

Proyecto entrega 1

Materia:

Introducción a la inteligencia artificial

Victor Gabriel Navarro Serna 1037577906

Profesor:

Raúl Ramos Pollan

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERIA



MEDELLIN 2022

## 1. Describas el problema predictivo a resolver

Vamos a predecir la relación entre pares de átomos en moléculas, dados los dos tipos de átomos (p. ej., C y H), el tipo de acoplamiento (p. ej., 2JHC) y cualquier característica que pueda crear a partir de la estructura de la molécula.

## 2. El dataset que vas a utilizar

- `train.csv`: El conjunto de entrenamiento, donde la primera columna (`molecule_name`) es el nombre de la molécula donde se origina la constante de acoplamiento (el archivo XYZ correspondiente se encuentra en `./structures/.xyz`), la segunda (`atom_index_0`) y la tercera columna (`atom_index_1`) son los índices de los átomos del par de átomos que crean el acoplamiento y la cuarta columna (`scalar_coupling_constant`) es la constante de acoplamiento escalar que queremos poder predecir
- `test.csv` - el conjunto de prueba; la misma información que el tren, sin la variable de destino
- `sample_submission.csv`: un archivo de envío de muestra en el formato correcto
- `estructuras.zip` - carpeta que contiene archivos de estructura molecular (xyz), donde la primera línea es el número de átomos en la molécula, seguida de una línea en blanco y luego una línea para cada átomo, donde la primera columna contiene el elemento atómico (H para hidrógeno, C para carbono, etc.) y las columnas restantes contienen las coordenadas cartesianas X, Y y Z (un formato estándar para químicos y programas de visualización molecular)
- `estructuras.csv`: este archivo contiene la misma información que los archivos de estructura xyz individuales, pero en un solo archivo

## 3. Las métricas de desempeño requeridas (de machine learning y de negocio)

Un modelo de predicción de interacción magnética debería tener un porcentaje mínimo de acierto de 85%, ya que la diferencia entre la teoría y la experimentación tienen un máximo 15% de diferencia para que sea considerado aceptable.

## 4. Un primer criterio sobre cuál sería el desempeño deseable en producción

No predecirá *todos* los pares de átomos en cada molécula, sino que solo necesitará predecir los pares que se enumeran explícitamente en los archivos de entrenamiento y prueba. Por ejemplo, algunas moléculas contienen flúor (F), pero no predecirá la constante de acoplamiento escalar para ningún par que incluya F.