# Protocolo seguido para generar el modelo de Inhibición CYP2D6

## Endpoint

Inhibidores o no inhibidores para CYP2D6

## Origen de los datos

Los datos vienen de la herramienta Deep-PK: <https://biosig.lab.uq.edu.au/deeppk/data>

## Tratamiento de los datos

Se obtienen un total de tres archivos csv, separados en test, train y validation. Para procesar estos archivos se ha empleado el script “prepare\_preprocessed\_and\_Hygieia”. El archivo procesado se dividió para llevar a cabo el cálculo de descriptores. Finalmente, el resultado de WOTAN se unión con el script “concat.py”.

Se llevó a cabo una eliminación de los OUTLIERS y, posteriormente un undersampling con el script “Elbow\_Undersampling”.

## Transformación de la “y”

No hay transformación

## Train/test ratio

Se ha realizado una partición del 80% para el train y del 20% para el test. Durante el proceso de partición se han mantenido los compuestos señalados como *quite dissimilar*.

## Scaler

Se ha usado el *Standard Scaler*