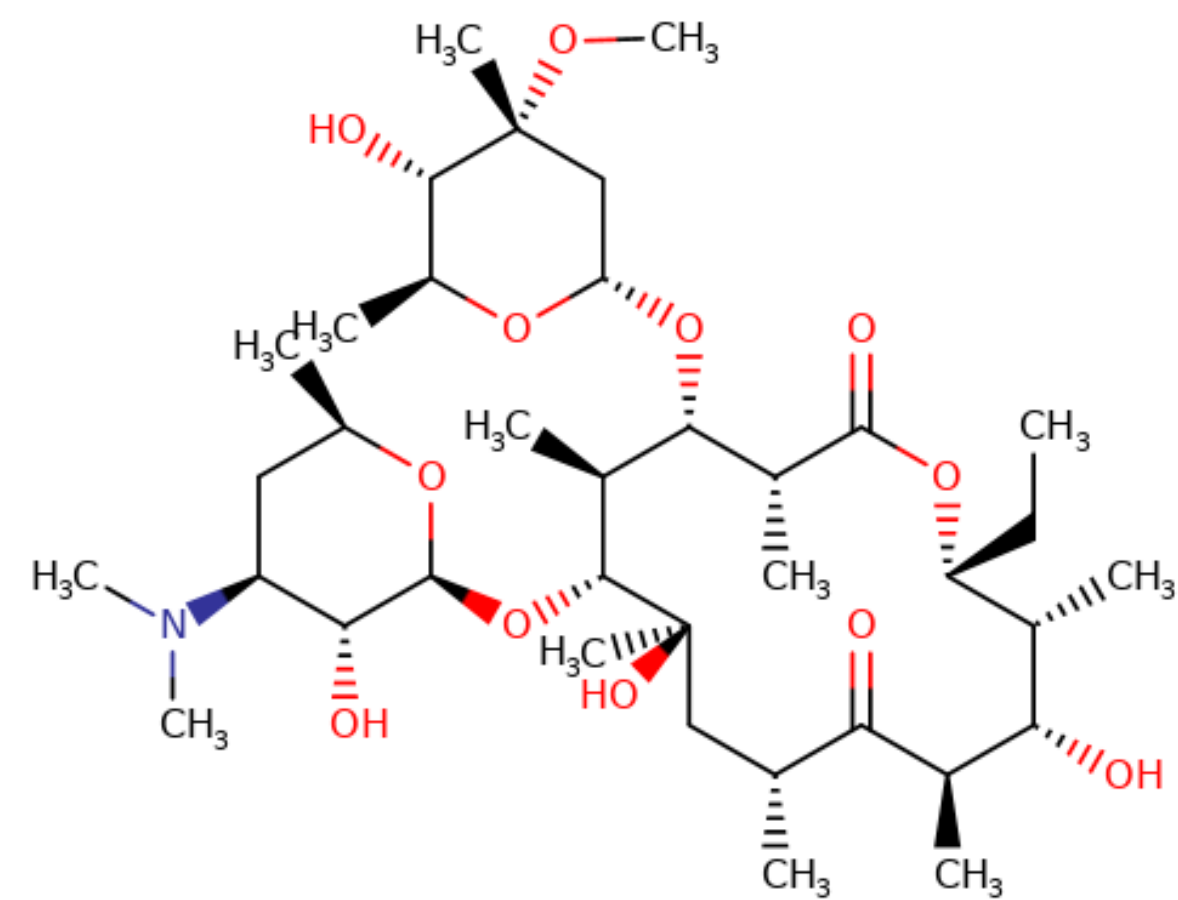


Imputacion de datos

Datos nulos dentro de dataset



Fase 1: Lectura y
Webscrapping

LOS DATOS REPORTADOS DE FORMULAS QUIMICAS COMPRENDEN MAS DE 4500 FORMULAS, ALGUNAS PROPIEDADES REPORTADAS NO SE ENCONTRARON TANTO EN LA FASE DE WEBSCRAPPING COMO EN EL DATASET.

Revision manual de
cada elemento

Filtrado de elementos

Filtrado de datos los cuales no se reporta formula quimica por webscrapping o alguna otra propiedad tal como masa molar

Imputacion de datos
categoricos y texto

Revision manual de formulas quimicas por segundo web scrapping hacia otra tabla y se imputaron datos con valores por default

Resumen

Se analizaron las 4500 formulas las cuales contenian errores tanto en valor reportado como datos faltantes no reportados los cuales en un segundo web scrapping se reemplazaron datos reales. **total de formulas con datos completos 3751**

Ejemplos:

`purity_category = "min 95%"`
`substance_category = "Supplied by Biosynth"`

