

M1 - ISD

Rapport de projet ML

Prédiction de popularité des chansons

Sarah IKHLEF

Sommaire

| | |
|--|----------|
| 1. Objectif du projet | 3 |
| 2. Données | 3 |
| 3. Variables d'entrée | 3 |
| 4. Variable de sortie | 3 |
| 5. Techniques | 3 |
| 6. Bilan de l'utilisation de ChatGPT | 4 |
| 7. Annexes: Conversation avec ChatGPT | 5 |

1. Objectif du projet

Ce projet vise à développer un modèle de prédiction de “popularité” des chansons Spotify en utilisant des techniques d'apprentissage supervisé (Régression, Decision Tree, Random Forests).

2. Données

Les données utilisées sont récupérées sur Kaggle sous format CSV :

<https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>

Le dataset comporte 20 features (colonnes) et 114000 lignes, chaque ligne représente une musique.

3. Variables d'entrée

Les différentes caractéristiques détaillées dans le lien ci-dessus (Artistes, album, durée de la chanson, énergie, ... etc).

4. Variable de sortie

Attribut “popularité” de la chanson qui prend ses valeurs entre 0 et 100.

5. Techniques

Ce projet est réalisé à l'aide de ChatGPT:

- Idée du projet
- Spécifications
- Étapes à suivre
- Code (à 90%) ... etc


6. Bilan de l'utilisation de ChatGPT


| <u>Points positifs</u> | <u>Points négatifs</u> | <u>Difficultés</u> |
|--|------------------------|--------------------|
| <ul style="list-style-type: none">- Il donne le code + explications- c'est lui qui a eu l'idée du projet + les spécification + les étapes à suivre- Il donne des conclusions sur les résultats obtenus | - | |




7. Annexes: Conversation avec ChatGPT

Les détails de la conversation avec ChatGPT est résumé ci dessous:


1) Choix du projet:

 Pour mon projet Machine Learning , j'ai pris ce data set sur Kaggle :
<https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset> pourriez vous me proposer des idées de projets à faire avec ?

 Bien sûr ! Voici quelques idées de projets que vous pouvez réaliser avec le jeu de données Spotify Tracks :

1. Prédiction de la popularité des chansons : Utilisez les différentes caractéristiques des chansons (comme l'énergie, la danseabilité, la durée, etc.) pour prédire leur popularité. Vous pouvez entraîner un modèle de régression pour prédire le nombre d'écoutes ou utiliser un modèle de classification pour prédire si une chanson sera populaire ou non.
2. Recommandation de chansons : Créez un système de recommandation de chansons personnalisées en utilisant des techniques de filtrage collaboratif ou de recommandation basée sur le contenu. Vous pouvez recommander des chansons similaires à celles aimées par un utilisateur donné en analysant les caractéristiques des chansons et en trouvant des correspondances.
3. Classification de genre musical : Utilisez les informations disponibles sur les chansons (tempo, timbre, mode, etc.) pour créer un modèle de classification capable de prédire le genre musical d'une chanson. Vous pouvez entraîner un modèle d'apprentissage automatique pour classer les chansons dans des catégories prédéfinies telles que le rock, le hip-hop, la musique élec
4. Analyse des tendances musical

 Regenerate response

2) Type apprentissage à utiliser:



Pour votre projet d'apprentissage utilisant le dataset Spotify Tracks, vous pouvez envisager d'utiliser une approche d'apprentissage supervisé. L'apprentissage supervisé consiste à entraîner un modèle sur un ensemble de données étiquetées (dans ce cas, les chansons avec leur popularité) afin qu'il puisse prédire la popularité des chansons non étiquetées.



Voici quelques algorithmes d'apprentissage supervisé que vous pouvez envisager pour ce projet :

1. **Régression linéaire** : Ce modèle permet de prédire une valeur continue (la popularité) en fonction des caractéristiques musicales. Il est utile lorsque vous souhaitez établir une relation linéaire entre les fonctionnalités et la popularité.
2. **Arbres de décision** : Les arbres de décision peuvent capturer des relations non linéaires entre les fonctionnalités et la popularité. Ils sont faciles à interpréter et peuvent être utilisés pour extraire des règles décisionnelles.
3. **Forêts aléatoires** : Les forêts aléatoires combinent plusieurs arbres de décision pour obtenir des prédictions plus robustes et réduire le surajustement. Elles peuvent être efficaces pour des ensembles de données complexes.
4. **Machines à vecteurs de support (SVM)** : Les SVM sont utilisées pour trouver une séparation optimale entre deux classes. Dans ce cas, vous pouvez les utiliser pour classer les chansons en fonction de leur popularité.
5. **Réseaux de neurones** : Les réseaux de neurones peuvent capturer des relations complexes entre les caractéristiques et la popularité. Ils sont particulièrement adaptés lorsque les données sont très riches et présentent des interactions non linéaires.

3) Spécifications du projet:



Certainement ! Voici les spécifications générales pour le projet de prédiction de popularité des chansons :



1. Collecte des données : Téléchargez le jeu de données Spotify Tracks à partir de Kaggle (<https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>) et familiarisez-vous avec sa structure et ses variables. Assurez-vous de comprendre la signification de chaque attribut.
2. Prétraitement des données : Effectuez une exploration des données et nettoyez-les si nécessaire. Traitez les valeurs manquantes, supprimez les doublons et convertissez les attributs pertinents dans le format approprié. Divisez le jeu de données en ensembles d'entraînement et de test.
3. Sélection des caractéristiques : Identifiez les attributs les plus pertinents pour prédire la popularité des chansons. Vous pouvez utiliser des techniques d'analyse statistique ou d'importance des variables pour effectuer cette sélection. Assurez-vous de conserver les attributs qui sont disponibles au moment de prédire la popularité d'une chanson.
4. Entraînement du modèle : Choisissez un algorithme d'apprentissage automatique adapté à votre problème de prédiction de popularité, comme la régression linéaire, la régression logistique, les arbres de décision, les forêts aléatoires, ou les réseaux de neurones. Entraînez votre modèle en utilisant l'ensemble d'entraînement et les attributs sélectionnés.

3. Sélection des caractéristiques : Identifiez les attributs les plus pertinents pour prédire la popularité des chansons. Vous pouvez utiliser des techniques d'analyse statistique ou d'importance des variables pour effectuer cette sélection. Assurez-vous de conserver les attributs qui sont disponibles au moment de prédire la popularité d'une chanson.
4. Entraînement du modèle : Choisissez un algorithme d'apprentissage automatique adapté à votre problème de prédiction de popularité, comme la régression linéaire, la régression logistique, les arbres de décision, les forêts aléatoires, ou les réseaux de neurones. Entraînez votre modèle en utilisant l'ensemble d'entraînement et les attributs sélectionnés.
5. Évaluation du modèle : Évaluez les performances de votre modèle en utilisant l'ensemble de test. Calculez des métriques telles que l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R^2), ou toute autre mesure appropriée pour évaluer la qualité de votre prédiction.
6. Optimisation du modèle : Si les performances de votre modèle ne sont pas satisfaisantes, explorez différentes techniques d'optimisation. Vous pouvez essayer différentes combinaisons d'attributs, utiliser des techniques de réduction de dimension comme l'analyse en composantes principales (PCA), ou ajuster les hyperparamètres de votre algorithme.
7. Prédiction de la popularité : Utilisez votre modèle entraîné pour faire des prédictions de popularité sur de nouvelles chansons ou sur des chansons du jeu de données qui n'ont pas été utilisées pour l'entraînement. Évaluez la précision de vos prédictions et comparez-les aux valeurs réelles de popularité si disponibles.

4) Code :