipo “descriptivo” porque se busca analizar **cómo es y cómo** se manifiesta un fenómeno (en este caso, el estudio de las Curvas de Demanda de Potencia en las provincias de Jujuy y Buenos Aires) y, sus componentes; permitiendo detallar dicho suceso a través del análisis y medición de sus atributos, buscando describir una situación en particular (Vargas et al., 2016).

El método de investigación utilizado se describe a continuación: 1) Observación, considerada una técnica científica que debe servir a un objetivo formulado, planificada de manera sistemática de manera que se especifique y defina lo que se pretende observar, e informar los resultados. Está sujeta a ser comprobada y validada. En el presente trabajo se realizan las siguientes fases:

1) Formulación de un problema y, planificación de un observación exploratoria y confirmatoria

El presente trabajo tiene como objeto describir las curvas de demanda de potencia en las provincias de Buenos Aires y Jujuy, mediante la aplicación de técnicas descriptivas y predictivas de Data Mining.

La observación exploratoria se realiza mediante el uso de gráficos estadísticos suministrados por la herramienta de Data Mining conocida como Rapidminer.

2) Recolección de datos

Para realizar el análisis exploratorio, se obtuvo el conjunto de datos de la página oficial de CAMMESA (Compañía Administradora del Mercado Mayorista Eléctrico) (). En el mismo se detallan los datos de **temperatura** y **demandas** de energía eléctrica, en un horario determinado.

3) Análisis e interpretación de los datos observados

El análisis de datos consiste en la extracción de información descriptiva, permitiendo relacionar las variables objeto de interés. Este análisis se detalla minuciosamente.

Los archivos que se obtuvieron de la página de CAMMESA no necesitan la aplicación de la etapa de preprocesamiento de datos, ya que se encuentran sin inconsistencia ni ruidos.

Se determinan las variables cuantitativas a analizar. También se realiza una transformación de datos adecuada, según los requerimientos de los algoritmos que se apliquen.

4) Visualización de resultados

Una vez que se obtiene el modelo, se visualizan los resultados obtenidos. Se muestran gráficos estadísticos que determinan la validez del modelo obtenido.

3. Desarrollo

3.1. Minería de Datos (Data Mining, DM por sus siglas en inglés)

El crecimiento exponencial de información en la web ha hecho posible que la información digitalizada sea fácil de capturar, almacenar, distribuir, y transmitir (Riquelme et al., 2006). Con los avances en las Ciencias Informáticas, tecnologías asociadas a ella y la expansión de su uso en

diferentes aspectos cotidianos, constantemente se recogen y almacenan en base de datos, gran cantidad de información.

Descubrir conocimiento en un volumen de datos de tamaño considerable, es un verdadero reto. La Minería de Datos (DM), es una aproximación de encontrarle sentido a la expansión de información que actualmente puede ser almacenada (Mittra y Acharya, 2003).

Los datos en bruto, raramente, son beneficiosos directamente. El valor de los mismos reside en: 1) la habilidad para extraer conocimiento útil para, la toma de decisiones o la exploración, y 2) y la comprensión del problema generado por la fuente de datos. En muchas áreas, el análisis de datos se trataba de un proceso manual. Pares de analistas, con la ayuda de técnicas estadísticas, proporcionaban resúmenes y generaban informes; es decir, el analista hacía de procesador de preguntas sofisticado. Pero, dicho enfoque cambió debido al crecimiento exponencial de datos que se almacenan en las bases de datos. Cuando la escala de manipulación de datos, exploración e inferencia va más allá de la capacidad humana, se necesita la ayuda de herramientas computacionales para automatizar el proceso (Riquelme et al., 2006).

Entonces se debe aplicar una metodología realice un análisis inteligente de datos, la cual pueda descubrir conocimiento útil en bases de datos. El Descubrimiento de Conocimiento en Base de Datos (KDD, por sus siglas en inglés), acuñado en 1989, se refiere a todo el proceso de extracción de conocimiento a partir de una base de datos; en donde la importancia del proceso reside en ser capaz de descubrir patrones desconocidos entre los datos. El KDD es un área de estudio basada en la Estadística, Sistema de Información, Aprendizaje Automático o Programación Paralela y que engloba al proceso que pretende extraer conocimiento a partir de los datos. Este proceso se define como:

Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos (Hernández Orallo, 2004).

En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído:

- Válido: hace referencia a que los patrones deben seguir siendo precisos para datos nuevos (con un cierto grado de certidumbre), y no sólo para aquellos que han sido usados en su obtención.
- Novedoso: que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.
- Potencialmente útil: la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- Comprensible: la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de su utilidad).

Los términos Minería de Datos y Descubrimiento de Conocimiento, frecuentemente, son usados indistintamente. En general se acepta que la Minería de Datos es un paso particular en el proceso KDD, y consiste en la aplicación de algoritmos de Aprendizaje Automático (ML, por sus siglas en inglés) para extraer patrones (modelos) de los datos. Otros pasos en el proceso KDD, son la preparación de los datos, la selección y limpieza de los mismos, la incorporación de conocimiento previo, y la propia interpretación de los resultados de minería. Estos pasos aplicados de una manera

iterativa e interactiva aseguran que un conocimiento útil se extraiga de los datos.

Las técnicas de la fase de minería de datos pueden ser descriptivas, (por ejemplo, descubrir patrones significativos o relaciones describiendo los datos), o predictivas (por ejemplo, clasificar un nuevo conjunto de datos, basándose en los anteriormente disponibles). En otras palabras, es un área interdisciplinar con el objetivo general de predecir las salidas y revelar relaciones en los datos (Mitra y Acharya, 2003). Para ello se utilizan herramientas automáticas que: 1) emplean algoritmos de Aprendizaje Automático (ML, por sus siglas en inglés) para descubrir principalmente patrones ocultos, asociaciones, anomalías, y/o estructuras de la gran cantidad de datos almacenados en los data warehouse u otros repositorios de información y, 2) filtran la información relevante y necesaria de las grandes bases de datos.

3.1.2. Extracción de conocimiento en bases de datos y Minería de Datos

El KDD es un proceso complejo que incluye no sólo la obtención de los modelos o patrones (el objetivo de la minería de datos), sino también la evaluación y posible interpretación de los mismos, tal como se refleja en la figura 1. Cada una de las fases del proceso KDD se sirve de una serie de herramientas para conseguir llevar a cabo su tarea. Sus fases se describen como sigue:

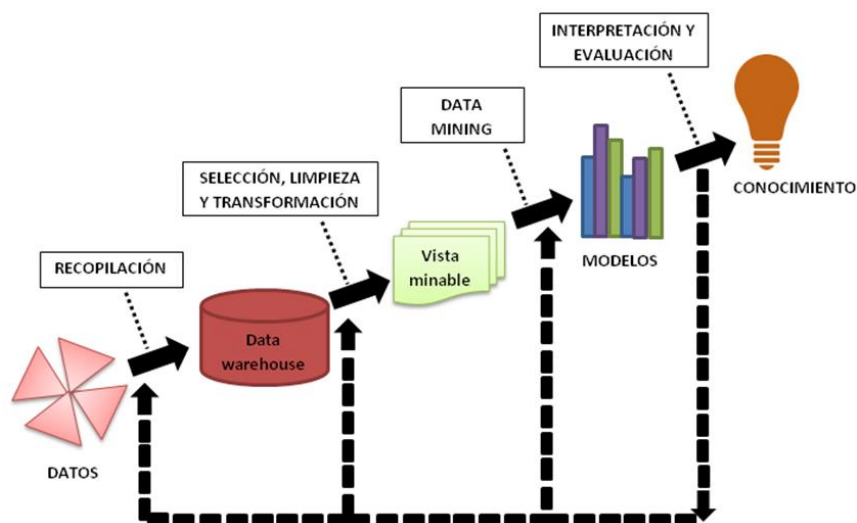


Figura 1. Proceso de KDD (Lara Torralbo, 2016)

- 1. Recopilación de datos.** En esta fase, se trata de elegir un subconjunto de variables o datos, procedentes de diferentes fuentes, que se integrarán en un mismo y único repositorio de datos, denominado almacén de datos, más conocido como data warehouse. El resultado final de esta fase es, precisamente, ese data warehouse. Esto se realiza con el fin de eliminar valores redundantes e inconsistentes en los datos de varias fuentes al juntarlos dentro de una sola base de datos.
- 2. Selección, limpieza y transformación de datos.** Sobre los datos recopilados en el almacén de datos no es posible realizar aún data mining, debido a que dichos datos pueden no estar limpios, pueden contener atributos irrelevantes, etc. Por ello, en la segunda fase del proceso

de KDD se realiza una selección de los datos integrados en el data warehouse. Esta fase incluye operaciones básicas sobre los datos, como el filtrado para reducir ruido, crear campos explícitos con relaciones entre los atributos conocidos que puedan hacer el análisis más sencillo, etc. Dichos datos se limpian y transforman de cara a fases posteriores. El resultado de esta fase es la denominada “vista minable”, que es un subconjunto limpio y transformado de los datos sobre el que ya se pueden aplicar las técnicas de data mining en la siguiente fase.

3. **Minería de datos:** fase principal del proceso KDD. Se aplican métodos inteligentes para extraer patrones de datos. Se eligen los algoritmos de minería adecuados en función de los datos y del tipo de información que se quiere descubrir.
4. **Interpretación y evaluación de modelos.** Los modelos obtenidos en la fase de data mining han de ser evaluados. Se trata de entender los resultados del análisis y sus implicaciones. Esto puede llevar a regresar a alguno de los pasos anteriores. Una vez comprobada la calidad de dichos modelos y de ser interpretados, a partir de ellos se obtiene el conocimiento. Este conocimiento.

El paso fundamental del proceso es el número 3. A continuación se comentan brevemente las tareas más comunes de la minería de datos, se enuncia un ejemplo y los métodos más usados en el estado del arte.

- **Clasificación:** El sistema de Minería de Datos, mediante la aplicación de un algoritmo de clasificación, se encarga de asignarle una clase o etiqueta a un objeto dado. Por ejemplo, detectar si una persona tiene cáncer de pulmón o no, mediante los atributos extraídos de las imágenes de radiografías de pulmones. Métodos: Decision Tree, Classification Rule, Neural Network, entre otros.
- **Regresión:** Se usa para predecir los valores ausentes de una variable basándose en su relación con otras variables del conjunto de datos. Por ejemplo, determinar el nivel de contaminación que puede emanar de un determinado automóvil, en función de sus atributos. Métodos: Regresión lineal, Regresión logística, Regresión Multivariada, entre otros.
- **Clustering:** Tienen como objetivo asignar los datos de entrada a grupos, de manera que exista un mayor parecido entre los elementos de un mismo grupo entre los elementos de otros grupos. Requiere de una medida de similitud o distancia. Por ejemplo, Identificar alumnos con rendimientos académicos similares, con el objetivo de reducir la deserción universitaria. Métodos: k-means, Nubes Dinámicas, Clustering Jerárquico, entre otros.
- **Reglas de asociación:** Permiten expresar relaciones entre ítems de una Base de Datos. Son aplicables a la toma de decisiones. Representan cuando un suceso X implica un suceso Y . Por ejemplo, los productos que se compran en un supermercado se pueden representar como transacciones; entonces se puede deducir mediante reglas, cuando una transacción X implica una transacción Y . Métodos: A priori, FP-Growth.

3.2. Curva de carga o de demanda

La demanda de una instalación o sistema de distribución, es la carga en las terminales receptoras tomadas en un valor medio en determinado intervalo, con esta definición se entiende por carga la que se mide en términos de potencia (aparente, activa, reactiva o compleja) o de

intensidad de corriente (1). Es la representación gráfica de cómo varía la demanda o carga eléctrica en el transcurso del tiempo (por ejemplo, Figura 2).

El intervalo de tiempo elegido para realizar el análisis puede ser: diario, semanal, mensual o anual. La carga no es constante en el período analizado.

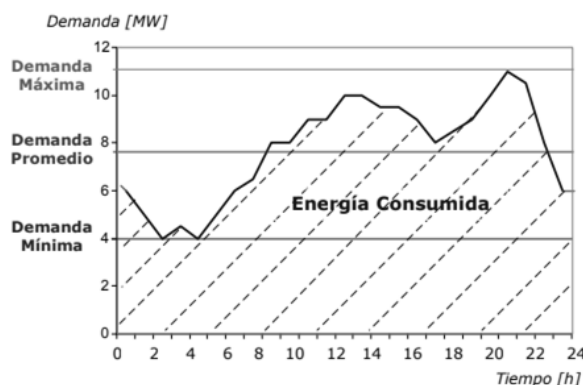


Figura 2: Curva de carga diaria

En las abscisas se representa el tiempo y en las ordenadas la potencia eléctrica demandada. El área que está por debajo de la curva formada, es la energía demandada.

La forma de la curva de carga, depende fundamentalmente si es una carga de tipo residencial, comercial, industrial, de un determinado día de la semana, de una estación en particular (invierno, verano) y de los factores climáticos (sobre todo de la temperatura).

3.2.1. Curva típica y de alta carga en Argentina

En las figuras 3 y 4, se puede ver una curva típica y de alta carga, correspondientes al invierno y verano 2015 en Argentina (3).

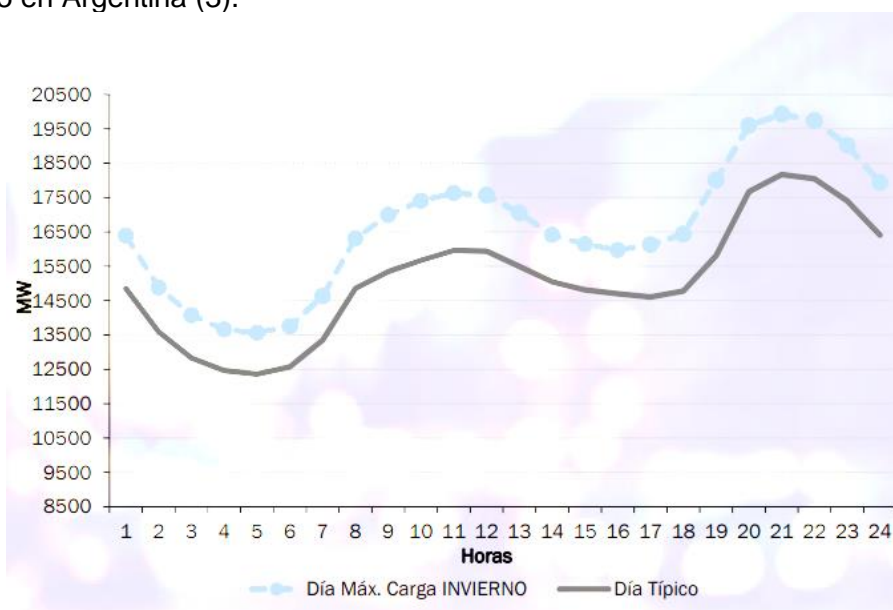


Figura 3: Curva típica y de máxima carga – INVIERNO 2015

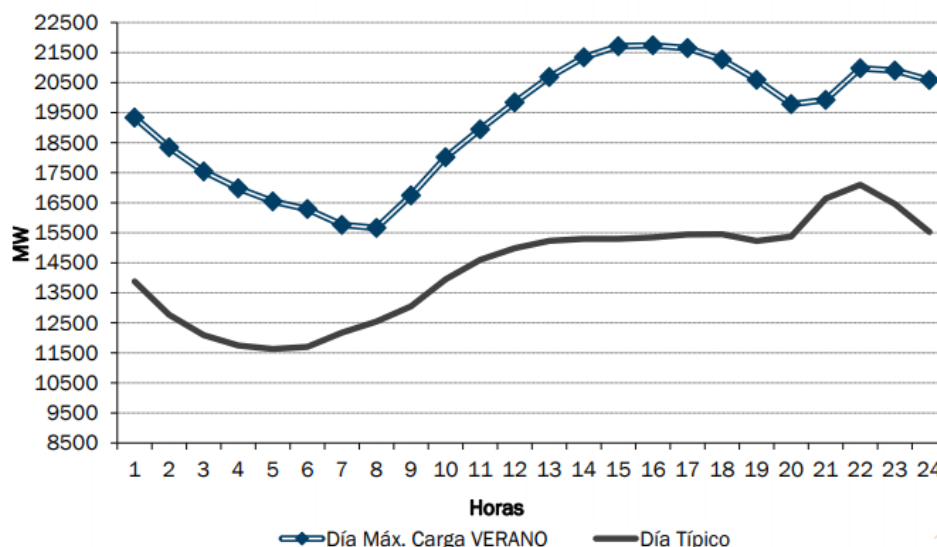


Figura 4: Curva típica y de máxima carga – VERANO 2015

3.2.2. Características de la demanda

En el contexto del funcionamiento de un sistema eléctrico de potencia, las actividades de generación, transporte y distribución cumplen la función de suministrar los requerimientos energéticos de la demanda, sin embargo atenderla constituye un desafío tecnológico que no tiene comparación con otros bienes o productos industriales, esto se debe a que las leyes físicas que rigen el comportamiento de un sistema eléctrico exigen que en todo momento exista un equilibrio entre producción y demanda, de lo contrario todo el sistema colapsaría. Esta característica obliga a que la producción se tenga que ajustar cada instante a los requerimientos de la demanda, teniendo que ingresar o salir bloques de generación.

Dependiendo de la tecnología de generación de las diferentes centrales será necesario contar con un cierto porcentaje de reserva en condiciones “standby” para cuando sea necesario, estas unidades en reserva se sincronizan con el sistema y tomen carga. El costo de esta generación “standby” o también denominada “spining reserve” o reserva rodante, es elevado y en función de la organización del sector, sus costos deben ser cubiertos por todos los agentes que se benefician de dicho servicio.

En razón de lo antes mencionado, se puede apreciar la importancia de conocer el comportamiento de la demanda, no obstante, el referido comportamiento depende de muchos factores como, por ejemplo, el tipo de segmento, si es residencial, comercial, industrial, entre otros, la zona o región geográfica donde se encuentra, la época del año, etc. Al tener muchas variables que la afectan, la predicción de su comportamiento no es posible conocerlo de manera exacta, por ello la importancia de disponer de registros históricos y de la utilización de un método de proyección que, con un grado razonable de error, permita predecir su comportamiento (Méndez Santos, 2013)

3.3. Usuarios

3.3.1. Tipos de usuarios

3.3.1.1. Residenciales

Incluye a la demanda de Distribuidores clasificada como:

- Residenciales menores o iguales 1000 KWh bimestral
- Residenciales mayores 1000 KWh y menores o iguales 1400 KWh bimestral
- Residenciales mayores 1400 KWh y menores o iguales 2800 KWh bimestral
- Residenciales mayores 2800 KWh bimestral

3.3.1.2. Menores

Incluye a la demanda de Distribuidores clasificada como:

- Menores 10 KW no residenciales con consumo Menor 4000 KWh bimestral.
- Menores 10kw no residenciales con consumo Mayor o igual 4000 KWh bimestral.
- Alambrado Público

3.3.1.3. Intermedios

Incluye a la demanda de Distribuidores clasificada como:

- Mayor o igual 10 KW y menor a 300 KW no residenciales.

3.3.1.4. Mayores

Incluye a la demanda de Distribuidores clasificada como:

- Mayor o igual a 300 KW no residenciales
- La Demanda de Grandes Usuarios Menores
- La Demanda de Grandes Usuarios Mayores [3].

3.3.2. Demanda mensual por Tipo de Usuario

La distribución de Energía Eléctrica en Argentina se divide principalmente en los sectores industrial, comercial, doméstico y de servicios, de los cuales el tipo de Usuario que más consumo de Potencia registra es el de Residenciales con un 42%, siguen Mayores (>300kW) con 29%, Intermedios (10 y 300kW) 15% y el tipo de usuario Menores (<10kW) con 13%.

3.4. La temperatura y su influencia en la demanda de Energía Eléctrica

En la figura 5, se observa como se modifica la curva de demanda de potencia en invierno, en presencia de un día frío respecto de uno moderado.

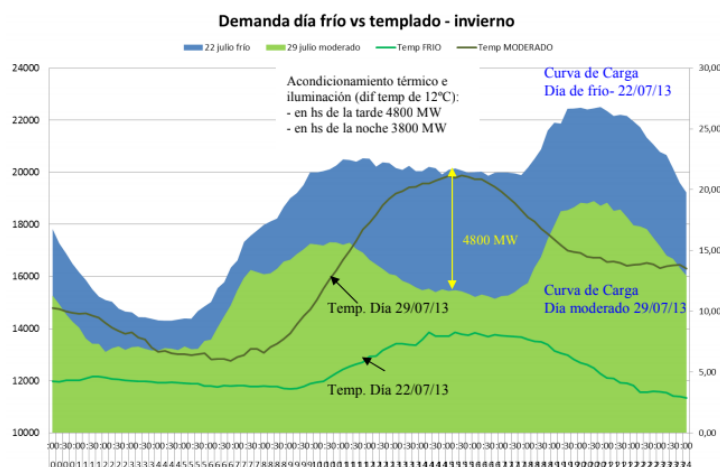


Figura 5: Influencia del acondicionador térmico en la demanda de invierno – 2013

Argentina tiene picos de demanda en verano (con altas temperaturas) y en invierno (con bajas temperaturas). Esto puede ser producto de la acumulación de aparatos de acondicionadores de aire y calefacción eléctrica. Dada la ausencia de mecanismos de incentivos al consumo eficiente, esta situación provoca un alto costo al sistema por la potencia demandada y las líneas de transporte y distribución que deben estar disponibles para ser utilizadas, con un bajo factor de utilización.

A su vez una curva de demanda típica semanal para el invierno 2015 en Argentina (Figura 6) y para el verano del 2015 (Figura 7), muestra las diferencias que existen entre los días hábiles, el sábado y el domingo.

Los trabajos relevados en la literatura encuentran una relación de no linealidad global entre la temperatura y la demanda eléctrica. Este efecto se confirma en Argentina dado que en meses invernales la relación se hace negativa mientras que en los meses estivales la relación de la demanda de electricidad con la temperatura es positiva y creciente. Ghalekhondabi (2016) realiza un relevamiento de la literatura más reciente (2005- 2015) con una descripción de la tipificación de los modelos más utilizados a nivel mundial en términos de predicción eléctrica. Si bien la mayoría de los modelos utilizan un enfoque econométrico de cross section (similar al que se presentará en las siguientes secciones) y de series de tiempo para la predicción de demanda anual, mensual, diaria y horaria, existen nuevas aplicaciones de la neurociencia alrededor del modelaje de redes neuronales artificiales (ANN, por sus siglas en inglés).

La importancia de la administración de los mercados de energía eléctrica a nivel global desencadenó en los últimos tiempos diferentes estudios que relacionan la energía eléctrica con componentes climáticas como factores importantes que determinan su nivel. La predicción de la demanda eléctrica con la mayor precisión posible resulta fundamental a la hora de administrar el mercado, dado la energía eléctrica que es consumida en un instante está siendo producida en ese mismo instante con la imposibilidad certera de poder interferir consumos.

Zotelo (2016) realiza análisis de regresión simple respecto a la energía y a la potencia demandada contrastándola con pronósticos climáticos a 3 meses con el objeto de analizar el desempeño de sus pronósticos del clima (temperaturas y olas de frío y calor) respecto al impacto en la demanda total de energía del país estimada.

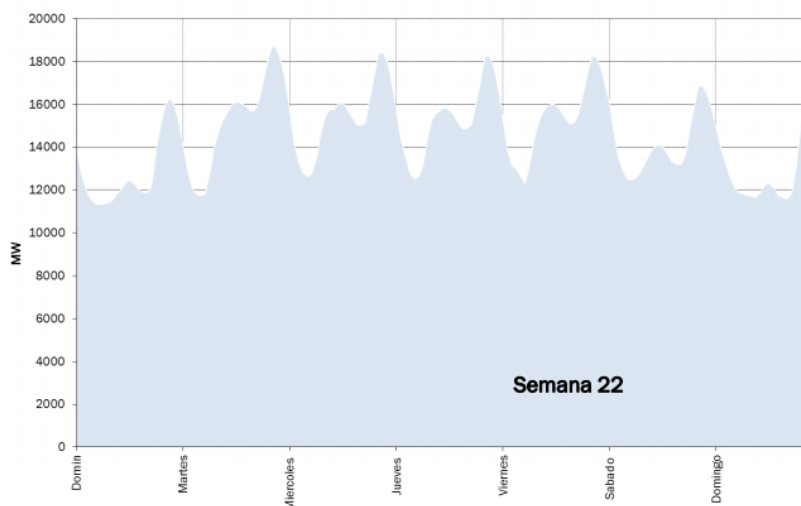


Figura 6: Demanda típica semanal de invierno

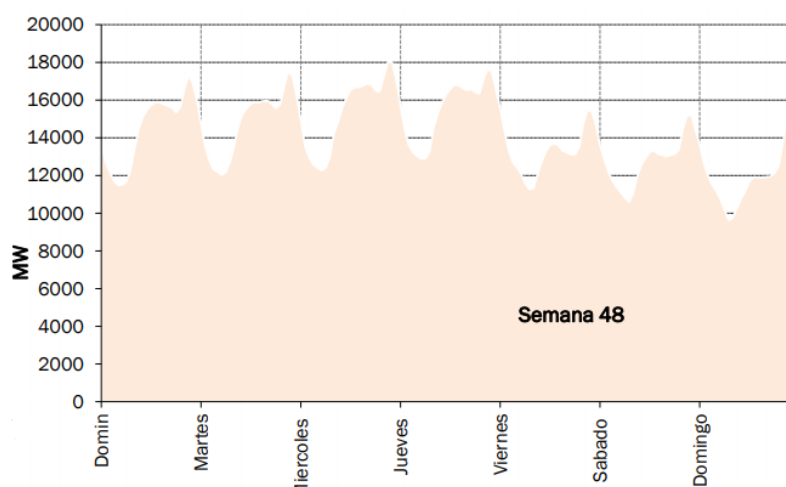


Figura 7: Demanda típica semanal de verano.

3.4.1. Relación entre temperatura y demanda de Energía Eléctrica

Si bien en la figura 8 no se observan efectos para diferentes años que computen el crecimiento interanual, se observa que para cada región a “bajas” temperaturas la demanda aumenta teniendo pendiente negativa y para “altas” temperaturas la demanda tiene una relación directa con la temperatura teniendo una pendiente positiva. Así mismo se observa gráficamente que las pendientes no son homogéneas en términos regionales lo que justifica una estimación por

distribuidora. También si se evalúa la relación general entre temperatura y energía, se puede llegar a estimar una función cuadrática, donde el componente cuadrático es positivo.

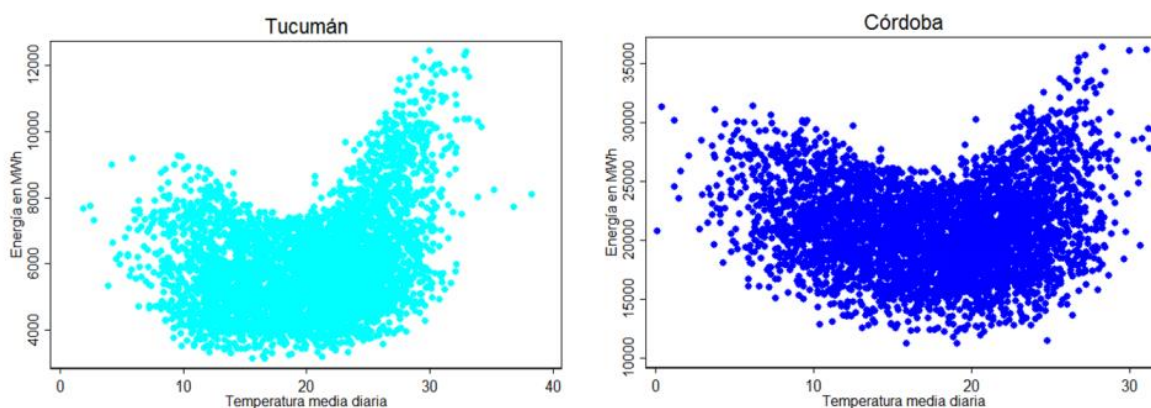


Figura 8. Partes semanales de demanda de CAMMESA e información meteorológica del SMN en algunas provincias de Argentina.

Otros factores relacionados con la actividad diaria influyen de manera directa en los niveles de energía consumida. Teniendo en cuenta las diferencias entre los días laborables y no laborables (fines de semana y feriados de cualquier índole) (Mastronardi et al., 2016).

3.5. Curvas de demanda y temperatura en Argentina

3.5.1. Análisis de exploratorio de la demanda de energía

3.5.1.1. Buenos Aires

Se analizaron los conjuntos de datos suministrados por C.A.M.E.S.A., los mismos pertenecen a la Provincia de Jujuy y Provincia de Buenos Aires. Estos datos corresponden a los días 24/11/2019 y 25/11/2019; y durante el transcurso de esos días no se registró ninguna eventualidad notable en esos sitios, es decir, no hubo un corte de luz que afecte a la distribución de Energía Eléctrica.

Se compara la demanda de potencia a través de diagramas de Boxplot (Figura 10) visualizados en Rapidminer, los cuales sirven para identificar el consumo de los distintos sitios analizados.

En la figura 9 se pueden observar los parámetros correspondientes a la demanda de energía, en el día 24/11/2019, en la provincia de Buenos Aires. Con $\bar{x} = 1933,536 \text{ kW}$ y $s = 112,555$.

En la figura 11 se pueden observar los parámetros correspondientes a la demanda de energía, en el día 25/11/2019, en la provincia de Buenos Aires. Con $\bar{x} = 1785,416 \text{ kW}$ y $s = 187,290$.

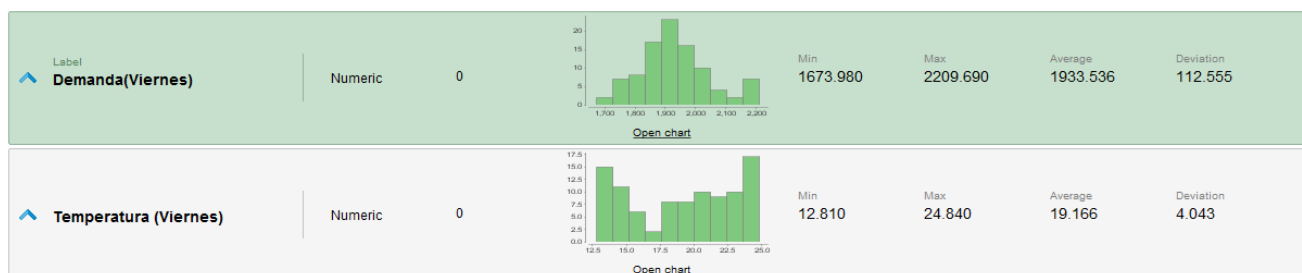


Figura 9. Parámetros (Demanda de Energía) correspondientes al día 24/11/2019 en la provincia de Buenos Aires

En el análisis exploratorio de datos no se omiten los valores atípicos porque el objeto de estudio de este trabajo es entender la naturaleza de dichos sucesos. A diferencia de la figura 10, en el diagrama de caja representado en la figura 12, se exhibe con claridad la cantidad de outliers que se reflejan en los valores de los parámetros obtenido, y que son representados en las figuras 9 y 11.

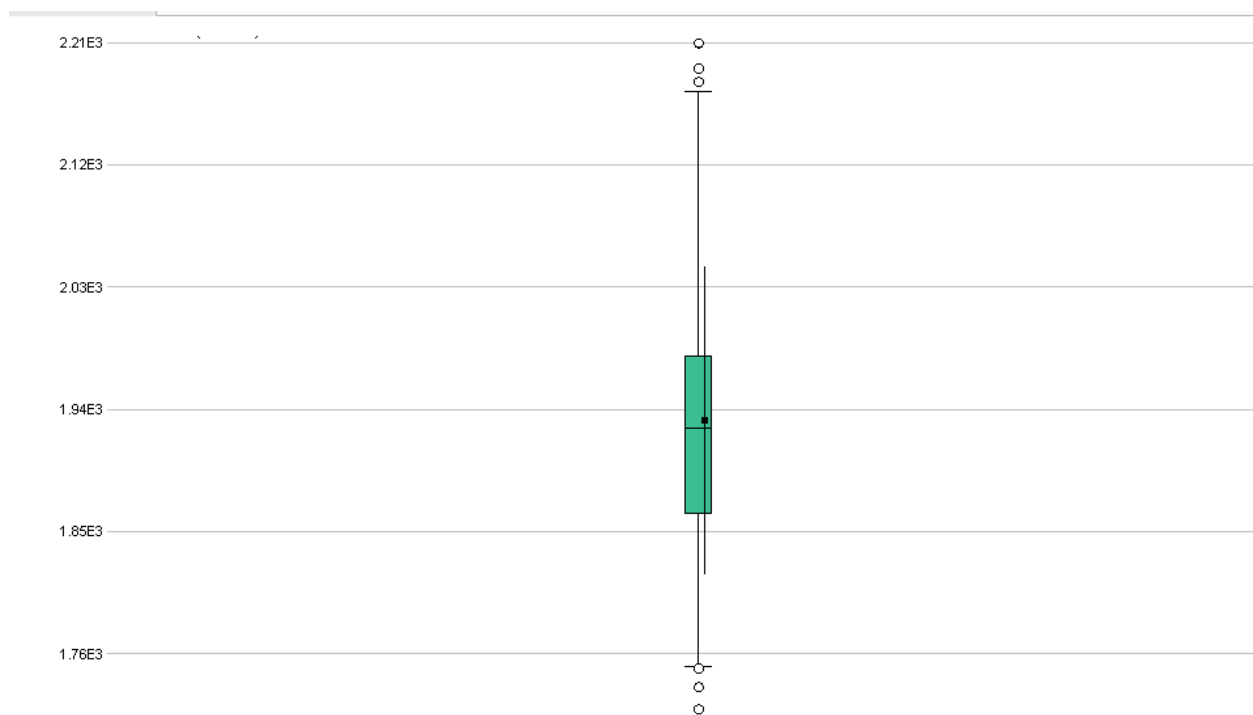


Figura 10. Boxplot de la Demanda de Energía correspondientes al día 24/11/2019 en la provincia de Buenos Aires.

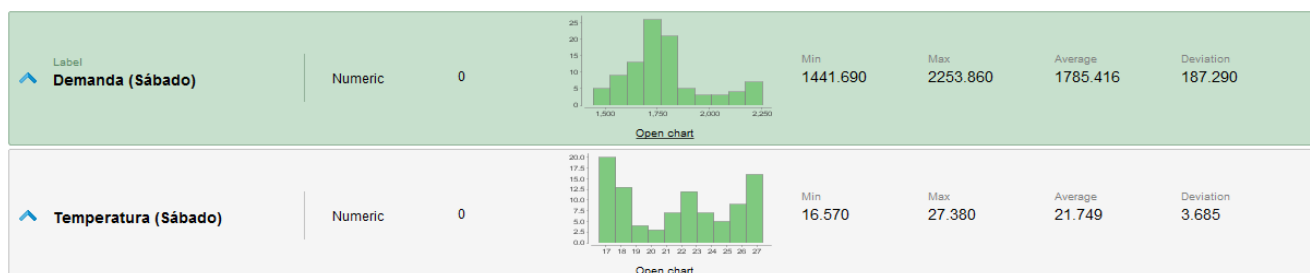


Figura 11. Parámetros (Demanda de Energía) correspondientes al día 25/11/2019 en la provincia de Buenos Aires

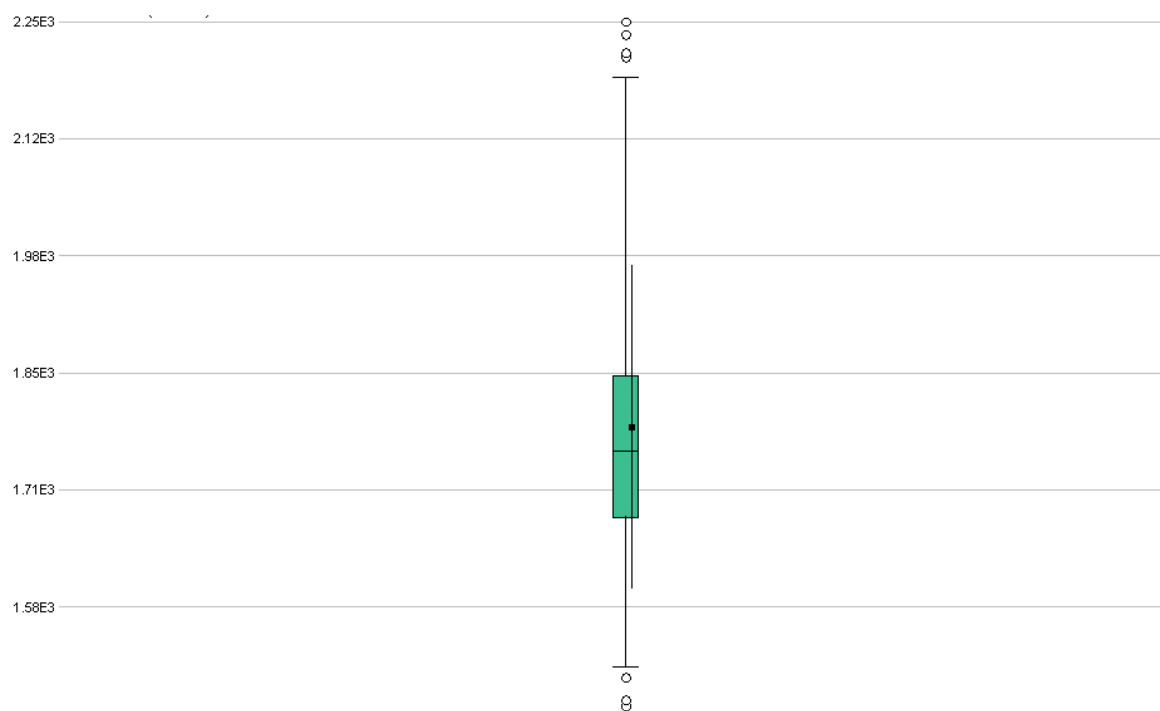


Figura 12. Boxplot de la Demanda de Energía correspondientes al día 25/11/2019 en la provincia de Buenos Aires.

3.5.1.2. Jujuy

A diferencia de Jujuy, el Gran Buenos Aires posee una demanda de potencia elevada. La diferencia radica en la cantidad de habitantes y en el sector industrial altamente desarrollado en la provincia de Buenos Aires. En la provincia de Jujuy se puede observar que la curva de demanda de potencia posee valores entre 86,130 kW y 142,38 kW; como lo muestra la figura 13.

El análisis realizado el día Sábado 25 de Noviembre (Figura 15) muestra que el rango (*valor máximo – valor mínimo*) de demanda de potencia es aproximadamente igual al del Viernes. Lo mismo ocurre con el rango de temperatura.

El día Sábado 25 presenta mayor dispersión de datos, y ello se ve reflejado en las figuras 13 y 15 respectivamente; en contraposición, la figura 14 presenta una mayor cantidad de outliers que la representación del boxplot interpretado en la figura 17.

Name	Type	Missing	Statistics			Filter (3 / 3 attributes): <input type="text" value="Search for Attributes"/>
Demanda(Viernes)	Numeric	0	Min 86.130	Max 142.380	Average 117.919	
Temperatura(Viernes)	Numeric	0	Min 12.810	Max 24.840	Average 19.189	

Figura 13. Parámetros (Demanda de Energía) correspondientes al día 24/11/2019 en la provincia de Jujuy

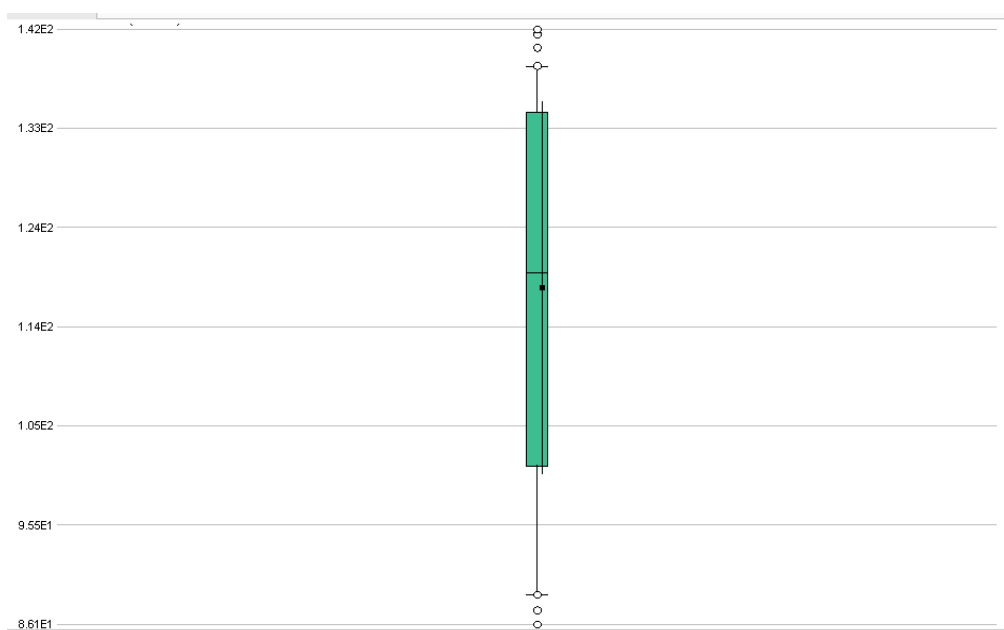


Figura 14. Boxplot de la Demanda de Energía correspondientes al día 24/11/2019 en la provincia de Jujuy.

Name	Type	Missing	Statistics			Filter (3 / 3 attributes): <input type="text" value="Search for Attributes"/>
Demanda(Sábado)	Numeric	0	Min 82.180	Max 138.830	Average 109.554	
Temperatura (Sábado)	Numeric	0	Min 16.570	Max 27.380	Average 21.780	

Figura 15. Parámetros (Demanda de Energía) correspondientes al día 25/11/2019 en la provincia de Jujuy

A través del estudio de gráficos estadísticos sencillos (Figura 16) podemos concluir que la demanda en punta puede reducirse notablemente mediante mejoras en la eficiencia de los equipos, es recomendable que las empresas distribuidoras de electricidad tomen estas medidas con adecuados diseños de tarifarias, fomentando así un uso menor de energía en las horas pico.

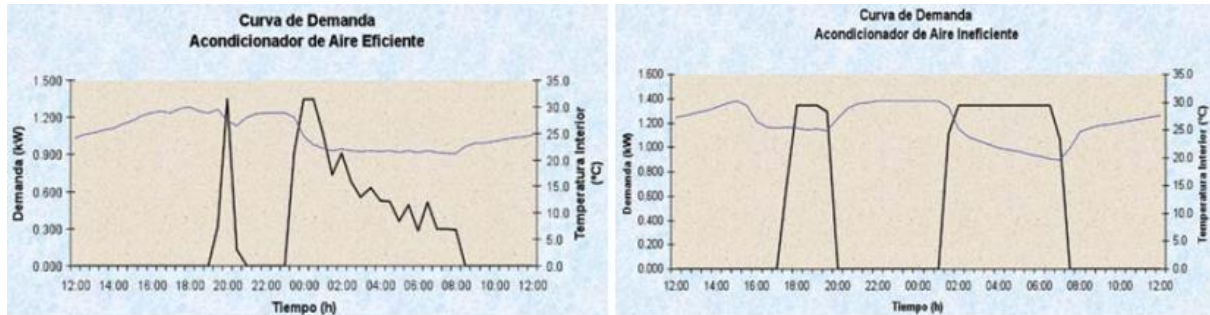


Figura 16: Demanda de equipos eficiente e ineficientemente

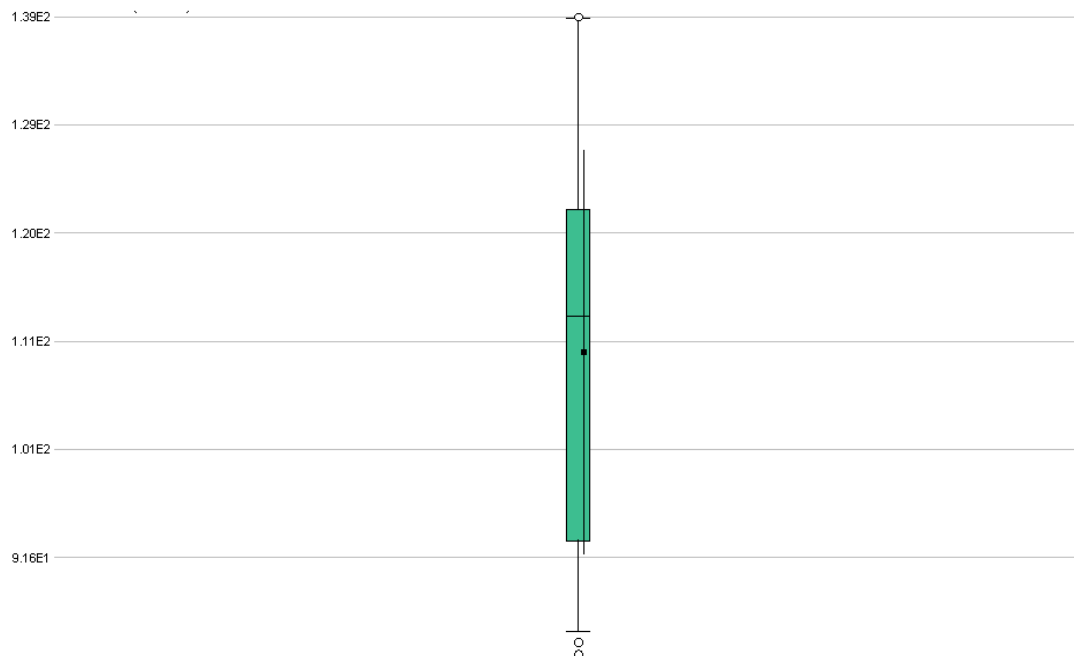


Figura 17. Boxplot de la Demanda de Energía correspondientes al día 25/11/2019 en la provincia de Jujuy.

3.5.2. Análisis de Correlación

La correlación mide el grado de asociación entre dos atributos X e Y ; y varía entre -1 y 1 (Tito y Mullicundo, 2010). Sea r_{xy} el coeficiente de correlación; el cual indica la relación entre dos atributos dados, entonces: 1) Si $r_{xy} = 0$, no hay relación entre los atributos; 2) Si $r_{xy} > 0$, indica que existe una asociación positiva entre los atributos. En este caso, los valores grandes de X tienden a asociarse con valores grandes de Y y los valores pequeños de X tienden a asociarse con valores pequeños de Y . 3) Si $r_{xy} < 0$, implica una asociación negativa o inversa. En este caso, los valores grandes de X tienden a asociarse con valores pequeños de Y y viceversa.

En las tablas 1, 2, 3 y 4; se observa la matriz de correlación para cada par de atributos (temperatura (x), demanda (y)), en Buenos Aires y Jujuy, para los días 24 y 25 de noviembre.

Tabla 1. Matriz de correlación. Buenos Aires. 24/11/2019

Atributos	Temperatura	Demanda
Temperatura	1	0.206
Demanda	0.206	1

Tabla 2. Matriz de correlación. Buenos Aires. 25/11/2019

Atributos	Temperatura	Demanda
Temperatura	1	0.326
Demanda	0.326	1

Tabla 3. Matriz de correlación. Jujuy. 24/11/2019

Atributos	Temperatura	Demanda
Temperatura	1	0.766
Demanda	0.766	1

Tabla 4. Matriz de correlación. Jujuy. 25/11/2019

Atributos	Temperatura	Demanda
Temperatura	1	0.422
Demanda	0.422	1

Las tablas indican que, en la provincia de Jujuy, los atributos tienen un mayor grado de asociación positiva que en la provincia de Buenos Aires. Una razón cuantitativa que explica dichos resultados, es la cantidad de outliers que presenta el conjunto de datos representativo de la demanda de energía en el Gran Buenos Aires. La correlación es sensible a valores atípicos.

3.5.3. Análisis confirmatorio de la curva de demanda de energía

Se parte de la siguiente hipótesis: “Las variaciones de demanda potencia son la principal causa de un mal aprovechamiento de las instalaciones”. En este apartado se utiliza como métrica de evaluación, la raíz del error cuadrático medio. La raíz del error cuadrático medio es sensible a valores atípicos (Pontius et al., 2008; Willmott y Matsuura, 2006).

En esta tarea se aplica un método de regresión polinomial utilizando validación cruzada. Los modelos obtenidos se representan mediante las ecuaciones (1), (2), (3) y (4).

Si la raíz del error cuadrático medio es igual a 0, es decir:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} = 0$$

entonces se dice que el modelo de predicción es perfecto. En otras palabras, cuanto menor sea su valor, mejor será el modelo obtenido. Es importante destacar que esta métrica de evaluación es apropiada para comparar diferentes modelos de predicción para un conjunto de datos en particular y no entre conjunto de datos, ya que depende de la escala (Hyndman y Koehler, 2006).

Para Buenos Aires:

24/11/2019:

Función obtenida: $D(T) = 94,037 * T + 72,548$ (1)

RMSE = 415.228 +/- 151.145 (mikro: 441.752 +/- 0.000)

25/11/2019:

Función obtenida: $D(T) = 83,711 * T - 77,353$ (2)

RMSE = 313.850 +/- 104.819 (mikro: 332.447 +/- 0.000)

Para Jujuy:

24/11/2019:

Función obtenida: $D(T) = 17,402 * T - 224,517$ (3)

RMSE = 60.067 +/- 45.450 (mikro: 75.897 +/- 0.000)

25/11/2019:

Función obtenida: $D(T) = 17,275 * T - 272,107$ (4)

RMSE = 52.836 +/- 37.151 (mikro: 65.072 +/- 0.000)

Si bien, la raíz del error cuadrático medio no es del todo adecuada para comparar dos conjuntos de datos diferentes; al analizar la escala en la que fueron medidos, los valores de demanda de energía y los resultados del modelo de predicción; se concluye que la Provincia de Jujuy es la provincia más afectada por los abruptos cambios de demanda de potencia. Este fenómeno se debe a un mal aprovechamiento en el uso de las maquinarias o equipos de energía.

4. Conclusión

En Argentina hay una desigualdad regional en el aprovisionamiento de energía eléctrica. La demanda de electricidad en una determinada red varía, como se observó, a lo largo del día.

La variación en las curvas de demanda de potencia depende indefectiblemente de variables como la provincia que se analiza, la región a la que pertenece, estación del año, habitantes totales y cultura, siendo importante destacar la actividad económica que se desarrolla en cada sitio. Se pudo observar a través de un análisis exploratorio y confirmatorio de datos que estas variaciones de potencia son la principal causa de un mal aprovechamiento de las instalaciones.

Cabe destacar que, entre los dos sitios analizados, se puede identificar fácilmente que la Provincia de Jujuy es la que menor demanda de potencia posee, esto se debe, en general, a la baja cantidad de industrias que desarrollan su actividad en la provincia, la cantidad de población y la cultura de sus habitantes. Estos últimos se caracterizan por llevar a cabo sus tareas durante la mañana y la tarde, dejando un pequeño lapso del medio día para la siesta y durante la noche descansan, lo que concuerda con la curva de demanda de potencia de la Provincia de Jujuy. Mientras que, en el Gran Buenos Aires, constituido por grandes industrias, la cultura de sus habitantes es notoriamente distinta, caracterizándose por desarrollar sus tareas en la tarde y durante la noche no descansar.

Jujuy, a diferencia de GBA, es la provincia más afectada, hay variaciones considerables de potencia y, por lo tanto, mal uso de la capacidad instalada, por ejemplo, equipos o maquinarias. Inclusive existe una reducción de la vida útil de los mismos, lo cual se traduce económicamente en pérdidas de dinero.

Como trabajo futuro se propone investigar el fenómeno de la curva de demanda de potencia con conjuntos de datos más extensos aplicando modelos robustos de redes neuronales; y se estudiarán otras métricas de evaluación de los modelos de regresión; tales como el error absoluto promedio (MAE), error de sesgo medio (MBE).

5. Referencias

1. Ajila Chapín, A. A. (2012). "Los Sistemas de Información Geográfico como Herramienta para la Proyección de la Demanda de Energía Eléctrica en CNEL (Corporación Nacional de Electricidad) Regional El Oro" (Tesis de Maestría). Universidad San Francisco, Quito, Ecuador. Obtenido de: <http://repositorio.usfq.edu.ec/bitstream/23000/2581/1/104345.pdf>
2. Ghalekhondabi, I, A. Ehsan, G. Weckman y W. Young (2016). "An overview of energy demand forecasting methods published in 2005–2015", Working paper of Ohio University, DOI: 10.1007/s12667-016-0203-y
3. Grainger, S. (2002). "Análisis de sistemas de potencia". Ed. M. Graw Hill. New York. 2002. pp. 56-124.
4. Hernández Orallo, J., Ramírez Quintana, J. y Ferri Ramírez, C. (2004). "Introducción a la Minería de Datos". Madrid, España: Pearson Educación S.A. ISBN 84-205-4091-9.
5. Hyndman, R. J. y Koehler, A. B. (2006). "Another look at measures of forecast accuracy". *International Journal of Forecasting* **22** (4): 679-688. doi:10.1016/j.ijforecast.2006.03.001
6. Király, F. J., Von Büнау, P., Saputra-Muller, J., Blythe, D. A. J., Meinecke, F. C. y Müller, K. R. (2013). "Regression for sets of polynomial equations". AISTATS, España. Obtenido de: <https://arxiv.org/abs/1110.4531>
7. Lara Torralbo, J. A. (2016). "Business intelligence". Madrid, España. Centro de Estudios Financieros (CEF): 978-8-445-43202-0
8. Mastronardi, L. J., Sfeir, A. y Sánchez, S. (2016). "La temperatura y su influencia en la demanda de energía eléctrica: Un análisis regional para Argentina usando modelos econométricos". Obtenido de: <https://scripts.minem.gob.ar/octopus/archivos.php?file=7287>
9. Méndez Santos, P. A. (2013). "Gestión de la demanda de energía eléctrica en la Empresa Cartopel de Cuenca" (tesis de maestría). Obtenido de: <http://dspace.ucuenca.edu.ec/handle/123456789/453>
10. Mitra, S. y Acharya, T (2003). "Data mining: multimedia, soft computing and bioinformatics". John Wiley & Sons, 2003.
11. Pontius, R., Thontteh, O. y Chen, H. (2008). "Components of information for multiple resolution comparison between maps that share a real variable". *Environmental Ecological Statistics* **15** (2): 111-142. doi:10.1007/s10651-007-0043-y
12. Riquelme Santos, J. C., Ruiz, R. y Gilbert, K. (2006). Riquelme Santos, J.C., Ruíz, R. y Gilbert, K. (2006). "Minería de Datos: Conceptos y Tendencias". *Inteligencia Artificial: Revista*

Iberoamericana de Inteligencia Artificial, 10 (29), 11-18. Obtenido de:
<https://idus.us.es/xmlui/bitstream/handle/11441/43290/Miner%c3%ada%20de%20datos.pdf?sequence=1&isAllowed=y>

13. Tito, L. and Mullicundo, F. (2010). "Rapidminer: Tutorial on-line + Operadores". Obtenido de:
https://www.dataprix.com/files/RapidMiner_Tutorial_online_Operadores.pdf
14. Vargas, L. V., Farfán, J. H., Rodríguez, M. E., Aramayo, F., Flores, H. J., Bustamante, S. E. R. y López, V. (2016). "Comparación de las principales herramientas de Data Mining y Análisis de Sábanas Telefónicas". II JATIC, Mar del Plata, Argentina. Obtenido de:
<http://jatic2016.ucaecemdp.edu.ar/trabajos/WCII411RI-VargasJATIC2016.pdf>
15. Willmott, C. y Matsuura, K. (2006). "On the use of dimensioned measures of error to evaluate the performance of spatial interpolators". International Journal of Geographic Information Science 20: 89-102. doi:10.1080/13658810500286976.
16. Zotelo, C. (2016). "Modelo de predicción de consumo eléctrico en función de la temperatura para AMBA" (mimeo). Documento de trabajo de la Subsecretaría de Escenarios y Evaluación de proyectos del Ministerio de Energía y Minería de la Nación Argentina.