

Cahier des charges :

Composition de l'équipe :

- Maxime Mezeray (**Référent**)
- Moncef Siala
- Anir Firaouni

Thématique du projet :

Catégorisation de réaction chimiques par changement d'énergie en électronVolt par électronégativité.

Caractéristique des données :

Le dataset se compose de 5 catégories, les deux réactifs sous forme SMILES (un ion et une molécule), les deux produits eux aussi puis la dernière catégorie est le target, la quantité électrique (en eV) afin de pouvoir réaliser la réaction

Fonctionnalités du projet :

- **A quelles questions, le système doit-il répondre ?**

Prédire la catégorie/classe en fonction de la quantité d'électronvolt nécessaire afin de stabiliser une molécule. (Catégorie 1 : 0-15, Catégorie 2 : 16-30, Catégorie 3 : 31-45, Catégorie 4 : 46-60)

- **Comportement de l'AI**

L'AI sera modélisé un arbre de décision basé sur une fonction de coût, elle s'entraînera sur le dataset traduit à l'aide d'un module qui peut lire l'écriture SMILES. Elle s'entraînera avec un algorithme de cross-validation. Une fois entraînée, elle peut prédire l'électronégativité en électronvolt nécessaire à réaliser la réaction complète

Caractéristiques techniques :

Nous utilisons la librairie sklearn afin de pouvoir prédire nos résultats et afin de modéliser le modèle. Le modèle est un arbre de décision basé une fonction d'erreur, qui s'entraînera sur un dataset préparé au préalable, pour l'entraînement nous utiliserons la cross-validation afin de s'assurer de l'autonomie du modèle par rapport au dataset

Afin de pouvoir lire les informations mais surtout afin que l'on puisse travailler avec et que l'on puisse l'exploiter avec notre IA, nous avons utilisé la librairie pandas qui permet de lire des fichiers en .csv et de les rendre lisible via Python sous forme d'une liste de listes

Nous utilisons aussi la librairie pysmiles afin de pouvoir lire la notation SMILES des données afin que l'on puisse par la suite traduire les molécules pour que les données soient lues plus facilement, puis nous représenterons les atomes qui constitue la molécule par leur numéro atomique.

Scénario d'usage :

Le modèle reçoit 4 données, représentant les deux réactifs et les deux produits, puis il prédit la classe à laquelle la réaction chimique appartient.

Type de Licence :



Signatures

A stylized black ink signature consisting of a large, loopy 'M' with a horizontal line extending to the right.

Maxime Mezeray

A stylized black ink signature featuring a large, rounded 'A' with a horizontal line and a small vertical line above it.

Anir Firaouni

A stylized blue ink signature with a large, loopy 'M' and a horizontal line extending to the right.

Moncef Siala