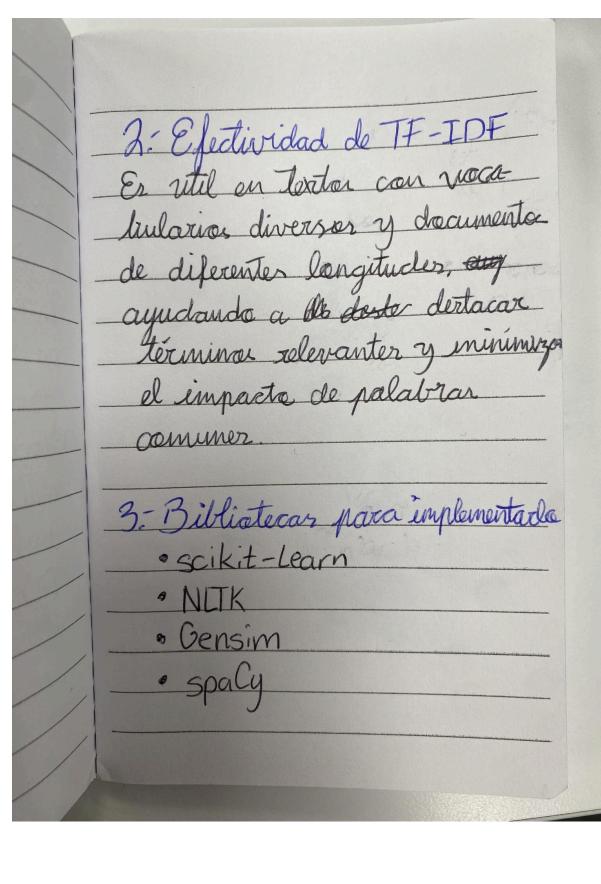
Taread Pt1 1. Investigue la estrategia de vetouzacion TF-IDF. Lécuie que mide la importancia de una palabra en un documento dentro de un conjunto documentos. Se calcula multiplicande la precuencia de la palatra en el documente (TF) por la frediencia inversa de esa palaba en todos las documentos (IDF)



4. Laplace Smoothing en /1-grams · Evita que 11-gramas no als about observador tengan una probabilidad de coro, la que podicio causar fallos en el modela · Urigna una probabilidad a todar los M-gramas, indura si aparecen en el corpur, anadiendo una constante a los frecuencias (generalmente 1) · Evita probabilidades nulas, perce puede asignar probabilità po realisto a sec. racos

5: Palabrar fuera de Vac. · Si una palabra del test set no ostá en d voc. del modela, no se tengan puede calcular la probabilidad. allor · Se puede usar suarvyando (como Laplace), etiquetar expecialer para DOV, a modelor que retroceden corpus a M-gramar de menor 2 los orden (back-off models) para estimar probabilidales ulas bability ron