

Concours kaggle

SPOTIFY, PRUT-ON PRÉDIRE LA POPULARITÉ D'UNE CHANSON ? - L2 MATH, IMIA -

L'évolution rapide de l'industrie musicale et la prévalence des plateformes numériques pour la consommation de musique soulignent l'importance de prédire la popularité des chansons. Ce challenge vise à construire un modèle prédictif le plus précis possible pour la popularité des chansons en explorant diverses architectures de réseaux neuronaux. Ces modèles offrent des informations cruciales sur le type de musique qui plait au public.



La popularité des morceaux est mesurée début avril 2025. Ce score est calculé sur la base de différents facteurs, notamment le nombre de flux, les interactions avec les utilisateurs telles que les sauvegardes et les partages, et l'inclusion du morceau dans les listes de lecture les plus populaires. Dans ce challenge nous nous concentrons sur les titres les mieux classés. Et nous utilisons des informations quantitatives telles que la durée, le potentiel pour la danse, le rythme, etc pour prédire la popularité.

Dans ce concours, nous proposons d'utiliser un ensemble de données d'apprentissage contenant 55000 titres. L'objectif est de construire un (ou plusieurs) modèle(s) permettant de prédire la popularité des morceaux de l'ensemble de test.

Pour rejoindre le concours kaggle, obtenir les règles du jeu, une description des données, etc suivre le lien <https://www.kaggle.com/t/c8089767d4d94efbb62b54e4923791ab>. Vous devrez vous inscrire pour participer au concours.

Un **exemple de code est disponible sur moodle** pour vous aider à démarrer. Il est proposé au format notebook et au format script python. Ce code déroule différentes étapes du concours : chargement des données, préparation des données, mise en oeuvre de la régression linéaire en utilisant la librairie **scikit learn**, mise en oeuvre d'un réseau de neurones en utilisant la librairie **scikit learn**, mise en oeuvre d'un réseau de neurones en utilisant la librairie **pytorch** pour ceux qui veulent aller plus loin.

En pratique, ce code doit être copié sur kaggle via un onglet ouvert dans le site de la compétition.

Votre travail ? Obtenir la meilleure prédiction (sur l'ensemble de test) en faisant varier les hyper-paramètres (architecture du réseau de neurones, taux d'apprentissage de l'algorithme de descente du gradient, etc). Vous pouvez aussi intervenir sur les variables explicatives ou essayer d'utiliser d'autres algorithmes si vous le souhaitez.

L'évaluation de ce TP sera fait sur la base de votre participation au concours, des scores obtenus et du code implémenté.