



GOBIERNO DEL PRINCIPADO DE ASTURIAS

CONSEJERÍA DE ECONOMÍA Y EMPLEO



Unión Europea

Fondo Europeo  
de Desarrollo Regional



## Resultados Potenciación y Refuerzo

Línea especialización análisis inteligente de datos

En Gijón , a 31 de Diciembre de 2017

## 1. MEMORIA TÉCNICA

### 2016

Durante el periodo que cubre este documento se ha realizado la ejecución completa de las tareas correspondientes al Hito 1 y se ha comenzado en paralelo la ejecución del Hito 2.

A continuación, se describen las tareas realizadas en cada uno de los hitos de trabajo mencionados anteriormente y los resultados conseguidos.

#### Hito 1: Revisión del estado del arte

El objetivo principal de este hito fue analizar los diferentes métodos recogidos en la literatura en relación a la predicción de consumo eléctrico, desglosando la actividad en tres tareas principales:

- T1.1: Análisis de requisitos
- T1.2: Recopilación de antecedentes y literatura científica
- T1.3: Análisis del estado de la técnica

##### T1.1: Análisis de requisitos

A lo largo de la primera tarea se han investigado y detallado los requisitos mínimos que debe reunir la herramienta desarrollada para que sea de fácil aplicación y parametrizable a edificios de diferentes características, valorando la posibilidad de extracción de modelos dinámicos que no necesiten de ajustes.

Se han identificado los principales requisitos del sistema, tanto funcionales como no funcionales, desde el punto de vista del almacenamiento y procesamiento de datos y la prioridad en cuanto a su aplicación (Alta/Obligatorio, Media/Recomendable, Baja/Opcional).

A continuación, se presentan los requisitos funcionales (RF) y no funcionales (RNF) del sistema que se han identificado:

| RF-1                                                                                                                                                                                                                                                        | Horizonte temporal y resolución | Alta |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|------|
| El sistema deberá realizar la predicción del consumo eléctrico de edificios con un horizonte temporal de un día y una resolución de una hora.                                                                                                               |                                 |      |
| RF-2                                                                                                                                                                                                                                                        | Parámetros característicos      | Alta |
| El sistema deberá utilizar como variables de entrada los parámetros característicos de la metodología o metodologías seleccionadas para realizar la predicción del consumo eléctrico de edificios. En la mayoría de los casos las variables de entrada son: |                                 |      |
| <ul style="list-style-type: none"><li>• Previsiones meteorológicas expuestas a través de servicios web (temperatura y</li></ul>                                                                                                                             |                                 |      |



|              |                                                                                                                                                                                                                                                                                                                                         |             |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
|              | precipitaciones).                                                                                                                                                                                                                                                                                                                       |             |
|              | <ul style="list-style-type: none"> <li>• Condiciones meteorológicas reales.</li> <li>• Consumo eléctrico histórico.</li> <li>• Consumo eléctrico real monitorizado con dispositivos de medida.</li> <li>• Ocupación de los edificios utilizando los sistemas de control de acceso, previsiones de los cuadros de vacaciones.</li> </ul> |             |
| <b>RF-3</b>  | <b>Almacenamiento y recogida de datos</b>                                                                                                                                                                                                                                                                                               | <b>Alta</b> |
|              | El sistema deberá almacenar los datos solicitados a servicios web de predicción meteorológica, los datos recogidos de consumo eléctrico, históricos de consumos de electricidad y las predicciones realizadas.                                                                                                                          |             |
| <b>RF-4</b>  | <b>Niveles de captación de datos</b>                                                                                                                                                                                                                                                                                                    | <b>Alta</b> |
|              | El sistema deberá recoger datos de consumo eléctrico de edificios de oficinas a diferentes niveles: despachos, salas, alas, plantas y edificio completo.                                                                                                                                                                                |             |
| <b>RF-5</b>  | <b>Integración</b>                                                                                                                                                                                                                                                                                                                      | <b>Alta</b> |
|              | El sistema deberá permitir integrar datos provenientes de diferentes fuentes como los sistemas de información y ocupación del edificio, servicios de predicción meteorológica, dispositivos de monitorización de consumos eléctricos y datos históricos de consumos eléctricos.                                                         |             |
| <b>RF-6</b>  | <b>Sistema de comunicación</b>                                                                                                                                                                                                                                                                                                          | <b>Alta</b> |
|              | Los dispositivos de monitorización del consumo eléctrico deberán comunicarse con el sistema de predicción vía Ethernet.                                                                                                                                                                                                                 |             |
| <b>RF-7</b>  | <b>Margen de error</b>                                                                                                                                                                                                                                                                                                                  | <b>Alta</b> |
|              | El sistema deberá calcular el margen de error entre la predicción del consumo eléctrico de los edificios y el consumo real, con el fin de minimizarlo.                                                                                                                                                                                  |             |
| <b>RNF-1</b> | <b>Extrapolabilidad</b>                                                                                                                                                                                                                                                                                                                 | <b>Alta</b> |
|              | El sistema será extrapolable a otros edificios de oficinas utilizando las correspondientes variables de entrada de los mismos.                                                                                                                                                                                                          |             |

**T1.2: Recopilación de antecedentes y literatura científica**

La predicción de la energía consumida por los edificios es un tópico fundamental para mejorar los sistemas de oferta/demanda energética, así como para mejorar la eficiencia de los edificios y reducir el impacto medioambiental. Sin embargo, el sistema de energía en los edificios es muy complejo, ya que

los tipos de energía y edificios son muy variados. En la literatura, las principales fuentes energéticas consideradas son los sistemas de enfriamiento/calentamiento, la gestión de agua caliente y el consumo eléctrico, mientras que en cuanto a tipos de edificios se suelen diferenciar oficinas, edificios residenciales y edificios dedicados a labores de ingeniería. Por otra parte, el comportamiento energético de un edificio está influenciado por muchos factores, como las condiciones meteorológicas, especialmente la temperatura, las características constructivas del edificio y las propiedades térmicas de los materiales utilizados, los ocupantes y su comportamiento, componentes de subsistemas como la luz, acondicionamiento, ventilación, etc.

Debido a la complejidad del problema, obtener una predicción precisa es complicado, y existe una amplia literatura científica cubriendo tanto el tópico relacionado con los sistemas de predicción, como el relacionado con la gestión de la demanda.

La investigación sobre los sistemas de gestión de la demanda, si bien es un tópico directamente relacionado con el proyecto, no forma parte de los objetivos fundamentales de éste. En los trabajos [1], [2] se puede consultar un resumen actualizado del estado del arte en este tipo de sistemas si se considerase necesario realizar algún tipo de análisis a ese respecto.

En general, desde el punto de vista de las técnicas para predicción de consumos, éstas se pueden agrupar en varias categorías: modelos de ingeniería, modelos estadísticos basados en aprendizaje automático, y modelos híbridos.

### Modelos de ingeniería

Los modelos de ingeniería son utilizados para analizar el comportamiento de los edificios, especialmente desde un punto de vista térmico y de comportamiento energético global.

Este tipo de modelos utilizan, en mayor o menor medida, ecuaciones y funciones de termodinámica para calcular el consumo de los edificios atendiendo a diversas fuentes de información, como la climatología exterior, características constructivas, tasas de utilización de equipos eléctricos, y datos de equipamientos de aire acondicionado y ventilación. Habitualmente se denominan mecanismos ‘de caja blanca’, ya que necesitan conocer de forma exhaustiva todos los parámetros del sistema a modelar, frente a las técnicas estadísticas y de aprendizaje automático, denominadas ‘de caja negra’.

En general, este tipo de sistemas buscan satisfacer la siguiente ecuación (ley de conservación de la energía):

$$\Phi_{int} + \Phi_{source} = \Phi_{out} + \Phi_{stock}$$

En dicha ecuación,  $\Phi_{int}$  es el flujo de calor entrante al sistema,  $\Phi_{source}$  es el flujo de calor de fuentes de calor eventuales,  $\Phi_{out}$  es el flujo de calor que abandona el sistema y  $\Phi_{stock}$  el flujo de calor almacenado. Los principales flujos a tener en cuenta dentro de la transferencia de energía serán los de conducción a través de paredes, convección, radiación y los sistemas de ventilación. Adicionalmente, se pueden suponer otro tipo de cargas en el sistema, como pueden ser las provenientes de iluminación u otros equipamientos eléctricos.

Dentro de esta categoría, atendiendo al grado de complejidad en los cálculos, se pueden diferenciar tres

aproximaciones [3][4]:

### ***Mecanismos basados en CFD<sup>1</sup>***

Este tipo de mecanismos permiten el modelado térmico del edificio a través de la descomposición de cada zona en un alto número de volúmenes de control definidos mediante una malla. Por lo tanto, esta técnica es considerada como una aproximación tridimensional a la resolución del problema.

El modelado mediante CFD se utiliza principalmente debido a su capacidad para producir una descripción detallada de diferentes flujos (aire, contaminantes, etc.) dentro de los edificios. Además, debido a su aplicación mediante volúmenes de control discretos, es capaz de estudiar geometrías de edificios complejas, optimizando el mallado necesario.

Existen múltiples paquetes de software CFD, entre los que se pueden mencionar FLUENT<sup>2</sup>, COMSOL Multiphysics<sup>3</sup> o PHOENICS-CFD<sup>4</sup> entre otros. Su campo de aplicación es muy grande y no siempre limitado a la simulación de edificios. Sin embargo, pueden ser aplicados a cualquier tipo de sistema si se detallan suficientemente las geometrías, las condiciones de contorno y los flujos a estudiar.

La principal limitación de los métodos de CFD reside en su alto tiempo de cálculo, debido a la necesidad de una malla 3D detallada. Sin embargo, es posible limitar su aplicación a las zonas de mayor flujo de aire, permitiendo reducir significativamente la complejidad de la simulación total. Esta primera simplificación del problema se denomina aproximación zonal y se describe en el siguiente apartado.

### ***Mecanismos basados en aproximación zonal***

Este tipo de método fue introducido por [5] y una amplia revisión al respecto se puede consultar en [6]. Fundamentalmente, esta aproximación se basa en la división de cada zona del edificio en varias celdas diferenciadas. Cada celda corresponderá a una zona pequeña de cada sala o habitación. Permite realizar un análisis rápido del ambiente interior y estimar una zona de confort térmico, utilizando una aproximación 2D.

La formulación zonal puede tratar volúmenes grandes, y el acoplamiento espacial entre el sistema y su entorno. Las ecuaciones físicas son resueltas para cada celda del sistema zonal. Uno de los paquetes software que utilizan dicha técnica es SimSPARK [7].

Como principales limitaciones de este método, se encuentran la necesidad de información previa de los flujos en el edificio para realizar una división adecuada, así como un menor detalle en la precisión de los resultados de las simulaciones.

### ***Mecanismos basados en aproximación/técnica multizona o método nodal***

Finalmente, la aproximación multizona o nodal considera cada zona del edificio como un volumen

<sup>1</sup> Computational Fluid Dynamics

<sup>2</sup> [http://www.ansys.com/es\\_es/Productos/Flagship+Technology/ANSYS+Fluent](http://www.ansys.com/es_es/Productos/Flagship+Technology/ANSYS+Fluent)

<sup>3</sup> <http://www.comsol.com/>

<sup>4</sup> <http://www.cham.co.uk/>

homogéneo caracterizado por variables de estado uniformes. De este modo, cada zona se aproxima a un nodo descrito por una única temperatura, presión, concentración, etc. Generalmente cada nodo representa una habitación, una pared o el exterior del edificio, pero puede ser más específico, representando cargas térmicas (ocupación interna o ganancias de equipamiento, sistemas de calefacción o refrigeración, etc). Las ecuaciones de transferencia térmica se resuelven para cada nodo del sistema.

Dentro de esta aproximación existen múltiples paquetes software que proveen dicha funcionalidad. Entre los más prominentes se pueden destacar TrnSys<sup>5</sup>, EnergyPlus<sup>6</sup>, IDA-ICE<sup>7</sup>, ESP-r<sup>8</sup> o BSim<sup>9</sup>.

La principal ventaja de este tipo de método consiste en la reducción de tiempo de cálculo necesaria para realizar la simulación de grandes volúmenes. También se adapta especialmente bien a la estimación del consumo energético y la evolución temporal de la temperatura para habitaciones aisladas.

Como principal desventaja existe la mayor dificultad para evaluar estancias grandes de forma individual debido a las simplificaciones impuestas por el propio método de simulación.

### **Modelos estadísticos basados en aprendizaje automático**

La particularidad de los modelos estadísticos en comparación con los de ingeniería es el hecho de que no requieren información previa sobre variables físicas. Esto incluye datos relativos a parámetros térmicos o eléctricos principalmente. Los modelos estadísticos se basan en la deducción de una función a partir de los datos de comportamiento observados en un sistema.

La gran ventaja de estas técnicas es que no necesitan conocer en detalle la geometría del edificio o los fenómenos físicos que tienen lugar en el interior para deducir un buen modelo de predicción, en cuanto a precisión y exactitud. Sin embargo, la necesidad de disponer de medidas experimentales puede convertirse en un verdadero problema, ya que en algunos casos no es fácil disponer de estos datos.

Entre las técnicas estadísticas utilizadas en la predicción de las necesidades energéticas de edificios se pueden destacar tres [3][4], que pertenecen al dominio de la inteligencia artificial.

### ***Técnicas de regresión lineal múltiple o análisis de la demanda condicional***

La técnica de análisis condicional de la demanda es una técnica de regresión lineal multivariable aplicada a la predicción de las necesidades energéticas de edificios, utilizada por vez primera en 1980 [8]. Se basa en deducir la demanda de energía a partir de la combinación lineal de numerosos consumos finales considerados como variables de entrada, añadiendo un término de ruido.

---

<sup>5</sup> <http://www.trnsys.com/>

<sup>6</sup> <http://apps1.eere.energy.gov/buildings/energyplus>

<sup>7</sup> <http://www.equa.se>

<sup>8</sup> <http://www.esru.strath.ac.uk/Programs/ESP-r.htm>

<sup>9</sup> <http://sbi.dk/en/bsim>

La principal ventaja de este método es la simplicidad de uso, ya que no precisa ajustes de ningún parámetro. Puede ser utilizada sin necesidad de experiencia previa.

Presenta como principal desventaja la imposibilidad de evaluar problemas no lineales, que puede darse en la predicción de las necesidades energéticas. Además, necesitan un gran volumen de datos.

### ***Técnicas basadas en algoritmos genéticos***

Los algoritmos genéticos son técnicas de optimización estocásticas deducidas a partir de una analogía con la teoría de evolución de Darwin. Esta técnica ha sido introducida por Holland en 1975 [9], pero no fue utilizada en el campo de la simulación de edificios hasta los años 90. Se basa en las capacidades que tienen unas especies dadas para adaptarse al ambiente natural y sobrevivir ante condiciones extremas. En el campo de las simulaciones de edificios, esta técnica se utiliza para encontrar modelos de predicción sencillos del consumo de energía y para la optimización de la demanda energética.

Entre las principales ventajas de los algoritmos genéticos cabe destacar la posibilidad de resolver cualquier problema si se dispone de una función detallada convexa y la capacidad para obtener varias soluciones para un problema complejo con un gran número de parámetros de entrada. Esta ventaja permite al usuario seleccionar la solución más probable según su criterio, aunque este hecho puede ser un inconveniente, ya que el usuario nunca estará seguro de haber seleccionado la mejor solución.

Además del inconveniente anterior, estas técnicas presentan otras desventajas como el elevado tiempo de cálculo, que puede ser reducido acoplando estos algoritmos con otros métodos estadísticos, el ajuste del algoritmo, la capacidad para generar óptimos locales en lugar de globales del sistema y la gran cantidad de datos que necesitan (bien a partir de simulaciones, bien experimentales).

### ***Técnicas basadas en redes neuronales***

Las redes neuronales pertenecen a las técnicas estadísticas no lineales, utilizadas principalmente para predicción. Están inspiradas en el esquema del sistema nervioso central con sus neuronas, dendritas, axones y sinapsis. Fueron representadas por ecuaciones matemáticas por McCulloch y Pitts en 1943, pero no lo publicaron hasta 1959 [10]. En las simulaciones de edificios, las redes neuronales son técnicas muy utilizadas y se aplican normalmente en el cálculo de la predicción del consumo de energía o de la demanda de calefacción y refrigeración, sin conocer la geometría o las propiedades térmicas del edificio. También se pueden utilizar para analizar y optimizar subsistemas energéticos de los edificios, como los sistemas de acondicionamiento. Se han publicado numerosos trabajos relacionados con la aplicación de estas técnicas a edificios, incluida una revisión en el año 2000 [11].

Entre las ventajas de las redes neuronales cabe destacar que no necesitan detectar la posible colinealidad, cuentan con una capacidad de aprendizaje que les permite deducir la relación entre diferentes variables a partir de los datos sin utilizar ninguna hipótesis, resuelven el problema de discretización, son capaces de gestionar la falta de fiabilidad de los datos, disponen de una gran variabilidad para el formato de las variables que se predicen (sí/no, binario 0 o 1, valor continuo, etc.) y un tiempo de simulación eficiente.

Como principal desventaja, al igual que en los casos anteriores, estas técnicas necesitan un volumen de datos considerable para la fase de aprendizaje y deben representar toda la casuística posible. Además,

las redes neuronales tienen un gran número de parámetros indeterminados.

### ***Técnicas basadas en máquinas de soporte vectorial***

Las máquinas de soporte vectorial fueron introducidas por Cortes y Vapnik en 1995 [12]. Es una técnica utilizada normalmente para resolver problemas de clasificación y de regresión, cuyo objetivo principal es encontrar un modelo óptimo generalizado que minimice una determinada función. Entre las ventajas de estas técnicas se encuentran el número reducido de parámetros que definen el modelo optimizado. Sin embargo, tienen ciertas dificultades como la selección de la mejor función kernel y el ajuste de los parámetros característicos.

Su aplicación en el campo de los edificios es bastante reciente, alrededor del 2005 aparecen los primeros estudios. Se utilizan fundamentalmente para la predicción del consumo de energía o de la temperatura. Pueden usar diferentes tipos de datos con varias escalas temporales y de diversa naturaleza. La única restricción es que se requieren datos en formato vectorial, pero presentan como ventaja la posibilidad de utilizar conjuntos de datos heterogéneos y no es necesario que todas las variables tengan el mismo número de datos.

### **Modelos híbridos**

Los modelos de caja blanca suponen que todas las características de los edificios (térmicas y geométricas) son conocidas. Este es el caso habitual en la fase de diseño de edificios nuevos, pero es difícil recopilar tanta información en el caso de edificios ya existentes. Sin embargo, toda esta información es necesaria cuando se diseñan estrategias de monitorización. Estas técnicas suponen que los mecanismos físicos que tienen lugar en el edificio se describen con cierta precisión. Sin embargo, aunque la mayoría de los fenómenos son bien conocidos, algunos de ellos son difíciles de modelar de forma precisa como es el caso de la ventilación natural que normalmente se describe mediante ecuaciones empíricas.

Los modelos de caja negra son de aplicación limitada debido fundamentalmente a la gran cantidad de datos que se necesitan para poder aplicarlos. Además, normalmente es difícil interpretar datos obtenidos mediante métodos estadísticos en términos físicos. Por otro lado, las técnicas de minería de datos son específicas para un edificio. De esta forma, el análisis de otro edificio necesita una nueva modelización. En cambio, debido a que las ecuaciones de transferencia de calor son generales, los métodos de caja blanca son normalmente aplicables de manera general a cualquier edificio.

Las limitaciones de los modelos de ingeniería o físicos y los estadísticos, comentados en los puntos anteriores, se pueden superar utilizando los modelos híbridos que acoplan ambas técnicas. En este sentido, se dispone de diversas estrategias. Una de ellas consiste en utilizar técnicas de aprendizaje automático para estimar parámetros físicos. Una segunda opción es utilizar las técnicas estadísticas para implementar un modelo de aprendizaje que describa el comportamiento del edificio, utilizando como base un modelo físico. Una tercera estrategia consiste en utilizar métodos estadísticos en campos donde los modelos físicos no son suficientemente efectivos ni precisos. Para cada una de estas estrategias se describen en [3][4] algunos ejemplos.

La principal ventaja de estos modelos es que no precisan un gran número de datos. Además, los parámetros de entrada no tienen que estar fijados en el instante inicial de la simulación y solo se

necesitan límites de los parámetros físicos. Con una descripción aproximada de los parámetros geométricos y térmicos es suficiente y por último, permiten conservar una interpretación física (limitación en el caso de los modelos estadísticos).

Como desventajas presentan las propias de cada una de las técnicas que se acoplan, como el ajuste de los parámetros para la herramienta estadística o el tiempo de cálculo necesario para ambos modelos (físicos o estadísticos).

Este método fue introducido en los años 90 para aplicaciones específicas como los sistemas de control automático. Posteriormente, se aplicaron también para sistemas de monitorización y para el cálculo de parámetros a partir de un conjunto de datos.

Una de las aplicaciones de los métodos híbridos es la estimación de parámetros. Al contrario que las técnicas estadísticas y/o físicas, estos métodos no sólo se utilizan para predecir el comportamiento térmico de un edificio, sino también para modificar las diferentes estrategias capaces de mejorar la eficiencia energética, ya que se pueden adaptar utilizando los sistemas de monitorización. Los métodos híbridos son capaces de seleccionar las ventajas de ambos métodos (físicos y estadísticos) y aplicarlas para desarrollar modelos eficientes orientados a aplicaciones de monitorización y control. Además, permiten mantener una parte de la interpretación física sin que requieran una descripción precisa de los fenómenos que ocurren en la transferencia de calor del edificio.

Estas técnicas son relativamente recientes y tienen un gran potencial de mejora en años futuros.

## Referencias

- [1] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 461–478, Feb. 2014.
- [2] J. Aghaei and M.-I. Alizadeh, "Demand response in smart electricity grids equipped with renewable energy sources: A review," *Renew. Sustain. Energy Rev.*, vol. 18, pp. 64–72, Feb. 2013.
- [3] A. Fouquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renew. Sustain. Energy Rev.*, vol. 23, pp. 272–288, Jul. 2013.
- [4] H. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012.
- [5] H. Bouia and P. Dalicieux, "Simplified modeling of air movements inside dwelling room," in *Proceedings of Building Simulation Conference*, 1991.
- [6] A. C. Megri and F. Haghighat, "Zonal Modeling for Simulating Indoor Environment of Buildings: Review, Recent Developments, and Applications," *HVAC&R Res.*, vol. 13, no. 6, pp. 887–905, Nov. 2007.
- [7] L. Rochelle, A. M. Crépeau, and L. R. Cedex, "COUPLED HEAT AND MASS TRANSFERS IN BUILDINGS," pp. 903–910, 2003.
- [8] M. Parti and C. Parti, "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," *Bell J. Econ.*, vol. 11, no. 1, pp. 309–321, 1980.
- [9] J. H. Holland, *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control,*

and artificial intelligence. Ann Arbor, 1975.

- [10] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, "What the Frog's Eye Tells the Frog's Brain," *Proc. Inst. Radio Eng. IEEE*, vol. 47, no. 1, pp. 1940–1951, 1940.
- [11] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Appl. Energy*, vol. 67, pp. 17–35, 2000.
- [12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

### T1.3: Análisis del estado de la técnica

En las dos tablas siguientes se señalan las principales características, así como las ventajas y desventajas de los métodos físicos y estadísticos existentes para realizar la predicción del consumo eléctrico de edificios de oficinas.

Tabla 1. Resumen de las especificaciones de los métodos físicos

| Técnica física      | Especificaciones / Ventajas / Desventajas                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Método CFD</b>   | <p><b>Especificaciones:</b></p> <ul style="list-style-type: none"> <li>• Una celda = un volumen de control (3D)</li> <li>• Variables de estado locales</li> </ul> <p><b>Ventajas:</b></p> <ul style="list-style-type: none"> <li>• Descripción detallada de los flujos de fluidos en el interior de los edificios</li> <li>• Geometrías de gran tamaño</li> </ul> <p><b>Desventajas:</b></p> <ul style="list-style-type: none"> <li>• Tiempo de cálculo elevado</li> <li>• Complejidad de la implementación del modelo</li> </ul>                                                      |
| <b>Método zonal</b> | <p><b>Especificaciones:</b></p> <ul style="list-style-type: none"> <li>• Una celda = una división de una habitación (2D)</li> <li>• Variables de estado locales</li> </ul> <p><b>Ventajas:</b></p> <ul style="list-style-type: none"> <li>• Distribución espacial y temporal de variables de estado locales (temperatura, concentración, presión, flujo de aire) en un gran volumen</li> </ul> <p><b>Desventajas:</b></p> <ul style="list-style-type: none"> <li>• Tiempo de cálculo elevado</li> <li>• Descripción detallada del campo de flujo y de los perfiles de flujo</li> </ul> |
| <b>Método</b>       | <b>Especificaciones:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |



|              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>nodal</b> | <ul style="list-style-type: none"> <li>• Una celda = una habitación (1D)</li> <li>• Variables de estado uniformes</li> </ul> <p><b>Ventajas:</b></p> <ul style="list-style-type: none"> <li>• Edificios multi-zona</li> <li>• Tiempo de cálculo razonable</li> <li>• Implementación sencilla</li> </ul> <p><b>Desventajas:</b></p> <ul style="list-style-type: none"> <li>• Dificultad para estudiar sistemas de gran volumen</li> <li>• Incapacidad para estudiar efectos locales como fuentes de calor o contaminantes</li> </ul> |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Tabla 2. Resumen de las especificaciones de los métodos estadísticos

| Técnica estadística                                | Especificaciones / Ventajas / Desventajas                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Análisis condicional de la demanda</b>          | <p><b>Especificaciones:</b></p> <ul style="list-style-type: none"> <li>• Hipótesis inicial: relación lineal entre las variables y la salida</li> </ul> <p><b>Ventajas:</b></p> <ul style="list-style-type: none"> <li>• Descripción del sistema mediante una regresión</li> </ul> <p><b>Desventajas:</b></p> <ul style="list-style-type: none"> <li>• Necesidad de un gran volumen de datos para el entrenamiento</li> <li>• No-colinealidad entre los datos</li> </ul>                                                                                                                                                                     |
| <b>Algoritmo genético: técnica de optimización</b> | <p><b>Especificaciones:</b></p> <ul style="list-style-type: none"> <li>• Hipótesis iniciales: tipo de ecuación impuesta por el usuario</li> <li>• Resultado final no es necesariamente la mejor solución</li> </ul> <p><b>Ventajas:</b></p> <ul style="list-style-type: none"> <li>• Descripción del sistema mediante una función</li> <li>• Algoritmo de optimización potente</li> </ul> <p><b>Desventajas:</b></p> <ul style="list-style-type: none"> <li>• Necesidad de un gran volumen de datos para el entrenamiento</li> <li>• Dificultades para ajustar los parámetros del algoritmo</li> <li>• Tiempo de cálculo elevado</li> </ul> |
| <b>Red neuronal: técnica de</b>                    | <p><b>Especificaciones:</b></p> <ul style="list-style-type: none"> <li>• No son necesarias hipótesis iniciales</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |



|                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>regresión</b>                                          | <ul style="list-style-type: none"> <li>Modelos de “caja negra” que no permiten interpretaciones físicas</li> </ul> <b>Ventajas:</b> <ul style="list-style-type: none"> <li>Gran capacidad de entrenamiento/aprendizaje</li> </ul> <b>Desventajas:</b> <ul style="list-style-type: none"> <li>Necesidad de un gran volumen de datos exhaustivo y representativo</li> <li>No permiten interpretaciones físicas</li> </ul>                                                                                                      |
| <b>Máquina de soporte vectorial: técnica de regresión</b> | <b>Especificaciones:</b> <ul style="list-style-type: none"> <li>Hipótesis inicial: función kernel impuesta por el usuario</li> </ul> <b>Ventajas:</b> <ul style="list-style-type: none"> <li>Volumen de datos razonable para entrenamiento principalmente en formato vectorial</li> <li>Problema de minimización basado en soporte vectorial</li> </ul> <b>Desventajas:</b> <ul style="list-style-type: none"> <li>Selección de la función kernel</li> <li>Dificultad para ajustar los parámetros característicos</li> </ul> |

Los métodos híbridos, como se ha señalado anteriormente, aprovechan las ventajas de los métodos anteriores. En la siguiente tabla se resumen de forma comparativa las principales metodologías que se pueden utilizar para predecir el consumo eléctrico en edificios.

Tabla 3. Resumen de las especificaciones de los métodos estadísticos

| Método                          | Características principales                                                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Físico/“caja blanca”</b>     | <b>Geometría del edificio:</b> <ul style="list-style-type: none"> <li>Se requiere una descripción detallada de la geometría</li> </ul> <b>Datos de entrenamiento/aprendizaje:</b> <ul style="list-style-type: none"> <li>No son requeridos</li> </ul> <b>Interpretación física:</b> <ul style="list-style-type: none"> <li>Los resultados se pueden interpretar en términos físicos</li> </ul> |
| <b>Estadístico/“caja negra”</b> | <b>Geometría del edificio:</b> <ul style="list-style-type: none"> <li>No es necesaria una descripción detallada de la geometría</li> </ul> <b>Datos de entrenamiento/aprendizaje:</b>                                                                                                                                                                                                          |

|                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |  |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                            | <ul style="list-style-type: none"> <li>Se requiere un gran volumen de datos de aprendizaje recogidos durante un período de tiempo elevado</li> </ul> <p><b>Interpretación física:</b></p> <ul style="list-style-type: none"> <li>Serías dificultades para interpretar los resultados en términos físicos</li> </ul>                                                                                                                                                                                  |  |
| <b>Híbrido/"caja gris"</b> | <p><b>Geometría del edificio:</b></p> <ul style="list-style-type: none"> <li>Descripción aproximada de la geometría es suficiente</li> </ul> <p><b>Datos de entrenamiento/aprendizaje:</b></p> <ul style="list-style-type: none"> <li>Se requiere un pequeño volumen de datos de aprendizaje recogidos durante un período de tiempo corto</li> </ul> <p><b>Interpretación física:</b></p> <ul style="list-style-type: none"> <li>Los resultados se pueden interpretar en términos físicos</li> </ul> |  |

**Hito 2: Metodología de predicción**

**T2.1: Diseño de mecanismos de predicción de consumo eléctrico**

Basándose en el estudio presentado anteriormente, se decidió comenzar a realizar la predicción del consumo eléctrico en edificios de oficinas utilizando algoritmos de aprendizaje automático, ya que existen casos similares resueltos con este tipo de métodos.

A continuación, se realizó un análisis de los datos que serían necesarios como entradas de los algoritmos de aprendizaje, seleccionando las siguientes variables de entrada: *año, mes, día, hora, número de personas del edificio* (ocupación) y *predicción meteorológica*.

Además de los datos de los que se disponía, se generaron variables calculadas en base a los datos de partida, como por ejemplo si el día analizado era un día laborable o no.

Una vez seleccionadas estas variables, en estos momentos se está procediendo a realizar una selección de atributos, con el objetivo de estudiar en profundidad aquellas variables que más influencia tengan en la variación del consumo eléctrico del edificio. Estas variables seleccionadas serán las que alimenten los algoritmos de predicción que se diseñarán y desarrollarán en las siguientes tareas.

**2017**

**Hito 2: Metodología de predicción**

**T2.1: Diseño de mecanismos de predicción de consumo eléctrico**

Se decidió comenzar a realizar la predicción del consumo eléctrico en edificios de oficinas utilizando algoritmos de aprendizaje automático, ya que como se señaló en ese documento existen casos similares resueltos con este tipo de métodos.

A continuación, se realizó un análisis de los datos que serían necesarios como entradas de los algoritmos de aprendizaje, seleccionando las siguientes variables de entrada: *año, mes, día, hora, número de personas del edificio* (ocupación) y *predicción meteorológica*. Además de los datos de los que se disponía, se generaron variables calculadas en base a los datos de partida, como por ejemplo si el día analizado era un día laborable o no. Una vez seleccionadas estas variables, se llevó a cabo un análisis de las correlaciones para detectar posibles relaciones entre ellas y seleccionar las más relevantes.

A continuación, se detallan cada una de estas subtareas:

### **Análisis de los datos**

A partir del catálogo de requisitos, se determinó las variables que se requería adquirir y almacenar: *el año, el mes, el día, la hora, la temperatura, la cantidad de lluvia, horario de verano o de invierno* (meses con horario intensivo distinto al resto del año), *si es o no viernes* (horario reducido, y por tanto menor consumo), *si es o no fin de semana, y la ocupación*.

### **Análisis de las variables**

En primer lugar, se analizó si todas las variables registradas influyen en el consumo, y en qué medida influyen. Algunas de las variables no parecían realmente relevantes a la hora de predecir, con lo cual se podría evitar su uso y así encontrar modelos más óptimos.

En este caso, se contaba con dos tipos distintos de variables de entrada:

- Variables discretas: *año, mes, día, hora, verano, fin de semana, viernes*.
- Variables continuas: *temperatura, precipitaciones, ocupación*.

Para realizar la selección de los atributos más relevantes se ha empleado la herramienta Weka. Utilizando el evaluador "*Cfs Subset Eval*" y el método de búsqueda "*Best First*", se ha obtenido que las variables que más influyen son la *temperatura, fin de semana y ocupación*.

Para corroborar la selección de las variables continuas se calculó la matriz de correlaciones. La Figura 1 muestra en la parte inferior de la diagonal principal las correlaciones existentes entre este tipo de variables, y en la parte superior una representación gráfica de las mismas. Cuanto más alargada sea la forma elíptica, más correlación. La inclinación a izquierda o derecha, así como el color de la forma, indican si la correlación es inversa o directa, respectivamente. Se puede observar que las variables continuas más influyentes son la *temperatura* y la *ocupación*, tal y como se había adelantado anteriormente.

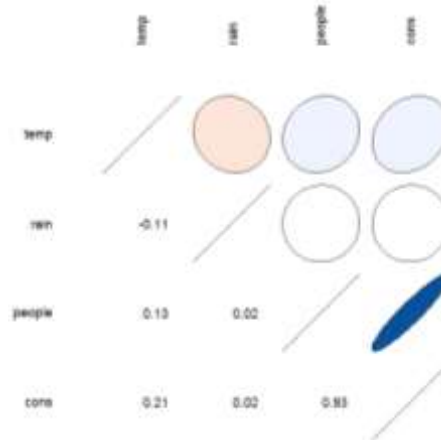


Figura 1. Matriz de correlación de las variables continuas

Por último, se analizaron en detalle las variables discretas *hora* y *viernes* que, aunque no han sido seleccionadas en el análisis anterior, sí pueden ser influyentes en el momento de entrenar los algoritmos. La Figura 2 muestra, mediante una representación gráfica con *boxplots*, la distribución del consumo para cada hora del día. Se puede comprobar que existe una variación relevante entre las distintas horas, por lo que se decidió incluirla en la selección de variables de entrada.

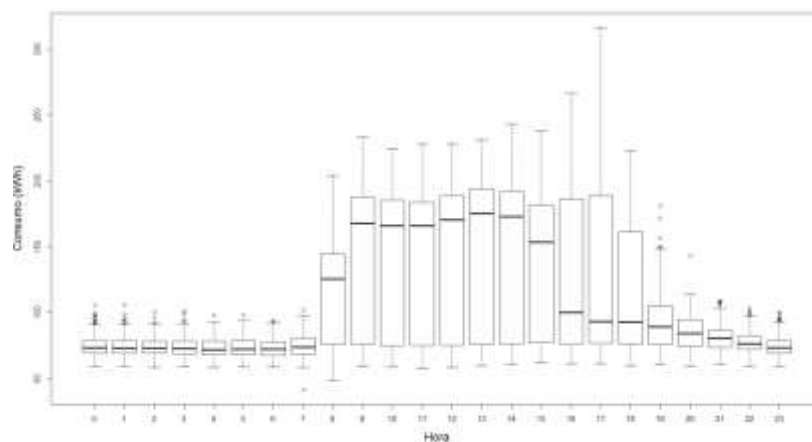


Figura 2. Análisis de la evolución del consumo medio para cada hora del día

Asimismo, y de manera similar, se pudo comprobar gráficamente la variación del consumo medio diario en función de si es día laborable o no. En la Figura 3 se puede comprobar fácilmente como el consumo disminuye sustancialmente durante el fin de semana (días 6 y 7).

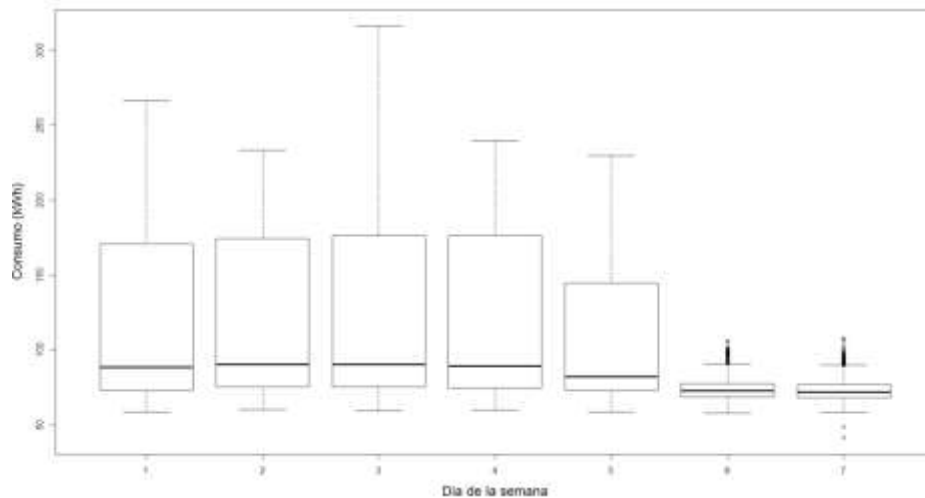


Figura 3. Análisis de la evolución del consumo medio para cada día de la semana

Por tanto, los modelos utilizados para la predicción del consumo han tenido en cuenta la *hora*, la *temperatura*, *si es o no fin de semana*, *si es o no viernes* y la *ocupación del edificio*.

### **T2.2: Recopilación de datos**

Durante esta tarea se recopilaban los datos necesarios para poder validar la metodología seleccionada y elegir los algoritmos de aprendizaje automático adecuados para resolver este problema.

Los datos fueron recopilados de tres fuentes diferentes y fueron etiquetados con el instante temporal al que pertenecen:

- Consumo eléctrico, usando los sensores disponibles y datos históricos. Para llevar a cabo la recopilación de datos, se trabajó en la creación de scripts que permitan su obtención automática. En el caso concreto del consumo eléctrico se trabajó con armarios siguiendo la arquitectura SISREM. Se muestra un ejemplo de implementación en la Figura 4. Este tipo de armarios fueron instalados para la captura del consumo eléctrico en tiempo real y la validación de los métodos utilizados para la predicción.

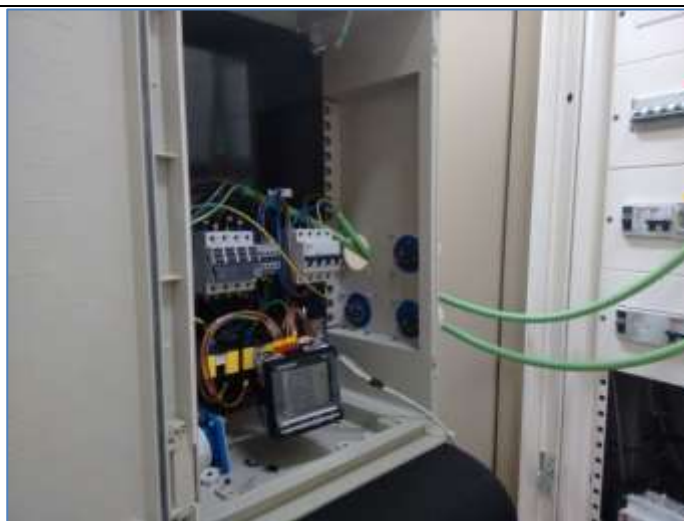


Figura 4. Ejemplo de armario utilizado para la captura de datos de consumo eléctrico.

- Ocupación del edificio, utilizando datos disponibles en las herramientas de gestión internas.
- Datos meteorológicos, predicciones tomadas de servicios meteorológicos y datos de observaciones. En particular, se han utilizado los datos de predicciones proporcionados por la página [www.wunderground.com](http://www.wunderground.com). Esta página ofrece una API gratuita para desarrolladores ([www.wunderground.com/weather/api/](http://www.wunderground.com/weather/api/)) que permite descargar predicciones del tiempo en cualquier parte del mundo, con una limitación de 500 peticiones al día y 10 peticiones por minuto. El formato de salida puede ser un archivo JSON o XML.

Todas las variables recopiladas se almacenaron en las bases de datos utilizadas por el sistema, de forma que se pueden recuperar posteriormente para ser utilizadas en los métodos de predicción propuestos.

Los datos recopilados fueron procesados para poder ser usados por los algoritmos desarrollados, formando tuplas de entrada con el siguiente formato:

|     |     |     |      |      |        |        |         |               |           |         |
|-----|-----|-----|------|------|--------|--------|---------|---------------|-----------|---------|
| año | mes | día | hora | temp | lluvia | verano | viernes | fin de semana | ocupación | consumo |
|-----|-----|-----|------|------|--------|--------|---------|---------------|-----------|---------|

### **T2.3: Desarrollo de algoritmos**

Durante esta tarea se realizaron pruebas con diferentes algoritmos de aprendizaje automático que permitieran validar la metodología y seleccionar el que mejor se adapte al problema de la predicción de consumo eléctrico en edificios de oficinas. Las técnicas que se consideraron fueron las siguientes:

- Regresión lineal. El modelo de regresión lineal es un modelo matemático que intenta predecir si la relación entre las variables consideradas es de tipo lineal. En este caso se analiza si existe una relación lineal entre el consumo y las variables que influyen en él.

- Árbol de regresión. Los árboles de regresión son modelos de predicción de aprendizaje automático que, teniendo una variable respuesta, permiten crear una serie de reglas basadas en variables predictoras.
- Redes de neuronas. Las redes neuronales son modelos de predicción de aprendizaje automático que están inspirados en el modo de funcionamiento del sistema nervioso de los animales.

Para realizar las pruebas con los diferentes métodos de predicción, se usó la herramienta de análisis estadístico R. El interés de usar esta herramienta reside en que proporciona una amplia gama de librerías, muy útiles en la fase de desarrollo, y que se pueden integrar de manera sencilla a cualquier desarrollo posterior del proyecto.

Todos los algoritmos comparten el mismo formato de entrada. La plataforma R permite importar archivos en formato .csv, uno de los formatos estándares en el mundo del análisis de datos.

Para cada modelo de predicción utilizado calculó el *Mean Absolute Error* (MAE) y el *Mean Absolute Percentage Error* (MAPE), con el objetivo de hacer una comparación y buscar el mejor de ellos. El MAE proporciona una media del promedio de los errores en valor absoluto, es decir, el promedio de la diferencia entre el consumo real y el estimado. Por otro lado, el MAPE proporciona una media del promedio de los errores relativos en valor absoluto, es decir, el promedio de la diferencia relativa porcentual entre el consumo real y el estimado.

Para evaluar los algoritmos desarrollados en la fase de entrenamiento se utilizaron técnicas de validación cruzada (*cross-validation*). Este método consiste en dividir el conjunto de datos en  $k$  subconjuntos, y utilizar  $(k-1)$  para construir el modelo (*training set*) y el restante para validarlo (*test set*), repitiendo este proceso  $k$  veces. Finalmente se realiza la media aritmética con los resultados de cada iteración para obtener un único resultado.

#### **T2.4: Pruebas iniciales**

Basándose en los resultados de las tareas anteriores (T2.1, T2.2 y T2.3), se realizaron las pruebas de los algoritmos desarrollados en la tarea T2.3 usando como conjunto de entrenamiento los datos correspondientes al edificio de CTIC. En la

Tabla 2 se detallan los resultados obtenidos los tres algoritmos explicados anteriormente:

Tabla 2. Errores cometidos en la fase de test.

| Algoritmo        | Resultados |        |
|------------------|------------|--------|
| Regresión lineal | MAE        | 11.92  |
|                  | MAPE       | 11.41% |
| Red de neuronas  | MAE        | 7.64   |

|                      |      |        |
|----------------------|------|--------|
|                      | MAPE | 7.41%  |
| <b>M5P</b>           | MAE  | 7.68   |
|                      | MAPE | 7.44%  |
| <b>Random Forest</b> | MAE  | 13.18  |
|                      | MAPE | 12.85% |

### Regresión lineal

En la tabla anterior se puede observar que este modelo fue de los que peor se ajustó al problema planteado. Se obtuvo un error medio absoluto de 11.92. El modelo de regresión lineal obtenido es el siguiente:

$$\text{consumo} = (8.17 \cdot \text{hora}) + (25.55 \cdot \text{temperatura}) - (3.49 \cdot \text{viernes}) - (7.81 \cdot \text{weekend}) + (106.64 \cdot \text{ocupacion}) + 64.25$$

### Redes de neuronas

Se utilizó como parámetro de entrada el número de iteraciones (además de los estándares), considerando un máximo de 5000. Se fijaron los otros parámetros a valores descritos en el estado del arte. Se obtuvo un error medio absoluto de 7.64, siendo el más bajo de todos los obtenidos en esta fase. La red de neuronas obtenida se representa gráficamente en la Figura 5.

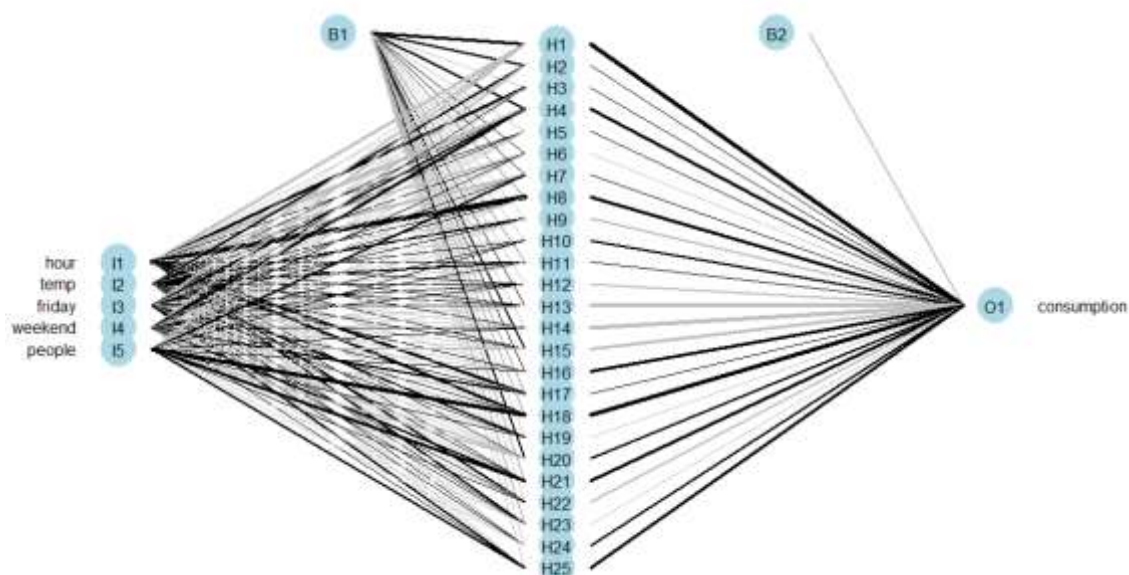


Figura 5. Red de neuronas obtenida.

### Árbol de regresión

Para analizar este problema se crearon dos modelos de árboles de regresión: el algoritmo M5P y el algoritmo Random Forest:

- El algoritmo M5P es una reconstrucción del algoritmo M5 desarrollado por Quinlan, que consiste en inducir árboles a partir de modelos de regresión. El M5P combina un árbol de decisión convencional con la posibilidad de añadir funciones de regresión lineal en cada uno de sus nodos. Con este algoritmo se obtuvo un error medio absoluto de 7.68. La Figura 6 muestra el árbol generado a partir del algoritmo M5P.

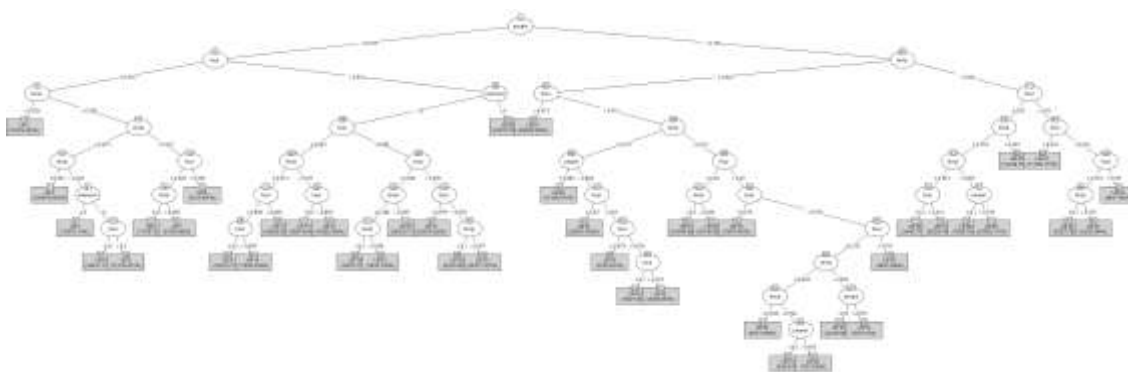


Figura 6. Árbol generado con el algoritmo M5P.

- El Random Forest es un algoritmo predictivo que usa la técnica de *bagging* para combinar diferentes árboles, donde cada árbol es construido con observaciones y variables aleatorias. Con este algoritmo multi-árbol se obtuvo el peor error medio absoluto, igual a 13.18.

Según los resultados obtenidos, el algoritmo que mejor se comportó en este problema con los datos de entrenamiento es la Red de Neuronas, seguido muy de cerca por el M5P.

### **Hito 3 Validación de la metodología de predicción**

#### **T3.1: Validación de la metodología**

A continuación, se detallan las principales características de los datos utilizados en la fase de test, así como los resultados obtenidos en la validación de los modelos.

#### **Recopilación de datos**

Para la fase de test se recopilaban datos a lo largo de un año, todos los días del año, y con una frecuencia horaria. La temperatura exterior ha oscilado entre 2 y 18°C, con un valor medio de 14,4°C. La lluvia ha alcanzado un máximo de 5 mm/m<sup>2</sup>, y el número de personas del edificio llegó a ser de 50, siendo el número medio de 40. El consumo eléctrico ha oscilado entre 34,45 kWh y 243 kWh, siendo el consumo medio de 104,38 kWh.

Al igual que ocurría con los datos de entrenamiento, los datos de test fueron recopilados desde tres

fuentes diferentes y fueron etiquetados con el instante temporal al que pertenecen:

- Consumo eléctrico, usando los sensores disponibles y datos históricos.
- Ocupación del edificio, utilizando datos disponibles en las herramientas de gestión internas.
- Datos meteorológicos, predicciones tomadas de servicios meteorológicos y datos de observaciones.

Todas las variables recopiladas se almacenaron en las bases de datos utilizadas por el sistema, de forma que pueden ser recuperadas posteriormente para ser utilizadas en los métodos de predicción propuestos.

Los datos recopilados fueron procesados para poder ser usados por los algoritmos desarrollados, formando tuplas de entrada con el siguiente formato:

|     |     |     |      |      |        |        |         |               |           |         |
|-----|-----|-----|------|------|--------|--------|---------|---------------|-----------|---------|
| año | mes | día | hora | temp | lluvia | verano | viernes | fin de semana | ocupación | consumo |
|-----|-----|-----|------|------|--------|--------|---------|---------------|-----------|---------|

Durante la fase de entrenamiento se concluyó que las variables de entrada relevantes en este problema eran la *hora*, la *temperatura*, *si es o no fin de semana*, *si es o no viernes* y la *ocupación del edificio*.

### Resultados de test

Para realizar la validación de los algoritmos de predicción obtenidos en la fase de entrenamiento, también se empleó la herramienta de análisis estadístico R.

Para cada modelo de predicción utilizado, se calculó el *Mean Absolute Error* (MAE) y el *Mean Absolute Percentage Error* (MAPE), con el objetivo de hacer una comparación y buscar el mejor de ellos. El MAE proporciona una media del promedio de los errores en valor absoluto, es decir, el promedio de la diferencia entre el consumo real y el estimado. Por otro lado, el MAPE proporciona una media del promedio de los errores relativos en valor absoluto, es decir, el promedio de la diferencia relativa porcentual entre el consumo real y el estimado.

En la Tabla 3 se detallan los resultados de test obtenidos con los algoritmos obtenidos en la fase de validación.

Tabla 3. Errores cometidos en la fase de test.

| Algoritmo        | Resultados |        |
|------------------|------------|--------|
| Regresión lineal | MAE        | 13.91  |
|                  | MAPE       | 13.35% |
| Red de neuronas  | MAE        | 9.09   |

|                      |      |              |
|----------------------|------|--------------|
|                      | MAPE | 8.76%        |
| <b>M5P</b>           | MAE  | <b>8.99</b>  |
|                      | MAPE | <b>8.65%</b> |
| <b>Random Forest</b> | MAE  | 15.05        |
|                      | MAPE | 14.69%       |

### Regresión lineal

En la tabla anterior se puede observar que este modelo fue de los que peor se ajusta al problema planteado. Se obtuvo un error medio absoluto de 13.91, con lo cual se descartó un modelo de regresión lineal.

### Redes de neuronas

Se obtuvo un error medio absoluto de 9.09.

### Árbol de regresión

A continuación, se detallan los resultados obtenidos con el algoritmo M5P y el algoritmo Random Forest:

- Con el algoritmo M5P se obtuvieron mejores resultados a los obtenidos con los algoritmos anteriores. Se obtuvo un error medio absoluto de 8.99, concluyendo que es el mejor modelo para predecir el consumo eléctrico.
- Con el algoritmo Random Forest se obtuvo el peor error medio absoluto, igual a 15.05. A la vista de los resultados obtenidos se ha comprobado que este algoritmo proporciona mejores resultados cuantas más variables de entrada reciba, es decir, la selección de atributos inicial provoca un empeoramiento de los resultados proporcionados. A pesar de ello, los resultados que se obtienen con todas las variables de entrada fueron similares a los obtenidos en el apartado anterior por el algoritmo M5P, siendo este último un algoritmo más interpretable y con menos probabilidades de sobreajuste.

### T3.2: Análisis de los resultados y conclusiones del proyecto

A continuación, se detallan los resultados obtenidos a lo largo de la presente tarea. Para ello, se analizaron los errores cometidos en la fase de test con el algoritmo que mejores resultados ha proporcionado, es decir, con el M5P (ver Figura 6). Se analiza la distribución de los errores con dos niveles de granularidad: horaria y diaria.

La Figura 7 muestra en color rojo la curva media de consumo para las 24 horas del día. Esta curva se ha obtenido agregando las horas de cada día para los datos de test. En verde se representa la curva media de consumo predicha por el algoritmo M5P entrenado.

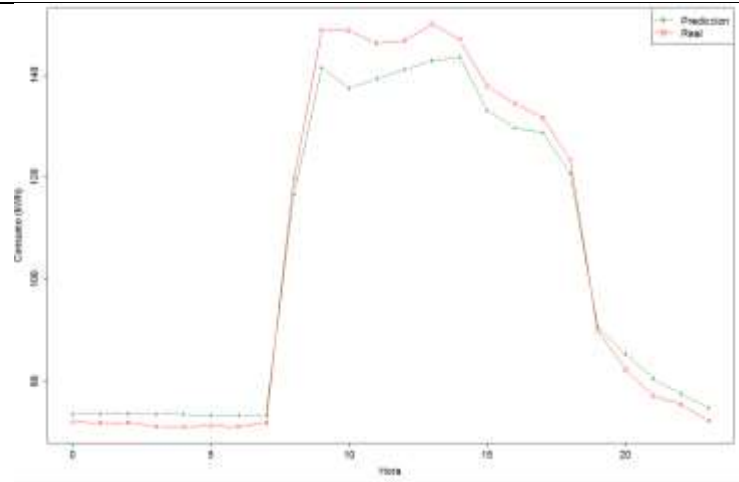


Figura 7. Comparación entre la curva media diaria predicha y la real.

Si cuantificamos el error medio horario cometido obtenemos los siguientes valores:

- Error mínimo: 0.0011
- Error máximo: 110.4740
- Media: 8.9946
- Mediana: 6.0853
- Desviación típica: 9.4672

La Figura 8 muestra mediante un boxplot la distribución de cuartiles del error cometido a nivel horario.

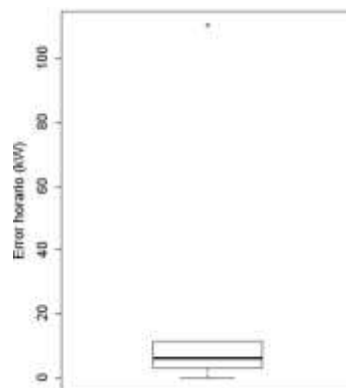


Figura 8. Representación de los cuartiles de error horario.

La Figura 9 desglosa este error medio para cada una de las horas del día.

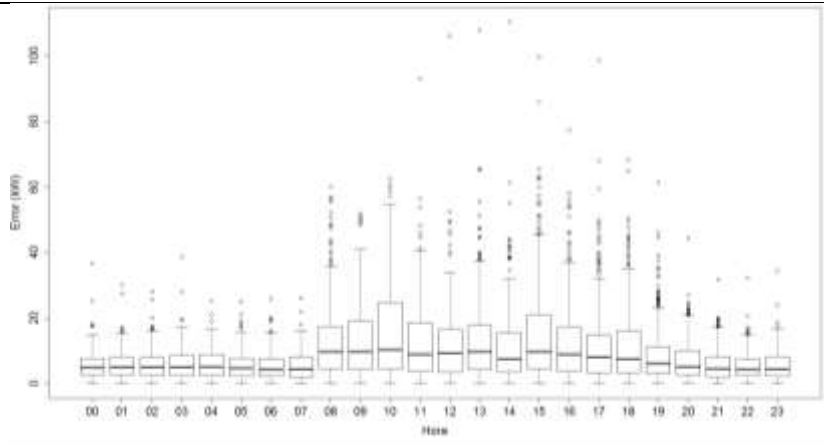


Figura 9. Evolución del error cometido para cada hora del día.

Por otra parte, la Figura 10 muestra en color rojo la curva media de consumo semanal. Esta curva se ha obtenido agregando el consumo de cada día de la semana para el conjunto de datos de test. Al igual que en el caso anterior, en verde se representa la curva media de consumo predicha por el algoritmo M5P entrenado.

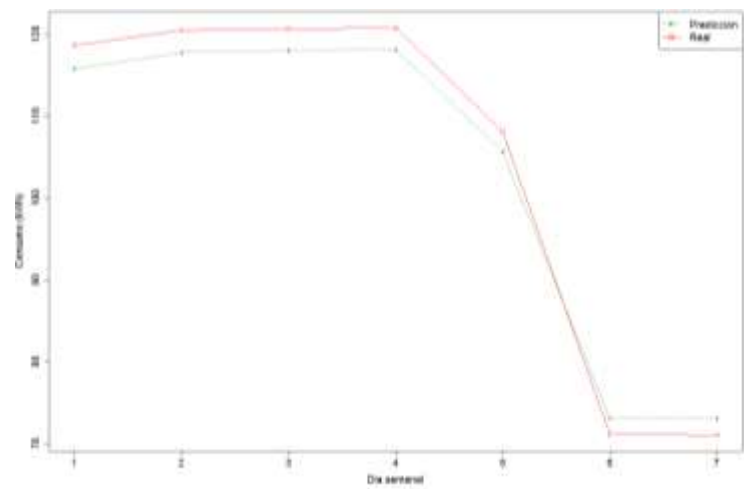


Figura 10. Comparación entre la curva media semanal predicha y la real.

Si cuantificamos el error medio diario cometido obtenemos los siguientes valores:

- Error mínimo: 0.0079
- Error máximo: 25.83
- Media: 5.2583
- Mediana: 4.2994
- Desviación típica: 4.1052

La Figura 11 muestra mediante la distribución de cuartiles del error cometido a nivel diario.

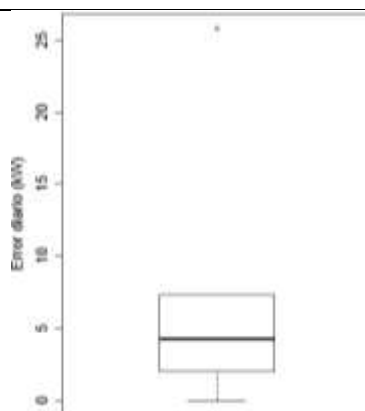


Figura 11. Representación de los cuartiles de error diario.

La Figura 12 desglosa este error medio para cada una de los días de la semana.

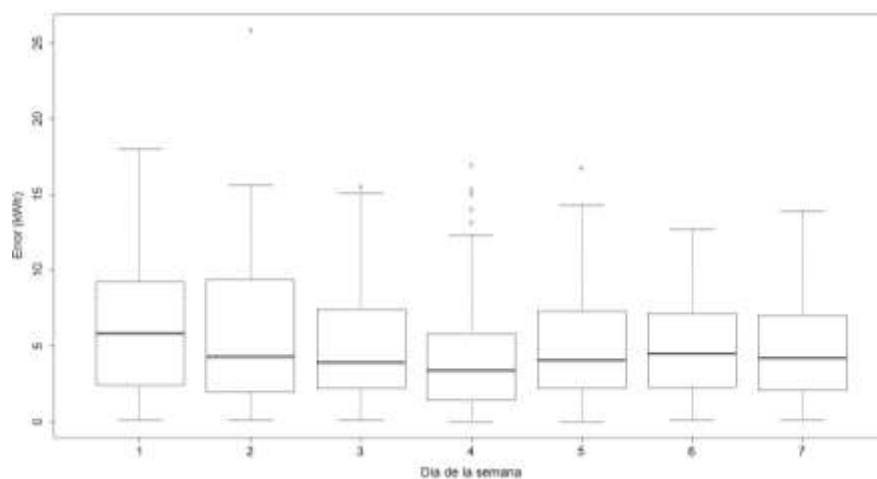


Figura 12. Evolución del error cometido para cada día de la semana.

A la vista de los resultados obtenidos, la principal conclusión del proyecto es que es viable la predicción del consumo eléctrico en edificios mediante técnicas de aprendizaje automático. En particular, y según los resultados obtenidos, el algoritmo que mejor se comporta en este problema es el M5P, seguido muy de cerca por la Red de Neuronas.



## 2. RESULTADOS CONSEGUIDOS

Durante el periodo de tiempo que cubre este informe se han conseguido todos los resultados planificados.

Desde el punto de vista del alcance técnico del proyecto, se han conseguido los siguientes resultados:

- Recolección de los requisitos mínimos que debe reunir la herramienta desarrollada para que pueda ser implantada en aquellas aplicaciones de análisis de la información desestructurada, opiniones y sentimientos.
- Recopilación de los antecedentes presentes en la literatura científica para la predicción del consumo eléctrico de edificios de oficinas.
- Estudio del arte comparativo entre todas las soluciones encontradas, destacando las ventajas e inconvenientes de cada una de ellas.
- Análisis de los principales factores de influencia en el consumo eléctrico de edificios de oficinas.
- Análisis de la metodología a seguir para realizar una predicción en el consumo eléctrico.
- Recopilación, integración y análisis de los datos recopilados desde las distintas fuentes de información consultadas.
- Se han recopilado los antecedentes presentes en la literatura científica para la predicción del consumo eléctrico de edificios de oficinas.
- Se han analizado los principales factores de influencia en el consumo eléctrico de edificios de oficinas.
- Se ha diseñado, desarrollado y validado una metodología de predicción del consumo eléctrico basada en aprendizaje automático. Para ello, se han recopilado, integrado y analizado los datos necesarios procedentes desde distintas fuentes de información.