

# Análisis del consumo de la energía de un edificio mediante técnicas predictivas de *Deep Learning*

**Profesor:** *Xavier Baró Solé*

**Tutor:** *David Isern Alarcón*

**Ángel García de la Chica Herrera**

Grado de Ingeniería Informática

**Inteligencia Artificial**

# Índice

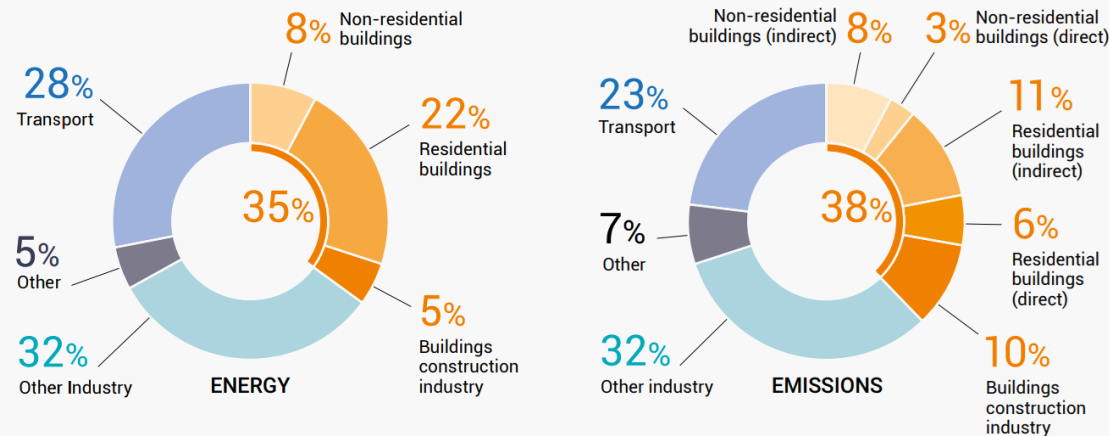
Índice	2
Introducción	3
Organización	4
Estado del arte	5
Presentación de los datos	6
Preprocesado de los datos 1	7
Preprocesado de los datos 2	8
Preprocesado de los datos 3	9
Preprocesado de los datos 4	10
Modelos	11
Modelo ANN Descripción	12
Modelo ANN Parámetros	13
Modelo ANN Implementaciones	14
Modelo Light Descripción	15
Modelo Light Parámetros	16
Modelo Light Implementaciones	17
Modelos un único edificio	18
Resumen de resultados	19
Conclusiones	20

# Introducción

## Contexto → Cambio climático

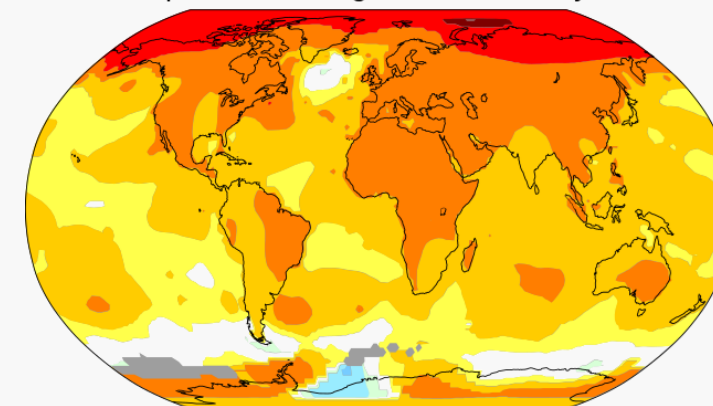
- Responsable: emisiones de gases de efecto invernadero
  - 35% de la energía total
  - 38 % de las emisiones de gases de efecto invernadero
- Solución: mejorar la eficiencia energética.
  - Normas, ayudas, sanciones por parte de los gobiernos y de la UE

Global share of buildings and construction final energy and emissions, 2019

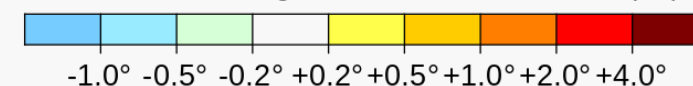


**Fuente:** 2020 GLOBAL STATUS REPORT FOR BUILDINGS AND CONSTRUCTION

Temperature change in the last 50 years



2011-2021 average vs 1956-1976 baseline (°C)



# Organización

## Objetivos Principales

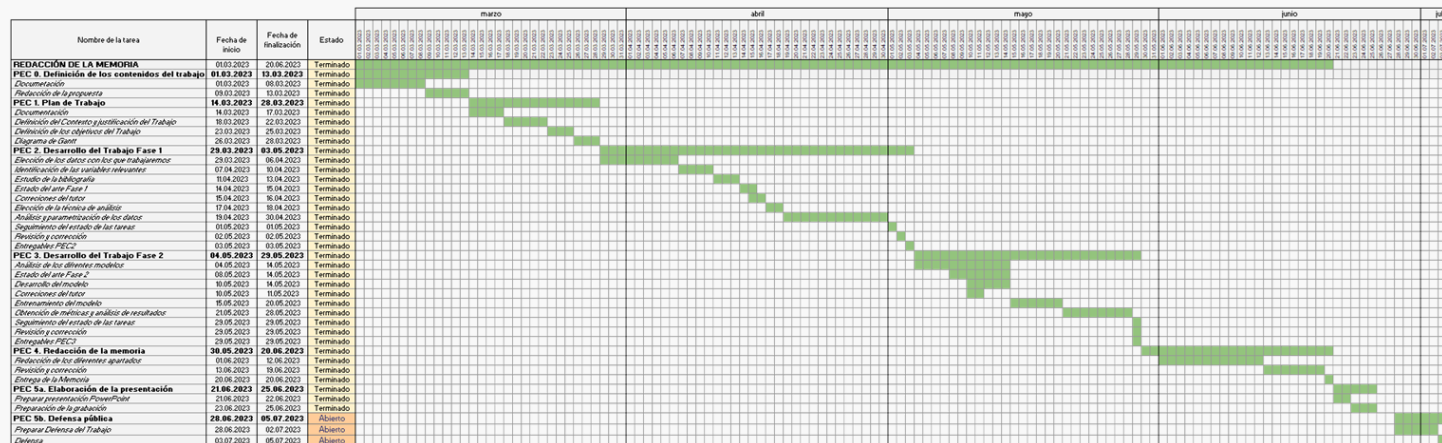
- Obtención de los datos.
- Identificación de las variables relevantes.
- Estudio de las diferentes técnicas de IA.

## Objetivos secundarios

- Desarrollar los modelos predictivos.
- Evaluar los resultados.

## Herramientas

- Microsoft Office 365
- Python (versión 3)
- Google COLAB Pro. Características:
  - Memoria: 25 GB de RAM y 166,8 GB de disco
  - Acelerador por Hardware GPU de 15 GB
- Disco de 166.8 GB
- Google Drive



# Estado del arte

## Enfoques

- Basado en reconocimiento de imágenes
- Factores de forma y materiales de los edificios
- Predicción de la etiqueta energética
- A partir de los datos históricos

## Modelos

Modelo	Número	Proporción
ANN	22	57 %
GPR	7	19 %
GMM	2	5 %
SVM	7	19 %

# Presentación de los datos

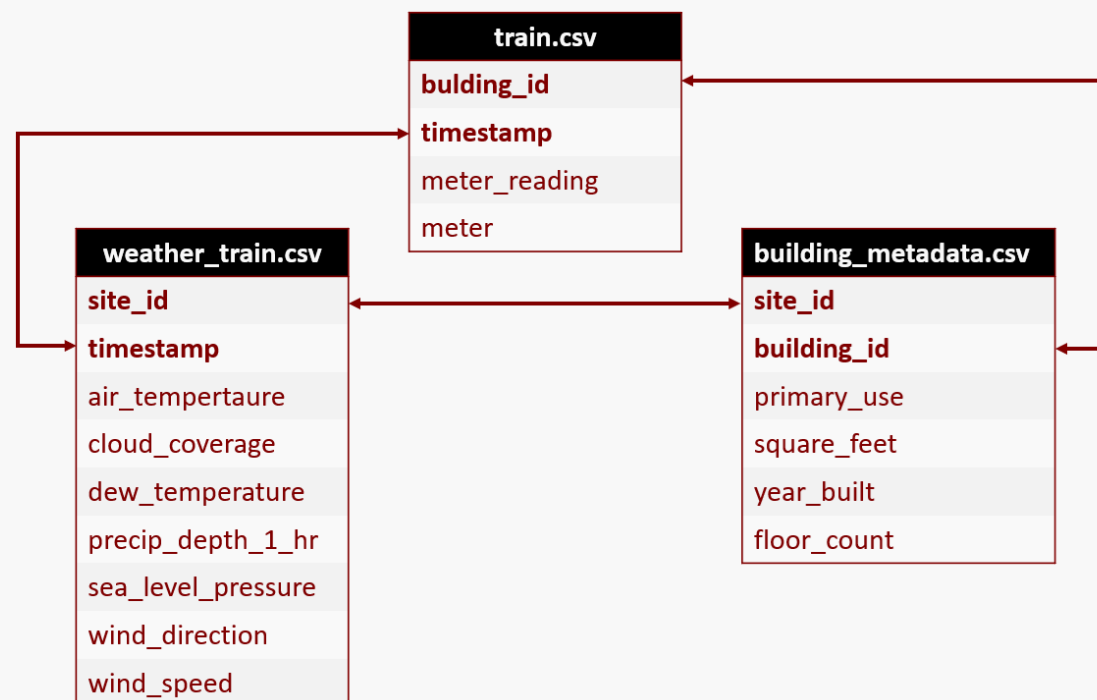
- Origen de los datos

ASHRAE: del inglés *American Society of Heating, Refrigerating and Air-Conditioning Engineers*

- Estructura de los datos



Shaping Tomorrow's Built Environment Today



# Preprocesado de los datos 1

## Limpieza de los datos

- Limpieza de los valores nulos o en blanco
  - Atributos con muchos nulos → se descartan
  - Atributos continuos → se sustituye por la media
  - Atributos discretos → se sustituye por la moda
- Limpieza de los valores atípicos

	Nulls (%)
building_id	0.000
timestamp	0.000
meter_reading	0.000

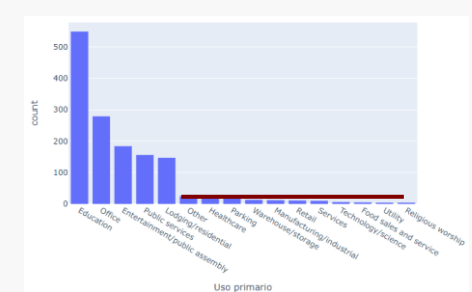
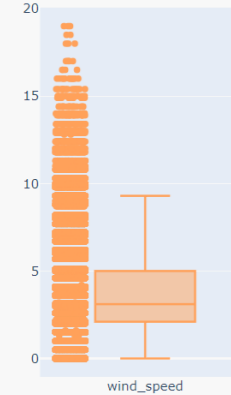
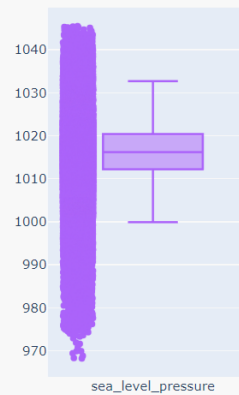
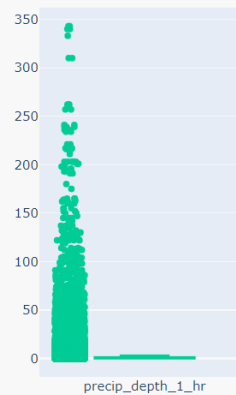
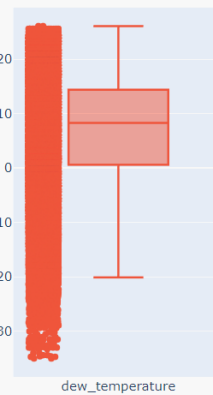
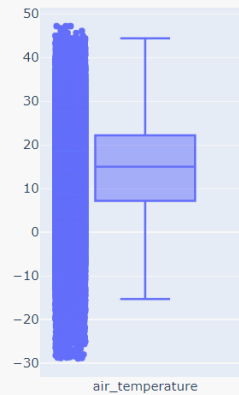
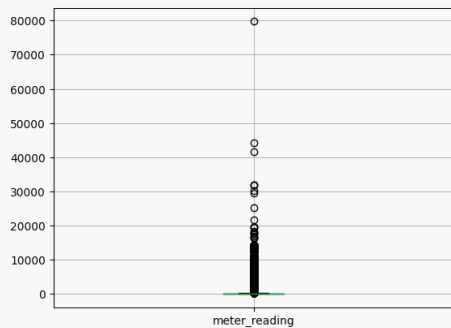
train.csv

	Null (%)
floor_count	75.500
year_built	53.416
site_id	0.000
building_id	0.000
primary_use	0.000
square_feet	0.000

building\_meta.csv

	Nulls (%)
cloud_coverage	49.490
precip_depth_1_hr	35.979
sea_level_pressure	7.597
wind_direction	4.484
wind_speed	0.217
dew_temperature	0.081
air_temperature	0.039
site_id	0.000
timestamp	0.000

weather\_train.csv



# Preprocesado de los datos 2

## Análisis de correlaciones entre los atributos

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\text{std}(x) \cdot \text{std}(Y)}$$

Donde:

$\rho_{x,y}$  es el coeficiente de correlación de Pearson

$\text{cov}(X,Y)$  es la covarianza de X e Y

$\text{std}(X)$  es la desviación típica de X

$\text{std}(Y)$  es la desviación típica de Y

	building_id	meter_reading	site_id	square_feet	air_temperature	cloud_coverage	dew_temperature	precip_depth_1_hr	sea_level_pressure	wind_direction	wind_speed
building_id	1.000000	0.004287	0.975536	0.086285	-0.282372	-0.280994	-0.125827	0.219620	0.037480	0.029676	0.000178
meter_reading	0.004287	1.000000	0.013694	0.564593	-0.002890	-0.044330	0.004957	0.050551	-0.009118	-0.002302	-0.022572
site_id	0.975536	0.013694	1.000000	0.094763	-0.266708	-0.266895	-0.145479	0.221214	0.013730	0.024033	-0.002046
square_feet	0.086285	0.564593	0.094763	1.000000	-0.005480	-0.061900	-0.025234	0.016122	-0.020227	-0.020058	-0.039734
air_temperature	-0.282372	-0.002890	-0.266708	-0.005480	1.000000	0.125875	0.757989	-0.077815	-0.279472	-0.104261	-0.081060
cloud_coverage	-0.280994	-0.044330	-0.266895	-0.061900	0.125875	1.000000	0.127123	-0.069448	-0.033959	0.006982	0.115901
dew_temperature	-0.125827	0.004957	-0.145479	-0.025234	0.757989	0.127123	1.000000	0.014468	-0.197973	-0.169826	-0.130879
precip_depth_1_hr	0.219620	0.050551	0.221214	0.016122	-0.077815	-0.069448	0.014468	1.000000	-0.033398	0.005704	0.031155
sea_level_pressure	0.037480	-0.009118	0.013730	-0.020227	-0.279472	-0.033959	-0.197973	-0.033398	1.000000	-0.095215	-0.197313
wind_direction	0.029676	-0.002302	0.024033	-0.020058	-0.104261	0.006982	-0.169826	0.005704	-0.095215	1.000000	0.411141
wind_speed	0.000178	-0.022572	-0.002046	-0.039734	-0.081060	0.115901	-0.130879	0.031155	-0.197313	0.411141	1.000000



# Preprocesado de los datos 3

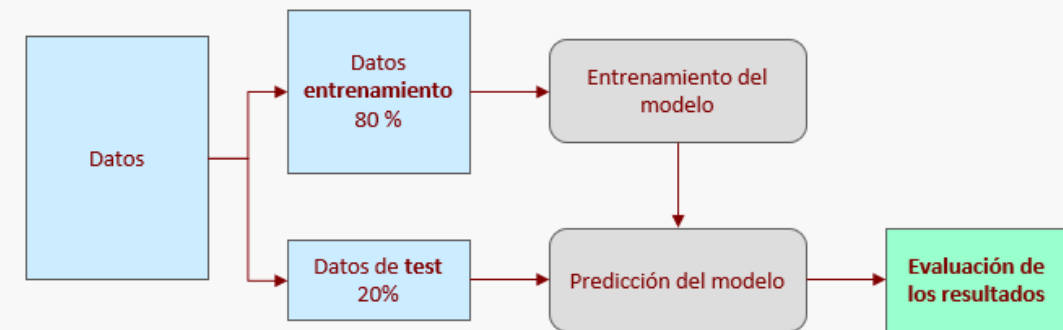
Codificación de las variables discretas

one-hot

PRIMARY_USE	PRIMARY_USE_ EDUCATION	PRIMARY_USE_ OFFICE	PRIMARY_USE_ ENTERTAINMENT/ PUBLIC_ASSEMBLY
EDUCATION	1	0	0
OFFICE	0	1	0
ENTERTAINMENT/PUBLIC ASSEMBLY	0	0	1

División del conjunto de datos

Train/Test



Normalización de los datos

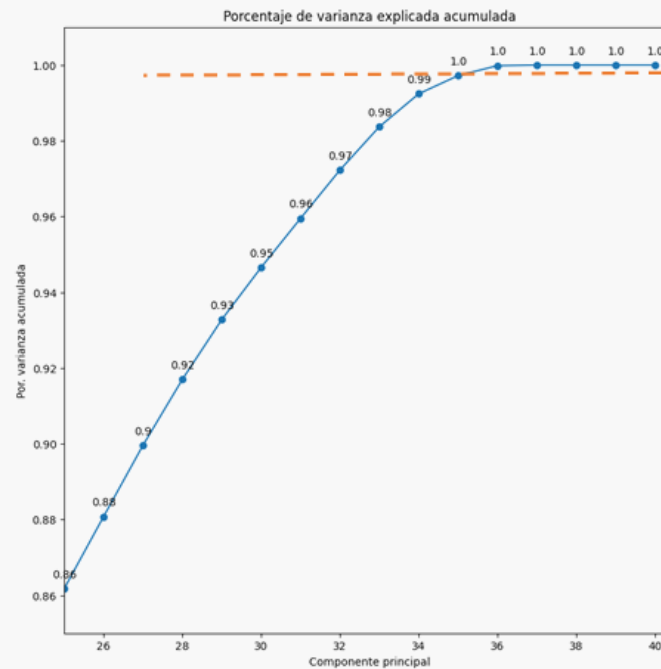
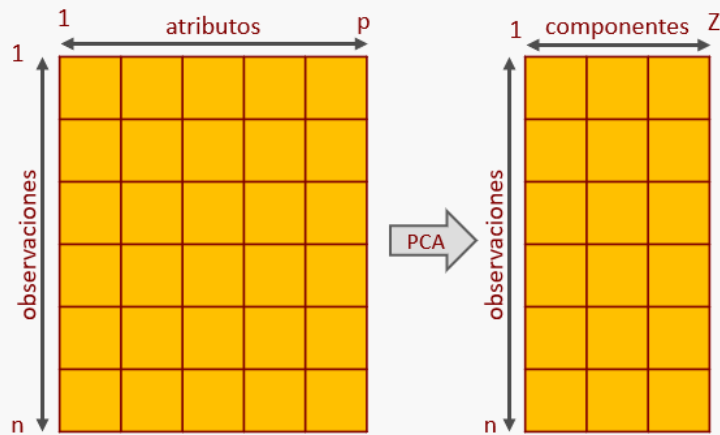
Estandarización

$$v' = \frac{(v - \text{media}(A(X)))}{\text{desviación}(A(X))}$$

Cada valor  $v$  asociado al atributo  $A$  para el objeto  $O$  se transforma en  $v'$ .

# Preprocesado de los datos 4

## Reducción de la dimensionalidad



Impacto en el uso de la memoria al reducir la dimensionalidad



# Modelos

## Evaluación de la calidad de los modelos

- Error Cuadrático medio

$$\text{ECM} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Donde:

**ECM:** Error Cuadrático Medio.

**n:** Número total de observaciones o datos.

**y:** Valor real de la variable dependiente en la i-ésima observación.

**$\hat{y}$ :** vector de n predicciones.

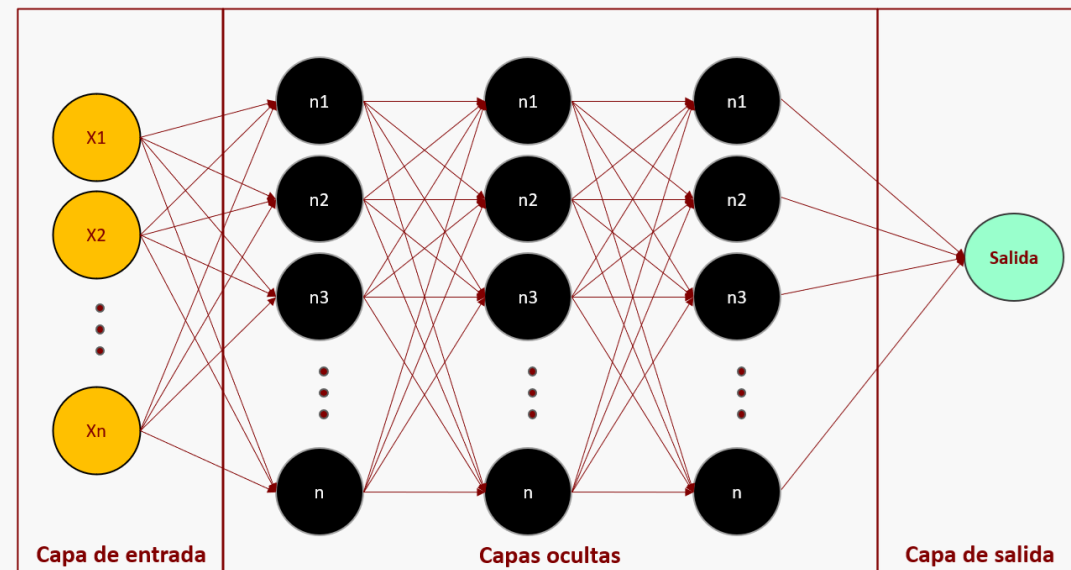
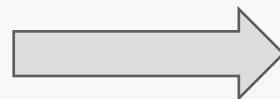
## Modelos entrenados

- Modelo ANN
- Modelo *LightGBM*

## Datos

- Todos los edificios sin reducir la dimensionalidad
- Todos los edificios reduciendo la dimensionalidad
- Un único edificio sin reducir la dimensionalidad
- Un único edificio reduciendo la dimensionalidad

# Modelo ANN. Introducción

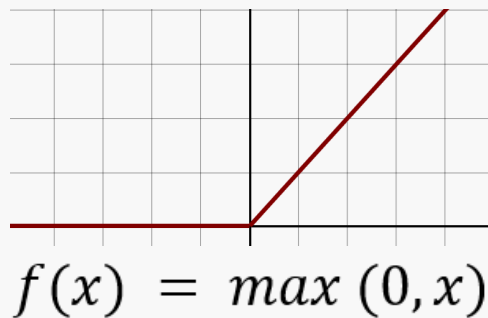


ANN (del inglés *Artificial Neural Network*) están inspiradas en la estructura y funcionamiento del cerebro humano. Estas redes se componen de múltiples capas de neuronas artificiales que procesan los datos de entrada de forma gradual y van refinando los resultados a medida que avanzan en las capas.

# Modelo ANN. Parámetros

## Constantes en todos los modelos implementados

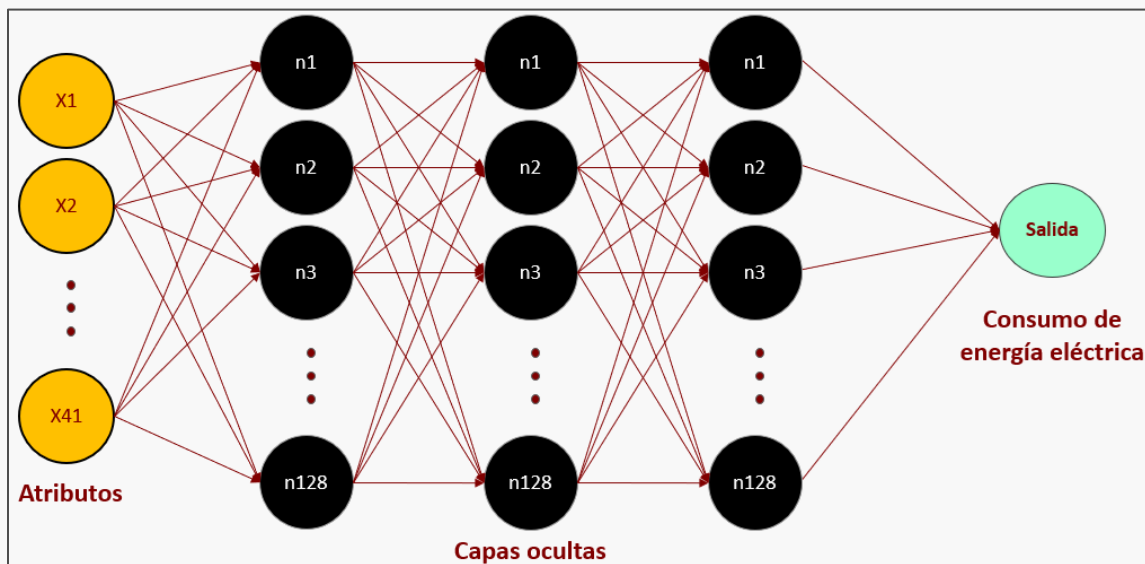
- Número de capas ocultas → 3 capas
- Optimizador → adam
- Tasa de aprendizaje → 0,01
- Función de pérdida → ECM (o mse)
- Función activación capas ocultas → ReLu



## Ajustes que se han realizado

- Tamaño de lote
- Épocas
- Número de neuronas
- Callbacks → *EarlyStopping (patience)*

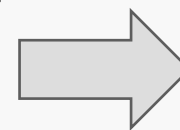
# Modelo ANN. Implementación



Esquema de la ANN

n1	n2	n3	batch_size	epochs	learning_rate	patience	ECM
64	32	16	256	100	0.05	∞	0.512
64	64	64	600	25	0.01	2	0.426
64	64	64	600	20	0.02	2	0.564
64	64	64	600	25	0.01	5	0.232
128	128	128	1024	25	0.01	5	0.360
128	128	128	1024	100	0.01	10	0.202

Ajustes



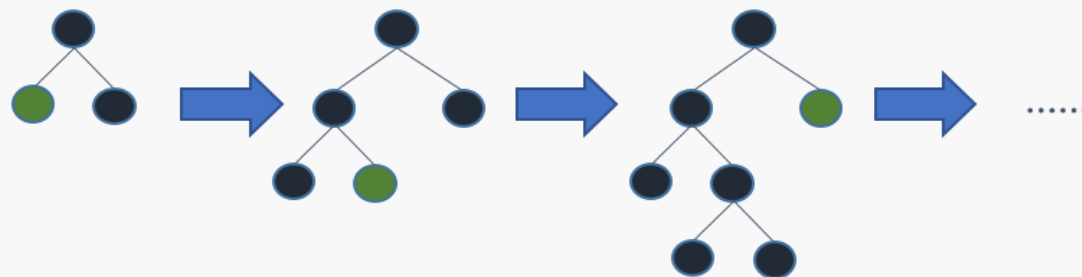
Modelo	ECM
Sin reducir la dimensionalidad	0.202
Reduciendo la dimensionalidad	0.242

Resultados

# Modelo *LightGBM*. Introducción



*LightGBM* (del inglés *Light Gradient-Boosting Machine*) es un algoritmo libre y de código abierto para el aprendizaje automático, desarrollado originalmente por Microsoft que se basa en algoritmos de árboles de decisión y se utiliza para la clasificación y otras tareas de aprendizaje automático.



**Fuente:** <https://lightgbm.readthedocs.io/>

# Modelo *LightGBM*. Parámetros

## Constantes en todos los modelos implementados

- `objective` → regression
- `boosting` → gbdt
- `metric` → ECM (o mse)
- `num_iterations` → 1500 iteraciones
- `early_stopping_round` → 50 iteraciones

## Ajustes que se han realizado

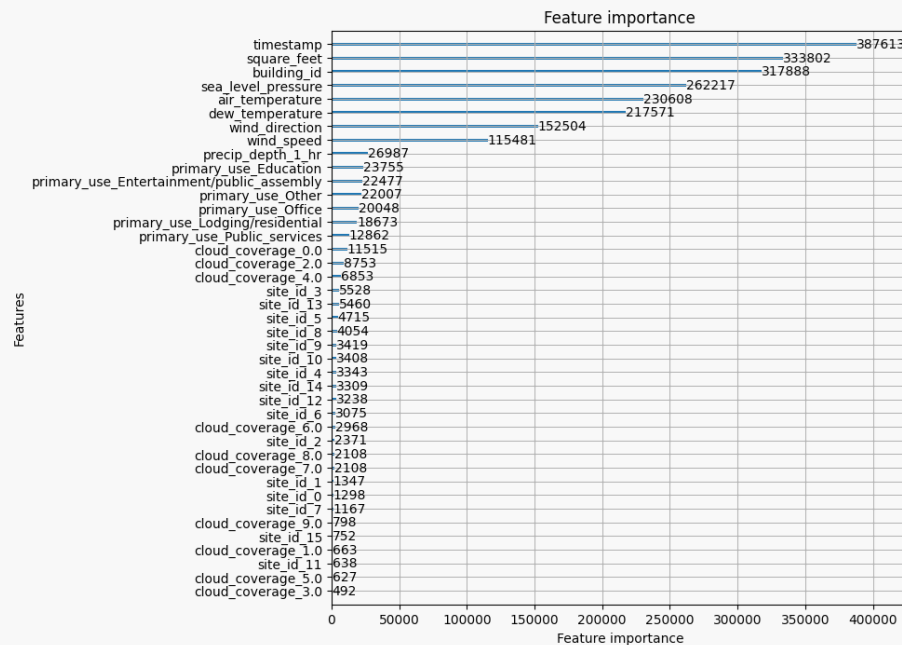
- `num_leaves`
- `learning_rate`
- `feature_fraction`
- `bagging_fraction`
- `max_depth`
- `reg_alpha`
- `reg_lambda`



# Modelo *LightGBM*. Implementación

<i>num_leaves</i>	<i>learning_rate</i>	<i>feature_fraction</i>	<i>bagging_fraction</i>	<i>max_depth</i>	<i>reg_lambda</i>	ECM
1500	0.05	0.85	1	-1	2	0.422
1500	0.1	0.8	0.8	-1	4	0.056
1500	0.15	0.9	0.9	-1	4	0.049
1500	0.2	0.8	0.8	-1	10	0.047
2500	0.25	0.9	0.9	12	10	0.041
4000	0.25	0.9	0.9	-1	10	0.042

Ajustes



Importancia de los atributos

Modelo

ECM

Sin reducir la dimensionalidad

0.042

Reduciendo la dimensionalidad

0.456

Resultados

# Modelos un solo edificio

## • Preprocesado de los datos de un edificio

- División del conjunto de los datos de test y entrenamiento
- Normalización → Estandarización
- Codificación de las variables discretas → *one-hot*
- Reducción de la dimensionalidad → PCA

Atributos → 13 a 11  
Memoria → -21%

## • ANN

n1	n2	n3	batch_size	epochs	learning_rate	patience
128	64	32	28	1000	0.001	10



Modelo	ECM
Sin reducir la dimensionalidad	0.080
Reduciendo la dimensionalidad	0.078

## • LightGBM

num_leaves	learning_rate	feature_fraction	bagging_fraction	max_depth	reg_lambda	num_iterations
128	0.05	0.9	0.9	-1	2	2000



Modelo	ECM
Sin reducir la dimensionalidad	0.030
Reduciendo la dimensionalidad	0.068

# Resumen de resultados

Modelo	Edificios	PCA	ECM	Memoria (MB)
ANN	Todos	No	0.202	3860
ANN	Todos	Sí	0.242	3216
ANN	Uno	No	0.080	0.983
ANN	Uno	Sí	0.078	0.772
<i>LightGBM</i>	Todos	No	0.042	3860
<i>LightGBM</i>	Todos	Sí	0.456	3216
<i>LightGBM</i>	Uno	No	0.030	0.983
<i>LightGBM</i>	Uno	Sí	0.068	0.772

Comparación de los modelos implementados

# Conclusiones

- **Objetivos**

- Cumplimiento de los objetivos principales y secundarios.

- **Desviaciones en la planificación**

- Dificultad para encontrar datos con los que realizar los modelos.
- Capacidad de memoria y de cómputo de Google Colab.

- **Trabajos futuros**

- Probar otros algoritmos para reducir la dimensionalidad.
- Probar a entrenar otros modelos.
- Añadir más metadatos de los edificios.

# Muchas gracias por su atención

**Google drive:**

[https://drive.google.com/drive/folders/1IFC\\_EYfa6YAgIPLDXstLDFdiEzpsUC5q?usp=sharing](https://drive.google.com/drive/folders/1IFC_EYfa6YAgIPLDXstLDFdiEzpsUC5q?usp=sharing)