

Proyecto entrega 2

Materia:

Introducción a la inteligencia artificial

Victor Gabriel Navarro Serna 1037577906

Profesor:

Raúl Ramos Pollan

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERIA



MEDELLIN 2022

## Introducción

Como sabemos realizar predicciones de problemas a nivel molecular son bastantes complicadas, lo que representa un gran desafío para la ciencia de datos. Nuestro objetivo es predecir las interacciones entre átomos, es decir, predecir la interacción magnética entre dos átomos en una molécula, para mejor comprensión lo que se desea hallar es la constante de acoplamiento escalar, que no es más que un número que determina la fuerza de la interacción respecto a la energía cinética. Esta constante depende de los electrones y de los enlaces químicos que forman la estructura tridimensional de una molécula. Así que el obtener un método fiable y rápido para predecir estas interacciones permitirá a los científicos comprender cómo la estructura química 3D de una molécula afecta a sus propiedades y comportamiento, con lo que finalmente se pueden realizar tareas celulares o ayudar a mejorar el desarrollo de fármacos.

### Desarrollo:

#### 1. Librerías:

Lo primero que vamos a hacer es importar las librerías, que son: numpy, os, pandas, seaborn, matplotlib.pyplot, warnings.filterwarnings, desde sklearn.metrics importamos make\_scorer y mean\_squared\_error, y finalmente desde sklearn.model\_selection importamos cross\_val\_score.

#### 2. Dataset:

Los dataset que vamos a utilizar son: energía potencial (potential\_energy.csv), la carga de Mulliken (mulliken\_charges.csv), test.csv, la contribución del acople escalar (scalar\_coupling\_contributions.csv), el momento dipolar (dipole\_moments.csv), el tensor de campo magnético (magnetic\_shielding\_tensor.csv), la estructura (structures.csv) y finalmente train.csv.

#### 3. Carga de datos:

```
pot_energy=pd.read_csv('../input/potential_energy.csv')
mulliken_charges=pd.read_csv('../input/mulliken_charges.csv')
train_df=pd.read_csv('../input/train.csv')
```

```
test_df=pd.read_csv('../input/test.csv')
magnetic_shield_tensor=pd.read_csv('../input/magnetic_shielding_tensors.csv')
dipole_moment=pd.read_csv('../input/dipole_moments.csv')
structures=pd.read_csv('../input/structures.csv')
scalar_coupling_cont=pd.read_csv('../input/scalar_coupling_contributions.csv')
')
```

#### 4. Ver el conjunto de los datos

```
print('Shape of potential energy dataset:',pot_energy.shape)
print('Shape of mulliken_charges dataset:',mulliken_charges.shape)
print('Shape of train dataset:',train_df.shape)
print('Shape of dipole moments dataset:',dipole_moment.shape)
print('Shape of structures dataset:',structures.shape)
print('Shape of test dataset:',test_df.shape)
print('Shape of magnetic shielding tensors dataset:',magnetic_shield_tensor.
shape)
print('Shape of scalar coupling contributions dataset:',scalar_coupling_cont
.shape)
```

#### 5. Exploración de datos

```
print('Data Types:\n',pot_energy.dtypes)
print('Descriptive statistics:\n',np.round(pot_energy.describe(),3))
pot_energy.head(6)
print('Data Types:\n',mulliken_charges.dtypes)
print('Descriptive statistics:\n',np.round(mulliken_charges.describe(),3))
mulliken_charges.head(6)
print('Data Types:\n',train_df.dtypes)
print('Descriptive statistics:\n',np.round(train_df.describe(),3))
train_df.head(6)
print('Data Types:\n',scalar_coupling_cont.dtypes)
print('Descriptive statistics:\n',np.round(scalar_coupling_cont.describe(),3
))
scalar_coupling_cont.head(6)
print('Data Types:\n',test_df.dtypes)
print('Descriptive statistics:\n',np.round(test_df.describe(),3))
test_df.head(6)
print('Data Types:\n',magnetic_shield_tensor.dtypes)
print('Descriptive statistics:\n',np.round(magnetic_shield_tensor.describe()
,3))
magnetic_shield_tensor.head(6)
print('Data Types:\n',structures.dtypes)
print('Descriptive statistics:\n',np.round(structures.describe(),3))
structures.head(6)
```

## 6. Asignación de datos de la estructura atómica en un conjunto de datos de entrenamiento y prueba

En esta parte lo que vamos a realizar es una primera prueba, que consiste en diseñar un vector de distancia entre átomos. Posteriormente, comprobamos nuestro conjunto de datos de entrenamiento y prueba, al realizar el primer intento no fueron exitosos, ya que se pasó mucho tiempo entendiendo el problema y los datos para poder ejecutarlos correctamente.

## 6. Asignación de datos de la estructura atómica en un conjunto de datos de entrenamiento y prueba.

En esta parte lo que se hizo fue realizar una primera prueba, la cual no fue exitosa.