



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

MICHELL JACKS DE OLIVEIRA NETTO

**ANÁLISE DE SENTIMENTO UTILIZANDO ALGORITMOS DE MACHINE
LEARNING EM STREAMING**

**CAMPINA GRANDE – PB
2021**

MICHELL JACKS DE OLIVEIRA NETTO

**ANÁLISE DE SENTIMENTO UTILIZANDO ALGORITMOS DE MACHINE
LEARNING EM STREAMING**

Trabalho de Conclusão de Curso ao Departamento do curso de Graduação em Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Computação.

Área de concentração: Ciência de dados e Inteligência Artificial.

Orientador: Prof. Dr. Wellington Candeia de Araújo

**CAMPINA GRANDE – PB
2021**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

O48a Oliveira Netto, Michell Jacks de.
Análise de sentimento utilizando algoritmos de *machine learning* em *streaming* [manuscrito] / Michell Jack S de Oliveira Netto. - 2021.
31 p. : il. colorido.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2021.
"Orientação : Prof. Dr. Wellington Candeia de Araújo ,
Coordenação do Curso de Computação - CCT."
1. Machine learning. 2. Análise de sentimentos. 3. Spotify.
I. Título

21. ed. CDD 006.31

MICHELL JACKS DE OLIVEIRA NETTO

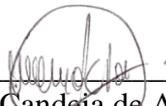
ANÁLISE DE SENTIMENTO UTILIZANDO ALGORITMOS DE MACHINE
LEARNING EM STREAMING

Trabalho de Conclusão de Curso ao Departamento do curso de Graduação em Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Computação.

Área de concentração: Ciência de dados e Inteligência Artificial.

Aprovada em: 20/10/2021.

BANCA EXAMINADORA



Prof. Dr. Wellington Candeia de Araujo (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Me. Francisco Anderson Mariano da Silva
Universidade Estadual da Paraíba (UEPB)



Prof. Vinicius Reuteman Feitoza Alves de Andrade
Universidade Estadual da Paraíba (UEPB)

A minha mãe, pela dedicação,
companheirismo e amizade, DEDICO.

“Sem a música, a vida seria um erro.”
(NIETZSCHE, 2006)

LISTA DE ILUSTRAÇÕES

Figura 1 – Aspecto da frequência auditiva.....	12
Figura 2 – Visão geral em um sistema de Banco de Dados.....	14
Figura 3 – Comparação entre MongoDB e Banco de Dados Relacional.....	15
Figura 4 – Etapas de Mineração de Texto.....	17
Figura 5 – Circulo de emoções de Plutchik.....	17
Figura 6 – Agrupamento de dados utilizando o K-means.....	19
Figura 7 – Avaliação de uma correlação.....	21
Figura 8 – Descrição dos atributos das faixas de música.....	22
Figura 9 – Atributos de cada música.....	22
Figura 10 – Correlação de Pearson.....	24
Figura 11 – Estudo comparativo entre correlações.....	25
Figura 12 – Correlação de Phik (ϕ_k).....	26

LISTA DE TABELAS

Tabela 1 – Amostra dos dados.....	27
-----------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
IR	Information Retrieval – Recuperação de Informação
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases - Conhecimento em descoberta de Databases
KDT	Knowledge Discovery in Texts – Conhecimento em descoberta de texto
LN	Linguagem Natural
KNN	K-Nearest Neighbor – K-Vizinho próximo
MSE	Mean Squared Error – Média do erro quadrado
PLN	Processamento de Linguagem Natural
SVM	Support Vector Machine – Máquina de Vetor de Suporte

SUMÁRIO

1 INTRODUÇÃO	10
2 REFERENCIAL TEÓRICO	11
2.1 Contextualização Musical	11
2.1.1 <i>Andamentos</i>	13
2.2.1 Spotify: Escutar muda tudo	13
2.2.2 Banco de Dados	13
2.2.3 NoSQL	14
2.2.4 MongoDB	15
2.2.5 Mineração de Textos	15
2.2.6 Emoções	17
2.2.7 Análise de dados	18
2.2.8 Clusterização	18
2.2.9 <i>Estatística Descritiva</i>	19
3 METODOLOGIA	19
3.1 Pesquisa de dados	20
3.2 Pesquisa exploratória	20
3.3 Correlação	20
4 RESULTADOS E DISCUSSÕES	21
4.1 Análise dos dados coletados	21
5 CONSIDERAÇÕES FINAIS	26
REFERÊNCIAS	27

FEELING ANALYSIS USING MACHINE LEARNING ALGORITHMS IN STREAMING

Michell Oliveira¹

RESUMO

Este trabalho tem como objetivo propor um modelo de apresentação de análise de sentimentos, utilizando algoritmos de *machine learning* e apresentar a transformação mercadológica dos meios musicais, por meio das plataformas de *streaming* e mostrar a relevância da análise pontual na influência no humor do usuário. Em poucos anos, pudemos acompanhar a evolução dos meios digitais, que nos permitiram sair dos rádios e fitas k7, para migrarmos aos CDs e dados digitais, com maior fidedignidade de áudio graças às evoluções tecnológicas. Temos ao nosso alcance com o advento da *internet*, o acesso a esses meios de reprodução em um grande banco de dados musicais e no contexto da nossa análise, podemos utilizar dados sensíveis intrínsecos a música como BPM (batidas por minuto), energia, dançabilidade, que permitem trazer um perfil do humor do usuário e utilizar uma base de histórico local para aumentar a acurácia do sistema. Empiricamente, muitos de nós possivelmente já tivemos a impressão de que a música tem influência sobre nossos sentidos e o nosso humor, a ciência por sua vez, veio por meio dessa discussão mostrar as evidências que já imaginávamos. Para a fase experimental, buscou-se coletar esses dados por meio das *playlists* públicas da biblioteca do Spotify através da linguagem de programação Python.

Palavras-Chave: Machine Learning. Análise de sentimentos. Spotify.

ABSTRACT

This work aims to propose a presentation model for analyzing feelings using machine learning algorithms and to present the market transformation of musical media through streaming platforms and show the relevance of punctual analysis in influencing the user's mood.

In just a few years, we were able to follow the evolution of digital media, which allowed us to move away from radios and k7 tapes, to migrate to CDs and digital data, with greater audio reliability thanks to technological developments. We have within our reach, with the advent of the *internet*, access to these means of reproduction in a large musical database and in the context of our analysis, we can use sensitive data intrinsic to music such as BPM (beats per minute), energy, danceability, that allow you to bring a profile of the user's mood and use a local history base to increase the accuracy of the system.

Empirically, many of us have possibly already had the impression that music has an influence on our senses and our mood, science, in turn, came through this discussion to show the evidence we already imagined. For the experimental phase, we sought to collect this data through public playlists in the Spotify library using the Python programming language.

Keywords: Machine Learning. Sentiment Analysis. Spotify.

¹ Autor: Michell Jack's de Oliveira Netto

1 INTRODUÇÃO

O que podia se escutar em fitas k7 e vinis, foi possível migrar também para os CD's e streaming. Com os novos conhecimentos científicos, pudemos acompanhar evolução tecnológica que permitiram o melhoramento contínuo de qualidade, armazenamento e portabilidade. Foi possível o transporte e armazenamento em dispositivos portáteis trazendo a comunicação e também entretenimento, como a música, que podem ser acessados de um *smartphone* ou *tablet* com acesso a *internet* (DAQUINO, 2012).

A primeira grande mudança na história dos hábitos se dá com a adaptação do rádio com a televisão, e em outro momento com a *internet*. A sociedade está vivendo um momento de migração de plataformas, onde emissoras de rádio podem ser escutadas através de ondas captáveis que são convertidas e devidamente processadas nos rádios, por celulares e tablets, e como resultado, temos 89% dos entrevistados em são ouvintes constantes em 13 regiões metropolitanas com 53 milhões de ouvintes com faixa etária de 30 a 39 anos, destes, 47% dos ouvintes buscam por músicas sem intervalos. O uso geralmente se dá por aparelhos de rádio comuns (65%) e 24% dos entrevistados escutam do carro e 16% de telefones móveis (THOMAZINI, 2016).

Ao longo dos anos temos acompanhado uma crescente utilização *internet* quando comparamos ao ano de 2018 (78,3%) e 2019 (82,7%) possuem acesso a *internet* pesquisa realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2019). É possível acompanhar o crescimento desta expansão e a melhora na qualidade de vida proporcionada em cada novo lar com acesso as vantagens do mundo conectado.

E dentre tantos gêneros, tantas opções musicais, de algo tão cotidiano, é difícil imaginar a mudança no humor e o impacto emocional que escutar algumas músicas preferidas no dia a dia interferem em nosso cotidiano. Nosso corpo reconhece o ritmo e diversas áreas são ativadas e assim, desencadeando reações secundárias, e nosso humor possui uma relação direta com neurotransmissores como a serotonina e dopamina. A serotonina, produzida na região do tronco encefálico, atua no apetite, no sono, no humor e inclusive no ritmo cardíaco. Já a dopamina, é produzida pela glândula adrenal, ela provoca a sensação de recompensa e atenção no corpo, assim as produções não suficientes desses dois hormônios podem causar cansaço, falta de memória, irritabilidade e até dificuldade no aprendizado (MOREIRA, 2017).

A dopamina acaba desencadeando uma sensação de recompensa e são ativados com o reconhecimento de acordes, tons e fala. Sendo que a quantidade de liberação da dopamina se modifica com a intensidade do reconhecimento e do prazer, em comparação com as medições realizadas ao escutar uma música "neutra". Durante o prazer associado à música, é ativado o núcleo "*accumbens*" que localiza-se numa posição central e interior do cérebro e constitui a principal parte do estriado ventral, que quando estimulado, sintetiza a euforia produzida pela ingestão de psicoestimulantes, assim, os pesquisadores concluíram dos resultados porque a música comumente pode ser utilizada em rituais, pelo marketing ou em filmes para induzir estados de humor (PRESSE, 2011).

A métrica utilizada pelo instituto se baseia na valência, que identifica traços mais melódicos em canções para classificá-las como mais ou menos felizes (BRÊDA, 2020).

Dentre as músicas utilizadas como base para identificar uma maior valência no top 200, temos hits que tocaram muito no carnaval (bregafunks como "Tudo OK", "Hit Contagante") do sertanejo ("Cem mil", de Gustavo Lima, e "Bebaça", de Marília Mendonça) no pop, Alok também emplacou nos Top 200, músicas menos entusiasmadas, como "Hear me Tonight" e "Symphonia". Músicas como da dupla Anavitória com temática da quarentena "Me conta da tua Janela" foi bastante ouvida e como "Você para sempre em mim" de Tiago Iorc (BRÊDA, 2020).

Diante deste contexto, traçamos a seguinte pergunta norteadora: Qual o sentimento das pessoas em relação as músicas tocadas nos *streaming* de música? Para responder esse questionamento traçamos o seguinte objetivo geral propor um modelo de apresentação de análise de sentimentos. Serão utilizados algoritmos de *machine learning* e apresentar a transformação mercadológica dos meios musicais, por meio das plataformas de *streaming* e mostrar a relevância da análise pontual na influência no humor do usuário.

O trabalho se justifica, tendo em vista que durante o distanciamento social provocado pela pandemia do coronavírus, o ânimo de muitas pessoas foi transformado por uma tristeza suave, substituindo um pouco os grandes hits com maior quantidade de batidas por minuto, nos topos de playlists. Onde o instituto DeltaFolha em um levantamento pode observar em cerca de 34 países incluindo o Brasil, nos tops 200 de cada país, listas de músicas que obtiveram uma menor energia, e entre as músicas mais escutadas são as que o algoritmo identifica com traços melódicos com atributo *valência* mais baixos, classificando estas como menos felizes (BRÊDA, 2020).

2 REFERENCIAL TEÓRICO

O objetivo desta seção é trazer uma breve contextualização das evoluções tecnológicas com alguns conceitos computacionais que permitiram uma nova forma de escutar e apreciar música, e o como surgiu o *streaming*.

2.1 Contextualização Musical

Temos diversos avanços na forma como escutamos e armazenamos a música, e têm nos acompanhado ao longo da história com o aproveitamento da energia elétrica. As diversidades de formas de armazenamento que foram proporcionados pelo menos nos últimos 70 anos, permitiram muita mobilidade e inclusive abstração como o armazenamento em “nuvem” (SANTOS; RAMOS, 2016).

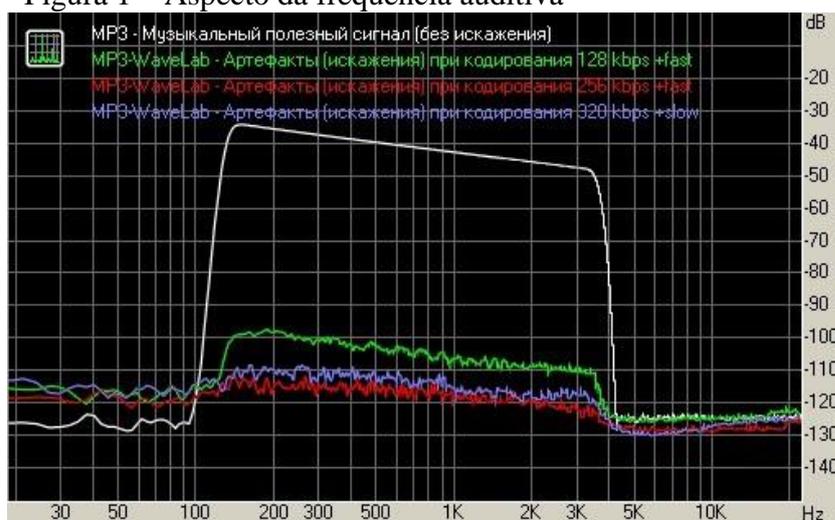
A música que nasce desde um instrumento musical, como o violão por exemplo, tem evoluído ao longo dos anos em sua qualidade, armazenamento, proteções contra cópias. O disco de vinil proporcionou a mobilidade e o armazenamento da música em meados de 1948, variando entre de 25 a 30 cm e com capacidade de até meia hora de cada lado, com isso, as indústrias fonográficas passaram a gravar, reproduzir, distribuir e divulgar os artistas. Em 1963, foi a vez das fitas cassetes, que gravavam de músicas às programações de rádio, e isso foi possível uma melhor entrega de qualidade e durabilidade e com possibilidade de escutar em um *walkman* portátil (SANTOS; RAMOS, 2016).

Um pouco mais a frente, temos os *Compact Disc* (CD), que trouxe inovações em questão de durabilidade, qualidade e leveza, e era utilizado com música com cerca de 74 a 80 minutos de gravação de dados como fotos e outros arquivos, também era possível utilizar aparelhos portáteis a pilha para reprodução de CD's de música nos *discman* e com a capacidade de 700mb, era possível ser reconhecível na maioria dos computadores e qualidade. E com essa sequencia de inovações, a *internet* transformou o mundo ao nosso redor, quebrando barreiras de armazenamento e limites físicos, onde permitiram novas experiências na forma de escutar música, quando as indústrias fonográficas descobriram o potencial de mercado, e distribuição via *streaming* (KLEINA, 2019).

O *mp3* é um projeto que é estudado desde 1979 pelo alemão o professor alemão Karlheinz Brandenburg, que pode registrar uma patente em que é possível transmitir fala em

alta qualidade ou música por linhas telefônicas. Este formato trouxe a mobilidade e facilidade na transmissão com redução da qualidade do sinal de uma forma com que o nosso ouvido não percebesse, assim em cerca de 1985, Karlheinz e a equipe aproveitam os avanços tecnológicos para estudar a psicoacústica, um campo onde é possível estudar a forma como o som é recebido pelo nosso aparelho auditivo (KLEINA, 2019).

Figura 1 – Aspecto da frequência auditiva



Fonte: Kleina, (2019).

Com a disponibilidade, assim como versatilidade do armazenamento, o *mp3 player*, foi o sucessor do *discman*, (KLEINA, 2019) com maior capacidade de armazenamento e que com a mobilidade, e durabilidade da bateria, uma vez que não existia o gasto com um motor de rotação e laser para leitura. Foram criadas diversas leis para proteger os direitos das produtoras, e dos artistas, alguns distribuem o seu material. É possível comprar discos inteiros ou apenas faixas em MP3 direto da *iTunes Store*, dos sites das gravadoras ou dos próprios artistas (COELHO, 2013).

Ao final dos anos 80, foi montada a equipe para criação da ISO, organização de padronizações internacionais na Alemanha, assim, foi possível a solicitação da criação de um padrão para o formato de áudio digital. E como resultado, o algoritmo OCF, que permite uma codificação otimizada da frequência, que logo em seguida foi substituído pelo ASPEC (*Adaptative Spectral Perceptual Entropy Coding*). E através da adaptação de três camadas, sendo a primeira camada de baixa frequência foi utilizada pelo algoritmo MUSICAM, desenvolvida pelo grupo de pesquisadores de vários países europeus (EUREKA), a segunda camada foi feita a otimização do MUSICAM e a terceira camada baseada no ASPEC para só em 1995 ser lançado o formato ISO MPEG Audio Layer 3, que seria popularizado como *.mp3*. assim como Lucas Françaolin (2013, p.36) diz:

Enquanto na era acústica se tornou possível à materialização do primeiro produto - que contém o registro do som em um suporte físico exclusivo -, a era digital possibilitou a existência de um registro do som que não está vinculado a um suporte físico em específico.

Por meio do avanço das tecnologias de acesso à *internet* por banda larga e conexões mais rápidas permitiram o avanço do streaming, onde os dados são utilizados sob demanda, e

não são armazenados no dispositivo. Trazendo uma acessibilidade deste conteúdo online como explica Braga et al. (2016). Com isso:

Seu modelo de negócio é baseado em uma experiência de consumo de conteúdos digitais que substitui a lógica da compra de um disco pelo acesso a fonogramas hospedados nas redes digitais, permitindo que seu desfrute possa ser realizado sem que se precise baixar, arquivar e organizar esse conteúdo em dispositivos individuais (KISCHINHEVSKY et al, 2019).

Em 2008 a startup sueca Spotify AB, iniciou suas atividades de distribuição, divulgação e de transmissão de som instantânea, sem necessariamente existir o download para o dispositivo, uma vez que as empresas podem também de acordo com suas políticas, oferecer o serviço de download para reprodução sem o acesso à *internet*. Esse tipo de reprodução inova e permite diversas funcionalidades ao ouvinte, como reprodução em um smartphone, tablet ou computador.

2.1.1 Andamentos

Ao se deparar com a plataforma Spotify pela primeira vez, o usuário se depara com uma interface evidente e que traz visualmente os conteúdos de forma análoga às formas das indústrias fonográficas e radiofônicas de se apresentar.

2.2.1 Spotify: Escutar muda tudo

Quando lembramos do cenário onde os CDs e mídias físicas eram mais caras, o Spotify em 2006, trouxe uma dinâmica totalmente nova. Daniel Ek e Martin Loretzon pensaram em uma plataforma que fosse possível um fácil manuseio e que pudesse trazer um grande acervo e pudesse gerar uma ponte entre os usuários e a empresa. (COUTINHO, 2019).

Após o seu registro na plataforma, é possível ter acesso a uma grande biblioteca musical dentre todas disponíveis no mercado. Onde é possível o cadastro de forma gratuita, trazendo algumas restrições de acesso, como por exemplo, comerciais e notícias da plataforma no decorrer da reprodução.

2.2.2 Banco de Dados

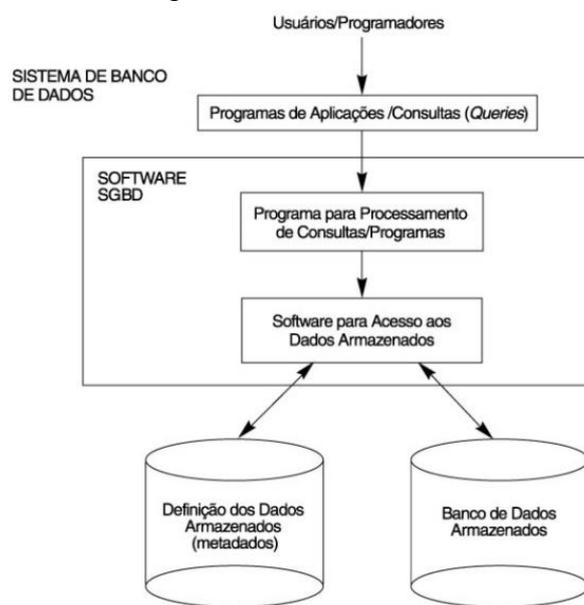
De acordo com Esmalri e Navathe (2011), quando temos uma coleção de dados que estão relacionados entre si, temos uma correlação entre os atributos, e esse vínculo pode ser de dados que possuem informações que tem representação em outra tabela. Assim, temos uma tabela que representa em colunas, sua descrição em dados no mundo real.

Um banco de dados representa algum aspecto do mundo real, às vezes chamado de minimundo ou de universo de discurso (UoD – Universe of Discourse). (ELMASRI; NAVATHE, 2011).

Assim, temos como elementos intrínsecos ao banco o próprio banco de dados e o software gerenciador do banco de dados, este por sua vez, tem a responsabilidade de

administrar o acesso a informação e garantir a integridade dos seus componentes que são os usuários e as aplicações (Figura 2).

Figura 2 – Visão geral em um sistema de Banco de Dados



Fonte: Elmasri & Navathe (2011).

Entre as informações armazenadas em um banco de dados e os usuários e programas consumidores da aplicação, existe um sistema gerenciador de banco de dados (SGBD). Este gerenciador compreende mais uma camada de segurança e confiabilidade no acesso aos dados garantindo o controle de concorrência dos dados de forma que não exista perda no processo de recuperação da informação solicitada.

2.2.3 NoSQL

Inicialmente utilizado por Carlo Strozzi em 1998, como um banco de dados relacional que não possuía interface SQL. O *NoSQL* foi projetado para vir como alternativa aos bancos de dados baseados em modelos relacionais, e trazer uma velocidade e disponibilidade elevada, suprimindo também a necessidade de exigir esquemas de tabela fixa. Dessa forma, existe uma tendência arquitetural para prover essa crescente escalabilidade, graças às grandes cargas de trabalho e volumes de dados (SOLAGNA; LAZZARETTI, 2016).

No uso do *NoSQL*, os dados podem ser classificados de acordo com 4 categorias: chave/valor, orientado às colunas, orientado a documentos e os que são baseados em grafos.

- Modelo Chave-valor (Key/Value): O seu uso acaba sendo semelhante ao uso de mapas relacionais ou dicionários, possuindo uma chave de valor único que permite a recuperação rápida e acesso aos dados estruturados ou não.
- Modelo em colunas: Em uma representação em colunas, quando acessados as informações conhecidas de uma coluna, como: nome, endereço, CEP. Sendo o acesso de forma atômica, e quando acessada, o retorno da consulta é a coluna

inteira. Sendo vantajosos em Data Warehouses¹, onde podem ser processadas uma grande quantidade de dados e agregações similares.

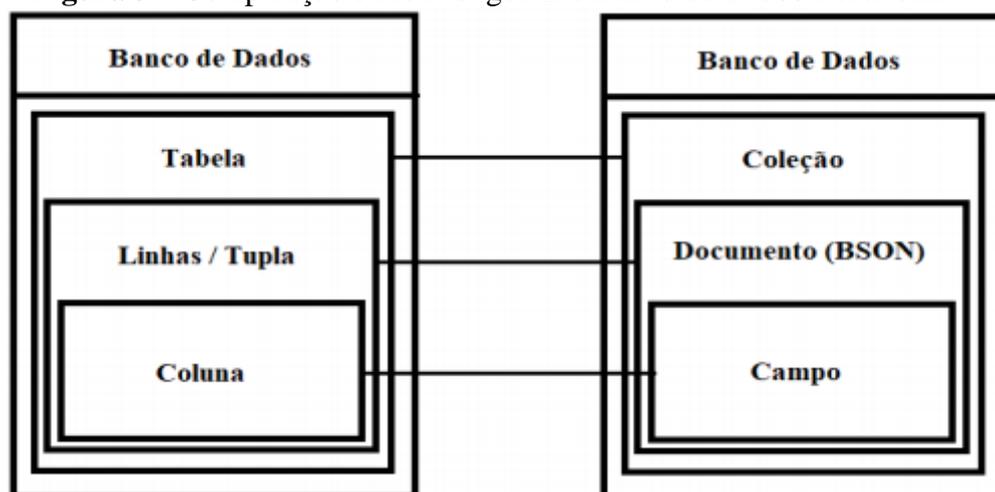
- Modelo baseado em grafos: Fundamentada pela teoria dos grafos, que consiste em nós e arestas e suas propriedades. Onde os nós representam as entidades, e as propriedades os respectivos atributos das relações.

Modelo baseado em documentos: O elemento dos dados é chamado de documento, e o *MongoDB* faz o armazenamento destes dados em um formato semelhante ao JSON (usado BSON – versão binária do JSON), que atribui à identificador de chave, atributos recuperáveis em dados semiestruturados (SOLAGNA; LAZZARETTI, 2016).

2.2.4 MongoDB

O banco de dados MongoDB é uma grande referência em dados não estruturados NoSQL, inicialmente desenvolvido como um componente de serviço da empresa, e acabou se transformando em um código open source em 2009, esta aplicação de alta *performance* orientada a documentos e desenvolvido em C++, busca diferenciar-se dos bancos relacionais com estruturas imutáveis.

Figura 3 – Comparação entre MongoDB e Banco de Dados Relacional



Fonte: Felisberto, (2019).

2.2.5 Mineração de Textos

Segundo Lazzarin (2017) a mineração de textos (MT) é uma extensão da mineração de dados (Data Mining), que consiste em um processo de exploração de padrões consistentes, e hipóteses que podem instrumentalizar uma solução ou desvendar tendências no processo conhecido como Descoberta de Conhecimento em Banco de Dados (Knowledge Discovery in Databases – KDD).

No processo de mineração de dados, é possível descobrir tendências por meio dos algoritmos, gerando informação através dos dados atômicos que extraem conteúdos mais relevantes e não triviais. Text Data Mining, Knowledge Discovery in Texts (KDT) (TAN,

¹ Sistema de gerenciamento de dados projetado para ativar e fornecer suporte às atividades como *business intelligence* (BI), por exemplo.

1999), e para reconhecimento de emoções básicas como raiva, medo, tristeza, surpresa e alegria em uma extração (DOSCIATTI, 2013).

No reconhecimento de grande parte dos trabalhos de mineração de dados, são seguidas algumas etapas para análise do domínio dos dados, o processamento desses dados, a transformação, processamento e interpretação nas etapas do KDD:

Quadro 1 – Etapas em Minerações de Textos

ETAPAS	CARACTERÍSTICAS
DADOS	O KDD se baseia no armazenamento dos dados de forma estruturada
SELEÇÃO DE DADOS	Após ter definido o domínio sobre o qual se pretende executar o processo de descoberta, a próxima etapa é selecionar e coletar o conjunto de dados ou variáveis necessárias
PROCESSAMENTO	Esta etapa é também conhecida como pré-processamento, visando eliminar os dados que não se adequam às informações, com base nos algoritmos, ou seja, dados incompletos, tratamento de nulos, sumarização, problemas de definição de tipos, eliminação de tuplas repetidas, etc.
TRANSFORMAÇÃO	Nesta etapa os dados deverão ser armazenados adequadamente para facilitar na utilização das técnicas de mineração de dados
MINERAÇÃO DE DADOS	A atividade de descoberta do conhecimento é onde são processados os algoritmos de aprendizado de máquina e de reconhecimento de padrões. A maioria dos métodos de Data Mining são baseados em conceitos de aprendizagem de máquina, reconhecimento de padrões, matemática, estatística, classificação, clusterização, modelos gráficos, redes neurais, <i>deep learning</i> , inteligência artificial, etc.
INTERPRETAÇÃO/AVALIAÇÃO	Nesta etapa final, os resultados do processo de descoberta do conhecimento podem ser mostrados de diversas formas, porém devem ser apresentadas de forma que o usuário possa entender e interpretar os resultados obtidos.

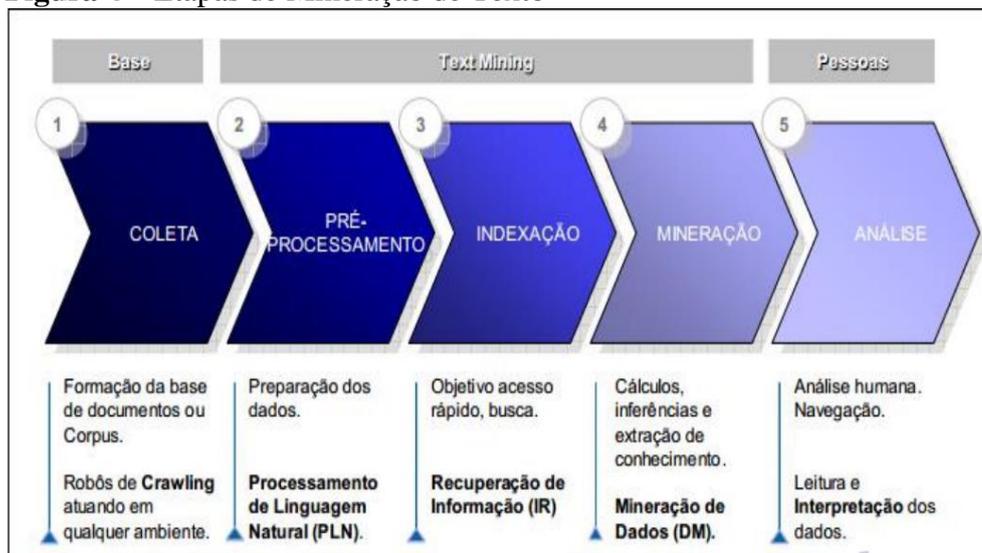
Fonte: Felisberto, (2019).

Na mineração de dados, é possível aplicar funções analíticas de linguagem natural (LN) de mineração de dados e técnicas de recuperação de informação – *Information Retrieval* (IR) (DORRE et al. 1999).

As utilizações das ferramentas de mineração de texto têm aplicabilidade:

- Extração de informações importantes de um documento, utilizando LN, IR e algoritmos e por DORRE (1999) correspondência padrão;
- Descobrir relações e tendências, comparando as informações extraídas dos documentos;
- Classificação e organização do documento observando o seu conteúdo (TKACH, 1998);
- Recuperação da informação contida nos documentos de acordo com seu conteúdo;
- Agrupar documentos por semelhança entre o seu conteúdo (CARVALHO FILHO, 2014).

Na Figura 4, é possível observar cada momento do processo de Mineração de Texto demonstrada por Carvalho Filho (2014):

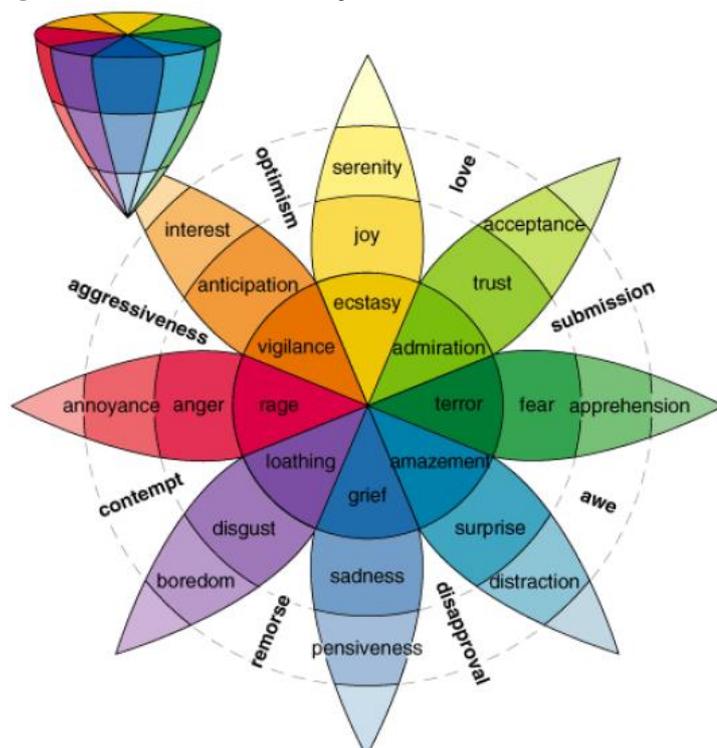
Figura 4 – Etapas de Mineração de Texto

Fonte: Carvalho Filho, (2014).

2.2.6 Emoções

O que temos de mais preciso no que se conhece por emoção é algo que ainda é tão incompleto quanto reconhecer a importância desse conhecimento (ROMAN, 2007).

Em 2001, Plutchik demonstrou um modelo circunflexo e tridimensional para representar os conceitos das emoções primárias. As oito emoções primárias são destacadas e divididas em seções indicando as interseções de duas emoções primárias e uma diagramação em cores.

Figura 5 – Circulo de emoções de Plutchik

Fonte: Plutchik (2001).

As emoções tem sido objeto de estudo da ciência da computação, especialmente quando se trata da interação entre humanos e máquinas.

2.2.7 Análise de dados

Ao analisarmos os dados, precisamos descrevê-los, e entender assim a representatividade numérica, como as relações entre eles, comparação em função do tempo. Assim podemos identificar de forma resumida para compor indicativos e previsões (GERHARDT, 2009).

Uma vez identificada a natureza dos dados, é possível distingui-los em dados quantitativos e qualitativos (MORAIS, 2005), visto no Quadro 2.

Quadro 2 – Natureza dos dados

Dados Quantitativos	Dados Qualitativos
<ul style="list-style-type: none"> • Os dados são representados por números ou símbolos. • As relações entre esses dados são dadas através de igualdade e desigualdade entre eles e possuem natureza discreta (números inteiros) ou contínua (números reais) (MORAIS, 2005). 	<ul style="list-style-type: none"> • Os valores qualitativos descrevem tipos, classes ou categorias, assumindo não medidas numéricas, mas sim uma classificação baseada em qualidade. (MORAIS, 2005).

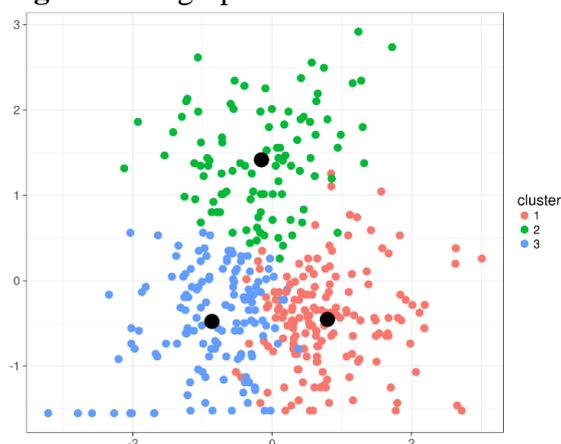
Fonte: Elaborada pelo autor, 2021.

Durante as análises dos dados é observada por (MORETTIN ET AL, 2019), conhecida como *Análise Exploratória de Dados* que explora a organização e por meio da inferência estatística, obter conclusões sobre uma população com o processamento de dados amostrais e é uma ferramenta imprescindível para observar tendências e resumi-los.

2.2.8 Clusterização

Quando precisamos descrever volumes de dados, e é necessário utilizar algum critério que possa melhorar o processo de visualização da informação ou classificação. Utilizando a técnica de K-ésimo vizinho mais próximo (*K-nearest neighbor*, KNN), é possível a classificação estimando a probabilidade condicional (MORETTIN ET AL, 2019).

Figura 6 – Agrupamento de dados utilizando o K-means



Fonte: RPubS (2021)

A Figura 6, demonstra uma nuvem de dados, e segundo Benito et al (2017), o processo de clusterização busca encontrar similaridades na estrutura em meio a coleção dos dados, o que inclui em criar clusters, onde cada cluster é representado pela centroide, que representa a média apresentada dos pontos com similaridade de características.

2.2.9 Estatística Descritiva

A partir da compreensão dos dados, é possível encontrar características importantes, a sua abrangência, permitem ao cientista de dados inferir informações, propor hipóteses e compreender melhor a essência dos dados. A estatística descritiva constitui um conjunto de ferramentas de análise utilizadas na Estatística para descrever de forma resumida um conjunto de dados, podendo ser organizados em um modelo de tabelas ou gráficos (MORAIS, 2005).

Por Guedes et al. (2005), através da estatística descritiva, é possível ter uma visão global das variações desses dados, organizando-os e descrevendo estes de três formas: tabelas, gráficos e de medidas descritivas. Onde assim é possível ressaltar características observadas isoladamente em tabelas ou em relação a outras, onde algumas tendências podem ser observadas por duas categorias: medidas de posição e medidas de dispersão.

3 METODOLOGIA

Em uma abordagem científica, a pesquisa pode apresentar diversos novos caminhos em que podem se trilhar, e com base em Wazlawick (2010) “Nem sempre um trabalho de pesquisa limita-se a um único tipo”. Sendo muitos trabalhos científicos servirem como pilares para o emprego de técnicas e novas discussões.

Visto a necessidade de classificação, a pesquisa exploratória busca pontuar porquês, e nesse caso analisar os fenômenos observados, dessa forma, “é a pesquisa científica por excelência porque, além de analisar os dados observados, ela busca suas causas e explicações, ou seja, os fatores determinantes destes dados” (WAZLAWICK, 2010).

Essa pesquisa foi utilizada com base na API disponível do *Spotify*, e desenvolvida com a ajuda da linguagem de programação *Python*. Através da *playlist* de maior relevância no Brasil, a “Viral 50” que trouxe uma pequena amostra e foram organizados em gráficos e tabelas para melhor interpretação e abstração.

Os dados mais relevantes foram filtrados com base na pesquisa para visibilizar tendências e enaltecer o potencial da mineração de dados como forma de constituir uma ferramenta com fundamentação teórica e matemática para sustentar uma possível correlação dos dados.

3.1 Pesquisa de dados

O processo de pesquisa formal caracteriza-se pelo processo sistemático de desenvolvimento científico, buscando dessa forma o progresso da ciência, que pode decorrer de razões intelectuais quando baseadas no desejo de conhecer e também quando busca desenvolver conhecimentos científicos sem preocupação direta com as suas aplicações e conseqüências práticas. E por seus objetivos, como o de evitar a subjetividade, os teóricos positivistas sugerem que as investigações dos fenômenos sociais restrinjam-se em cada uma das etapas, que pode ser subdividida em outras mais específicas embora, quanto mais são as observações, mais elas se afastam da realidade física, onde é sugerida a restrição ao que se sugere ser investigado (GIL, 2008).

Nas pesquisas científicas descritivas, em sua maior preocupação, busca descobrir a possível indicação de relação entre as variáveis, e quando pretende também determinar a natureza dessa relação, a pesquisa descritiva se aproxima da explicativa. Assim como as pesquisas exploratórias, que acabam servindo mais para proporcionar uma nova visão do problema (GIL, 2008).

3.2 Pesquisa exploratória

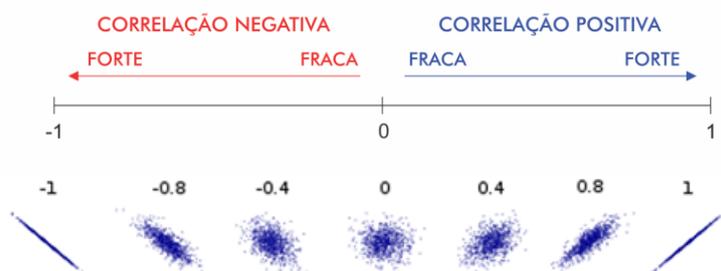
Em uma pesquisa social, é comum possuir um objetivo definido, assim, é possível organizar as diversas pesquisas em agrupamentos amplos. Onde, seu desenvolvimento tende a ser objetivando a generalização, que permitam enriquecer o levantamento bibliográfico e documental. Pesquisas exploratórias buscam esclarecer desde a formulação de problemas mais precisos a hipóteses pesquisáveis, dessa forma, buscam apresentar um panorama acerca do tema, assim como são pesquisas descritivas (GIL, 2008).

3.3 Correlação

Entre as operações matemáticas que podem convalidar informações de dados distintos, por meio de cálculos estatísticos são as correlações. O coeficiente de correlação de Pearson (r), ele avalia o grau de relação entre duas variáveis quantitativas e exprime o grau de correlação através de valores situados entre -1 e 1. Seja X_1, X_2, \dots, X_n o conjunto das medidas de uma das variáveis (período das ondas), e seja Y_1, Y_2, \dots, Y_n , as medidas da outra variável (diâmetro médio de sedimentos), onde, a partir do cálculo de distância e dados intrínsecos, é possível obter o coeficiente de correlação que permite inferir informações,

propor hipóteses e compreender melhor a essência dos dados (OLIVEIRA, 2019). Sendo o sinal positivo indicativo de que as variáveis são diretamente proporcionais, enquanto que o sinal negativo indica que a relação entre as variáveis é inversamente proporcional. Sendo este coeficiente adimensional, logo não é afetado pelas unidades de medidas das variáveis X e Y (OLIVEIRA, 2019).

Figura 7 – Avaliação de uma correlação



Fonte: Elaborada pelo autor, 2021.

4 RESULTADOS E DISCUSSÕES

Uma vez delimitada a fonte dos dados, a *playlist* “Viral 50”, é observado os atributos principais como a relevância dos dados distintos, quantidade de zeros no grupo de estudo, e a presença de dados *outliers* que ultrapassam muito da média e dos desvios padrões do conjunto principal de dados, buscando uma maior homogeneidade das informações da pesquisa e não trazer uma média que proporcione uma irregular análise do conjunto principal.

Com base na coleta de dados, resultado da análise e informações principais fornecidos pela plataforma Spotify, é possível listá-los e por meio de bibliotecas de análise de dados em python como o Pandas entre outras, é possível ter um arsenal de ferramentas que podem auxiliar o processo de salvar os dados para novas análises e gerar comparativos. Assim, conseguimos visualizar tendências por meio de correlações entre os atributos, sejam essas correlações positivas fortes ou negativas, temos um forte indicativo associando atributos como valência e energia em relação ao possível humor do usuário.

4.1 Análise dos dados coletados

Uma vez obtido o *login* e autorização para uso da plataforma API do *Spotify*, é possível delimitar o a abrangência da pesquisa e os dados que podemos observar. Os atributos das músicas representam suas maiores informações intrínsecas, como propriedades das músicas, a presença ou não de instrumentos e em relação ao humor, descritas na Figura 8.

Figura 8 – Descrição dos atributos das faixas de música

Audio Features & Analysis

Explore audio features and in-depth audio analysis of tracks

With the Spotify Developer Platform, you're able to read calculated audio features of tracks to learn about its danceability, energy, valence, and more. For more advanced use cases, it is possible to read in-depth analysis data about tracks such as the segments, tatum, bars, beats, pitches, and more.

- Mood: Danceability, Valence, Energy, Tempo
- Properties: Loudness, Speechiness, Instrumentalness
- Context: Liveness, Acousticness
- Segments, Tatum, Bars, Beats, Pitches, Timbre, and more

[READ THE DOCS](#)

Fonte: Spotify, 2021.

Uma vez coletada as informações da *playlist*, e por meio dos dados intrínsecos às músicas, podemos obter informações necessárias para a análise, assim como na Figura 8:

Figura 9 – Atributos de cada música

```
Data columns (total 14 columns):
# Column Non-Null Count Dtype
---
0 energy 50 non-null float64
1 liveness 50 non-null float64
2 tempo 50 non-null float64
3 speechiness 50 non-null float64
4 acousticness 50 non-null float64
5 instrumentalness 50 non-null float64
6 time_signature 50 non-null int64
7 danceability 50 non-null float64
8 key 50 non-null int64
9 duration_ms 50 non-null int64
10 loudness 50 non-null float64
11 valence 50 non-null float64
12 mode 50 non-null int64
13 type 50 non-null object
dtypes: float64(9), int64(4), object(1)
memory usage: 5.6+ KB
```

Fonte: Elaborada pelo autor, 2021.

Onde é possível observar uma amostra das informações coletadas de cada música pela Tabela 1.

Tabela 1 – Amostra dos dados

	energy	liveness	tempo	speechiness	acousticness	instrumentalness	time_signature	danceability	key	duration_ms	loudness	valence
0	0.871	0.9700	151.175	0.2580	0.6170	0.000000	4	0.486	3	153410	-3.642	0.663
1	0.595	0.0882	82.492	0.2460	0.1380	0.000000	4	0.806	8	187636	-6.166	0.677
2	0.887	0.2570	151.920	0.0564	0.3690	0.000001	4	0.668	11	157176	-4.404	0.907
3	0.318	0.1900	86.453	0.0507	0.8070	0.101000	5	0.628	9	221538	-8.887	0.492
4	0.693	0.0829	129.882	0.0559	0.2660	0.000000	4	0.608	11	153740	-4.057	0.308
5	0.772	0.0710	75.663	0.0645	0.5610	0.000000	4	0.596	10	154662	-4.721	0.591
6	0.669	0.1720	109.999	0.0341	0.2730	0.000000	4	0.575	4	168645	-4.750	0.524
7	0.775	0.0724	160.067	0.0429	0.3000	0.000000	4	0.681	2	174938	-4.153	0.960
8	0.731	0.0891	86.959	0.0887	0.1800	0.000000	4	0.526	11	193221	-4.662	0.476
9	0.732	0.2430	135.005	0.0308	0.1080	0.000000	4	0.774	9	192725	-4.227	0.734
10	0.704	0.0501	149.995	0.0615	0.0203	0.000000	4	0.736	3	212000	-7.409	0.894
11	0.807	0.9770	115.010	0.1000	0.3810	0.000000	4	0.811	4	164949	-2.681	0.766
12	0.792	0.9200	154.027	0.0393	0.6070	0.000024	4	0.650	6	225672	-3.825	0.872
13	0.873	0.0946	150.006	0.0449	0.4020	0.000000	4	0.682	6	147072	-4.163	0.964
14	0.499	0.1310	122.070	0.4810	0.3850	0.000000	4	0.621	6	145574	-6.998	0.421
15	0.743	0.0674	166.043	0.0629	0.2510	0.000000	4	0.746	7	172621	-3.349	0.938

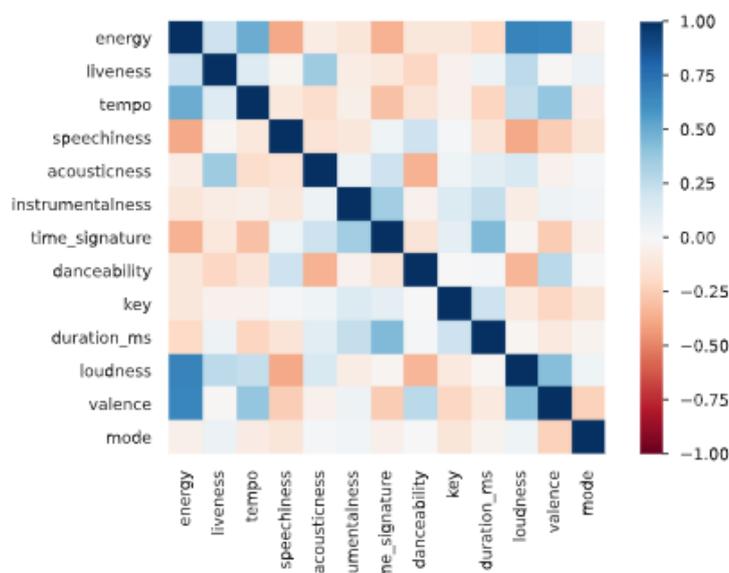
Fonte: Elaborada pelo autor, 2021.

- O atributo *Energy* (energia), ele representa uma medida perceptual de intensidade e atividade. Normalmente, faixas energéticas parecem rápidas, altas e barulhentas.
- O *Liveness* (vivacidade) este valor descreve a probabilidade de a música ter sido gravada com um público ao vivo. De acordo com a documentação oficial um valor acima de 0,8 fornece grande probabilidade de a faixa estar ao vivo
- No *speechiness* (acústica). este valor descreve o quão acústica é uma música. Uma pontuação de 1,0 significa que é mais provável que a música seja acústica.
- A *acousticness* (fala), detecta a presença de palavras faladas em uma faixa. Se a fala de uma música estiver acima de 0,66, provavelmente é composta de palavras faladas, uma pontuação entre 0,33 e 0,66 é uma música que pode conter música e palavras e uma pontuação abaixo de 0,33 significa que a música não tem fala.
- Na *instrumentalness* (instrumentalidade), Este valor representa a quantidade de vocais na música. Quanto mais próximo estiver de 1.0, mais instrumental será a música.
- A *Danceability* (dançabilidade) descreve o quão adequada uma faixa é para dançar com base em uma combinação de elementos musicais incluindo tempo, estabilidade de ritmo, força de batida e regularidade geral. Um valor de 0,0 é menos dançável e 1,0 é mais dançante
- O *Loudness* (sonoridade), detecta a presença de um público na gravação. Valores de vivacidade mais altos representam um aumento na probabilidade de a trilha ter sido executada ao vivo.
- A *Valence* (valência) apresenta uma medida de 0,0 a 1,0 que descreve a positividade musical transmitida por uma faixa. Faixas com alta valência soam mais positivas (por exemplo, feliz, alegre, eufórico), enquanto faixas com baixa valência soam mais negativas (por exemplo, triste, deprimido, com raiva).

Entre as nuvens de pontos da figura 10, podemos analisar que em situações não lineares, o coeficiente de Phik (ϕ_k), apresentou resultados diferentes e maiores que zero

Assim como a correlação de Pearson, é possível observar a correlação de Phik (ϕ_k) onde demonstra a correlação com variação entre 0.0 e 1.0, dada pela Figura 10:

Figura 10 – Correlação de Pearson



Fonte: Elaborada pelo autor, 2021.

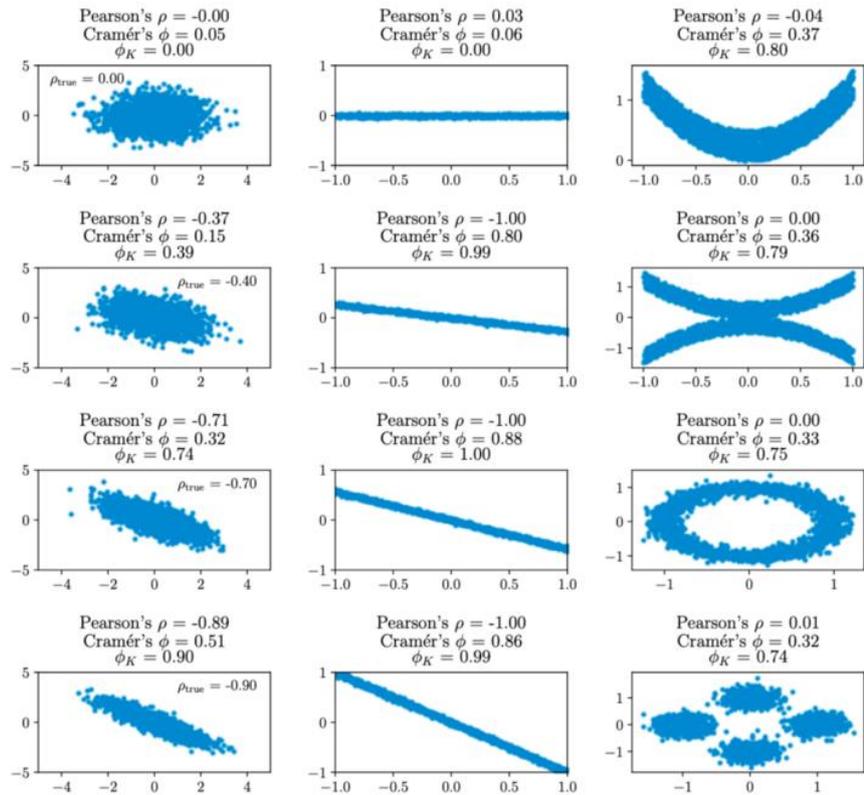
Sendo seu cálculo demonstrado pelo resultado da fórmula abaixo:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}}$$

Nas informações obtidas, é possível obter atributos para cada uma das músicas, onde o atributo Energia (**Energy**), em nossa análise indicou maiores correlações. Com base na correlação de Pearson que demonstra um modelo matemático com correlação positiva ou negativa para demonstrar a afinidade dos dados.

No coeficiente de correlação Phik (ϕ_k), obtemos um novo coeficiente de correlação e prático que funciona consistentemente entre variáveis categóricas, ordinais e de intervalo de captura sem dependência linear (BAAK, 2020). Assim, podemos ver que Phik (ϕ_k) faz um bom trabalho na detecção de padrões não lineares perdidos por outros coeficientes conforme a figura abaixo sobre um comparativo entre alguns modelos matemáticos em que existem dados não linearmente separáveis:

Figura 11 – Estudo comparativo entre correlações

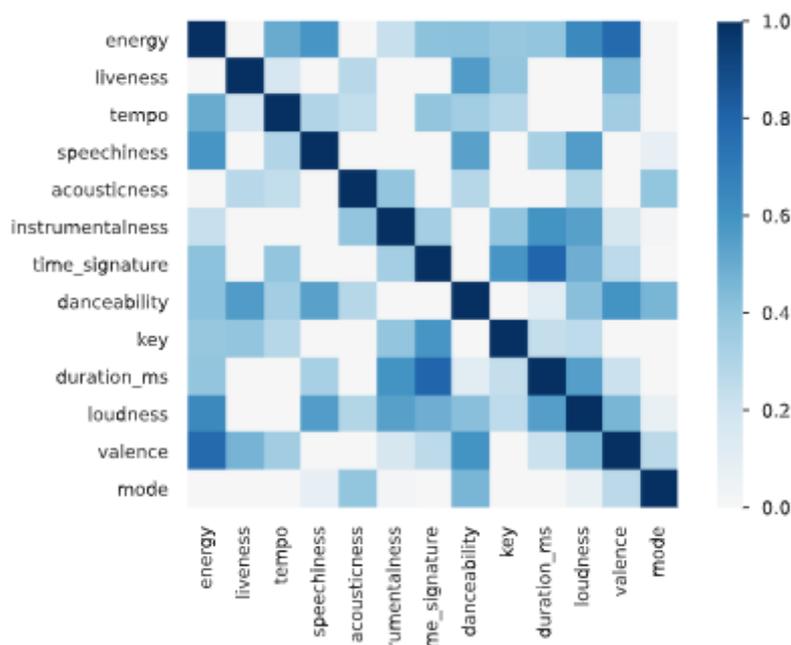


Fonte: BAAK, 2020.

Entre as nuvens de pontos da figura 10, podemos analisar que em situações não lineares, o coeficiente de Phik (ϕ_k), apresentou resultados expressivos e diferentes de zero.

Assim como a correlação de Pearson, é possível observar a correlação de Phik (ϕ_k) onde demonstra a correlação com variação entre 0.0 e 1.0, acrescentando uma nova perspectiva do mesmo conjunto de dados, mostrado pela Figura 11:

Figura 12 – Correlação de Phik (ϕ_k)



Fonte: Elaborada pelo autor, 2021.

Onde os pontos com tons de azuis mais escuros, tendem a revelar uma maior correlação, apresentando informações com tendências e com potencial de análise, excluindo-se a diagonal principal com valores iguais a um, onde são suas representações dos próprios atributos com eles mesmos.

5 CONSIDERAÇÕES FINAIS

Neste estudo, foi possível apresentar como acontece o processo de mineração de dados, para estruturar e criar um modelo de classificação com base nos dados fornecidos pela plataforma de streaming *Spotify*, assim, permitiu reconhecer a melhoria dos meios de distribuição e armazenamento até chegar ao seu ouvinte. É possível observar também o ganho de qualidade associado a cada tipo de mídia e a capacidade de armazenamento. Com a evolução das mídias e com maior acessibilidade, é observada uma nova forma de consumir músicas em diferentes estilos e gêneros acerca das músicas.

Em meio às transições de mídia, as plataformas de *streaming* tiveram um papel fundamental na difusão e acesso as mídias. Fornecendo ao usuário a liberdade de escutar suas músicas inicialmente por meio da *internet*, mas, também permitindo o seu conteúdo seja escutado em um ambiente sem esse acesso. O Spotify é uma Big Tech sueca com milhões de usuários ativos globais, que podem modelar o seu ambiente com suas playlists favoritas, e poder ver a atividade dos amigos semelhante a uma rede social.

Com um grande crescimento de usuários e uma ampla aplicabilidade da ciência de dados, é possível através de o streaming compreender características intrínsecas da música e compreender possíveis tendências sendo a princípio a principal dificuldade em relação à delimitação do trabalho.

Por conseguinte, é possível um direcionamento de sugestões e publicidades, quando identificados as principais características de um grupo de ouvintes, onde a partir dessa análise, seria possível uma inferência do humor do usuário pela aprendizagem de máquina utilizando como base, os aspectos de cada cluster de usuários.

REFERÊNCIAS

BRAGA, Vitor et al. **O streaming de música como um estímulo para a ampliação do consumo musical: um estudo do Spotify**. In: XXXIX CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO. 05 a 09 de setembro de 2016. Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. São Paulo, SP, 2016.

BRÊDA, Lucas; MARIANI, Daniel; YUKARI, Diana. **Brasil é o país que mais passou a ouvir músicas tristes na quarentena**. 2020. Disponível em: <https://www1.folha.uol.com.br/ilustrada/2020/06/brasil-e-o-pais-que-mais-passou-a-ouvir-musicas-tristes-na-quarentena.shtml>. Acesso em: 01 de mar. 2021.

CARVALHO FILHO, José Adail. **Mineração de textos: análise de sentimentos utilizando Tweets referentes à Copa do Mundo**. 2014. 46f.: il. Monografia (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Engenharia de Software, Quixadá, 2014. Disponível em: <http://www.repositoriobib.ufc.br/000017/0000179f.pdf>. Acesso em: 24 de abr. 2021.

COELHO, Taysa. **Fita cassete completa 50 anos; veja o que mudou em cinco décadas**. 2013. Disponível em: <https://www.techtudo.com.br/noticias/noticia/2013/09/fita-cassete-completa-50-anos-veja-o-que-mudou-em-cinco-decadas.html> Acesso em: 03 de abr. 2021.

COUTINHO, Thiago. **Spotify: saiba como tudo começou** 2019. Disponível em: <https://www.voitto.com.br/blog/artigo/spotify> Acesso em: 18 de abr. 2021.

DAQUINO, Fernando. **A evolução do armazenamento de músicas**. 2012. Disponível em: <https://www.tecmundo.com.br/infografico/30658-a-evolucao-do-armazenamento-de-musicas-infografico-.htm>. Acesso em: 01 de abr. 2021.

DOSCIATTI, Mariza Miola; FERREIRA, Lohann Paterno Coutinho; PARAISO, Emerson Cabrera. **Identificando emoções em textos em português do Brasil usando máquina de vetores de suporte em solução multiclasse. ENIAC Encontro Nacional de Inteligência Artificial e Computacional**. Fortaleza, Brasil, 2013. Disponível em: <https://www.ppgia.pucpr.br/~paraiso/mineracaodeemocoes/recursos/emocoensENIAC2013.pdf>. Acesso em: 24 de abr. 2021.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. Tradução: Daniel Vieira; revisão técnica Enzo Scraphim e Thatyana de Faria Piola Scraphim. 6. ed. - São Paulo: Pearson Addison Wesley, 2011.

FELISBERTO, Adriano Araújo. **Análise de dados de músicas utilizando a mídia social Spotify**. 2019. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação). Curso de Tecnologia em Ciência da Computação. Universidade Estadual da Paraíba, Campus Campina Grande Disponível em: <http://dspace.bc.uepb.edu.br/jspui/handle/123456789/22061>. Acesso em: 23 de abr. 2021.

NIETZSCHE, F. **Crepúsculo dos Ídolos**. São Paulo: Cia das Letras, 2006.

GIL, A. C. Métodos e técnicas de pesquisa social. 6. ed. São Paulo: Atlas, 2008

GERHARDT, Tatiana; SILVEIRA, Denise. **Métodos de pesquisa. 2009. Monografia (II Curso de Especialização em Banco de Dados) – Universidade Federal do Rio Grande do Sul.** 2009. Disponível em: <http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>. Acesso em: 24 de abr. 2021.

GUEDES, Terezinha; ACORSI, Clédina; MARTINS, Ana; JANEIRO, Vanderly. **Estatística Descritiva – Projeto de Ensino.** Revista Rumores, São Paulo, v. 12, n.2, 2005. Disponível em <http://www.each.usp.br/rvicente/Guedes_etal_Estatistica_Descritiva.pdf>. Acesso em: 06 de mai. 2021.

IBGE. **Uso de Internet, televisão e celular no Brasil, 2019.** Disponível em < <https://educa.ibge.gov.br/jovens/materias-especiais/20787-uso-de-internet-televisao-e-celular-no-brasil.html>>. Acesso em: 21 de out. 2021.

KISCHINHEVSKY, Marcelo et al. **Música infinita: serviços de streaming como espaços híbridos de comunicação e consumo musical.** Compós. Associação Nacional dos Programas de Pós-Graduação em Comunicação. Disponível em: http://www.compos.org.br/biblioteca/test_2860.pdf. Acesso em: 26 out. 2019.

KLEINA, Nilton. **A história do MP3, o formato que espalhou a música digital.** 2019. Disponível em: <https://www.tecmundo.com.br/mercado/139452-historia-mp3-formato-espalhou-musica-digital-video.htm> Acesso em: 03 de abr. 2021.

LAZZARIN, Lilian N. A. **Técnica para mineração de textos na análise de sentimentos: Um estudo de caso em uma Instituição de Ensino Técnico, Tecnológico e Superior.** 2017. Monografia (II Curso de Especialização em Banco de Dados) – Universidade Tecnológica Federal do Paraná. Pato Branco, 2017.

MORAIS, Carlos. **Escalas de medida, estatística descritiva e inferência estatística.** Trabalho de Conclusão de Curso. Curso Instituto Politécnico de Bragança, Escola Superior de Educação, 2005. Disponível em: <https://bibliotecadigital.ipb.pt/bitstream/10198/7325/1/estdescr.pdf>. Acesso em: 23 de abr. 2021.

MOREIRA, Norma. **Ansiedade e Transtornos de Humor.** 2017. Disponível em: http://bio-neuro-psicologia.usuarios.rdc.puc-rio.br/assets/17_ansiedade_e_transtornos_de_humor.pdf. Acesso em: 13 de mar. 2021.

MORETTIN, Pedro; SINGER, Julio. **Introdução à Ciência de Dados: Fundamentos e Aplicações.** São Paulo: USP, 2019. Disponível em: <https://www.ime.usp.br/~jmsinger/MAE5755/cdados2019ago06.pdf>. Acesso em: 02 de mai. 2021.

NIETZSCHE, F. **Crepúsculo dos Ídolos.** Companhia de Bolso. 1889. Disponível em: <https://www.afoiceomartelo.com.br/posfsa/Autores/Nietzsche,%20Friedrich/Friedrich%20Ni>

etzsche%20-%20Crep%C3%BAsculo%20dos%20%C3%8Ddolos.pdf. Acesso em: 25 de out. 2021.

NORMAN-HAIGNERE, Sam; KANWISHER, Nancy; MCDERMOTT, Josh. **Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition**. 2015. Disponível em: [https://www.cell.com/neuron/fulltext/S0896-6273\(15\)01071-5](https://www.cell.com/neuron/fulltext/S0896-6273(15)01071-5). Acesso em: 01 de mar. 2021.

PRESSE, France. **Ouvir música causa liberação de dopamina, diz pesquisa**. 2011. Disponível em: <http://g1.globo.com/ciencia-e-saude/noticia/2011/01/ouvir-musica-causa-liberacao-de-dopamina-diz-pesquisa.html>. Acesso em: 01 de abr. 2021.

PLUTCHIK, P. (2001) **A Nature of Emotions**. *American Scientis*. 89:344-350.

RIBEIRO, Eduardo. **Qual é o papel do CD no mercado da música hoje?** 2017. Disponível em: https://www.vice.com/pt_br/article/wj8ygn/cd-mercado-fonograficoestudando-cena-3. Acesso em: 01 de abr. 2021.

ROMAN, N. T. **Emoção e a Sumarização Automática de Diálogos**. Tese de doutorado do Instituto de Computação da Unicamp. 2007. Disponível em: http://repositorio.unicamp.br/bitstream/REPOSIP/276233/1/Roman_NortonTrevisan_D.pdf. Acesso em: 24 de abr. 2021.

SANTOS, Mylena C. Gadelha; RAMOS, Rebecca Costa. **Aplicativos de música: o Spotify, as mudanças no mercado fonográfico e os filtros-bolha**. In: **XXXIX CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO**. 05 a 09 de setembro de 2016. Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. São Paulo, SP, 2016. Disponível em: <http://portalintercom.org.br/anais/nacional2016/resumos/R11-2913-1.pdf>. Acesso em: 02 de abr. 2021.

SOLAGNA, Emanuele Araujo; LAZZARETTI, Alexandre Tagliari. **Um estudo comparativo entre O mongodb e o postgresql**. 2016. 23f. Trabalho de Conclusão de Curso (Graduação em Tecnólogo em Sistemas para Internet). Curso de Tecnologia em Sistemas para Internet. Instituto Federal Sul-Rio-Grandense, Campus Passo Fundo, 2016. Disponível em: <https://painel.passofundo.ifsul.edu.br/uploads/arq/201607111805501015914198.pdf>. Acesso em: 23 de abr. 2021.

TAN, Ah-Hwee et al. **Text mining: The state of the art and the challenges**. In: **Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases**. 1999. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.6973&rep=rep1&type=pdf>. Acesso em: 24 de abr. 2021.

TKACH, D. (1998). **“Turning information into knowledge.”** a white paper from IBM. Disponível em: <http://www.partnersmarket.com/downloads/IBMResearchTextMiningWP.pdf>. Acesso em: 24 de abr. 2021.

THOMAZINI, Thamyres. **Mesmo com internet, uso do rádio é constante no Brasil**. 2017. Disponível em: <http://mescla.cc/2017/03/15/mesmo-com-internet-uso-radio-e-constante-no-brasil>. Acesso em: 01 de mar. 2021.

VICENTE, Eduardo. **Indústria da música ou indústria do disco? A questão dos suportes e de sua desmaterialização no meio musical.** Revista Rumores, São Paulo, v. 12, n.2, 2012. Disponível em <http://www3.usp.br/rumores/pdf/rumores12_12.pdf>. Acesso em: 15 de abr. 2021.

WAZLAWICK, Raul Sidnei. **Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico.** Revista de Sistemas de Informação da FSMA, n. 6, 2010, p. 3-10.

YOBERO, Czar. **K-Means Clustering Tutorial.** 2018. Disponível em: <https://rpubs.com/cyobero/k-means> Acesso em: 01 mai. 2021.

AGRADECIMENTOS

A princípio a Deus, pela força, saúde e por ajudar a superar as adversidades durante esse percurso.

A minha mãe (*in memoriam*), embora fisicamente ausente, sentia sua presença ao meu lado, dando-me força.

Ao professor Wellington Candeia de Araújo pelas leituras sugeridas ao longo dessa orientação, pela dedicação e por toda paciência envolvida durante o processo.

Aos professores do Curso de Ciência da Computação da UEPB, que contribuíram ao longo desses anos, por meio das disciplinas e debates, para o desenvolvimento desta pesquisa.

Aos funcionários da UEPB, pela presteza e atendimento quando nos foi necessário.

Aos colegas de classe pelos momentos de amizade e apoio.