

Resultados de la primera entrega:

Recall: 0.3367003367003367

Fbetascore: 0.860096267655646

Precision: 1

Threshold: 1128.962334

Siendo el recall = $0.3367 = TP / TP + FN$

He calculado el número de secuencias que clasificamos como anómalas, que es lo mismo que TP ya que la precisión que tenemos es 1. Con TP = 100 despejo el recall para obtener los FN.

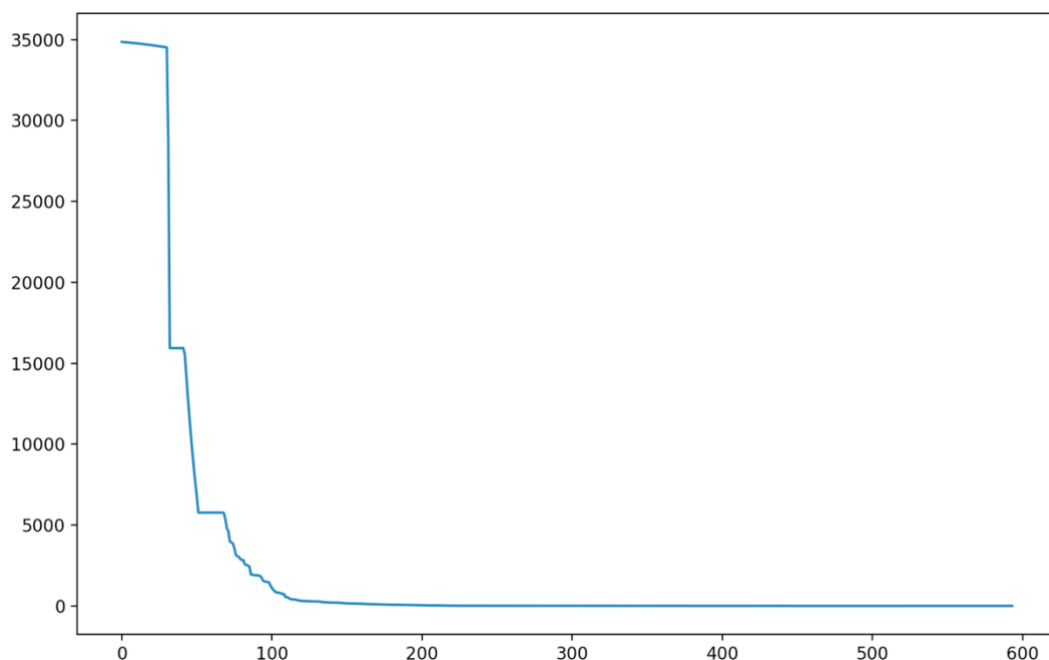
FN = 197, es decir, **hay 297 anomalías en validación de 594 secuencias**. Ahora, dónde ponemos el threshold para intentar capturarlas todas?

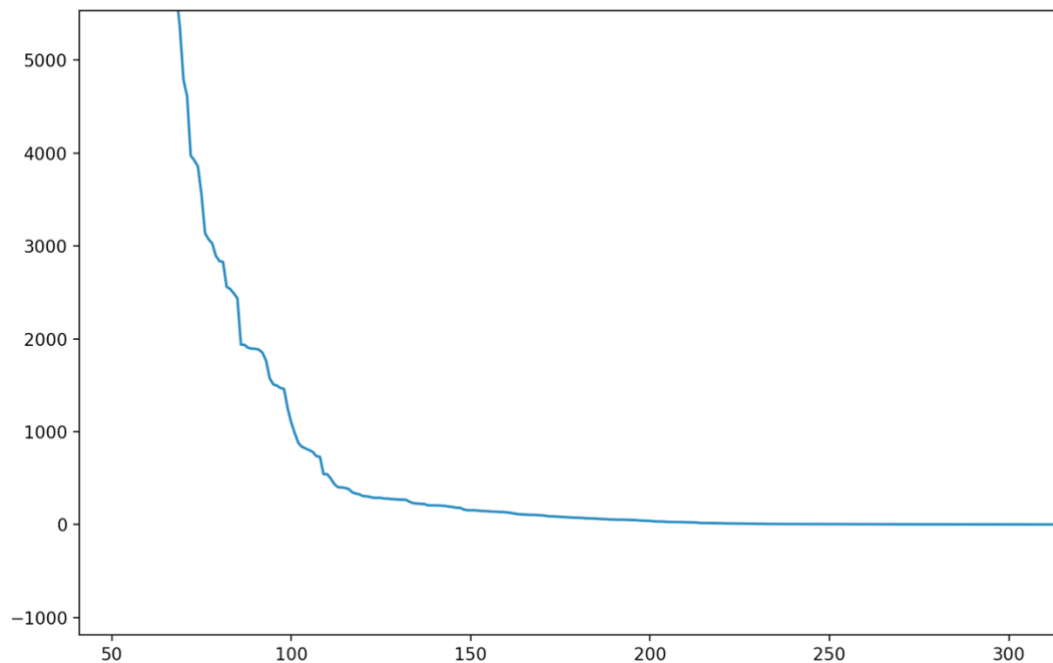
Poner el threshold mirando los valores de la gráfica en train

En train teníamos un escalón cerca del valor 416. Poniendo el threshold ahí obtenemos 119 positivos, solo 19 más que antes. He ido bajando el threshold mirando los escalones de train llegando al final a un threshold de 58.834819 y obteniendo 188 positivos.

Poner el threshold mirando los valores de la gráfica en valid

Mirando la gráfica en validación para tener las 297 secuencias más anómalas nos encontramos con esto (secuencias en el eje x, suma de distancias en el eje y):





Para obtener 297 anomalías en el modelo hay que poner un threshold de más o menos 2.2, lo cuál no se si vaya a funcionar del todo bien.

Pienso probar mañana el threshold de 58.834819 para tener clasificadas 188 anomalías y ver si no perdemos mucha precisión y podemos bajar aún más el threshold.