
Jesus Emmanuel Ramos Davila

Maestría en ciencia de datos

Facultad de Ciencias Fisico Matematicas

Regression para prediccion de masa molar usando CNN

12th Noviembre 2023

Script

Este detalla los pasos realizados en todo el ciclo de desarrollo tanto del modelo como de las obtención , procesamiento y limpieza de datos

1. Análisis y parseo de el origen de los datos en formato XML, convertir datos a formato plano en CSV
2. Fase 1 web scrapping consiste en obtener por la URL obtenida del paso 1 otras propiedades reportadas de su respectiva fórmula química, tales como masa molar, fórmula química, y categoría de pureza
3. Fase 2 web scrapping consiste en obtener la imagen de sus estructura química
4. Creación de modelo CNN para regresión usando las imágenes y la variable masa molar como variable de respuesta
5. Resultados
6. Mejoras al modelo
7. Trabajo a futuro (Incluir la parte de uso de pictogramas de seguridad para un cambio de enfoque en la variable de repuesta a ser un modelo de clasificación 'multi categorías')