**UNIVERSIDADE VIRTUAL DO ESTADO DE SÃO PAULO**

LEONARDO SILVA FRANCISCO

**SISTEMA DE RECOMENDAÇÃO DE MÚSICAS BASEADO EM CLUSTER**

**MAUÁ**

**2023**

**UNIVERSIDADE VIRTUAL DO ESTADO DE SÃO PAULO**

**SISTEMA DE RECOMENDAÇÃO DE MÚSICAS BASEADO EM CLUSTER**

Relatório Técnico - Científico apresentado na disciplina de Projeto Integrador para os cursos do Eixo de Computação da Universidade Virtual do Estado de São Paulo (UNIVESP).

**MAUÁ**

**2023**

FRANCISCO, Leonardo Silva; **Sistema de recomendação de músicas baseado em Cluster.** 00f. Relatório Técnico-Científico. Eixo de Computação – **Universidade Virtual do Estado de São Paulo**. Tutor: Guilherme Bugatti dos Santos. Polo Mauá, 2023.

**RESUMO**

Este projeto tem como objetivo a implementação de uma interface visual desenvolvida na linguagem Python, afim de otimizar o sistema de recomendação de músicas da plataforma de streaming Spotify. Para melhor entendimento desta demanda, foi realizada uma pesquisa com usuários assíduos da plataforma, com o intuito de obter insumos que guiassem o andamento do trabalho. Com isso, foi avaliada a necessidade da criação de uma interface para apresentar as recomendações feitas a partir da análise dos dados e do uso de algoritmos de agrupamento de acordo com as variáveis que julgarmos mais relevantes. Em seguida foi feita a pesquisa bibliográfica para embasar aplicação de conceitos aprendidos durante o curso, e assim iniciar a construção do sistema.

**PALAVRAS-CHAVE:** Análise de dados; aprendizado de máquina; Interface visual; recomendação.

**SUMÁRIO**

[1. INTRODUÇÃO 5](#_Toc133070181)

[2. DESENVOLVIMENTO 6](#_Toc133070182)

[2.1. OBJETIVOS 6](#_Toc133070183)

[2.2. JUSTIFICATIVA E DELIMITAÇÃO DO PROBLEMA 6](#_Toc133070184)

[2.3. FUNDAMENTAÇÃO TEÓRICA 7](#_Toc133070185)

[2.4. METODOLOGIA 8](#_Toc133070186)

[3. REFERÊNCIAS BIBLIOGRÁFICAS 10](#_Toc133070187)

# INTRODUÇÃO

Com o passar dos anos e com a concorrência cada vez mais acirrada, o ramo da tecnologia se vê desafiado constantemente. Novas ferramentas devem ser desenvolvidas afim de aprimorar cada vez mais os serviços ofertados, e nem sempre é possível de se realizar toda otimização dos produtos manualmente, com isso o uso do aprendizado de máquinas vem se mostrando cada vez mais essencial.

A solução apresentada por este relatório propõe a utilização de uma interface que possa apresentar aos usuários listas personalizadas baseadas em grupos de similaridades utilizando um algoritmo de agrupamento.

Foram analisadas algumas das ferramentas que possibilitariam a realização desse projeto. Por fim, foi definido o uso da linguagem Python, através do ambiente Jupyter Notebook para realização das análises e uso de aprendizado de máquina, e da IDE PyCharm que é um ambiente de desenvolvimento integrado usado para programação em Python. No desenvolvimento desta aplicação usaremos o algoritmo de aprendizado de máquina chamado K-means, que realiza o agrupamento dos dados. E para interface visual, iremos utilizar a biblioteca TKinter que permite a apresentação dos dados em uma janela pop-up.

# DESENVOLVIMENTO

# OBJETIVOS

Objetivo Geral:

Elaborar uma interface utilizando a biblioteca Tkinter do Python, que apresente os dados já tratados pós análise, posteriormente sejam validadosutilizando o algoritmo k-means do contexto de aprendizado de máquinas para o agrupamento de acordo com sua similaridade.

Objetivo Específico:

* Realizar o tratamento e análise dos dados;
* Elaborar a interface e aplicar o algoritmo;
* Analisar os resultados da aplicação;
* Aprimorar o sistema de recomendação das músicas.

# JUSTIFICATIVA E DELIMITAÇÃO DO PROBLEMA

Para entender o problema específico que orienta o presente Projeto Integrador (PI) foi necessário observar a atividade de alguns usuários na plataforma e se estes avaliam positivamente o que o serviço entrega no quesito de indicação para o ouvinte de acordo com seu gosto musical. O Spotify é a plataforma de streaming de músicas com o maior número de usuários do mundo e é uma das referências no uso de Ciência de Dados para potencializar a experiência do usuário.

Esses recursos de recomendação de faixas só existem em versões pagas dos serviços de streaming. Mediante isso, existe a necessidade de ter um sistema assertivo e que compense o valor pago.

Continuando com a análise dos pontos levantados em diálogo com a comunidade foi identificado o incômodo com algumas das recomendações da aplicação, e a partir disso traçar um plano de ação que resulte em uma interface concisa que reúna tais informações, a fim de elencá-las e dar o devido tratamento a elas, fazendo com que estas cheguem de maneira mais personalizada ao público.

Ao longo das entrevistas com notamos que uma parte dos entrevistados não estava satisfeita com a sequência automática reproduzida pela aplicação e foi sugerida a criação de um sistema similar para solução do problema, mas que realize tal tarefa de forma automatizada e que aprimore gradativamente sua função.

Esta aplicação tornará a experiência e satisfação do usuário cada vez maior já que o sistema irá se aperfeiçoar de acordo com o uso.

# FUNDAMENTAÇÃO TEÓRICA

O Spotify, é o líder mundial em popularidade entre os serviços de streaming de áudio. Porém seus concorrentes estão se aproximando cada vez mais.

Lançado em 2008, numa época onde a pirataria dominava a internet, o serviço veio para competir com outras opções legais que era muito caras ou limitadas. O intuito da plataforma foi proporcionar aos usuários uma plataforma de fácil manejo, com um enorme acervo de músicas e que fosse acessível. Para artistas e gravadoras, uma alternativa para aumentar seus ganhos e adentrarem o mundo digital (COUTINHO, 2019).

Corroborando com o levantamento de Coutinho (2019), o serviço tem a opção da versão gratuita, onde são reproduzidos anúncios durante os áudios e o usuário não pode pular músicas, e a versão paga, onde os assinantes tem maior autonomia nas reproduções podendo escutar na ordem que optarem e sem anúncios. Além da opção de fazer download dos áudios para escutarem offline.

E dados sempre fizeram parte da cultura do Spotify, a empresa sempre teve times robustos e dedicados a fazerem suas análises. Com isso, os dados dão suporte à tomada de decisões, informações para previsões e segmentação para ofertas de publicidade. E a maior parte das ações são centradas na pessoa usuária, com isso a análise e ciência dos dados vêm forte para fornecer as recomendações, por exemplo, escolher a próxima faixa musical a ser executada (REDAÇÃO TERA, 2018).

Quando se trata de tratamento e análise de dados, devemos nos atentar a qual estratégia adotar para tais tarefas. Ao partirmos para uma análise quantitativa, infere-se que trataremos um grande número de variáveis e de observações. Por outro lado, se utilizarmos uma abordagem qualitativa devemos buscar técnicas mais aprofundadas que possibilitem uma visão mais ampla do que se parece tão vasto. Deste modo, a análise quantitativa busca instrumentalizar conceitos, alinhar relações de causalidade e generalizar conclusões. Enquanto a abordagem qualitativa opera por um modo de reforço e acúmulo de informações que proporcionem a generalização prática das conclusões (RODRIGUES, 2012, p. 172).

Como ferramentas mais utilizadas atualmente, no contexto de análise de dados, podemos citar a linguagem de programação Python, o Microsoft Excel, o Microsoft Power BI e o Tableau Server.

Afim de auxiliar no entendimento dos dados utilizaremos o conceito de aprendizado de máquina juntamente com a análise dos dados.

# METODOLOGIA

Para a realização desse Projeto Integrador (PI), adotou-se a metodologia de resolução de problemas baseada nos conceitos do *Design Thinking* - ouvir, criar, prototipar e testar - além da realização de pesquisa bibliográfica. Em todo o desenvolvimento do trabalho foram respeitadas as questões éticas em pesquisas com seres humanos.

Num primeiro momento de reflexão, partindo-se do tema definido pela UNIVESP “Desenvolver análise de dados em escala utilizando algum conjunto de dados existente e aprendizado de máquina”, procurou-se definir qual seria o público-alvo do projeto. Após algumas entrevistas, definiu-se que a proposta seria aplicada junto a um público assíduo da plataforma Spotify.

Partindo-se dessa primeira definição, para selecionar o problema específico a ser trabalhado o grupo utilizou-se de análise os dados disponibilizados no site Kaggle (www.kaggle.com) onde encontram-se diversos tipos de base de dados, e esses dados encontram-se geralmente em formato CSV (valores separados por vírgula) e foi esse o formato do dataset que iremos utilizar.

Como alternativa para exibição deste dataset, foi utilizado o software Anaconda Navigator, e em uma de suas extensões usamos o Jupyter Notebook que é um ambiente para desenvolvimento em Python amplamente utilizado por Cientistas de Dados.

Após essa etapa de delimitação do problema, a ideaçãofoi feitalevantando possíveis estratégias a serem utilizadas na construção do protótipo, dando-se andamento ao processo de criação. Nessa etapa, foram avaliadas as condições específicas apontadas no tema proposto pela UNIVESP para esse Projeto Integrador, em especial a necessidade do uso de aprendizado de máquina.

Paralelamente, iniciou-se também um processo de imersão na temática definida, por meio de uma pesquisa bibliográfica, para levantamento dos principais estudos em torno do tema. Nesse processo, buscaram-se referências em fontes como artigos científicos, livros, revistas especializadas, dissertações e teses.

# REFERÊNCIAS BIBLIOGRÁFICAS

RODRIGUES, Miguel Ângelo Vilela. O tratamento e análise de dados. *In*: SILVESTRE, Hugo Consciência; ARAÚJO, Joaquim Filipe. **Metodologia para a Investigação Social**. Lisboa - Portugal: Escolar Editora, 2012. cap. 9, p. 171-172. ISBN 978-972-592-329-0.

COUTINHO, Thiago. **Você conhece a história do Spotify? Aprenda tudo sobre a plataforma!.** 2. 2. ed. [S. l.]: Thiago Coutinho, 11 jun. 2019. Disponível em: <https://www.voitto.com.br/blog/artigo/spotify>. Acesso em: 15 abr. 2023.

REDAÇÃO TERA (Brasil). **Como o Spotify usa Ciência de Dados para engajar usuários**. 1. 1. ed. Brasil: Redação Tera, 30 jul. 2018. Disponível em: https://blog.somostera.com/data-science/como-spotify-usa-ciencia-de-dados-para-engajar-usuarios. Acesso em: 17 abr. 2023.