



Projeto da Disciplina

Aplicação de Inteligência Artificial em Problema Real

Individual ou em Grupo

1) Título: Análise de Fatores que Influenciam a Popularidade das Músicas no Spotify:
Uma Abordagem Baseada em Características de Áudio, Gêneros e Artistas

2) Integrantes: Isabella Abreu Comelli (10400807@mackenzista.com.br), Pedro Henrique Wege Barbosa (10400851@mackenzista.com.br)

Link do Github: https://github.com/pedrohwbarbosa/projeto_IA_mack

3) Resumo

Este projeto tem como objetivo identificar os fatores que influenciam a popularidade das músicas no Spotify, com foco nas características de áudio presentes no banco de dados da plataforma. Além disso, busca-se explorar a influência de gêneros musicais e artistas, complementando a análise sonora. Para o estudo, foram selecionadas as 100 músicas mais populares do Spotify, permitindo uma investigação detalhada sobre os elementos que contribuem para seu sucesso.

Os módulos pandas e scikit-learn, disponíveis na linguagem de programação Python, serão utilizados para a manipulação dos dados e o treinamento de modelos de aprendizado de máquina, analisando atributos como energia, valência e tempo. A eficácia do modelo será avaliada com base na precisão das associações identificadas entre as características sonoras e a popularidade das faixas.

A pesquisa busca contribuir para a compreensão dos padrões que tornam certas músicas mais populares.

4) Introdução

A popularidade de uma música no Spotify é influenciada por uma série de fatores, desde aspectos subjetivos, como preferências do público e tendências culturais, até características mais técnicas, como sua estrutura sonora. Com o avanço das tecnologias



UNIVERSIDADE PRESBITERIANA MACKENZIE

Faculdade de Computação e Informática

Prof. Dr. Leandro Zerbinatti

Inteligência Artificial – 7ºJ SI - Noite



de análise de dados, tornou-se possível identificar padrões que contribuem para o sucesso de determinadas faixas na plataforma. A compreensão desses padrões pode beneficiar tanto artistas e produtores, que buscam otimizar suas produções para alcançar um público maior, quanto o próprio Spotify, que pode aprimorar suas estratégias de recomendação e marketing.

O mercado musical está em constante transformação, e entender os fatores que impulsionam o sucesso de uma música pode ser um diferencial para artistas, gravadoras e plataformas de streaming. Identificar características acústicas comuns em músicas populares pode fornecer insights valiosos sobre tendências sonoras e preferências do público. Além disso, a análise da influência de gêneros musicais e da notoriedade dos artistas na popularidade das músicas pode auxiliar na formulação de estratégias de lançamento e promoção. Ao utilizar dados públicos do Spotify e técnicas de aprendizado de máquina, este estudo busca contribuir para uma melhor compreensão do que torna uma música popular.

Este projeto tem como objetivo principal identificar os fatores que influenciam a popularidade das músicas no Spotify, com ênfase nas características de áudio extraídas diretamente da plataforma. Além disso, busca-se explorar a relação entre gêneros musicais e a popularidade, bem como o impacto do reconhecimento prévio dos artistas no sucesso de suas faixas. Para isso, foram selecionadas as 100 músicas mais populares do Spotify, permitindo uma investigação detalhada dos elementos que contribuem para seu sucesso.

O estudo será conduzido por meio da coleta e análise de dados extraídos da API pública do Spotify, organizados em dois conjuntos principais: um contendo informações gerais das músicas, como nome, artista e popularidade, e outro focado nas características de áudio, incluindo dançabilidade, energia e tonalidade. As técnicas de aprendizado de máquina serão aplicadas para identificar padrões entre esses atributos e o sucesso das músicas. O projeto será implementado na linguagem Python, utilizando bibliotecas como pandas, numpy, matplotlib e scikit-learn para manipulação de dados, visualização e modelagem preditiva.

5) Descrição do Problema

A popularidade das músicas no Spotify é um fenômeno complexo, influenciado por diversos fatores, desde a exposição midiática até as características sonoras das faixas. Com o crescimento das plataformas de streaming, compreender o que torna uma música popular tornou-se uma questão estratégica para artistas, gravadoras e para o próprio



UNIVERSIDADE PRESBITERIANA MACKENZIE
Faculdade de Computação e Informática
Prof. Dr. Leandro Zerbinatti
Inteligência Artificial – 7ºJ SI - Noite



Spotify. No entanto, a definição de popularidade muitas vezes é subjetiva e depende de uma combinação de fatores que vão além do simples número de reproduções.

As características de áudio, como energia, dançabilidade e valência, podem desempenhar um papel significativo na recepção do público. Entretanto, a relação entre esses atributos e o sucesso de uma música ainda não é totalmente compreendida. Da mesma forma, a influência dos gêneros musicais e da notoriedade dos artistas na popularidade das faixas é um aspecto relevante que carece de uma análise aprofundada.

Este estudo busca responder a questões fundamentais: Quais características sonoras estão mais associadas à popularidade das músicas? Existe um padrão comum entre as faixas que alcançam maior sucesso? A notoriedade do artista impacta diretamente a popularidade da música, independentemente de suas características sonoras? Para responder a essas perguntas, será utilizada a API do Spotify, permitindo a extração e análise de dados de músicas populares na plataforma.

A utilização de APIs como ferramenta de extração de dados tem se tornado uma prática comum em diversas áreas, incluindo a análise musical. Segundo OpenAI (2024) e Camelo (2023), as APIs possibilitam acesso a grandes volumes de informações, permitindo automação e processamento eficiente dos dados coletados. No contexto deste projeto, a API do Spotify será fundamental para coletar dados sobre as músicas e suas características sonoras, viabilizando uma abordagem analítica baseada em evidências.

Dessa forma, este estudo visa explorar a relação entre atributos sonoros e popularidade no Spotify, utilizando aprendizado de máquina e análise estatística para identificar padrões e tendências que possam contribuir para uma melhor compreensão do sucesso musical na era do streaming.

6) Discutir a respeito dos aspectos Éticos do uso da IA e a sua Responsabilidade no desenvolvimento da solução

O uso de inteligência artificial (IA) para análise de dados e tomada de decisões tem se tornado cada vez mais comum, trazendo benefícios significativos para diversas áreas, incluindo o setor musical. No entanto, a aplicação dessas tecnologias levanta questões éticas importantes, especialmente no que diz respeito à transparência, vies algorítmico e impacto na indústria da música.

Um dos principais desafios éticos na implementação de soluções baseadas em IA é garantir que os modelos utilizados sejam justos e imparciais. Algoritmos treinados com dados históricos podem perpetuar padrões existentes, favorecendo determinados



UNIVERSIDADE PRESBITERIANA MACKENZIE

Faculdade de Computação e Informática

Prof. Dr. Leandro Zerbinatti

Inteligência Artificial – 7ºJ SI - Noite



gêneros musicais, artistas ou características sonoras em detrimento de outros. Isso pode reforçar desigualdades no mercado, prejudicando músicos independentes ou gêneros menos populares. Para mitigar esse risco, é fundamental que os dados utilizados no estudo sejam representativos e que haja um monitoramento contínuo dos resultados gerados pela IA.

Além disso, a transparência no desenvolvimento da solução é um aspecto essencial. Usuários e profissionais da indústria musical devem compreender os critérios utilizados pelos modelos para determinar padrões de popularidade. Soluções de IA que influenciam a recomendação de músicas, por exemplo, podem impactar diretamente a visibilidade de artistas, tornando necessário um compromisso ético na construção dos algoritmos. De acordo com OpenAI (2024) e Camelo (2023), o uso de APIs e IA deve ser pautado por boas práticas, garantindo que as decisões automatizadas sejam explicáveis e auditáveis.

Outro ponto relevante é a responsabilidade no uso dos dados. Como o projeto envolve a coleta de informações por meio da API do Spotify, é essencial garantir a conformidade com políticas de privacidade e uso responsável dos dados. A anonimização e o respeito às diretrizes da plataforma são medidas necessárias para evitar o uso indevido das informações.

Por fim, a implementação da IA na análise de popularidade musical deve ser feita com o compromisso de agregar valor à indústria da música, sem substituir a criatividade humana. O objetivo não é criar um modelo que dite o que deve ser produzido, mas sim fornecer insights que possam auxiliar artistas, produtores e plataformas de streaming na compreensão das tendências musicais de forma ética e responsável.

7) Dataset, se for o caso (anonimizados quando necessário), descrição detalhada do seu conteúdo/origem, análise exploratória e preparação dos dados em Python;

Os dados utilizados neste estudo foram coletados por meio da API pública do Spotify, sendo organizados em dois dataframes principais: `df_tracks` e `df_audio_features`. O primeiro contém informações gerais sobre as músicas, como título, popularidade, gênero musical, duração, nome do artista e do álbum. Já o segundo abrange as características de áudio de cada faixa, incluindo atributos como dançabilidade, energia, tom e instrumentalidade.

A base de dados extraída da API apresentou uma qualidade satisfatória, com informações completas para a maioria das músicas. Foram identificadas poucas ocorrências de valores ausentes, principalmente em colunas relacionadas a álbuns, e



UNIVERSIDADE PRESBITERIANA MACKENZIE

Faculdade de Computação e Informática

Prof. Dr. Leandro Zerbinatti

Inteligência Artificial – 7ºJ SI - Noite



algumas duplicatas que precisaram ser removidas. Como o foco da análise é a relação entre características de áudio e popularidade, os outliers foram considerados músicas populares que não apresentavam padrões acústicos similares às demais faixas de destaque.

Durante a etapa de pré-processamento, algumas transformações foram aplicadas para garantir a padronização e a qualidade dos dados. No dataframe `df_tracks`, a coluna `track_explicit` foi convertida de um formato booleano para inteiro, e os subgêneros musicais foram agrupados dentro de gêneros principais, como no caso de synth-pop, que foi classificado simplesmente como pop. Além disso, todas as linhas duplicadas foram removidas.

Após a junção dos dois dataframes, algumas colunas numéricas que continham valores ausentes foram preenchidas com zero, garantindo a consistência do dataset. Foi realizada a conversão de valores em notação científica na coluna `track_instrumentalness`, transformando-os em valores de ponto flutuante (float). Também foi criada a coluna `track_main_genre_id`, que classifica as músicas de acordo com gêneros musicais padronizados. Algumas colunas foram descartadas por não agregarem valor à análise, como informações sobre álbuns e subgêneros, além de atributos de áudio como `track_mode`, `track_energy`, `track_acousticness`, `track_liveness`, `track_valence`, `track_loudness` e `track_time_signature`.

Para facilitar a análise dos dados, algumas colunas foram convertidas em variáveis categóricas. Atributos como `track_danceability`, `track_speechiness` e `track_instrumentalness` foram segmentados em duas categorias, sendo classificados com valores binários (0 ou 1) com base em seus percentis medianos. Além disso, a coluna `track_popularity` foi transformada em categorias de popularidade, agrupando os valores por faixas. Por exemplo, uma música com popularidade 85 foi classificada na "casa dos 80" e recebeu o valor 8, enquanto uma faixa com popularidade 74 foi categorizada na "casa dos 70" e recebeu o valor 7.

A preparação dos dados foi conduzida utilizando Python e as bibliotecas `pandas`, `numpy` e `matplotlib`, permitindo a manipulação eficiente das tabelas, a aplicação de transformações e a visualização inicial dos padrões nos dados. A análise exploratória incluiu estatísticas descritivas para identificar distribuições e possíveis correlações entre os atributos de áudio e a popularidade das músicas. Esse processo foi fundamental para garantir a qualidade dos dados utilizados no estudo e a confiabilidade dos resultados obtidos.



8) Metodologia e Resultados Esperados: apresentar a abordagem que pretende empregar na resolução do problema e quais são os resultados esperados.

A metodologia adotada para este estudo baseia-se em análise de dados e aprendizado de máquina para identificar padrões entre as características de áudio das músicas e sua popularidade no Spotify. O primeiro passo foi a coleta de dados por meio da API pública do Spotify, que forneceu informações detalhadas sobre faixas musicais, incluindo atributos sonoros e métricas de popularidade. Após a obtenção dos dados, foi realizada uma etapa de pré-processamento para padronizar os valores, eliminar duplicatas e categorizar variáveis de interesse.

A análise exploratória buscou compreender a relação entre elementos como dançabilidade, tom e batidas com a popularidade das músicas. Com base nesses insights, foram aplicados algoritmos de aprendizado de máquina para prever a probabilidade de uma música se tornar popular. Para isso, foram utilizados os modelos K-Nearest Neighbors (KNN) e Decision Tree, escolhidos pela sua capacidade de lidar com classificações complexas e interpretar padrões nos dados.

Os resultados demonstraram que músicas com maior energia e dançabilidade tendem a alcançar maior popularidade. Além disso, faixas do gênero pop se destacam, independentemente de serem exclusivamente desse estilo ou de possuírem fusões com outros gêneros musicais. No entanto, prever com exatidão quais músicas se tornarão sucessos continua sendo um desafio, pois fatores externos, como a fama do artista e tendências do mercado, exercem grande influência.

Os modelos apresentaram taxas de acurácia entre 66% e 73%, sendo que, mesmo ao embaralhar os dados, a estabilidade dos resultados se manteve em torno de 70%. Esse desempenho foi otimizado através da categorização dos dados, que facilitou o aprendizado dos algoritmos. Apesar da variação nos valores e da presença de alguns dados singulares, os resultados foram considerados satisfatórios, pois as características selecionadas para a análise estavam entre as mais relevantes para a popularidade das músicas.

Espera-se que a abordagem adotada contribua para a compreensão dos fatores que influenciam o sucesso de uma música, oferecendo insights para artistas, produtores e plataformas de streaming sobre quais elementos podem aumentar a aceitação do público. Além disso, a metodologia utilizada pode ser aprimorada no futuro com o uso de técnicas mais avançadas, como redes neurais e análise de sentimentos das letras das músicas, possibilitando previsões ainda mais precisas.



UNIVERSIDADE PRESBITERIANA MACKENZIE
Faculdade de Computação e Informática
Prof. Dr. Leandro Zerbinatti
Inteligência Artificial – 7ºJ SI - Noite



9) Referências: citadas dentro do texto do projeto

CAMELO, L. API do ChatGPT: como usar todo o poder da inteligência artificial? Agosto. 2023. Disponível em: <https://pluga.co/blog/api-chatgpt/>. Acesso em: 10 de fevereiro de 2025.

OpenAI. How can I access the ChatGPT API? Agosto. 2024. Disponível em: <https://help.openai.com/en/articles/7039783-how-can-i-access-the-chatgpt-api>.

SPOTIFY FOR DEVELOPERS. Web API. Retrieve metadata from Spotify content, control playback or get recommendations. Sweden: Spotify, 2024. Disponível em: <https://developer.spotify.com/documentation/web-api>.

10) Bibliografia: na qual o grupo se baseará para criar e desenvolver o projeto e/ou Referências: citadas dentro do texto do projeto.

Pandas documentation. Pandas, 20 set. 2024. Disponível em: <https://pandas.pydata.org/docs/index.html>.

Matplotlib 3.9.2 documentation. Matplotlib, 15 mai. 2024. Disponível em: <https://matplotlib.org/stable/index.html>.

API Reference. Scikit-learn, 1 set. 2024. Disponível em: <https://scikit-learn.org/stable/api/index.html>.