

# EFICIENCIA ENERGÉTICA

Miguel Porras  
Karen Valdez  
Santos López

## PROBLEMA A RESOLVER ↺

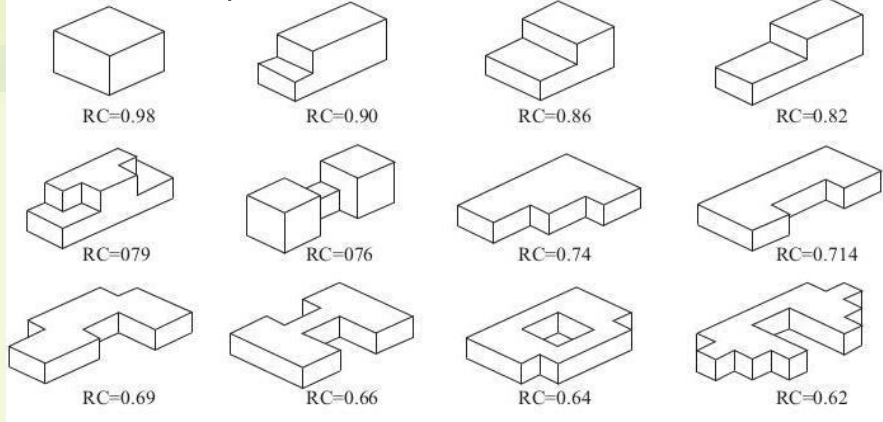
Predecir qué tanta temperatura debemos aumentar o disminuir dentro de un edificio para tener un balance en las instalaciones. Con la intención de mejorar y disminuir la contaminación que produce mantener la temperatura dentro de un edificio no adecuada y reducir la huella de carbono.

## ENTENDIENDO EL DATASET

Podemos considerar que es un dataset pequeño ya que solo contiene 768 observaciones con 10 columnas

### RELATIVE COMPACTNESS

Esta variable nos dice la forma que va a tener el edificio (esta variable es categorica)



### SURFACE AREA

Esta variable nos dice el area superficial que tiene el edificio, cabe denotar de que el area superficial del edificio esta muy correlacionada con el Relative Compactness.

### ROOT AREA

Esta variable nos dice el area de el techo que tiene el edificio, cabe denotar de que el area del techo del edificio esta muy correlacionada con el Relative Compactness.

## HIPOTESIS

- ¿El modelo para predecir el heating se puede usar para el cooling?

Según los resultados encontramos que, si se puede predecir los dos casos con el mismo modelo, pero las variables tienen un comportamiento distinto por lo cual si deseamos un resultado optimo deberíamos de generar el modelo adecuado para predecir cada variable. Concluyendo que en el caso de Heating el modelo más optimo fue Random forest y para Cooling fue XGBoost.

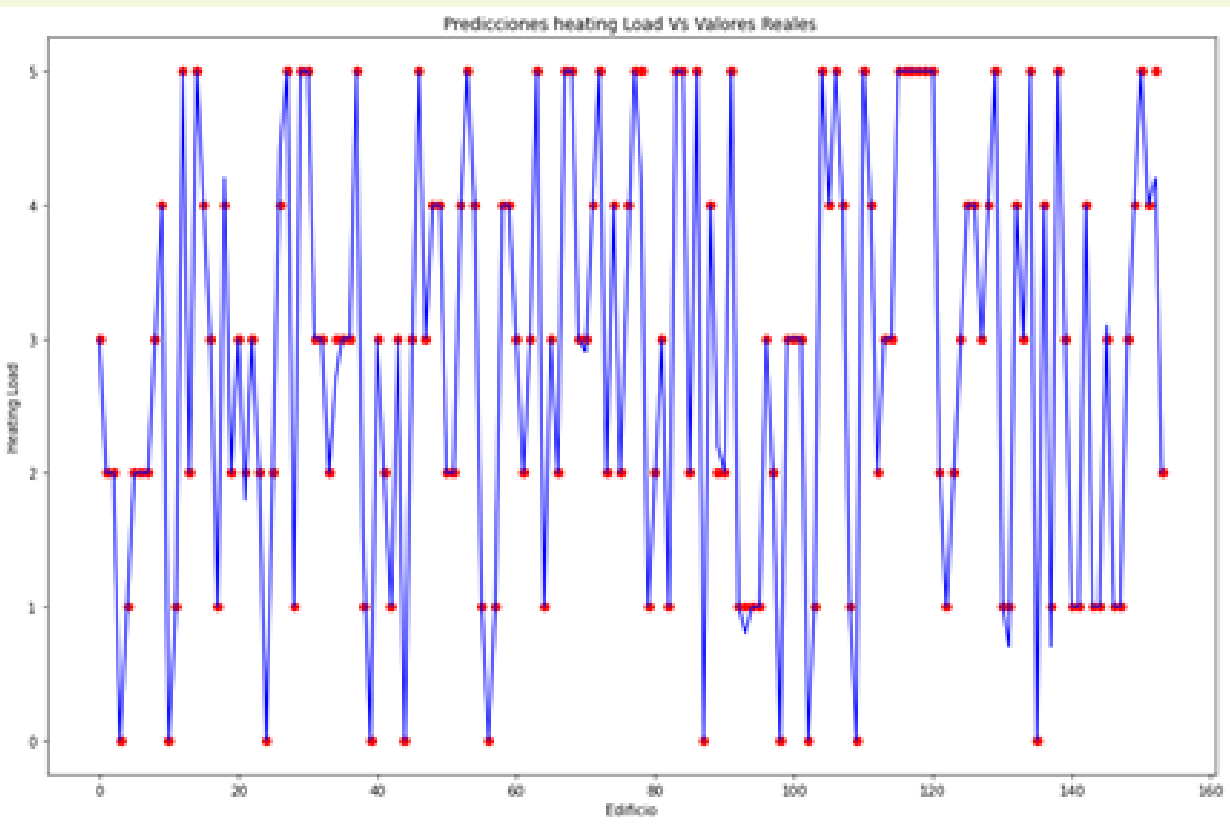
- ¿Un modelo simple de regresión lineal es peor que un modelo complejo formado por muchas capas neuronales?

Notamos que la regresión lineal tuvo un valor mas bajo que una red neuronal sabiendo que se considera un modelo complejo. Concluyendo que en este caso una regresión lineal simple puede explicar a primera vista el problema mejor que una red neuronal, tomando en cuenta que no siempre una respuesta complicada va a ser siempre la mejor

- ¿Alguno de nuestros modelos tiene un mejor resultado que el modelo que se utilizó en el paper (Random Forest)?

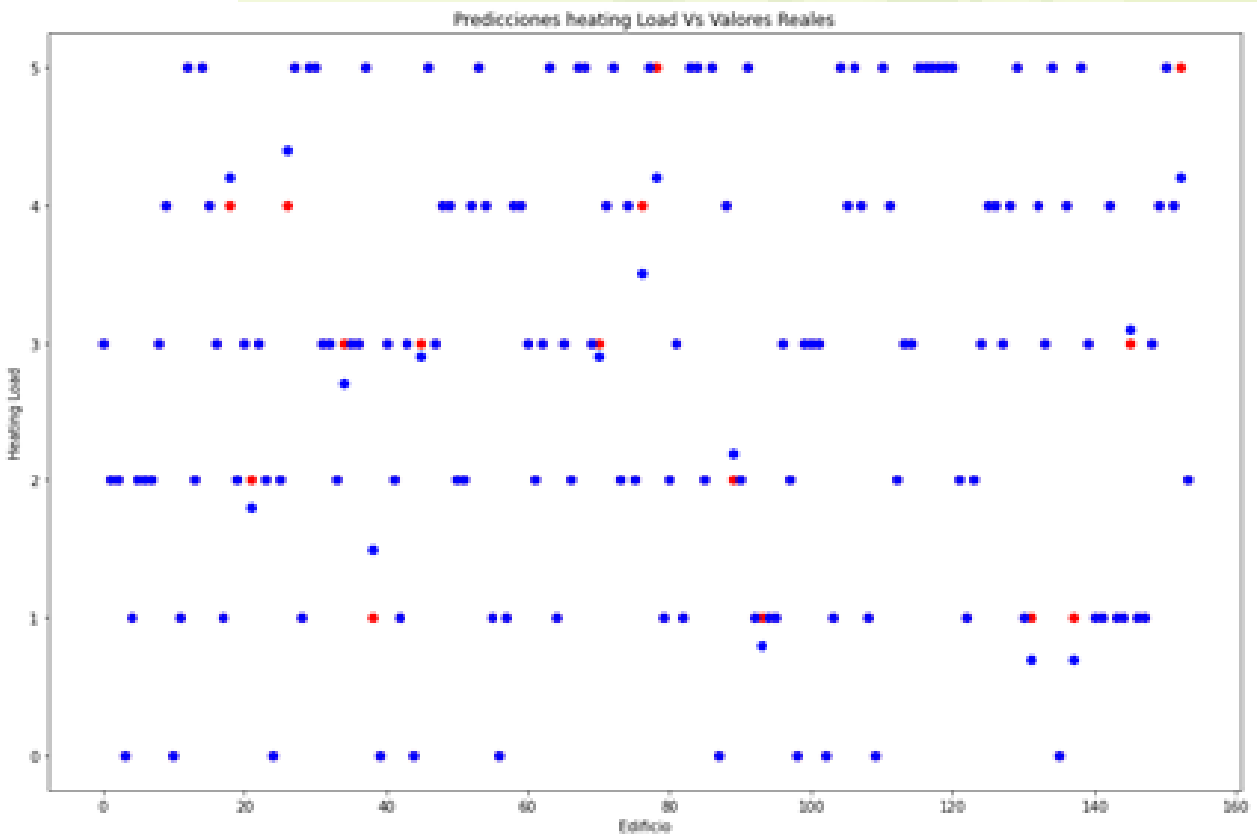
Según los resultados encontramos que nuestros modelos tienen mejores resultados que en modelo que se utilizó en el paper. Concluyendo que en el caso de Cooling el modelo más optimo fue XGBoosting y para el caso de heating fue random forest

## HEATING

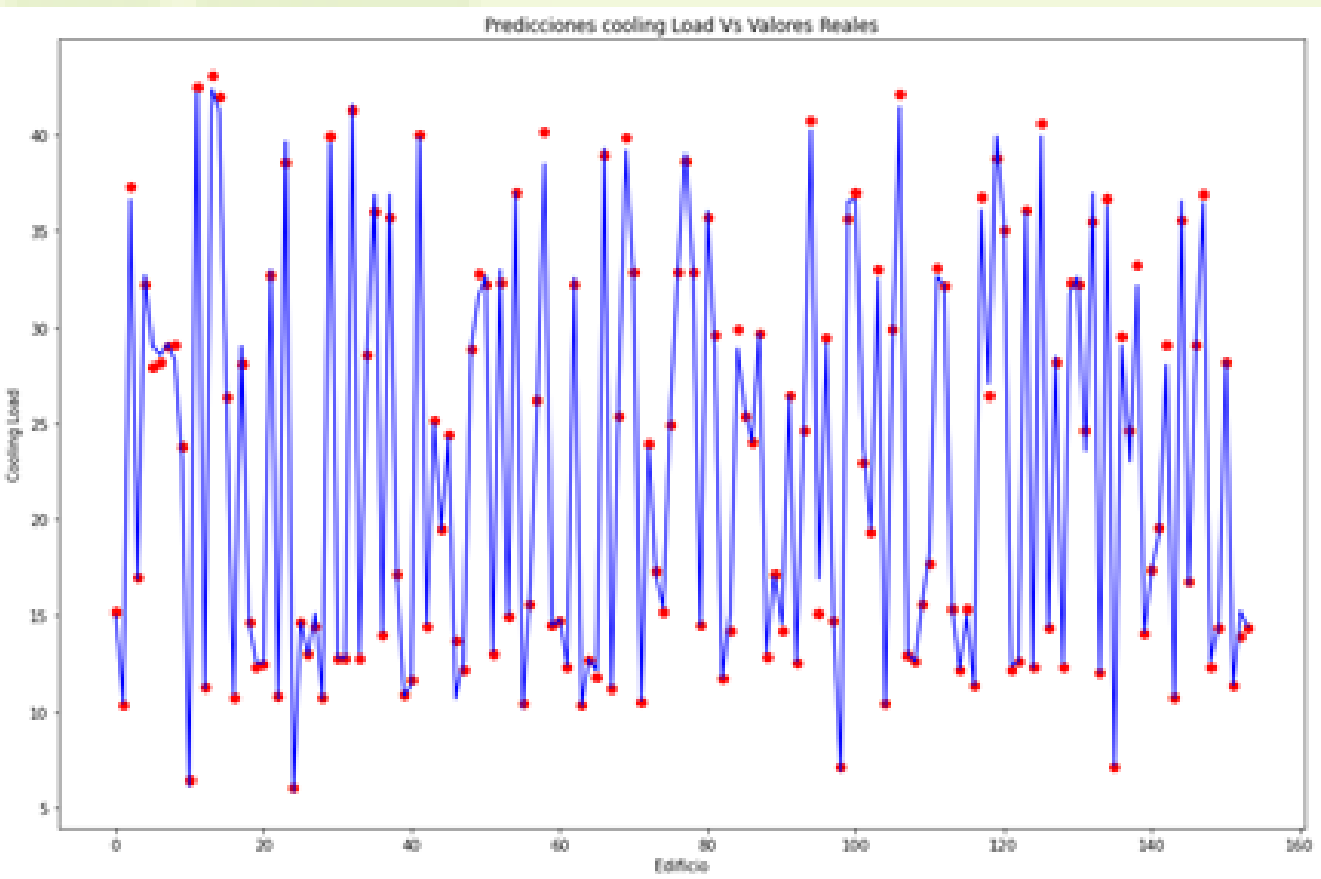


### RANDOM FOREST

EL modelo Random Forest es el modelo que más se ajusta a Heating ya que pareciera que al reducir la varianza logra ajustarse mejor

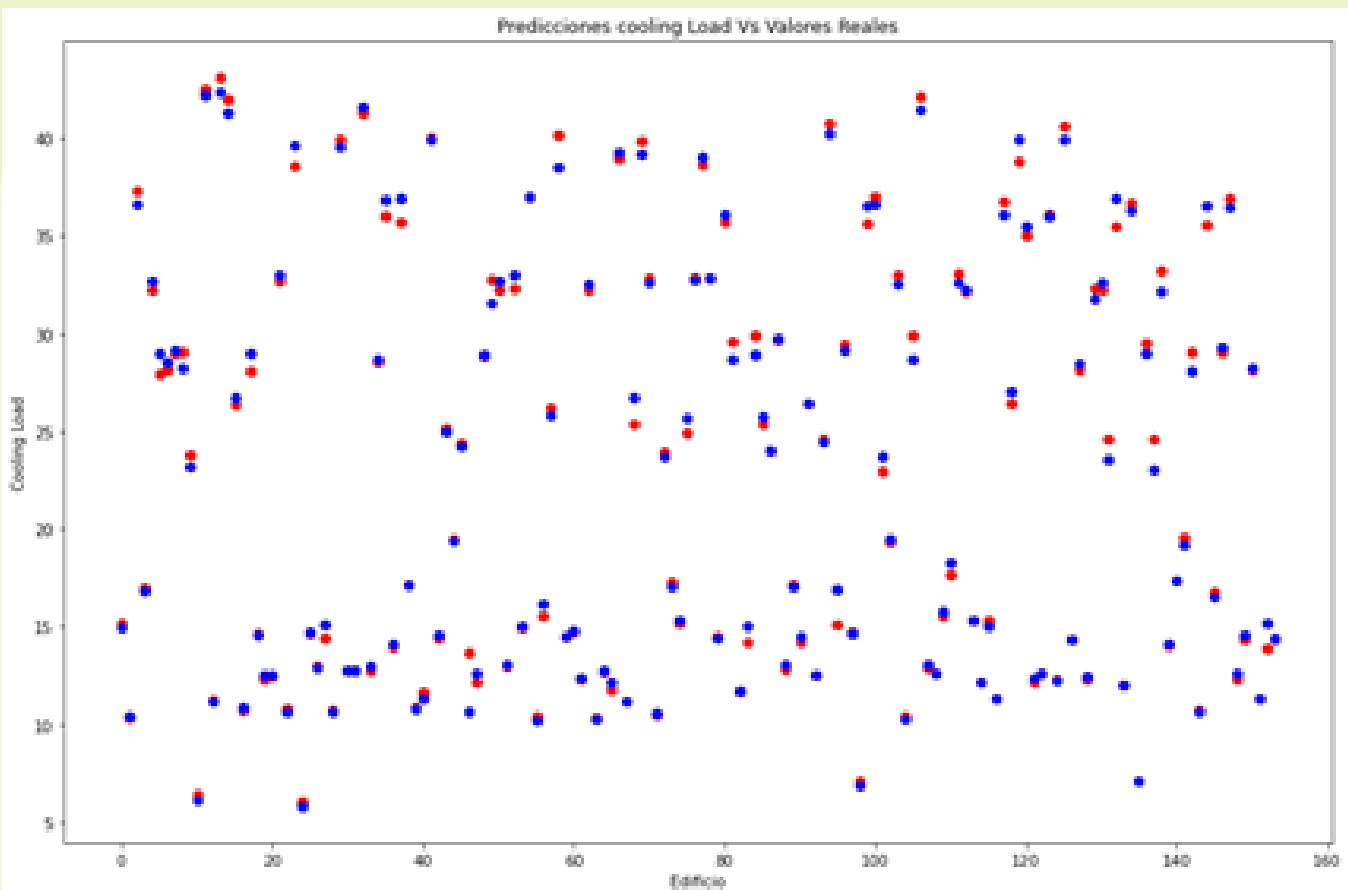


## COOLING

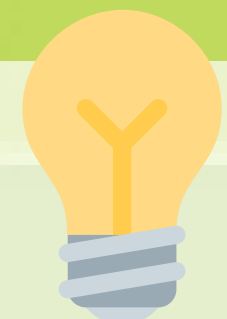


### XGBOOST

EL modelo XGBOOST es el modelo que mas se ajusta a Cooling ya que pareciera que el método de predecir el error logro ajustarse mejor a la forma de Cooling



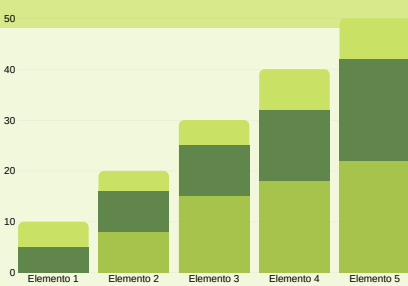
## METODOLOGÍA



Nuestra metodología consiste en entrenar 5 modelos diferentes (Regresión Lineal, SVR, Redes Neuronales, XGBoost y un ensamble de modelos) y luego comparamos el resultado contra el modelo de random forest que se presento en el paper.

La comparación de los modelos la realizamos empleando medidas de comparación como el MSE, MAE y RMSE las cuales son métricas nos permiten explicar que tan cercanos son los valores de nuestro modelo contra los valores reales, al igual que con una visualización que compara nuestras predicciones contra el valor real que tenia nuestro modelo.

## MODELOS



### Regresion lineal

la cual básicamente consiste en utilizar una linea para poder separar los valores o en nuestro caso poder predecir cuál va a ser el valor relativo a las otras variables que tiene nuestro dataset.

### SVR

El cual también crea una linea para poder predecir nuestro valor de heating y cooling load utilizando otras variables pero este utiliza una “calle” la cual funciona como un margen de aceptación para realizar la linea.

### Redes Neuronales

el cual consiste en un grupo de nodos llamados neuronas las cuales tienen una función de activación y se encuentran conectadas unas con otras. Al juntar todas las neuronas podemos tener una red de predicción muy robusta la cual consiste de varias funciones matemáticas que afectan en la decisión de la predicción que queremos realizar

### Arboles de Decisión

es un modelo el cual nos ayuda a partir nuestras variables en preguntas de Si/No y nos ayuda a tomar decisiones de cómo es que vamos a predecir nuestros valores.

### XGBoost

el cual se basa en un ensamble de modelos de arboles de decisión, este modelo se ajusta muy bien a datasets grandes y pequeños ya que tiene mucha flexibilidad en términos de la cantidad de arboles que vamos a crear, su profundidad, etc.

### Ensamble de modelos

es cuando entrenamos varios modelos y luego los resultados los proporcionamos para poder predecir los resultados tomando en cuenta la opinión de cada modelo.

### Random Forest

es un ensamble de modelos de arboles de decisión básicamente crea muchos modelos para poder predecir de varias maneras el valor de nuestro modelo.