

# 01

Fecha de presentación: mayo, 2023

Fecha de aceptación: julio, 2023

Fecha de publicación: noviembre, 2023

## MINERIA DE DATOS

SOBRE LOS EFECTOS CLIMATICOS Y BIOCLIMATICOS EN EL CONSUMO DOMESTICO DE AGUA EN LA EMPRESA AGUAS DE LA HABANA

### DATA MINING ABOUT THE CLIMATIC AND BIOCLIMATIC EFFECTS ON DOMESTIC WATER CONSUMPTION IN HAVANA WATERS COMPANY

Mario Ramos Joseph<sup>1</sup>

E-mail: [mramos@ahabana.co.cu](mailto:mramos@ahabana.co.cu)

ORCID: <https://orcid.org/0000-0002-2784-206X>

Alcides León Méndez<sup>2</sup>

E-mail: [aleonm@civil.cujae.edu.cu](mailto:aleonm@civil.cujae.edu.cu)

ORCID: <https://orcid.org/0000-0002-5072-6438>

Raisa Socorro Llanes<sup>2</sup>

E-mail: [raisa@ceis.cujae.edu.cu](mailto:raisa@ceis.cujae.edu.cu)

ORCID: <https://orcid.org/0000-0002-2627-1912>

Reniel Carvajal Alfonso<sup>1</sup>

E-mail: [rcarvajal@ahabana.co.cu](mailto:rcarvajal@ahabana.co.cu)

ORCID: <https://orcid.org/0000-0002-8656-4709>

<sup>1</sup>Aguas de La Habana. Cuba.

<sup>2</sup>Universidad Tecnológica de la Habana. Cuba.

<sup>2</sup> Universidad de Cienfuegos “Carlos Rafael Rodríguez” Cuba.

Cita sugerida (APA, séptima edición)

Ramos Joseph, M., León Méndez, A., Socorro Llanes, R., & Carvajal Alfonso, R. (2023). Minería de datos sobre efectos climáticos y bioclimáticos en el consumo doméstico de agua en la Empresa Aguas de la Habana. *Universidad y Sociedad*, 15(6),8-18.

#### RESUMEN

La aplicación de novedosas técnicas de inteligencia artificial se ha convertido en una poderosa herramienta de trabajo que soluciona múltiples problemas en todas las esferas del conocimiento, en el campo de la hidráulica en general y en los sistemas de abasto de agua en lo particular. En este reporte se aplican técnicas de minería de datos a un sector hidráulico de Aguas de La Habana, para establecer la relación entre el consumo doméstico de agua y algunas variables climáticas y bioclimáticas. En el trabajo se obtienen las principales variables climáticas que influyen en el consumo doméstico. Este resultado se constituye en la antesala necesaria para la construcción de un modelo de predicción del consumo doméstico de agua a partir del comportamiento de la variabilidad climática, una herramienta fundamental para actividades como la operación, el mantenimiento y el planeamiento de la demanda futura.

**Palabras clave:** Consumo doméstico, variables bioclimáticas, minería de datos.

#### ABSTRACT

The application of innovative Artificial Intelligence techniques has become a powerful work tool that solves multiple problems in all spheres of knowledge, in the field of hydraulics in general and in water supply systems in particular. This report applies data mining techniques to the Aguas de La Habana hydraulic sector to establish the relationship between domestic water consumption and some climatic and bioclimatic variables. In the work the main climatic variables that influence domestic consumption are obtained. This result constitutes the necessary prelude to the construction of a prediction model for domestic water consumption based on the behavior of climate variability, a fundamental tool for activities such as operation, maintenance, and planning of future water demand.

**Keywords:** Domestic consumption, bioclimatic variable, data mining.

## INTRODUCCIÓN

Tal como plantea Lizcano (2022), debido al aumento de la población mundial, la aparición de continuos eventos extremos asociados al cambio climático ha elevado el consumo del preciado líquido a niveles importantes.

Hoy se aprecia que la demanda de agua está aumentando y las fuentes de la misma están disminuyendo, por ende, se evidencia que está en escasez y aunque el 70% del planeta está cubierto de agua solo el 2,5% de la misma es agua potable, y de esta se tiene acceso a menos del 1% para uso humano de acuerdo a lo que plantean Adamowski (2012); Morote (2017)

Se aprecian factores que inciden en esta situación tales como: infraestructuras en deterioro, presiones altas, servicio intermitente, errores de medición, uso ilegal, operaciones y prácticas de gestión ineficiente, según plantean Haque (2015); Ramos (2016)

Con la aparición de nuevos protagonistas como el cambio climático, y el aumento en magnitud y frecuencia de la aparición de eventos extremos entre ellos las sequías, los estudios que relacionan el consumo residencial con los factores climáticos se convierten en una imperiosa necesidad para toda empresa de abasto, de acuerdo a lo planteado por Ramos (2016)

Cuba que cuenta con sistemas de abasto de agua de más de 100 años de antigüedad, no está ajena a la problemática de las pérdidas de agua, de acuerdo a estimaciones del INRH, se pierden 2 500 hectómetros cúbicos anualmente, que equivalen a 730 millones de pesos.

Aguas de La Habana es la empresa que gestiona el abasto de agua de los 15 municipios de la capital, y también presenta un gran volumen de pérdidas de agua. De acuerdo a sus estimaciones este volumen alcanza niveles de 333,2 hectómetros cúbicos anuales.

El consumo del agua puede verse afectado por varias causas, se puede apreciar que la mayoría de los estudios precedentes del tema son fundamentalmente realizados en países desarrollados USA, Reino Unido y Australia de acuerdo a Soto-Montes (2016); Bich-Noc (2018).

Se pueden reconocer un conjunto de variables que aparecen sistemáticamente en los diferentes estudios consultados. Por ejemplo, se identifica el nivel de ingreso, la edad, la presencia de jubilados y la cantidad de personas por vivienda en algunos casos según plantea Bich-Noc (2018) y el impacto de la temperatura, las precipitaciones según lo plantea en su trabajo House-Peters (2010).

Otros autores, tal como lo hace Morote (2017), hablan de una psicología que determina el consumo. De Ore

(1999), relaciona el consumo con características de las viviendas como su antigüedad, la cantidad de habitaciones, baños o su nivel de consumo interno. Otra nueva arista es la que plantea la existencia de niveles de consumo por un componente social, según exponen Morales & Gori (2021). En otras palabras, el consumo doméstico se ha convertido en una variable multifactorial tomando en cuenta todos los factores que son capaces de influir en el mismo.

Según Soto-Montes (2016) la variabilidad climática toma cada vez mayor importancia y aporta seguimiento de estudiosos y científicos debido a su nexo actual con el cambio climático, en un breve análisis de algunos de los principales estudios realizados sobre el tema se tiene lo siguiente.

En un estudio realizado por Balling (2006), sobre la variabilidad del clima y uso residencial del agua, sobre la base de una serie de datos tomados durante 25 años (1980 a 2004), se obtuvieron coeficientes de correlación entre el uso del agua con la temperatura media anual y la precipitación anual de +0,55 y -0,69 respectivamente, lo que ilustra un aumento del uso del agua con altas temperaturas y baja precipitación.

Mientras tanto en Alemania, en un estudio sobre los factores determinantes en la demanda de agua, encontraron, al emplear dos diferentes modelos, que en uno de los modelos la lluvia resultó estadísticamente significativa, con un signo negativo y también se verificó que un descenso de la lluvia en el verano podría resultar en un aumento del consumo de agua diario por persona en cerca de 0,7 litros de acuerdo con el primer modelo, que aplica una única ecuación de regresión y técnicas de instrumentación variable. De otra parte, de acuerdo con el segundo modelo, solo utiliza técnicas de instrumentación variable, el resultado fue de cerca de 1,2 litros, mientras que la temperatura no fue estadísticamente significativa en el consumo de agua (Schleich, 2007).

En otro escenario, a partir de un estudio realizado en la ciudad de Hillsboro, Oregón, en el área metropolitana de Portland, se concluyó que, las características de las propiedades físicas y las variables socioeconómicas tienen un impacto en la demanda de agua y esta relación varía en dependencia de la estación del año y el promedio de la precipitación (House-Peters, 2010).

En un estudio realizado en Canadá, que incluyó tres ciudades (Calgary, Ottawa y Montreal), todas con características climáticas y socioeconómicas distintas, se llegó a la conclusión de que la demanda urbana de agua tenía una correlación lineal con la temperatura media diaria. En contrapartida, para los tres casos no se encontró ninguna

correlación entre la demanda de agua y la precipitación (Adamowski J., 2012).

Sarker (2013), realizó un estudio en Greater Melbourne (Australia), sobre los valores límites de la temperatura y lluvia a partir de los cuales se producen efectos en el consumo del agua. Utilizando una serie de datos diarios sobre el consumo de agua, temperatura y lluvia de 30 años (1980 a 2009), de sus análisis se puede concluir que el valor límite de la temperatura para el cual se producía efecto en el consumo del agua era de 15,53°C, demostrando que el uso del agua aumenta al aumentar la temperatura y para valores de temperatura por debajo del valor límite no ejercían influencia. Para la lluvia el valor límite fue de 4,08 mm, y se apreció que para valores de precipitación debajo de este, el uso del agua aumenta al decaer la precipitación.

Chang (2014), en análisis con bases de datos mensuales, concluye que la temperatura máxima y la precipitación en conjunto explican menos de la mitad del consumo de agua mensual, por lo que admite que existen otras variables hidrometeorológicas, que explican mejor las variaciones adicionales en el consumo de agua mensual.

En Australia en un estudio realizado en New South Wales, empleando Análisis de los Componentes Principales, encontraron que la temperatura máxima media mensual y la lluvia, presentan una baja correlación con el consumo de agua mensual por vivienda, siendo positiva para la temperatura y negativa para la lluvia. La temperatura fue la variable de mayor efecto comparativamente con la lluvia, indicando un aumento de la demanda, con el aumento de esta última (Haque, 2015).

Brandner (2016), en Suecia, realizó un estudio donde se utilizó un modelo de análisis de regresión lineal simple para tres áreas municipales estudiadas. Los resultados indicaron que el consumo de agua tendía a aumentar con el aumento de la temperatura.

Del estudio realizado por Soto-Montes (2016) en la ciudad de México (caracterizada por un clima en el norte semiárido con altas temperaturas y moderado húmedo con bajas temperaturas al sur), encontró que los efectos sobre el consumo de agua, son influenciados positivamente por la temperatura y negativamente por la precipitación. El 10% de aumento en la temperatura, está asociado con un aumento de 5% en el consumo doméstico de agua y el aumento de 10% en la precipitación, está asociado a un descenso de 1,2% en la cantidad demandada. La temperatura media anual en la zona de estudio fue de 16,7°C para la región norte y de 11°C para la región sur.

Bich-Ngoc (2018), en una revisión bibliográfica, concluye en la misma que los usos finales del agua en el interior de la vivienda varían con el clima, evidenciándose que la ocurrencia de la lluvia provoca una inmediata disminución de los consumos promedios de agua y además, que la mayoría de los estudios analizados se enfocaron a zonas con clima árido y semiárido, dejando un vacío de conocimiento para otras áreas geográficas con climas diferentes.

Mientras tanto en Cuba, se realizó un estudio que por primera vez determina la relación entre el consumo per cápita doméstico y variables climáticas y bioclimáticas. Es relevante que estas últimas son variables que incorporan el impacto de la sensación térmica en relación con el consumo, el estudio concluyó que la humedad relativa máxima y la temperatura máxima aisladamente afectan el consumo per cápita con una razón creciente en el orden de 15,9% y 16,7% respectivamente (Guilherme Foquiço, 2017).

Resulta relevante, que el efecto simultáneo de ambas variables, provocó aumentos significativos del consumo per cápita con una correspondencia de 28,6%. La precipitación media causa variación sobre el consumo per cápita provocando disminución en el mismo en un orden de 2,1% con el aumento de la precipitación media. Combinada con la humedad relativa media y la temperatura media, es la variable de menor efecto, pero con tendencia siempre de disminuir el consumo con su aumento. El consumo per cápita aumenta con el aumento de la variable bioclimática Temperatura Efectiva (TE) en el orden de 10,7%. De otra parte, la TE combinada con la precipitación media provoca aumento del consumo, contrario a la precipitación, siendo la razón de variación del consumo respecto a las dos variables de 12,3%. Mientras tanto, la Temperatura Efectiva Equivalente (TEE), otra variable bioclimática empleada, afecta al consumo per cápita de forma positiva en un orden de 13,1%, en escenario simultáneo con la precipitación provoca aumento del consumo con la subida de la TEE y la disminución de la precipitación con una razón de variación del consumo de 14,6%. Para este estudio se concluye que la TEE es la variable bioclimática de mayor incidencia en el consumo (Guilherme Foquiço, 2017).

De acuerdo al criterio de expertos del Instituto de Meteorología de Cuba (INSMET), el archipiélago cubano presenta un clima tropical estacionalmente húmedo con influencia marítima y rasgos de semi continentalidad y con una elevada radiación solar en el año, además de la influencia estacional de fenómenos meteorológicos tropicales y extra tropicales.

Algo más reciente, Pomares (2022), trabajando con datos de la empresa cubana Aguas de La Habana, incluyó otra variable bioclimática, la Amplitud Térmica (AT). Para el procesamiento de las bases de datos obtenidos en la empresa empleó técnicas de minería de datos, logrando obtener un coeficiente de correlación de 29% para explicar la relación entre el consumo doméstico y un grupo de variables climáticas y bioclimáticas. La novedad en este caso, consiste en la inclusión de la TE, TEE y AT como variables independientes en este estudio, resultando una variable más que las empleadas por Guilherme (2017).

Rodríguez (2022), plantea que ha sido amplio el uso en Cuba, de índices para evaluar las sensaciones térmicas de las personas desde los años 80 del pasado siglo, entre ellos se destacan, el cálculo del Balance de Calor del Cuerpo Humano, el Índice Térmico Universal, Índice de Calor Sofocante y en especial el Índice de Temperatura Efectiva (TE) y el Índice de Temperatura Efectiva Equivalente (TEE). Hoy en día el índice TEE se encuentra difundido en todo el mundo, tiene gran alcance en el análisis del comportamiento de la sensación térmica ya que resulta más completo, considerando al individuo expuesto al movimiento del aire en un medio exterior a la sombra. Los primeros estudios, se desarrollaron desde el punto de vista climático descriptivo con el análisis de la variación geográfica del confort térmico según la TEE.

Todos estos trabajos y criterios constituyen la base necesaria para poder establecer un conocimiento más preciso de la multilateralidad de factores que determinan el consumo doméstico. Este conocimiento se puede convertir en una extraordinaria herramienta para la caracterización y previsión del consumo en cada sistema de abasto.

El estudio que aquí se presenta, tiene como objetivo profundizar en el conocimiento de los principales factores bioclimáticos y climáticos que intervienen en la demanda doméstica de agua y su peso relativo mediante la aplicación de técnicas de minería de datos.

## MATERIALES Y MÉTODOS

El trabajo se enmarca el estudio del sector hidráulico denominado AB entre los sectores que gestiona la empresa Aguas de La Habana.

Este sector se encuentra enclavado en uno de los municipios más poblados de La Habana. Presenta acometidas metradas con un historial de consumo desde el año 2000; se caracterizan por un nivel socioeconómico muy similar; está conformada por viviendas de tipología media similar; existencia de al menos de un baño y cerca perimetral,

con una baja presencia de viviendas unifamiliares y predominio de viviendas multifamiliares.

Cuenta con red de servicio continuo, todas las acometidas se encuentran metradas, con un total de 159 clientes. Del total de clientes hay 46 metrados en viviendas unifamiliares y 113 metrados en viviendas multifamiliares (Guilherme Foquiço, 2017).

Para el estudio de los clientes metrados se eliminaron todos aquellos que por diferentes causas no mantenían las mismas condiciones durante todo el periodo en cuanto a calidad de las mediciones, cambios de uso y propiedad, por contadores sustituidos, o rotos durante periodos que hacían muy difícil la determinación precisa de su registro mensual, después de esta depuración se seleccionaron 34 clientes que son los que definitivamente conforman la muestra del estudio utilizando sus datos mensuales y obteniendo la media de consumo de los 34 clientes para llegar a un único valor medio mensual expresado en (lppm) litros por persona por mes.

Los datos de consumo mensual de los 34 clientes fueron aportados por la Dirección Comercial de la empresa, a partir de los análisis de consumo que realiza la empresa para garantizar la facturación sobre la base de registros medidos.

En el caso de las variables meteorológicas, los datos se obtuvieron de los registros diarios de la Estación Meteorológica de Casablanca, adscrita a la red climatológica del Instituto de Meteorología (INSMET) a partir de su frecuencia diaria se transformaron a valores medios para su posterior contraste con el consumo medio mensual.

Para el procesamiento y análisis de la información recopilada, así como la obtención de asociaciones entre las variables, con base a criterios matemáticos y estadísticos; se utilizó la herramienta Knime.

El estudio comprende un período que abarca desde 1ro enero de 2008 hasta 31 diciembre de 2020, en función del objetivo trazado se define el consumo en (lppm) como variable dependiente.

Los consumos fueron aportados en metros cúbicos por mes para cada cliente o vivienda de la muestra, y se transformaron en litros por persona por mes (lppm) para cada cliente obteniendo un promedio mensual para su mejor análisis y comparación con estudios similares de otras latitudes, este promedio mensual durante los 13 años que comprende el estudio aporta un total de 156 filas a la vista minable. Este consumo representa el volumen de agua consumido por una persona en un mes. El resto de las variables analizadas se describen a continuación como variables independientes:

- Lámina de precipitación (Lluvia): representa que cantidad de lluvia es precipitada en un área determinada y es medida en mm con ayuda de un pluviómetro en la estación meteorológica. Se obtiene en forma de registros diarios y se calculó su promedio mensual.
- Ocurrencia de Lluvia (OcLI): esta variable se obtiene de los registros de lámina de precipitación. Se transformó a valor binario. El 0 representa ausencia de lluvia y el 1 significa que llovió, es adimensional. Se incluye esta variable como cuantitativa con un peso relativo que solo toma 2 valores y se puede utilizar en una ecuación de regresión.
- Intensidad de precipitación (r 24h): representa la cantidad de agua que se precipita por unidad de tiempo y está expresada en milímetros por hora. Se obtiene a partir de registros diarios los cuales son transformados obteniendo su promedio en función de la unidad temporal.
- Temperaturas mínimas (T mín.), medias (T med.) y máximas (T máx.): representa los registros históricos de temperatura atmosférica medida con termómetro en la estación meteorológica en °C. Se obtiene en forma de registros diarios y se transforman obteniendo su promedio en función de la unidad temporal.
- Humedades relativas mínimas (HR mín.), medias (HR med.) y máximas (HR máx.): es la relación entre la cantidad de vapor de agua contenida en el aire y la máxima cantidad que el aire sería capaz de contener a esa temperatura dada en porcentaje se expresa en %. Se obtiene en forma de registros diarios y se transforman obteniendo su promedio en función de la unidad temporal.
- Velocidad del viento media (FF med.) y máxima (FF máx.): mide la componente horizontal del desplazamiento del viento en un punto e instante determinados y está muy relacionada con la sensación térmica. Esta variable se mide mediante un anemómetro y está dada en kilómetros por hora. Se obtiene en forma de registros diarios y se transforma obteniendo su promedio en función de la unidad temporal propuesta.
- Dirección del viento (DD máx.): mide la componente horizontal de la velocidad del viento expresada en grados sexagesimales contados en el sentido de las manecillas del reloj a partir del norte geográfico utilizando los rumbos de la Rosa de los Vientos. Se obtuvieron en forma de registros diarios y se transformaron obteniendo su promedio en función de la unidad temporal que se propuso.
- Amplitud Humedad Relativa (AHr): es la diferencia numérica entre los valores máximos y mínimos de la humedad registrados en un punto dado durante un período de tiempo. La contemplación de esta variable fue concebida gracias al aporte de expertos del INSMET, se expresa en %.
- Temperatura efectiva (TE): se define como la sensación que se experimenta a una temperatura y humedad dadas, con aire saturado y en calma. Es válido para un sujeto joven, sano, aclimatado al sitio, con vestimenta ligera y sometida a una actividad ligera. Esta variable está dada en °C clasifica como una variable bioclimática y es un aporte novedoso en la relación entre el consumo doméstico y las variables climatológicas.
- Temperatura efectiva equivalente (TEE): se define como la temperatura del aire a la que el balance de energía humano, para unas condiciones asumidas bajo techo, está equilibrado con unas mismas temperaturas de la piel y tasas de sudoración como las calculadas en condiciones a cielo abierto. Esta variable está expresada en °C, al igual que la anterior es una variable bioclimática al incorporar las sensaciones térmicas de los seres humanos al estudio.
- Amplitud Térmica (AT): representa el diferencial de temperatura presente en la atmósfera aportando mayor peso a la variabilidad de la misma y resaltando comportamientos anómalos de fluctuaciones drásticas, se expresa en °C. Es la diferencia numérica entre los valores máximos y mínimos de temperatura observados en un punto dado durante un período de tiempo. La contemplación de esta variable fue concebida gracias al aporte de los expertos del INSMET, por tanto al igual que las dos anteriores es una variable bioclimática obtenida mediante transformación de variables, es afectada por factores como la influencia del mar, la topografía, la nubosidad, y la latitud, es muy importante para la agricultura y se ha incluido por su impacto en los seres vivos en este caso las personas, no se ha encontrado este enfoque en estudios anteriores que relacionen clima y consumo de agua urbano.

Finalmente se obtiene un conjunto de datos que conforma la vista minable compuesta por 156 filas o registros y 19 columnas, desagregadas en 16 variables meteorológicas independientes, 2 variables temporales y 1 variable dependiente.

Es importante reconocer que la temperatura efectiva, la temperatura equivalente, y la amplitud térmica clasifican como variables bioclimáticas obtenidas por transformación de otras variables. Las 3 tienen en común que implican una respuesta a los cambios en sensación térmica que es capaz de apreciar el cuerpo humano ante la variabilidad climática. En este estudio se analiza su relación con el consumo doméstico. No se ha encontrado en la literatura consultada referencias que utilicen este análisis, siendo este un aspecto relevante a tener en cuenta por su incidencia en el consumo doméstico.

En esta sección se presenta una propuesta para determinar la influencia de las variables climáticas y bioclimáticas en la predicción o estimación del consumo doméstico.

Esto se logra utilizando técnicas de selección de variables tales como: el de introducción progresiva de variables (forward) y el de eliminación progresiva de variables (backward) durante el proceso de estimación del consumo. del contraste entre ambos métodos se determinan las variables que mayor influencia tienen en la estimación del consumo.

El modelo de predicción del consumo se realizó utilizando diferentes técnicas de regresión, de las descritas en Witten (2016), entre las que se encuentran:

**RBF Regressor:** Algoritmo que utiliza las redes neuronales de base radial para la clasificación entrenado en la supervisión más completa utilizando el optimizador de WEKA, minimizando el error cuadrado con el método de BFGS. En la aplicación todos los atributos deben ser normalizados en una escala de (0,1)

**Isotonic Regression:** Este algoritmo aprende de un modelo de regresión isotónico. Los picos en el atributo dan como resultado el mínimo del error cuadrático. No se permiten datos faltantes. Solo se puede tratar con atributos numéricos.

**Linear Regression:** Clasificador que utiliza la regresión lineal para predecir. Utiliza el criterio de Akaike para seleccionar el mejor modelo y puede tratar con ejemplos pesados (basado en WEKA 3,7)

**Pace Regression:** Es un clasificador que construye modelos lineales de regresión por pasos para utilizarlos para predicción. En condiciones de regularidad la regresión por pasos es óptima cuando el número de coeficientes tiende a infinito. Consiste en un grupo de estimadores que son globalmente óptimos u óptimos bajo ciertas condiciones. En condiciones normales la regresión de paso no maneja los valores perdidos y los atributos nominales no binarios.

**Additive Regression:** Clasificador de Meta que refuerza la actuación de un clasificador de base de regresión. Cada iteración encaja un modelo para los residuales dejados por el clasificador en la iteración previa. La predicción es cumplida agregando la predicción de cada clasificador.

**Regression by discretization:** Esquema de la regresión que emplea cualquier clasificador en una copia de los datos que tienen el discretizado de atributo de clase

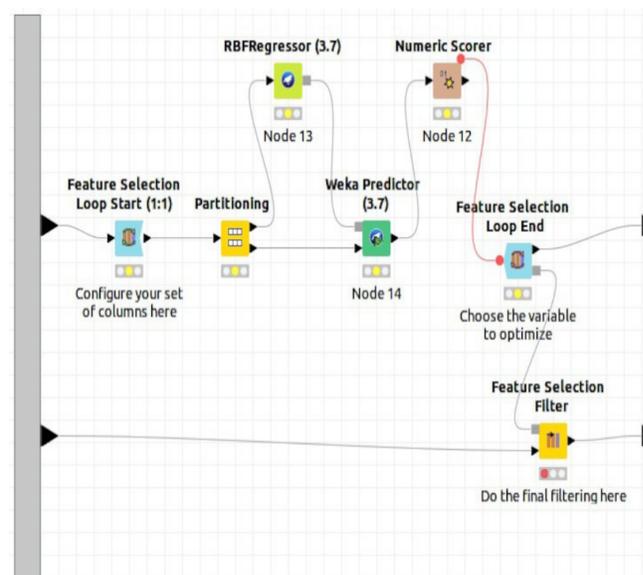
El valor predicho es el valor esperado del valor de la media de la clase para cada intervalo discretizado, basado

en las probabilidades de predicción de cada intervalo. Esta clase también apoya la estimación de la densidad condicional construyendo un estimador de densidad univariado de los valores designados en los datos de entrenamiento pesado por las probabilidades de la clase.

**Ensemble regressor:** Este algoritmo combina las acciones de varios clasificadores que usan el método de selección de conjunto o ensemble.

**M5P:** En este clasificador los implementos basan las rutinas para generar un modelo de árboles y reglas. El algoritmo original fue creado por R y Quinlan y Yong Wang hicieron las mejoras.

En la Figura 1 se muestra un ejemplo del flujo Knime utilizado para realizar la selección de variables en forward empleando la técnica RBF Regresor.



**Figura 1.** Ejemplo de flujo Knime para el forward con algoritmo de RBF Regresor.

Fuente: Elaboración propia

Una vez ejecutado el método de selección de variables se seleccionan aquellos escenarios que tienen un valor de **R-Squared** ajustado por encima de 54,7% para el forward y 50% para el backward. Las frecuencias de aparición de las variables en cada una de las técnicas de regresión analizadas se muestran en las Tablas 1 y 2 respectivamente. En ambas se presentan en columnas el número de escenarios en que aparece cada variable según la técnica aplicada. En la columna total se cuantifica el número de escenarios en que aparece cada variable y el porcentaje representa la frecuencia total de aparición de cada variable con respecto al total de escenarios.

## RESULTADOS Y DISCUSIÓN

Los modelos propuestos en la sección anterior se aplicaron al conjunto de datos descritos en la sección Área de estudio. Los modelos se construyen utilizando el 80% de los datos disponibles seleccionados de manera aleatoria y se utiliza el 20% restante para evaluar la calidad del modelo en cada caso. En estos experimentos se utilizó el *R-Squared* ajustado para evaluar la calidad de la estimación obtenida por cada uno de los modelos. En este caso solo se analizan aquellas combinaciones de variables (escenarios) en los que el modelo de predicción construido a partir de ellas logra predecir con un valor de *R-Squared* igual o mayor a 54,7% para el forward y 50% para el backward esto garantiza que al menos un 50% de los valores de la variable dependiente responden a este modelo.

En la (Tabla 1) se presenta el número de escenarios que cumplen los requisitos explicados anteriormente, además se presenta la frecuencia de aparición de las variables utilizadas en cada escenario lo que permitirá reconocer aquellas que predominan.

Tabla 1. Frecuencia de aparición de las variables en los escenarios obtenidos con el método de selección de variables Forward.

| Variable/<br>Escenarios | RBF Regressor<br>(6) | Isotonic Regression<br>(15) | Linear Regression<br>(6) | Pace Regression<br>(10) | Ensemble Regressor<br>(13) | M5P (10) | Total (60) | % Aparición |
|-------------------------|----------------------|-----------------------------|--------------------------|-------------------------|----------------------------|----------|------------|-------------|
| LPPM                    | 6                    | 15                          | 6                        | 10                      | 13                         | 10       | 60         | 100,00      |
| T mín                   | 5                    | 2                           | 0                        | 10                      | 3                          | 7        | 27         | 45,00       |
| T med                   | 2                    | 8                           | 1                        | 4                       | 9                          | 0        | 24         | 40,00       |
| T máx                   | 0                    | 10                          | 2                        | 9                       | 7                          | 3        | 31         | 51,67       |
| AT                      | 4                    | 13                          | 5                        | 1                       | 3                          | 0        | 26         | 43,33       |
| Hr mín                  | 0                    | 5                           | 4                        | 3                       | 11                         | 10       | 33         | 55,00       |
| HRmed                   | 2                    | 4                           | 1                        | 8                       | 3                          | 8        | 26         | 43,33       |
| Hr máx                  | 6                    | 14                          | 5                        | 10                      | 4                          | 0        | 39         | 65,00       |
| Ahr                     | 2                    | 11                          | 0                        | 5                       | 5                          | 6        | 29         | 48,33       |
| OcLI                    | 3                    | 12                          | 6                        | 8                       | 1                          | 1        | 31         | 51,67       |
| TE                      | 2                    | 1                           | 3                        | 10                      | 9                          | 0        | 25         | 41,67       |
| TEE                     | 2                    | 0                           | 6                        | 6                       | 12                         | 0        | 26         | 43,33       |
| r 24h                   | 3                    | 9                           | 4                        | 10                      | 4                          | 5        | 35         | 58,33       |
| Rain                    | 1                    | 0                           | 6                        | 2                       | 10                         | 9        | 28         | 46,67       |
| FFmed                   | 3                    | 6                           | 5                        | 8                       | 6                          | 0        | 28         | 46,67       |
| FFmáx                   | 0                    | 15                          | 4                        | 10                      | 5                          | 0        | 34         | 56,67       |
| DDmáx                   | 0                    | 7                           | 6                        | 7                       | 3                          | 4        | 27         | 45,00       |

Fuente. Elaboración propia.

Del análisis de la (Tabla 1) se presenta el Forward para diferentes algoritmos. Para cada algoritmo se muestra en paréntesis el número de escenarios con un R-Squared mayor igual a 54,7%.

Se puede resumir que todas las variables seleccionadas influyen en el consumo residencial con una frecuencia de aparición que transita desde el 40% hasta el 65 %, las de mayor peso son Hrmax, r24h, FFmax y Hrmin.

En el caso de las variables bioclimáticas se aprecia que las de mayor frecuencia son la AT y la TEE ambas con el mismo valor seguidas por la TE en un rango que abarca desde el 43,33% hasta el 41,67%.

Las principales técnicas que justifican los escenarios utilizados para predecir las variables necesarias de ese universo se muestran a continuación:

En la (Tabla 2) se aprecian las principales técnicas utilizadas en el proceso de la selección de variables utilizando el forward, como se muestra en función de varias métricas lo que permite reconocer la mejor técnica en función del mismo.

Tabla 2. Principales técnicas utilizadas en la selección de variables.

| Modelo Forward      |          |           |             |         |
|---------------------|----------|-----------|-------------|---------|
| Técnicas empleadas  | RMSE     | R-Squared | MSE         | MAE     |
| Isotonic Regression | 909,577  | 0,612     | 827329,971  | 741,794 |
| Pace Regression     | 1179,373 | 0,405     | 1390921,441 | 843,374 |
| Ensemble Regressor  | 738,663  | 0,535     | 545623,727  | 601,746 |
| RBF Regressor       | 1139,657 | 0,257     | 1298817,513 | 917,707 |
| Linear Regression   | 1112,163 | 0,27      | 1236905,809 | 925,586 |
| M5P                 | 871,273  | 0,541     | 759117,408  | 702,587 |

Fuente. Elaboración propia.

De estas técnicas como se puede apreciar en la tabla 2 del modelo forward el mejor desempeño lo tuvo la técnica de Isotonic Regression a partir de utilizar el *R-Squared* como métrica.

En la (Tabla 3) se muestra el número de escenarios que cumplen los requisitos. Además, se presenta la frecuencia de aparición de las variables utilizadas en cada escenario lo que permitirá reconocer todas aquellas que predominan.

Tabla 3. Frecuencia de aparición de las variables en los escenarios obtenidos con el método de selección de variables Backward.

| Variables/<br>Escenarios | RBF Regressor<br>(6) | Isotonic Regression<br>(13) | Linear Regression (7) | Pace Regression (10) | Ensemble Regressor<br>(11) | M5P (8) | Total (55) | % Aparición |
|--------------------------|----------------------|-----------------------------|-----------------------|----------------------|----------------------------|---------|------------|-------------|
| LPPM                     | 6                    | 13                          | 7                     | 10                   | 11                         | 8       | 55         | 100,00      |
| T mín                    | 3                    | 13                          | 0                     | 1                    | 5                          | 0       | 22         | 40,00       |
| T med                    | 0                    | 8                           | 2                     | 9                    | 3                          | 8       | 30         | 54,55       |
| T máx                    | 0                    | 12                          | 2                     | 6                    | 11                         | 8       | 39         | 70,91       |
| AT                       | 5                    | 11                          | 2                     | 10                   | 8                          | 8       | 44         | 80,00       |
| HRmín                    | 6                    | 0                           | 1                     | 7                    | 6                          | 1       | 21         | 38,18       |
| HRmed                    | 4                    | 2                           | 2                     | 8                    | 4                          | 6       | 26         | 47,27       |
| Hrmáx                    | 4                    | 6                           | 0                     | 10                   | 11                         | 2       | 33         | 60,00       |
| Ahr                      | 1                    | 13                          | 0                     | 5                    | 11                         | 5       | 35         | 63,64       |
| OcLI                     | 3                    | 5                           | 4                     | 9                    | 9                          | 8       | 38         | 69,09       |

|       |   |    |   |    |    |   |    |       |
|-------|---|----|---|----|----|---|----|-------|
| TE    | 1 | 4  | 0 | 10 | 10 | 8 | 33 | 60,00 |
| TEE   | 0 | 10 | 4 | 5  | 0  | 8 | 27 | 49,09 |
| r 24h | 6 | 9  | 4 | 10 | 11 | 4 | 44 | 80,00 |
| Rain  | 5 | 13 | 0 | 3  | 5  | 8 | 34 | 61,82 |
| FFmed | 6 | 1  | 3 | 2  | 11 | 3 | 26 | 47,27 |
| FFmáx | 5 | 11 | 0 | 9  | 11 | 8 | 44 | 80,00 |
| DDmáx | 1 | 7  | 4 | 4  | 1  | 8 | 25 | 45,45 |

Fuente. Elaboración propia

La (Tabla 3) muestra el Backward para diferentes algoritmos. Para cada algoritmo se presenta en paréntesis el número de escenarios con un R-Squared mayor igual a 50%. A partir de la misma se puede apreciar que las variables seleccionadas influyen en el consumo residencial con una frecuencia de aparición que transita desde el 38,2% hasta el 80%, las de mayor peso son AT, r24h, y FFmax.

El análisis de las variables bioclimáticas muestra que la de mayor frecuencia de aparición es la AT seguida por la TE y la TEE en valores que abarcan desde los 49,1% hasta el 80%, se aprecia en este algoritmo un mayor impacto de estas variables en el consumo doméstico.

Tabla 4. Principales técnicas utilizadas en la selección de variables.

| Modelo Backward     |          |           |             |         |
|---------------------|----------|-----------|-------------|---------|
| Técnicas empleadas  | RMSE     | R-Squared | MSE         | MAE     |
| Isotonic Regression | 897,542  | 0,484     | 805581,861  | 693,815 |
| Pace Regression     | 1017,907 | 0,305     | 1036134,27  | 717,14  |
| Ensemble Regressor  | 1044,45  | 0,387     | 1090875,703 | 813,569 |
| RBF Regressor       | 790,929  | 0,442     | 625569,104  | 653,92  |
| Linear Regression   | 1195,058 | 0,311     | 1428162,54  | 917,851 |
| M5P                 | 1026,63  | 0,408     | 1052806,138 | 826,275 |

Fuente. Elaboración propia

El análisis de la (Tabla 4) muestra las principales técnicas utilizadas para el Backward en las que similar a lo que ocurre en la tabla 2 se aprecia que la mejor técnica en función del *R-Squared* es la de Isotonic Regression, siendo este resultado similar al obtenido con la aplicación del forward, todo ello con el *R-Squared* como métrica.

El análisis integral de los resultados de las (Tablas 1 y 3) indica que en ambos métodos de selección se destacan con mayor frecuencia la r24h, y la FFmax que son las variables con mayor frecuencia de aparición seguidas por Hrmax, Hrmin, y la AT, lo que permite reconocerlas como las que se deben tener en cuenta para la confección de un modelo de predicción del consumo doméstico, en otras palabras, se reconocen como variables predictoras, para la base de datos existente del caso de estudio analizado. La FFmax es una variable que está muy relacionada con la sensación térmica lo que alimenta el criterio de la importancia que tienen para explicar el consumo doméstico aquellas variables que incorporan el comportamiento bioclimático de las personas.

Un análisis de las variables bioclimáticas seleccionadas por ambos métodos forward y backward muestra que AT presenta la mayor frecuencia de aparición y en un caso seguida por la TEE y la TE o por la TE y la TEE de acuerdo al caso. Este resultado indica que pueden ser utilizadas como variables predictoras en la conformación de un modelo de predicción del consumo doméstico, ya sea combinadas con las de mayor frecuencia o como bloque único de elementos con efectos sensoriales en su relación con el consumo residencial, este resultado es la primera vez que se obtiene relacionando el consumo doméstico con variables bioclimáticas.

Acorde a este resultado en Cuba existen estudios que han demostrado cambios en el comportamiento de los valores extremos de temperaturas como los desarrollados por Rodríguez (2022). En estos trabajos se demuestra un incremento en la temperatura mínima, al mismo tiempo que existen condiciones meteorológicas que tienen características de continentalidad sobre todo en aquellas regiones que se encuentran alejadas de las costas. A la vez que existen condiciones similares a las Islas de Calor que se producen en las grandes ciudades. Todo lo antes expuesto contribuye al incremento de la diferencia que se produce entre la temperatura máxima del día y los valores mínimos de esta variable, lo que conlleva al incremento de la AT.

Los resultados que se obtienen están en concordancia con estudios previos tales como los presentados por Balling (2006); House-Peters (2010); Bradner (2016) entre otros, que tienen en cuenta los efectos de la temperatura en el consumo, presentes en la AT como variable bioclimática transformada, mientras que para la  $r_{24}$  es similar a los estudios de Sarker (2013); Chang (2014); Haque (2015). También son consecuentes con los resultados anteriores obtenidos en Cuba por Guilherme (2017); Pomares (2022).

## CONCLUSIONES

La mayoría de los estudios realizados a nivel global en este campo están dirigidos a la caracterización del consumo doméstico y su aplicación como una potente herramienta de predicción. Destacándose el carácter multifactorial del consumo doméstico y la necesidad de un enfoque holístico que permita obtener los mayores beneficios a partir del conocimiento adquirido en su correcta caracterización. En Cuba, a pesar de su prioridad para conocer cuáles son las principales variables meteorológicas que definen el comportamiento del consumo doméstico, algo fundamental para actividades como el planeamiento y la operación de los sistemas de distribución, aun los estudios son insuficientes.

El trabajo presentado tiene utiliza por primera vez de variables bioclimáticas en su relación con el consumo doméstico, para lo cual no se encontró precedente en la literatura consultada y se están aportando herramientas para predecir el comportamiento de la demanda futura en función de la variabilidad climática.

El uso de técnicas de minería de datos en la selección de variables permitió determinar aquellas variables climáticas y bioclimáticas que impactan en el consumo destacando las que más lo afectan y se destacan como

predictoras en una muestra del sector AB de Aguas de La Habana.

Los experimentos fueron realizados para 6 técnicas de regresión, se logró identificar que las variables que más frecuencia de aparición tienen en el modelo, son:  $r_{24}$ , FFmax, AT, Hrmax y Hrmin y de acuerdo al modelo son variables predictoras.

Relevante es el caso de las variables bioclimáticas, de las cuales la que más aparece es la AT por lo que definitivamente es la variable que se impone en este acápite seguida por la TE y la TEE, por lo que son las que se proponen para incluir en un modelo de predicción del consumo que se está construyendo y que debe dar respuestas en breve plazo. A destacar el hecho de que la FFmax tiene en cuenta el comportamiento sensorial lo que demuestra la importancia de las variables bioclimáticas en la predicción del consumo doméstico.

## REFERENCIAS BIBLIOGRÁFICAS

- Adamowski, J. A. (2012). "A Spectral Analysis Based Methodology to Detect Climatological Influences on Daily Urban Water Demand", *International Association for Mathematical Geosciences*. <https://doi.org/10.1007/s11004-012-9427-0>. Canada.
- Balling Jr., R. C. (2006). "Climate Variability and Residential Water Use in The City of Phoenix, Arizona", *School of Geographical Sciences, Arizona State University, Tempe, Arizona. Journal of Applied Meteorology and Climatology*, 46. [https://doi.org/www.mcgill.ca/bioeng/files/bioeng/a\\_spectral\\_analysis\\_based\\_methodology\\_to\\_detect\\_climatological\\_influences\\_on\\_daily\\_urban\\_water\\_demand.pdf](https://doi.org/www.mcgill.ca/bioeng/files/bioeng/a_spectral_analysis_based_methodology_to_detect_climatological_influences_on_daily_urban_water_demand.pdf)
- Bich-Ngoc, N. &. (2018). "A Review of Residential Water Consumption Determinant". <https://orbi.uliege.be/handle/2268/226248>
- Brandner, H. (2016). "Identifying the Influential Factors of the Temporal Variation of Water Consumption: A Case Study using Multiple Linear Regression Analysis". *KTH Architecture and the Built Environment, Environmental Engineering and Sustainable Infrastructure Program (EESI)*. <https://kth.diva-portal.org/smash/get/diva2:971606/Fulltext01.pdf>.
- Chang, H. P. (2014). "Sensitivity of Urban Water Consumption to Weather and Climate Variability at Multiple Temporal Scales: The Case of Portland, Oregon". *International Journal of Geospatial and Environmental Research*, vol 1(No. 1, article 7). <https://dc.uwm.edu/ijger/vol1/iss1/7>

- De Oreo, W. B. (1999). "Residential End Uses of Water Study". *Aquacraft, Inc. Water Engineering and Management Water Research Foundation, U.S.A.* [https://www.waterdm.com/sites/default/files/WRF%20\(1999\)%20Residential%20End%20Uses%20of%20Water.pdf](https://www.waterdm.com/sites/default/files/WRF%20(1999)%20Residential%20End%20Uses%20of%20Water.pdf).
- Guilherme Foquição, E. R. (2017). "Determinación de patrones de consumo doméstico y su relación con los factores climáticos, en un sector de Aguas de La Habana". Tesis de maestría, Instituto Superior Politécnico José Antonio Echeverría, La Habana. Cuba.
- Haque, M. M. (2015). "Assessing the significance of climate and community factors on urban water demand". *Gulf Organization for Research and Development, International Journal of Sustainable Built Environment, New South Wales, Australia*. <https://doi.org/10.1016/j.ijbsbe.2015.11.001>
- House-Peters, L. P. (2010). "Effects of Urban Spatial Structure, Sociodemographics, and Climate on Residential Water Consumption in Hillsboro, Oregon". *Journal of the American Water Resources Association (JAWRA)*, 46(3):461-472. <https://doi.org/10.1111/j.1752-1688.2009.00415.x>
- Lizcano Chapeta, C. J. (2022). Disposiciones legales de la gestión comunitaria del agua y los pueblos indígenas en el Ecuador. *Universidad y Sociedad*, 14(4), 514-522.
- Morales Martínez, D., & Gori Maia, A. (2021). "The Effect of Social Behavior on Residential Water Consumption". *Water2021*, 13, 1184. <https://doi.org/https://doi.org/10.3390/w13091184>
- Morote Seguido, Á. F. (2017). "Factores que inciden en el consumo de agua doméstico. Estudio a partir de un análisis bibliométrico". *Estudios Geográficos*, LXXVIII/282, pp. 257-281. Enero-junio 2017. ISSN: 0014-1496 e ISSN: 1988-8546. <https://doi.org/10.3989/estgeogr.201709>, <https://estudiosgeograficos.revistas.csic.es/index.php/estudiosgeograficos/article/view/511>
- Pomares, P. J. (2022). Principales factores climáticos que afectan la demanda doméstica. Tesis de diploma, Instituto Superior Politécnico José Antonio Echeverría, La Habana. Cuba.
- Ramos, J. M. (2016). "Gestión Integral de pérdidas de agua. Un caso de estudio". *Ingeniería Hidráulica y Ambiental*, Vol 37, No3. ISSN 1680-0338. La Habana. Cuba.
- Rodríguez Díaz Y., R. V. (2022). Efecto de las sensaciones térmicas en la enfermedad cerebrovascular en la región Occidental de Cuba. 2001-2012. *Revista Cubana De Meteorología*, 28(3). <http://rcm.insmet.cu/index.php/rcm/article/view/643>.
- Sarker, R. G.-T. (2013). "Temperature and Rainfall Thresholds corresponding to water consumption in Greater Melbourne, Australia". *20th International Congress on Modelling and Simulation, Adelaide, Australia*. [https://doi.org/https://www.researchgate.net/publication/281042716\\_Temperature\\_and\\_Rainfall\\_Thresholds\\_corresponding\\_to\\_water\\_consumption\\_in\\_Greater\\_Melbourne\\_Australi](https://doi.org/https://www.researchgate.net/publication/281042716_Temperature_and_Rainfall_Thresholds_corresponding_to_water_consumption_in_Greater_Melbourne_Australi).
- Schleich, J. &. (2007). "Determinants of Residential Water Demand in Germany". *Working Paper Sustainability and Innovation No. S 3/2007 Fraunhofer, Institute Systems and Innovation Research, Germany*. <https://doi.org/https://www.econstor.eu/bitstream/10419/28515/1/538778458.pdf>
- Soto-Montes, G. H.-P. (2016). "Implications of Climate Change on Water Resource Management in Megacities in Developing Countries: Mexico City Case Study". <https://doi.org/doi:10.5296/emsd.v5i1.8807>. Environmental Management and Sustainable Development ISSN 2164-7682. 2016, Vol. 5, No. 1. México. DOI: 10.5296/emsd.v5i1.8807
- Witten, H. I. (2016). Data mining practical machine learning tools and techniques. *Elsevier, Amsterdam, 4th ed. edition*.