

Modelo de Estimación de Demanda Diaria de Distribuidoras de Gas Natural



www.enargas.gov.ar



© 2020 ENARGAS

Última revisión: septiembre 2020

Suipacha 636
C1008AAN, Buenos Aires
República Argentina
www.enargas.gob.ar

*Se autoriza la reproducción total o parcial del presente documento siempre que se cite la fuente.
En ese caso, solicitamos enviar una reproducción o enlace de la publicación a info@enargas.gob.ar*

AUTORIDADES ENARGAS

Intervención

Lic. Federico Bernal

Gerencia General

Dr. Osvaldo Felipe Pitrau

Gerencia Secretaría del Directorio

Dra. María Tereza Pittorino Díaz

Unidad de Auditoría Interna

Cra. Beatríz Suárez Trillo

Gerencia de Protección del Usuario

Dr. Héctor Domingo Maya

Gerencia de Innovación Tecnológica

Ing. Ana Carina Buccieri

Gerencia de Distribución

Ing. Jorge Osvaldo José Deferrari

Gerencia de Transmisión

Ing. Luis María Buisel

Gerencia de Despacho de Gas

Sr. Favio Marcelo Pezzullo

Gerencia de Gas Natural Vehicular

Lic. María Fernanda Martínez

Gerencia de Asuntos Legales

Dra. Carolina Inés Guerra Bianciotti

Gerencia de Desempeño y Economía

Lic. Néstor Daniel Touzet

Gerencia Regional Delegaciones

Dra. Natalia Judith Lascano

Gerencia de Recursos Humanos y Relaciones Institucionales

Sr. Eric Oscar Salomone Strunz

Dr. Claudio Oscar de la Fuente (adjunto)

Gerencia de Administración

Cr. Darío Javier Echazú

Gerencia de Tecnologías de la Información y Comunicación

Ing. Ramiro Pigliapoco

El presente documento fue desarrollado por el Lic. Sebastián Baigorria, de la Gerencia de Despacho de Gas, ENARGAS.

CONTENIDOS

Introducción	5
Objetivo y alcance	7
Metodología	9
Revisión bibliográfica sobre proyección de demanda de gas natural	10
Gas Demand Forecasting Methodology, de National Grid (NGrid), en el Reino Unido (2016)	10
Gas Demand Forecasting Methodology Information Paper, Australian Energy Market Operator (2019).....	11
Estimaciones de funciones de demanda de gas natural para consumo residencial en Argentina, Arrufat y Nader (Universidad Nacional de Córdoba, 1994)	12
Modelo generalizado de predicción de consumos de gas natural a mediano y corto plazo, S.Gil, J. Deferrari y L. Duperron (ENARGAS, 2002)	13
Conclusiones sobre la revisión bibliográfica	14
Modelos de Proyección de Demanda Diaria	15
Variables explicativas de la demanda	15
Tarifas	15
Sábados, domingos y feriados	16
Estimador mensual de actividad económica	16
Temperatura	17
Factor estacional.....	19
Modelo Econométrico	22
Modelo Machine Learning	24
Comparación entre los modelos Econométrico y Machine Learning	29
Aplicación del Modelo Machine Learning para Pronóstico de Demanda a Nivel País	31
Conclusiones.....	32

INTRODUCCIÓN

El gas natural constituye la principal fuente de energía en Argentina¹. Las centrales eléctricas, que constituyen el sector de mayor consumo de gas natural, generan a partir de este combustible un 55% de la energía eléctrica del país². La industria, el segundo sector consumidor de gas natural, obtiene del mismo un 56% de sus requerimientos energéticos³.

Los hogares, el tercer segmento consumidor de gas natural, obtienen más del 60% de sus requerimientos de energía a partir del mismo⁴.

Ilustración 1: Consumo de gas natural promedio mensual - 2018

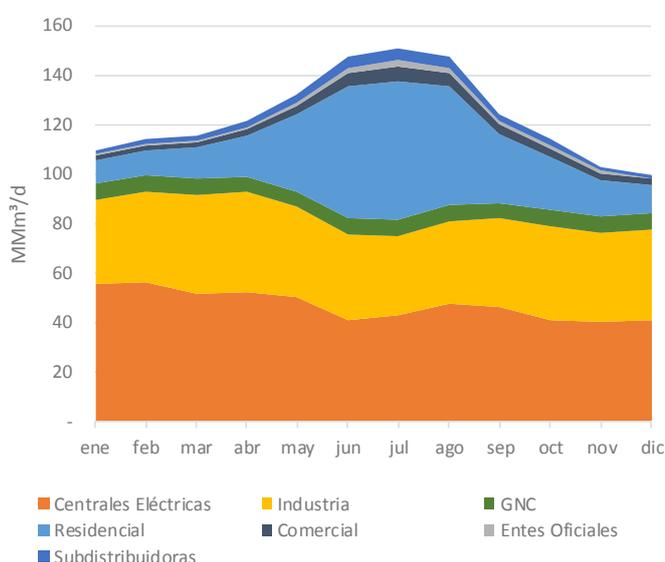
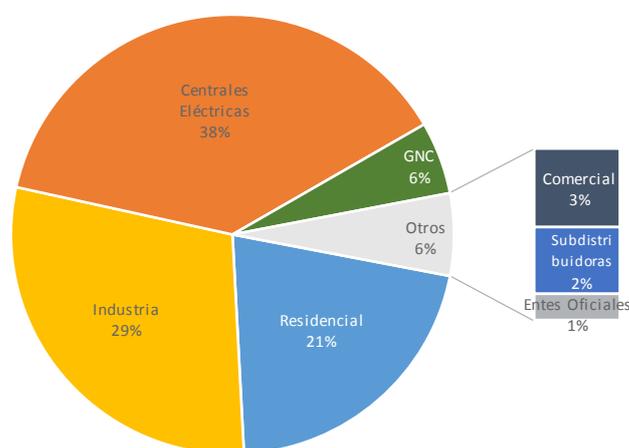


Ilustración 2: Consumo de gas por tipo de usuario - 2018



Fuente: Datos Operativos, ENARGAS

Una característica particular de los hogares, denominado sector residencial, es la marcada estacionalidad de su demanda. Durante los meses de verano el consumo de gas residencial es explicado por las necesidades de cocción de alimentos y de agua sanitaria. Mientras que en el período invernal el gas natural se requiere además para la calefacción de los hogares. Otros sectores que muestran un comportamiento estacional similar son los comercios y los entes oficiales (escuelas, hospitales, etc.). En los siguientes gráficos se muestra la evolución de los consumos mensuales de gas natural para estos segmentos de demanda a partir del año 2011.

1. Con una participación del 58% en la oferta interna total de energía de acuerdo a Escenarios Energéticos 2030 de la entonces Subsecretaría de Planeamiento Energético, ex Secretaría de Gobierno de Energía, 2019.
 2. Estimado para 2018 a partir de Datos Síntesis Mensual, CAMMESA.
 3. Estimado para 2018 a partir del Balance Energético Nacional, ex Secretaría de Gobierno de Energía.
 4. Estimado para 2018 a partir del Balance Energético Nacional, ex Secretaría de Gobierno de Energía.

Ilustración 3: Evolución de la demanda residencial

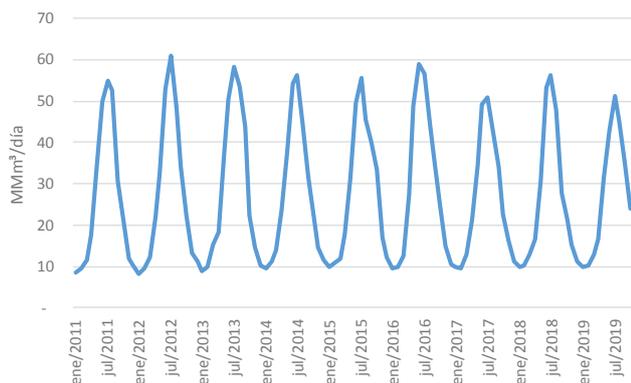
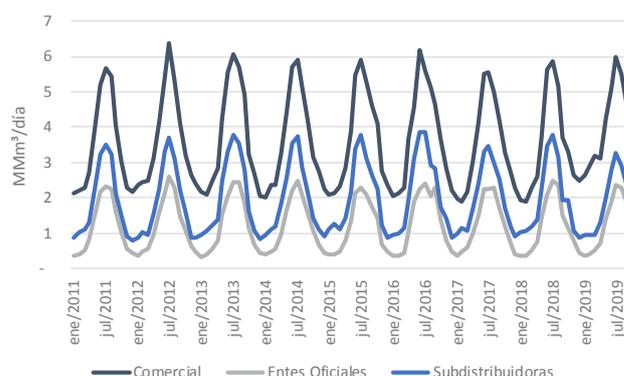


Ilustración 4: Evolución de demandas Comercial, Entes Oficiales y Subdistribuidoras⁵



Fuente: Datos Operativos, ENARGAS

Analizando las series de consumos mensuales, el mes invernal de mayor consumo representa aproximadamente cinco veces el consumo del mes más cálido. Cuando se realiza un análisis similar, pero con información diaria, un día de consumo pico en el invierno puede representar hasta seis veces el consumo de un día de verano. Así entonces, los segmentos residencial, comercial y entes oficiales llegan a representar más del 60% de la demanda total del país en un día frío de invierno.

Es por ello que se considera que la estimación de la demanda diaria de gas natural de estos segmentos es de mayor importancia para la operación eficiente y segura de los sistemas de transporte y distribución.

Por otra parte, su estimación es un requisito indispensable para el diseño de la cadena de suministro de gas natural: producción, importación, almacenamiento, transporte y distribución.

5. En general, las subdistribuidoras abastecen a usuarios residenciales y comerciales. En algunos casos abastecen a pequeñas industrias.

OBJETIVO Y ALCANCE

Considerando que el marco regulatorio de la industria del gas natural, Ley 24.076, tiene como objetivos, entre otros:

- Propender a una mejor operación y confiabilidad de los servicios e instalaciones de transporte y gas natural;
- Incentivar la eficiencia en el transporte, almacenamiento, distribución y uso del gas natural;
- Incentivar el uso racional del gas natural, velando por la adecuada protección del medio ambiente.

Por ello, el ENARGAS precisa contar con herramientas que permitan conocer los factores que determinan el comportamiento de la demanda de gas natural, con la finalidad de lograr los objetivos estipulados por la Ley.

El objetivo del presente documento es proponer modelos de estimación de demanda diaria de los usuarios de servicio completo, concepto que aglutina a los segmentos residencial, comercial y entes oficiales.

Un desafío al respecto, es la disponibilidad de información sobre mediciones diarias de consumo. Se puede dividir a los usuarios de gas natural en dos grandes grupos:

- Usuarios con medición diaria
- Usuarios con medición periódica

El grupo de los usuarios con medición diaria está compuesto por grandes industrias, centrales eléctricas, pequeñas industrias y estaciones de GNC. Su número es relativamente reducido, pero su consumo unitario es sustancialmente mayor al consumo de los usuarios con medición periódica. El consumo de los usuarios industriales y estaciones de GNC es relativamente estable a lo largo del año. La demanda de las usinas térmicas tiene cierta estacionalidad, con un pico en los meses de verano por la mayor demanda de energía eléctrica para refrigeración. En 2019, la cantidad de usuarios con medición diaria en Argentina era de aproximadamente 4.000. Estos usuarios, en general, deben comprarse el gas natural por su cuenta, salvo las estaciones de GNC, que pueden optar por ser abastecidas por las distribuidoras.

El grupo de los usuarios con medición periódica está compuesto principalmente por usuarios residenciales y comerciales. En general, la medición de consumo de los usuarios residenciales se realiza en forma bimestral y la medición de los usuarios comerciales se realiza en forma mensual. En 2019, se contabilizaban en Argentina aproximadamente 8,8 millones de usuarios con medición periódica. Las distribuidoras deben procurar el gas natural de estos segmentos de demanda, que la normativa denomina “demanda prioritaria”⁶⁻⁷.

Al no contar con mediciones diarias de sus usuarios de servicio completo⁸, las distribuidoras de gas natural realizan una estimación de su consumo aparente en forma agregada. Cada licenciataria de distribución cuenta con la medición del gas total entregado en su zona (mediciones de las transportistas de gas). A este consumo le detrae el consumo diario del resto de los segmentos, de los cuales posee mediciones de consumos diarios para obtener así, una estimación del consumo de ese día de la Demanda Prioritaria.

Demanda Prioritaria = Entregas totales en la zona - Consumos de usuarios con medición diaria

Cabe aclarar que el volumen obtenido también incluye el gas natural no contabilizado (GNNC), que representa conceptos tales como: fugas, robos, venteos y errores de medición. Sin embargo, es un volumen que también debe ser “abastecido” por las distribuidoras. La proporción del GNNC respecto a la demanda prioritaria es variable en cada zona de distribución, pero en promedio es del 7%⁹. Además, tiene un comportamiento similar a estas demandas. Por lo tanto, la demanda a ser explicada por los modelos propuestos en el presente documento será la demanda diaria bruta, que también incluye el GNNC.

Por otra parte, se encuentra el concepto de demanda unitaria, consumo específico o demanda por usuario, que se calcula dividiendo a la demanda bruta reportada diariamente por la cantidad de usuarios. Las Licenciatarias informan mensualmente la cantidad de usuarios de cada categoría. Entonces, podría calcularse la demanda unitaria de la siguiente manera:

$$Demanda\ Unitaria_{(día/mes/año)} = \frac{Demanda\ Bruta_{(día/mes/año)}}{Usuarios_{(mes/año)}}$$

En general, la incorporación de nuevos usuarios en un mes es ínfima o marginal respecto al stock, por lo tanto, se considera válida la simplificación anterior.

Finalmente, cuando en este documento se refiera a la variable demanda, debe entenderse que es la demanda bruta (incluye GNNC) unitaria.

6. Definida por la Resolución SE N° 599/2007.

7. A partir de la Resolución SGE N°175/2019, Resolución ENARGAS N° 750/2019 y Resolución ENARGAS N° 838/2019, el abastecimiento de los usuarios SGP3 por parte de las distribuidoras es optativo.

8. Excluyendo a las estaciones de GNC de servicio completo.

9. Corresponde a la relación verificada en el año 2018.

METODOLOGÍA

En primer término, se realizará una revisión bibliográfica de los modelos propuestos para realizar estimaciones de demanda sensible a la temperatura, tanto en la literatura internacional como local.

A partir de las prácticas identificadas en la revisión bibliográfica y de la base de datos disponible se propondrán modelos para estimar la demanda prioritaria diaria.

Para la estimación de los modelos se utilizarán las siguientes herramientas:

- Regresión con técnicas econométricas para series de tiempo;
- Estimación mediante el modelo Machine Learning (aprendizaje automático) denominado Boosting Machine.

El software estadístico a utilizar es R¹⁰. Para el análisis de series de tiempo se utilizará la librería “Forecast”¹¹ y para el modelo de Machine Learning se utilizará la librería “GBM”¹².

Para evaluar el rendimiento de una herramienta de aprendizaje estadístico sobre un conjunto de datos dado, se necesita alguna forma de medir qué tan bien sus predicciones realmente coinciden con los datos observados. Es decir, es necesario cuantificar en qué medida el valor pronosticado para una observación dada se aproxima al verdadero de esa observación. En el presente documento se utilizará la medida del Error Cuadrático Medio (ECM), que está dado por:

$$ECM = \frac{1}{n} \sum_{(i=1)}^n (y_i - \hat{y}_i)^2$$

Donde, n es la cantidad de observaciones, y_i es el valor real de la demanda e \hat{y}_i es el valor estimado de la demanda.

10. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

11. <https://CRAN.R-project.org/package=forecast>

12. <https://CRAN.R-project.org/package=gbm>

REVISIÓN BIBLIOGRÁFICA SOBRE PROYECCIÓN DE DEMANDA DE GAS NATURAL

A continuación, se presenta una síntesis de la revisión bibliográfica realizada sobre el tema.

Gas Demand Forecasting Methodology, de National Grid (NGrid), en el Reino Unido (2016)

NGrid¹³ es la empresa operadora de los sistemas de transmisión eléctrica y de gas en el Reino Unido (Inglaterra y Gales).

El modelo de NGrid para proyectar la demanda no medida en forma diaria utiliza a la temperatura como principal variable explicativa, pero en conjunto con otras variables climatológicas, que denomina variable meteorológica compuesta (Composite Weather Variable - CWV).

La CWV toma en consideración no solo la temperatura, sino también la velocidad del viento, la temperatura efectiva y la temperatura efectiva estacional normal.

La temperatura efectiva es definida como la media móvil entre la temperatura del día y la temperatura efectiva del día anterior.

La temperatura efectiva estacional es representativa de los valores históricos esperados de temperatura para cada época del año (modelada con funciones trigonométricas).

La relación entre el consumo de gas y la CWV resulta lineal. Para lograr esta relación lineal con el consumo de gas se consideran tres umbrales de transición para la variable CWV.

13. <https://www.nationalgrid.com/>

Ilustración 5: Relación lineal demanda - CWV

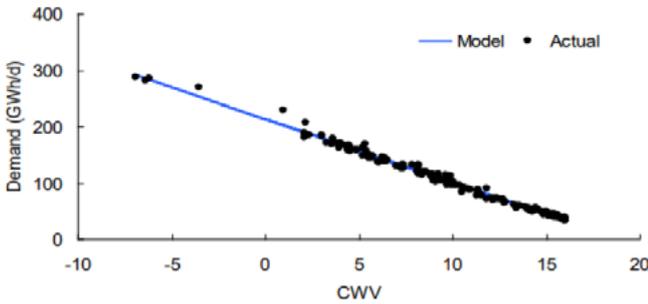
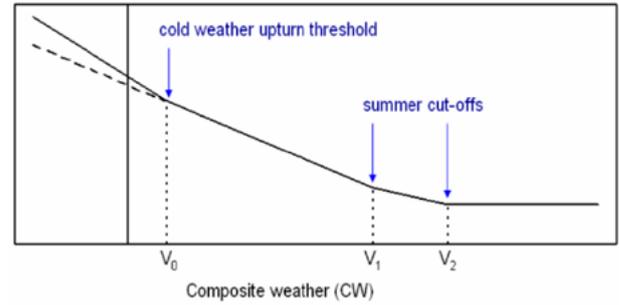


Ilustración 6: Umbrales de transición para CWV¹⁴



Fuente: National Grid

También son tomados en consideración los consumos durante fines de semanas y vacaciones. Luego, el modelo de regresión para la demanda es el siguiente:

$$Demanda_t = a + b * CWV_t + u_t$$

Donde $Demanda_t$ es la demanda pronosticada, a es la ordenada de origen (el valor de la demanda cuando el resto de los parámetros son igual a cero), b es el coeficiente de regresión de la variable CWV (cuánto se incrementa la demanda cuando se incrementa una unidad la variable CWV).

El error tiene la siguiente estructura: $u_t = \rho u_{t-1} + e_t$. Donde e_t es un término de error aleatorio con distribución normal, de media cero y varianza σ_e^2 .

Gas Demand Forecasting Methodology Information Paper, Australian Energy Market Operator (2019)

Australian Energy Market Operator (AEMO) realiza una serie de funciones relacionadas a la operación, desarrollo y planificación en los mercados del gas natural y la electricidad.

Para realizar sus estimaciones de demanda AEMO también utiliza variables climáticas: Heating Degree Days (HDD) y Effective Degree Days (EDD).

El parámetro HDD se usa como un indicador de los niveles de temperatura exterior por debajo de lo que se considera una temperatura confortable. Si la temperatura diaria promedio cae por debajo de los niveles de confort (temperatura base), se requiere calefacción.

14. LDZ Aggregate NDM Demand: Demanda agregada no medida en forma diaria (NDM) en la zona de distribución local (LDZ).

Por otra parte, el parámetro EDD es utilizado para cuantificar el impacto de otras variables meteorológicas, como velocidad del viento, radiación solar y estacionalidad.

El factor estacionalidad representa la respuesta del consumidor a diferentes condiciones climáticas. Los consumidores tendrían diferentes hábitos energéticos en invierno que fuera del invierno a pesar de que los días fueran de la misma temperatura (modelada con funciones trigonométricas).

Antes de realizar las estimaciones de demanda diaria, se deducen elementos de tendencia y efectos precio de la serie original de demanda.

$$\text{Demanda} = f(\text{HDD ó EDD}, \text{día de la semana}, \text{Vacaciones}, \text{Tendencia})$$

Donde *Demanda* representa a la demanda diaria y *f* hace referencia a un modelo o función de el déficit grado día (*HDD*), otras variables metodológicas (*EDD*), el día de la semana, si es un período vacacional y un componente tendencial.

Estimaciones de funciones de demanda de gas natural para consumo residencial en Argentina, Arrufat y Neder (Universidad Nacional de Córdoba, 1994)

Los autores proponen un modelo lineal para la estimación de la demanda mensual de cada distribuidora del país:

$$\text{Demanda} = \alpha + \beta * \text{HHD}_{\text{ponderado}}$$

$$\text{HDD} = \max(18,33 - \text{Temperatura media}, 0)$$

Donde *HHD*_{ponderado} resulta de calcular los déficits grado día (respecto a una temperatura base de 18,33°C) de las localidades más importantes de una zona de distribución, y ponderarlos por su importancia relativa.

Los autores también proponen un modelo lineal a trozos, considerando un segundo umbral, donde la pendiente de la curva de demanda es más suavizada porque los hogares se encuentran cerca de saturar la capacidad instalada de calefacción.

$$HDD2 = \max(\text{segundo umbral} - \text{Temperatura media}, 0)$$

$$HDD1 = HDD - HDD2$$

$$\text{Demanda} = \alpha + \beta_1 * HDD1 + \beta_2 * HDD2$$

Donde α es la ordenada ar origen, β_1 es el coeficiente para temperaturas entre los umbrales 1 y 2, y β_2 es el coeficiente para temperaturas menores al umbral 2.

Modelo generalizado de predicción de consumos de gas natural a mediano y corto plazo, S.Gil, J. Deferrari y L. Duperron (ENARGAS, 2002)

Los autores utilizan el concepto de temperatura efectiva, entendido como la media móvil entre la temperatura promedio del día y el promedio de las temperaturas medias de los n días anteriores (usualmente, n=4 o n=5)

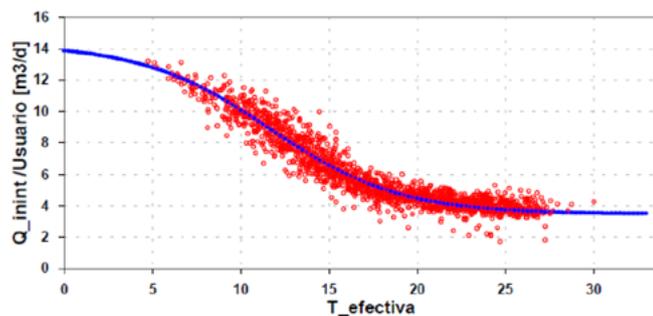
$$Tef_t = w * T_t + (1-w) * T_n$$

Donde Tef_t es la temperatura media efectiva en el momento t , w es un ponderador, T_t es la temperatura media del día y T_n es la temperatura promedio de los n días anteriores.

La temperatura efectiva automáticamente incorpora el efecto de retardo que se manifiesta en el comportamiento típico de los usuarios de gas natural. Más específicamente, este parámetro permite incluir en los modelos de consumo el hecho de que los usuarios de gas no encienden la calefacción inmediatamente después que la temperatura desciende, sino una vez que la misma se mantuvo baja por algunos días. Similarmente los calefactores no se apagan inmediatamente apenas sube la temperatura sino una vez que dicho aumento persistió por unos días.

El modelo para predecir la demanda es no lineal (tangente hiperbólica). A continuación, se muestra un ejemplo de la curva para la zona del Gran Buenos Aires.

Ilustración 7: Curva tangente hiperbólica para la zona del Gran Buenos Aires



Fuente: ENARGAS, 2002

Conclusiones sobre la revisión bibliográfica

Todos los documentos consultados consideran a la variable temperatura como predictor del consumo de gas natural.

Tanto NGrid como el modelo de ENARGAS de 2002 introducen el concepto de “temperatura efectiva”, calculada como la media móvil de las temperaturas medias de los días precedentes.

Todos los documentos introducen el concepto de Heating Degree Days (HDD), umbrales con cambio de pendiente, o formas funcionales no lineales para modelizar la relación entre temperatura y consumo. Considerar la temperatura a través de HDD es equivalente a modelizar la relación a través de umbrales, donde el nivel del umbral está definido por la temperatura base.

Por otra parte, NGrid y AEMO consideran en sus estimaciones los efectos de los días sábados, domingos y feriados.

Finalmente, AEMO tiene en cuenta el impacto del precio del gas antes de realizar sus predicciones.

MODELOS DE PROYECCIÓN DE DEMANDA DIARIA

Variables explicativas de la demanda

En función a la bibliografía revisada e información disponible para el estudio se ensayarán modelos con las siguientes variables explicativas:

- Tarifas pagadas por los usuarios;
- Días sábado, domingo y feriados;
- Estimador mensual de actividad económica;
- Temperatura media del día, como valores rezagados de temperatura. También se consideran los valores en nivel como en diferencias respecto a una temperatura base (HDD);
- Demanda estacional normal para el día;

La base de datos disponible para el estudio corresponde al período comprendido entre enero 2011 y julio 2019 (3.118 registros) para la distribuidora Naturgy BAN, la cual es una de las mayores distribuidoras de la zona metropolitana del país. Más adelante se presentará una aplicación de modelos de proyección para todas las distribuidoras del país.

Tarifas

En la actualidad existen ocho tipos de tarifas para usuarios residenciales y otros seis tipos de tarifas para usuarios comerciales.

Si bien cada categoría tarifaria posee una diferente composición de cargos fijos y variables, y pesos relativos distintos de las componentes gas, transporte y distribución, en general, las variaciones han sido similares. Por ello, se tomará como referencia la tarifa de una sola categoría, Residencial R3 2°, como valor de la mediana del universo de usuarios residenciales. Sólo se pretende tener un índice representativo del movimiento real de las tarifas, que luego pueda ser utilizado en forma simplificada para realizar proyecciones.

Las tarifas serán ingresadas a los modelos en valores reales, deflactados por inflación.

Ilustración 8: Tarifa real para usuarios R3 2 de Naturgy BAN



Fuente: Elaboración propia

Sábados, domingos y feriados

Se consideran variables dummy¹⁵ para indicar:

- Días sábado: algunas industrias y comercios funcionan en horarios limitados durante los días sábados y el consumo hogareño suele ser diferente.
- Días domingo y feriados: algunas industrias y comercios no trabajan durante los días domingos o feriados, y el consumo hogareño suele ser diferente.

Si un día sábado es feriado, la variable dummy indicadora de los sábados será igual a cero.

Estimador mensual de actividad económica

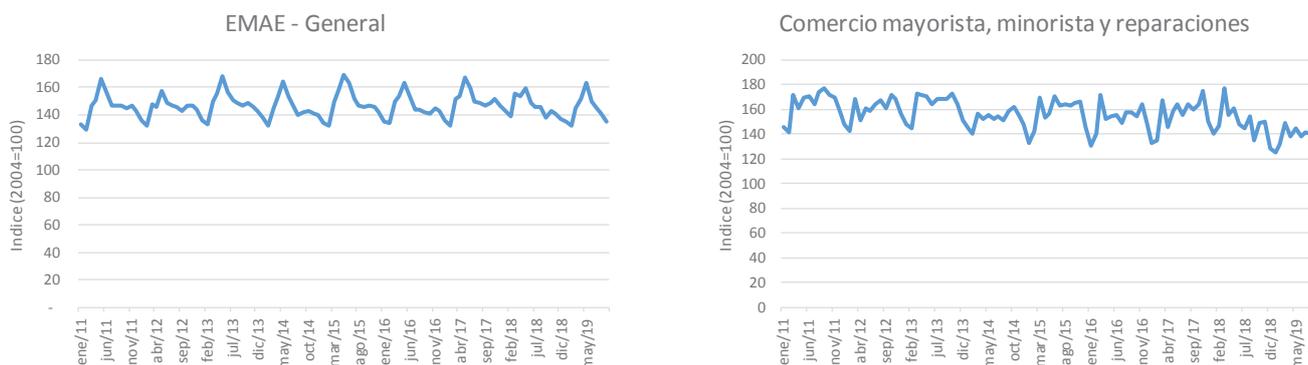
El estimador mensual de actividad económica (EMAE) elaborado por el Instituto de Estadísticas y Censo (INDEC)¹⁶ refleja la evolución mensual de la actividad económica del conjunto de los sectores productivos a nivel nacional.

Dado que la demanda prioritaria incluye comercios y pequeñas industrias, se incluirán indicadores del nivel de actividad de dichos sectores.

15. Las variables dummy son variables binarias que toman valor 0,1.

16. <https://www.indec.gob.ar/indec/web/Nivel4-Tema-3-9-48>

Ilustración 9: Evolución del EMAE



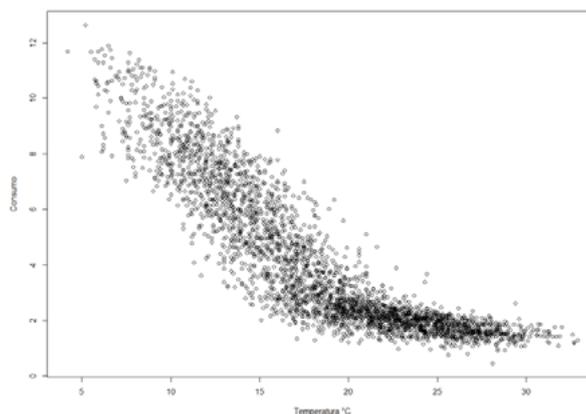
Fuente: INDEC

Temperatura

Las temperaturas utilizadas corresponden al Servicio Meteorológico Nacional (SMN), para el Observatorio de Central de Buenos Aires.

A continuación, se presenta gráficamente la relación entre el consumo diario registrado para la distribuidora de gas y la temperatura media ente enero 2011 y julio 2019.

Ilustración 10: Relación consumo y temperatura



Fuente: Elaboración propia a partir de temperaturas del SMN y consumos de Naturgy BAN

Cabe aclarar que la temperatura media es calculada como el promedio simple entre las temperaturas mínima y máxima.

En el gráfico anterior se observan claramente dos secciones:

- Para temperaturas medias mayores a $\sim 20^{\circ}\text{C}$, el consumo de gas es poco sensible a la temperatura, asociado a consumos de cocción y agua sanitaria.
- Para temperaturas medias menores a $\sim 20^{\circ}\text{C}$, el consumo de gas es muy sensible a la temperatura, asociado a calefacción de hogares y locales comerciales.

Mediante una regresión logística se estimó el umbral a partir del cual existe un comportamiento diferente del consumo de gas respecto a la temperatura media, es decir los valores aproximados de temperatura media que limitan esas zonas de distinto comportamiento del consumo. Dicho punto se encontraría, para la distribuidora en cuestión, en el nivel de $19,4^{\circ}\text{C}$. Esta temperatura se encuentra en el rango de temperaturas base utilizadas en Australia por el AEMO.

Cuando se consideran temperaturas menores a $19,4^{\circ}\text{C}$, existiría un segundo umbral de consumo en un nivel de $11,5^{\circ}\text{C}$.

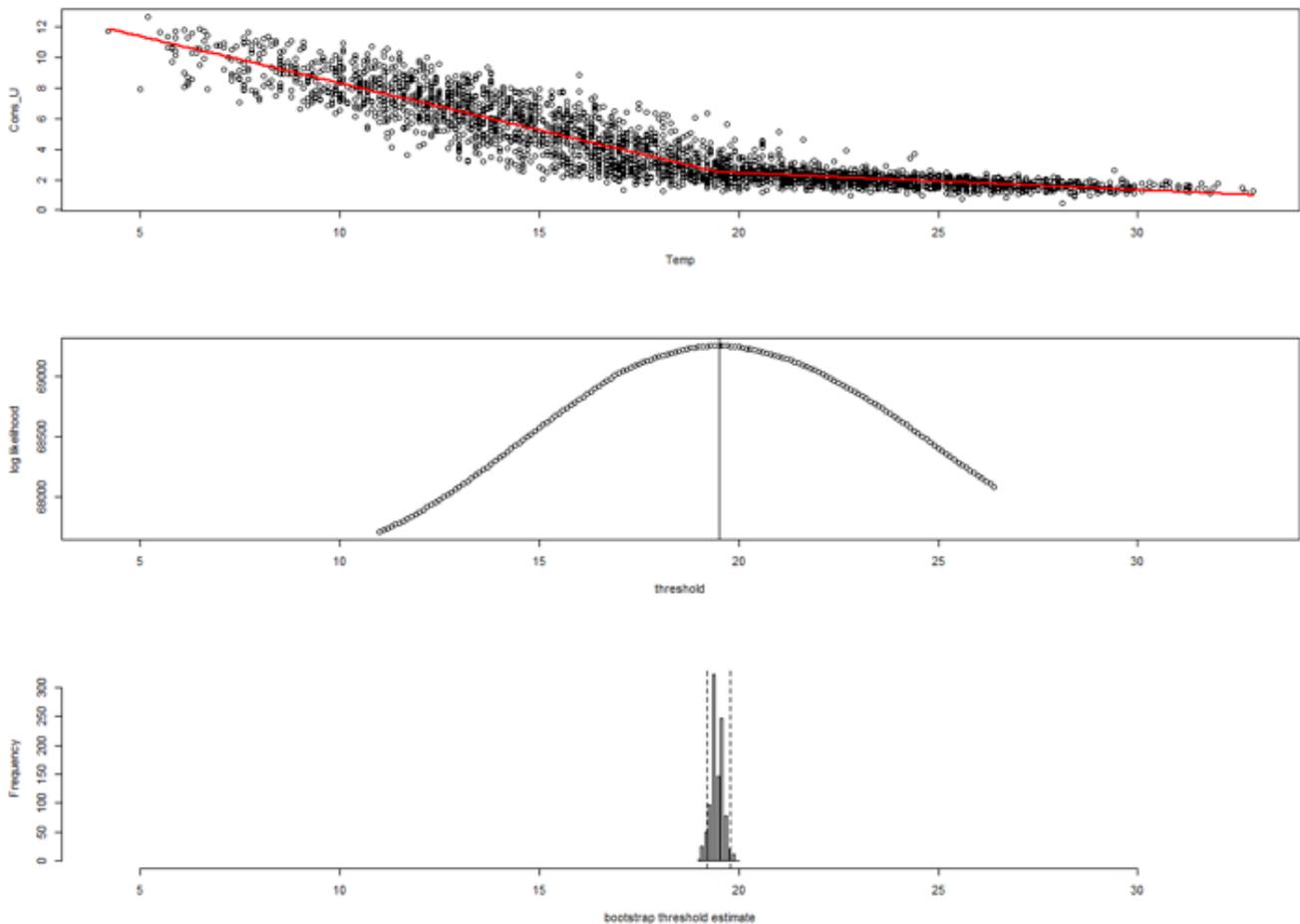
Así entonces, la variable temperatura será ingresada a los modelos en diferentes configuraciones:

- Temperatura media: T_{media_t}
- Déficit grado día: HDD_t

$$HDD_t = \max(19,4 - T_{media_t}; 0)$$

- Curva segmentada: $T_{media_t} + HDD_t$
- Déficit grado día para temperaturas frías (umbral 2)
- $HDD_t = \max(11,5 - T_{media_t}; 0)$
- Temperatura media y déficit grado día al cuadrado: TSQ_{media} y $HDDSQ$
- Temperatura media y déficit grado día rezagados: $T_{media_{t-r}}$ y HDD_{t-r}

Ilustración 11: Estimación del umbral de temperaturas



Fuente: Salida del paquete "chngptm"¹⁷

Factor estacional

La variable factor estacional representaría el comportamiento de los usuarios de encender comúnmente la calefacción en el invierno y de no hacerlo fuera de temporada, aunque la temperatura descienda más allá de la temperatura base (19,4°C en este caso).

En primer término, se modeló la temperatura normal para cada día del año a partir del siguiente modelo:

$$T_{media,t} = A + B * t + C * \text{sen} \left(\frac{2\pi}{365} t \right) + D * \text{cos} \left(\frac{2\pi}{365} t \right) + \varepsilon_t$$

17. *Estimation and Hypothesis Testing for Threshold Regression*, Youyi Fong, 2019.

Donde t es el número de día de la serie o la tendencia, A es la ordenada al origen, B es el coeficiente de la tendencia, C y D son coeficientes de las funciones seno y coseno con las cuales se representa la estacionalidad anual. Dado que existe correlación serial en la serie de temperaturas se aplicó un modelo autorregresivo integrado de media móvil (ARIMA), ya que la temperatura respondería a un modelo dinámico (la temperatura de hoy depende de la temperatura de ayer). Si se obviara la correlación serial de la serie de tiempo podrían obtenerse valores sesgados de los parámetros del modelo.

El modelo que mejor se ajustó a los datos fue el ARIMA (1,0,1), siendo los coeficientes:

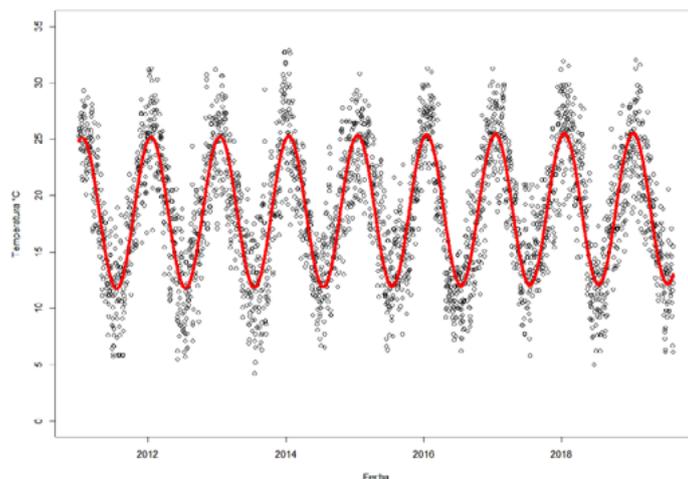
COEFICIENTE	A	B	C	D
VALOR	18,42623*	0,00013895**	2.00819*	6.39627*

* Significativo al 1%.

** Significativo al 5%.

Según este modelo, la temperatura media en la Ciudad de Buenos Aires aumentaría en 1,57°C hacia 2050. A las predicciones de temperatura de este modelo se denominará “temperatura normal”.

Ilustración 12: Temperatura normal



Fuente: Elaboración propia a partir de temperaturas del SMN

En una segunda etapa, se construyó un modelo de regresión cuya variable dependiente es la demanda diaria por usuario. Los regresores son la temperatura normal, sábados, domingos y feriados. El modelo que mejor se ajustó a los datos fue un ARIMA (3,1,1) con la siguiente estructura:

$$Demanda_t = T_{normal_t} + HDD_{normal_t} + HDD_{normal_t}^2 + Sab_t + D y F_t$$

Donde:

T_{normal_t} : temperatura estacional normal del día t

HDD_{normal_t} : HDD estacional normal del día t

$HDD_{normal_t}^2$: HDD estacional normal del día t elevado al cuadrado

Sab_t : si es día sábado

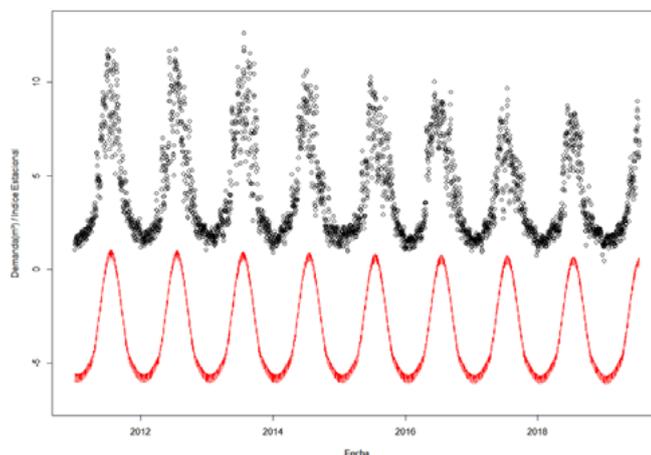
$D y F_t$: si es día domingo o feriado

COEFICIENTE	T_{normal_t}	HDD_{normal_t}	$HDD_{normal_t}^2$	Sab_t	$D y F_t$
VALOR	-0.1848*	0.2148**	0.0433*	-0.2409*	-0.4361*

* Significativo al 1%. ** Significativo al 5%.

En el siguiente gráfico se muestran las observaciones de demanda y las proyecciones obtenidas para la variable de factor estacional, que es la proyección del modelo anterior (línea roja).

Ilustración 13: Factor Estacional



Fuente: Elaboración propia

Modelo Econométrico

Tomando en consideración las variables explicativas abordadas con anterioridad, el criterio de información de Akaike y el comportamiento de los residuos, el modelo con mejor desempeño sería un ARIMA (2,0,1) con la siguiente estructura:

$$Demanda_t = \alpha + F_{estacional_t} + dTEF_t + dTEF_t^2 + HDD_t^{TEF} + dHDD_t + HDD_t^{frío} + dT_{media_t} + LTarifa_{t-30}^{HDD} + LTarifa_{t-30}^{no\ HDD} + EMAE_{Comer_t} + EMAE_{Comer_t}^2$$

Donde:

α : ordenada al origen

$F_{estacional_t}$: factor estacional del día t

$dTEF_t$: diferencia entre la temperatura efectiva registrada en el día t y su temperatura normal

$dTEF_t^2$: diferencia entre la temperatura efectiva registrada en el día t y su temperatura normal, elevada al cuadrado

HDD_t^{TEF} : diferencia entre el déficit grado día calculado respecto a la temperatura efectiva y su HDD normal

$dHDD_t$: diferencia entre el déficit grado día calculado respecto a la temperatura del día y su HDD normal

$HDD_t^{frío}$: déficit grado día calculado respecto a la temperatura base de 11,5 °C

dT_{media_t} : diferencia entre la temperatura media registrada en el día t y su temperatura normal

$LTarifa_{t-30}^{HDD}$: logaritmo de la tarifa vigente en el mes anterior al día t, si HDD es mayor que cero

$LTarifa_{t-30}^{no\ HDD}$: logaritmo de la tarifa vigente en el mes anterior al día t, si HDD es igual a cero

$EMAE_{Comer_t}$: índice EMAE para la actividad de Comercio mayorista, minorista y reparaciones

$EMAE_{Comer_t}^2$: índice EMAE para la actividad de Comercio mayorista, minorista y reparaciones, elevado al cuadrado

Los valores de los coeficientes del modelo se muestran en la siguiente tabla.

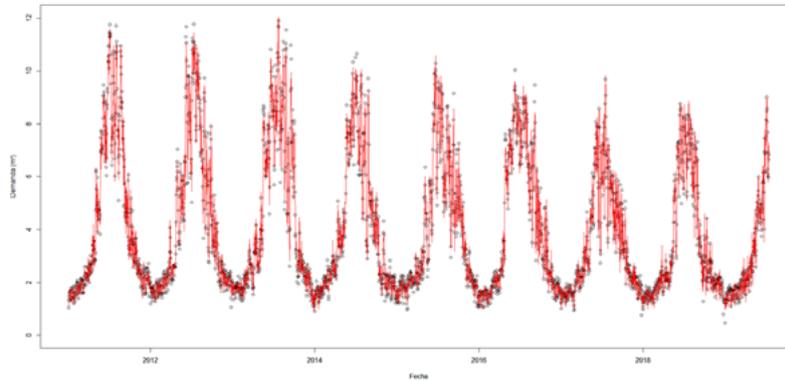
COEFICIENTE	VALOR
α	12.4905*
$F_{estacional,t}$	0.9848*
$dTEF_t$	-0.0216*
$dTEF_t^2$	-0.0095*
HDD_t^{TEF}	0.2473*
$dHDD_t$	0.1000*
$HDD_t^{frío}$	0.0630**
$dT_{media,t}$	-0.0171**
$LTarifa_{t-30}^{HDD}$	-0.5498*
$LTarifa_{t-30}^{no HDD}$	-0.1031*
$EMAE_{Comer,t}$	0.0002***
$EMAE_{Comer,t}^2$	-0.0629***

* Significativo al 1%.
 ** Significativo al 5%.
 *** Significativo al 10%.

El error cuadrático medio del modelo es de 0.1099.

En el siguiente gráfico se muestran las observaciones de demanda (puntos) y las predicciones del modelo (línea roja).

Ilustración 14: Predicciones de demanda del modelo econométrico



Fuente: Elaboración propia

Modelo Machine Learning

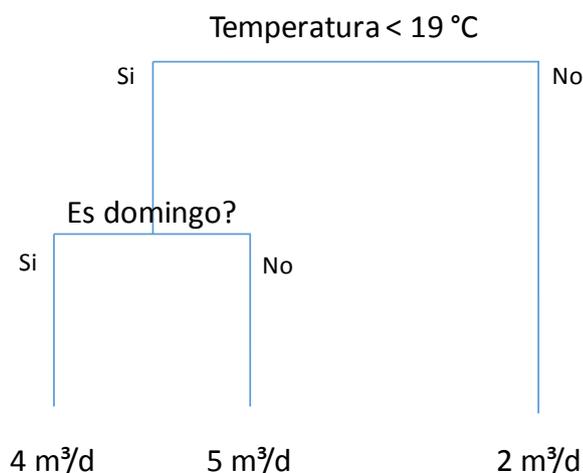
El Machine Learning es un método de análisis de datos que automatiza la construcción de modelos analíticos. Es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de datos, identificar patrones y tomar decisiones con mínima intervención humana.

Dentro del Machine Learning, se propone utilizar el modelo Boosting Machine (Boosting). Antes de explicar el funcionamiento de Boosting, es conveniente comenzar con los árboles de decisión.

A continuación, se muestra un ejemplo de árbol de regresión, que consiste en una serie de reglas de decisión. Inicialmente, en la cima del árbol, se realiza una separación de las observaciones, asignando al brazo derecho si la variable temperatura es mayor o igual a 19° C. El consumo predicho para dichas observaciones está dado por la media de las observaciones con temperaturas mayores o iguales a 19 °C, que es de 2 m³/d.

Observaciones con temperaturas inferiores a 19° C son asignadas sobre la rama izquierda y luego, son nuevamente separadas en dos grupos, dependiendo si es día domingo. Se dice entonces que esta rama tiene dos nodos terminales. Si es día domingo, el consumo predicho es de 4 m³/d, y si no es domingo el consumo predicho es de 5 m³/d. Estos valores pronosticados se calculan como el promedio o la moda sobre cada subconjunto de observaciones.

Ilustración 15: Ejemplo de árbol de decisión



Fuente: Elaboración propia

Boosting Machine es una técnica para mejorar el poder predictivo de los árboles de decisión. Consiste en crear numerosos árboles en forma secuencial, donde cada árbol toma información de los árboles anteriores. Se arman pequeños árboles de decisión, cuyos parámetros se van ajustando en función de los residuos resultantes de los modelos anteriores.

A diferencia del modelo econométrico, con Boosting Machine no se obtienen los coeficientes β , que representan la respuesta de la variable dependiente (demanda de gas) ante cambios en las variables explicativas, si no que directamente se obtiene la predicción de la variable dependiente.

En el caso del modelo Boosting Machine, se incluyeron las siguientes variables explicativas:

- $Seno = \sin \left(\frac{2\pi}{365} t \right)$

- $Coseno = \cos \left(\frac{2\pi}{365} t \right)$

- $Temperatura\ efectiva = 0.5 * Temperatura_{media} + 0.5 * Temperatura_{efectiva,t-1}$

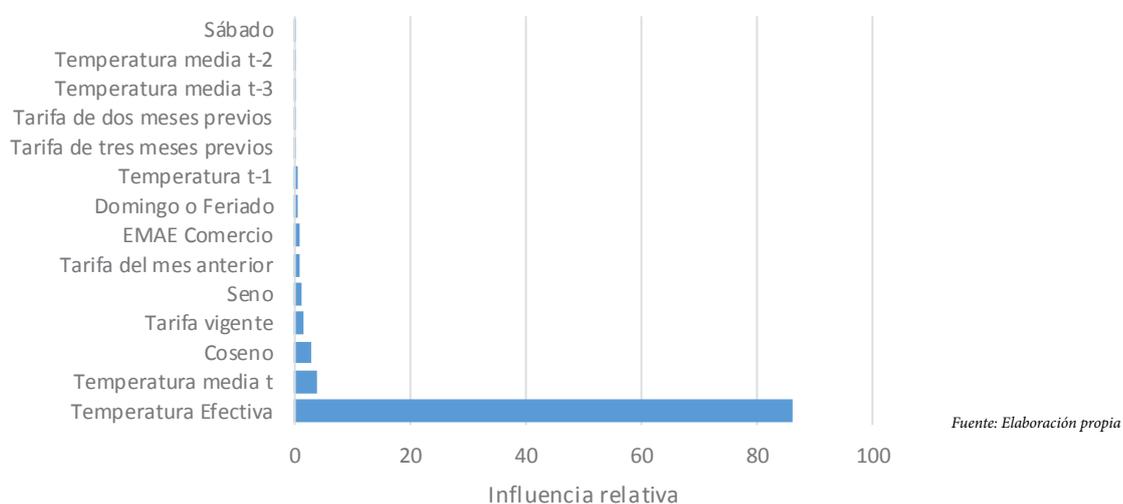
- Temperatura media del día
- Temperatura media de uno, dos y tres días previos
- Tarifa vigente en términos reales
- Tarifa vigente en términos reales de uno, dos y tres meses previos
- Si es día sábado
- Si es día domingo o feriado
- Incide EMAE para la actividad de Comercio mayorista, minorista y reparaciones

De las 3.118 observaciones de la muestra, se seleccionó aleatoriamente un 75% de las mismas para el conjunto de entrenamiento, a partir del cual se parametriza el modelo. El 25% restante, se utiliza para testear las predicciones realizadas por el modelo.

El mejor modelo que se ajustó a los datos, realizando validación cruzada de sus parámetros, está compuesto por 1.962 árboles, de 20 niveles, y con un mínimo de 10 observaciones en los nodos terminales.

En el siguiente gráfico se muestra la importancia relativa de cada variable a la hora de predecir la demanda de gas.

Ilustración 16: Influencia relativa de las variables en el modelo Boosting Machine



Tal como sucede en los modelos de NGrid y ENARGAS (2002), la variable Temperatura Efectiva tiene una importancia preponderante para predecir la demanda de gas (86% de influencia relativa). En segundo lugar, se encuentra la temperatura del día, 4% de influencia relativa.

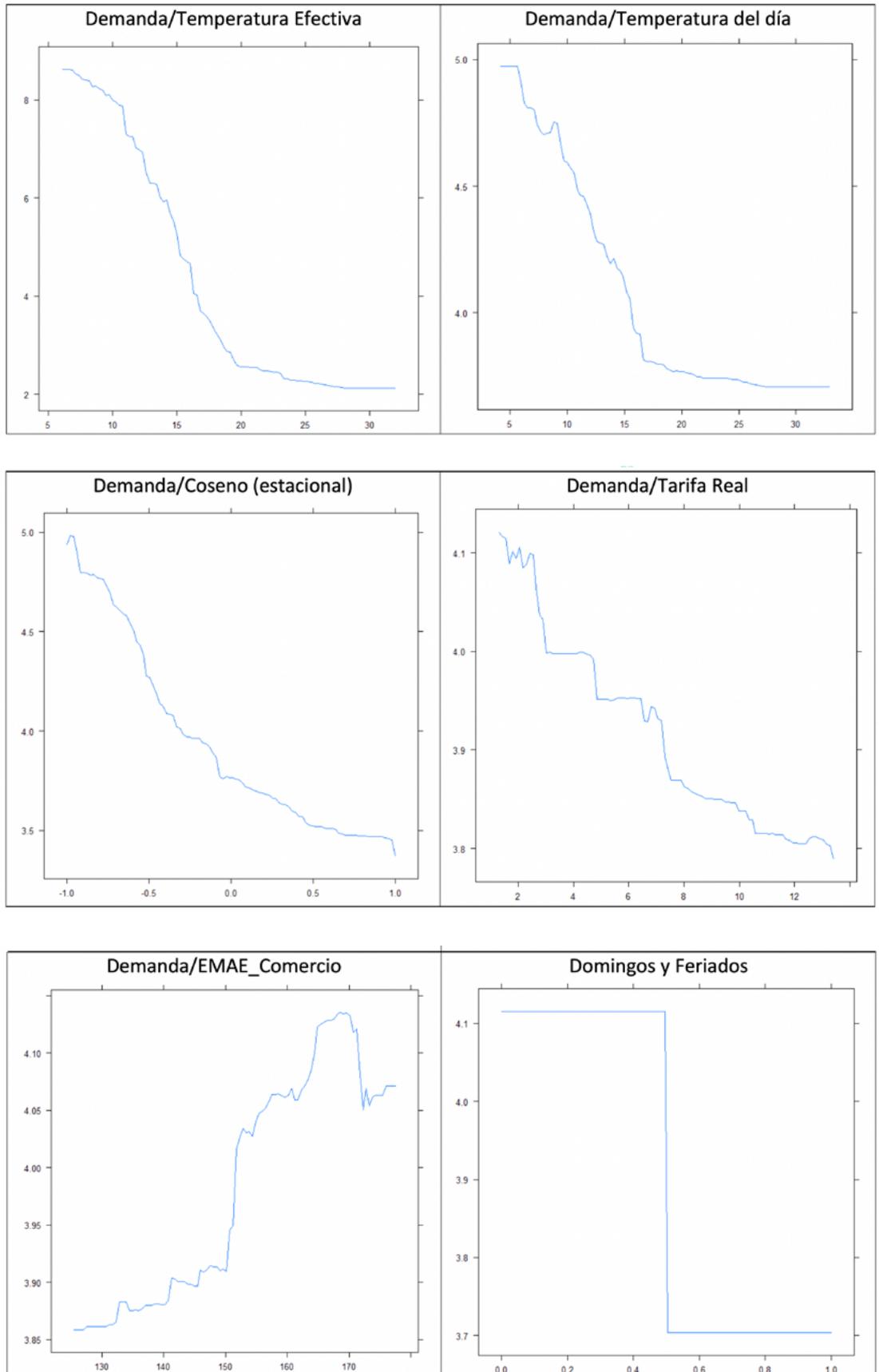
La variable Coseno, que representa la componente estacional (junto con Seno), se encuentra en el tercer lugar de importancia (2% de influencia relativa).

El cuarto lugar de importancia es la tarifa vigente en términos reales (1,4% de influencia relativa).

Luego, todas las otras variables tienen valor predictivo, pero con participaciones sustancialmente menores.

Si bien el modelo Boosting Machine no arroja coeficientes β , se pueden obtener gráficos de cómo se comporta la demanda para diferentes valores de las variables explicativas.

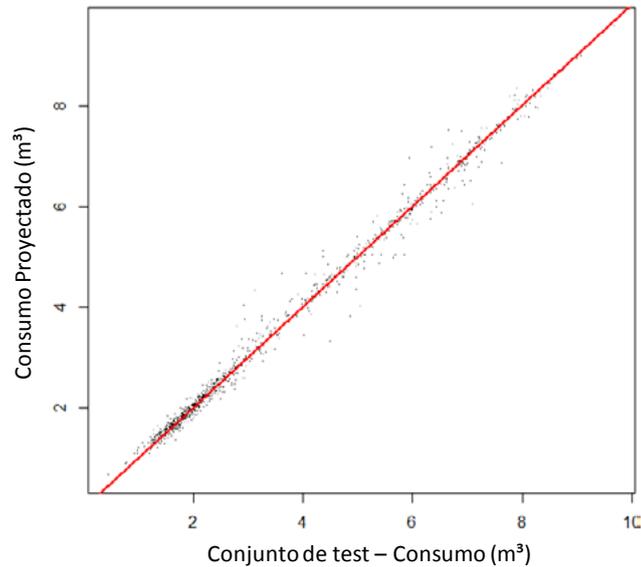
Ilustración 17: Dependencia parcial entre demanda y variables explicativas en el modelo Boosting Machine



Fuente: Elaboración propia

El error cuadrático medio (ECM) para el conjunto de entrenamiento es de 0,0359 y para el conjunto de test es de 0,1309. Si bien el comportamiento del modelo sobre el conjunto de test no es tan bueno como en el conjunto de entrenamiento, sus predicciones son satisfactorias.

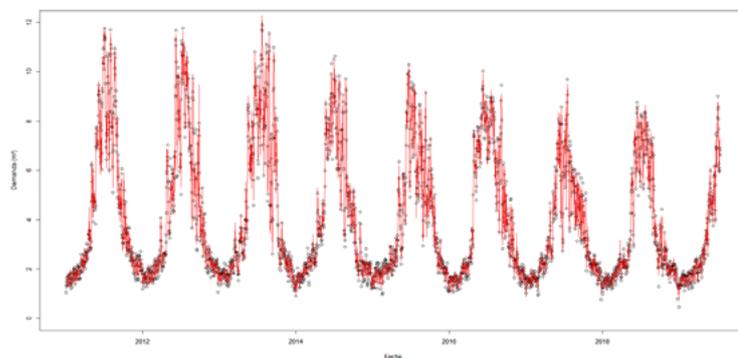
Ilustración 18: Consumo real y proyectado en la muestra de test



Fuente: Elaboración propia

Por último, se corrió el modelo sobre la misma muestra utilizada para el modelo econométrico (muestras de entrenamiento y test) y el ECM obtenido fue de 0,059 (es menor que en el caso del modelo econométrico). En el siguiente gráfico se muestran las observaciones de demanda (puntos) y las predicciones del modelo Boosting Machine (línea roja).

Ilustración 19: Predicciones de demanda del modelo Boosting Machine

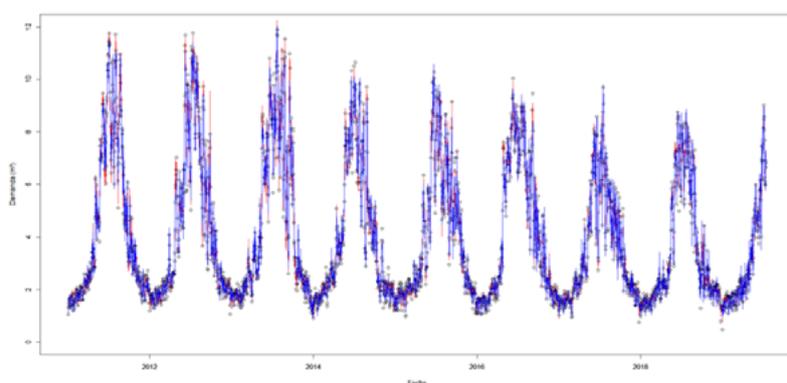


Fuente: Elaboración propia

Comparación entre los modelos Econométrico y Machine Learning

En el siguiente gráfico se muestran las observaciones de demanda (puntos) y las predicciones del modelo econométrico (azul) y las predicciones del modelo Boosting Machine (rojo) correspondiente a las muestras de entrenamiento de los modelos.

Ilustración 20: Comparación de predicciones entre modelos



Fuente: Elaboración propia

Si bien la mayor parte de las predicciones coinciden (línea azul superpuesta a la roja), el modelo Boosting Machine se encuentra más cerca de las observaciones reales en los días de alta demanda (línea roja supera a la azul en los días de demanda alta).

Como se mencionó anteriormente, la medida del ECM da la ventaja al modelo Boosting Machine.

MEDIDA/MODELO	ECONOMÉTRICO	BOOSTING MACHINE
ECM	<i>0.110</i>	<i>0.059</i>
RECM	<i>0.331</i>	<i>0.150</i>

La raíz cuadrada del error cuadrático medio (RECM) puede ser interpretada como una medida del error expresada en la misma unidad que la variable pronosticada (demanda medida en m³/d).

En la siguiente tabla se muestra el cociente entre la RECM y la demanda promedio para cada modelo, a nivel global, durante los meses de invierno (mayo-septiembre) y durante los meses de verano (octubre-abril).

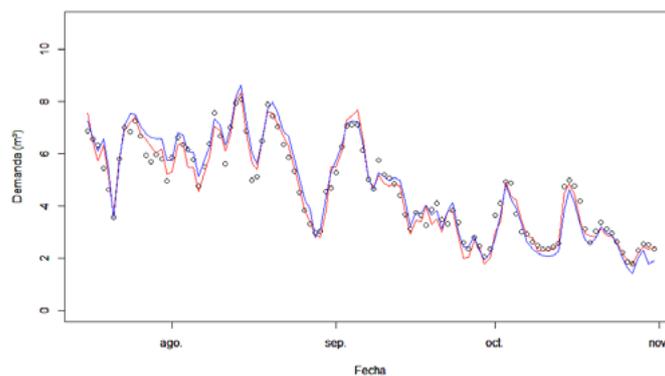
	ECONOMÉTRICO	BOOSTING MACHINE
GLOBAL	5.7%	3.8%
INVIERNO	5.2%	3.3%
VERANO	6.7%	4.7%

En ambos modelos, el error relativo es menor durante el invierno. Y el modelo Boosting Machine tiene mejor error relativo en cada temporada y a nivel global.

Por otra parte, se evaluaron los modelos a través de una muestra de test que no fue incluida en el entrenamiento de los mismos. Corresponde a las observaciones de consumo para el período comprendido entre el 16/07/2019 y el 31/10/2019 (108 observaciones).

En el siguiente gráfico se muestran las observaciones de demanda y las predicciones del modelo econométrico (azul) y las predicciones del modelo Boosting (rojo) correspondiente a las muestras de test de los modelos.

Ilustración 21: Comparación de predicciones entre modelos en sobre muestra de test



Fuente: Elaboración propia

Al igual que en el caso anterior, el ECM favorece el desempeño del modelo Boosting Machine, aunque es más elevado en comparación con la muestra de entrenamiento.

	ECONOMÉTRICO	BOOSTING MACHINE
ECM	0.187	0.133
RECM	0.346	0.288
RECM/DEMANDA	7.4%	6.2%

APLICACIÓN DEL MODELO MACHINE LEARNING PARA PRONÓSTICO DE DEMANDA A NIVEL PAÍS

Finalmente, se presenta una aplicación del modelo Boosting Machine para el pronóstico de la demanda prioritaria de gas natural a nivel país.

Se seleccionó este modelo ya que su aplicación es más simple, no requiere la transformación de variables ni cálculos intermedios, tales como los umbrales o HDD, que sí requiere el modelo econométrico.

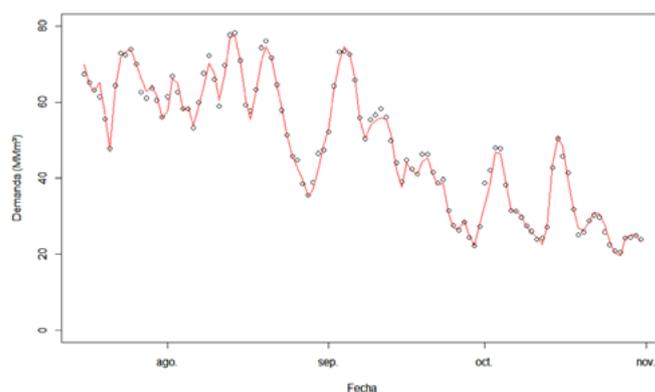
Se entrenaron 16 modelos Boosting Machine para diferentes subzonas del país: Gran Buenos Aires (Metrogas y Naturgy BAN), Litoral, Gasnea, Salta, Tucumán, Centro Norte, Centro Sur, Cuyana, Buenos Aires, Bahía Blanca, Neuquén, Buenos Aires Sur, Chubut, Santa Cruz y Tierra del Fuego.

Se evaluaron los modelos a través de una muestra de test que no fue incluida en el entrenamiento de los mismos. Corresponde a las observaciones de consumo para el período comprendido entre el 16/07/2019 y el 31/10/2019 (108 observaciones).

La demanda total país pronosticada surge de la suma de las demandas de cada subzona. La demanda de cada subzona surge de multiplicar el consumo específico de cada una por la cantidad de usuarios.

En el siguiente gráfico se muestran las observaciones de demanda total país (puntos) y las proyecciones de demanda (línea roja), correspondiente a las muestras de test de los modelos.

Ilustración 22: Demanda Prioritaria Total País



Fuente: Elaboración propia

En promedio, el desvío de los valores pronosticados respecto a la demanda real es del 3,5% (RECM/Demanda).

CONCLUSIONES

Los modelos recabados en la literatura nacional e internacional pudieron ser adaptados para modelar la demanda diaria de gas de la demanda prioritaria, en función de la base de datos disponible.

La variable temperatura, ingresada en los modelos de diferentes formas (temperatura efectiva, HDD, temperatura del día o temperatura rezagada), es la principal variable explicativa de la demanda de gas.

También se pudo verificar la importancia de la variable estacional (mayor propensión de los usuarios a utilizar la calefacción durante el período invernal que fuera del mismo). El tipo de día (sábado, domingo o feriado), también explica en nivel de consumo de gas.

Por otra parte, la tarifa del servicio de gas en términos reales es una variable sensible para los usuarios de la zona metropolitana para determinar su nivel de consumo, ya sea la vigente en el período o un período rezagada.

Respecto de los modelos evaluados, el modelo Boosting ha mostrado mejores resultados que el modelo Econométrico, evaluados con la medida del error cuadrático medio.

Ello no implica necesariamente que una técnica sea mejor que la otra, sino que las relaciones planteadas en modelo econométrico propuesto podrían no captar acabadamente la realidad. Por ejemplo, podrían existir relaciones funcionales que no sean lineales (podrían ser cuadráticas o productos cruzados entre variables) o existir más de dos umbrales (que fue el modelo propuesto).

Sin embargo, el modelo Boosting Machine es fácil de implementar y no requiere de mucho pre-proceso de los datos (determinar umbrales o calcular HDD).

BIBLIOGRAFÍA

Revisión bibliográfica sobre proyección de demanda de gas natural:

- *Estimación de funciones de demanda de gas natural para consumo residencial en la Argentina*. J. Arrufat, A. Neder. Universidad Nacional de Córdoba, 1994.
- *Gas Demand Forecasting Methodology*. National Grid, 2016.
- *Gas Demand Forecasting Methodology Information Paper*. Australian Energy Market Operator, 2019.
- *Modelo generalizado de predicción de consumos de gas natural a mediano y corto plazo*. S. Gil, J. Deferrari, L. Duperron. ENARGAS, 2002.

Modelo Econométrico:

- *Forecast: Forecasting functions for time series and linear models*. R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeeen. CRAN website, 2019.
- *Time Series Analysis*. J. Hamilton. Princeton University Press, 1994.
- *Tutorial for the R package chngpt*. Y. Fong. CRAN website, 2019.

Modelo Machine Learning:

- *An Introduction to Statistical Learning with Applications in R*. G. James, D. Witten, T. Hastie, R. Tibshirani. Springer, 2013.
- *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. J. Faraway. CRC Press, 2016.
- *GBM: Generalized Boosted Regression Models*. B. Greenwell, B. Boehmke, J. Cunningham. CRAN website, 2019.
- *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. T. Hastie, R. Tibshirani, J. Friedman. Springer, 2017.