

Ce document réunit le fruit de nos recherches bibliographiques. Dans un premier temps nous avons lu des articles généralistes sur les IA de prédiction du top 50. Une fois quelques connaissances et quelques mots de vocabulaire acquis, nous avons cherché des articles scientifiques afin de voir leurs approches et nous aider à sélectionner les paramètres importants pour trouver une musique populaire.

I Articles de vulgarisation, journaux

Hitwizard :

articles généraux sur ce projet d'une IA qui devine le top 50 sous forme d'application

- <https://www.numerama.com/tech/299256-hitwizard.html> : davantage focus sur l'aspect technique - rythmique et pulsations/min
- <https://www.tomsguide.fr/une-ia-capable-de-deviner-vos-gouts-musicaux/> : résultats - publiés sur Arxiv.org
 - elle enregistrait les options préférées des utilisateurs : le tempo, le volume ou encore la popularité des artistes.
 - Un modèle fiable à plus de 90%
 - répondre aux questions du Goldsmiths Musical Sophistication Index (Gold-MSI) prend en compte : l'engagement actif (le temps et l'argent consacrés à la musique) les émotions (les réactions émotionnelles engendrées)
 - 61 sujets, l'IA a réussi à prédire les émotions dans 95% des cas et elle est parvenue à déterminer les préférences de 56 personnes
 - Hitwizard avait trouvé 66% des hits, il était plus précis avec les flops, la précision avait atteint 93%.
- <https://www.jedha.co/blog/predire-le-succes-dune-chanson-sur-spotify> (article plus général, une personne qui s'est visiblement lancé le même type d'exercice que nous, je pense qu'il y a des infos à tirer :
 - popularité noté de 1 à 100, avec une répartition vu en histogramme
 - si popularité=0, conseille de supprimer
 - nettoyage des données, notamment supprimer des valeurs de durée extrême, pas très représentatif de l'échantillon
 - voir si les noms d'artistes sont pertinents, des fois l'échantillon montre une répartition pas du tout homogène de chaque morceau d'un même artiste
 - dans ce cas de figure, les données gardées sont : accoustiness, danceability, energy, liveness, key, duration_ms, loudness, mode, tempo, time-signature, valence, instrumentality, speechiness
 - Il nous faudra choisir un modèle, est-ce qu'on considère comme 'hit' une chanson qui dépasse 60/100 ? 70/100 ? 80/100 ?
 - attention à ne pas tomber dans un modèle qui nous donnerait un échantillon très faible - de 2% ...
 - évoque aussi un problème de précision, mais je n'en comprend pas plus et je ne sais pas si c'est notre priorité en si peu de temps et vu notre niveau
 - Écrit par Gabriel Ropers - Senior Consultant - Data Analyst)

II Littérature scientifique

- Ni, Y., Santos-Rodriguez, R., Mcvicar, M., & De Bie, T. (2011, December). Hit song science once again a science. In *4th International Workshop on Machine Learning and Music*.

lien:<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.412.9732&rep=rep1&type=pdf>

1. the new century music listeners prefer faster songs.
2. In every decade simple songs are usually hits. This trend is captured by the Time Signature Binary and Harmonic Simplicity features.
3. There is a clear tendency for music to become relatively longer, indicated by the evolving importance of weights of duration.
4. Hits are getting relatively louder. The increasing importance through time shows that the loudness of a song is becoming more useful at distinguishing a hit from a non-hit.
Our more optimistic results can be explained in various ways. The first is that the task description is different - we are trying to distinguish top 5 hits from top 30-40 hits whereas other authors attempted to learn a label of low, medium or high popularity

- Borg, N., & Hokkanen, G. (2011). What makes for a hit pop song? What makes for a pop song. *Unpublished Thesis, Stanford University, California, USA*.

lien:<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.374.8012&rep=rep1&type=pdf>

Point autour de la popularité, une fois qu'un artiste est dit populaire, il peut produire une musique d'un genre différent mais ce nouveau son peut devenir un hit plus tard par le simple fait de la popularité de l'artiste. Voir, devenir potentiellement un top 10 des genres populaires. Évoque la notion de random forest - un arbre de décision, choisit aléatoirement des caractéristiques et fait des observations.

- Reiman, M., & Örnell, P. (2018). Predicting hit songs with machine learning. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1214146&dsid=-8354>
La page 16 est la plus intéressante car elle définit les paramètres mentionnés dans les tables SQL.
- Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The million song dataset. <https://academiccommons.columbia.edu/doi/10.7916/D8NZ8J07>
As an illustration, we present year prediction as an example application, a task that has, until now, been difficult to study owing to the absence of a large set of suitable data. We show positive results on year prediction, and discuss more generally the future development of the dataset.2012

Synthèse

- le tempo de la chanson : ex jusqu'en 90 c'était les **tempos** longs qui étaient en tête maintenant...
- la simplicité de la musique, + elle l'est, + elle cartonne, représentée par : **time signature binary** and **harmonic simplicity features**.
- tendance pour les musiques à devenir longues : **duration**
- loudness** : the more the better through years

- une fois qu'un artiste est dit populaire, il peut produire une musique d'un genre différent mais ce nouveau son peut devenir un hit plus tard par le simple fait de la popularité de l'artiste **artist popularity**
- Importance d'étudier par années pour avoir les tendances