Salut @[DS P7] Maxime Le Gall ! La valeur du paramètre lambda ne dépend pas de la taille du jeu de données (qui est d'ailleurs plutôt conséquent pour ce projet), mais des contraintes métier.

En ce qui me concerne, je n'ai pas utilisé les indices de Kulczynski et Dice pour une raison triviale : je ne les connaissais pas. Du coup, je te remercie pour la référence et je vais aller regarder leur documentation.

Ma démarche a été la suivante : une société de crédit est une société "à but lucratif". J'ai donc supposé qu'elle cherche à optimiser ses revenus. J'ai ensuite associé à chacune des quatre situations possibles (FP, TP, FN, TN) une "valeur" (gain ou coût pour l'entreprise). En l'occurrence, j'ai considéré : TP\_value = FP\_value = 0 ; TN\_value = 1 et FN\_value = -10. Ce sont des valeurs arbitraires qui devraient être revues par l'entreprise en fonction de la "réalité métier" (cela revient à fixer le lambda de tes métriques). À partir de cela, j'ai construit le gain (fonction d'utilité si tu préfère) de l'entreprise.

Et j'ai normalisé pour obtenir une valeur entre 0 (modèle naïf qui prédit la classe non-défaut à tous) et 1 (modèle parfait qui prédit les vrais classes). Cela me donne donc une "métrique métier", qui peut facilement être paramétrisée avec TN\_value et FN\_value.

