UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIA DA COMPUTAÇÃO

LUIZ DELANDO SANTOS MOREIRA JÚNIOR

Comparando Predição de Popularidade de Podcast a Partir de Metadados, Conteúdo e Características de Áudio

Recife

2021

LUIZ DELANDO SANTOS MOREIRA JÚNIOR

Comparando Predição de Popularidade de Podcast a Partir de Metadados, Conteúdo e Características de Áudio

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Processamento de Sinais e Reconhecimento de Padrões.

Orientador (a): Prof. Dr. Giordano Ribeiro Eulálio Cabral

Coorientador (a): Dr. Ricardo Enrique Pereira Scholz

Recife

2021

Catalogação na fonte
Bibliotecária Fernanda Bernardo Ferreira, CRB4-2165

M838c Moreira Júnior, Luiz Delando Santos
Comparando predição de popularidade de Podcast a partir de metadados, conteúdo e características de áudio / Luiz Delando Santos Moreira Júnior. – 2021.
76 f.: il., fig., tab.

Orientador: Giordano Ribeiro Eulálio Cabral.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2021.
Inclui referências e apêndices.

1. Processamento de Sinais e Reconhecimento de Padrões. 2. Podcast. 3. Aprendizado de Máquina. I. Cabral, Giordano Ribeiro Eulálio (orientador). II. Título.

006.4 CDD (23. ed.) UFPE- CCEN 2021 - 104

Luiz Delando Santos Moreira Júnior

“Comparando Predição de Popularidade de Podcast a Partir de Metadados,Conteúdo e Características de Áudio”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 09/03/2021.

BANCA EXAMINADORA

Prof. Dr. Geber Lisboa Ramalho
Centro de Informática/ UFPE

Prof. Dr. Tiago Fernandes Tavares
Departamento de Engenharia de Computação e
Automação Industrial / UNICAMP

Prof. Dr. Giordano Ribeiro Eulalio Cabral
Centro de Informática / UFPE
(Orientador)

Dedico este trabalho aos meus pais que sempre me apoiaram e minha esposa e filha que são inspirações para eu finalizar este ciclo.

AGRADECIMENTOS

Parte da jornada é o fim. E aqui finalizo mais um dos vários ciclos que vivemos durante a vida, demonstrando minha eterna gratidão por todos que me acompanharam de forma direta e indireta durante todo esse processo.

Em especial agradeço a:

A Deus, sempre presente no meu coração e em todos momentos difíceis dessa caminhada.

A minha Namorada/Noiva/Esposa Karina Viegas, ela conseguiu ser tudo que era possível durante todo esse processo, por termos dividido todos os momentos de pressão, desespero e alegria percorrido durante este ciclo e por nos permitir ganhar o maior presente que poderíamos ganhar que é nossa filha Ada.

Aos meus Pais e Irmã, que sempre me apoiaram não somente neste ciclo mas em todos os outros vividos, sempre com palavras e apoio durante toda a vida.

Ao meu orientador Giordano Cabral que me acompanhou durante toda a graduação e topou está junto comigo neste ciclo e ao meu coorientador Ricardo Scholz que mesmo chegando no final da jornada, foi de uma ajuda incomparável e sem ele nada disso teria acontecido.

A Bernardo Júnior, que compartilhou vários momentos da pesquisa, entrando na madrugada programando e ajudando em tudo que fosse possível.

Aos meus amigos Flaviano, Jessyca, Mago, Cleidane, Beth, Eli, Clau, Ni, Diogo, Rafael, Teago, Daivid, que sempre me apoiaram para que eu conseguisse terminar logo esse ciclo.

Ao grupo de pesquisa MUSTIC por em todas reuniões sempre estarem dispostos a ajudar.

Ao corpo docente e tecnico do CIn e ao CNPq.

RESUMO

Podcasts são uma forma relativamente nova de conteúdo, que vem ganhando cada vez mais popularidade no cenário nacional e mundial. Por exemplo, a Spotify introduziu os podcasts em seu aplicativo e investiu 340 milhões de dólares na compra de produtoras como a Gimlet e a Anchor. Além disso, em 2019, o grupo Globo lançou diversos dos seus programas e telejornais nesse novo formato, feito para o público ouvir quando e onde quiser. Mas, criar podcasts não é uma particularidade desses grandes grupos. Devido à facilidade no processo de produção e publicação, produtores independentes também ocupam uma fatia desse mercado, fazendo com que exista um grande volume de conteúdo sendo produzido diariamente. Segundo a Whitner, em 2020 foi alcançada a marca de 1 milhão de canais ativos e mais de 30 milhões de episódios publicados. Identificar quais são os podcasts com mais chances de sucesso é uma atividade não trivial é extremamente importante, pois essa informação pode ser usada para modificar a forma com que os episódios são exibidos nos aplicativos ou definir investimentos a serem realizados, por exemplo. O objetivo principal deste trabalho é identificar quais caracerísticas apresentam melhor desempenho na predição da popularidade dos episódios. Para isso, construímos uma base com mais de 800 episódios, totalizando mais de 600 horas de podcasts transcritos. Então, comparamos a predição de popularidade dos podcasts utilizando diversas estratégias de aprendizagem de máquina, bem como características de áudio, do conteúdo falado e dos metadados disponíveis no RSS Feed de cada canal. Estes últimos são utilizados pelos próprios produtores para inserir informações relevantes sobre o podcast.

Palavras-chaves: Podcast. Aprendizado de máquina. Processamento de texto. Processamento de Áudio. Predição de popularidade. Seleção de características.

ABSTRACT

Podcasts are a relatively new form of content, which has been gaining more and more popularity in the national and worldwide scene. For example, Spotify introduced podcasts to its app and invested $340 million to buy producers like Gimlet and Anchor. In addition, in 2019, Grupo Globo made several of its programs and news in this new format, made for the public to listen when and where they want. But, creating podcasts is not a big group feature. Limited to the ease in the production and publication process, independent producers also occupy a slice of this market, making it possible to have a large volume of food content daily. According to Whitner, in 2020 it reached the mark of 1 million active channels and more than 30 million episodes published. Identifying which podcasts are most likely to succeed is a non-trivial activity is extremely important, as this information can be used to modify the way episodes are refined in applications or define investments to be made, for example. The main objective of this work is to identify which characteristics have the best performance in predicting the popularity of episodes. For that, we built a base with more than 800 episodes, totaling more than 600 hours of transcribed podcasts. Then, it compares the popularity prediction of podcasts using different machine learning methodologies, as well as audio, spoken content and metadata characteristics available in each channel's RSS Feed. The latter are used by the producers themselves to insert relevant information about the podcast.

Keywords: Podcast. Machine learning. Text processing. Audio processing. Popularity prediction, Feature selection.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Neste capítulo, contextualizamos e apresentamos o problema abordado na pesquisa e, em seguida, apresentamos as motivações e objetivos que orientam nossa investigação.

1.1 CONTEXTUALIZAÇÃO E MOTIVAÇÃO

Segundo Cold (2006), os podcasts são uma forma relativamente nova de conteúdo de áudio disponibilizado em um feed RSS, um protocolo que permite o compartilhamento de conteúdo com outras aplicações. Isso permite ao usuário acompanhar informações de diversos canais em um único lugar, com aplicativos chamados de agregadores. Cada vez mais os podcasts vem ganhando popularidade atingindo marcas expressivas de consumo e produção. Isso se reflete no interesse de grandes marcas por esse tipo de mídia. Por exemplo, recentemente, Google lançou um agregador próprio para ouvir podcasts enquanto o Spotify inseriu esta funcionalidade em seu aplicativo e vem oferecendo cada vez mais destaque a ela em sua interface.

As plataformas de streaming estão se consolidando definitivamente como ferramentas para ouvir podcasts. Particularmente, 2019 foi um ano muito importante para a consolidação dos podcasts no cenário mundial e brasileiro. Por exemplo, o Spotify realizou um invetimento de 340 milhões de dólares na compra da Gimlet, uma das maiores produtoras de podcast do mundo (BESIK, 2020), enquanto o Grupo Globo de Comunicação anunciou sua entrada nesse nicho de mercado transformando muitos dos seus programas de jornalismo, esporte e entretenimento em canais de podcasts (G1, 2019).

Os podcasts são compartilhados em arquivos de áudio, que são disponibilizados através de um link dentro do RSS Feed, um arquivo XML, estruturado para conter os metadados do canal, como os produtores, os apresentadores, o resumo, o título e as descrições. Além dessas informações, ele contém uma listagem com todos os episódios disponíveis, elencando, para cada episódio, informações como o título, o resumo, a descrição, as palavras-chave ou o link para download e reprodução. A partir desses dados, os agregadores podem indexar os podcasts.

Porém, como o processo de criação de podcasts não necessita de equipamentos caros, além dos muitos podcasts produzidos e distribuídos profissionalmente, há também uma grande quantidade de produtores independentes que cuidam de todo o processo de roteirização, gravação,

edição e publicação. Isso traz uma enorme diversidade aos episódios, desde a variedade de gírias, sotaques e formatos - como uma conversa em uma mesa de bar, entrevistas elaboradas ou até minisséries - até a qualidade da produção e dos efeitos sonoros, com vinhetas que podem aumentar o engajamento e o nível da produção, por exemplo (BESSER; LARSON; HOFMANN, 2010).

Segundo Besser, Larson e Hofmann (2010), essa mídia envolve dois tipos de fontes principais de informações que podem ser exploradas por um sistema de recomendação. Primeiro, o próprio áudio do episódio. Esse conteúdo pode ser explorado de algumas formas diferentes, como a extração de características de áudio, ou até do texto, que pode ser extraído a partir de uma transcrição automática. Abordaremos essas possibilidades nos próximos capítulos. Segundo, os metadados, preenchidos pelo criador do podcast em dois níveis: os detalhes do programa, contendo informações como nome e descrição, e as informações do episódio, como título, resumo, duração e etc., presentes no RSS Feed, conforme se vê no exemplo do [apêndice 1].

Os podcasts ainda têm sido pouco estudados(CLIFTON et al., 2020), como veremos no segundo capitulo. Existem poucos trabalhos relacionados ao tema de como processar os podcasts, seja para extração de características, sumarização do conteúdo ou recomendação. Esse cenário nos motivou a explorar como as ferramentas disponíveis nas áreas de Recuperação de Informação Musical (MIR) e Processamento de Linguagem Natural (PLN) podem nos ajudar a entender os podcasts.

1.2 PROBLEMA DE PESQUISA E HIPÓTESES

Identificar o sucesso de uma postagem em uma rede social é um desafio complexo, pois depende de interações complexas entre a qualidade do conteúdo, as interações com os usuários (LERMAN; HOGG, 2010), Mas predizer essa popularidade pode influenciar a forma como as plataformas exibem essa publicação ou o investimento que uma música, por exemplo, pode receber conquistando uma vantagem competitiva para os produtores. Esse é um tema estudado em diversos cenários, como a predição de popularidade para ranquear reportagens online. Nesse nicho, Alexandru (TATAR et al., 2014) identificou que os comentários na publicação são uma das características importantes a serem analisadas, estudando a interação dos usuários para medir o compartilhamento de vídeos. Haitao Li (LI et al., 2013) segue uma abordagem voltada também para a interação dos usuários e estuda a relação entre as curtidas dadas no início da

postagem e a sua popularidade. Já Francesco Gelli (GELLI et al., 2015) segue uma abordagem voltada para o conteúdo, em que estuda a predição de popularidade de imagens nas redes sociais, analisando a imagem e extraindo característica de sentimento para realizar a predição. Traçando um paralelo entre esses estudos e o aumento no consumo de podcasts, vemos que os podcasts começam a enfrentar problemas semelhantes, para os quais a predição de popularidade pode ajudar, desde como os episódios são apresentados para o público, até no auxílio a sistemas de recomendação.

Portanto, neste trabalho, assumimos como hipóteses: Que o processo de predição de popularidade passa não só pela forma que o episódio é divulgado na rede, mas pelo seu conteúdo. Podem ser características que auxiliem no processo de predição de popularidade as informações que estão contidas no episódio, como todo o conteúdo falado, as características presentes no áudio - como qualidade e informações acústicas - e até as informações previamente disponibilizadas pelos produtores - por exemplo, a presença de título, descrição, duração e tipo do episódio.

1.3 OBJETIVOS

1.3.1 Objetivos Gerais

Este trabalho visa explorar a mídia podcast com a ótica da computação. Para alcançar isto, temos como objetivo principal realizar uma análise comparativa de sistemas de aprendizagem de máquina para predição de popularidade dos Podcasts, utilizando as características textuais, as características de áudio e os metadados.

1.3.2 Objetivos Especificos

1. Criar um sistema para coletar os dados do Rss Feeds dos episódios populares e não populares;
2. Construir uma base de episódios em MP3, rotulados por popularidade;
3. Construir uma base com os textos do conteúdo falado durante os episódios;
4. Criar um sistema capaz de extrair as características de áudio e texto dos episódios;

5. Definir as métricas e usá-las para comparar a classificação de popularidade utilizando abordagens de aprendizagem de máquina.

2 CARACTERIZAÇÃO DO PROBLEMA

Neste capítulo vamos apresentar a área de podcast e quais os problemas que essa aplicação traz para área de computação e uma visão sobre os requisitos para construção de uma solução.

Podcasts são uma áudio série disponibilizadas na internet através de um RSS-Feed (TSAGKIAS; LARSON; RIJKE, 2010), comumente publicados com uma cadência regular nos mais diversos níveis de formalidade. Este tipo de conteúdo tem ganhado cada vez mais relevância no cenário mundial e nacional. Somente nos Estados Unidos, 75% da população é familiarizada com o termo podcast, e 55% já ouviu pelo menos um episódio. Em 2020, os podcasts atingiram a marca de 1 milhão de canais ativos e mais de 30 milhões de episódios publicados (WHITNER, 2021). O perfil de consumo do ouvinte brasileiro segundo a podpesquisa 2018 é de que 77% entrevistados foram levados a ouvir podcasts pois lhe permitem realizar outras atividades enquanto escutam os episódios, o que reflete diretamente com um dos principais slogans de divulgação deste formato que é: "Escute quando, como e onde quiser". Segundo a mesma pesquisa, os ouvintes gastam entre 2 e 3 horas do seu dia escutando podcasts, o que representa uma parcela significativa do tempo acordado de uma pessoa, essas estatísticas reforçam a importância que os podcasts estão ganhando na vida das pessoas, concorrendo com escutar músicas e outros hobbies.

Com o aumento da audiência cada vez mais pessoas têm tentando produzir conteúdo para este tipo de mídia, acarretando na criação de novos shows dos mais diverso níveis de produção desde de publicações amadoras, com equipamentos como um celular para gravação dos episódios, até gravações profissionais que se utilizam de grandes estúdios para realizarem suas gravações onde existe toda uma preocupação com a qualidade do áudio e o roteiro dos episódios. Este processo gera uma série de desafios, por exemplo, como definir a qualidade de um podcast? Quais as chances de um podcast se tornar popular? Como definir a categoria dos podcasts automaticamente? Como realizar um processo de filtragem de episódios? Isto são problemáticas não visíveis para o usuário mas que podem impactar na sua relação de consumo. Diferente como acontece com os filmes e músicas onde os sistemas de recomendação já são a principal forma de conhecer novos episódios e programas, no universo dos podcasts mais de 80% dos ouvintes descobrem um novo podcast a partir da indicação de outros hosts. Entender esse processo de recomendação e classificação dos episódios e programas pode ser um grande vantagem competitiva para os grandes players de streaming como Deezer, Google e Spotify

que realizam uma corrida para conseguir oferecer os melhores algoritmos de recomendação e propiciar uma melhor experiência para os seus usuários.

2.1 DESAFIOS

Atualmente um dos grandes desafios da computação está em entender o gosto das pessoas, cada vez mais os sistemas de recomendação tem se tornado indispensável para o processo de seleção de informações dado o oceano de opções que são oferecidos diariamente para o usuário, sejam eles nos filmes ou músicas e com o crescimento dos podcasts isto não seria diferente. Para isso, conhecer a fundo as características de um podcasts é de extrema importância para que com base nisso possamos relacionar os gostos com as informações individuais de cada episódio trazendo diversos benefícios para a área de podcasts e com isso temos diversos outros desafios. Como a criação de uma base de dados rotulados, visto que ainda não existem muitos trabalhos nesse campo. Por isso a criação de uma base rotulada é um grande desafio, pois como podemos definir quais são os podcasts populares? É possível definir que os 50 podcasts mais ouvidos são os mais populares e qual o grau de separação dos podcasts populares para os não populares? Podemos dizer que todos os podcasts abaixo dos 50 mais ouvidos são os não populares? Determinar essa linha de corte é uma atividade bastante complexa e pode trazer vieses para a nossa análise.

O problema do produtor é algo encontrado na análise de músicas que possuem o mesmo estúdio e produtor onde os resultados podem estar vinculados não a música mas sim ao estúdio que realizou a gravação, e esse desafio pode ser encontrado também no nosso cenário dos podcasts onde os diversos canais são editados por um único produtor, por exemplo o estúdio Radiofobia que editam diversos programas. O que gera mais uma dificuldade para o nosso cenário de realizar a rotulação dos programas e identificação de quais as características que podem distinguir os podcasts populares dos não populares

Além dos desafios mais gerais de preparação da base dos dados cada uma das subáreas como o processamento do áudio, processamento do texto e o processo de recomendação de episódios possuem um campo de estudo específico e este trabalho tem como principal provocação unir todas essas linhas e realizar a análise de predição de podcasts, a seguir vamos apresentar alguns dos desafios presentes em cada um destes campos de estudos:

1. Processamento de áudio: A essência do podcast é ser programas de áudios e aqui temos

os primeiros desafios que é a utilização de técnicas comumente utilizadas na área de MIR(Music Information Retrieval) é com base nessas bibliotecas que os primeiros estudos foram baseados para tentar entender o conteúdo do áudio dos podcasts então como definir a qualidade do áudio de um podcasts nessas bibliotecas e qual a influência dessas características para a popularidade ou para a retenção do usuário dentro de um episódio, isso são perguntas que o processamento do áudio pode nos ajudar a resolver. Uma outra dificuldade está relacionada com a classificação automática dos podcasts, na área de música podemos definir, gênero, dançabilidade tudo a partir do áudio da música então quais seriam os atributos do áudio que poderiam definir o humor de um podcasts, fazer a distinção e rotulação automática dos episódios com relação a gênero, reconhecimento dos apresentadores, presença ou não de vinheta entre outras características.

2. Processamento do texto: O primeiro desafio ligado a esta área é a recuperação do texto que para o processo de transcrição comumente são utilizados ferramentas de speech to text capazes de recuperar o texto de toda a fala dentro de um arquivo de áudio, grandes empresas como Google, IBM, Amazon oferecem esse serviços em suas plataformas de cloud as a service, mas um grande problema está no curto de realizar essa transcrição para o grande volume de podcasts. Mas de posse dessa informação o processamento de texto tem um grande potencial de enriquecer os dados dos podcasts pois os estudos voltados para esta área podem auxiliar no processo de entendimento dos episódios como é o exemplo da sumarização do episódio, os podcasts não possuem limite de duração então um grande desafio está em automaticamente conseguir sumarizar criando um trailer do podcasts para efeito de divulgação. Identificação de palavras chaves com objetivo de melhorar o processo de descoberta de um podcast, atualmente a principal fonte de informação de um episódio são os seus metadados, disponibilizados em seu rss-feed mas muitas às vezes, os resumos e descrições inseridos pelo produtor do programa não contemplam todo conteúdo do podcast além de que o processo de identificação do trecho do episódio que contempla determinado conteúdo também é um desafio que o processamento de texto pode auxiliar a entender.

E a integração de todas essas áreas de estudos ligados diretamente ao podcasts trazem um grande desafio de multidisciplinaridade para este contexto em oferecer uma solução capaz de unir os diversos campos de estudos para a melhoria do processo de recomendação e classificação dos podcasts.

2.2 POTENCIAIS USUÁRIOS

Este trabalho traz uma leitura inicial sobre a área de podcasts e a aplicaçãoque um sistema de podcasts podem trazer para os seguintes perfis de usuários:

1. Os produtores que estão interessados em extrair informações técnicas a respeito da sua gravação com o objetivo de comparar a qualidade do material produzido e obter dicas de como melhorar o processo de gravação para que seja possível alcançar mais ouvintes. Então essa ferramenta pode ser bastante útil para esse perfil de usuário e nosso estudo irá experimentar características do áudio e mapear qual a sua relação com a popularidade, servindo como alicerce para trabalhos futuros mapearem e criarem visualizações com esse objetivo. Um outro caso de uso ainda para os produtores é o processo de transcrição dos podcasts que é de grande uso para o processo de pós-produção onde eles podem utilizar o texto do episódio para o processo de criação de resumo e divulgação de cada um dos episódios.

2. Ouvintes: O processo de busca de um novo podcast ou seleção de um episódio que aborda um tema específico não é uma atividade das mais simples, pois o processo de recomendação de podcast ainda está em fase inicial e vem se evoluindo rapidamente com empresas como Spotify e Deezer entrando neste mercado. Mas essa pesquisa tem como objetivo estudar características que descrevem um podcasts que podem auxiliar no processo de seleção de podcasts para cada um dos usuários.

2.3 CRITÉRIOS DE ACEITAÇÃO

No processo da análise e criação de uma classificador de podcasts populares é esperado que esses sistemas ofereçam informações para os perfis citados anteriormente. Trazendo uma análise comparativa entre os diversos tipos de dados que podem ser utilizados para avaliar os episódios fornecendo uma série de informações sobre os resultados do processamento via áudio, texto e metadados. E para isso é necessário realizar algumas etapas durante o processo de desenvolvimento que serão descritas a seguir:

- Possui uma base de dados rotulados por popularidade, para que seja possível aplicar técnicas de aprendizagem de máquina supervisionado.

- No processamento de áudio, Identificar quais características de áudio podem ser utilizadas para mensurar a qualidade de um áudio?
- Utilização das seguintes parâmetros de avaliação para comparar os resultados obtidos com os experimentos de aprendizagem de máquina, por exemplo, acurácia, precisão, recall e F1-score;
- Transcrever os episódios de podcasts, dado é uma mídia essencialmente de áudio para realizar as análises de processamento de texto é necessário que sejam aplicadas técnicas de speech to text(JUANG; RABINER, 2005) com o objetivo de obter a transcrição dos episódios;
- Identificar as palavras chaves e traçar relações com os podcasts populares objetivo deste trabalho com intuito de identificar se algumas palavras são mais utilizadas além de ser um primeiro passo para entender qual a influência do conteúdo na popularidade do podcasts
- Comparar os resultados obtidos a partir do áudio com os resultados alcançados a partir de todo conteúdo falado.

3 ESTADO DA ARTE

Este capítulo faz uma revisão da literatura, sobre os trabalhos relacionados, o que são os podcasts, e o seu cenário atual. O capítulo também aborda as técnicas de processamento de linguagem natural que utilizamos para processar o texto dos episódios e as técnicas de processamento de áudio que são comumente utilizadas na indústria da música, e que utilizamos neste trabalho. Por fim, faremos uma revisão sobre o aprendizagem de máquina e as métricas que utilizamos para avaliar o experimento

3.1 TRABALHOS RELACIONADOS

Neste capítulo, abordaremos os principais trabalhos relacionados a podcasts na área da computação. O aumento no consumo de podcasts nos últimos anos fez com que grandes empresas começassem a investir nesse tipo de mídia, trazendo novos produtores de conteúdo e pessoas interessadas em estudar a área. Por exemplo, o Spotify, uma das maiores empresas do ramo, patrocinou a conferência RecSys em 2020, organizando um workshop específico para o tema de podcasts, no qual foram apresentados novos trabalhos para a área.

Os trabalhos de Tsagkias (TSAGKIAS; LARSON; RIJKE, 2010; TSAGKIAS et al., 2008) propõem um framework para análise das preferências de podcast, fornecendo um conjunto de características que pudessem definir a preferência do usuário. Essas características são divididas em quatro grandes partes:

- Conteúdo do podcast: reúne aspectos relacionados ao conteúdo, como, por exemplo, se o episódio é sobre um tema específico ou o canal sempre trata esse tema, se contém discussões e opiniões, se são enciclopédias e se as fontes são citadas além da regularidade em que os episódios são publicados.

- Apresentador: são as características pessoais dele, como, por exemplo, algum sotaque que possua, qual é a sua fluência e velocidade na fala, qual tipo de sentimento e humor ele costuma transmitir e seu currículo - a quem ele está ligado e se existe influência fora do contexto do podcast.

- Contexto do podcast: quais informações externas refletem no canal, como, por exemplo, se existe algum grande parceiro do podcast, se é um programa republicado das rádios,

se recebem muitos comentários.

- Execução técnica: aspectos como presença de efeitos sonoros, vinhetas, sons ambientes presentes na gravação ou preenchimento de todas as informações para divulgação.

Para validação desse framework, eles utilizaram as características presentes no RSS Feed de cada episódio, realizando uma classificação binária entre podcasts populares e não populares.

Já Mizuno, Ogata e Goto (2008) utilizaram uma ferramenta denominada PodCastle para propor um método de busca de episódios similares, utilizando o conteúdo falado durante o episódio, pois, até então, o mais comum era a utilização de informações bibliográficas ou tags definidas pelo usuário. Para isso, eles realizaram um pré-processamento nas buscas, usando uma versão melhorada do tradicional TF-IDF, para que fosse possível minimizar a interferência dos erros de transcrição que acontecem no processo. Para validação, o modelo foi testado e comparado com o estado da arte para extração de palavras chaves, e obteve um aumento de performance na maioria dos cenários. Os cenários em que não houve melhoria foram os casos em que o processo de transcrição não obteve uma acuracidade satisfatória nesse processo.

No trabalho de Yang et al. (2019), há a proposição de um sistema de recomendação focado nas características não textuais dos podcasts. Nele, eles apresentam o ALPR (Adversarial Learning-based Podcast Representation), um método para capturar características comumente encontradas no cenário da música como "energia" e "seriedade". Para definição dessas características, foi usada uma abordagem colaborativa utilizando a ferramenta Amazon Mechanical Tork, tornando possível obter um conjunto de trechos de áudio rotulados, indicando o nível de seriedade e energia de cada trecho.

Para validação desse modelo, foi proposta a utilização do ranking de podcast disponibilizado pelo site Chartable (uma ferramenta de analytics de Podcast) para obtenção de uma lista dos episódios populares e não populares utilizados no processo de classificação. Por fim, o artigo combina características textuais para entender qual a melhora causada pela sua hipótese, fazendo uma predição de popularidade a partir da combinação do texto e do áudio, e realizando variações na quantidade de tempo - entre 1 e 10 minutos - utilizado para analisar os episódios. Os resultados foram favoráveis ao modelo combinado, em comparação aos demais.

A pesquisa de Clifton et al. (2020) apresenta a maior base de dados relacionados a podcasts, construída em colaboração com a TREC Text Retrieval Conference. Neste trabalho, é apresentada uma base com mais de 100.000 episódios, juntamente com seus arquivos de áudio, totalizando mais de 47.000 horas. Além dos arquivos de áudio, a base fornece também

as transcrições automáticas geradas a partir da Google Cloud Plataform Speech-to-Text. Estes episódios foram separados aleatoriamente, com uma grande variedade nos tamanhos, regiões de produção e qualidade do áudio. A base selecionou episódios publicados entre janeiro de 2019 e Março de 2020, filtrados por idiomas da língua inglesa. Para os episódios não profissionais, limitou-se o tamanho máximo em 90 minutos, removendo-se, ainda, episódios que possuíam um percentual de fala menor que 50%. Isso foi possível devido a um classificador automático que verifica a presença de ruído, fala ou música a cada segundo. Esta base ainda não foi disponibilizada publicamente. Ela fez parte de um desafio lançado pelo Spotify na conferência TREC 2020.

Nazari et al. (2020) desenvolveram um sistema de recomendação de podcasts baseado nas preferências musicais do usuário. O sistema se propõe a ser uma alternativa para o problema de "cold-start", ou seja, o que recomendar para usuários que acabaram de entrar na plataforma, e sobre os quais não temos um histórico dos podcasts escutados. Os experimentos foram conduzidos com uma base de 17 milhões de usuários, com informações demográficas, metadados e o gosto musical, além dos podcasts de interesse fornecidos pelo Spotify. Os experimentos offline e online indicaram que os gostos musicais podem predizer os podcasts de interesse.

O trabalho de Sharpe (2020) traz uma revisão sobre o RSS Feed, formato utilizado para publicação e divulgação dos podcasts online. Neste trabalho, podemos observar como os criadores de podcasts têm usado, atualmente, os campos disponíveis nos metadados. Por exemplo, no preenchimento do campo "categorias", existe uma tendência dos produtores de incluir mais de uma categoria, resultando em podcasts presentes em várias categorias, o que dificulta o processo de recomendação.

Um outro ponto interessante é o campo "itunes:type", que possui dois tipos, os "sequenciais"e o "episódio". Esses tipos deveriam estar ligados à forma com que consumimos cada um desses canais. Pois os sequenciais deveriam ser ouvidos dos mais velhos aos mais novos, enquanto o outro tipo não possui uma recomendação forte, sendo mais comumente escutado do mais novo para o mais velho. A observação importante aqui é que há indícios de que existem problemas no preenchimento desses metadados, já que, de acordo com o catálogo do Spotify, 97,5% de todos os episódios listados são categorizados como do tipo "episódio".

3.2 PROCESSAMENTO DE LINGUAGEM NATURAL

Comumente chamado por sua sigla em inglês, NLP (Natural Language Processing) é uma área da inteligência artificial e da linguística direcionada para o entendimento da linguagem humana (KHURANA et al., 2017). Essa área surgiu pela necessidade de facilitar a comunicação humana com a máquina. No início, para nos comunicarmos com as máquinas, tínhamos que saber instruções específicas responsáveis por disparar uma série de execuções que resultavam na realização de uma função. Atualmente, estão em alta os chatbots, programas de computador desenvolvidos para interagirem com o usuário utilizando linguagem natural (SHAWAR; ATWELL, 2007). Inclusive, existe uma competição anual de inteligência artificial, o Prêmio Loebner, que premia a interação mais próxima com a de um ser humano.

Segundo Khurana et al. (2017), o processamento de linguagem natural pode ser classificado em duas partes: a primeira é a geração de linguagem natural, e a segunda é o entendimento da linguagem natural. A imagem X mostra as tarefas envolvidas em cada uma dessas partes.

- Geração de linguagem natural é um processo de construção de frases e textos, com o objetivo de se comunicar com o interlocutor. É diferente de usar a língua portuguesa em um programa de computador, pois o fato de gerar uma mensagem de erro ou um texto pré-programado não faz disso uma geração de linguagem. A geração de linguagem acontece quando se usam artifícios gramaticais para produzir um texto, com base em um conjunto de entradas. Essa área é muito utilizada no jornalismo, para criação de materiais como o trabalho de Chen e Mooney (2008), que relata a criação de um comentarista de futebol para partidas simuladas. Já Thomason et al. (2014) aplicou a geração automática de textos para criar descrições de acordo com as imagens que lhe eram apresentadas.

- Compreensão da Linguagem Natural, muito conhecida pelo seu termo em inglês, Natural Language Understanding (NLU), é o processo de entendimento dos textos, atualmente muito utilizado por grandes empresas que oferecem serviços cognitivos, nos quais é necessário que a máquina possua a capacidade de entender contextos, entidades e intenções. Exemplos atuais são serviços como o Watson, da IBM, e o Luis, da Microsoft, voltados principalmente para criação de chatbots (ZUBANI; IVAN; GEREVINI, 2020). Aplicações, como reconhecimento de emoção e de sentimentos, podem se utilizar de NLU na predição, com base nas palavras capturadas da fala do usuário (POPOVIC et al., 2020).

O Processamento de Linguagem Natural é aplicado nos mais diversos campos, como medicina (SANDOVAL et al., 2019), direito (ALETRAS et al., 2016) e jornalismo (OSHIKAWA; QIAN; WANG, 2018). Mas independente da área de aplicação, existem algumas técnicas que são comumente utilizadas para pré-processamento, extração das características e classificação. O objetivo desta seção é apresentar algumas das principais técnicas, como Part-of-speech Taggers, chunking, Named Entity Recognition, Lemmatization, stopwords, Bag-of-Words e etc.

Stopwords são palavras que possuem a mesma probabilidade de aparecer em documentos relevantes e em documentos não relevantes. Assim, comumente aparecem com grande frequência, mas são considerados irrelevantes para os resultados, um exemplo das palavras são: "as", "e", "os", "de"etc. A remoção dessas palavras, antes de realizar a análise dos textos, pode diminuir seu tamanho, diminuindo a complexidade do processamento e aumentando a efetividade dos processos de classificação (WILBUR; SIROTKIN, 1992). Segundo Schäuble (2012), de 30% a 50% das palavras contidas em um texto são stopwords.

Lemmatization é a tarefa responsável por encontrar o radical das palavras, retirando a conjugação verbal e padronizando o gênero dos adjetivos e substantivos. Por exemplo, "Naturalmente» "Natural» "Nat"ou "gatosgatas» "gato"ou "tivertenho» "tem". Esse processo evita que algoritmos considerem palavras com o mesmo significado como palavras diferentes, possibilitando uma redução na complexidade e um aumento da eficácia (INGASON et al., 2008).

Chunking, também conhecido como parsing, é uma série de processos em que, primeiramente, dentro de todo o texto são identificados “Chunks” segmentos, para em seguida, eles sejam transformados em frases para que seja realizada uma classificação, normalmente, como verbal ou nominal. (KUDO; MATSUMOTO, 2001).

Named Entity Recognition (NER) consiste em rotular as palavras em categorias como “Local”, “Data”, “Pessoa”, “Organização”, entre outras (RITTER et al., 2011). É possível propor uma rotulação específica de entidades para o domínio dos dados, pois, em alguns casos, as pessoas não utilizam a linguagem da forma padrão. Em sites como o twitter, é muito comum ver um uso diferente da linguagem. Assim, uma anotação específica pode permitir um ganho de performance ao classificar esses textos.

Part-of-speech Tagging é a tarefa de rotular cada palavra com uma marcação única que indica a função sintática dela na frase em que está inserida. Nessa abordagem, as palavras são rotuladas comumente como pronome, advérbio, adjetivo, verbo, sujeito, entre outras (VOUTILAINEN, 2003).

Bag-of-words (BoW) é um método que tem por objetivo gerar uma representação nu-

mérica de um conjunto de palavras, de forma que cada palavra única seja apresentada com um número, dentro de uma sequência de palavras que compõem o texto em questão. Esse número é calculado de acordo com a quantidade de vezes que a palavra se repete no texto. O método não considera as características sintáticas. Ele é comumente utilizado em algoritmos estatísticos para recuperação de informação (IRIE; SCHLÜTER; NEY, 2015).

Term Frequency-inverse document frequency (TF-IDF), assim como na BoW, esse método tem por objetivo gerar uma representação numérica do texto, mas, ao invés de considerar somente a quantidade de repetições no texto, como no BoW, essa estratégia adiciona mais uma camada de processamento, com intuito de não beneficiar as palavras que se repetem muito, em detrimento das demais, no processo de classificação. Essa técnica pode ser dividida em duas partes:

1. *Term frequency*: representação da frequência das palavras, dentro de um texto ou documento.

2. *Inverse document frequency*: definição de uma métrica para a quantidade de informação que cada palavra representa para o texto, gerando um "peso"da relevância da palavra na prepresentação numérica dela no texto (RAMOS et al., 2003).

Análise de sentimento é uma técnica para extrair o sentimento de um texto sobre um determinado assunto. Comumente, os textos são classificados entre positivos e negativos (KHURANA et al., 2017). Existem algumas técnicas diferentes para determinar o sentimento de um texto, mas em boa parte delas o princípio está em rotular as palavras ou sentenças em uma escala de negativo e positivo e, em seguida, definir o sentimento do texto com base nessas características. Alguns exemplos de métodos são: Emoticons (HUANG; YEN; ZHANG, 2008), LIWC (FILHO; PARDO; ALUÍSIO, 2013), Sentistrength (THELWALL et al., 2010), SentiWordNet (DENECKE, 2008), entre outros (ARAÚJO et al., 2013).

Detecção de emoção possui uma abordagem semelhante à utilizada na análise de sentimento, porém as classificações são feitas de acordo com modelos de emoções, que fornecem uma base de conhecimento capaz de associar os resultados das análises semânticas com possíveis emoções. Sua utilização está bastante associada às mídias sociais, com intuito de fornecer uma interpretação sobre os textos fornecidos pelos usuários (EZHILARASI; MINU, 2012).

3.3 ANÁLISE DE ÁUDIO

Humanos classificam áudio a todo tempo, sem esforço, reconhecendo uma voz ao telefone ou identificando a diferença entre uma campainha e um toque de celular. Essas são atividades rotineiras para o cérebro humano. A área de processamento de áudio visa tornar as máquinas capazes de realizar essas tarefas, cobrindo diversos cenários, como a fala (YU; DENG, 2016), a música (DOWNIE, 2003) e os sons ambientes (SALAMON; BELLO, 2015).

A forma como consumimos os podcasts é similar à forma como consumimos música, pois ambos são escutados, normalmente, com o objetivo de entretenimento. Portanto, algoritmos e técnicas utilizadas para modelar música podem indicar caminhos a serem seguidos, visto que a análise de áudio voltada para podcasts ainda é um tema pouco investigado (MARTIKAINEN, 2020). Existem poucos trabalhos que associam a extração de características do áudio do podcast com a sua popularidade, logo, representações comuns, como MFCC(LOGAN et al., 2000), utilizadas no processamento de áudio em geral, tem potencial para atingir bons resultados na análise de podcasts. Nesta seção, iremos abordar os métodos de reconhecimento automático de fala e MFCC (YANG et al., 2019)

3.3.1 Reconhecimento Automático de Fala

Mais conhecido pela sigla em inglês, ASR (Automatic Speech Recognition) é o ramo da computação que estuda o processo de transcrição da voz humana para texto, também chamado de speech to text (STT). Diversas áreas comerciais se utilizam desse tipo de tecnologia, como a transcrição e os serviços de automação telefônica, muito conhecidos como Unidades de Resposta Audível (URA).

Um exemplo muito presente no dia a dia de várias pessoas são os assistentes pessoais, sejam eles o Google Assistant ou a Siri (LÓPEZ; QUESADA; GUERRERO, 2017), que têm como principal entrada de informação a fala. Esse é o principal foco do seu uso, manter uma interação com usuário somente usando a fala. E esse é, também, um dos principais exemplos do uso dessa tecnologia, além de que essas assistentes estão ganhando cada vez mais notoriedade nas vidas dos usuários, pois elas tem saído do celular e entrado nas salas de cada um de nós através de dispositivos como Alexa e Google Home, permitindo que diversos lares, nas mais variadas línguas, interajam com essa tecnologia (BESACIER et al., 2014).

As primeiras pesquisas relacionadas ao sistema de reconhecimento automático de fala uti-

lizaram, principalmente, a teoria fonética acústica (PIKE, 1972). Ela representa os elementos fonéticos da fala, isto é, os sons fundamentais da língua, procurando compreender como são realizados acusticamente. Tais modelos naturais de ressonância são chamados de "formants" ou "formant frequencies" e demonstram-se pela concentração de energia no espectro de potência da fala.

Davis, Biddulph e Balashek (1952) construíram uma aplicação de conhecimento isolado de reconhecimento de dígitos por uma única pessoa. Na mesma época, uma pesquisa realizada nos laboratórios da RCA desenvolveu um sistema preparado para reconhecer dez sílabas, pronunciadas por uma única pessoa. Nos primeiros estudos, os modelos eram altamente especialistas e todas as pesquisas apontavam para a criação de modelos capazes de reconhecer a voz de uma única pessoa. Um ano mais tarde, essa restrição já não se verificava mais (DAVIS; BIDDULPH; BALASHEK, 1952).

Em meados de 1970, Tom Martin fundou a empresa Threshold Technology, que desenvolveu o primeiro produto de speech to text, o VIP-100 System. Na prática, o sistema não tinha muitas aplicações, mas as principais eram para as fabricantes de televisão e para o Fedex, para classificação de pacotes na transportadora (JUANG; RABINER, 2005). Mas o principal papel dessa empresa não está ligado diretamente aos seus produtos, mas à forma com que ela influenciou a ARPA - Advanced Research Projects Agency, do departamento de defesa. Em conjunto com a universidade Carnegie Mellon, eles foram capazes de reconhecer a fala utilizando um vocabulário de 1,011 palavras. Outros sistemas importantes desenvolvidos nesses anos pela DARPA's foram o CMU's Hearsay(-II) e o BBN's HWIM (LOWERRE, 1990).

Já nas décadas de 1980 e 1990, as pesquisas nessa área começaram a seguir por um caminho diferente do que estava sendo realizado até o momento. Os modelos de Cadeias de Markov Escondidas (Hidden Markov Models - HMM) começaram a ser estudados por alguns laboratórios, como o IBM e o IDA (Institute for Defense Analyses), apontando um caminho promissor. Alguns anos depois, HMM se tornou o principal método para realização do reconhecimento de fala.

Com a dedicação da IBM, capitaneada por Fred Jelinek, tinham como meta desenvolver uma máquina de escrever ativada por voz. Sua principal função seria converter a fala em uma sequência de letras e palavras para serem exibidas em um display ou escritas em um papel. O foco desse modelo, assim como os anteriores, era dependente do locutor, ou seja, um treinamento específico precisava ser realizado para cada usuário. Mas, diferente dos anteriores, já existia um vocabulário maior que conseguia ser identificado pelo sistema, batizado por

Tangora (LEE; SOONG; PALIWAL, 2012).

Rapidamente, HMM e Máquinas de Vetor de Suporte (Support Vector Machines - SVM) foram se consolidando como as principais técnicas. Elas levaram ao desenvolvimento de sistemas mais genéricos e com um maior vocabulário de reconhecimento. Alguns desses sistemas ainda são utilizados atualmente, como, por exemplo, o Sphinx, um sistema da CMU, que apresentou resultados notáveis para a fala de amplo vocabulário (JUANG; RABINER, 2005).

Atualmente, mais de 6.900 linguagens no mundo possuem este tipo de tecnologia, em produtos como Alexa e Google Home, cada vez mais presentes no dia a dia das pessoas. Utilizar técnicas de ASR sem nenhum treinamento prévio dos modelos para locutor ou vocabulário tem sido cada vez mais acessível aos desenvolvedores, com os serviços de speech-to-text disponibilizados pelas grandes empresas de tecnologia em formato de API, em seus sistemas cognitivos, como o Watson, da IBM, o Azure, da Microsoft, e o Cloud Platform, da Google. O Google Claud Platform foi o sistema utilizado para realizar a transcrição dos podcasts neste trabalho (JUANG; RABINER, 2005).

3.3.1.1 Google Cloud Speech-to-Text API

Cloud Speech-to-Text é a ferramenta do Google para realização da conversão de áudio em texto. Esta aplicação fica localizada dentro das ferramentas de aprendizagem de máquina no sistema em nuvem da empresa, chamado de "Google Cloud", e é distribuída no formato SaaS (Software as a Service). Assim, a forma de cobrança pelo serviço é no formato "pay as you go", que permite uma flexibilidade em seu uso e um baixo custo para testes de utilização.

Este serviço oferece a possibilidade de transcrição do áudio em mais de 120 idiomas e variantes, e é extremamente consolidada no mercado. Segundo a Gartner (GOOGLE, 2021), foi considerado o líder no relatório Quadrante Mágico da Gartner em serviços de Inteligência Artificial em 2020. O tempo médio das transcrições é a metade do tempo do arquivo original, e, segundo Clifton et al. (2020), foi obtida uma taxa de erro de palavra de 18,1%, considerando a variedade do universo dos podcast, com os mais diversos vocabulários, sotaques e qualidades do áudio.

A ferramenta oferece uma série de modelos de aprendizagem de máquina pré-treinados pelo Google para realizar a transcrição dos áudios, que devem ser selecionados de acordo com a origem, a fim de melhorar os resultados obtidos pela conversão do áudio em texto.

1. *Video*

Este modelo deve ser utilizado para transcrever o áudio de vídeo ou que inclua vários locutores, os melhores resultados podem ser obtidos fornecendo áudios gravados com uma taxa de amostragem superior a 16.000 Hz.

2. *phone_callx*

Devido à baixa taxa de amostragem das ligações telefônicas realizadas por smartphones, cerca de 8.000 Hz, este modelo é indicado para uso quando o objetivo da transcrição são chamadas de telefone.

3. *command_and_search*

Este modelo deve ser utilizado para transcrição de áudios mais curtos, como, por exemplo, comandos de voz ou pesquisas de voz muito utilizadas nos assistentes pessoais.

4. *default*

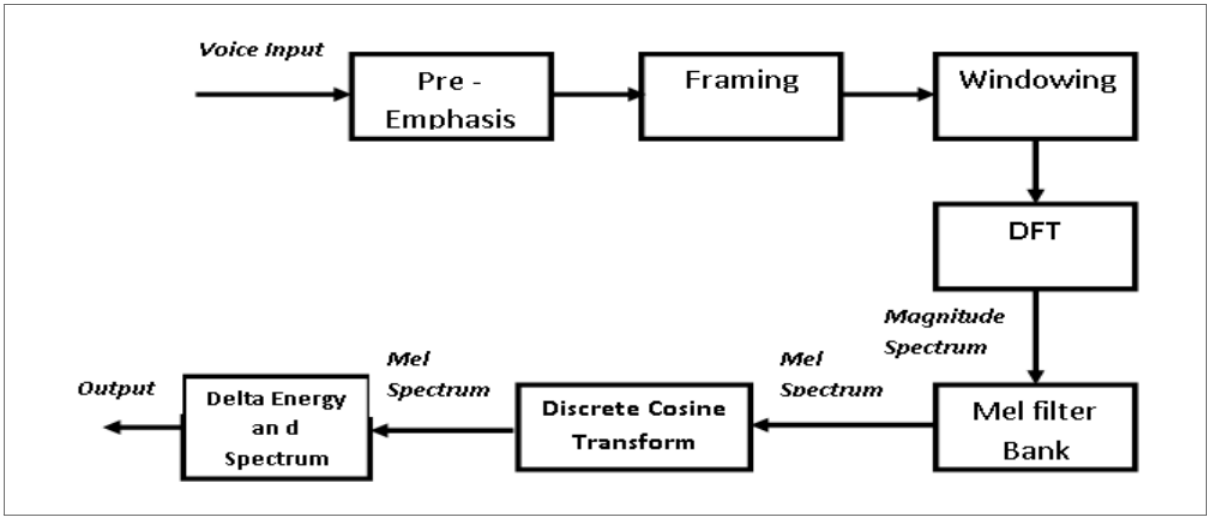
Este é o modelo padrão de áudio que deve ser utilizado quando nenhum dos modelos descritos anteriormente se encaixar, como, por exemplo, áudios mais longos e com uma boa taxa de amostragem.

3.3.2 Mel Frequency Cepstral Coefficients - MFCC

Um dos mais populares métodos para a recuperação de características utilizadas nas mais diversas áreas de processamento de áudio, seja em reconhecimento de voz ou recuperação de informação musical, é o MFCC. Ele é baseado na percepção dos ouvidos humanos, que não conseguem ouvir frequências acima de 1 KHz. Ou seja, MFCC se baseia nos limites de frequência conhecidamente detectáveis pelo ouvido humano (MUDA; BEGAM; ELAMVAZUTHI, 2010).

A extração dessas características possui os seguintes passos: pré-ênfase, divisão das janelas do áudio, transformada rápida de fourier, aplicação da escala de Mel no espectro e inversa da transformada discreta do cosseno para a geração dos valores contendo as características para as janelas de áudio geradas. Na imagem 1, temos uma visão geral do processo (ITTICHAICHAREON; SUKSRI; YINGTHAWORNSUK, 2012).

Figura 1 – Diagrama MFCC



Fonte: MUDA; BEGAM; ELAMVAZUTHI (2010)

1. *Pre-Emphasis*

É o primeiro passo do processo de extração das características de MFCC, em que o áudio passa por um filtro para enfatizar as frequências mais altas.

2. *Framing*

Processo de segmentação do áudio, no qual são utilizadas janelas entre 20 e 40 milissegundos para dividir a entrada em vários segmentos menores do tamanho da janela definida.

3. *Hamming Windowing*

No Framing, é muito comum que os as janelas formadas tenham divisões muito definidas, o objetivo desta etapa é sobrepor essas janelas, para diminuir os efeitos dessa divisão.

4. *FFT - Transformada Rápida de Fourier*

Essa etapa é aplicada em cada amostra de áudio gerada, para converter do domínio do tempo para o domínio da frequência.

5. *Mel Filter Bank Processing*

O resultado da FFT possui um alcance das frequências muito largos e o sinal da voz não segue uma escala linear. Por isso, o retorno da FFT é passado pelo filtro na escala Mel. Cada filtro

tem como saída a soma dos seus componentes espectrais filtrados. Depois disso, é aproximada a frequência Mel em Hz que é uma unidade de medida da frequência percebida de um tom (RIBEIRO et al., 2014).

6. *Discrete Cosine Transform*

É o processo para converter o log Mel para a frequência do tempo novamente. O resultado dessa transformação é chamado de MFCC (Mel Frequency Cepstrum Coefficient) e o conjunto desses coeficientes é denominado vetor acústico.

7. *Delta Energy and Delta Spectrum*

O sinal da voz e as janelas mudaram, devido às transições entre as amostras do áudio. Por isso, é necessário adicionar características relacionadas com a mudança de características cepstrais no tempo. A energia de uma janela para um sinal x em uma amostra de tempo t1 para a amostra de tempo t2.

3.4 APRENDIZAGEM DE MÁQUINA

Aprendizagem de Máquina é uma área da computação que tem por objetivo prover ao computador a capacidade de aprendizado similar à dos humanos, de forma que a solução para os problemas seja realizada sem a necessidade de uma configuração expressa de como o programa deve se comportar. Nesses casos, o método computacional baseia-se nas experiências passadas para melhorar a performance ou a acurácia da solução de determinado problema.

Nesse tipo de abordagem, os dados utilizados para treinamento do algoritmo são de extrema importância. Logo, o sucesso da aprendizagem de máquina está ligado diretamente às entradas utilizadas. Para isso, algumas etapas são comumente realizadas: primeiro, define-se o problema e coletam-se os dados, em seguida, eles são pré-processados com o objetivo de eliminar inconsistências e padronizá-los. No entanto, muitas vezes os dados ainda não estão prontos para serem inseridos no classificador, então eles passam por uma etapa de extração de características, na qual cada fonte de informação pode possuir um processamento distinto.

Com todas as informações extraídas, pode-se, então, selecionar qual modelo será utilizado pela aplicação. Mais adiante, neste capítulo, abordaremos alguns modelos utilizados na aprendizagem de máquina supervisionada. Após a escolha do modelo, passamos para a fase de seleção dos parâmetros, pois cada modelo possui um conjunto de parâmetros, e à medida

que eles são ajustados, o resultado pode melhorar ou piorar. Após todas essas fases, temos o processo de treinamento e validação do modelo. Neste momento, o nosso conjunto de dados é comumente dividido em três partes: treinamento, validação e testes. Com isso, são realizados testes nas bases para extrair métricas, como acurácia, precisão e recall, que nos ajudam a entender a performance do modelo. (MOHRI; ROSTAMIZADEH; TALWALKAR, 2018)

Comumente, dividimos a aprendizagem de máquina em supervisionada, não supervisionada e semi-supervisionada. A aprendizagem de máquina supervisionada é um dos métodos mais utilizados. Para seu uso, é necessário possuir os dados históricos rotulados, ou seja, cada informação que vai ser inserida no algoritmo deve possuir um "rótulo"identificando a qual classe o dado pertence. Para que seja possível desenvolver um modelo matemático genérico para lidar com os dados já conhecidos e com os novos exemplos de forma satisfatória (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007).

Já a aprendizagem não supervisionada é aplicada nos casos em que os dados não possuem uma classe definida ou que não seja possível rotulá-los antes do treinamento. O objetivo é agrupar os dados de forma a identificar padrões. Isso é possível utilizando algoritmos que procurem identificar as características que sejam capazes de agrupar todo o conjunto de dados em clusters de menores dimensões. Com base nos grupos criados, é possível que os especialistas identifiquem esses grupos ou tentem extrair quais padrões estão sendo representados em cada grupo (DAYAN; SAHANI; DEBACK, 1999) . Um algoritmo muito utilizado neste tipo de abordagem é o K-means (STEINBACH; KARYPIS; KUMAR, 2000).

3.4.1 Aprendizagem de Máquina Supervisionada

Dentro da aprendizagem de máquina, existem diversos algoritmos que utilizam deste princípio da supervisão. Existem os algoritmos com base lógica, estatística, em neurónios (redes neurais) e os com base em máquinas de vetores de suporte, muito conhecidas como Support Vector Machines (SVM). Todas essas abordagens seguem o mesmo princípio: elas utilizam uma base de dados rotulados antes de iniciar o treinamento. A seguir, alguns exemplos de algoritmos que usaremos neste trabalho.

3.4.1.1 Naive Bayes

Este algoritmo, apesar de simples, é muito efetivo em inúmeros casos práticos, como a classificação de texto, o que o torna bastante popular. Naive Bayes é um método estatístico baseado no Teorema de Bayes, que tem como premissa que os atributos de cada instância são independentes da classe respectiva. Por exemplo, se determinado animal é considerado um cachorro por possuir quatro patas, ter pelos e rabo, o algoritmo não irá considerar a correlação entre as diversas características fornecidas ao modelo. Assim, cada atributo será tratado de forma independente. (KIM et al., 2006)

Para realizar a classificação, o algoritmo Naive Bayes calcula as probabilidades a posteriori e a priori de cada um dos eventos acontecerem e, em seguida, normaliza com as probabilidades do evento da outra classe ocorrer, para, então, determinar a chance dessa nova entrada pertencer a uma classe ou a outra (RISH et al., 2001).

3.4.1.2 Árvore de Decisão

Um método com base lógica, que cria uma árvore (estrutura de dados) capaz de classificar as instâncias de acordo com valores do vetor, ou seja, com base nas suas características. Cada nó da árvore representa uma característica do vetor, e cada galho da árvore representa o valor que a característica pode assumir (RISH et al., 2001).

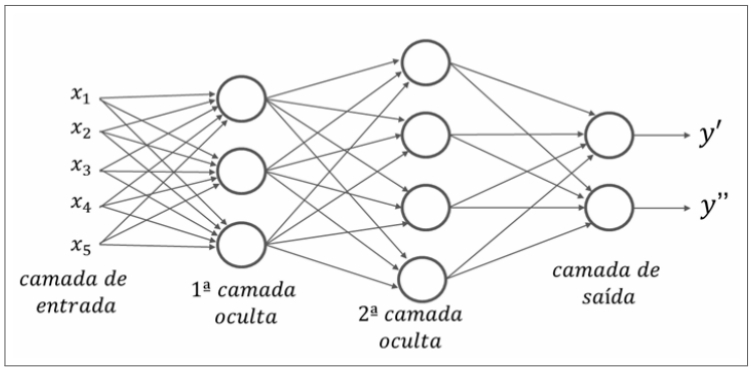
A seleção dos atributos a serem escolhidos é realizada pela entropia, que é o nível de desordem do sistema. Assim, para uma árvore de decisão ter uma boa classificação, é desejável que seu nível de entropia seja baixo, ou seja, cada característica representada por um nó da árvore consiga separar bem as classes.(RISH et al., 2001)

3.4.1.3 Multilayer Perceptron - MLP

Método inspirado no cérebro humano, o MLP é um exemplo de redes neurais. Ele consiste em uma grande quantidade de neurônios 2 que são ligados entre si, formando múltiplas camadas. Normalmente, os neurônios são separados em três camadas distintas: a camada de entrada, que recebe as informações a serem processadas; a camada de saída, que é onde os resultados obtidos do processamento são exibidos; e as camadas que estão no meio desse processo, chamadas de camadas escondidas, nas quais cada uma dos neurônios presentes possui

um peso associado. O processo de aprendizagem da rede neural está em encontrar os valores ideias entre as camadas de entrada e as de saída (NORIEGA, 2005).

Figura 2 – Multilayer Perceptron - MLP



Fonte: Elaborada pelo autor (2021)

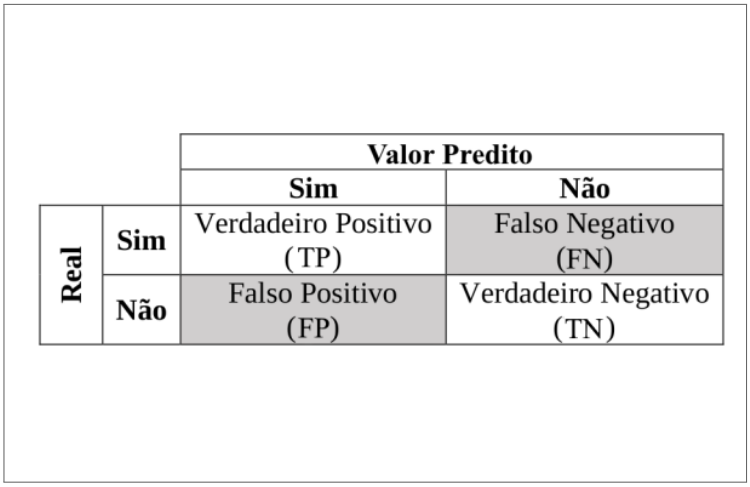
O processo de treinamento tem como objetivo calcular os pesos que minimizem o erro da rede. Para isso, são inseridos aleatoriamente valores entre -1 e 1 para todos os pesos da rede, e apresentados exemplos de entrada, calculando-se a saída gerada. Com isso, é possível medir o erro de cada nó e, com base nesses valores, ajustar levemente os pesos em cada uma das camadas. Em seguida, é apresentado um novo padrão de entrada, e esse passo é repetido até que se chegue ao valor mínimo (NORIEGA, 2005). Existem vários algoritmos que podem ser utilizados para estimar os valores dos pesos, mas o mais conhecido é o algoritmo Backpropagation, no qual o erro é propagado da camada de saída para a camada de entrada (RAMCHOUN et al., 2016).

3.4.1.4 Métricas de Avaliação

Ao realizar o treinamento e a classificação de uma tarefa usando aprendizagem de máquina, é de extrema importância que possamos medir qual é o resultado do modelo treinado, para que seja possível avaliar o desempenho do classificador. Para isso, existe um conjunto de métricas que podem ser calculadas com base no conjunto de teste e validação, comparando os resultados preditos com os valores rotulados anteriormente. Para ajudar no entendimento, podemos usar uma matriz de confusão, que permite visualizar os acertos e erros da classificação, gerando uma comparação entre o valor real e o valor predito, como mostrado na Figura 3 (JUNKER; HOCH; DENGEL, 1999).

1. *Acurácia*

Figura 3 – Matriz de Confusão



Fonte: Elaborada pelo autor (2021)

É a taxa que mede a performance do classificador, ou seja, quantos dos valores foram preditos corretamente, independente da classe. A taxa é dada pela fórmula (JUNKER; HOCH; DENGEL, 1999):

$$Acuracia = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precisão*

É a taxa que mede a quantidade de acertos em determinada classe. Definida pela razão entre os verdadeiros positivos e a soma de todos os classificados como positivos, sejam eles verdadeiros ou falsos (JUNKER; HOCH; DENGEL, 1999) Assim, ela mede a probabilidade de um exemplo ser, de fato, positivo, por exemplo dos que eu classifiquei como positivo, quantos realmente eram?

$$Precisao = \frac{TP}{TP + FP}$$

3. *Recall*

É entendido como a frequência que eu encontro exemplos de determinada classe. Ela é dada pela razão entre a quantidade de verdadeiros positivos pelo total de todos os exemplos da classe, ou seja, verdadeiros positivos + falsos negativos (JUNKER; HOCH; DENGEL, 1999). Desta forma conseguimos responder, quando realmente um exemplo é da classe positiva, o quão frequente eu classifico como positivo.

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-Score*

É a média harmônica entre recall e precisão, que tem por objetivo trazer um unico valor que represente a qualidade do seu modelo. É definido pela fórmula:

$$F1Score = 2 * \frac{Precisao * Recall}{Precisao + Recall}$$

4 METODOLOGIA

Neste capítulo, apresentaremos como este trabalho foi desenvolvido. Primeiro, mostraremos o processo de criação da base de dados com os episódios dos podcasts para análise. Em seguida, apresentaremos como foi realizada a conversão e transcrição dos episódios. Depois, a segunda etapa do processo será detalhada, com a extração de características de texto e de áudio, e a criação do modelo de predição de popularidade.

Segundo Gil et al. (2002), esta pesquisa é definida como exploratória, pois tem como objetivo se familiarizar sobre a forma com que a computação pode se relacionar com os podcasts, trazendo possibilidades do uso de processamento de texto e de áudio para o estudo dessa mídia.

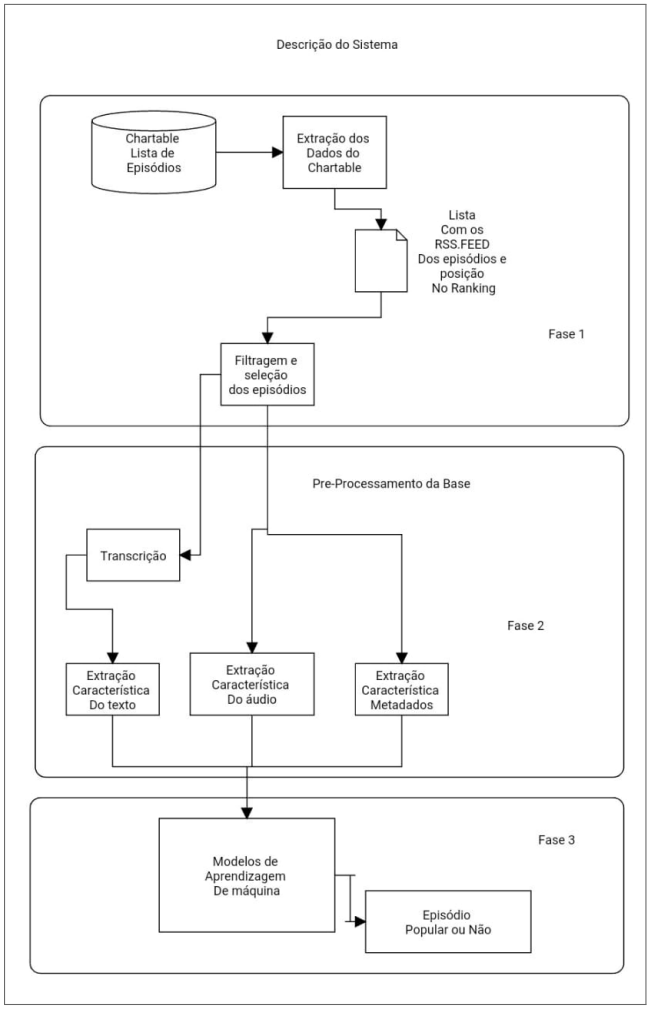
Para alcançar o objetivo da pesquisa, o experimento foi dividido em três fases, como mostrado na Figura 4.

A primeira fase consistiu na criação de uma base de dados com os podcasts classificados por popularidade. Para isso, foram coletados os podcasts da plataforma Itunes disponíveis no Chartable, o que possibilitou recuperar todos os RSS Feed dos canais e selecioná-los de acordo com o idioma, a disponibilidade e a data do último episódio publicado.

Após a seleção, na segunda fase, extraímos do áudio o MFCC (TIWARI, 2010) e a relação sinal ruído (CHEN; WANG; WANG, 2014). Ainda utilizando o áudio, realizamos um processo de transcrição do episódio, semelhante ao realizado por Clifton et al. (2020). Após a transformação em texto, aplicamos as técnicas Term Frequency Inverse Document Frequency (TF-IDF) (RAMOS et al., 2003) e Bag of Words (BoW) (ZHANG; JIN; ZHOU, 2010) para a extração de características do texto. A partir do RSS Feed dos episódios, selecionamos também os metadados para utilização na última fase.

Assim, na terceira fase, realizamos a execução dos modelos de aprendizagem de máquina para predição de popularidade. Comparamos, então, os resultados de acuracidade, precisão, recall e F1-Score de cada uma dos três grupos de características selecionadas na segunda fase, separadamente, e de todas elas simultaneamente.

Figura 4 – Visão geral do experimento



Fonte: Elaborada pelo autor (2021)

4.1 BASE DE DADOS

Para o desenvolvimento deste trabalho, utilizamos a aprendizagem de máquina supervisionada, que necessita de uma amostra de dados rotulados. No nosso caso, o objetivo era a criação de um base de dados dos canais ranqueados pelo nível de popularidade. Devido à característica de distribuição descentralizada dos podcasts, é difícil precisar quais são as audiências totais dos episódios, pois eles podem ser acessados pelo site ou aplicativo dos próprios canais (como o SoundCloud, site de hospedagem de mídias) ou por agregadores de podcasts (como o Google Podcast, o Podcast Addict, o aplicativo Podcast da Apple Store). Cada uma dessas ferramentas possui uma política de contabilização de visualizações que, na maioria das vezes, não disponibiliza os resultados para o público.

No nosso trabalho, para rotular os episódios com base em sua popularidade, utilizamos a

plataforma Chartable, que disponibiliza uma lista dos canais mais populares nas plataformas Apple Podcast, Spotify e Google Podcast. Segundo o próprio Chartable, cada uma dessas plataformas possui um método específico para recuperar esses dados, que não estão ligados diretamente com a quantidade de downloads realizados, por exemplo.

Para essa pesquisa, escolhemos o ranking do Apple Podcast, por que, segundo a podpesquisa 2018 (ABPOD, 2018), a última disponível no momento da extração da listagem, este era o aplicativo mais utilizado entre as plataformas que tínhamos acesso às informações de popularidade, com 20,3% dos usuários. O Google Podcast e o Spotify - segundo e terceiro colocados, respectivamente - tinham 14,5% e 11,0%. Além disso, as categorias utilizadas para segmentação dos canais pelo Apple Podcast eram as mais referenciadas em outros trabalhos (TSAGKIAS; LARSON; RIJKE, 2010; TSAGKIAS et al., 2008; SHARPE, 2020).

No processo de construção da nossa base de podcasts, foram realizadas consultas à listagem de podcasts mais populares no Apple Podcast, no dia 20 de novembro de 2019, considerando episódios das dezenove categorias existentes.

- Arts
- Business
- Comedy
- Education
- Fiction
- Government
- Health & Fitness
- History
- Kids & Family
- Leisure
- Music
- News
- Religion & Spirituality

- Science
- Society & Culture
- Sports
- TV & Film
- Technology
- True Crime

Cada uma dessas categorias continha até 250 canais, com o seu respectivo nome e URL de uma página do Chartable, onde havia informações mais detalhadas de cada canal. Obtivemos uma listagem com 3.932 URLs dos RSS Feeds de cada canal. No entanto, fomos obrigados a excluir 131 canais, por indisponibilidade dos servidores no momento de recuperar as informações dos RSS Feeds. Ao final dessa etapa, obtivemos 3.801 canais, com as seguintes informações de cada um deles:

- Nome;
- XML do RSS Feed;
- Categoria (uma das listadas acima)
- Classificação (entre 1 e 250)

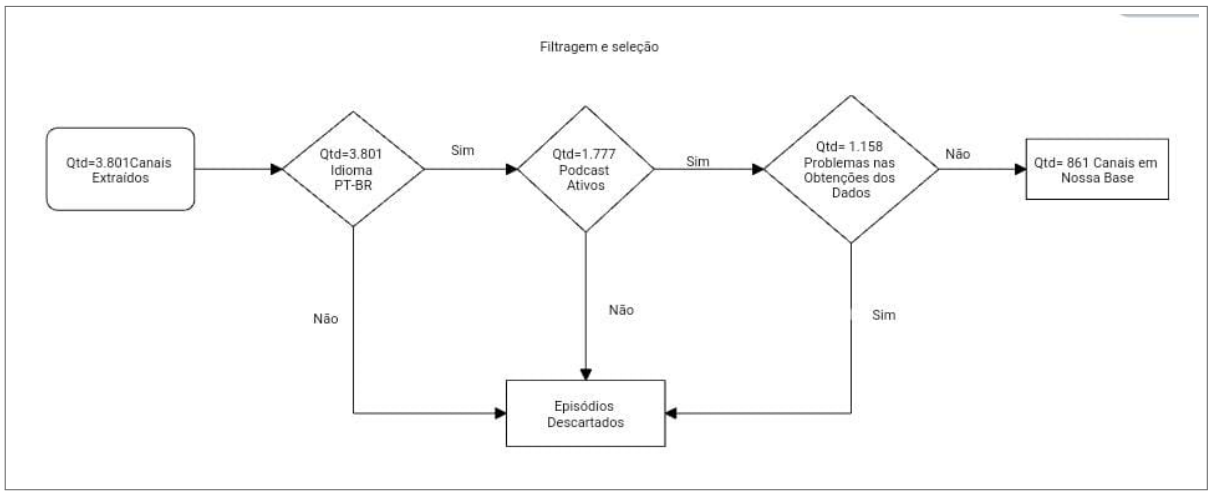
4.1.1 Filtragem e Seleção dos Episódios

Nesta seção, realizamos a seleção dos episódios que foram utilizados durante todo o experimento. Para isso, utilizamos o RSS Feed como fonte para os recortes. Utilizando a biblioteca xmltodict, desenvolvemos um script em Python para analisar os arquivos XML dos canais e filtrá-los de acordo com a Figura 5.

- Podcasts no Idioma Português

Como recorte de pesquisa, decidimos apenas selecionar os podcasts na língua portuguesa falada no Brasil. Isso porque, segundo a Podpesquisa 2018 (ABPOD, 2018), 91,5% dos brasileiros escutam podcasts de origem nacional. Selecionamos os episódios cuja tag <language>,

Figura 5 – Etapas da filtragem e seleção dos episódios



Fonte: Elaborada pelo autor (2021)

dentro da tag raiz <channel>, possuia o valor "pt-BR". A Figura 6 exemplifica isso, mostrando as informações indexadas do canal. Após isso, restaram 1.777 canais que possuíam a tag com o idioma desejado.

Figura 6 – Exemplo da tag <language> no XML de um canal podcast

```
<channel>
        <title>AGAMENON | 45 Minutos</title>
        <atom:link href="https://controle.podcast45minutos.com.br/agamenon/feed/" rel="self" type="application/rss+xml" />
        <link>https://podcast45minutos.com.br/agamenon/</link>
        <description>Para falar sobre futebol nós já temos o 45 Minutos. Aqui falar do maior esporte do mundo é (quase) proibido. No Agamenon é onde a conversa rola solta e a censura
não existe. Por isso, nosso lema é: Ouça se quiser!</description>
        <lastBuildDate>Fri, 23 Oct 2020 05:58:49 +0000</lastBuildDate>
        <language>pt-BR</language>
        <sy:updatePeriod>
        hourly  </sy:updatePeriod>
        <sy:updateFrequency>
        1       </sy:updateFrequency>
        <generator>https://wordpress.org/?v=5.5.1</generator>
```

Fonte: Elaborada pelo autor (2021)

- Podcasts Ativos

Com intuito de assegurar que os canais estivessem ativos no momento da análise, foram selecionados apenas os canais cuja última publicação era mais recente do que duas semanas, ou seja, o último episódio continha data de publicação entre 06/11/2019 e 20/11/2019, esta última sendo a data de coleta da lista (YANG et al., 2019).

A data de publicação do episódio pôde ser extraída através da tag <pubDate> 7 do primeiro episódio listado na tag <item>, dentro da tag raiz <channel>. Após a realização dessa etapa, restaram 1.158 podcasts ativos do idioma português.

- Erro na Obtenção dos Dados

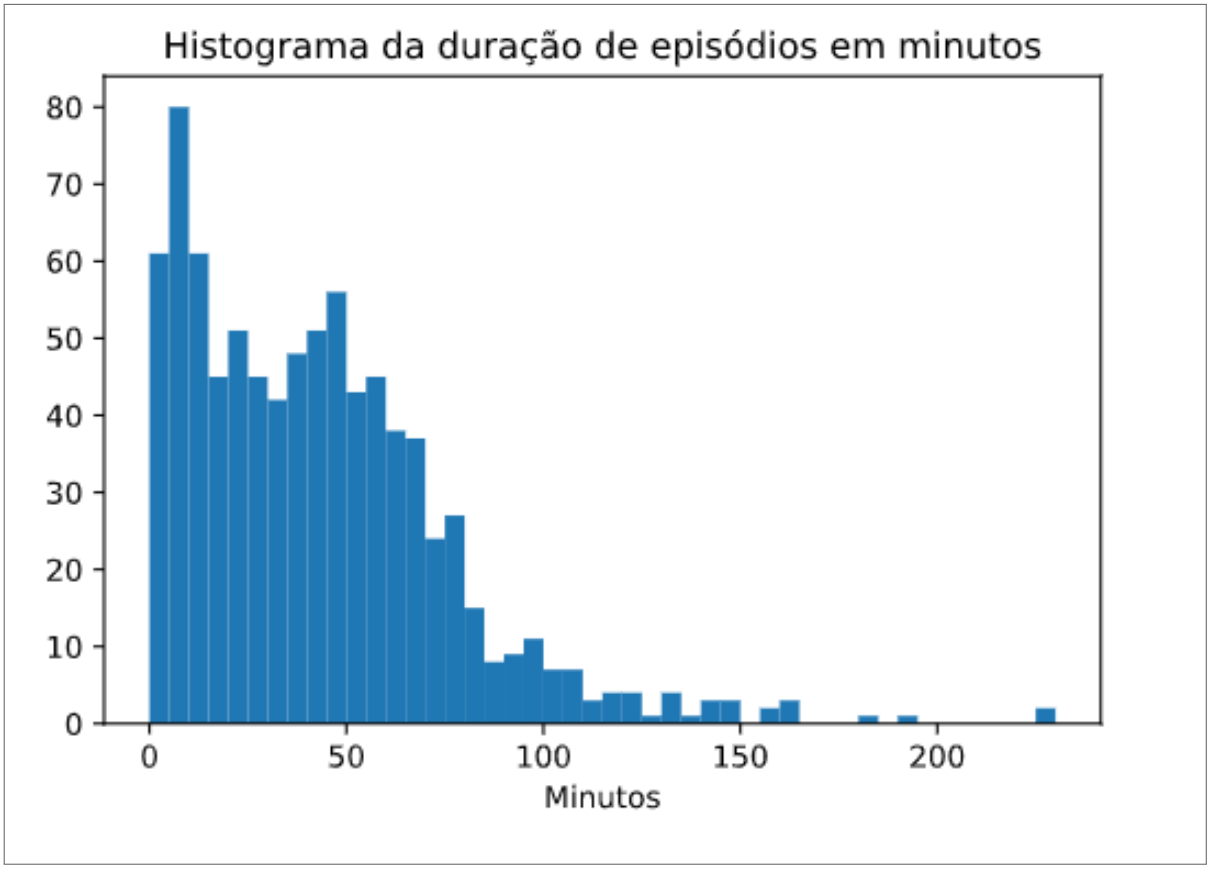
Figura 7 – Exemplo da tag <pubdate> no XML de um canal podcast

Fonte: Elaborada pelo autor (2021)

Nessa etapa retiramos os canais que por algum motivo apresentaram problema para recuperação das informações. Alguns canais não possuíam uma das tags utilizadas anteriormente ou apresentavam falha no download do episódio em MP3. Além de alguns arquivos ao serem baixados mostrarem problemas relacionados aos codecs na qual a biblioteca usada para leitura não foi capaz de finalizar a tarefa, por causa disto estes canais foram excluídos da base por não oferecerem as informações suficientes para classificarmos.

Finalizado o processo de seleção dos episódios, obtivemos um total de 861 arquivos de áudio no formato MP3 (um episódio por canal), totalizando 623 horas, 34 minutos e 9 segundos de áudios, com uma média de 42m53s por episódio e um desvio padrão de 33m12s. O maior episódio tinha 3h47m30s de duração o menor tinha 60 segundos. A Figura 8 mostra a distribuição da duração dos episódios.

Figura 8 – Histograma da duração dos episódios em minutos



Fonte: Elaborada pelo autor (2021)

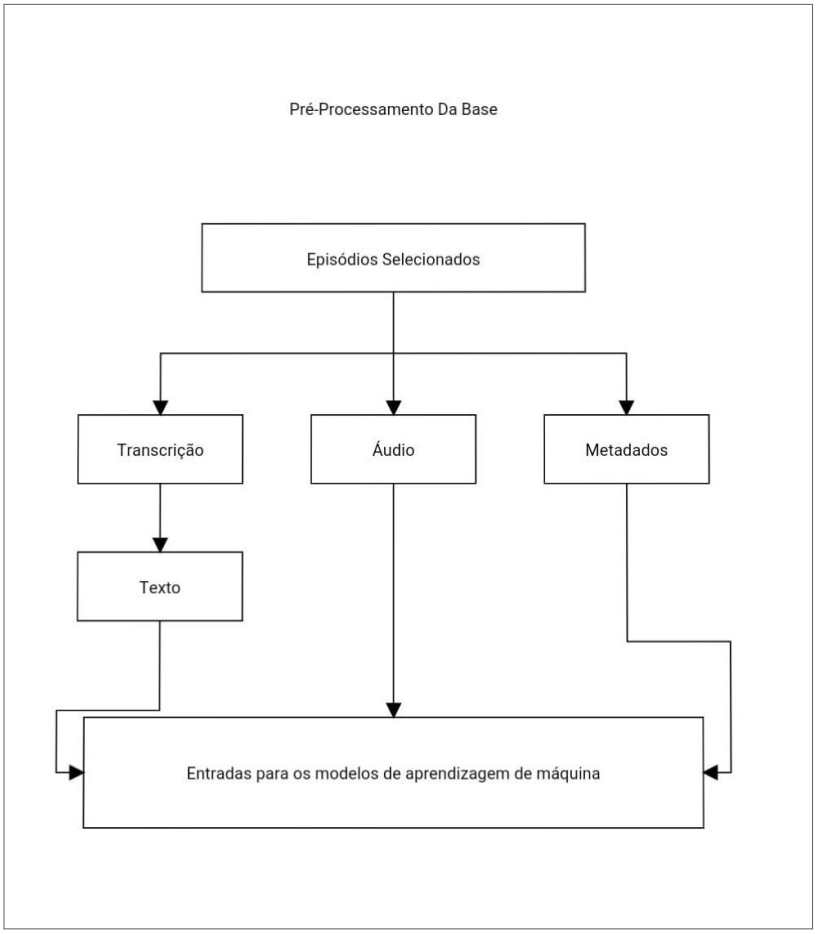
4.2 PRÉ-PROCESSAMENTO

No aprendizado de máquina supervisionado, possuir uma base de dados de qualidade é essencial para obter um classificador eficiente. Por isso, além de garantir uma quantidade suficiente de informações, é necessário pré-processar os dados, pois cada tipo de dado pode conter uma informação e ser processado de uma forma distinta, antes de ser analisado pelo algoritmo de aprendizagem de máquina. Nesta seção, iremos descrever quais foram as etapas e os parâmetros utilizados para realizar a transcrição, o pré-processamento e o processamento dos dados, extraindo as características de cada um dos tipos de dados. A Figura 9 mostra uma visão geral do processo.

4.2.1 Transcrição

O processo de transcrição dos 861 arquivos, com mais de 623 horas, foi um desafio a parte. Diante da necessidade de executar a transcrição com qualidade e sem um custo demasiado,

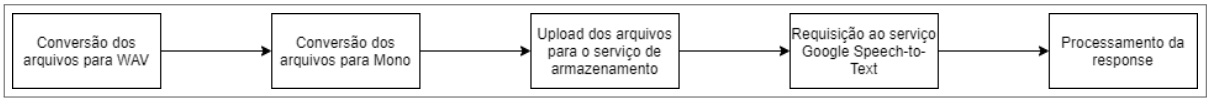
Figura 9 – Visão geral do pré-processamento



Fonte: Elaborada pelo autor (2021)

escolhemos o serviço Google Speech to Text.

Figura 10 – Etapas do processo de transcrição



Fonte: Elaborada pelo autor (2021)

Para a realização da transcrição, foram necessários cinco passos, como mostra a Figura 10, a seguir.

4.2.1.1 Conversão dos arquivos para WAV

Para o serviço de transcrição realizado pelo *Google Speech-to Text,* é recomendado que os áudios enviados estejam nos formatos *"WAV"* ou *"FLAC"*. Então, todos os arquivos MP3 foram convertidos para o formato *"WAV"*. Para esse processo, utilizamos a biblioteca *"py-*

dub" (JIAARO, 2020), na qual os arquivos foram carregados como um *"AudioSegment"* e, assim, puderam ser convertidos e exportados para o formato desejado, por meio do comando *"export(nome_do_arquivo.wav, format='wav')"*

4.2.1.2 Conversão de todos os arquivos para mono

Embora no Google Speech to Text seja possível processar a transcrição com mais de um locutor, o custo de processamento aumentava demasiadamente. Assim, neste trabalho, as transcrições foram realizadas utilizando a configuração padrão para somente um locutor. Para seguir as diretrizes do serviço, realizamos a conversão para que todo o áudio estivesse em um único canal. Para este processo, utilizamos o método *"set_channels(parâmetro)"*, com parâmetro igual a 1, da biblioteca *"pydub"* (JIAARO, 2020).

4.2.1.3 Upload dos arquivos convertidos para o serviço de armazenamento.

Para a transcrição dos episódios, foi necessário que o arquivo estivesse disponível dentro do sistema de armazenamento *Cloud Storage do Google*. Este processo foi realizado manualmente devido a sua simplicidade.

4.2.1.4 Requisição ao serviço de transcrição

O formato de requisição assíncrona é o recomendado pela documentação do *Google Speech-to-Text*, para realizar o processo de transcrição de áudios de longa duração, maiores do que 1 min. Para essa etapa, utilizamos a biblioteca *"google-cloud-speech"*, disponibilizada oficialmente pelo próprio Google. Nela, utilizamos o método *"long_running_recognize()"* da classe *"SpeechClient"*, no qual os parâmetros *"config"* e *"audio_url"* são objetos do tipo *"RecogntionConfig"*, com dados como a URL do áudio dentro do Cloud Storage e o encoding *(LINEAR16)* dos arquivos.

4.2.1.5 Criação do arquivo do episódio transcrito

Após o processamento da requisição pelos servidores do Google, a API nos retorna o resultado da requisição com algumas alternativas possíveis do texto transcrito e os valores de

início e fim de cada palavra no áudio. Após o retorno, realizamos o processamento do retorno da API, para obter um arquivo de texto para cada episódio, no qual cada linha indicava uma palavra transcita, bem como seus tempos de início e término no áudio. Esse processo facilitou o armazenamento e o posterior compartilhamento desses dados.

Ao final dessas etapas, restaram 861 episódios de canais diferentes, cada um contendo seu respectivo RSS Feed, o arquivo de áudio do último episódio disponível e um arquivo de texto com a transcrição gerada pelo *Google Speech-to-Text*. A distribuição final da quantidade de episódios pela categoria é listada na figura 11, na qual também é possível ver a relação de episódios populares e não populares. É possível notar que, para algumas classes, a relação difere bastante das demais. Por exemplo, as categorias "Fiction", "Governament", "History" e "True Crime"possuem uma baixa proporção de episódios não populares, em comparação com os populares. A proporção média dessa relação é de 1 episódio popular para 3.28 não populares.

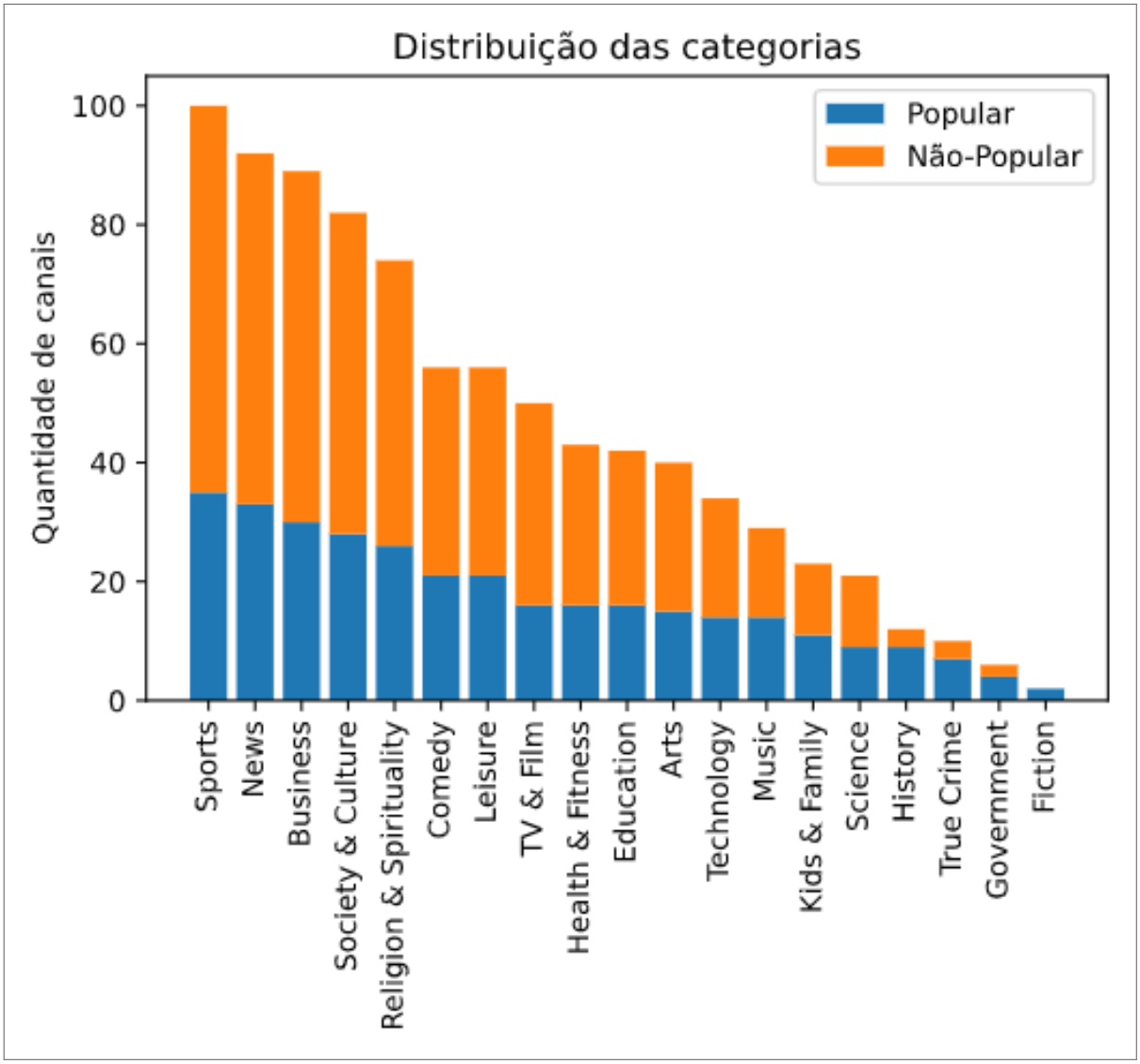
4.2.2 Texto

Para o processamento do texto, utilizamos a linguagem de programação Python e as seguintes bibliotecas: "Scikit-Learn" (ferramenta de aprendizado de máquina (PEDREGOSA et al., 2011a)), "Spacy" com o modelo de dados "pt_core_news_md" (ferramenta para processamento avançado de linguagem natural) (SRINIVASA-DESIKAN, 2018)) e "pandas" (ferramenta para manipulação e análise de dados) (MCKINNEY et al., 2011)). As seguintes etapas foram realizadas conforme figura 12.

Primeiro, os textos transcritos foram armazenados cada um em um arquivo de texto, contendo, em cada linha, a palavra e seu horário de início e fim no áudio. Para realizar a análise de todo conteúdo falado durante o podcast, foi necessário juntar todas essas palavras, para que fosse possível prosseguir com as etapas de pré-processamento.

Com o Data Frame (estrutura de dados utilizada pelo pandas para representar as tabelas) contendo as informações de todo o texto transcrito de cada episódios, utilizamos a biblioteca spacy para quebra-lo em tokens. Em seguida, para cada token, verificamos se ela é uma stop word através do atributo *"is_stop"* e removemos ela da sentença. E os sinais de pontuação através do atributo *"is_punct"* onde também foram removidos. Por fim, realizamos a lematização dessas palavras a partir do atributo *"lemma_"* da biblioteca do spacy que já nos fornece esse conjunto de ferramentas.

Figura 11 – Distribuição das categorias na base com separação por popularidade



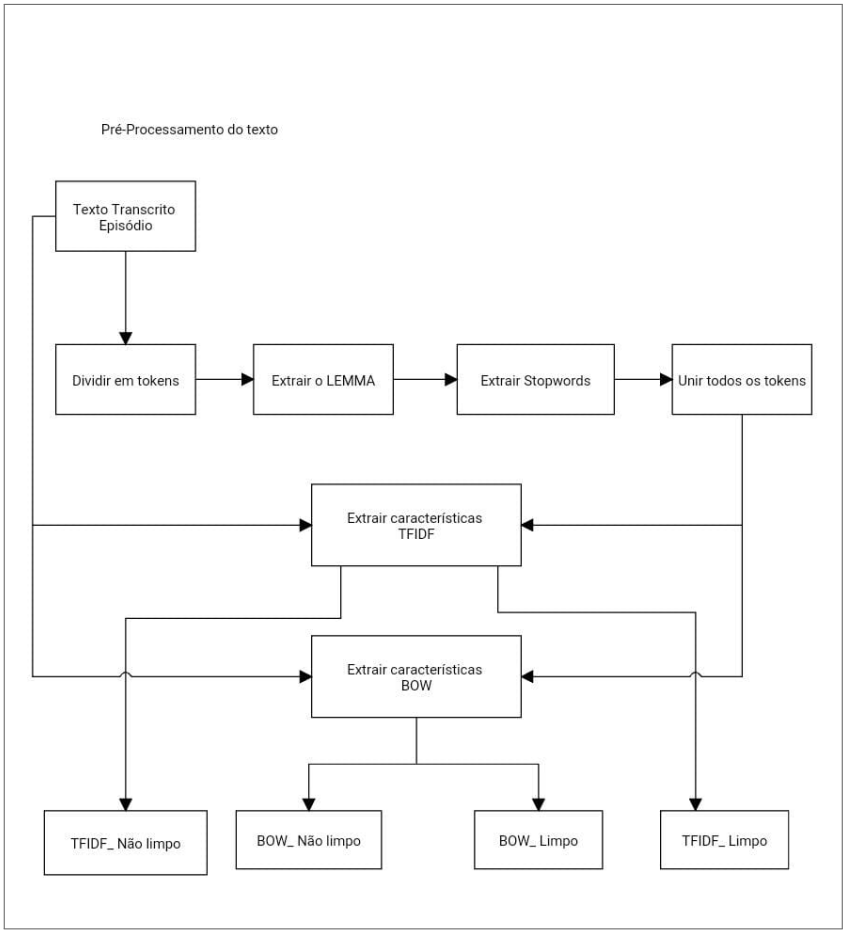
Fonte: Elaborada pelo autor (2021)

Após removidas todas as stop words e recuperados os radicais das palavras, extraímos as características que foram utilizadas no algoritmo de aprendizagem de máquina. Para isso, utilizamos os métodos TfidfVectorizer() - para gerar um vetor de característica utilizando a abordagem TF-IDF - e CountVectorizer() - que utiliza o conceito de Bag-of-Words.

4.2.3 Áudio

Assim como no processamento do texto existem bibliotecas específicas, para trabalhar com áudio elas também existem. Escolhemos a linguagem de programação Python, com a biblioteca librosa (concebida para análise de áudio e música), para realizar o processamento

Figura 12 – visão geral do pré-processamento do texto



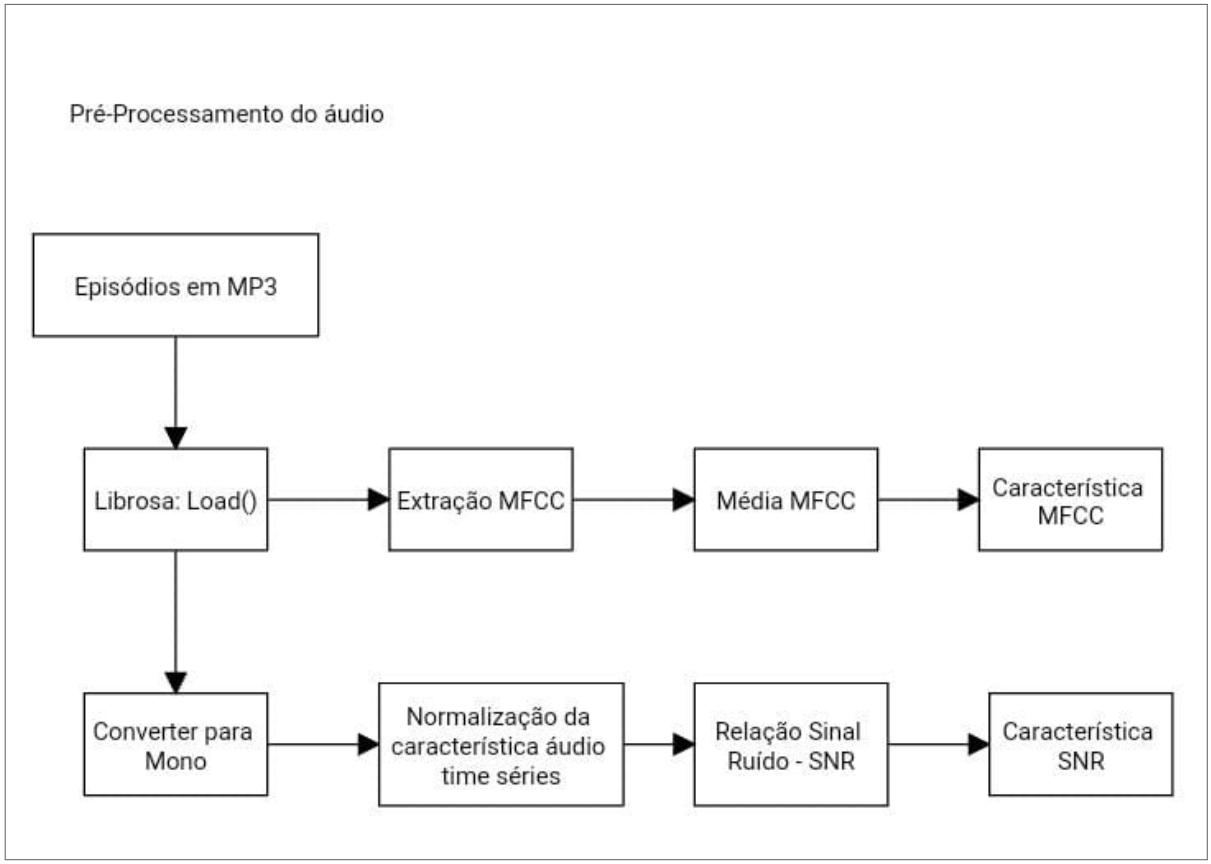
Fonte: Elaborada pelo autor (2021)

e extração de características dos episódios(MCFEE et al., 2015) conforme figura 13. Nessa biblioteca, existem vários métodos que auxiliam nas etapas de processamento do áudio. Neste trabalho, as características utilizadas foram MFCC e relação sinal ruído, conhecida como SNR.

Para a extração das características MFCC, utilizamos o áudio MP3 completo de todos os episódios selecionados na base de dados e, para cada arquivo processado, realizamos a leitura do MP3, recuperando os seguintes parâmetros: "audio time series", nomeado de y, e "sampling rate of y". Para isso, utilizamos o método "librosa.feature.mfcc()", passando os parâmetros "time_series" e "sampling_rate". Foram usados os argumentos padrão da biblioteca, com a quantidade de features n_mfcc=20, obtendo como retorno vinte características para cada trecho de áudio selecionado.

Com essas informações, realizamos o processo de normalização dos dados. Para isso, utilizamos o método "sklearn.preprocessing.scale(mfccs, axis=1)". Em seguida, para utilização dessas informações pelos algoritmos de aprendizagem de máquina, calculamos a média de

Figura 13 – visão geral do pré-processamento do áudio



Fonte: Elaborada pelo autor (2021)

cada uma das 20 características MFCC. Isso ocorreu para que fosse possível representar todos os arquivos com o mesma quantidade de características, uma vez que esse processo não seria possível, devido a diferença na duração de cada episódio. Com isso, cada episódio foi representado por um conjunto de vinte características. Além do MFCC, utilizamos a relação sinal ruído de cada episódio, definida pela média dividida pelo desvio padrão das amostras de áudio extraídas pelo método "librosa_load()".

4.2.4 Metadados

Os metadados estão presentes no arquivo RSS Feed dos canais e utilizam o padrão XML. Para realizar o pré-processamento dessas informações, utilizamos a biblioteca xmltodict do Python, visando facilitar o processo de criação das características. Nesse processo, selecionamos as informações referentes aos episódios e separamos dois tipos de características: as que definem a presença ou não de uma informação e as que são calculadas - como, por exemplo, o tamanho total do título, em caracteres. A Tabela 1 exibe as características utilizadas e seus

Tabela 1 – Tabela de característica dos metadados

Característica	Descrição	Tipo
has_title	Titulo do episódio	presença
len_title	Quantidade de caracteres do titulo do episódio	calculado
has_description	Descrição dos episódios	presença
len_description	Quantidade de caracteres da descrição do episódio	calculado
duration	Duração em segundo dos episódios	calculado
len_keywords	Quantidade de palavras chaves do episódio	calculado
len_categories	Quantidade de categorias em que o podcast está associado	calculado
has_link	Link encaminhando, geralmente para página do episódio	presença
has_itunes:image	Foto de capa para o feed do Itunes	presença
has_itunes:episodeType	Tipo do episódio, que pode ser: "full- conteúdo completo do episódio -, "trailer- uma pequena parte ou conteúdo promocional do episódio -, ou "bonus- um conteúdo a parte do seu canal, como, por exemplo, o making-off.	presença
has_itunes:author	Nome dos criadores dos podcasts	presença
has_itunes:explicit	Possui conteudo explicito	presença
has_itunes:title	Titulo do episódio, geralmente igual a tag title	presença
has_itunes:sumary	Descrição do episódio, geralmente igual a tag description	presença
has_titleApp	Titulo do episódio especifico para alguns agregadores.	presença

Fonte: Elaborada pelo autor (2021)

respectivos tipos.

4.2.5 Execução dos Modelos de Aprendizagem de Máquina

Nesta seção, são apresentadas as condições em que os experimentos foram realizados, utilizando a base de dados mencionada nos subcapítulos anteriores. Inicialmente, rodamos os modelos de aprendizagem de máquina (Naive Bayes, MLP e Decision Tree) com cada grupo de características isoladamente: áudio(snr e mfcc), texto (TFIDF e BoW) e metadados. Para cada grupo e modelo, computamos as métricas de acurácia, precisão, recall e F1-Score de cada uma das execuções.

Para testar as hipóteses, o primeiro passo foi separar nossa base de dados em subconjuntos de treino e teste, utilizando o Método Stratified KFold com o número de folds igual a 4, ou seja, todo o conjunto de dados foi dividido em 4 partes iguais, respeitando a porcentagem de amostras de cada uma das classes. No nosso caso, populares e não populares, dado que nossos exemplos não possuíam dados balanceados (SECHIDIS; TSOUMAKAS; VLAHAVAS, 2011). Em seguida, com o objetivo de avaliar quais características traziam melhores resultados para a predição de popularidade, executamos os modelos apresentados a seguir para cada uma das partes separadas pelo Stratified KFold, utilizando as características extraídas do texto, do

áudio e dos metadados

Na execução dos modelos de aprendizagem de máquina, foram utilizadas as implementações da biblioteca SkLearn do Python (PEDREGOSA et al., 2011b) com os seguintes métodos e parâmetros:

1. **Naive Bayes**
    - Método: GaussianNB
    - Não possui parâmetros
2. **Árvore de Decisão**
    - Método: DecisionTreeClassifier
    - Mínimo número de objetos: 1
    - Poda: Não
    - Critério: Entropy
3. **Rede Neural**
    - Método: MLPClassifier
    - Número de neurônios na camada escondida: 100
    - Função de ativação: Rectifier
    - Taxa de aprendizado: 0.001
    - Épocas: até 200
    - Momentum: 0.9

5 RESULTADOS E DISCUSSÕES

Nesta seção, apresentamos as métricas de avaliação de cada modelo, para as características que foram nomeadas nos experimentos e tabelas da seguinte maneira: meta (metadado do episódio processado), bow_nao_limpo (Bag-of-words sem realizar o processamento do texto), bow_limpo (Bag-of-Words após realizar o processamento do texto), tfidf_nao_limpo (TF-IDF sem realizar o processamento do texto), tfidf_limpo (TF-IDF após realizar o processamento do texto), mfcc (20 características do MFCC), snr (característica da relação sinal ruído do áudio).

5.1 RESULTADOS

No primeiro momento, treinamos os modelos com cada uma das características de forma individual. Os resultados estão disponíveis na Tabela 2.

Tabela 2 – Resultados dos modelos com uma característica

Feature	GNB				Dtree				MLP			
	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score
meta	43,09%	38,31%	81,65%	52,15%	53,54%	39,23%	40,67%	39,94%	**59,93%**	**44,51%**	22,32%	29,74%
bow_nao_limpo	59,93%	40,00%	11,01%	17,27%	55,17%	41,25%	42,51%	41,87%	42,86%	38,81%	87,46%	53,76%
bow_limpo	58,30%	37,30%	14,37%	20,75%	54,82%	40,49%	40,37%	40,43%	42,74%	38,28%	82,87%	52,37%
tfidf_nao_limpo	59,70%	38,10%	9,79%	15,57%	53,31%	36,65%	31,50%	33,88%	45,53%	39,14%	78,29%	52,19%
tfidf_limpo	58,54%	37,70%	14,07%	20,49%	54,36%	39,75%	39,14%	39,45%	47,85%	40,07%	75,23%	52,28%
mfcc	56,91%	33,58%	13,76%	19,52%	52,85%	38,21%	39,14%	38,67%	52,73%	35,61%	30,28%	32,73%
snr	50,99%	37,97%	45,87%	41,55%	54,24%	39,50%	38,53%	39,01%	**61,90%**	33,33%	**0,31%**	0,61%
desvio_padrão	5,74%	1,81%	25,06%	13,13%	0,78%	1,40%	3,24%	2,32%	7,31%	3,26%	32,61%	18,31%

Fonte: Elaborada pelo autor (2021)

A característica que mais se destaca é o SNR computado sobre o áudio, com 61,90% de acurácia, utilizando o modelo MLP, único resultado em todos os experimentos a superar 60%. Por outro lado, esse modelo apresenta recall próximo de zero (0,31%), o que nos mostra uma capacidade baixa de identificar os objetos relevantes da classe "populares". Ao executar novamente o experimento, agora buscando o recall da classe "não populares", obtivemos um resultado de 99,6% para identificação dos objetos relevantes dessa classe. Já utilizando somente os metadados, o mesmo modelo chegou a 59,93% de acurácia, com precisão de 44,51% na identificação de podcasts da classe "populares".

Note-se, ainda, que as características textuais apresentaram resultados semelhantes nos modelos Naive Bayes e Árvore de Decisão, com a diferença de acurácia não ultrapassando 2%

em cada um dos modelos.

Em seguida, treinamos os modelos combinando as características duas a duas. Os resultados são apresentados na Tabela 3.

Tabela 3 – Resultados dos modelos com duas características

Features	GNB				dtree				MLP			
	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score
meta_bow_nao_limpo	59,93%	40,00%	11,01%	17,27%	56,33%	42,64%	43,43%	43,03%	42,97%	38,83%	87,16%	53,72%
meta_bow_limpo	58,30%	37,30%	14,37%	20,75%	55,17%	40,51%	38,53%	39,50%	41,93%	37,97%	83,49%	52,20%
meta_tdidf_nao_limpo	59,70%	38,10%	9,79%	15,57%	53,77%	37,72%	33,33%	35,39%	46,92%	39,78%	77,37%	52,54%
meta_tdidf_limpo	58,54%	37,70%	14,07%	20,49%	53,31%	38,32%	37,61%	37,96%	48,43%	40,46%	75,84%	52,77%
meta_mfcc	47,15%	39,64%	74,92%	51,85%	50,99%	36,31%	38,53%	37,39%	55,40%	39,93%	34,56%	37,05%
meta_snr	42,28%	38,13%	83,49%	52,35%	55,63%	41,21%	39,45%	40,31%	59,00%	43,30%	25,69%	32,25%
bow_nao_limpo_mfcc	59,93%	40,00%	11,01%	17,27%	52,73%	37,50%	36,70%	37,09%	49,48%	38,05%	52,60%	44,16%
bow_nao_limpo_snr	59,93%	40,00%	11,01%	17,27%	55,75%	41,12%	38,23%	39,62%	42,86%	38,81%	87,46%	53,76%
bow_limpo_mfcc	58,30%	37,30%	14,37%	20,75%	51,34%	35,63%	34,86%	35,24%	42,97%	38,45%	83,49%	52,65%
bow_limpo_snr	58,30%	37,30%	14,37%	20,75%	54,59%	40,00%	39,14%	39,57%	49,01%	37,56%	51,68%	43,50%
tfidf_nao_limpo_mfcc	59,70%	38,10%	9,79%	15,57%	55,28%	40,14%	36,09%	38,00%	54,36%	34,43%	22,32%	27,09%
tfidf_nao_limpo_snr	59,70%	38,10%	9,79%	15,57%	54,59%	39,26%	35,78%	37,44%	46,57%	39,29%	74,62%	51,48%
tfidf_limpo_mfcc	58,54%	37,70%	14,07%	20,49%	52,15%	37,08%	37,31%	37,20%	46,57%	39,59%	77,37%	52,38%
tfidf_limpo_snr	58,54%	37,70%	14,07%	20,49%	53,31%	38,87%	40,06%	39,46%	54,70%	40,60%	41,59%	41,09%
mfcc_snr	55,98%	42,40%	44,34%	43,35%	54,24%	40,06%	41,28%	40,66%	52,38%	35,93%	32,42%	34,08%
desvio_padrão	5,74%	1,81%	25,06%	13,13%	0,78%	1,40%	3,24%	2,32%	7,31%	3,26%	32,61%	18,31%

Fonte: Elaborada pelo autor (2021)

Ao combinarmos as características textuais como meta_bow_limpo e bow_limpo_mfcc no modelo DTree foran encontrados valores iguais para acurácia, precisão, recall e F1-Score com os respectivos valores (58,30% 37,30% 14,37% 20,75%). Esse fato se repetiu tanto com o algoritmo sendo executando antes e depois do processo de limpeza. O algoritmo TF-IDF apresentou o mesmo comportamento quando comparado meta_tfidf_nao_limpo com tfidf_nao_limpo_snr, os valores de acurácia, precisão, recall e F1-Score encontrados respectivamente foram (59,70% 38,10% 9,79% 15,57%)

Por fim, executamos os modelos com as combinações entre os três conjuntos de características (resultados mostrados na Tabela 4) e com todas as características (resultados mostrados na Tabela 5).

Observando as Tabelas 4 e 5, tomando como base os cenários anteriores das Tabelas 2 e 3, observamos que não houve melhora nos resultados, ao combinarmos todas as características simultaneamente.

Tabela 4 – Resultados dos modelos com três características

Feature	GNB				Dtree				MLP			
	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score
meta_mfcc_snr	44,02%	38,55%	79,82%	51,99%	55,05%	41,33%	43,73%	42,50%	55,05%	38,81%	31,80%	34,96%
meta_bow_nao_limpo_mfcc	59,93%	40,00%	11,01%	17,27%	54,82%	40,25%	39,14%	39,69%	42,97%	38,92%	88,07%	53,98%
meta_bow_nao_limpo_snr	59,93%	40,00%	11,01%	17,27%	56,10%	42,25%	42,51%	42,38%	42,62%	39,03%	90,83%	54,60%
meta_bow_limpo_mfcc	58,30%	37,30%	14,37%	20,75%	51,34%	35,06%	33,03%	34,02%	42,74%	38,18%	81,96%	52,09%
meta_bow_limpo_snr	58,30%	37,30%	14,37%	20,75%	53,08%	37,29%	34,56%	35,87%	41,81%	38,05%	84,71%	52,51%
meta_tfidf_nao_limpo_mfcc	59,70%	38,10%	9,79%	15,57%	54,12%	39,24%	37,92%	38,57%	47,27%	39,77%	75,54%	52,11%
meta_tfidf_nao_limpo_snr	59,70%	38,10%	9,79%	15,57%	54,70%	39,10%	34,56%	36,69%	46,92%	39,38%	73,70%	51,33%
meta_tfidf_limpo_mfcc	58,54%	37,70%	14,07%	20,49%	53,89%	39,46%	40,06%	39,76%	47,62%	40,10%	76,76%	52,68%
meta_tfidf_limpo_snr	58,54%	37,70%	14,07%	20,49%	55,28%	41,52%	43,43%	42,45%	47,27%	39,90%	76,76%	52,51%
tfidf_limpo_mfcc_snr	58,54%	37,70%	14,07%	20,49%	54,24%	39,82%	40,06%	39,94%	47,62%	40,06%	76,45%	52,58%
tfidf_nao_limpo_mfcc_snr	59,70%	38,10%	9,79%	15,57%	54,59%	39,19%	35,47%	37,24%	45,30%	39,09%	78,90%	52,28%
bow_limpo_mfcc_snr	58,30%	37,30%	14,37%	20,75%	50,52%	33,98%	32,11%	33,02%	42,97%	38,45%	83,49%	52,65%
bow_nao_limpo_mfcc_snr	59,93%	40,00%	11,01%	17,27%	53,66%	38,68%	37,61%	38,14%	42,04%	38,28%	85,93%	52,97%
desvio_padrão	4,08%	1,00%	18,09%	9,16%	1,49%	2,30%	3,74%	2,96%	3,54%	0,69%	14,08%	4,79%

Fonte: Elaborada pelo autor (2021)

Tabela 5 – Resultados dos modelos com todas características

Feature	GNB				Dtree				MLP			
	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score
meta_tfidf_nao_limpo_mfcc_snr	59,70%	38,10%	9,79%	15,57%	54,82%	40,06%	38,23%	39,12%	46,81%	39,69%	77,06%	52,39%
meta_tfidf_limpo_mfcc_snr	58,54%	37,70%	14,07%	20,49%	53,54%	38,41%	37,00%	37,69%	45,99%	39,35%	77,98%	52,31%
meta_bow_limpo_mfcc_snr	58,30%	37,30%	14,37%	20,75%	52,85%	37,54%	36,39%	36,96%	41,35%	37,71%	83,49%	51,95%
meta_bow_nao_limpo_mfcc_snr	59,93%	40,00%	11,01%	17,27%	55,63%	41,21%	39,45%	40,31%	42,51%	38,65%	87,46%	53,61%
desvio_padrão	0,71%	1,03%	1,96%	2,19%	1,08%	1,43%	1,17%	1,29%	2,29%	0,76%	4,23%	0,62%

Fonte: Elaborada pelo autor (2021)

Tabela 6 – Desvio Padrão dos algoritmos que continham o bow_nao_limpo

Features	GNB				dtree				MLP			
	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score
bow_nao_limpo	59,93%	40,00%	11,01%	17,27%	55,17%	41,25%	42,51%	41,87%	42,86%	38,81%	87,46%	53,76%
meta_bow_nao_limpo	59,93%	40,00%	11,01%	17,27%	56,33%	42,64%	43,43%	43,03%	42,97%	38,83%	87,16%	53,72%
bow_nao_limpo_mfcc	59,93%	40,00%	11,01%	17,27%	52,73%	37,50%	36,70%	37,09%	49,48%	38,05%	52,60%	44,16%
bow_nao_limpo_snr	59,93%	40,00%	11,01%	17,27%	55,75%	41,12%	38,23%	39,62%	42,86%	38,81%	87,46%	53,76%
meta_bow_nao_limpo_mfcc	59,93%	40,00%	11,01%	17,27%	54,82%	40,25%	39,14%	39,69%	42,97%	38,92%	88,07%	53,98%
meta_bow_nao_limpo_snr	59,93%	40,00%	11,01%	17,27%	56,10%	42,25%	42,51%	42,38%	42,62%	39,03%	90,83%	54,60%
bow_nao_limpo_mfcc_snr	59,93%	40,00%	11,01%	17,27%	53,66%	38,68%	37,61%	38,14%	42,04%	38,28%	85,93%	52,97%
meta_bow_nao_limpo_mfcc_snr	59,93%	40,00%	11,01%	17,27%	55,63%	41,21%	39,45%	40,31%	42,51%	38,65%	87,46%	53,61%
desvio_padrão	0,00%	0,00%	0,00%	0,00%	1,17%	1,63%	2,37%	1,94%	2,26%	0,31%	11,70%	3,21%

Fonte: Elaborada pelo autor (2021)

5.2 DISCUSSÃO

Nesta seção discutiremos os resultados e o processo de pesquisa.

5.2.1 Treinamento com Características Individuais

Em relação ao treinamento dos modelos de aprendizagem de máquina utilizando uma única característica (Tabela 2), é possível destacar alguns resultados que chamaram atenção

no processo de predição de popularidade. As características do áudio (SNR) e metadados, no modelo MLP, demonstraram os maiores valores de acurácia para esta tarefa. Utilizando o SNR, obtivemos o maior valor de acurácia entre todos os cenários (61,90%), o que poderia nos mostrar um bom resultado com o uso de uma simples característica. Principalmente por que o SNR é uma característica que está ligada diretamente à qualidade de captação do áudio. Isso nos sugere que os podcasts mais populares seriam os que possuem melhores qualidades de gravação, com melhores equipamentos. No entanto, ao realizar uma análise de correlação, não identificamos relação entre essa característica e popularidade, indicando que podemos estar diante de um caso de co-ocorrência com o nível de produção geral do podcast, como preocupação com roteiro, convidados e etc.

Ainda, o recall próximo a zero (0,31%) sugere que, praticamente em todos os casos, o episódio foi classificado como não popular. Isso, na pratica, oferece um baixo poder de separação das classes, indicando que o valor elevado da acurácia pode ter ocorrido, na verdade, devido à desproporcionalidade entre os casos populares e os não populares.

Já com os metadados tivemos um valor de acurácia (59,93%) muito próximo ao encontrado no SNR, mas com um valor de precisão e recall muito maior, com valores iguais a (44,51% e 22,32%) respectivamente. Sugerindo ser uma boa característica para analisar os podcasts.

É importante salientar que, em ambos os cenários, existem evoluções possíveis. Por exemplo, quanto aos metadados, podemos utilizar os dados processados dos canais - como a regularidade dos episódios publicados, o total de episódios, o tempo ativo, entre outras - para melhorar os resultados. Já quanto às características extraídas do áudio, pode-se acrescentar métricas sobre o silêncio, a energia e a seriedade (vide (YANG et al., 2019)), além de características subjetivas (como as criadas para música por Andersen (2014)).

As características textuais foram as que apresentaram o maior valor de recall no modelo MLP com o algoritmo bow_nao_limpo com 87,46% e F1-Score de 53,76%. O que poderia representar um bom resultado, mas ao analisarmos em conjunto com a acurácia de 42,86% podemos perceber que esses valores se aproximam muito com o classificador que prediz todos os episódios como positivos.

Sugerindo que as características textuais não são interessantes no cenário de predição de popularidade, visto que, quando analisadas individualmente, as características de áudio e metadados obtiveram resultados similares entre os modelos, por exemplo, as características textuais obtiveram para o modelo GNB uma acurácia semelhante aos encontradas nos metadados e

características de áudio no modelo MLP.

Mesmo levando em consideração o melhor F1-Score dos treinamentos com o BoW limpo e não limpo, dos modelos de Árvore de Decisão e MLP, as diferenças de performance podem não compensar o custo de extração dessas características, pois o download, a conversão e a transcrição dos áudios pode durar mais que a metade do tempo de duração de cada episódio. Isso eleva o custo da formação da base de dados, para se obter um resultado muito semelhante às outras características.

Ressaltamos aqui que o processo de transcrição, mesmo não demonstrando um potencial para predição de popularidade, possibilita sumarizar o episódio (ZHENG et al., 2020a; SONG et al., 2020; VARTAKAVI; GARG, 2020; ZHENG et al., 2020b). Isso é, apontar quais são os principais temas abordados no programa e indicar o tempo que cada assunto é abordado, nos permitindo criar visualizações prévias, entre outras aplicações.

Ainda analisando as características textuais, o algorítimo BoW nos revelou um cenário interessante no contexto dos podcasts, no qual os resultados após o processo de limpeza, não obteve uma melhora em todas as métricas. Nos modelos (GNB, DTree e MLP), ao compararmos os resultados, identificamos que não houve melhora na maioria das métricas, como podemos ver na Tabela 7, na qual comparamos a diferença entre as estratégias. Já usando o algorítimo TF-IDF tivemos uma piora na acurácia no modelo GNB, saindo de 59,70% para 58,54%, e uma melhora esperada nos modelos DTree e MLP, como pode ser visto na Tabela 3.

Tabela 7 – Diferença das estratégias bow_nao_limpo vs bow_limpo

Feature	GNB				Dtree				MLP			
	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score	Acuracia	Precisão	Recall	F1-Score
bow_nao_limpo	59,93%	40,00%	11,01%	17,27%	55,17%	41,25%	42,51%	41,87%	42,86%	38,81%	87,46%	53,76%
bow_limpo	58,30%	37,30%	14,37%	20,75%	54,82%	40,49%	40,37%	40,43%	42,74%	38,28%	82,87%	52,37%
bow_nao_limpo - bow_limpo	1,63%	2,70%	-3,36%	-3,48%	0,35%	0,76%	2,14%	1,44%	0,12%	0,53%	4,59%	1,39%

Fonte: Elaborada pelo autor (2021)

5.2.2 Treinamento com Características Combinadas

Ao combinarmos as características textuais (BoW e TF-IDF) com uma das outras, os valores encontrados para acurácia, precisão e recall foram semelhantes ou iguais. No caso do modelo GNB, os resultados apresentaram-se exatamente iguais, o que ocorreu novamente ao executarmos combinando mais características.

Isso ocorreu, devido a desproporcionalidade entre a quantidade de características textuais disponíveis e as características de áudio e metadados, no qual os vetores gerados pelos algoritmos Bow e TF-IDF possuem aproximadamente 65 mil colunas na matriz, enquanto o MFCC possui apenas vinte colunas, o SNR, uma, e os metadados, quinze. Assim, no modelo GNB, todos os resultados se repetem quando as características textuais estão presentes. Mostrando-nos a necessidade de utilizar técnicas de seleção de características, ao comparar características de dimensionalidades muito distintas. Na Tabela 6, é possível visualizar o desvio padrão de todos os resultados que envolviam o algorítimo bow_nao_limpo e notar que não houve variação maior que 2,5% entre todos os valores de acurácia.

Considerando as características textuais, dois cenários se destacaram, quando combinamos as características duas a duas:

1. O uso combinado dos metadados com a característica bow_não_limpo resultou em uma melhora, se comparado aos resultados obtidos com as duas características individualmente. Na Tabela 4, vemos que as acurácias foram de 53,54% e 55,17% para 39,23% e 41,25% para os metadados e o bow_não_limpo, respectivamente. Ao combinarmos as duas características, tivemos um aumento, chegando a 56,33% de acurácia e 42,64% de precisão. Isso sugere que o uso combinado dessas técnicas pode aumentar o poder de predição dos modelos de Árvore de Decisão.

2. O uso combinado do tfidf_nao_limpo e mfcc no modelo MLP resultou em uma melhora na acurácia, que, individualmente, foi de 45,53% e 52,73%, respectivamente, e, com as duas características juntas, chegou a 54,36%. Porém, em compensação, houve uma piora nos valores de precisão e recall, o que indica que o aumento na capacidade de acertar a classificação dos episódios vem ao custo da perda de capacidade em distinguir quais são os episódios de determinada classe. Assim, o uso dessas características combinadas pode ser útil em modelos que já possuam boa precisão e recall, e nos quais deseje-se melhorar a acurácia.

Ao utilizarmos as duas características com melhor desempenho individual (metadados e SNR), tivemos uma redução na acurácia, considerando o modelo MLP (de 59,93% para metadados e 61,90% para SNR para 59,00%, para as duas características combinadas). Em contrapartida, houve uma melhora significativa nas métricas de recall e F1-Score, o que indica uma melhor capacidade do modelo em classificar como popular os episódios que são real-

mente populares. Isso pode sugerir o uso combinado dessas características para predição de popularidade.

Essa melhoria também foi observada no modelo de Árvore de Decisão. Nele, o SNR apresentou 54,24% de acurácia, e os metadados, 53,54%, enquanto o uso conjunto dessas características resultou em uma acurácia de 55,63%, com melhorias também significativas nas métricas de precisão e F1-Score. Isso pode ser um indício de que a combinação das características extraídas do áudio com as características provenientes dos metadados pode trazer resultados promissores. Além disso, essas características apresentam um menor custo de desenvolvimento, pois as características geradas a partir dos metadados são obtidas a partir de um recorte do RSS feed ou de um processamento simples de somatórios e médias sobre as informações do RSS Feed. As características do áudio também não exigem uma preparação dos episódios, podendo ser obtidas sobre o próprio arquivo MP3, minimizando o custo do processo de preparação dos dados, ao contrário do que acontece com a obtenção das características textuais, para as quais é necessária a transcrição prévia dos episódios.

5.2.3 Sensibilidade dos Modelos

Há indícios que, no contexto de predição de popularidade de podcasts, há modelos mais suscetíveis aos dados de entrada, como o MLP. Nesse modelo, a cada nova execução, com um conjunto de características distintos, observamos uma maior variação nos valores, o que se constata pelo desvio padrão encontrado a partir de todas as execuções. A acurácia teve desvio padrão de 5,39%, a precisão, de 1,92%, o recall, de 23,91%, e o F1-Score, de 10,92%. Já no modelo de Árvore de Decisão mostrou uma menor variação, quando treinado com diferentes características ou combinações de características. O maior desvio padrão ocorreu na métrica de recall (3,02%), ou seja para o contexto de árvore de decisão as características utilizadas possuem um menor impacto no processo de classificação.

6 CONCLUSÃO

Esta seção tem como objetivo apresentar as considerações finais sobre os principais tópicos abordados nesta dissertação, incluindo as contribuições alcançadas e indicações para trabalhos futuros.

Há alguns anos escutamos que este é o ano do podcast mas pelos movimentos que aconteceram principalmente no Brasil o ano de 2019 foi um ano determinante para a popularização deste tipo de mídia com grandes players entrando neste mercado e em 2020 foi a consolidação disso tudo demonstrando um movimento sem volta com um aumento de investimentos e de conteúdos produzidos e com isso a demanda por entender os podcasts só crescem e diante deste cenário que nossa pesquisa tem por objetivo explorar como grandes áreas da computação como processamento de texto e processamento de áudio podem contribuir para o processo de conhecimento dos podcasts vendo a viabilidade da utilização do texto e do áudio para predição de popularidade. Ao longo deste trabalho vimos quais as dificuldades para utilização de cada uma destas características no contexto dos podcasts onde segundo os resultados a relação sinal ruído (snr) e os metadados se demonstraram boas características para predizer o sucesso de determinado episódio com o público. Já as características textuais trouxeram um resultado um pouco abaixo dos encontrados com as outras features, o que nos desencoraja em trabalhos futuros a utilizar estas informações para a predição de popularidade principalmente pelo fato de que para utilizar o conteúdo textual se mostra necessário transcrever o áudio, atividade esta que existe um custo associado que utilizando os serviços do Google Cloud Plataform está estimado em média metade do tempo da duração total do episódio o que representa um custo bem maior as outras características exploradas durante este trabalho.

6.1 PRINCIPAIS CONTRIBUIÇÕES

Diante dos resultados obtidos, acreditamos que este trabalho contribui para um maior entendimento dos podcasts e dos fatores que tornam um determinado episódio popular ou não. Além disso, outras contribuições importantes podem ser elencadas:

- O avanço no entendimento sobre as características mais relevantes para a predição de popularidade, entre as características acústicas, a transcrição do áudio e os metadados fornecidos pelos editores;

- Um processo para recuperação e organização de uma base de dados rotulada de poscasts;
- Uma base de dados com rótulos de popularidade de podcasts, contendo os áudios, sua transcrição e os metadados associados;
- Um maior entendimento do impacto que cada característica tem no processo de predição de popularidade dos podcasts.

6.2 LIMITAÇÕES E TRABALHOS FUTUROS

Neste trabalho, utilizamos duas características para analisar os áudios - MFCC e SNR. Em trabalhos futuros, desejamos ampliar essa quantidade, com o objetivo de tentar identificar outros aspectos dos episódios - como a presença de trilha sonora, a quantidade de tempo em silêncio, a quantidade de participantes, ou até mesmo o humor de cada um. Um outro aspecto a se mapear são os produtores de cada um dos podcasts pois como avaliamos métricas do áudio, podemos encontrar um viés de produção, no qual os produtores de podcasts são peças em comum entre vários programas. A partir disso, realizar um estudo mais aprofundado dos efeitos dessas características na popularidade dos podcasts tentando minimizar os possíveis viés durante a análise.

Com o objetivo de minimizar o viés encontrado nas características textuais, pretende-se aplicar técnicas de redução de dimensionalidade, por exemplo, PCA ou Qui-quadrado nas mais de 65 mil características geradas a partir dos algorítimos TF-IDF e BoW, com proposito de melhorar os resultados individuais dessas características e diminuir a influencia delas, quando em conjunto com os metadados e as características de áudio.

Para os próximos trabalhos, é desejável aplicar o teste t pareado para verificar a significância estatística do ganho ou perda de performance de cada modelo, em relação a um modelo-base. Essa comparação pode ser feita utilizando como benchmark um modelo matemático que classifica aleatoriamente cada exemplo como popular ou não popular, com igual probabilidade.

Ao aplicar o processo de transcrição dos episódios, tivemos um custo financeiro e de processamento associado à conversão de áudio para texto, de cada um dos episódios. Isso limitou o tamanho da base de dados que foi construída neste trabalho. Assim, para o futuro, será preciso adotar um processo de transcrição dos episódios menos custoso, permitindo ampliar a base de dados e coletar, durante um espaço de tempo maior, os episódios populares pelo Chartable. Outra possível melhoria é a investigação envolvendo a comparação com informações

de popularidade de outras plataformas, como o Spotify, caso venham a tornar-se públicas, para verificar se os resultados são consistentes com os encontrados com a base construída sobre o Chartable. Liberações de mais dados podem vir a permitir, inclusive, evoluir o trabalho para tentar predizer a quantidade de ouvintes ou qual posição do ranking cada episódio ocupará, ao invés da utilização de um rótulo binário (popular/não popular), como foi o caso deste trabalho.

Por fim, no processo de predição de popularidade, a influência que os hosts e os entrevistados possuem podem ser um fator determinante para realizar a predição. Como trabalho futuro, planejamos mapear o perfil de cada um dos apresentadores e entrevistados nas mídias sociais, com o objetivo de entender o grau de relação entre a popularidade dos participantes e a do episódio.

REFERÊNCIAS

ABPOD. *PodPesquisa 2018 - Podcast no Brasil*. 2018. [Acessado 27 janeiro 2021]. Disponível em: <https://www.slideshare.net/greicematos/podpesquisa-2018-podcast-no-brasil>.

ALETRAS, N.; TSARAPATSANIS, D.; PREOŢIUC-PIETRO, D.; LAMPOS, V. Predicting judicial decisions of the european court of human rights: A natural language processing perspective. *PeerJ Computer Science*, PeerJ Inc., v. 2, p. e93, 2016.

ANDERSEN, J. S. Using the echo nest's automatically extracted music features for a musicological purpose. In: IEEE. *2014 4th International workshop on cognitive information processing (CIP)*. [S.l.], 2014. p. 1–6.

ARAÚJO, M.; GONÇALVES, P.; BENEVENUTO, F.; CHA, M. Métodos para análise de sentimentos no twitter. In: SN. *Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)*. [S.l.], 2013. p. 19.

BESACIER, L.; BARNARD, E.; KARPOV, A.; SCHULTZ, T. Automatic speech recognition for under-resourced languages: A survey. *Speech communication*, Elsevier, v. 56, p. 85–100, 2014.

BESIK, L. *Google named a Leader in the Gartner 2020 Magic Quadrant for Cloud AI Developer Services*. 2020. [Acessado 27 janeiro 2021]. Disponível em: <https://cloud.google.com/blog/products/ai-machine-learning/google-named-leader-gartner-2020-magic-quadrant-cloud-ai-developer-services>.

BESSER, J.; LARSON, M.; HOFMANN, K. Podcast search: User goals and retrieval technologies. *Online information review*, Emerald Group Publishing Limited, 2010.

CHEN, D. L.; MOONEY, R. J. Learning to sportscast: a test of grounded language acquisition. In: *Proceedings of the 25th international conference on Machine learning*. [S.l.: s.n.], 2008. p. 128–135.

CHEN, J.; WANG, Y.; WANG, D. A feature study for classification-based speech separation at low signal-to-noise ratios. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, IEEE, v. 22, n. 12, p. 1993–2002, 2014.

CLIFTON, A.; REDDY, S.; YU, Y.; PAPPU, A.; REZAPOUR, R.; BONAB, H.; ESKEVICH, M.; JONES, G.; KARLGREN, J.; CARTERETTE, B. et al. 100,000 podcasts: A spoken english document corpus. In: *Proceedings of the 28th International Conference on Computational Linguistics*. [S.l.: s.n.], 2020. p. 5903–5917.

COLD, S. J. Using really simple syndication (rss) to enhance student research. *ACM SIGITE Newsletter*, ACM New York, NY, USA, v. 3, n. 1, p. 6–9, 2006.

DAVIS, K. H.; BIDDULPH, R.; BALASHEK, S. Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 24, n. 6, p. 637–642, 1952.

DAYAN, P.; SAHANI, M.; DEBACK, G. Unsupervised learning. *The MIT encyclopedia of the cognitive sciences*, MIT Press, p. 857–859, 1999.

DENECKE, K. Using sentiwordnet for multilingual sentiment analysis. In: IEEE. *2008 IEEE 24th international conference on data engineering workshop*. [S.l.], 2008. p. 507–512.

DOWNIE, J. S. Music information retrieval. *Annual review of information science and technology*, Wiley Online Library, v. 37, n. 1, p. 295–340, 2003.

EZHILARASI, R.; MINU, R. Automatic emotion recognition and classification. *Procedia Engineering*, Elsevier, v. 38, p. 21–26, 2012.

FILHO, P. B.; PARDO, T. A. S.; ALUÍSIO, S. An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis. In: *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*. [S.l.: s.n.], 2013.

G1. *Jornalismo da Globo lança novos podcasts*. 2019. [Acessado 27 janeiro 2021]. Disponível em: <https://g1.globo.com/podcast/noticia/2019/08/25/jornalismo-da-globo-lanca-novos-podcasts.ghtml>.

GELLI, F.; URICCHIO, T.; BERTINI, M.; BIMBO, A. D.; CHANG, S.-F. Image popularity prediction in social media using sentiment and context features. In: *Proceedings of the 23rd ACM international conference on Multimedia*. [S.l.: s.n.], 2015. p. 907–910.

GIL, A. C. et al. *Como elaborar projetos de pesquisa*. [S.l.]: Atlas São Paulo, 2002. v. 4.

GOOGLE. *https://cloud.google.com/gartner-caids?hl=pt-br*. 2021. [Acessado 27 janeiro 2021]. Disponível em: <https://cloud.google.com/gartner-caids?hl=pt-br>.

HUANG, A. H.; YEN, D. C.; ZHANG, X. Exploring the potential effects of emoticons. *Information & Management*, Elsevier, v. 45, n. 7, p. 466–473, 2008.

INGASON, A. K.; HELGADÓTTIR, S.; LOFTSSON, H.; RÖGNVALDSSON, E. A mixed method lemmatization algorithm using a hierarchy of linguistic identities (holi). In: SPRINGER. *International Conference on Natural Language Processing*. [S.l.], 2008. p. 205–216.

IRIE, K.; SCHLÜTER, R.; NEY, H. Bag-of-words input for long history representation in neural network-based language models for speech recognition. In: *Sixteenth Annual Conference of the International Speech Communication Association*. [S.l.: s.n.], 2015.

ITTICHAICHAREON, C.; SUKSRI, S.; YINGTHAWORNSUK, T. Speech recognition using mfcc. In: *International conference on computer graphics, simulation and modeling*. [S.l.: s.n.], 2012. p. 135–138.

JIAARO. *PyDub - Manipulate audio with a simple and easy high level interface*. 2020. [Acessado 27 janeiro 2021]. Disponível em: <https://github.com/jiaaro/pydub>.

JUANG, B.-H.; RABINER, L. R. Automatic speech recognition–a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, v. 1, p. 67, 2005.

JUNKER, M.; HOCH, R.; DENGEL, A. On the evaluation of document analysis components by recall, precision, and accuracy. In: IEEE. *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*. [S.l.], 1999. p. 713–716.

KHURANA, D.; KOLI, A.; KHATTER, K.; SINGH, S. Natural language processing: State of the art, current trends and challenges. *arXiv preprint arXiv:1708.05148*, 2017.

KIM, S.-B.; HAN, K.-S.; RIM, H.-C.; MYAENG, S. H. Some effective techniques for naive bayes text classification. *IEEE transactions on knowledge and data engineering*, IEEE, v. 18, n. 11, p. 1457–1466, 2006.

KOTSIANTIS, S. B.; ZAHARAKIS, I.; PINTELAS, P. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, Amsterdam, v. 160, n. 1, p. 3–24, 2007.

KUDO, T.; MATSUMOTO, Y. Chunking with support vector machines. In: *Second Meeting of the North American Chapter of the Association for Computational Linguistics*. [S.l.: s.n.], 2001.

LEE, C.-H.; SOONG, F. K.; PALIWAL, K. K. *Automatic speech and speaker recognition: advanced topics*. [S.l.]: Springer Science & Business Media, 2012. v. 355.

LERMAN, K.; HOGG, T. Using a model of social dynamics to predict popularity of news. In: *Proceedings of the 19th international conference on World wide web*. [S.l.: s.n.], 2010. p. 621–630.

LI, H.; MA, X.; WANG, F.; LIU, J.; XU, K. On popularity prediction of videos shared in online social networks. In: *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. [S.l.: s.n.], 2013. p. 169–178.

LOGAN, B. et al. Mel frequency cepstral coefficients for music modeling. In: CITESEER. *Ismir*. [S.l.], 2000. v. 270, p. 1–11.

LÓPEZ, G.; QUESADA, L.; GUERRERO, L. A. Alexa vs. siri vs. cortana vs. google assistant: a comparison of speech-based natural user interfaces. In: SPRINGER. *International Conference on Applied Human Factors and Ergonomics*. [S.l.], 2017. p. 241–250.

LOWERRE, B. The harpy speech understanding system. In: *Readings in speech recognition*. [S.l.: s.n.], 1990. p. 576–586.

MARTIKAINEN, K. *Audio-based stylistic characteristics of Podcasts for search and recommendation: a user and computational analysis*. Dissertação (Mestrado) — University of Twente, 2020.

MCFEE, B.; RAFFEL, C.; LIANG, D.; ELLIS, D. P.; MCVICAR, M.; BATTENBERG, E.; NIETO, O. librosa: Audio and music signal analysis in python. In: CITESEER. *Proceedings of the 14th python in science conference*. [S.l.], 2015. v. 8, p. 18–25.

MCKINNEY, W. et al. pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, Seattle, v. 14, n. 9, p. 1–9, 2011.

MIZUNO, J.; OGATA, J.; GOTO, M. A similar content retrieval method for podcast episodes. In: IEEE. *2008 IEEE Spoken Language Technology Workshop*. [S.l.], 2008. p. 297–300.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of machine learning*. [S.l.]: MIT press, 2018.

MUDA, L.; BEGAM, M.; ELAMVAZUTHI, I. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *arXiv preprint arXiv:1003.4083*, 2010.

NAZARI, Z.; CHARBUILLET, C.; PAGES, J.; LAURENT, M.; CHARRIER, D.; VECCHIONE, B.; CARTERETTE, B. Recommending podcasts for cold-start users based on music listening and taste. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. [S.l.: s.n.], 2020. p. 1041–1050.

NORIEGA, L. Multilayer perceptron tutorial. *School of Computing. Staffordshire University*, Citeseer, 2005.

OSHIKAWA, R.; QIAN, J.; WANG, W. Y. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*, 2018.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, JMLR. org, v. 12, p. 2825–2830, 2011.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PIKE, K. L. Phonetics: A critical analysis of phonetic theory and a technic for the practical description of sounds. ERIC, 1972.

POPOVIC, I.; CULIBRK, D.; MIRKOVIC, M.; VUKMIROVIC, S. Automatic speech recognition and natural language understanding for emotion detection in multi-party conversations. In: *Proceedings of the 1st International Workshop on Multimodal Conversational AI*. [S.l.: s.n.], 2020. p. 31–38.

RAMCHOUN, H.; IDRISSI, M. A. J.; GHANOU, Y.; ETTAOUIL, M. Multilayer perceptron: Architecture optimization and training. *IJIMAI*, v. 4, n. 1, p. 26–30, 2016.

RAMOS, J. et al. Using tf-idf to determine word relevance in document queries. In: CITESEER. *Proceedings of the first instructional conference on machine learning*. [S.l.], 2003. v. 242, n. 1, p. 29–48.

RIBEIRO, G.; GOMES, R.; COSTA, S.; COSTA, W. Análise mel-cepstral na discriminação de patologias laríngeas. In: *XXIV Congresso Brasileiro de Engenharia Biomédica*. [S.l.: s.n.], 2014.

RISH, I. et al. An empirical study of the naive bayes classifier. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*. [S.l.: s.n.], 2001. v. 3, n. 22, p. 41–46.

RITTER, A.; CLARK, S.; ETZIONI, O. et al. Named entity recognition in tweets: an experimental study. In: *Proceedings of the 2011 conference on empirical methods in natural language processing*. [S.l.: s.n.], 2011. p. 1524–1534.

SALAMON, J.; BELLO, J. P. Unsupervised feature learning for urban sound classification. In: IEEE. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2015. p. 171–175.

SANDOVAL, A. M.; DÍAZ, J.; LLANOS, L. C.; REDONDO, T. Biomedical term extraction: Nlp techniques in computational medicine. *IJIMAI*, UNIR-Universidad Internacional de La Rioja, v. 5, n. 4, p. 51–59, 2019.

SCHÄUBLE, P. *Multimedia information retrieval: content-based information retrieval from large text and audio databases*. [S.l.]: Springer Science & Business Media, 2012. v. 397.

SECHIDIS, K.; TSOUMAKAS, G.; VLAHAVAS, I. On the stratification of multi-label data. In: SPRINGER. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. [S.l.], 2011. p. 145–158.

SHARPE, M. A review of metadata fields associated with podcast rss feeds. *arXiv preprint arXiv:2009.12298*, 2020.

SHAWAR, B. A.; ATWELL, E. Chatbots: are they really useful? In: *Ldv forum*. [S.l.: s.n.], 2007. v. 22, n. 1, p. 29–49.

SONG, K.; LI, C.; WANG, X.; YU, D.; LIU, F. The ucf podcast summarization system at trec 2020. *arXiv preprint arXiv:2011.04132*, 2020.

SRINIVASA-DESIKAN, B. *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. [S.l.]: Packt Publishing Ltd, 2018.

STEINBACH, M.; KARYPIS, G.; KUMAR, V. A comparison of document clustering techniques. 2000.

TATAR, A.; ANTONIADIS, P.; AMORIM, M. D. D.; FDIDA, S. From popularity prediction to ranking online news. *Social Network Analysis and Mining*, Springer, v. 4, n. 1, p. 174, 2014.

THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G.; CAI, D.; KAPPAS, A. Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, Wiley Online Library, v. 61, n. 12, p. 2544–2558, 2010.

THOMASON, J.; VENUGOPALAN, S.; GUADARRAMA, S.; SAENKO, K.; MOONEY, R. Integrating language and vision to generate natural language descriptions of videos in the wild. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. [S.l.: s.n.], 2014. p. 1218–1227.

TIWARI, V. Mfcc and its applications in speaker recognition. *International journal on emerging technologies*, Citeseer, v. 1, n. 1, p. 19–22, 2010.

TSAGKIAS, M.; LARSON, M.; RIJKE, M. D. Predicting podcast preference: An analysis framework and its application. *Journal of the American Society for information Science and Technology*, Wiley Online Library, v. 61, n. 2, p. 374–391, 2010.

TSAGKIAS, M.; LARSON, M.; WEERKAMP, W.; RIJKE, M. D. Podcred: A framework for analyzing podcast preference. In: *Proceedings of the 2nd ACM workshop on Information credibility on the web*. [S.l.: s.n.], 2008. p. 67–74.

VARTAKAVI, A.; GARG, A. Podsumm–podcast audio summarization. *arXiv preprint arXiv:2009.10315*, 2020.

VOUTILAINEN, A. Part-of-speech tagging. *The Oxford handbook of computational linguistics*, University Press, Oxford, p. 219–232, 2003.

WHITNER, G. *The meteoric rise of podcasting*. 2021. [Acessado 27 janeiro 2021]. Disponível em: <https://musicoomph.com/podcast-statistics/>.

WILBUR, W. J.; SIROTKIN, K. The automatic identification of stop words. *Journal of information science*, Sage Publications Sage CA: Thousand Oaks, CA, v. 18, n. 1, p. 45–55, 1992.

YANG, L.; WANG, Y.; DUNNE, D.; SOBOLEV, M.; NAAMAN, M.; ESTRIN, D. More than just words: Modeling non-textual characteristics of podcasts. In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. [S.l.: s.n.], 2019. p. 276–284.

YU, D.; DENG, L. *AUTOMATIC SPEECH RECOGNITION*. [S.l.]: Springer, 2016.

ZHANG, Y.; JIN, R.; ZHOU, Z.-H. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, Springer, v. 1, n. 1-4, p. 43–52, 2010.

ZHENG, C.; WANG, H. J.; ZHANG, K.; FAN, L. A baseline analysis for podcast abstractive summarization. *arXiv preprint arXiv:2008.10648*, 2020.

ZHENG, C.; ZHANG, K.; WANG, H. J.; FAN, L. A two-phase approach for abstractive podcast summarization. *arXiv preprint arXiv:2011.08291*, 2020.

ZUBANI, M.; IVAN, S.; GEREVINI, A. E. A comparison of natural language understanding services to build a chatbot in italian. 2020.

APÊNDICE A – RSS FEED PODCAST SOMOS CINTIA

Código Fonte 1 – Parte do arquivo XML do RSS Feed

```
<?xml version="1.0" encoding="UTF-8"?><rss xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:atom="http://www.w3.org/2005/Atom" version="2.0" xmlns:itunes="http://www.itunes.com/dtds/podcast-1.0.dtd" xmlns:anchor="https://anchor.fm/xmlns">
    <channel>
        <title><![CDATA[Somos Cintia]]></title>
        <description><![CDATA[O Podcast do Grupo Cintia - Ciência e Tecnologia da Informação com Elas - que faz parte do Centro de Informática da UFPE, é um programa quinzenal de discussão e divulgação científica com foco em mulheres e tecnologia. Aqui vamos conversar sobre pesquisas e ações realizadas por mulheres do Cintia, do Brasil e do mundo. Tudo regado com as questões envolvendo equidade de gênero, tanto no campo da educação quanto do trabalho. Para saber mais e interagir conosco, siga e comente nas nossas redes sociais. Instagram e Twitter: @grupocintia
#SomosCintia #MulheresPodcasters]]></description>
        <link>http://www.cin.ufpe.br/~cintia</link>
        <image>
            <url>https://d3t3ozftmdmh3i.cloudfront.net/production/podcast_uploaded/2290545/2290545-1589417606349-4f9171067e195.jpg</url>
            <title>Somos Cintia</title>
            <link>http://www.cin.ufpe.br/~cintia</link>
        </image>
        <generator>Anchor Podcasts</generator>
        <lastBuildDate>Mon, 01 Mar 2021 21:34:44 GMT</lastBuildDate>
        <atom:link href="https://anchor.fm/s/e3faea4/podcast/rss" rel="self" type="application/rss+xml"/>
        <author><![CDATA[Grupo Cintia]]></author>
        <copyright><![CDATA[Grupo Cintia]]></copyright>
        <language><![CDATA[pt-br]]></language>
        <atom:link rel="hub" href="https://pubsubhubbub.appspot.com/"/>
        <itunes:author>Grupo Cintia</itunes:author>
        <itunes:summary>O Podcast do Grupo Cintia - Ciência e Tecnologia da Informação com Elas - que faz parte do Centro de Informática da UFPE, é um programa quinzenal de discussão e divulgação científica com foco em mulheres e tecnologia. Aqui vamos conversar sobre pesquisas e ações realizadas por mulheres do Cintia, do Brasil e do mundo. Tudo regado com as questões envolvendo equidade de gênero, tanto no campo da educação quanto do trabalho. Para saber mais e interagir conosco, siga e comente nas nossas redes sociais. Instagram e Twitter: @grupocintia
```

```
#SomosCintia #MulheresPodcasters</itunes:summary>
        <itunes:type>episodic</itunes:type>
        <itunes:owner>
            <itunes:name>Grupo Cintia</itunes:name>
            <itunes:email>podcasts19+e3faea4@anchor.fm</itunes:email>
        </itunes:owner>
        <itunes:explicit>No</itunes:explicit>
        <itunes:category text="Technology"/>
        <itunes:image href="https://d3t3ozftmdmh3i.cloudfront.net/production/
            podcast_uploaded/2290545/2290545-1589417606349-4f9171067e195.jpg"/>
        <item>
            <title><![CDATA[Me seguraaa, tem uma comunidade ali e eu quero
                participar!!!]]></title>
            <description><![CDATA[<p><strong>#S02E32 -</strong> Oláaaaa!!!!
                Depois de um tempinho de recesso, estamos de volta com uma
                convidada que já estava no nosso radar fazia tempo! E a demora
                pra gravar com ela foi recompensada com esse ep que ficou uma
                lindeza :D Conversamos com Ana Cecília Vieira que é membra ativa
                da comunidade PyLadies Recife (e do Conselho Global PyLadies
                2020-2022), embaixadora do Women in Data Science (WiDS) Recife,
                graduada em Biblioteconomia pela UFPE e atualmente servidora pú
                blica na mesma instituição. É uma entusiasta das comunidades e
                tem uma gatinha preta chamada Nísia Floresta. Escuta pra se
                apaixonar por ela como a gente se apaixonou :D E depois de ouvir
                não esquece de compartilhar esse ep usando as hashtags #
                SomosCintia e #MulheresPodcasters</p>
<p><strong>Referências citadas nesse episódio</strong>:</p>
<p><a href="https://www.instagram.com/pyladiesrecife">Pyladies Recife</a></p>
<p><a href="https://anchor.fm/grupo-cintia/episodes/Mulheres-que-Levantam-outras-
    Mulheres-e9abtn">Ep de Andreza</a></p>
<p><a href="https://www.widsrecife.com.br/">Women in Data Science Recife</a></p>
<p><a href="http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.
    htm">Lei de Acesso a Informação</a></p>
<p><a href="http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777
    .htm">Política de Dados Abertos do Poder Executivo</a></p>
<p><a href="https://www.ok.org.br/noticia/open-knowledge-brasil-abre-chamada-para
    -realizar-censo-de-diarios-municipais/">Open Knowledge Brasil abre chamada
    para realizar censo de diários municipais</a></p>
<p><a href="https://www.pucminas.br/PucVirtual/Pos-Graduacao/Paginas/Inteligencia
    -Artificial-e-Aprendizado-de-Maquina.aspx?moda=1&amp;polo=1&amp;area=11&amp;
    curso=2975&amp;situ=1">Inteligência Artificial e Aprendizado de Máquina (PUC
    Minas)</a></p>
<p><a href="https://elections.pyladies.com/pt/pages/council.html">Conselho Global
     Pyladies</a></p>
<p><a href="https://serenata.ai/">Serenata de Amor</a></p>
<p><strong>Dicas Empoderadas:</strong></p>
<p>Coleção Feminismos Plurais - <a href="https://feminismosplurais.com.br/">https
```

```
    ://feminismosplurais.com.br/</a></p>
<p>Flaira Ferro - <a href="https://youtu.be/H3jv4Vlh844">https://youtu.be/
    H3jv4Vlh844</a></p>
<p>Amanda Gânimo - <a href="https://www.youtube.com/user/amandabas1mc">https
    ://www.youtube.com/user/amandabas1mc</a></p>
<p>Tatuagem - <a href="https://youtu.be/Y7bwGO_f_4Q">https://youtu.be/Y7bwGO_f_4Q
    </a></p>
<p>Podcast TecnoPolítica - <a href="http://tecnopolitica.blog.br/">http://
    tecnopolitica.blog.br/</a></p>
<p>Podcast Cirandeiras - <a href="https://anchor.fm/cirandeiraspodcast">https://
    anchor.fm/cirandeiraspodcast</a></p>
<p>Podcast Arrumadinho - <a href="https://open.spotify.com/show/3
    nn1mVTH30T9HzZHW0Itd8">https://open.spotify.com/show/3nn1mVTH30T9HzZHW0Itd8</
    a></p>
<p>Bridgerton - <a href="https://youtu.be/pyi8QAlHR8k">https://youtu.be/
    pyi8QAlHR8k</a> <a href="https://www.netflix.com/br/title/80232398">https://
    www.netflix.com/br/title/80232398</a></p>
<p>Podcast Pizza de Dados - <a href="https://pizzadedados.com/">https://
    pizzadedados.com/</a></p>
<p>Lupin - <a href="https://youtu.be/WpSweeyuHb8">https://youtu.be/WpSweeyuHb8</a
    ></p>
<p>Podcast Mamede Connection - <a href="https://podcasts.google.com/feed/
    aHR0cHM6Ly9hbmNob3IuZm0vcy8yYzNiNDJiMC9wb2RjYXN0L3Jzcw==">https://podcasts.
    google.com/feed/aHR0cHM6Ly9hbmNob3IuZm0vcy8yYzNiNDJiMC9wb2RjYXN0L3Jzcw==</a
    ></p>
]]></description>
            <link>https://anchor.fm/grupo-cintia/episodes/Me-seguraaa--tem-uma-
                comunidade-ali-e-eu-quero-participar-eqnkoe</link>
            <guid isPermaLink="false">257abea5-4967-44e6-8ad0-441207eda303</guid>
            <dc:creator><![CDATA[Grupo Cintia]]></dc:creator>
            <pubDate>Thu, 25 Feb 2021 21:00:00 GMT</pubDate>
            <enclosure url="https://anchor.fm/s/e3faea4/podcast/play/27037902/
                https%3A%2F%2Fd3ctxlq1ktw2nl.cloudfront.net%2Fstaging%2F2021
                -1-21%2F4e283046-20e2-248c-7429-cea8b5d6986c.mp3" length="
                36390434" type="audio/mpeg"/>
            <itunes:summary>&lt;p&gt;&lt;strong&gt;#S02E32 -&lt;/strong&gt; Olá
                aaaa!!!! Depois de um tempinho de recesso, estamos de volta com
                uma convidada que já estava no nosso radar fazia tempo! E a
                demora pra gravar com ela foi recompensada com esse ep que ficou
                uma lindeza :D Conversamos com Ana Cecília Vieira que é membra
                ativa da comunidade PyLadies Recife (e do Conselho Global
                PyLadies 2020-2022), embaixadora do Women in Data Science (WiDS)
                Recife, graduada em Biblioteconomia pela UFPE e atualmente
                servidora pública na mesma instituição. É uma entusiasta das
                comunidades e tem uma gatinha preta chamada Nísia Floresta.
                Escuta pra se apaixonar por ela como a gente se apaixonou :D E
                depois de ouvir não esquece de compartilhar esse ep usando as
```

```
                hashtags #SomosCintia e #MulheresPodcasters&lt;/p&gt;
&lt;p&gt;&lt;strong&gt;Referências citadas nesse episódio&lt;/strong&gt;:&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://www.instagram.com/pyladiesrecife&quot;&gt;Pyladies Recife&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://anchor.fm/grupo-cintia/episodes/Mulheres-que-Levantam-outras-Mulheres-e9abtn&quot;&gt;Ep de Andreza&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://www.widsrecife.com.br/&quot;&gt;Women in Data Science Recife&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm&quot;&gt;Lei de Acesso a Informação&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm&quot;&gt;Política de Dados Abertos do Poder Executivo&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://www.ok.org.br/noticia/open-knowledge-brasil-abre-chamada-para-realizar-censo-de-diarios-municipais/&quot;&gt;Open Knowledge Brasil abre chamada para realizar censo de diários municipais&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://www.pucminas.br/PucVirtual/Pos-Graduacao/Paginas/Inteligencia-Artificial-e-Aprendizado-de-Maquina.aspx?moda=1&amp;amp;amp;polo=1&amp;amp;amp;area=11&amp;amp;amp;curso=2975&amp;amp;amp;situ=1&amp;quot;&gt;Inteligência Artificial e Aprendizado de Máquina (PUC Minas)&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://elections.pyladies.com/pt/pages/council.html&quot;&gt;Conselho Global Pyladies&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://serenata.ai/&quot;&gt;Serenata de Amor&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;strong&gt;Dicas Empoderadas:&lt;/strong&gt;&lt;/p&gt;
&lt;p&gt;Coleção Feminismos Plurais - &lt;a href=&quot;https://feminismosplurais.com.br/&quot;&gt;https://feminismosplurais.com.br/&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Flaira Ferro -&amp;amp;nbsp;&lt;a href=&quot;https://youtu.be/H3jv4Vlh844&quot;&gt;https://youtu.be/H3jv4Vlh844&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Amanda Gânimo -&amp;amp;nbsp;&lt;a href=&quot;https://www.youtube.com/user/amandabas1mc&quot;&gt;https://www.youtube.com/user/amandabas1mc&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Tatuagem - &lt;a href=&quot;https://youtu.be/Y7bwGO_f_4Q&quot;&gt;https://youtu.be/Y7bwGO_f_4Q&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Podcast TecnoPolítica - &lt;a href=&quot;http://tecnopolitica.blog.br/&quot;&gt;http://tecnopolitica.blog.br/&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Podcast Cirandeiras - &lt;a href=&quot;https://anchor.fm/cirandeiraspodcast&quot;&gt;https://anchor.fm/cirandeiraspodcast&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Podcast Arrumadinho - &lt;a href=&quot;https://open.spotify.com/show/3nn1mVTH30T9HzZHW0Itd8&quot;&gt;https://open.spotify.com/show/3nn1mVTH30T9HzZHW0Itd8&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;Bridgerton - &lt;a href=&quot;https://youtu.be/pyi8QAlHR8k&quot;&gt;https://youtu.be/pyi8QAlHR8k&lt;/a&gt; &lt;a href=&quot;https://www.netflix.
```

```
    com/br/title/80232398&amp;quot;&amp;gt;https://www.netflix.com/br/title/80232398&amp;lt;/
    a&amp;gt;&amp;lt;/p&amp;gt;
&amp;lt;p&amp;gt;Podcast Pizza de Dados - &amp;lt;a href=&amp;quot;https://pizzadedados.com/&amp;quot
    ;&amp;gt;https://pizzadedados.com/&amp;lt;/a&amp;gt;&amp;lt;/p&amp;gt;
&amp;lt;p&amp;gt;Lupin - &amp;lt;a href=&amp;quot;https://youtu.be/WpSweeyuHb8&amp;quot;&amp;gt;https://
    youtu.be/WpSweeyuHb8&amp;lt;/a&amp;gt;&amp;lt;/p&amp;gt;
&amp;lt;p&amp;gt;Podcast Mamede Connection - &amp;lt;a href=&amp;quot;https://podcasts.google.com
    /feed/aHR0cHM6Ly9hbmNob3IuZm0vcy8yYzNiNDJiMC9wb2RjYXN0L3Jzcw==&amp;quot;&amp;gt;https
    ://podcasts.google.com/feed/
    aHR0cHM6Ly9hbmNob3IuZm0vcy8yYzNiNDJiMC9wb2RjYXN0L3Jzcw==&amp;lt;/a&amp;gt;&amp;lt;/p&amp;gt;
</itunes:summary>
            <itunes:explicit>No</itunes:explicit>
            <itunes:duration>3394</itunes:duration>
            <itunes:image href="https://d3t3ozftmdmh3i.cloudfront.net/production/
                podcast_uploaded_episode/2290545/2290545-1614201191582-
                e3585e5c3fbea.jpg"/>
            <itunes:season>2</itunes:season>
            <itunes:episode>32</itunes:episode>
            <itunes:episodeType>full</itunes:episodeType>
        </item>
        <item>
            <title><![CDATA[Em 2020 a gente Meteu a Colher... e você?]]></title>
            <description><![CDATA[<p><strong>#S02E31 -</strong> Nosso último epis
                ódio do ano tem a participação da maravilhosa <a href="https://
                www.instagram.com/renatalbertim/">Renata Albertim</a>, Diretora e
                 fundadora do <a href="https://www.instagram.com/appmeteacolher
                /">Mete a Colher</a>, jornalista, mestra, sonhadora e realizadora
                 de sonhos. Uma mulher que levanta outras mulheres e enxerga a
                tecnologia como aliada na luta por equidade de gênero e fim da
                violência contra as mulheres. Vem se inspirar e depois de ouvir n
                ão esquece de compartilhar esse ep usando as hashtags #
                SomosCintia e #MulheresPodcasters - P.S.: VOLTAMOS EM FEVEREIRO
                DE 2021 :D Feliz Natal e Feliz Ano Novo ;)<br>
</p>
<p><strong>Referências citadas nesse episódio</strong>:</p>
<p><a href="https://blogs.ne10.uol.com.br/mundobit/tag/startup-weekend/"><strong>
    Startup Weekend Women</strong></a></p>
<p><a href="https://labcodes.com.br/"><strong>Labcodes</strong></a><a href="https
    ://www.leeds.ac.uk/"><strong>&amp;nbsp;</strong></a></p>
<p><a href="https://youtu.be/9Rq-aL1UAbY"><strong>Como estrelas na Terra</strong
    ></a></p>
<p><strong>Dicas Empoderadas:</strong></p>
<p>Emicida: AmarElo É Tudo pra Ontem - <a href="https://www.netflix.com/br/title
    /81306298">https://www.netflix.com/br/title/81306298</a></p>
<p>A máquina do ódio - Patrícia Campos Mello - <a href="https://www.
    companhiadasletras.com.br/detalhe.php?codigo=14883">https://www.
    companhiadasletras.com.br/detalhe.php?codigo=14883</a></p>
```

```
<p>Quem se importa - <a href="https://youtu.be/P8j67-yR37I">https://youtu.be/P8j67-yR37I</a></p>
<p>A vida não é útil - Ailton Krenak - <a href="https://www.companhiadasletras.com.br/detalhe.php?codigo=14911">https://www.companhiadasletras.com.br/detalhe.php?codigo=14911</a></p>
<p>Tudo bem no natal que vem - <a href="https://www.netflix.com/br/title/81160045">https://www.netflix.com/br/title/81160045</a></p>
<p>Story of Seasons: Friends of Mineral Town - <a href="https://store.steampowered.com/app/978780/STORY_OF_SEASONS_Friends_of_Mineral_Town">https://store.steampowered.com/app/978780/STORY_OF_SEASONS_Friends_of_Mineral_Town</a></p>
<p>8 em Istambul - <a href="https://www.netflix.com/br/title/81106900 ">https://www.netflix.com/br/title/81106900 </a><br>
<br>
<strong>Sorteada do Curso da Udemy:</strong></p>
<p>Ana Alice Oliveira - <a href="https://www.instagram.com/alice___oliver">@alice___oliver</a></p>
<p><strong>Referência de música:</strong></p>
<p><a href="https://youtu.be/8ahD-2LK_Zw">Jingle Bells - Kevin MacLeod (No Copyright Music)</a></p>
]]></description>
            <link>https://anchor.fm/grupo-cintia/episodes/Em-2020-a-gente-Meteu-a-Colher----e-voc-eo4sio</link>
            <guid isPermaLink="false">7a7082ea-bca8-479d-ac72-960e03a0a77f</guid>
            <dc:creator><![CDATA[Grupo Cintia]]></dc:creator>
            <pubDate>Thu, 24 Dec 2020 21:00:00 GMT</pubDate>
            <enclosure url="https://anchor.fm/s/e3faea4/podcast/play/24326168/https%3A%2F%2Fd3ctxlq1ktw2nl.cloudfront.net%2Fstaging%2F2020-11-24%2Faf7774fa-22e4-8c29-8383-0d13396d25c8.mp3" length="51058321" type="audio/mpeg"/>
            <itunes:summary>&lt;p&gt;&lt;strong&gt;#S02E31 -&lt;/strong&gt; Nosso último episódio do ano tem a participação da maravilhosa &lt;a href=&quot;https://www.instagram.com/renatalbertim/&quot;&gt;Renata Albertim&lt;/a&gt;, Diretora e fundadora do &lt;a href=&quot;https://www.instagram.com/appmeteacolher/&quot;&gt;Mete a Colher&lt;/a&gt;, jornalista, mestra, sonhadora e realizadora de sonhos. Uma mulher que levanta outras mulheres e enxerga a tecnologia como aliada na luta por equidade de gênero e fim da violência contra as mulheres. Vem se inspirar e depois de ouvir não esquece de compartilhar esse ep usando as hashtags #SomosCintia e #MulheresPodcasters - P.S.: VOLTAMOS EM FEVEREIRO DE 2021 :D Feliz Natal e Feliz Ano Novo ;)&lt;br&gt;
&lt;/p&gt;
&lt;p&gt;&lt;strong&gt;Referências citadas nesse episódio&lt;/strong&gt;:&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://blogs.ne10.uol.com.br/mundobit/tag/startup-weekend/&quot;&gt;&lt;strong&gt;Startup Weekend Women&lt;/strong&gt;&lt;/a&gt
```

```
    ;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://labcodes.com.br/&quot;&gt;&lt;strong&gt;
    Labcodes&lt;/strong&gt;&lt;/a&gt;&lt;a href=&quot;https://www.leeds.ac.uk/&
    quot;&gt;&lt;strong&gt;&amp;nbsp;&lt;/strong&gt;&lt;/a&gt;&lt;/p&gt;
125 &lt;p&gt;&lt;a href=&quot;https://youtu.be/9Rq-aL1UAbY&quot;&gt;&lt;strong&gt;
    Como estrelas na Terra&lt;/strong&gt;&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;&lt;strong&gt;Dicas Empoderadas:&lt;/strong&gt;&lt;/p&gt;
127 &lt;p&gt;Emicida: AmarElo É Tudo pra Ontem - &lt;a href=&quot;https://www.netflix
    .com/br/title/81306298&quot;&gt;https://www.netflix.com/br/title/81306298&lt
    ;/a&gt;&lt;/p&gt;
&lt;p&gt;A máquina do ódio - Patrícia Campos Mello - &lt;a href=&quot;https://www
    .companhiadasletras.com.br/detalhe.php?codigo=14883&quot;&gt;https://www.
    companhiadasletras.com.br/detalhe.php?codigo=14883&lt;/a&gt;&lt;/p&gt;
129 &lt;p&gt;Quem se importa - &lt;a href=&quot;https://youtu.be/P8j67-yR37I&quot;&gt
    ;https://youtu.be/P8j67-yR37I&lt;/a&gt;&lt;/p&gt;
&lt;p&gt;A vida não é útil - Ailton Krenak - &lt;a href=&quot;https://www.
    companhiadasletras.com.br/detalhe.php?codigo=14911&quot;&gt;https://www.
    companhiadasletras.com.br/detalhe.php?codigo=14911&lt;/a&gt;&lt;/p&gt;
131 &lt;p&gt;Tudo bem no natal que vem - &lt;a href=&quot;https://www.netflix.com/br/
    title/81160045&quot;&gt;https://www.netflix.com/br/title/81160045&lt;/a&gt;&
    lt;/p&gt;
&lt;p&gt;Story of Seasons: Friends of Mineral Town - &lt;a href=&quot;https://
    store.steampowered.com/app/978780/STORY_OF_SEASONS_Friends_of_Mineral_Town&
    quot;&gt;https://store.steampowered.com/app/978780/
    STORY_OF_SEASONS_Friends_of_Mineral_Town&lt;/a&gt;&lt;/p&gt;
133 &lt;p&gt;8 em Istambul - &lt;a href=&quot;https://www.netflix.com/br/title
    /81106900 &quot;&gt;https://www.netflix.com/br/title/81106900&amp;nbsp;&lt;/a
    &gt;&lt;br&gt;
&lt;br&gt;
135 &lt;strong&gt;Sorteada do Curso da Udemy:&lt;/strong&gt;&lt;/p&gt;
&lt;p&gt;Ana Alice Oliveira - &lt;a href=&quot;https://www.instagram.com/
    alice___oliver&quot;&gt;@alice___oliver&lt;/a&gt;&lt;/p&gt;
137 &lt;p&gt;&lt;strong&gt;Referência de música:&lt;/strong&gt;&lt;/p&gt;
&lt;p&gt;&lt;a href=&quot;https://youtu.be/8ahD-2LK_Zw&quot;&gt;Jingle Bells -
    Kevin MacLeod (No Copyright Music)&lt;/a&gt;&lt;/p&gt;
139 </itunes:summary>
            <itunes:explicit>No</itunes:explicit>
141         <itunes:duration>3266</itunes:duration>
            <itunes:image href="https://d3t3ozftmdmh3i.cloudfront.net/production/
                podcast_uploaded_episode/2290545/2290545-1608685829665-3
                ed5edc00f8df.jpg"/>
143         <itunes:season>2</itunes:season>
            <itunes:episode>31</itunes:episode>
145         <itunes:episodeType>full</itunes:episodeType>
        </item>
147     <item>
            <title><![CDATA[Mais um podcast?]]></title>
```

```
149             <description><![CDATA[<p>#S01Ep00 - Chegamos na Podosfera... já
                    pensasse? ;) Mais um podcast!!! Mais um? ;) Huum... escuta aí o
                    nosso "Teaser" e vê o que tu acha :) E pra saber mais sobre a
                    gente, pra comentar o que tu achou, pra dar sugestão ou mesmo um
                    likezinho pra gente saber que tu passou por aqui, acessa nossas
                    redes nos endereços:  https://www.cin.ufpe.br/~cintia/ https
                    ://www.instagram.com/grupocintia/ https://twitter.com/grupocintia
                     Daqui a 15 dias a gente volta com nosso primeiro episódio \o/ #
                    SomosCintia #MulheresPodcasters A música que você escuta durante
                    o episódio foi criada por Kara Square. Mais informações: 2018
                    Kara Square Licensed to the public under http://creativecommons.
                    org/licenses/by-nc/3.0/ Verify at http://ccmixter.org/files/
                    mindmapthat/53770</p>
]]></description>
151             <link>https://anchor.fm/grupo-cintia/episodes/Mais-um-podcast-e5dt1t
                    </link>
                <guid isPermaLink="false">c96e2b28-ae84-02be-0556-7e6e354cbf54</guid>
153             <dc:creator><![CDATA[Grupo Cintia]]></dc:creator>
                <pubDate>Thu, 26 Sep 2019 21:00:00 GMT</pubDate>
155             <enclosure url="https://anchor.fm/s/e3faea4/podcast/play/4698621/
                    https%3A%2F%2Fd3ctxlq1ktw2nl.cloudfront.net%2Fproduction%2F2019
                    -8-22%2F24614359-44100-2-8e2245861b458.mp3" length="7665051" type
                    ="audio/mpeg"/>
                <itunes:summary>&lt;p&gt;#S01Ep00 - Chegamos na Podosfera... já
                    pensasse? ;) Mais um podcast!!! Mais um? ;) Huum... escuta aí o
                    nosso &quot;Teaser&quot; e vê o que tu acha :) E pra saber mais
                    sobre a gente, pra comentar o que tu achou, pra dar sugestão ou
                    mesmo um likezinho pra gente saber que tu passou por aqui, acessa
                     nossas redes nos endereços:&amp;nbsp; https://www.cin.ufpe.br/~
                    cintia/ https://www.instagram.com/grupocintia/ https://twitter.
                    com/grupocintia Daqui a 15 dias a gente volta com nosso primeiro
                    episódio \o/ #SomosCintia #MulheresPodcasters A música que você
                    escuta durante o episódio foi criada por Kara Square. Mais
                    informações: 2018 Kara Square Licensed to the public under http
                    ://creativecommons.org/licenses/by-nc/3.0/ Verify at http://
                    ccmixter.org/files/mindmapthat/53770&lt;/p&gt;
157 </itunes:summary>
                <itunes:explicit>No</itunes:explicit>
159             <itunes:duration>317</itunes:duration>
                <itunes:image href="https://d3t3ozftmdmh3i.cloudfront.net/production/
                    podcast_uploaded_episode/2290545/2290545-1568929132499-
                    b2ab3fa9392ef.jpg"/>
161             <itunes:season>1</itunes:season>
                <itunes:episodeType>trailer</itunes:episodeType>
163         </item>
        </channel>
165 </rss>
```