

UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO EN INGENIERÍA DE
TECNOLOGÍAS Y SERVICIOS DE
TELECOMUNICACIÓN**

TRABAJO FIN DE GRADO

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN
DE REDES NEURONALES PARA LA
PREDICCIÓN DE TEMPERATURA EN
CENTROS DE PROCESADO DE DATOS**

**LUIS CASTEJÓN LOZANO
2021**

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Título: Análisis, diseño e implementación de redes neuronales para la predicción de temperatura en centros de procesamiento de datos.

Autor: D. Luis Castejón Lozano

Tutor: D. José Manuel Moya Fernández

Ponente: D.

Departamento: Departamento de Ingeniería Electrónica

MIEMBROS DEL TRIBUNAL

Presidente: D.

Vocal: D.

Secretario: D.

Suplente: D.

Los miembros del tribunal arriba nombrados acuerdan otorgar la calificación de:

Madrid, a de de 20...

UNIVERSIDAD POLITÉCNICA DE MADRID

**ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN**



**GRADO EN INGENIERÍA DE TECNOLOGÍAS Y
SERVICIOS DE TELECOMUNICACIÓN**

TRABAJO FIN DE GRADO

**ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE
REDES NEURONALES PARA LA
PREDICCIÓN DE TEMPERATURA EN
CENTROS DE PROCESADO DE DATOS**

LUIS CASTEJÓN LOZANO

2021

Agradecimientos

Agradecer a las personas que me han ayudado a lo largo del desarrollo de este Trabajo Fin de Grado. En primer lugar a mi tutor, José Manuel, por proporcionarme este trabajo y por la guía. A Claudia por su apoyo incondicional y por los ánimos en todo momento. A mis padres, Luis y Beatriz, por estar siempre a mi lado y suponer una ayuda fundamental a lo largo de todo el Grado, en los buenos y malos momentos. Por último, a mis hermanos, al resto de mi familia y amigos.

RESUMEN

Hoy en día en los centros de datos se desperdician grandes cantidades de energía y, por tanto, de dinero, en la refrigeración de los mismos, debido a la ausencia del correcto análisis de datos. En torno a un 2 % de las emisiones globales de dióxido de carbono corresponden al sector TIC (al que pertenecen los centros de datos), debido al consumo energético.

Se pretende hacer una transformación digital basada en el análisis de datos y la predicción automática de variables, con el objetivo de alcanzar una mayor eficiencia energética en los centros de datos, dando lugar al ahorro de gastos en refrigeración, ayudando esto tanto al medio ambiente, como a la economía de los propietarios de los centros. Para ello, será necesario conocer la situación actual del sector, así como las tecnologías del estado del arte.

Se emplean redes neuronales, construidas con Python a través de las herramientas proporcionadas por TensorFlow y Keras, utilizando librerías como Pandas, para predecir la temperatura en los distintos puntos del centro de datos, con el objetivo de reducir el coste de refrigeración. Se obtiene un sistema capaz de predecir de forma fiable la temperatura que va a haber en un determinado instante de tiempo, permitiendo así regular la temperatura de climatización. De esta forma, los servidores se encontrarán a la temperatura óptima para desempeñar su función. Estas redes son entrenadas con mediciones obtenidas de los sensores de un centro de datos. Se llega a resultados de calidad, obteniendo un error cuadrático medio en la predicción en torno a 0,002.

Además, se realiza un estudio relacionado con los sensores establecidos. Este estudio tiene como misión comprobar tanto el número y la colocación óptima de estos sensores. En esta fase, llevando a cabo una reducción del número de mediciones por sensor y del número de sensores, se consigue lograr resultados similares a los obtenidos con la totalidad de los datos.

PALABRAS CLAVE

Aprendizaje automático, redes neuronales artificiales, sensor, base de datos, temperatura, redes neuronales recurrentes

SUMMARY

Nowadays, massive quantities of energy and, therefore money, are being wasted in data center refrigeration, as an effect of the lack of an effective data analysis system. In addition, a fact that must be considered, is that about a 2 % of the carbon dioxide emissions are caused by TIC services (in whose data centers are included).

The purpose of this project is to carry out a digital transformation based in data analysis and automatic prediction of variables, trying to reach a higher energy efficiency in data centers, aiming to save expenses in refrigeration, being this beneficial for the environment and the data centers owners. It is necessary to acquire knowledge about the current situation of the sector, and the main new technologies.

Artificial Neural Networks, built using Python, leaning on TensorFlow and Keras and libraries such as Pandas, are the main tools used to predict temperature reliably in every point of the data center, to adapt the refrigeration system along time, achieving that wage reduction. In this way, servers will be kept in the optimum temperature in order to assure a good efficiency, as they need to work in an adequate range of temperature. The network is trained with measures taken by different sensors disposed in a center. Quality results are reached, obtaining a mean squared error about 0,002.

Another mission is to study the most suitable number and colocation of those sensors. In this phase, similar results are obtained carrying out a reduction of the number of sensors and the measurements taken by each one.

KEYWORDS

Machine learning, artificial neural networks, sensor, database, temperature, recurrent neural networks

Abreviaturas

- **ANN** - Red Neuronal Artificial
- **CNN** - Red Neuronal Convolucional
- **CRAC** - Aire Acondicionado para la Sala de Cómputo
- **GRU** - Gated Recurrent Unit
- **IT** - Tecnologías de la Información
- **MSE** - Error Cuadrático Medio
- **PUE** - Power Usage Effectiveness
- **RNN** - Red Neuronal Recurrente

Índice

1. INTRODUCCIÓN Y OBJETIVOS	1
1.1. NECESIDAD	1
1.2. OBJETIVOS	3
1.3. CENTROS DE PROCESADO DE DATOS	4
1.3.1. REFRIGERACIÓN	5
1.4. APRENDIZAJE AUTOMÁTICO	7
1.5. REDES NEURONALES ARTIFICIALES	8
1.5.1. REDES NEURONALES RECURRENTE (RNN)	13
1.5.2. CÉLULAS GRU	14
1.5.3. REDES NEURONALES CONVOLUCIONALES (CNN)	15
2. ESTADO DEL ARTE	16
2.1. GREEN COMPUTING	16
2.2. MACHINE LEARNING PARA REDUCIR ENERGÍA DE REFRIGERACIÓN . . .	17
2.3. FREE COOLING TECHNIQUES	18
2.4. OTROS	20
2.5. CONCLUSIONES	21
3. DESARROLLO	22
3.1. DATOS DE ENTRADA	22
3.2. RED NEURONAL	25
3.2.1. DISEÑO	25
3.2.2. ENTRENAMIENTO	28
4. RESULTADOS	30
4.1. EXPERIMENTO GLOBAL	30
4.2. EXPERIMENTO REDUCIENDO MEDICIONES	33
4.3. EXPERIMENTO REDUCIENDO SENSORES	38
5. CONCLUSIONES Y LÍNEAS FUTURAS	41
5.1. CONCLUSIONES	41
5.2. LÍNEAS FUTURAS	42
6. BIBLIOGRAFÍA	43

7. ANEXO A: ASPECTOS ÉTICOS, ECONÓMICOS, SOCIALES Y AMBIEN- TALES	46
7.1. INTRODUCCIÓN	46
7.2. DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PRO- YECTO	46
7.3. ANÁLISIS DETALLADO DE ALGUNO DE LOS PRINCIPALES IMPACTOS: MEDIOAMBIENTAL Y ECONÓMICO	47
7.4. CONCLUSIONES	48
8. ANEXO B: PRESUPUESTO ECONÓMICO	49
9. ANEXO C: CÓDIGO SOFTWARE	50

Índice de figuras

1.	Evolución de la demanda de energía del sector TIC. [17]	2
2.	Evolución de la demanda global de energía.[17]	2
3.	Racks en un centro de datos. [4]	5
4.	Organización pasillo frío y caliente [4]	6
5.	Planta de una sala de un CPD [23]	6
6.	El perceptrón	9
7.	Función sigmoide	10
8.	Red Neuronal Artificial. [2]	11
9.	Descenso de gradiente.	12
10.	Una capa de una RNN a lo largo del tiempo. [15]	13
11.	Células de memoria en una red neuronal [15]	14
12.	Célula GRU. [15]	14
13.	Funcionamiento de las CNN	15
14.	Evolución del PUE de Google en función del uso de aprendizaje automático. [26]	18
15.	Refrigeración por Aire Directo. [28]	19
16.	Centro de datos de Microsoft sumergido en el agua. [22]	20
17.	Diagrama de bloques del proceso de desarrollo.	22
18.	Centro de datos objeto de estudio.	23
19.	Regresión lineal.[3]	26
20.	Arquitectura de la red neuronal.	27
21.	Evolución del MSE obtenido en el entrenamiento en función de los distintos repartos para entrenamiento y test.	31
22.	Predicciones realizadas por el modelo en experimento Global.	32
23.	Predicciones de Inlet reduciendo la cantidad de mediciones.	36
24.	Predicciones de Outlet reduciendo la cantidad de mediciones.	37
25.	Distribución de sensores para el estudio.	38
26.	Predicciones temperatura Inlet con distribución uniforme de sensores.	39
27.	Predicciones temperatura Outlet con distribución uniforme de sensores.	39

1. INTRODUCCIÓN Y OBJETIVOS

1.1. NECESIDAD

Los centros de procesamiento de datos son estrictamente fundamentales para el mundo actual, debido a la necesidad creciente que se tiene de almacenar y procesar datos para fines diversos. Millones de datos se producen cada día con todo tipo de finalidades, en una sociedad cada vez más digital, con una creciente utilización de la computación en la nube e Internet, en general.

El proceso de digitalización ha sido acelerado por la pandemia causada por el COVID-19. Esta nueva situación da lugar a un cambio radical en la forma de vida de la sociedad, generando efectos como el hecho de que cada día las empresas y trabajadores recurran de manera habitual al teletrabajo, y la sociedad en general a sistemas conectados a Internet para el ocio y relaciones interpersonales, lo cual da lugar a un crecimiento de la cantidad de datos.

Este proceso produce un gasto energético perjudicial para el medio ambiente, ya que es muy cuantioso en los centros de datos, sobre todo en los que poseen grandes dimensiones. Según estudios de la Comisión Europea [24], el sector TIC (incluyendo los CPD) genera un 2 % de las emisiones globales de CO_2 . La huella de carbono generada por los centros de datos es la que presenta un crecimiento más rápido en todo el sector, ya que consumen cantidades ingentes de energía, debido a la gran actividad que realizan de forma continuada, siendo esta necesaria cada día del año. Además, el PUE¹ de una gran parte de los centros empeora año a año.

La revista científica *Nature* [17], muestra las previsiones de la demanda de energía del sector TIC en la presente década, en la gráfica reflejada en la figura 1. Como se puede observar, la demanda de los centros de datos ha aumentado notablemente desde el 2010, pudiendo ser, en 2030, la más importante, junto a la correspondiente a las redes.

¹Es la unidad utilizada para medir la eficiencia energética de un CPD. Se explica detalladamente en la sección 1.3.

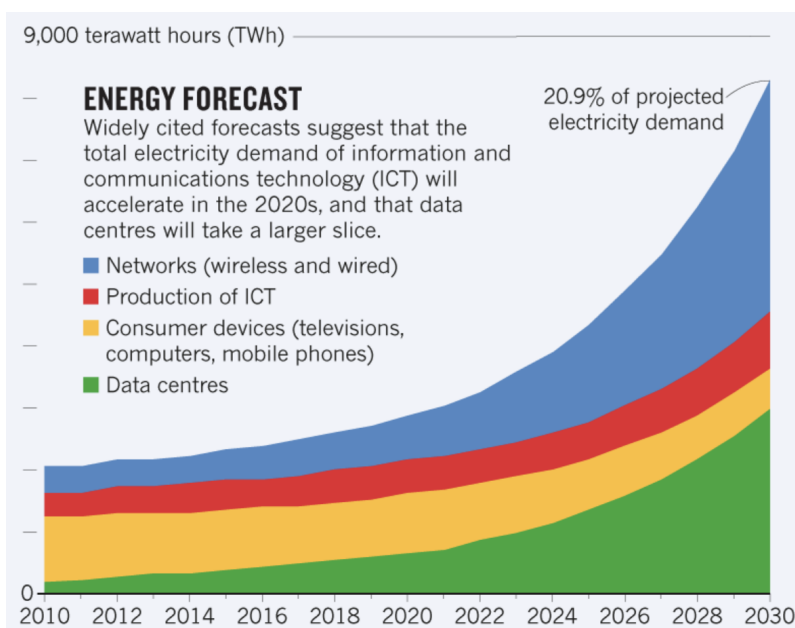


Figura 1: Evolución de la demanda de energía del sector TIC. [17]

En la figura 2, se refleja la evolución de la demanda global de energía. A lo largo del tiempo, el porcentaje correspondiente a las TIC ha ido aumentando, siendo preocupante la cantidad de energía que se espera en 2030. Solo en el mejor de los casos, poniendo en funcionamiento las mejores prácticas de optimización energética, para ese año se puede conseguir un mejor resultado (en el que se aprecia la gran parte que supondrán los CPD).

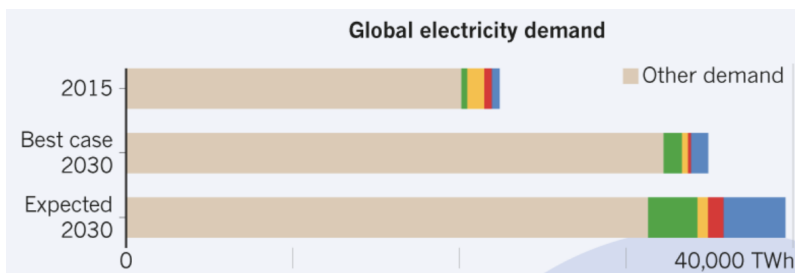


Figura 2: Evolución de la demanda global de energía.[17]

Esta situación, unida a otras causas de la aceleración del cambio climático, han dado lugar a una necesidad de reducir la emisiones y la contaminación, y debido a la preocupación existente, surge el Pacto Verde Europeo, lanzado por la Unión Europea. Con ello, la UE se compromete a reducir sus emisiones como mínimo en un 55 % antes de 2030, tratando de ser el primer continente neutro. Uno de los efectos provocados por el acuerdo es que las compañías del sector tengan la obligación de reducir la demanda de energía de la forma más rápida posible, y de comenzar a

poner en funcionamiento diversas prácticas que les permita mantener un compromiso con el medio ambiente.

La energía consumida en los CPD es, en gran parte, destinada a la alimentación de los sistemas integrados, en especial los servidores, sin embargo, existe otro factor diferencial relacionado con este tema: la refrigeración.

La refrigeración es un proceso clave en el mantenimiento de los centros de datos. La continua actividad del sistema produce un calor que perjudica al centro, reduciendo el rendimiento de los servidores y aumentando el riesgo de generar daños permanentes en los procesadores, existiendo la posibilidad de recalentamiento de los servidores al superar cierta temperatura. La alimentación de los sistemas de refrigeración supone en torno al 40 % de la energía consumida por el centro de datos, bajando este porcentaje hasta el 24 % en los sistemas más eficientes e incrementándose hasta un 61 % en los menos eficientes. [24]

La temperatura de los servidores no es fija, varía en función de factores como la carga de trabajo que tengan en un instante determinado, la temperatura exterior, la humedad, etc. Por ello, otra necesidad a tener en cuenta es la relacionada con el plano económico (aparte del medioambiental, ya tratado) de las empresas, ya que en ocasiones, derrochan energía, y por tanto, dinero, debido a que, las salas se enfrían de manera insuficiente por momentos, o incluso de más, ignorando la efectividad y rendimiento de los servidores en función de la temperatura. Es necesario un sistema que permita detectar la temperatura en todo momento, para poder adaptar la refrigeración a las condiciones existentes.

1.2. OBJETIVOS

Teniendo en cuenta todas las necesidades expuestas en el apartado 1.1, es evidente la importancia de descubrir métodos que permitan reducir estos gastos de energía, con el fin de que los centros de datos de las empresas funcionen con un compromiso con el medio ambiente y de forma económica para las mismas. Para crear una solución, se definen distintos objetivos:

- El objetivo principal de este proyecto es obtener un sistema que permita optimizar el funcionamiento de los sistemas de refrigeración, ya que supone una porción importante de la energía consumida. Para ello se diseñará un modelo de aprendizaje automático con la capacidad de predecir de forma fiable la temperatura que habrá en un punto determinado del centro de datos, habilitando así, la adaptación continua de los elementos encargados de la

refrigeración en cada instante.

- Adquirir conocimientos sobre el entorno de los CPD, la situación actual, su funcionamiento general y el de los sistemas de refrigeración. También, se investigará acerca de las tecnologías del estado del arte con el mismo objetivo de este proyecto, reducir el consumo energético de los centros.
- Comprender las bases y profundizar en el estudio del aprendizaje automático y las redes neuronales a utilizar, así como su aplicación a este novedoso problema, en concreto. Particularmente, analizar las Redes Neuronales Recurrentes y las Convolucionales, ya que se prevee que serán adoptadas para el diseño de la red. Además, aprender a manejar correctamente el lenguaje de programación (*TensorFlow* y *Keras*). Con el conocimiento adquirido, se pretende diseñar una ANN que ofrezca unos resultados de calidad.
- Hacer una investigación acerca de si se podría reducir la cantidad de mediciones realizadas por los sensores, obteniendo un resultado similar al obtenido con la base de datos completa. Con este proceso se podrían reducir muchos costes relacionados con el mantenimiento de los centros de datos.
 - Se tratará de comprobar, a partir de los datos disponibles, la cantidad de sensores necesarios para obtener resultados aceptables. En caso positivo, se obtendrá un importante ahorro a la hora de adquirir los sensores.
 - Se estudiará la cantidad de mediciones realizadas por los sensores, con el propósito de intentar reducirla. Este decremento de mediciones alargaría la batería de los sensores, por lo que su vida útil incrementaría.

1.3. CENTROS DE PROCESADO DE DATOS

Los centros de procesamiento de datos son lugares que contienen los recursos de computación de una o varias organizaciones. Poseen servidores que desempeñan la función de sistemas de procesamiento y almacenamiento de datos, apilados en racks. Estos racks se colocan de manera contigua, formando filas. Los servidores están conectados entre sí a través de múltiples conmutadores Ethernet, que se conectan a su vez con un conmutador central.

Para medir la eficiencia energética en la industria de este sector se usa la métrica PUE. Cuando el valor de esta unidad es cercano a 1,0 casi toda la energía suministrada al centro se emplea para



Figura 3: Racks en un centro de datos. [4]

alimentar a los ordenadores. Si el valor está próximo a 2,0 significa que por cada vatio de potencia de alimentación a los sistemas informáticos, se consume un vatio adicional para otros propósitos como la refrigeración, distribución de energía al resto del equipo, etc. Por ello, lo deseable es que el valor de este coeficiente sea lo más próximo posible a 1. El PUE se calcula utilizando las energías consumidas a lo largo de un año, correspondientes al total del centro de datos, y a la consumida por los sistemas IT:

$$PUE = \frac{EnergiaTotalCPD}{EnergiaSistemasIT}$$

1.3.1. REFRIGERACIÓN

Uno de los métodos de refrigeración más utilizados es el desempeñado por las unidades CRAC (Aire Acondicionado Para Sala De Computadoras). Funcionan con un compresor que aspira aire a través de un componente con refrigerante. Son bastante ineficientes en términos de uso de energía, pero el equipo en sí es relativamente económico.

Para que todo este proceso sea efectivo, los racks se deben colocar en filas de forma que los pasillos que las separan se alternen entre los denominados pasillos fríos y pasillos calientes.

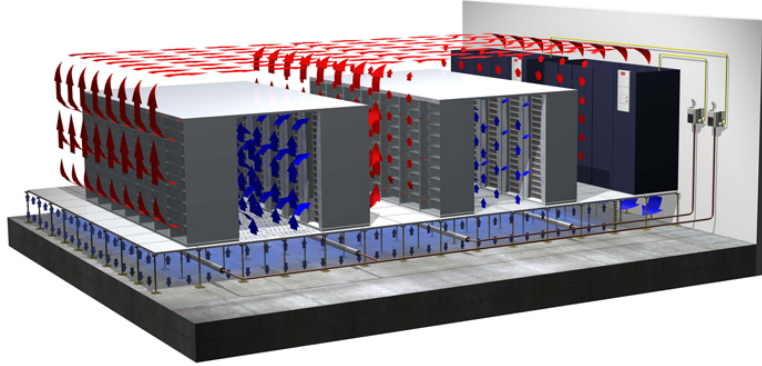


Figura 4: Organización pasillo frío y caliente [4]

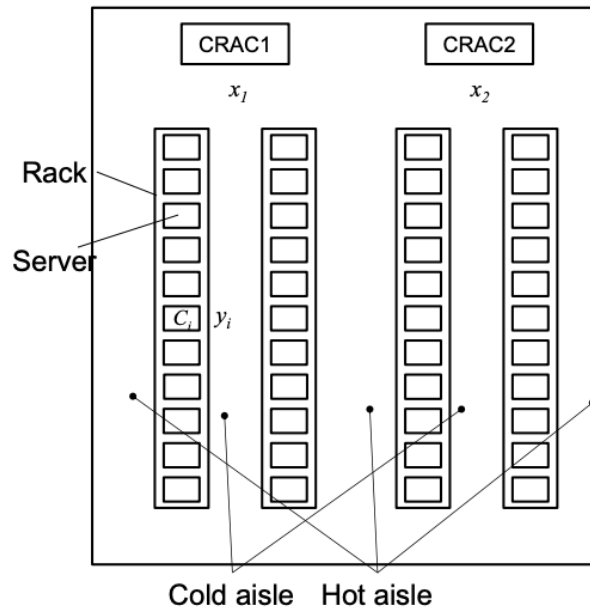


Figura 5: Planta de una sala de un CPD [23]

El pasillo frío inyecta aire frío, procedente del refrigerador, en la parte delantera de los servidores. El pasillo caliente extrae el aire caliente expulsado por la parte trasera de los servidores, introduciéndolo en el aire acondicionado, donde será refrigerado. Posteriormente, una vez refrigerado, ese aire pasará al pasillo frío de nuevo.

Para el desarrollo de este proyecto, se hablará de tres temperaturas:

- *Inlet*: es la temperatura entrante en los servidores medida en los sensores dispuestos.

- *Outlet*: representa la temperatura que sale de los servidores hacia el pasillo caliente.
- *Return*: la temperatura que retorna a los CRAC.

1.4. APRENDIZAJE AUTOMÁTICO

Para comprender la herramienta principal de este proyecto, las redes neuronales, se debe conocer las bases del aprendizaje automático. El aprendizaje automático (*Machine Learning*) es el conjunto de técnicas que tratan de conseguir que los computadores, mediante algoritmos, adquieran la habilidad de realizar una tarea determinada por sí mismos, a partir de la introducción de datos. El aprendizaje consiste en desempeñar esta tarea de forma cada vez más correcta, minimizando progresivamente el error en la medida de lo posible.

El científico informático Tom Mitchell proporciona una definición muy acertada de Machine Learning, orientada al mundo ingenieril:

A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E .

Por ejemplo, una aplicación posible de esta técnica sería el diseño de un sistema dotado de la capacidad de dictaminar la patología que sufre un paciente, disponiendo de sus síntomas.

El modelo de aprendizaje automático recibe un conjunto de datos de entrada (múltiples casos clínicos), que le servirá para adquirir experiencia y aprender a clasificar nuevos datos (diagnosticar patologías). Este conjunto de datos de entrada se denomina datos de entrenamiento (*training set*), y cada uno de estos, recibe la denominación de *instancia* o *ejemplo*. Siguiendo el ejemplo anterior, una instancia representaría el caso clínico de un paciente. Cada una de estas instancias está caracterizada por una serie de *atributos*, que serían los síntomas del paciente en el ejemplo.

El entrenamiento o aprendizaje puede ser de dos clases: [19]

- **No supervisado**: el sistema recibe datos de entrada, a partir de los cuales debe aprender a reconocer patrones. Estos estarán presentes en las distintas combinaciones de atributos.
- **Supervisado**: en este método las instancias del data-set deben estar etiquetadas, es decir, se deben insertar las instancias con sus atributos y su *clase*. Esta clase se trata de la clasificación en sí del ejemplo, prosiguiendo con el ejemplo anterior, la clase sería la patología sufrida

por el paciente. Mediante este tipo de *learning*, se supervisa al algoritmo de aprendizaje, permitiendo medir su precisión durante la fase de entrenamiento y ganando experiencia a través del tiempo. Se trata de métodos más precisos que los no supervisados, con la desventaja de la necesidad del ser humano para etiquetar los datos de entrada.

Dependiendo de la clase a predecir, el sistema puede ser de clasificación o regresión:

- **Clasificación:** el resultado final es una clase de entre un número limitado de posibilidades. Por ejemplo, si el resultado puede ser únicamente sí o no.
- **Regresión:** el resultado a predecir es numérico, pudiendo haber un número infinito de posibilidades.

Una vez se ha completado el entrenamiento, se pasa a la etapa de evaluación del modelo. Para ello, se introduce un conjunto de datos de test (*training set*) sin etiquetar y el sistema entrenado realizará una predicción de la clase en base a los parámetros de las instancias introducidas. Completada tal predicción, el diseñador del modelo debe comprobar su calidad, comparando las predicciones con los valores reales de las clases de las instancias del conjunto de test.

1.5. REDES NEURONALES ARTIFICIALES

Las redes neuronales son un método de aprendizaje automático que simulan el comportamiento de las neuronas de los seres vivos. Estas se conectan entre sí, desempeñando una tarea en conjunto, sin tener asignada ninguna de ellas una tarea única. Las neuronas biológicas poseen bajas capacidades de procesamiento, sin embargo toda su capacidad cognitiva se sustenta en la conectividad de éstas. La experiencia que adquieren la emplean para reforzar ciertas conexiones, lo cual mejorará el funcionamiento de la red. [20]

Como método de aprendizaje automático que es, su utilidad reside en la capacidad de reconocer patrones en los parámetros de los datos introducidos. La red trata de aprender a predecir o clasificar los datos de forma correcta, a partir de las combinaciones de los distintos parámetros de las instancias introducidas. Esto se lleva a cabo ajustando los pesos progresivamente, de forma que el error entre la salida de la red y la salida deseada, sea mínimo.

La utilidad de las RNN reside en problemas en los que se tienen montones de datos caracterizados por una serie de atributos. Por ejemplo, predecir el rendimiento de un jugador de fútbol teniendo en cuenta una serie de parámetros de sus partidos jugados, predecir el éxito que va a tener una

canción según características de su sonido, o incluso clasificar plantas según sus medidas y atributos físicos.

Para comprender las redes neuronales artificiales, se debe comenzar por la unidad más simple: el perceptrón.

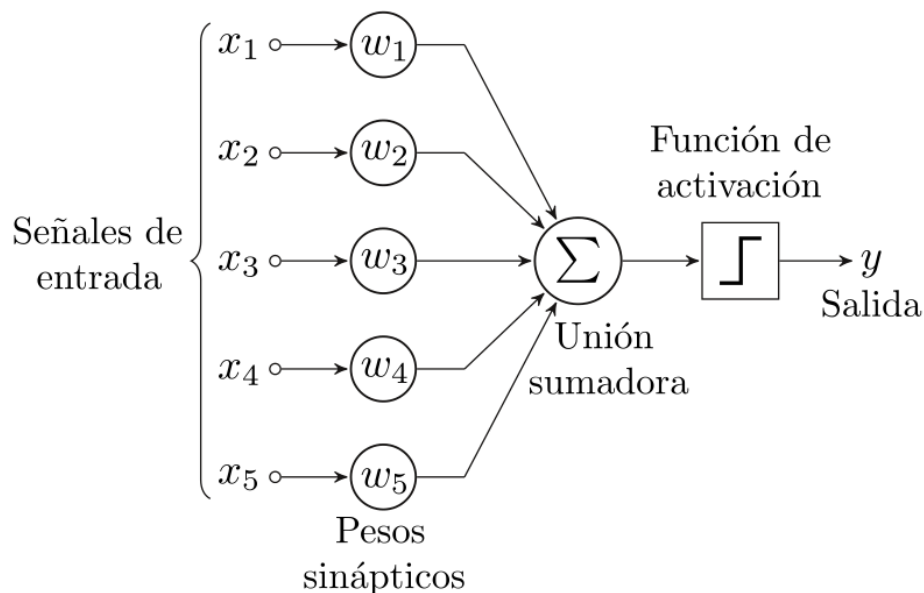


Figura 6: El perceptrón

Las entradas corresponden a los valores de los distintos atributos de cada instancia de los datos introducidos:

$$x = (x_1, \dots, x_n)^T$$

Los pesos son la simulación de las conexiones sinápticas de las neuronas de los seres vivos. Determinan el efecto que generará un nodo en el nodo siguiente. Son asociados directamente con las conexiones de las ANN. El valor de este peso determina la intensidad de esta conexión entre neuronas. Cada uno de ellos comienza siendo un valor aleatorio o constante. Al ser entrenada, la red tratará de encontrar los valores óptimos de estos pesos para obtener el mejor resultado. El conjunto de los pesos forma lo que se denomina el vector de pesos:

$$w = (w_1, \dots, w_n)^T$$

Como se puede ver en la figura 6, esta unidad mínima, realiza el sumatorio procedente del producto

escalar del vector de pesos y las entradas. El resultado de la anterior operación es el siguiente:

$$y' = \sum_{i=1}^n x_i w_i$$

Posteriormente, esta y' pasa por una función de activación, dando lugar a la salida final:

$$y = \gamma \sum_{i=1}^n x_i w_i$$

Esta función de activación es una función no lineal, que controla el comportamiento de la neurona. Transforma la salida del sumatorio que combina los productos del peso con la entrada, de forma que adquiera un valor dentro de un rango adecuado, como puede ser $[0,1]$. Estas funciones propagan la salida de cada neurona de una capa, hacia la siguiente capa.

Existe la posibilidad de escoger la función de activación entre una infinidad de opciones, pero las más empleadas son:

- *Sigmoide*. La función sigmoide adapta a la escala $(0,1)$ los valores de la entrada que recibe. Los valores más cercanos al máximo tenderán asintóticamente a 1, y los más bajos tenderán a 0. Es de especial utilidad en casos en los que se tienen datos con valores comprendidos en un intervalo entre un valor máximo y un valor mínimo.

A continuación en la figura 7, se puede observar su apariencia y su fórmula:

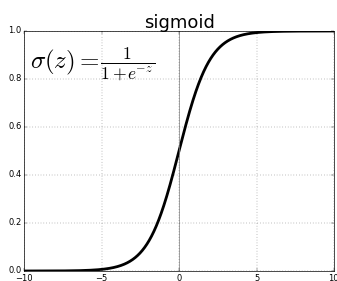


Figura 7: Función sigmoide

- *Rectified Lineal Unit (ReLU)* Esta función transforma los valores de entrada de manera que anula los valores negativos y no produce modificaciones en los valores positivos. Se suele usar en redes neuronales con aplicaciones relacionadas con el tratamiento de imágenes.

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

- *Tangente hiperbólica* La tangente hiperbólica transforma la entrada a valores dentro del intervalo $(-1,1)$. Los valores que tienden al máximo, tenderán a 1 asintóticamente y los cercanos al mínimo, a -1.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1$$

A continuación, se muestra un esquema de la unión de varias neuronas para observar la apariencia de una red neuronal artificial completa:

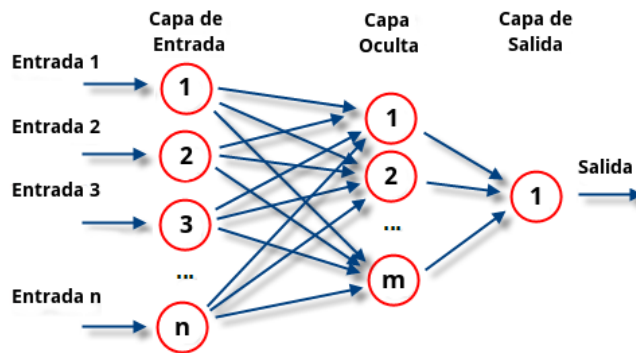


Figura 8: Red Neuronal Artificial. [2]

- *Capa de entrada.* Las neuronas de esta capa reciben directamente los datos de entrada. No se produce procesamiento de los datos en esta capa.
- *Capa oculta.* Esta capa recibe la salida de capas anteriores como entrada, y sus neuronas procesan los datos entregándolos a la siguiente capa. Puede haber más de una capa oculta, el número óptimo de estas depende de la aplicación.
- *Capa de salida.* La salida de esta capa de neuronas corresponde a la salida total de la red. Puede contener una o varias neuronas, dependiendo de si es un modelo de regresión (tendría una sola neurona) o clasificación con varias clases (varias neuronas en la capa de salida).

Las redes neuronales de aprendizaje profundo son capaces de optimizar su funcionamiento a lo largo del entrenamiento gracias al proceso denominado *optimización*. Este concepto consiste en tratar de minimizar el error por medio de la modificación de parámetros de la red como los pesos. Tal tarea se ejecuta mediante el descenso de gradiente. Este proceso consiste en varias fases: [11]

1. Introducción de N ejemplos en la red, dando lugar unas predicciones tras los cálculos llevados a cabo por la red.
2. Cálculo de la función de pérdidas (*"loss function"*) en las distintas iteraciones. Se evalúa la diferencia existente entre las predicciones y los valores reales de la clase. Se calcula mediante una función matemática y da lugar al valor que la red tratará de minimizar en cada paso. Una vez calculada, en cada iteración la red recalculará los pesos para reducirla. La función de pérdidas a utilizar es elegida por el usuario dependiendo del problema a resolver.
3. Cálculo del gradiente, que es el conjunto de las derivadas parciales de la función de pérdidas respecto a los parámetros de la red. Es un vector que indica la dirección en la que la *"loss function"* incrementa de manera más rápida, de forma que se tratará de mover en la dirección opuesta para minimizarla. Se trata de hallar el mínimo global de la forma más rápida posible. Gráficamente:

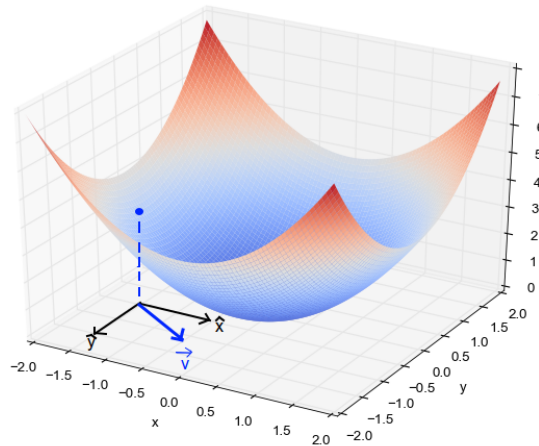


Figura 9: Descenso de gradiente.

En la práctica es algo mucho más complejo que lo mostrado en la figura 9, ya que se realiza el descenso de gradiente con un número mucho mayor de variables (en la figura únicamente hay dos), dando lugar a confusión de mínimos globales con los locales. Para simplificar tal problema se toma la técnica conocida como *"back-propagation"*, consistente en el cálculo de las derivadas parciales de la función de pérdidas respecto a los parámetros de la última capa, para posteriormente calcular las derivadas parciales de la *loss function* respecto a la anterior capa, de forma sucesiva hasta llegar a la primera capa de la red.

4. Por último, tras haber calculado el vector gradiente, se actualizan los parámetros de la red,

restándoles el valor del gradiente multiplicado por la tasa de aprendizaje. Esta tasa permite controlar el impacto que crea cada una de las iteraciones en el proceso global. Se lleva a cabo una resta del gradiente, debido a que, como se ha mencionado, se desea ir en contra del mismo, para que disminuya la función de coste.

1.5.1. REDES NEURONALES RECURRENTE (RNN)

Las RNN son una tipología de redes neuronales empleadas para predecir el futuro, a partir de series temporales. Para su funcionamiento, no son necesarios unos datos de entrada con longitud fija, por lo que son especialmente útiles para campos como el procesamiento del lenguaje.[15]

Como se aprecia en la figura 10, en cada iteración de tiempo ("*timestep*"), cada neurona recibe los datos entrada $x(t)$ y la propia salida de la capa en la iteración anterior $y(t-1)$.

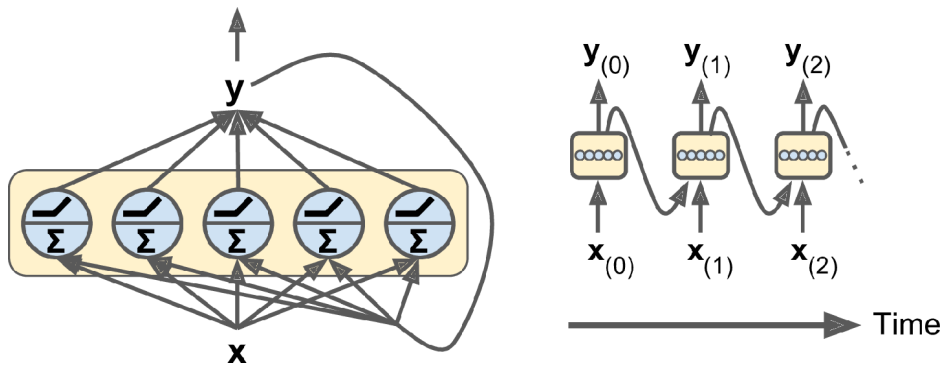


Figura 10: Una capa de una RNN a lo largo del tiempo. [15]

Resulta la siguiente salida, siendo b el sesgo, y γ la función de activación:

$$y(t) = \gamma(W_x^t x(t) + W_y^t y(t-1) + b)$$

W_x y W_y representan respectivamente los vectores de pesos de los datos de entrada $x(t)$ y de la salida $y(t-1)$.

Las neuronas y las capas de esta clase de ANN reciben el nombre de células de memoria, ya que son capaces de conservar el estado de la red a lo largo del tiempo. El estado de una célula a lo largo del tiempo se designa como $h(t)$, y depende de la entrada $x(t)$ y del estado del instante de tiempo anterior $h(t-1)$. De forma que la salida de $y(t)$ de cada célula queda determinada por el estado anterior y por la entrada.

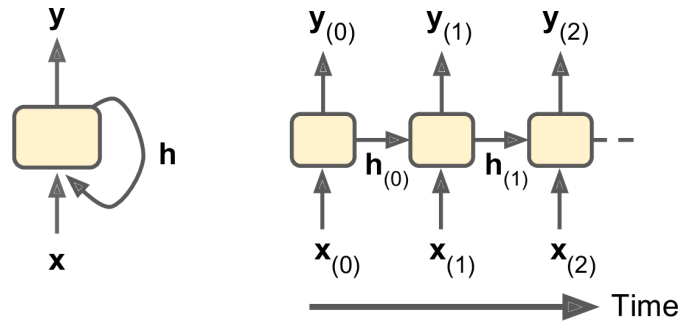


Figura 11: Células de memoria en una red neuronal [15]

Sin embargo, estas células son muy simples y no son capaces de memorizar estados complejos. Con el fin de conseguir células con mayor memoria, surgen las LSTM y las GRU. Se profundizará sobre estas últimas en la sección 1.5.2.

1.5.2. CÉLULAS GRU

Las células de memoria Gated Recurrent Unit (GRU) fueron planteadas por Kyunghyun Cho en 2014.

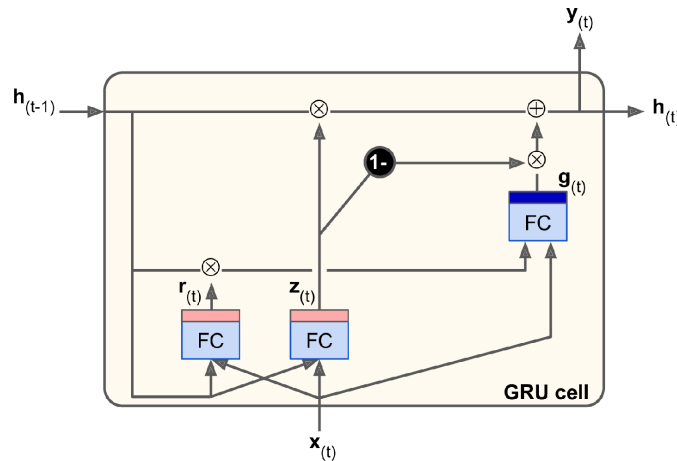


Figura 12: Célula GRU. [15]

- $h(t)$ une la entrada y el estado anterior.
- $z(t)$ se encarga de que cuando sea necesario almacenar memoria, se borra lo que haya almacenado en el lugar donde vaya a ser guardada la nueva memoria.
- $r(t)$ controla la parte del estado que recibe $g(t)$

Ofrecen un resultado bastante mejor que las RNN simples con secuencias de gran longitud.

1.5.3. REDES NEURONALES CONVOLUCIONALES (CNN)

Las Redes Neuronales Convolucionales, son una clase de ANN que surge como forma de mejorar el procesamiento de imágenes. Estas tratan de solucionar inconvenientes propios de las redes neuronales convencionales, decrementando la cantidad de información procesada por la red, reduciendo así la capacidad computacional requerida y, en la mayoría de casos, dando lugar a un mejor rendimiento del modelo.

Las capas ocultas convolucionales son capaces de encontrar patrones en los datos de entrada. Para cumplir tal objetivo, aplica filtros sobre el *input*, que recorren la totalidad de los datos en busca de patrones, los cuales son extraídos a través de "parches". Así, se permite construir redes neuronales más sencillas.

Cada uno de los filtros aplicados tendrá asignada determinada información que deberá extraer del total de datos. El tamaño de la salida de estas capas serán tantas matrices convolucionales como filtros se apliquen.

Para comprender mejor este proceso, se refleja en la figura 13 un ejemplo para el procesamiento de imágenes:

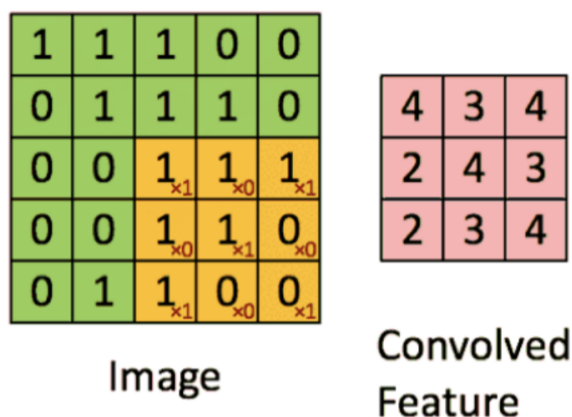


Figura 13: Funcionamiento de las CNN

Es fundamental comprender también el parámetro denominado "*stride*" o paso. El paso indica la cantidad de movimiento que realiza el filtro sobre los datos introducidos en cada paso.

2. ESTADO DEL ARTE

Ante la situación ya relatada, del crecimiento acelerado de la generación, procesado y almacenamiento de datos, y por tanto, de los CPD, desde los inicios de la década pasada, comienza a surgir una conciencia, cada vez más generalizada, que tiene como objetivo paliar los efectos medioambientales que supone esta tesitura. Se proponen, por ello, una gran cantidad de soluciones que gozan de mucha diversidad.

En este apartado se tratarán de reflejar algunas de las técnicas más novedosas destinadas a la reducción del consumo de energía destinada a la refrigeración de los centros de datos. Se analizarán los resultados de proyectos similares al de este Trabajo de Fin de Grado, que proponen el *Machine Learning* como herramienta para optimizar la refrigeración. Sin embargo, también serán expuestas otro tipo de tecnologías totalmente distintas, con el fin de que se cree una vista global de las propuestas más innovadoras e importantes en el sector.

2.1. GREEN COMPUTING

En la actualidad, debido a la gran preocupación que existe debido al impacto que generará el gran crecimiento del uso de los recursos informáticos en el medio ambiente (como se ha explicado en la sección 1), surge el concepto de "*Green Computing*".

Green Computing es un conjunto de métodos a realizar, con el fin de reducir el consumo de energía y de la huella de carbono. Al efecto, se debe recurrir a las energías renovables, y al mantenimiento de los centros de la forma más ecológica posible. Se deben seguir varios puntos para cumplir con ello:

- Uso inteligente de la energía y de los sistemas de información.
- Reutilización y reciclaje de los componentes informáticos, así como reducir los residuos generados.
- Empleo de componentes manufacturados de forma conveniente para el medio ambiente.

2.2. MACHINE LEARNING PARA REDUCIR ENERGÍA DE REFRIGERACIÓN

Otra de las soluciones estudiadas es la que emplea el aprendizaje automático. En diversos estudios,[18] se comprueba que emplear métodos de aprendizaje automático con parámetros interdependientes de los centros de datos, ofrece buenos resultados. Se trata de obtener un modelo que consiga controlar la temperatura de consigna, a partir del procesamiento de distintos parámetros de las salas de servidores.

En 2016, en la Universidad de Osaka, se completó un estudio [29] para la *IEEE International Conference on Cloud Engineering*, con el objetivo de reducir el consumo de potencia en los CPD a través de la predicción de temperatura mediante técnicas de aprendizaje automático, como la regresión lineal o *random forest*. Para ello, se empleaban principalmente parámetros como las temperaturas *inlet* y *outlet* de los sensores colocados en los servidores, la temperatura saliente de los aires acondicionados, entre otros.

A través de la aplicación de este sistema en un centro de datos experimental en Japón, implantando un aire acondicionado controlado proactivamente con el modelo entrenado, se consiguió reducir un 17% el consumo medio de potencia de los aires acondicionados. Además, se obtuvo un ahorro del 30% en el momento de consumo máximo. Se concluyó así, con esta investigación, que la implantación del aprendizaje automático para el control proactivo del aire acondicionado en los centros de datos, es una medida eficiente para reducir el consumo.

En líneas similares, se puede ver el progreso de *Google* mediante su proyecto de aprendizaje automático, *DeepMind*, con el que ha conseguido reducir hasta un 40% la energía consumida por la refrigeración de sus centros de datos, promediando un 15% de reducción. El proyecto consiste en entrenar redes neuronales a partir de grandes cantidades de datos medidos por sensores colocados por el centro. Las mediciones recogen atributos como la temperatura, la potencia, etc. Concretamente, se entrena una red que predice el PUE futuro, otras dos redes que predicen la temperatura y la presión existentes dentro de una hora. De esta manera, se sabrá el estado inminente del centro y se tendrá la posibilidad de poner en marcha las prácticas necesarias en cada momento.

En la figura 14 se observa el cambio sufrido por el PUE a causa de la puesta en marcha de las redes neuronales entrenadas.

En un estudio realizado por Jim Gao para *Google*, también se trata de predecir el PUE de centros de datos de la compañía, a través de redes neuronales artificiales. Se consiguen evidencias de que la aplicación de técnicas de aprendizaje automático son muy adecuadas para mejorar la eficiencia

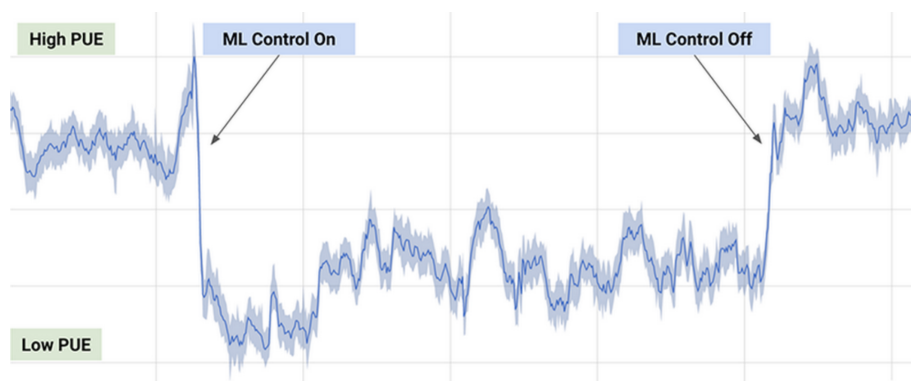


Figura 14: Evolución del PUE de Google en función del uso de aprendizaje automático. [26]

energética de los CPD, usando datos que provienen de los sensores dispuestos a lo largo del centro. Mediante este modelo, se logra obtener un PUE de 1,1 y, por tanto, grandes ahorros de coste. Como limitación se menciona la necesidad de disponer de grandes cantidades de datos.[13]

Aparte de las ANN, otros algoritmos de aprendizaje, como *Naïve-Bayes* o SVM, están siendo usados para resolver este problema. Algunos de los parámetros más influyentes para alcanzar buenos funcionamientos de estos sistemas son: las dimensiones de la sala, la altura del techo, la anchura del pasillo, la distancia entre los servidores y la temperatura Inlet.[25]

2.3. FREE COOLING TECHNIQUES

Otras tecnologías utilizadas para reducir estos gastos, son las denominadas *Free Cooling Technologies*, que pretenden enfriar los centros de datos aprovechando recursos naturales de refrigeración sin coste o con coste bajo. Estas están muy generalizadas actualmente, y se usan en gran parte de los centros, ya que las compañías tienen en cuenta este factor a la hora de disponer el CPD.

Si la temperatura exterior es suficientemente inferior a la del centro de datos, el calor sale hacia el exterior, sin necesidad de disponer de aire acondicionado de compresión. Además de reducir los costes energéticos, también genera ahorro de capital, permitiendo reducir el tamaño de las salas de refrigeración. Por este motivo, se sitúan los centros de datos en localizaciones geográficas secas con temperaturas bajas. [24]

Las soluciones se clasifican en 4 campos principales:

- *Aire Directo (DA)*: se sirve del aire del exterior directamente para refrigerar.

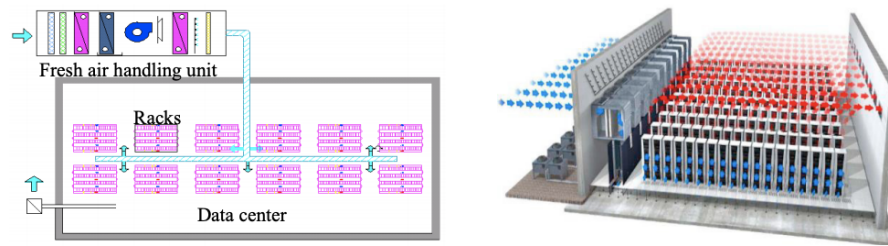


Figura 15: Refrigeración por Aire Directo. [28]

- *Aire Indirecto (IA)*: a través de un intercambiador de aire, recircula el aire interior con el exterior, eliminando el calor.
- *Direct Water (DW)*: emplea agua enfriada por el aire del exterior de forma directa para refrigerar la sala.
- *Indirect Water(IW)*: utiliza un intercambiador de calor para enfriar el agua a través de la temperatura exterior.

Volviendo a Google, cabe mencionar que llevan a cabo prácticas de este tipo, elevando la temperatura de sus centros a 26°C y refrigeran utilizando el aire del exterior. Con eso, sumado a un conjunto de otras buenas prácticas, como la alimentación a través otras energías renovables como la eólica, o el, ya tratado, uso de aprendizaje automático, han llegado a alcanzar un PUE de 1,11 de media. Además, han completado una reducción del consumo energético del 11 %. [14]

También se ha de destacar la opción seleccionada por Microsoft, que opta por sumergir los centros de datos en el fondo del mar. Lleva a cabo la refrigeración utilizando una energía renovable, como son las frías aguas del mar, utilizando un sistema de tuberías de intercambio de calor, reduciendo de esta forma los costes de operación. En primavera de 2018, se introdujo un centro de datos en el mar y se han supervisado sus resultados hasta que en 2020 se subió a la superficie, donde se comprobó que la tasa de fallos ha sido ocho veces inferior a la de un CPD convencional, debido a que el contenedor estaba sellado y se inyectó nitrógeno en su interior. Además, mejora la latencia de las zonas costeras próximas. [22]

La compañía estudia la posibilidad de alimentar los servidores con energía eólica, gracias a los vientos de la costa, consiguiendo que sean centros aún más "verdes". [12]

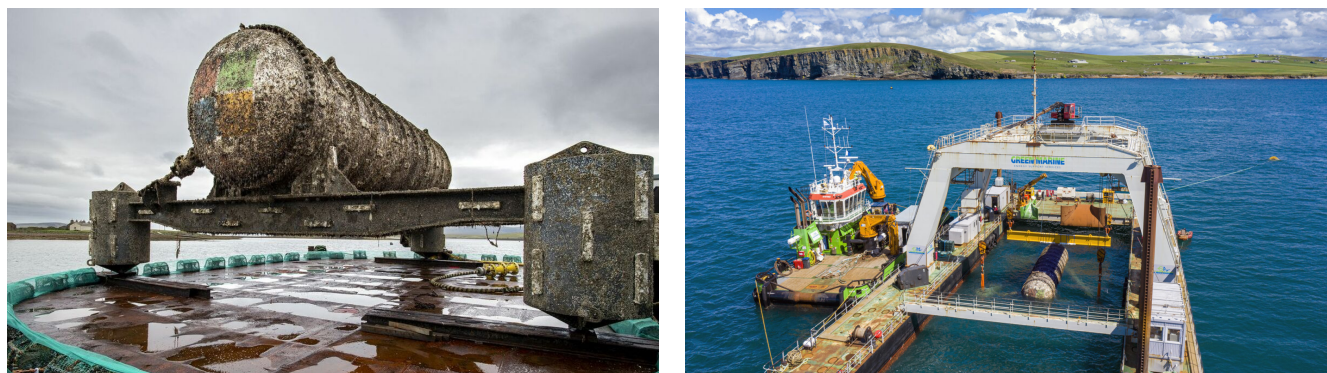


Figura 16: Centro de datos de Microsoft sumergido en el agua. [22]

En el plano nacional, se encuentra el caso de la empresa española con mayor plantilla, *Mercadona*, que también apuesta por localizar uno de sus dos CPD en un clima frío. Concretamente, se ubica en Villadangos del Páramo, un pequeño pueblo de León, cuyo clima aporta una reducción significativa del consumo eléctrico por el ahorro en climatización. [10]

2.4. OTROS

Por ejemplo, *Apple* con su programa de eficiencia energética trata de adaptar de la mejor forma sus centros a las necesidades específicas. Para ello, se desarrolla un sistema de seguimiento de los sistemas de energía que identifica los posibles fallos. Las nuevas construcciones se hacen atendiendo a factores como la temperatura exterior, humedad, etc. Además, tratan de reorganizar y renovar sus localizaciones antiguas, asegurando un buen rendimiento en todos los centros. Para tener un completo control, se monitoriza continuamente el funcionamiento de los centros, de esta forma se pueden arreglar los fallos. [7]

Gracias a este programa, la compañía ha logrado trabajar un 60 % mejor que las especificaciones dadas por la ASHRAE², reduciendo en 2020 el uso energético en 13,9 MWh anuales. Con estas medidas, se calcula evitar la creación de unas 4.900 toneladas de CO_2 al año.

²ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) crea estándares globales de diseño energético eficiente, incluyendo el estándar específico para los centros de datos.

2.5. CONCLUSIONES

De esta diversidad de aplicaciones enfocadas a satisfacer la misma necesidad, se sacan varias conclusiones:

- La clara insistencia de compañías tan grandes como *Google*, *Apple* o *Microsoft*, manifiesta el hecho de que es absolutamente necesario invertir en reducir los gastos energéticos en refrigeración, para conseguir un PUE lo más cercano posible a 1. La conciencia creada es tranquilizadora, ya que está generando iniciativas como *Green computing*, que tratan de decrementar los efectos negativos que posee el ecosistema de las tecnologías de la información y la computación, en el medio ambiente.
- Las conocidas como *Free Cooling Techniques* producen satisfactorios aportes, siendo muy amigables con el medio ambiente. Sin embargo, algunas de las soluciones expuestas de este tipo, no están al alcance de cualquier compañía, como por ejemplo las basadas en sumergir los centros de datos, para lo cual se necesita la construcción y el despliegue de infraestructuras muy caras. Las compañías tienen en cuenta este factor, que es diferencial, ya que causa perceptibles efectos sobre el funcionamiento de los centros.
- La aplicación de tecnologías basadas en aprendizaje automático es totalmente beneficiosa, como se ha visto en la sección 2.2. Usando este tipo de soluciones se han conseguido grandes reducciones de los gastos energéticos relacionados con la refrigeración, alcanzando hasta un 40 % en estudios reflejados. Posee, en adición, la ventaja de ser más económica que otras resoluciones expuestas, ya que está basada principalmente en software y en la disposición de sensores de medición de las condiciones por las salas de servidores. El hecho de que varias empresas importantes apuesten por esta técnica, aporta consistencia a la idea de elegir esta opción. Todo esto, notoriamente, deja evidencias de que la utilización de *Machine learning* es un buen camino a seguir para resolver el problema expuesto, motivando estudios como el objeto de este Trabajo de Fin de Grado.

3. DESARROLLO

Para la creación, entrenamiento y predicción de la red neuronal, se utiliza el lenguaje de programación *Python* tomando la biblioteca de aprendizaje profundo ("*Deep Learning*"), *Keras* incluida en la distribución oficial de *TensorFlow*. Esta herramienta es una API de alto nivel utilizada para la implementación y entrenamiento de modelos de aprendizaje profundo. Está diseñada para ofrecer APIs consistentes y simples, y reducir la carga cognitiva. Ha sido escogida como herramienta principal debido a su facilidad de uso, ya que da la posibilidad de crear y entrenar una red neuronal de forma eficiente e intuitiva, con técnicas del estado del arte. Posee una buena modularidad, ya que los bloques construidos se conectan fácilmente entre sí, sin muchas restricciones. En adición, dispone de una gran cantidad de documentación y guías de desarrollo. [1]

La librería *Numpy* ha sido utilizada de forma recurrente. Esta es una librería *Python* muy relevante para el uso científico de los ordenadores, ya que permite la manipulación eficiente de *arrays* multidimensionales y la aplicación de operaciones matemáticas, lógicas, de filtrado, etc.

Toda la programación ha sido realizada sobre el entorno interactivo *Google Colaboratory*, que permite escribir y ejecutar programas escritos en *Python* en el navegador, con la ventaja de que da acceso al usuario a GPUs de forma gratuita. Ejecuta el código en los servidores en la nube de *Google*, aportando la potencia de su hardware, por tanto, funcionará independientemente de la potencia del equipo en el que se use. La elección de este entorno se debe principalmente a la comodidad de la interfaz y al factor de los servidores en la nube. [6]

El proceso de desarrollo del proyecto posee las siguientes fases:

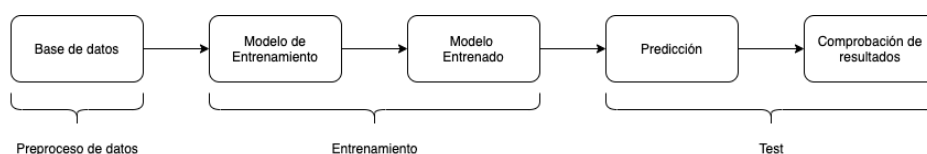


Figura 17: Diagrama de bloques del proceso de desarrollo.

3.1. DATOS DE ENTRADA

Para entrenar la red neuronal con el objetivo de que sea capaz de predecir la temperatura de forma acertada, es necesario disponer de un gran número de instancias de entrada. Los datos provienen de una sala de un centro de datos que tiene la siguiente apariencia:

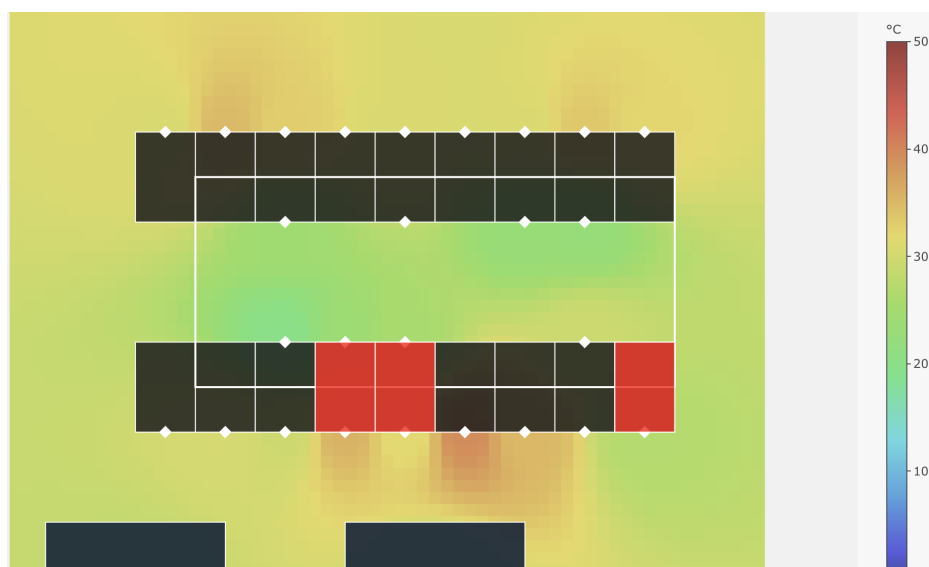


Figura 18: Centro de datos objeto de estudio.

En la figura 18 se ve la planta de la sala. Esta consta de dos filas de nueve racks cada una, separadas por un pasillo frío. En la otra cara de las filas, se observan los dos pasillos calientes. También se ven dos CRAC que realizarán el intercambio de aire frío y caliente.

Estos datos inicialmente, se dividen en dos bases de datos en formato *parquet*. Una de ellas, contiene las mediciones de parámetros realizadas por cada sensor; la otra recoge la posición de cada uno de ellos dentro del centro de datos a estudiar, a través de coordenadas x, y, z.

Para leer estas bases de datos, se toma como herramienta la librería *Pandas*. *Pandas* es una librería software dedicada al análisis y la manipulación de datos. Se ha optado por trabajar con ella a lo largo del proyecto, principalmente, debido a su potencia, flexibilidad, rapidez, simplicidad y a la numerosa documentación existente sobre ella.

Una vez leídas, se fusionan ambas tablas en una sola, obteniendo una base de datos más completa y útil, que posee los siguientes parámetros:

```
timeID, deviceID_x, ClientID_x, date, sensors, macaddress,  
timestamp, lostPackages, transferid, temperature, humidity,  
pressure, rssi, battery, battid, battime, Ambient, Building,  
ClientID_y, Inlet, Model Name, Outlet, Return, Supply, U,  
aisle, deviceID_y, rack, room, x, x_max, y, y_max, z
```

Para simplificar el manejo del instante en que se realizan las mediciones, se toma la columna

`timeID` (que tiene un formato complejo) para crear la columna `time`, que representa el tiempo en base a los minutos del día. Esta característica será útil para la asignación de la clase a cada instancia (proceso explicado en la sección 3.2.2).

De cara al entrenamiento, de entre todas las características iniciales, se seleccionan las siguientes, las cuales serán las más determinantes para la predicción de la temperatura:

`humidity, temperature, Inlet, Outlet, Return, x, y, z`

Nótese que los atributos de las instancias introducidas poseen distintos rangos de valor, lo cual genera un peor funcionamiento de la red, por lo que para completar la preparación de los datos, se lleva a cabo la normalización entre 0 y 1 de todos los parámetros. Las indicaciones de la documentación sobre ANN *TensorFlow* [9] son claras respecto a esto, en ellas se recomienda introducir valores pertenecientes a dicho rango en las redes de neuronales, ya que la optimización por gradiente descendiente ofrece un mejor rendimiento cuando los valores de las entradas pertenecen a rangos similares, causando que el proceso tarde menos tiempo en converger.[16] Además, el uso de valores menores permite alcanzar mejores resultados.

El gradiente descendiente funciona mucho mejor cuando los datos tienen una escala similar. Esto se debe a la forma geométrica de la función de coste. Cuando todas las dimensiones tiene una escala similar, se puede esperar la forma de cuenco multidimensional. Sin embargo, si distintas dimensiones tienen una escala diferente, la forma geométrica será muy diferente. Esto hace que el gradiente descendiente tarde mucho en converger.

`Inlet`, `Outlet` y `Return` tendrán únicamente valores 0 o 1. Cuando uno de estos sea igual a 1, indicará que la temperatura medida reflejada en los datos, será de esa tipología.

Así pues, anticipadamente a la fase de entrenamiento, las mediciones se clasifican en 3 bases de datos distintas, correspondientes a las distintos tipos de temperaturas. Una vez hecho esto, se eliminan las columnas de `Inlet`, `Outlet` y `Return`.

El total de instancias es separado entre los conjuntos de entrenamiento y *test*. Será beneficioso para el proceso encontrar el porcentaje óptimo de datos a seleccionar para cada uno de ellos. En este estudio hay que tener en cuenta principalmente dos factores: por un lado está el hecho de que es necesario una cantidad numerosa de datos, que sea suficiente para que el modelo sea capaz de predecir correctamente. Por otro lado, se tiene el caso contrario, que se conoce como sobreaprendizaje o sobreajuste (*'overfitting'*) .

El sobreajuste tiene lugar cuando el número de datos es excesivo, produciendo un "sobre-entrenamiento".

Esto quiere decir que el modelo se ajusta demasiado al conjunto de datos introducido, por lo que no será capaz de predecir de forma correcta los datos nuevos que se introduzcan a posteriori.

La elección final del reparto de instancias entre entrenamiento y test, se expone en la sección 4.

3.2. RED NEURONAL

3.2.1. DISEÑO

El objetivo del proyecto es conseguir predecir la temperatura, por lo que al ser un valor numérico, el modelo será de regresión. La regresión lineal es un método estadístico empleado para modelar la relación entre una variable dependiente y una o varias independientes. Estudia la correlación existente entre las variables y predice la de salida del sistema basándose en ellas. Es muy utilizado para la predicción de valores futuros y modelos de series temporales, como es el caso.

Calcula la salida en función de pesos de sus valores numéricos de los atributos de la entrada:

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Siendo β_i los pesos y x_i los atributos de entrada.

Estos pesos son obtenidos mediante el proceso de minimizar el error cuadrático medio (sección 3.2.2) entre la clase real y la predicha por la red, con la siguiente fórmula (siendo y_i el valor predicho.):

$$\sum_{i=1}^n (y_i - \sum_{j=1}^k \beta_j x_{ji})^2$$

Gráficamente, se muestra como una recta que trata de minimizar la distancia con los datos disponibles. De esta manera, genera el mejor comportamiento a la hora de predecir la salida a partir de los valores de los atributos introducidos, como se puede observar en la figura 19.

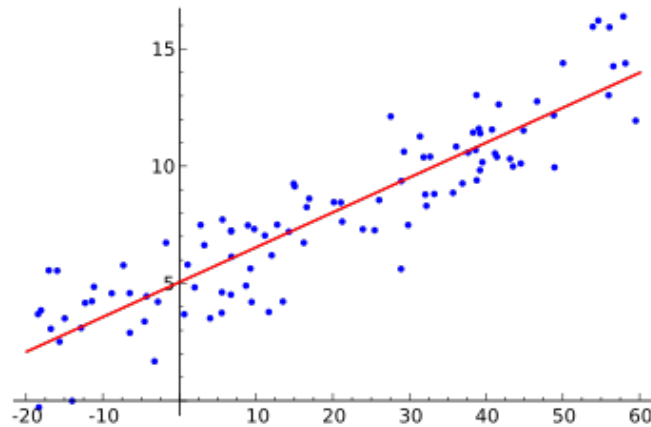


Figura 19: Regresión lineal.[3]

En este modelo se pretende conocer la temperatura que habrá en la sala de servidores dentro de 10 minutos, por tanto, se está tratando de predecir un valor de un momento cercano que está por llegar. Las redes que mejor funcionamiento aportan a la hora de predecir valores futuros son las Redes Neuronales Recurrentes (desarrolladas en la sección 1.5.1), debido a la memoria que poseen, por ello son la tipología de redes escogidas para este trabajo. Dentro de las RNN se seleccionan las células GRU (previamente explicadas en la sección 1.5.2), que ofrecen mejores resultados que las RNN simples, debido a que poseen mayor memoria. Estas células ofrecen mejor rendimiento para números reducidos de instancias que otras RNN, como por ejemplo las LSTM, que requieren un mayor tamaño de la base de datos. [15]

Sin embargo, a pesar de que las GRU presentan una memoria aceptable, esta puede ser mejorada para alcanzar resultados de mayor calidad. Con este fin, se decide colocar una capa convolucional unidimensional ("*1D Convolutional Layer*") anterior a las GRU. Ayudará a las GRU a conservar la información más relevante para la red, ignorando la que no es influyente para el modelo.

La capa extenderá varios kernel a lo largo de la secuencia introducida, dando lugar a un mapa unidimensional de atributos por kernel. Cada uno de estos kernel detectará un patrón de la secuencia con el tamaño de kernel (`kernel_size`) como máximo. Esta capa tendrá una salida con longitud igual al número de kernel.

La última capa posee una única neurona, debido a que el modelo predice únicamente un valor numérico (el de la temperatura). Para la capa de salida se selecciona una *Dense Layer*, que proporciona una ejecución algo más rápida, sin perder apenas precisión, dando libertad de escoger la función de activación que se desee.

Por tanto, la ANN diseñada consta de 5 capas:

- *1D-Convolutional Layer*: 5 neuronas
- Capas ocultas:
 - *GRU Layer*: 32 neuronas
 - *GRU Layer*: 32 neuronas
 - *GRU Layer*: 16 neuronas
- Capa de salida:
 - *Dense Layer*: 1 neurona

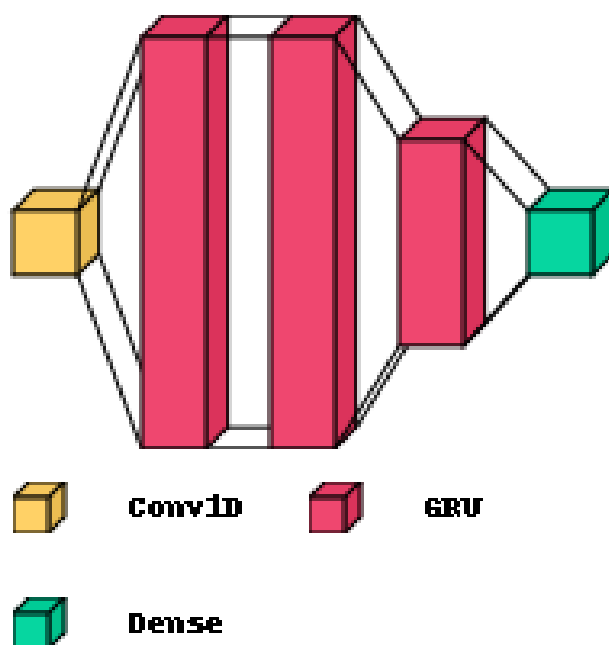


Figura 20: Arquitectura de la red neuronal.

Capa	Tipo	Tamaño Kernel	Stride	Función Activación	Size/ N° Neuronas
Conv1D	CNN	4	1	relu	(None,1)
GRU	RNN	N/A	N/A	tanh	32
GRU	RNN	N/A	N/A	tanh	32
GRU	RNN	N/A	N/A	tanh	16
Dense	Salida	N/A	N/A	Sigmoid	1

Tabla 1: Resumen del diseño de la red neuronal.

3.2.2. ENTRENAMIENTO

De la fase de procesado de datos, expuesta en la sección 3.1, resultan 3 conjuntos distintos de datos que poseen 5 atributos (`humidity`, `temperature`, `x`, `y`, `z`). Por consiguiente, los resultados de la red serán determinados por la temperatura, la humedad y la posición en el centro.

El entrenamiento será supervisado, por lo que es necesario suministrar a la red la clase a predecir. Por esta razón, antes de comenzar con el aprendizaje, se procede a introducirla en las bases de datos para calcularla, con el nombre de `temperature10min`, que será la temperatura real que habrá dentro de 10 minutos en el punto en el que se localiza cada sensor. Para ello, en los valores de este parámetro en cada instancia, se introducen los correspondientes al atributo `temperature` 10 minutos después, que será la temperatura real en ese instante futuro.

Lo siguiente será separar los atributos de la base de datos completa entre los atributos de entrenamiento y la salida esperada. Por un lado se crea `x_train` que contiene los valores de los atributos que se utilizan para entrenar a la red. Por el otro lado, se forma `y_train` que dispone de los valores de salida reales.

Al tratarse de un modelo de regresión lineal, se emplea el error cuadrático medio (MSE) como método para cuantificar el acierto de la red durante el entrenamiento, ya que es el más utilizado. El MSE mide la media de los cuadrados de los errores. El error indica la diferencia entre el valor real y el predicho por el modelo. Se aplica el cuadrado de esta diferencia para que no resulte un valor negativo.

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

\hat{y}_i representa el valor predicho e y_i el valor real.

La función `model.compile()` [5] de *Keras*, permite configurar los parámetros de entrenamiento del modelo. Esta función permite establecer:

- *optimizer*: recoge el valor del método de optimización del entrenamiento de la red. Se escoge 'adam' que utiliza el algoritmo de Adam, el cual optimiza el aprendizaje a través el descenso de gradiente estocástico (sección 1.5) basándose en la estimación adaptativa de momentos de primer o segundo orden.
- *loss*: este parámetro permite indicar al usuario la manera en que se desea calcular la función

de pérdidas (sección 1.5), imprescindible para la optimización. Por ser el modelo de regresión, se establece el `'mean_squared_error'` (MSE).

- *metrics*: recoge los resultados del acierto del modelo. Es similar a la función de pérdidas, mas este resultado no influye el entrenamiento.[21] Se elige `'mean_squared_error'` por la misma razón que en el parámetro *loss* .

Preparados los datos de entrada, la red neuronal y los parámetros de entrenamiento, se da comienzo a la fase entrenamiento. Se ejecuta la función `model.fit()` de *Keras*, en la que se introducen el modelo diseñado (*model*), *x_train*, *y_train*. En adición, permite establecer:

- *batch_size*: representa el número de instancias por cada actualización del gradiente. Se asigna el tamaño por defecto, igual a 32.
- *epochs*: número de epochs (iteraciones sobre el conjunto completo de *x* e *y*) para entrenar el modelo. Se emplean 5 epochs.

4. RESULTADOS

Tras la fase anterior de diseño y construcción del modelo, se obtiene el código disponible en la sección 9, con el que también se realiza todo el preproceso de datos, y se entrena y testea la red.

Superada la fase de entrenamiento de la red neuronal diseñada, se analizan los resultados obtenidos. Como se expone en la sección 1.2, se completarán varios experimentos. En primer lugar, se comprobarán el funcionamiento de la red entrenada con la totalidad de ejemplos de la base de datos. Después se contemplará la posible reducción del número de sensores dentro de la sala de servidores. En siguiente lugar, se decrementará el número de mediciones del entorno objeto de estudio.

Como se ha aclarado en el apartado anterior, se tomará el error cuadrático medio como método para medir los resultados de la red, observando a través de ello la calidad de las predicciones realizadas por la red. El método de *Keras*, `evaluate`, proporciona el MSE final al introducir los conjuntos de test (`x_test` e `y_test`) aplicando la función al modelo entrenado.

Se hará uso de la función `predict` de *Keras*, que proporcionará las predicciones que genera el modelo entrenado al introducir los datos de test reservados. Será útil para analizar si la red es capaz de pronosticar de manera fiable la temperatura futura, es decir, para examinar si existe o no sobreaprendizaje. Se construirán gráficas para los distintos experimentos en las que se compararán los valores reales de `temperature10min` y los predichos por la red.

4.1. EXPERIMENTO GLOBAL

En primer lugar, se desea investigar qué porcentaje es el óptimo para utilizar en el conjunto de entrenamiento y test. Tras analizar los resultados correspondientes a los entrenamientos contando con distintas divisiones de datos (figura 21), se concluye que el reparto óptimo de los mismos será tomando el 33 % de las instancias para test, y el resto para entrenamiento, ya que posee un MSE bajo y la utilización de más datos de entrenamiento causa overfitting.

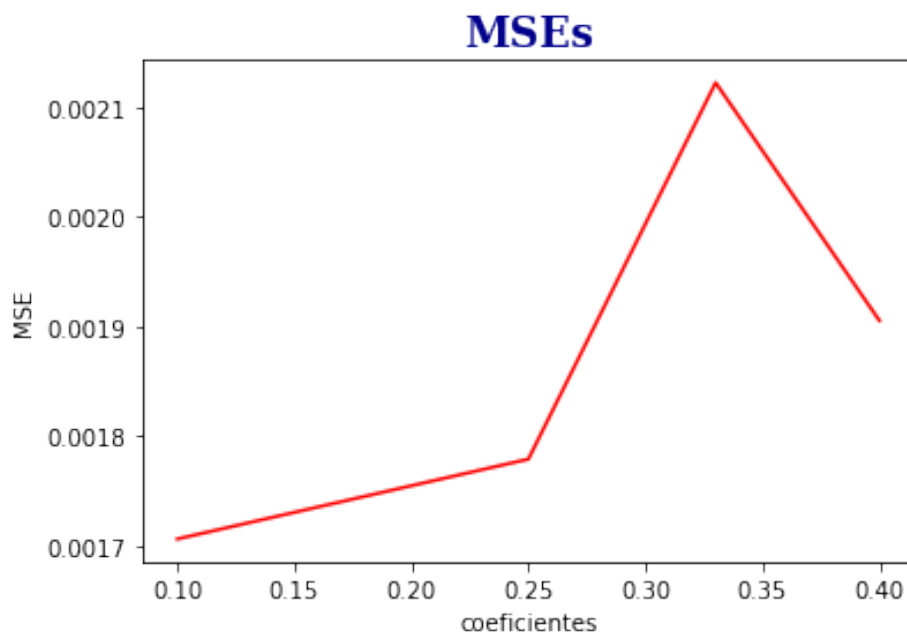
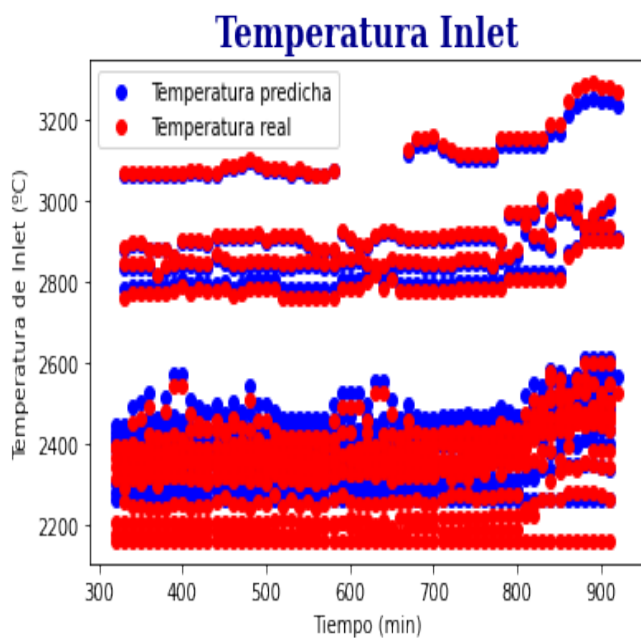


Figura 21: Evolución del MSE obtenido en el entrenamiento en función de los distintos repartos para entrenamiento y test.

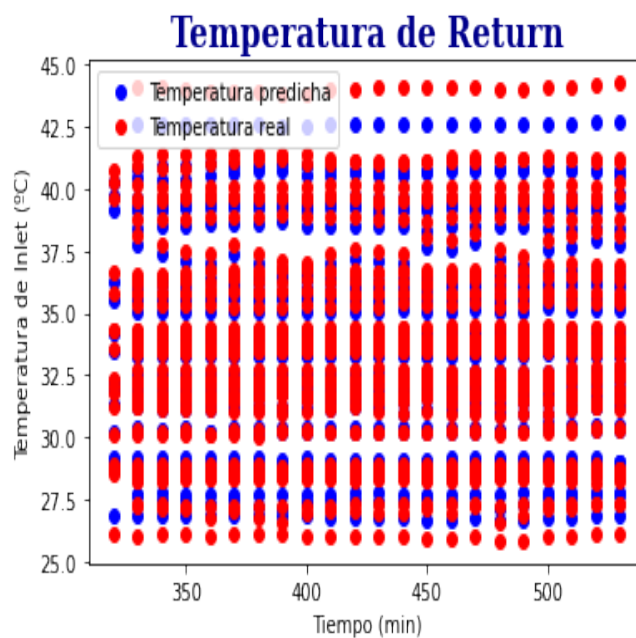
En este apartado se analiza el resultado alcanzado por la red al introducir el conjunto de entrenamiento al completo. Se trata de verificar que, efectivamente, es posible pronosticar de forma eficaz la temperatura. A continuación la siguiente tabla recoge la calidad del modelo en función de la base de datos correspondiente a las tres distintas temperaturas usadas:

Temperatura	Tiempo ejecución	MSE
Inlet	44 min 47 seg	0,001805
Outlet	1 h 3 min 0 seg	0,001020
Return	1 min 45 seg	0,000701

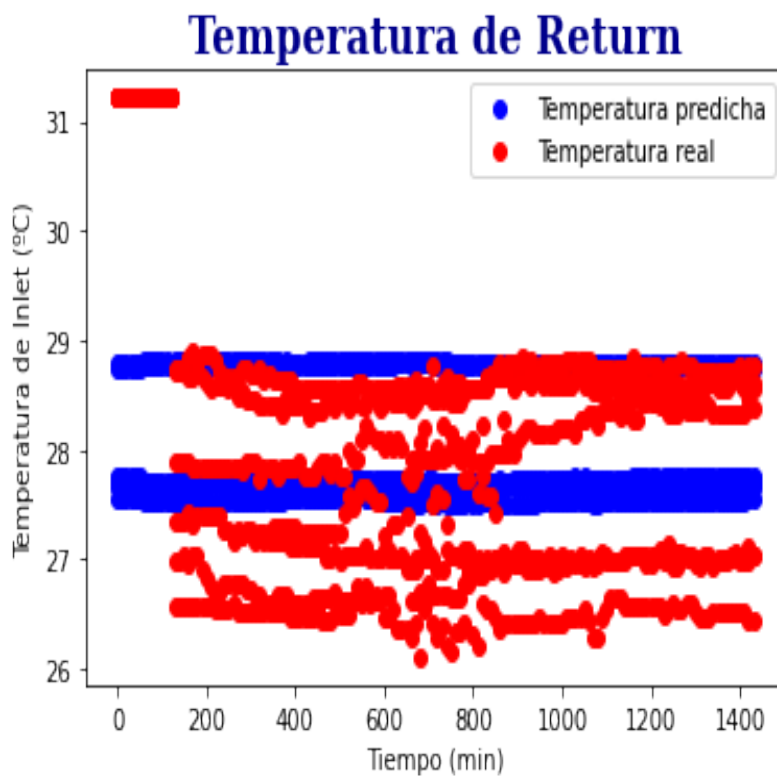
Tabla 2: Resumen resultados del experimento global.



(a) Predicción temperatura Inlet.



(b) Predicción temperatura Outlet.



(c) Predicción temperatura Return.

Figura 22: Predicciones realizadas por el modelo en experimento Global.

Temperatura	Instancias entrenamiento	Instancias test
Inlet	178.595	88.269
Outlet	493.096	243.143
Return	14.093	6.822

Tabla 3: Tamaño bases de datos de entrenamiento.

Los resultados son claramente mejores en el entrenamiento y predicción de temperatura *Outlet*, ya que el número de instancias es bastante mayor, lo que permite a la red encontrar patrones con mayor precisión en los datos de entrada. No es malo tampoco el error cuadrático medio correspondiente al finalizar el proceso con la temperatura *Inlet*, a pesar de disponer menos datos de este tipo para entrenar el modelo.

Tras analizar lo anterior, se puede ver un error bastante bajo tanto entre la temperatura real y la predicha, tanto en el caso de la temperatura *Inlet*, como en la de *Outlet*, siendo estas dos temperaturas las clave del estudio. Se puede observar, además, en ambas gráficas que la temperatura, efectivamente, puede ser predicha con 10 minutos de antelación, permitiendo así la capacidad de aplicarlo al sistema de refrigeración (en este caso a través de los CRAC). Así, la potencia consumida por este proceso se vería reducida, pudiendo adaptarla a la demanda. La duración de la ejecución del entrenamiento es mayor en la temperatura *Outlet*, aunque tampoco es exageradamente alto.

Sin embargo, el éxito no es tal en la predicción de la temperatura de *Return*, a pesar de que el MSE es muy bajo. En la gráfica se ve que la red ha conseguido situar dos claros rangos temperaturas que consiguen reducir el error, mas la predicción de temperatura no es buena. Esto es debido a que se dispone de una cantidad de datos muy reducida de este tipo de temperatura, hecho que da lugar a un tiempo de ejecución muy inferior a los demás.

4.2. EXPERIMENTO REDUCIENDO MEDICIONES

La batería de los sensores de los centros de datos es limitada y dura varios años tomando medidas. Cada vez que tiene lugar una medición de la temperatura actual, humedad y posición del sensor, se produce un gasto de la duración de la batería. Consiguientemente, cuanto mayor es el número de mediciones tomadas y enviadas a la base de datos, mayor es el desgaste.

El fin de este experimento es examinar la mínima cantidad de instancias con la que un modelo de aprendizaje profundo es capaz de predecir satisfactoriamente una futura temperatura. Al decrementar la frecuencia con la que se miden los parámetros, aumentaría la duración de la vida de la

batería de estos sensores. Por tanto, la posible reducción de la frecuencia de medición supondría un considerable ahorro económico para los propietarios de los centros que dispongan de sistemas similares de aprendizaje automático, que podrían gozar del trabajo de estos sensores durante un mayor periodo de tiempo.

El desarrollo de esta fase consiste en la introducción en la red de distintos porcentajes de datos pertenecientes al *dataframe* global. Los datos seleccionados para el entrenamiento llevado a cabo con cada porcentaje son completamente aleatorios. Se tanteará la calidad del modelo entrenado con el 90 %, 75 %, 66 % o el 50 % de los datos. Para testear la predicción se emplearán los conjuntos reservados para test.

No se ha llevado a cabo este estudio con la base de datos que contiene medidas de temperatura de *Return*. Nótese que en la sección anterior, los resultados con el número total de datos no ofrecía muy buenos resultados (figura 22c), por lo que no interesa reducir aún más las instancias de entrenamiento. Por tanto, se contemplará únicamente el desarrollo seguido para las otras dos clases de temperatura.

Temperatura	% datos	Tiempo ejecución	MSE
Inlet	90	10 min 16 seg	0,002044
	75	9 min 8 seg	0,002113
	66	7 min 6 seg	0,001774
	50	4 min 16 seg	0,001794
Outlet	90	51 min 46 seg	0,001115
	75	42 min 15 seg	0,000971
	66	35 min 26 seg	0,001012
	50	33 min 54 seg	0,001027

Tabla 4: Resultados entrenamiento reduciendo mediciones

Analizando los resultados mostrados en la tabla 4 y las gráficas contenidas en la figura 23, se extraen las siguientes afirmaciones para la temperatura de *Inlet*:

- Claramente es posible reducir el conjunto total de mediciones disponibles de esta clase. Al utilizar el 90 % y el 75 % de los datos, se ve que el MSE resultante crece ligeramente respecto al experimento global. Sin embargo, aporta unos resultados aceptables, la precisión de la predicción, visible en las figuras, aporta consistencia a esta afirmación.
- Se aprecia un descenso del error cuadrático medio al emplear para el entrenamiento de la red un 66 % del conjunto total, además, en la gráfica se ve que es capaz de predecir con gran

calidad y precisión.

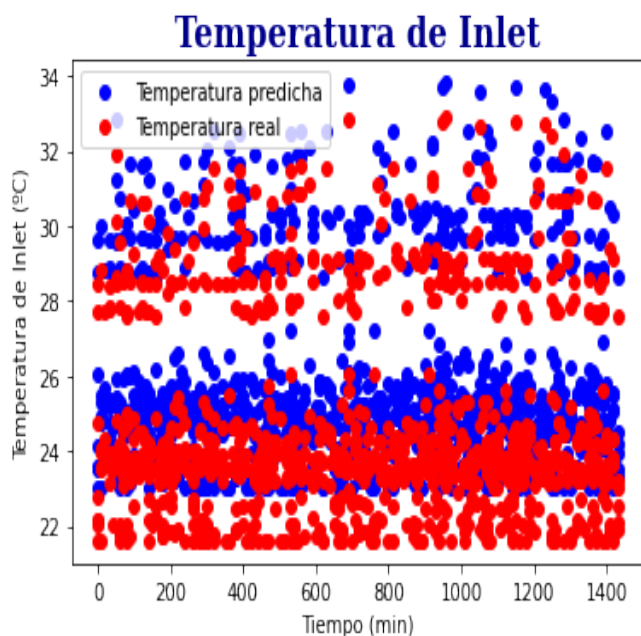
- Se observa como, en el entrenamiento con el 50 % de los datos, el error prácticamente se mantiene, mas en la gráfica, a simple vista, tiene menos exactitud que los anteriores, aunque tampoco es un decremento muy pronunciado de calidad.

Se continua ahora con los resultados del entrenamiento del modelo con las instancias correspondientes a la temperatura *Outlet*, reflejados en la tabla 4 y las gráficas de la figura 24:

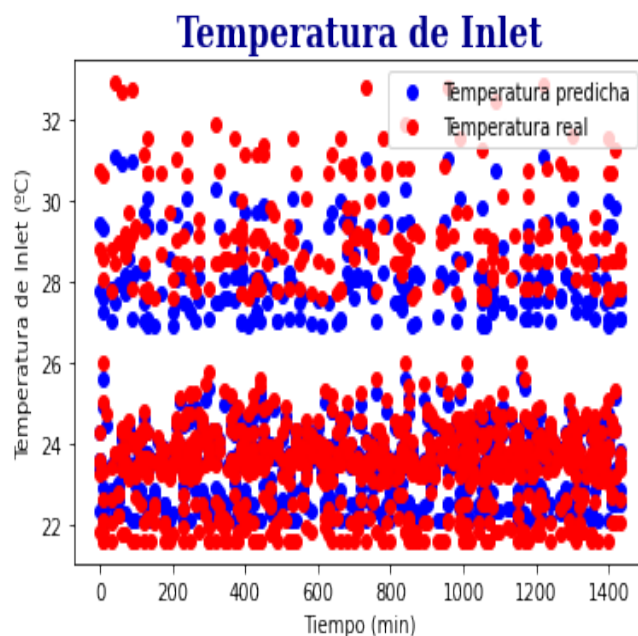
- Al disponer de un número notablemente mayor que en las otras dos clases, de temperatura, existe un mayor margen para eliminar instancias.
- El error cuadrático medio disminuye respecto al experimento global conforme se van quitando instancias.
- Las gráficas reflejan el gran acierto de las predicciones llevadas a cabo por el modelo entrenado.

En consecuencia, para ambos tipos de medición de temperatura, se puede contemplar la posibilidad una reducción cuantiosa de los datos introducidos en el entrenamiento de la ANN. Para medir *Inlet*, con la utilización del 66 % de la base de datos, se llega a los mejores resultados. Respecto a la temperatura *Outlet*, sería posible reducir hasta un 50 % las mediciones realizadas. Así, se reduciría la frecuencia de los mediciones tomadas, y se conseguiría un aumento importante de la vida útil de la batería de los sensores, dando lugar a un ahorro económico para las compañías.

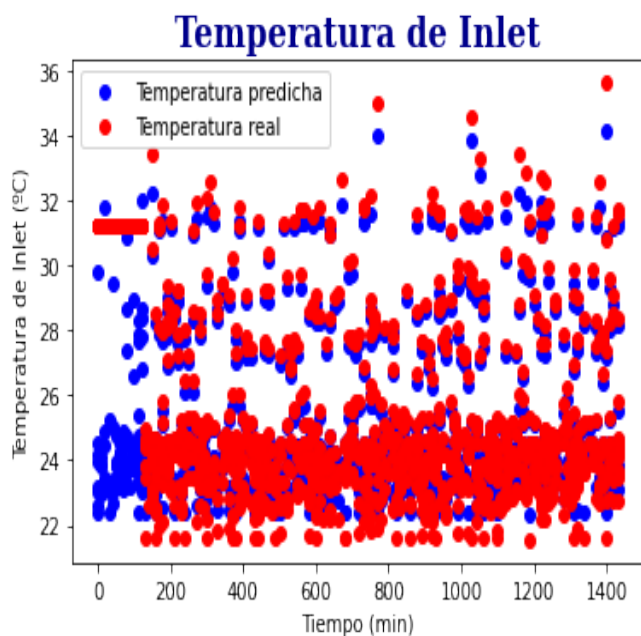
Con esta reducción de instancias, además, se alcanzaría un tiempo de ejecución menor del entrenamiento del modelo, lo cual es otra clara ventaja de la satisfactoria finalización de este experimento.



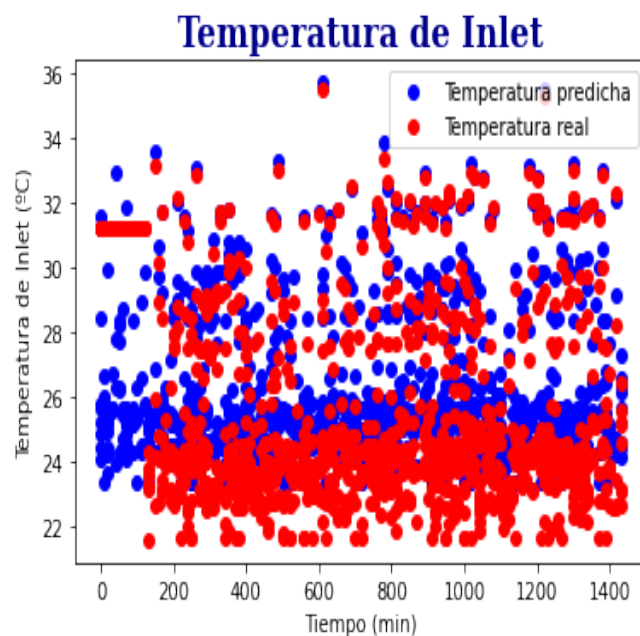
(a) Predicción con el 90 %.



(b) Predicción con el 75 %.

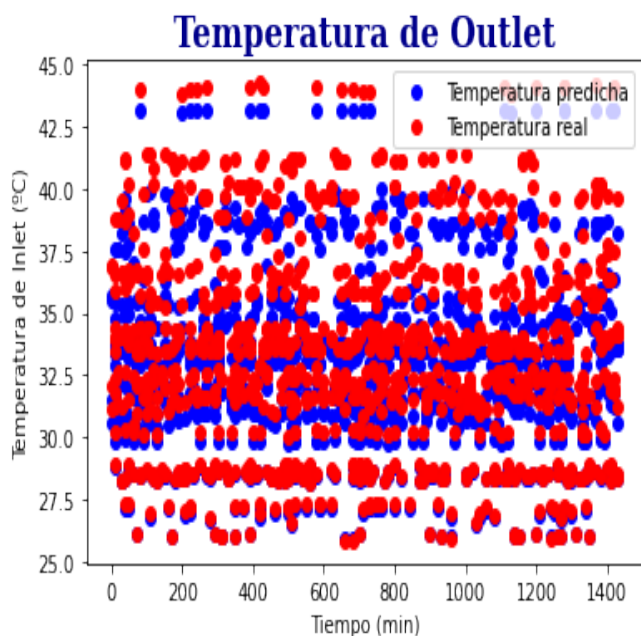


(c) Predicción con el 66 %.

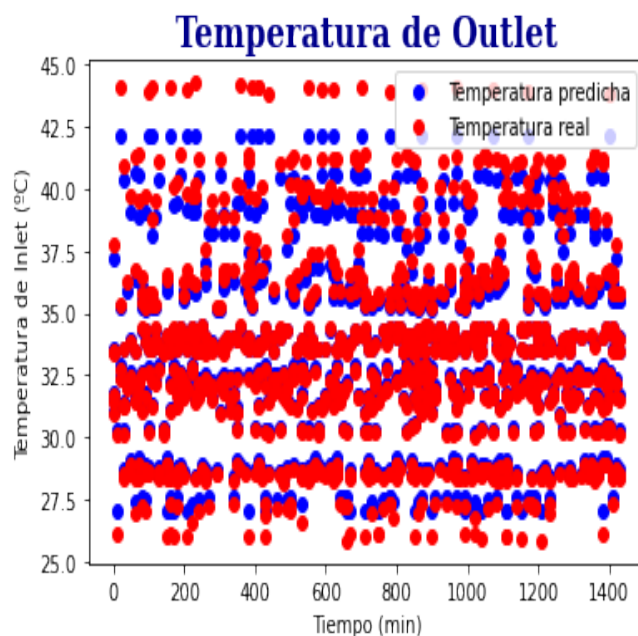


(d) Predicción con el 50 %..

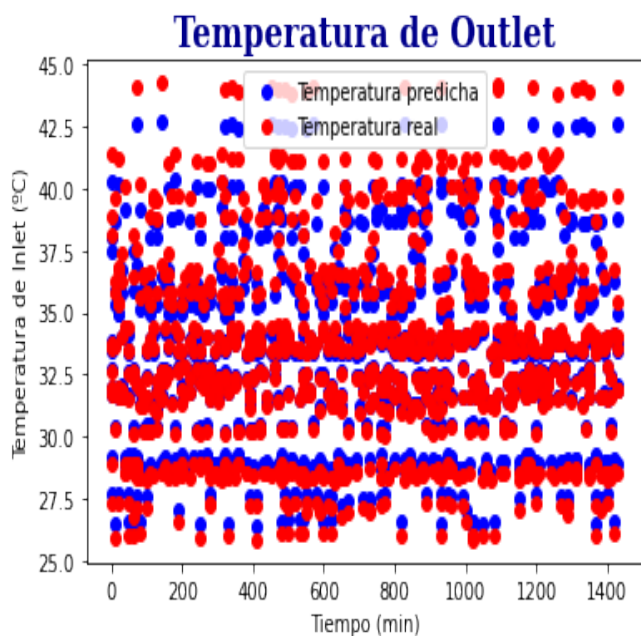
Figura 23: Predicciones de Inlet reduciendo la cantidad de mediciones.



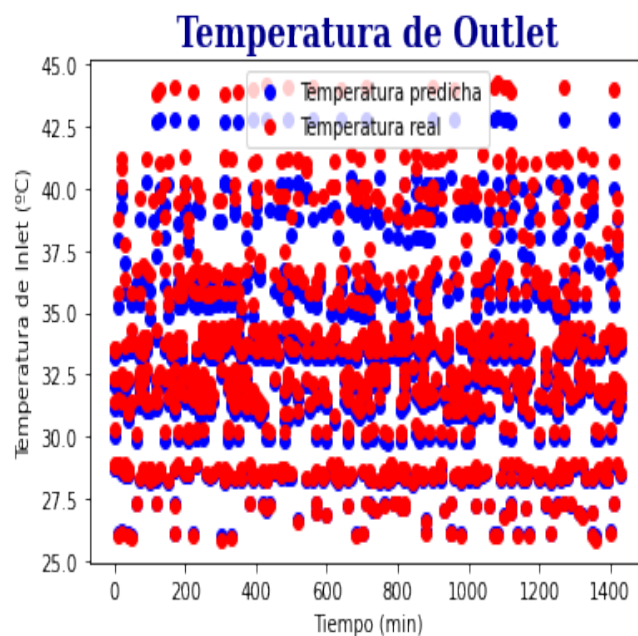
(a) Predicción con el 90 %.



(b) Predicción con el 75 %.



(c) Predicción con el 66 %.



(d) Predicción con el 50 %..

Figura 24: Predicciones de Outlet reduciendo la cantidad de mediciones.

4.3. EXPERIMENTO REDUCIENDO SENSORES

Otro punto remarcable que puede generar grandes ahorros económicos cuando se pongan en práctica soluciones basadas en *Deep Learning*, es el uso de un menor número de sensores. Para entornos de escasa superficie no supone un gran coste, sin embargo, para controlar la temperatura de un centro de datos de grandes extensiones, será necesario un gran esfuerzo económico para disponer los sensores a lo largo de toda la sala.

La intención de este apartado es hallar una distribución con un número menor de sensores que los dispuestos en el entorno estudiado, con el que resulten unas predicciones igual de fiables. Esto supondría un ahorro, tanto para el momento en el que se adquieran todos los dispositivos, como para el mantenimiento y recambio de los mismos.

Se establece una distribución uniforme por racks, para disponer de una cobertura más o menos constante del pasillo. Para ello, en el primer rack de la fila, se mantienen los sensores, mas en el contiguo se eliminan, en el siguiente permanecerán los sensores, y así sucesivamente. La figura 25 muestra una clara representación de ello :

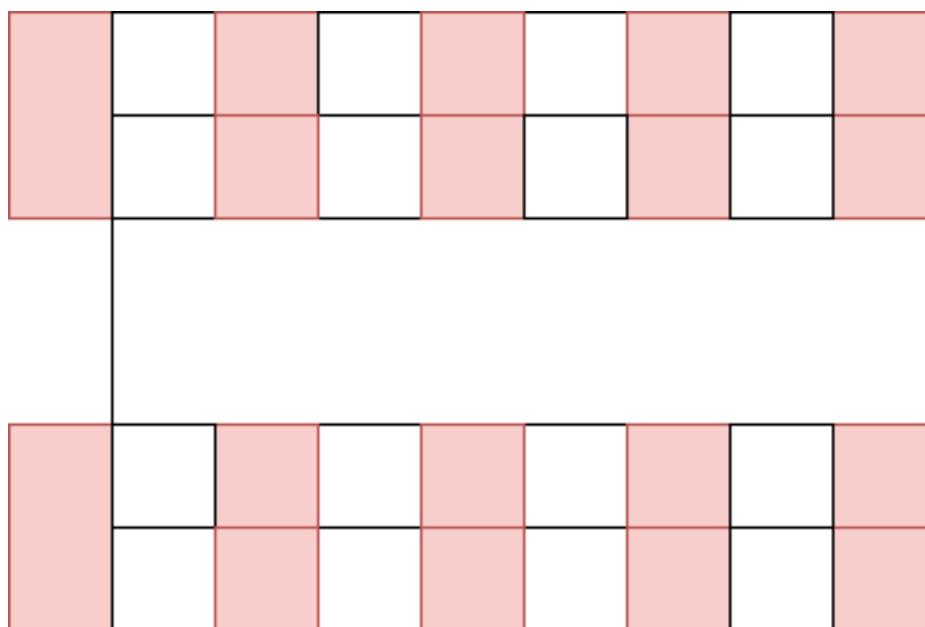


Figura 25: Distribución de sensores para el estudio.

Una vez seleccionados las instancias pertenecientes a los racks elegidos, se pasa al entrenamiento del modelo con los dos distintos conjuntos de datos (correspondientes a las temperaturas Inlet y Outlet):

Temperatura	Tiempo ejecución	MSE
Inlet	3 min 1 seg	0,001787
Outlet	38 min 30 seg	0,001031

Tabla 5: Resumen resultados del experimento de reducción de sensores.

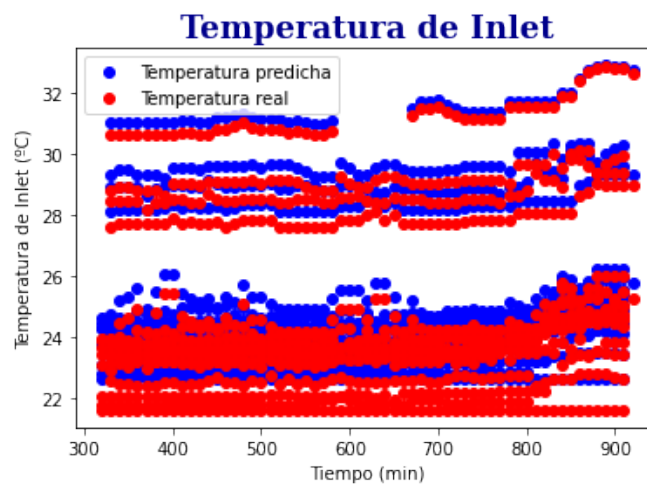


Figura 26: Predicciones temperatura Inlet con distribución uniforme de sensores.

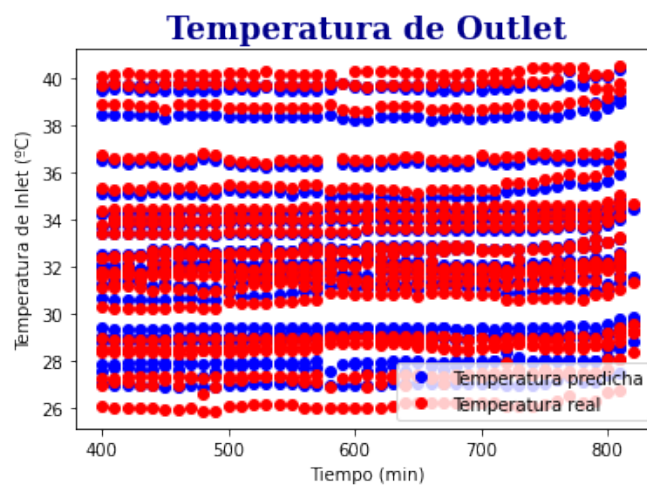


Figura 27: Predicciones temperatura Outlet con distribución uniforme de sensores.

Como se puede ver, los resultados son satisfactorios, demostrando que es completamente posible reducir el número de sensores colocados, siempre y cuando se establezca una distribución coherente, que cubra de forma lo mejor posible la longitud de los pasillos. Se llega a valores del error cuadrático medio menores que en el experimento global, en ambos tipos de entrenamiento. Por tanto, al aplicar tales cambios en la distribución, se lograría un ahorro económico debido a la reducción de sensores utilizados.

5. CONCLUSIONES Y LÍNEAS FUTURAS

5.1. CONCLUSIONES

La cada vez más creciente digitalización de la sociedad causa la creación masiva de datos, que deben ser procesados y almacenados, generando una dependencia muy grande de los centros de datos. Esto se traduce en un uso intenso de los mismos las 24 horas del día, cada día del año, y por lo tanto, en un consumo de grandes cantidades de energía, que son perjudiciales para el medio ambiente, a la par que un coste económico enorme. Esta situación ha dado lugar a que las emisiones de CO_2 correspondientes al sector TIC (en el que se incluyen los centros de datos) supongan el 2% de las totales actualmente.

Debido a esto, surgen tecnologías destinadas a mitigar el problema. En este trabajo se ha propuesto una posible solución. Tras el desarrollo de este Trabajo Fin de Grado, respecto a los objetivos iniciales marcados en la sección 1.2, se llega a las siguientes conclusiones:

- Se ha realizado un estudio extenso sobre los centros de datos, llegando a conocer la situación actual del sector, así como la problemática existente en relación con el medio ambiente. Se ha comprendido el funcionamiento y la estructura de los mismos, y del sistema de refrigeración, punto principal para este proyecto. Además, se ha logrado un entendimiento de las tecnologías novedosas aplicadas a las necesidades expuestas, siendo esto útil para inspirar la solución desarrollada.
- Se ha llevado a cabo un análisis sobre del aprendizaje automático y las redes neuronales, estrictamente necesario para poder desarrollar la solución. Se ha adquirido también, un buen dominio del lenguaje de programación Python, junto a la herramienta *Keras* y *TensorFlow*.
- Tras la finalización de este Trabajo Fin de Grado se comprueba que el modelo obtenido genera muy buenos resultados, y predice la temperatura de forma fiable, con un error cuadrático medio en torno al 0,002. Tanto para temperatura *Inlet*, como para *Outlet*, el entrenamiento no supone un tiempo de ejecución muy alto. Tras este entrenamiento, ambos modelos alcanzan gran precisión a la hora de realizar predicciones. El trabajo realizado, a pesar de tener margen de mejora, podría ser empleado para el control, a lo largo del tiempo, de sistemas de refrigeración como los CRAC, aumentando o reduciendo la potencia suministrada a los mismos. Gracias a la predicción realizada por el modelo, se podría adaptar la temperatura

de la sala, de forma automática, ya que esta es variable en función a condiciones externas como la ocupación de los servidores, la humedad y la temperatura actual.

- Se concluye mediante el estudio del número de instancias introducidas en la red, que se podría llegar a incrementar la duración de vida útil de los sensores, ya que se proporcionan evidencias de que es posible utilizar un número reducido de instancias para realizar el entrenamiento de las redes neuronales, lo que generará un ahorro en los costes de los sensores respecto a la utilización actual.
- Se halla una distribución uniforme con menor número de sensores colocados en la sala, que da unos resultados similares en la predicción de la temperatura, a los obtenidos usando para el entrenamiento la totalidad de sensores iniciales.

5.2. LÍNEAS FUTURAS

El siguiente paso a seguir de este proyecto es la puesta en práctica del modelo desarrollado en un centro de datos, con el objetivo de comprobar que la temperatura se predice de forma correcta. Es necesario verificar que no comete errores graves que supongan un decremento del rendimiento de los servidores o su sobrecalentamiento. Además, sería interesante calcular el ahorro energético y el PUE resultantes tras la aplicación del sistema.

Como mejora se podría estudiar la posibilidad introducir la posibilidad al usuario de predecir la temperatura en el momento que se desee (con ciertos límites). También sería interesante mejorar la predicción de la denominada temperatura de *Return*, para lo que sería indispensable recopilar un mayor número de mediciones de esta clase.

6. BIBLIOGRAFÍA

Referencias

- [1] Keras. <https://keras.io/>.
- [2] Perceptrón multicapa. https://es.wikipedia.org/wiki/Perceptr%C3%B3n_multicapa.
- [3] Regresión lineal. https://es.wikipedia.org/wiki/Regresi%C3%B3n_lineal.
- [4] Sistemas de aire acondicionado de precisión para salas de cómputo y comunicaciones. <https://www.mundohvacr.com.mx/2009/04/sistemas-de-aire-acondicionado-de-precision-para-salas-de-computo-y-comunicaciones/>.
- [5] tf.keras.model. https://www.tensorflow.org/api_docs/python/tf/keras/Model.
- [6] ¿qué es colab? https://colab.research.google.com/notebooks/welcome.ipynb?hl=es#scrollTo=5fCEDCU_qrC0.
- [7] APPLE. Environmental progress report. https://www.apple.com/environment/pdf/Apple_Environmental_Progress_Report_2021.pdf.
- [8] BERNADREFRIGERACIÓN. Todo lo que necesitas saber sobre la refrigeración de los centros de datos. <https://www.josebernad.com/todo-lo-que-necesitas-saber-sobre-la-refrigeracion-de-los-centros-de-datos/>.
- [9] CHOLLET, F. Clasificación básica: Predecir una imagen de moda. <https://www.tensorflow.org/tutorials/keras/classification>.
- [10] DELGADO, J. Los secretos del 'búnker' tecnológico de mercadona en la España vaciada. https://www.vozpopuli.com/economia_y_finanzas/mercadona-centro-leon.html.
- [11] DURÁN, J. Everything you need to know about gradient descent applied to neural networks. <https://medium.com/yottabytes/everything-you-need-to-know-about-gradient-descent-applied-to-neural-networks-d70f85e0>.
- [12] FOXWELL, D. Underwater data centres could be powered by offshore wind, Microsoft says. <https://www.rivieramm.com/news-content-hub/news-content-hub/underwater-data-centres-could-be-powered-by-offshore-wind-microsoft-says-60917>.
- [13] GAO, J. Machine learning applications for data center optimization. <https://research.google/pubs/pub42542/>.

- [14] GOOGLE. Centros de datos. <https://www.google.com/about/datacenters/efficiency/#water-and-cooling>.
- [15] GÉRON, A. Hands-on machine learning with scikit-learn, keras, and tensorflow.
- [16] HERAS, J. M. Gradiente descendiente para aprendizaje automático. <https://www.iartificial.net/gradiente-descendiente-para-aprendizaje-automatiko/>.
- [17] JONES, N. How to stop data centres from gobbling up the world's electricity. <https://www.nature.com/articles/d41586-018-06610-y>.
- [18] JOSEP BERRAL, ÍÑIGO GOIRI, R. N. F. J. J. G. R. G., AND TORRES, J. Towards energy-aware scheduling in data centers using machine learning, 2010.
- [19] JULIANNA DELUA, IBM, A. Supervised vs. unsupervised learning: What's the difference? <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning#:~:text=The%20main%20difference%20between%20supervised,unsupervised%20learning%20algorithm%20does%20not>.
- [20] JULIÁN, G. Las redes neuronales: qué son y por qué están volviendo. <https://www.xataka.com/robotica-e-ia/las-redes-neuronales-que-son-y-por-que-estan-volviendo/>.
- [21] KERAS. Metrics. <https://keras.io/api/metrics/#:~:text=A%20metric%20is%20a%20function,loss%20function%20as%20a%20metric>.
- [22] MICROSOFT. Proyecto natick, el futuro de los centros de datos bajo el mar es fiable, práctico y sostenible. <https://news.microsoft.com/es-es/2020/09/15/proyecto-natick-el-futuro-de-los-centros-de-datos-bajo-el-mar-es-fiable-practico-y-so>.
- [23] NAKAMURA, M. Learning and optimization models for energy efficient cooling control in data center. <https://ieeexplore.ieee.org/document/7749090>, 2016.
- [24] PAOLO BERTOLDI, MARIA AVGERINOU, L. C. Trends in data centre energy consumption under the european code of conduct for data centre energy efficiency. <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC108354/kjna28874enn.pdf>.
- [25] RAJENDRA KUMAR, SUNIL KUMAR KHATRI, M. J. D. Effect of cooling systems on the energy efficiency of data centers: Machine learning optimisation. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9200088>.

- [26] RICH EVANS, J. G. Deepmind ai reduces energy used for cooling google data centers by 40 %, 2016.
- [27] SALAS, R. Redes neuronales artificiales. https://d1wqtxts1xzle7.cloudfront.net/50358783/Redes_Neuronales_Artificiales.pdf?1479332205=&response-content-disposition=inline%3B+filename%3DRedes_Neuronales_Artificiales.pdf&Expires=1619196991&Signature=KBE5cZTBr sdmcoevqwQP xG5ePfYg-I~EDewpB4wSsod1e5Ee8ij0G5xgJjwjcElirhk4PIV8nqqY8Rn5fpDWRtI _&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA.
- [28] YIN ZHANG, ZHIYUAN WEI, M. Z. Free cooling technologies for data centers: energy saving mechanism and applications. <https://www.sciencedirect.com/science/article/pii/S1876610217364676>.
- [29] YUYA TARUTANI, KAZUYUKI HASHIMOTO, G. H. Y. N. T. T. K. M. M. M. Reducing power consumption in data center by predicting temperature distribution and air conditioner efficiency with machine learning. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7484193>.

7. ANEXO A: ASPECTOS ÉTICOS, ECONÓMICOS, SOCIALES

7.1. INTRODUCCIÓN

Hoy en día, la búsqueda de nuevas aplicaciones del aprendizaje automático está a la orden del día, lo cual tiene mucha repercusión. El *machine learning* trata de simular el aprendizaje humano, y si estos sistemas siguen avanzando pueden llegar a ser autónomos en algún momento, lo cual genera un impacto grande y, en ocasiones, dudas en la sociedad. Sin embargo, la aplicación de estos sistemas puede ser beneficiosa, dando la posibilidad de completar tareas inimaginables en el pasado, e incluso trabajar ayudando al ser humano a realizarlas. En concreto, en el caso de este Trabajo Fin de Grado, la predicción de temperatura en centros de datos a través de redes neuronales tendrá una serie de impactos importantes, que serán tratados en el subapartado posterior.

7.2. DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PROYECTO

Impacto medioambiental

El impacto más notable de este proyecto es de esta tipología debido a que trata de reducir la energía consumida por los centros de datos. Trabajos como este, deben ayudar a alimentar la iniciativa de los propietarios a optimizar el PUE, y por tanto la huella de carbono producida por los centros de datos, a través del análisis de datos. Además, para el estudio realizado es únicamente necesario un ordenador portátil y una serie de sensores, el resto es software, lo cual no genera gastos energéticos elevados, ni emisiones, ni residuos, por lo que se trata de un desarrollo sostenible.

Impacto ético

El respeto hacia el medioambiente es un valor ético que todo ser humano debe poseer. Es una obligación respetar la naturaleza y poner en práctica las mejores acciones para contribuir lo menos posible ante el calentamiento global y tratar de reducirlo. Por ello, este trabajo posee un impacto ético, debido a que su mayor finalidad es la de reducir la huella de carbono generada por los centros de datos, y por tanto, trata de reducir el calentamiento global. Además, el estudio realizado en este trabajo se ha llevado a cabo sin incurrir en la privacidad de las empresas, ni individuos.

Impacto económico

En el plano económico, este proyecto también supone un impacto bastante importante. Las empresas y propietarios de centros de datos gastan grandes cantidades de dinero en su alimentación energética, derrochando un porcentaje de esta en una refrigeración incorrecta. Por tanto, la aplicación de un sistema de predicción que gobierne el funcionamiento de la refrigeración para adaptarse al momento, generará grandes ahorros económicos. Además, el coste de este tipo de técnicas no es tan elevado, ya que no necesita grandes despliegues de infraestructura, ni grandes costes adicionales.

Impacto social

El impacto social de este proyecto reside sobre todo en la mejora de la calidad de vida de la sociedad producida por la reducción del calentamiento global que tendría lugar si se comienza a extender la implantación de sistemas de este tipo. Conforme vayan pasando los años venideros, debido al continuo crecimiento de los datos, mayor importancia tendrán los centros de datos y mayor carga de trabajo, por lo que la energía consumida y la huella de carbono aumentarán, si no se toman medidas. Esta posible situación concluiría en un daño irreparable en la sociedad.

7.3. ANÁLISIS DETALLADO DE ALGUNO DE LOS PRINCIPALES IMPACTOS: MEDIOAMBIENTAL Y ECONÓMICO

Los centros de datos suponen un 2 % del consumo energético global, por lo que el sector TIC debe esforzarse en trabajar de forma sostenible y hay una continua búsqueda de formas de lograrlo de la manera más barata posible. Sistemas como el diseñado, gozan de un desarrollo que no supone altas cantidades económicas y que generan el efecto que el sector busca. Por ello, el impacto de esto es grande, ya que se está comprobando que es estrictamente necesario.

En proyectos similares se ha conseguido reducir la potencia consumida hasta un 40 %, lo cual comprueba que la predicción de la temperatura para controlar la refrigeración en las salas de servidores es una técnica óptima para mejorar los sistemas. Se ha logrado mediante la aplicación de redes neuronales, en concreto, reducir el PUE hasta un valor de 1,1, lo cual invita a pensar que un sistema como el desarrollado, puede reducir la energía consumida de forma considerable ahorrando miles de euros al sector y reduciendo la huella de carbono.

Además, el estudio acerca de los sensores y las mediciones realizadas por los mismos, verifican que el gasto en ellos no es muy elevado, ya que no son muy caros y realmente, en salas de servidores

pequeñas como la estudiada, se puede reducir el número de sensores dispuestos.

7.4. CONCLUSIONES

El aprendizaje automático es una herramienta muy potente que va a ser clave para la completa transformación digital de la sociedad, la aplicación de esta a resolver problemas de todo tipo va a cambiar la forma de trabajar y de vivir de las personas. Optimizar la refrigeración de los centros de datos claramente traerá grandes beneficios, ayudando sobre todo al medioambiente, con un proceso de desarrollo del sistema de forma sostenible, teniendo grandes efectos en la economía.

8. ANEXO B: PRESUPUESTO ECONÓMICO

		horas	Precio/hora	TOTAL
COSTE DE MANO DE OBRA (coste directo)		300	50 €	15.000 €
COSTE DE RECURSOS MATERIALES (coste directo)				
	Precio de compra	Uso en meses	Amortiz. en años	TOTAL
Ordenador personal (Software incluido)	1.500,00 €	2,28	5	57,00 €
TOTAL				57,00 €
GASTOS GENERALES (costes indirectos)	15%	sobre CD		2.258,55 €
BENEFICIO INDUSTRIAL	30%	sobre CD+CI		5.194,67 €
SUBTOTAL PRESUPUESTO				22.510,22 €
IVA APLICABLE			21%	4.727,15 €
TOTAL PRESUPUESTO				27.237,36 €

9. ANEXO C: CÓDIGO SOFTWARE

Código disponible en GitHub:

<https://github.com/castejon777/ANNTemperaturePrediction>