

UTILIZANDO INTELIGENTEMENTE LA ENERGÍA

Miguel Porras 17001003
Karen Valdez 16006196
Santos López 15002241

MOTIVACIÓN

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 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 con 8 columnas independientes y 2 columnas dependientes

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 área superficial que tiene el edificio, cabe denotar de que el área superficial del edificio esta muy correlacionada con el Relative Compactness.

ROOF AREA

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

GLAZING AREA

Esta variable nos dice el área de acristalamiento que tiene el edificio, en otras palabras nos dice el área de vidrios y espejos que contiene el edificio.

OVERALL HEIGHT

Esta variable nos va a decir la altura general que tiene el edificio.

WALL AREA

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

AREA DISTRIBUTION

Esta variable nos dice la distribución del área de acristalamiento dentro del edificio.

ORIENTATION

Esta variable nos dice la orientación en la que se encuentra el edificio.

HEATING LOAD

Esta variable es una de las variables que vamos a predecir y nos dice que tanto debemos de aumentar la temperatura para que mantenga un rango.

COOLING LOAD

Esta variable es una de las variables que vamos a predecir y nos dice que tanto debemos de disminuir la temperatura para que mantenga un rango.

HIPOTESIS

- ¿El modelo para predecir el heating se puede usar para el cooling?
- ¿Un modelo simple de regresión lineal es peor que un modelo complejo formado por muchas capas neuronales?
- ¿Alguno de nuestros modelos tiene un mejor resultado que el modelo que se utilizó en el paper (RandomForest)?

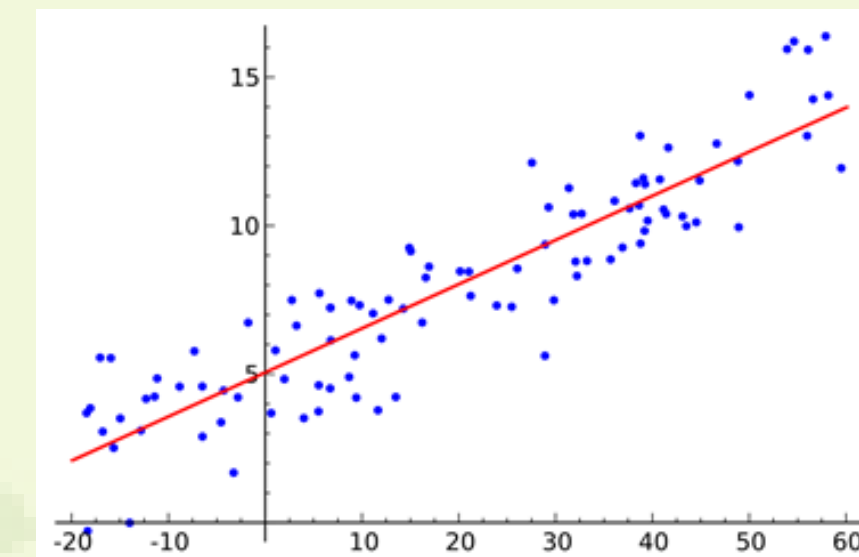
METODOLOGÍA

Nuestra metodología consiste en entrenar 4 modelos diferentes (Regresión Lineal, Redes Neuronales, XGBoost y arboles de decisión) y luego comparamos el resultado contra el modelo de random forest que se presenta 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 tenía nuestro modelo.

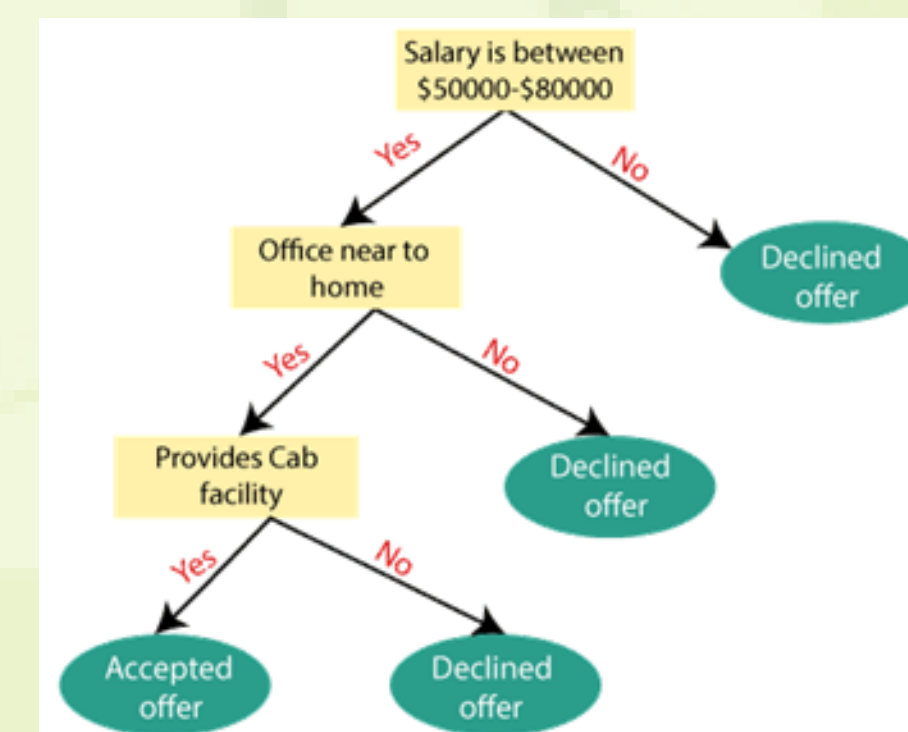
ARQUITECTURA

REGRESIÓN LINEAL

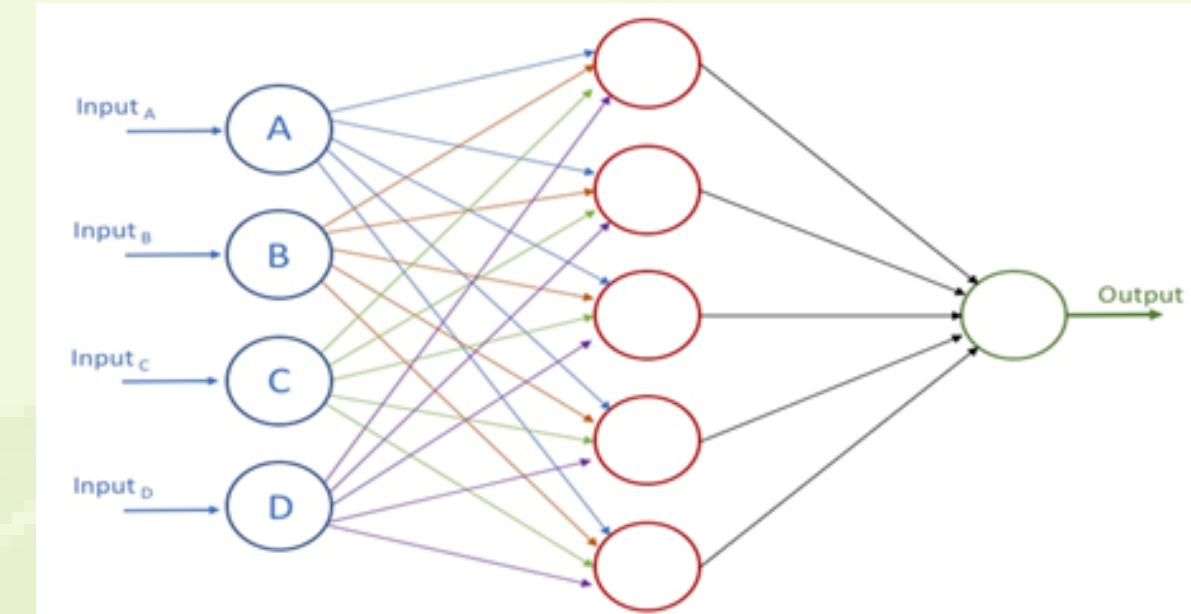


la cual básicamente consiste en utilizar una línea para poder separar los valores o en nuestro caso poder predecir cuál va a ser el valor relativo a las otras variables que tiene nuestro dataset.

ARBOLES DE DECISIÓN

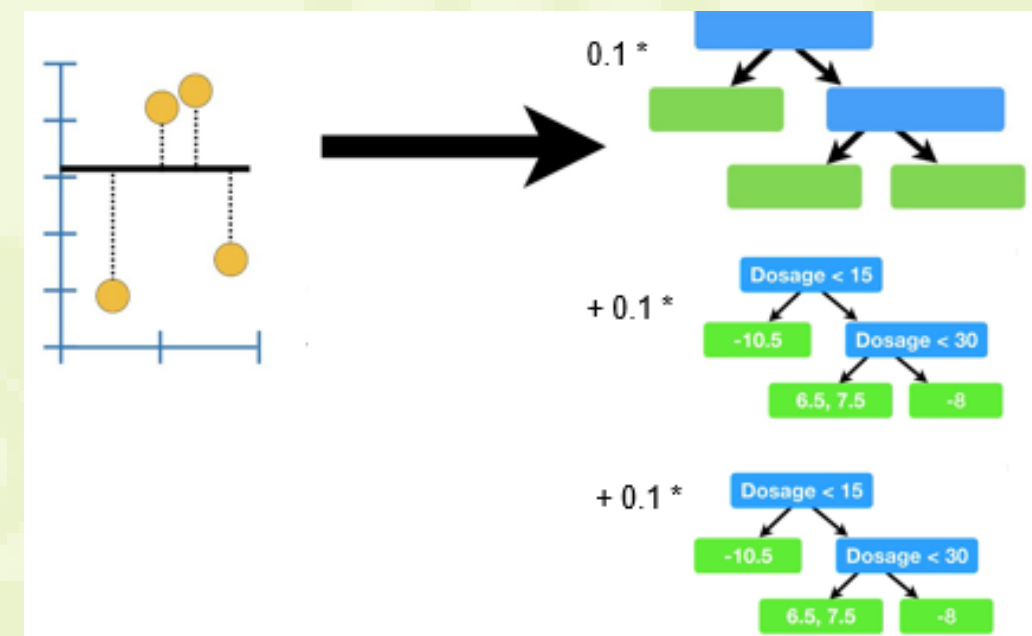


REDES NEURONALES

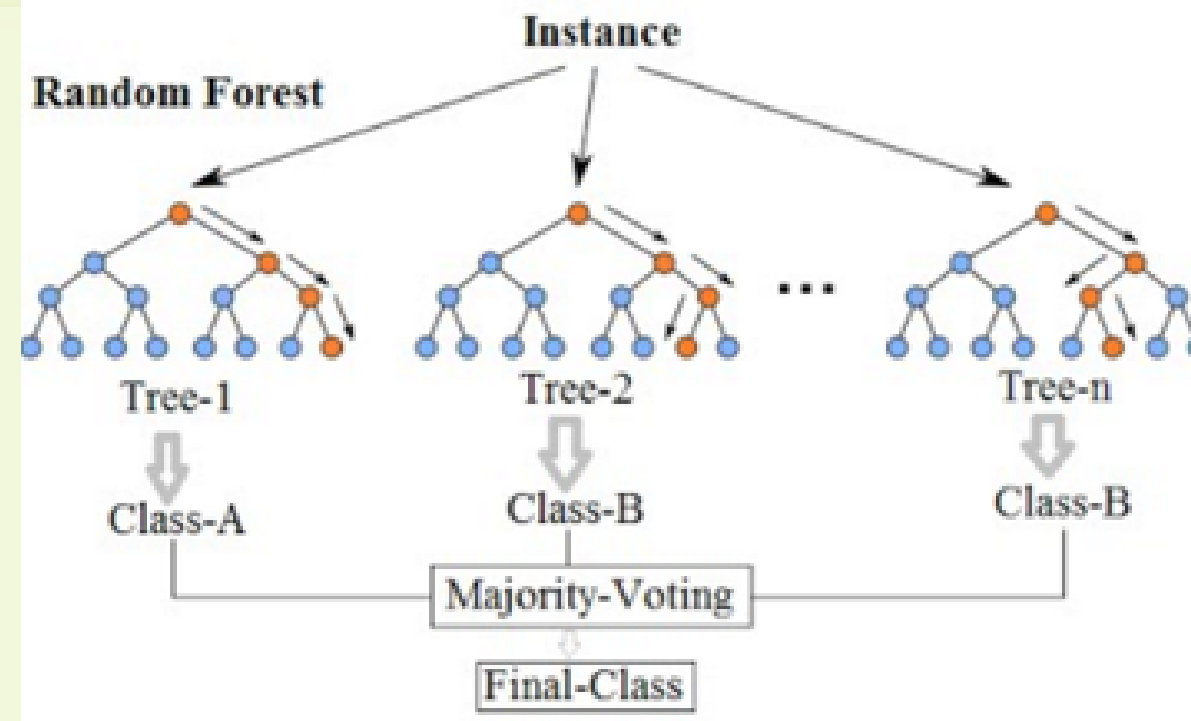


el cual consiste en un grupo de nodos llamados neuronas las cuales tienen una función de activación y se encuentran conectadas unas con otras. Al juntar todas las neuronas podemos tener una red de predicción muy robusta la cual consiste de varias funciones matemáticas que afectan en la decisión de la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

XGBOOST



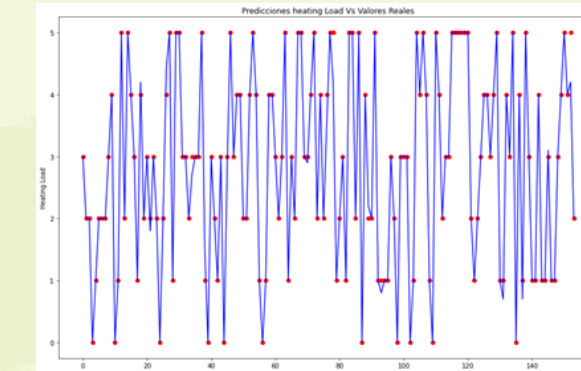
RANDOM FOREST



Es un ensemble de modelos de arboles de decisión básicamente crea muchos modelos para poder predecir de varias maneras el valor de nuestro modelo, luego pone a todos los arboles a predecir y la predicción que tenga la mayor cantidad de votos va ser nuestra predicción final

RESULTADOS

HEATING

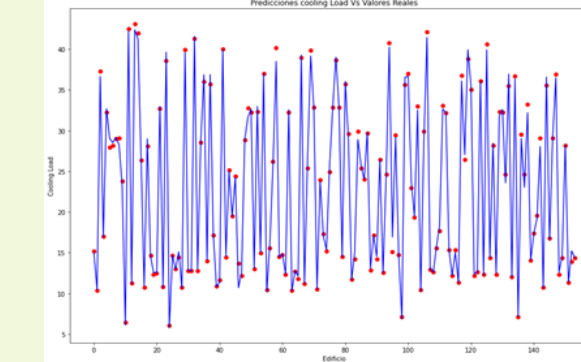


RANDOM FOREST

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

Modelo	Pred. Heat
Reg. Lineal	0.766172285
Arbol de decisión	0.080582296
Red Neuronal	1.828542613
XGBoost	0.273621641
Random Forest	0.124837557

COOLING



XGBOOST

El modelo XGBOOST es el modelo que más se ajusta a Cooling ya que pareciera que el método de predecir el error logra ajustarse mejor a la forma de Cooling

Modelo	Pred. Cooling
Reg. Lineal	3.227311309
Arbol de decisión	0.635354263
Red Neuronal	19.40496753
XGBoost	0.593202956
Random Forest	0.628409092

CONCLUSIONES

- Logramos concluir que si se puede predecir correctamente las dos variables con un solo modelo
- En el caso de Heating el modelo optimo fue Random forest y para Cooling fue XGBoost.
- 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

HALLAZGOS

- Que un modelo de regresion lineal puede expresar mejor el problema que una red neuronal
- Que la variable de Heating y Cooling no tiene el mismo comportamiento ya que pareciera que heating tiene un comportamiento lineal y cooling tiene un comportamiento no lineal ya que los algoritmos lineales lograron predecir mejor heating que cooling

TRABAJOS A FUTURO

- Investigar otros modelos
- Podríamos Tratar de optimizar más los parámetros de XGBoost y de Random Forest
- Probar hacer un ensemble entre Random Forest y XGBoost

REFERENCIAS

- <https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>
- <https://www.sciencedirect.com/science/article/abs/pii/S037877881200151X>
- <https://xgboost.readthedocs.io/en/latest/>
- http://andrei.clubcisco.ro/cursuri/fff-sym/5master/aac-nnga/AI_neural_nets.pdf