

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE AUTOMAÇÃO E SISTEMAS

Eduardo Campos Kneipp

Análise de áudios para Predição de No-Show

Florianópolis, Abril de 2019

Eduardo Campos Kneipp

Análise de áudios para Predição de No-Show

Relatório submetido à Universidade Federal de Santa Catarina como requisito para a aprovação na disciplina **DAS 5511: Projeto de Fim de Curso** do curso de Graduação em Engenharia de Controle e Automação.

Orientador(a): Prof Eduardo Camponogara

Florianópolis, Abril de 2019

Eduardo Campos Kneipp

Análise de áudios para Predição de No-Show

Esta monografia foi julgada no contexto da disciplina DAS5511: Projeto de Fim de Curso e aprovada na sua forma final pelo Curso de Engenharia de Controle e Automação.

Florianópolis(DATA)

Banca Examinadora:

Guilherme Larroyd
Orientador na Empresa
Zygo Tecnologia

Prof Eduardo Camponogara
Orientador no Curso
Universidade Federal de Santa Catarina

Prof ;nome do professor avaliador;
Avaliador
Universidade Federal de Santa Catarina

;nome aluno 1;
Debatedor
Universidade Federal de Santa Catarina

;nome aluno 2;
Debatedor
Universidade Federal de Santa Catarina

Dedicatória

Dedico esse trabalho aos meus pais, Rosângela Campos de Souza Kneipp e Carlos Eduardo Kneipp. Eles me proporcionaram as condições para seguir esta carreira que eu almejo tanto, mas a principal contribuição foi na construção do meu caráter. Sempre me apoiando nos tempos difíceis e dando liberdade para tomar minhas próprias decisões e cometer meus próprios erros. Obrigado por tudo.

Agradecimentos

Quero agradecer ao meu orientador Professor Eduardo Camponogara, por sempre indicar o caminho mais plausível para os planejamentos desse trabalho, além da compreensão e apoio durante o processo de execução. Quero agradecer também ao meu líder e colegas dentro da Zygo Tecnologia, Guilherme Larroyd, João Paulo Garavello Ghisio, e Bruna Borges. Eles proporcionaram um ambiente de trabalho onde pude me desenvolver muito e me deram todo apoio de forma que pude realizar muitas tarefas além das que envolvem este projeto.

Por fim quero agradecer também a todos aqueles que me proporcionaram base emocional para lidar com os desafios desse semestre. Meus amigos Diandra S. Moraes e Lucas Battistella, minha terapeuta, meus dois grupos de RPG que possuem meus amigos mais estimados e meus colegas de apartamento que lidaram com a minha bagunça.

I see now that the circumstances of one's birth are irrelevant. It is what you do with the gift of life that determines who you are.

Mewtwo

Resumo

Dentro da empresa Zygo Tecnologia existe uma área interna de Vendas chamada de **Inside Sales**, e um dos maiores problemas desta área é a ocorrência de **No-Show**. O **No-Show** é o não comparecimento de um possível cliente para a reunião de venda do nosso produto, estas reuniões são marcadas por ligações telefônicas realizadas pelos pré-vendas e o índice de **No-Show** por trimestre é em torno de 40%. O intuito deste projeto é desenvolver uma base sólida de conhecimento para analisar os arquivos de áudio destas ligações, procurando um padrão que possa trazer novas descobertas para a forma com que vendas realiza seus processos, e além disso aplicar algum **aprendizado de máquina** para criar um preditor de **No-Show**. Para este propósito os estudos foram direcionados a projetos que envolviam **análise de sentimentos**, e como realizar **o processamento digital dos sinais de fala**. Foi utilizada a biblioteca em Python **LibROSA** para extrair características de áudio e analisar a relevância destas características referente à incidência ou não de **No-Show**.

Palavras-chave: aprendizado de máquina, LibROSA, No-Show, Inside Sales, análise de sentimentos, processamento digital de sinais de fala.

Abstract

Within the company Zygo Technology there is an internal sales area called **Inside Sales** which one of the biggest problems in this area is the occurrence of **No-Show**. The **No-Show** is a non-attendance of a possible customer for the sales meeting of our product, these meetings are marked by phone calls made by pre-sales and the **No-Show** index per quarter is around 40 %. The purpose of this project is to develop a solid foundation of knowledge to analyze the audio files of these calls. Also look for patterns that can bring new discoveries to the way sales carry out their processes, and also apply some **machine learning** to create a **No-Show** Predictor. For this purpose the studies were directed to projects that involved **sentiment analysis**, such as performing the **digital processing of speech signals**. We used the library in Python **LibROSA** to extract audio features and analyze the relevance of these features regarding the incidence or not of **No-Show**.

Keywords: machine learning, LibROSA, No-Shows Predictor, Inside Sales, sentiment analysis, digital processing of speech signals.

Lista de Figuras

| | | |
|-----|---|----|
| 1.1 | Linha Temporal da Zygo | 13 |
| 2.1 | Visão geral do Lead Soul Burguer no SumForce | 19 |
| 2.2 | Visão geral do Lead Delicia Natural no SumForce | 20 |
| 2.3 | Sistema Final | 21 |
| 3.1 | Linha Temporal da Zygo | 30 |
| 4.1 | Gráfico de ritmo | 36 |
| 4.2 | Gráfico de potencia | 37 |
| 4.3 | Espectograma na escala Mel | 38 |
| 4.4 | Diagramas de dispersão par a par | 41 |
| 4.5 | Gráfico de tempo de duração pelo ritmo | 42 |
| 4.6 | Gráfico de tempo de duração pela potência | 43 |

Lista de Abreviaturas

- PLG - Product Led Growth
- BDR - Business Development Representative
- SDR - Sales Development Representative
- AWS - Amazon Web Services
- SQL - Structured Query Language
- CSV - Comma Separated Values
- MIR - Music Information Retrieval
- SPIN - Situation Problem Implicatiion Need-Payoff
- API - Application programming interface
- LPC - linear predictor coefficients
- SMC - short-time coherence method
- TEO - Teafer-energy-operator
- S3 - Simple Storage Service
- WAV - Waveform Audio File Format
- LPCM - Linear pulse-code modulation

Conteúdo

| | | |
|----------|--|-----------|
| 1 | Introdução | 12 |
| 1.1 | Zygo Tecnologia | 13 |
| 1.2 | Caracterização do Problema | 14 |
| 1.3 | Solução | 15 |
| 1.4 | Objetivos | 15 |
| 1.5 | Relevância Acadêmica | 16 |
| 2 | Definição do Problema | 17 |
| 2.1 | Inside Sales | 17 |
| 2.1.a | Pré-vendedores | 18 |
| 2.1.b | SumForce | 18 |
| 2.2 | Casos reais | 19 |
| 2.3 | Soul Jazz Burguer | 19 |
| 2.4 | Delicia Natural | 20 |
| 2.5 | Sistematização do Problema | 20 |
| 3 | Embasamento Teórico e Ferramentas | 22 |

| | | |
|----------|---|-----------|
| 3.1 | Ferramentas | 22 |
| 3.1.a | SciPy | 23 |
| 3.1.b | LibROSA | 24 |
| 3.1.c | Scikit-learn | 24 |
| 3.1.d | PyAudioAnalysis | 25 |
| 3.2 | Embasamento Teórico | 25 |
| 3.2.a | Softwares de Transcrição de Áudio | 26 |
| 3.2.b | Análise de Sentimentos | 27 |
| 3.2.c | SPIN Selling | 29 |
| 3.2.d | Pré-Processamento de Dados | 31 |
| 3.3 | Escolha das Características do LibROSA | 31 |
| 4 | Aplicação à Predição de <i>Shows/No-Shows</i> e Resultados | 33 |
| 4.1 | Aplicação à Predição de <i>No-Shows</i> | 33 |
| 4.1.a | PostgreSQL e o SumForce | 33 |
| 4.1.b | Amazon S3 | 34 |
| 4.1.c | Separação de Canais | 35 |
| 4.1.d | Retirada das Características | 35 |
| 4.2 | Análise e Resultados | 39 |
| 5 | Conclusões | 44 |

Capítulo 1

Introdução

Existe uma tendência dentro de empresas de tecnologia de automatizar e aperfeiçoar o maior número de processos possíveis. A tecnologia está tão presente que as maiores empresas de tecnologia estão agora adotando o *PLG Product Led Growth*. Essa abordagem de mercado muda completamente a estrutura padrão utilizada nas empresas pelo mundo, a forma padrão de gerar crescimento na sua empresa segue a seguinte ordem: Marketing gera recursos para vendas convencer os clientes a comprar, para que então o sucesso do cliente implemente o produto para que só neste momento o cliente comece a receber os retornos de sua compra. O *PLG* é o novo modelo onde a Jornada do Cliente está sempre ligada ao produto, desde o primeiro momento que o cliente ganha conhecimento de uma empresa que segue o *PLG*, esse cliente estará colhendo retornos, o produto é a porta de entrada do cliente, muitas vezes em uma versão grátis e com o crescimento das necessidades a evolução do cliente em planos mais avançados.

Este novo conceito obriga todas as áreas a serem mais dedicadas a dados, a compreender as vantagens de uma automação bem feita, e até mesmo a criação de áreas nunca vistas antes como *People Analytics*, a utilização de ciência de dados em gestão de talentos.

Estes feitos geram uma nova necessidade para os engenheiros e cientistas da computação que desenvolvem essas ferramentas, a necessidade conhecer cada vez mais processos de diferentes áreas como administração, marketing, psicologia e comércio. Um Engenheiro de Controle e Automação, porém está acostumado a sistematizar diferentes modelos, sejam eles relacionados à elétrica, química, mecânica ou até mesmo computacional. Esta visão sistêmica está nos dando um diferencial nas mais vari-

adas necessidades que surgem dentro destas empresas. Foi este diferencial que me direcionou ao projeto que estou encarregado dentro da Zygo.

1.1 Zygo Tecnologia

A história da Zygo esta marcada pela mudança, o que representa exatamente a capacidade de seguir e superar as tendências do mercado. Mudar é tão presente dentro da Zygo que compõe nosso propósito: "Pessoas que se reinventam, reinventam tudo ao redor o tempo todo." por este motivo é importante conhecer a jornada da empresa. Este histórico segue o caminho dos fundadores até a marca Zygo.

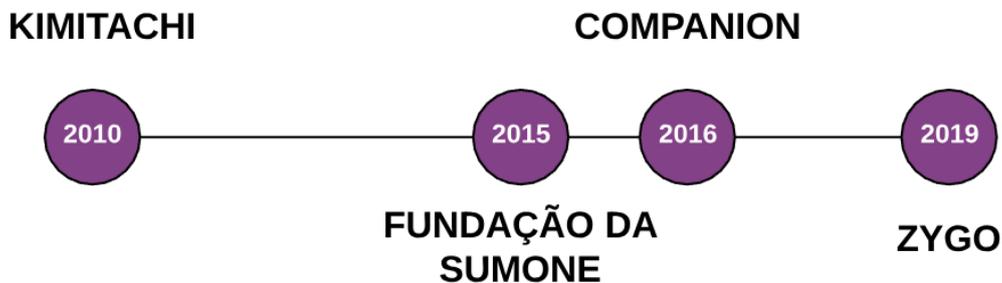


Figura 1.1: Linha Temporal da Zygo

2010 – KIMITACHI

Antes de ser uma empresa de marketing para negócios locais os fundadores experimentaram na pele o que é ter um negócio e os enormes desafios de mantê-lo sempre cheio e vendendo. O Kimitachi um pequeno delivery de comida japonesa em Florianópolis que só vendia online, recebeu diversos prêmios internacionais de inovação por ter revolucionado o mercado com sua tecnologia. Permitia, entre outras coisas, que o cliente assistisse seu pedido sendo feito ao vivo e online. Foi fazendo os restaurantes crescerem que entenderam a importância da tecnologia e do relacionamento com os clientes. Chegaram a oito unidades e foram comprados pelo Grupo Trigo, um dos maiores franqueadores de restaurantes do país.

2015 - SUMONE

No Rio de Janeiro, enquanto faziam a transição dos restaurantes para o Grupo

Trigo, perceberam o imenso potencial que as ferramentas que desenvolveram para o mundo online poderiam ter em centenas de outros tipos de negócios locais. A porta de entrada para esse mundo apareceu através de um aplicativo, chamado Bonuz, capitaneado por um novo sócio. Rapidamente viram a sinergia e resolveram se juntar. Aí nascia a SumOne.

2016 - COMPANION

Depois de diversas investidas com o Bonuz no mercado, perceberam que para entregar crescimento de verdade para os negócios locais era preciso ir além do aplicativo. Foi neste período em que começaram a desenvolver a primeira versão do sistema de marketing para restaurantes, o Companion, construído para fazer os negócios crescerem cuidando dos seus clientes.

2019 - ZYGO

Três anos depois de iniciarem essa jornada, foi gerado crescimento para milhares de negócios em todo o país e atingimos mais de nove milhões de consumidores. Para chegar até o sonho de ter centenas de milhares de negócios locais usando estas soluções, resolveram criar um novo nome e ampliar o produto. O novo nome, Zygo, veio da palavra grega zygotos, que significa célula reprodutora, a união que dá origem a novos seres. É isso que querem gerar: novos negócios, novas pessoas, novas perspectivas, todas orientadas ao crescimento dos negócios locais.

A Zygo Tecnologia é uma empresa que possui uma base de dados muito grande, esta base esta majoritariamente relacionada aos nossos clientes e por este motivo existe muito espaço para diferentes aplicações de algoritmos de classificação em ferramentas para estes usuários. Por outro lado nossos clientes não buscam por ferramentas complexas, pois muitas vezes não conseguem compreendê-las, e até mesmo desconfiam de "tecnologias diferentes". Considerando então que este tipo de aplicação não é prioridade para nosso mercado foi estabelecido que desenvolveríamos ferramentas internas, para ganhar este conhecimento e futuramente aplicar da forma mais intuitiva possível para os nossos clientes.

1.2 Caracterização do Problema

A Zygo Tecnologia é uma empresa que possui sua própria máquina de crescimento interna, o sistema de vendas que nós chamamos de *Inside Sales*. Este sistema

de vendas possui pré-vendedores que ligam para os possíveis clientes (*Leads*) e marcam um vídeo conferencia com os vendedores, porém em cerca de 40% dessas reuniões marcadas os leads não aparecem (*No-show*). Os *No-shows* são um prejuízo para a Zygo, visto que nossos vendedores não são capazes de realizar nenhum outro trabalho dentro do período que estava reservado para aquela reunião. Considerando este problema foi iniciado o projeto no qual eu estou encarregado.

1.3 Solução

O projeto é criar um preditor de *No-Shows* utilizando como dado às ligações dos pré-vendedores para tentar encontrar padrões que nos permitam classificar as reuniões futuras como mais prováveis de acontecer ou mais prováveis de ocorrer *no-show*. Será utilizada neste projeto uma biblioteca em Python para manipular e extrair características(*features*) dos áudios das ligações e utilizar essas características para classificar os áudios com um sistema simples de *Machine Learning*.

1.4 Objetivos

Um preditor de *No-Shows* pode beneficiar o trabalho dos Vendedores, dos pré-vendedores e futuramente até mesmo da área de Sucesso do Cliente da Zygo. Focando nos Vendedores, a possibilidade de um *no-show* permitira esses vendedores de planejarem outras atividades no período da reunião, em casos mais extremos até mesmo marcar duas reuniões no mesmo horário como um processo de *Overbooking* de companhias aéreas. Já para os pré-vendedores a métrica de *No-Show* é importante para classificar a qualidade do seu trabalho, assim caso fosse detectada a possibilidade de *No-Show*, outras interações poderiam ser feitas com este *lead* antes da reunião ser realizada para gerar mais engajamento. E por fim com esta ferramenta funcionando dentro de *Inside Sales*, ela poderia ser adaptada para Sucesso de Cliente, que é outra área da empresa que perde muita eficiência em *No-Show* de clientes em ligações de acompanhamento ou implantação.

1.5 Relevância Acadêmica

O projeto Preditor de *No-Shows* envolveu diversos tópicos estudados durante minha graduação em Engenharia de Controle de Automação, além disso esta intimamente ligada ao propósito do curso de criar estudantes que possuem uma visão sistêmica que ao lidarem com um problema conseguem racionalizar uma forma automatizada de solucionar. Indicando quais são as entradas necessárias e as saídas desejadas é possível encontrar a melhor solução.

Uma das matérias mais relacionadas a este projeto é Sinais e Sistemas Lineares que me permitiu compreender os sinais de áudio por me fornecer a base em sinais digitais e transformada de Fourier. Além disso a matéria Fundamentos da Estrutura da Informação me permitiu compreender como utilizar e explorar as bibliotecas utilizadas nesse projeto. Por fim a matéria Inteligência Artificial Aplicada a Controle e Automação me permitiu compreender quais eram os insumos necessários para aplicar um sistema de aprendizado de máquina, o que foi o foco deste trabalho.

Além dessas matérias outras foram importantes, por exemplo, a obtenção dos arquivos de áudio só foi possível aplicando o conhecimento da disciplina Fundamentos de Sistemas de Banco de Dados, para explorar os bancos de dados da Zygo. E mesmo explorando devidamente as bibliotecas utilizadas no sistema, a compreensão só foi possível pela base em cálculo numérico fornecida pela matéria Cálculo Numérico para Controle e Automação.

Capítulo 2

Definição do Problema

Uma empresa de tecnologia de alto crescimento recebe investimentos para que possa gastar muito dinheiro expandindo o seu mercado, para isso geralmente essas empresas criam o próprio sistema de vendas interno, no caso da Zygo não foi diferente. O time de Inside Sales da Zygo é resultado de uma equipe sempre disposta a mudanças que já atribuiu vários sistemas diferentes, absorveu ideias de consultores e comparou os resultados com outras empresas similares. O time esta ainda em busca de melhorias mas o sistema atual, chamado de time de especialistas, esta bem consolidado e será explicado para dar um contexto melhor ao problema.

2.1 Inside Sales

Inside sales ou vendas internas é um modelo de vendas que funciona internamente na empresa. Este modelo se utiliza de e-mail, telefone, redes sociais e videoconferências para se comunicar com os possíveis clientes de forma mais barata e eficiente do que uma visita presencial. Dependendo do tipo de produto que esta sendo vendido não é viável realizar o Inside Sales, mas em uma empresa de software como a Zygo consegue utilizar todo o potencial deste sistema.

Os contatos dos leads são enviados para o time de Inside Sales através de uma ferramenta interna chamada de SumForce, nesta plataforma o lead vai ser tratado em duas etapas, primeiro este lead será enriquecido e aquecido pelos pré-vendedores e sem seguida será enviado para os vendedores para realizar a venda.

2.1.a Pré-vendedores

As funções dos pré-vendedores são de filtrar os leads que não possuem capacidades para se tornarem nossos clientes e criar interesse e engajamento desses leads no nosso produto para que cheguem mais interessados no vendedor. Os times de pré-vendedores são formados dos BDR que tem a função de conseguir o contato da pessoa que toma as decisões no estabelecimento, dos SDR que realizam uma ligação para o decisor deste estabelecimento, fazem o filtro e o engajamento, e do Quarterback que é encarregado de gerenciar os horários de reunião dos vendedores e garantir que o lead apareça na reunião.

O projeto deste documento, o preditor de No-shows, teria um impacto direto no trabalho dos Quarterbacks, que poderiam direcionar para os SDR melhorarem o engajamento dos leads antes da reunião.

2.1.b SumForce

O Sumforce é um uma plataforma web onde os pré-vendedores e vendedores podem organizar os leads, adicionar informações obtidas e até mesmo criar categorizações para diferentes tipos de Leads. Essa ferramenta é acessada por todo o time de Inside Sales permitindo que todas as informações estejam disponíveis em tempo real. Contudo pelo fato da ferramenta ter como objetivo melhorar o trabalho do time de Inside Sales muitas funções foram criadas perante necessidades, sem serem pensadas formas otimizada de mensurar estas classificações. Dentro desta plataforma existem informações redundantes e classificações diferentes para um mesmo propósito, por este motivo para realizar algumas pesquisas várias tabelas diferentes são necessárias e tratamentos manuais precisam ser aplicados devidas as repetições.

Processos para padronizar as atividades do Sumforce estão sendo realizados, mas independente destes dados a utilização dos áudios dentro da empresa é algo novo que interessa vários departamentos. Estes audios estão todos armazenados dentro do Victory, o banco de dados do Sumforce.

2.2 Casos reais

Os pré-vendedores acabam desenvolvendo uma intuição em relação aos leads, dependendo da forma como a ligação termina eles já preveem se a reunião será realizada ou se terá um no-show, porém não existe nenhuma métrica na empresa que possa mensurar essa intuição. Conversando com um dos times de pré-vendedores da empresa foram levantados dois exemplos desta intuição.

2.3 Soul Jazz Burguer

Este lead é uma hamburgueria de Belo Horizonte, o próprio SDR conseguiu o contato deste lead em uma ligação para o estabelecimento, depois disso em uma ligação de 13 minutos e 30 segundos conseguiu agendar uma reunião. O lead falou pouco durante toda a ligação mas demonstrou interesse no produto, e por mais que as respostas do lead tenham sido poucas a SDR responsável passou para o Quarterback que este seria uma reunião bastante provável de ocorrer. E assim como a intuição do SDR previu a reunião foi realizada.

Pipe: Pipe de Vendas Etapa: Demonstração SDR: Sara Faria Vendedor: Pedro Bernardi

Visão Geral

Nome da empresa: Soul Jazz Burger
Telefone Primário: 3125281232
Endereço: Rua Noraldino de Lima 387 - Aeroporto Belo Horizonte - MG 31270-650
Cidade: Belo Horizonte
Estado: MG
E-Mail: pessoal@souljazz.com.br
Site: http://souljazzburger.negocio.site/
Criado em: 26 de Fevereiro, 19:42 (5 meses atrás)

Contatos

Maira:
pessoal@souljazz.com.br 31991545681
Bruno:
financeiro@souljazz.com.br 31984713165

Informações

Origem: Outbound
Sub-Origem: 3068HE
Fan page: https://www.facebook.com/souljazzburger/
Likes: 1224
Segmento: Burger Restaurant
Conta: Conta B
Marketing: Sim
Quantidade de Curtidas TripAdvisor: 47
score: 90127.2
Horário de funcionamento: Noite

Timeline

- Foi feita uma interação do tipo "" por Jorge Cruz em 04 de Junho, 16:42 (aproximadamente 1 mês)
- Foi feita uma interação do tipo "" por Jorge Cruz em 04 de Junho, 16:24 (aproximadamente 1 mês)
- Foi feita uma interação do tipo "" por Jorge Cruz em 04 de Junho, 16:23 (aproximadamente 1 mês)
- Foi feita uma interação do tipo "Salesman - 1 Demo". por Pedro Bernardi em 01 de Abril, 10:57 (3 meses)
- Aplicado o scoring **Feedback de Reunião 1.0** com o score 20. por Pedro Bernardi em 01 de Abril, 10:56 (3 meses)

Figura 2.1: Visão geral do Lead Soul Burguer no SumForce

2.4 Delicia Natural

Este lead é um restaurante buffet de Porto Alegre, onde o BDR conseguiu o contato do decisor pelo Facebook e o SDR realizou o agendamento em um ligação de 16 min e 39 segundos. O lead foi muito responsivo e falou por muito mais tempo que o SDR, contudo ao finalizar o agendamento o SDR comunicou ao Quarterback que esta reunião provavelmente não seria realizada. Como o SDR previu houve o No-show e conversando melhor com o SDR ele explicou que o interesse e engajamento do lead era muito alto durante toda a ligação, mas durante o processo de marcar a reunião ele foi capaz de perceber menos interesse do Lead.

The image shows a screenshot of the SumForce CRM interface. On the left, there are three columns of information: 'Pipe: Pipe Padrão Etapa: No-Show (reagendar)', 'SDR: Mateus Santana', and 'Vendedor: Não definido'. Below these are sections for 'Visão Geral', 'Contatos', and 'Informações'. The 'Visão Geral' section includes company name, phone, address, city, state, email, and website. The 'Contatos' section lists 'Vanessa' as a contact with email and phone. The 'Informações' section includes origin, sub-origin, fan page, likes, segment, and category. On the right, a 'Timeline' window shows a list of interactions: 'Foi feita uma interação do tipo "Ninguém atendeu"', 'Foi feita uma interação do tipo "SDR - No-show!"', 'Aplicado o scoring No-show #j com o score 102', 'Foi feita uma interação do tipo "Enviei um e-mail"', and 'Foi feita uma interação do tipo "Não falei com o decisor"'. Each interaction entry includes the date and time, and the user who performed the action.

Figura 2.2: Visão geral do Lead Delicia Natural no SumForce

2.5 Sistematização do Problema

O projeto deste documento foi planejado seguindo os passos necessários para uma boa execução, esses passos serão melhor detalhados e acompanhados de seus embasamentos teóricos no capítulo seguinte. Por fim será mostrado como é deve funcionar a estrutura final do projeto.

Primeiramente foram selecionados e filtrados os áudios que seriam utilizados para o treinamento do sistema de aprendizado de máquina. Estes dados possuem um registro no banco da dados da empresa, porém os arquivos são armazenados em um

banco de dados na nuvem da AWS. Obtidos os dados foi iniciado o pré-processamento, a separação dos canais do áudio e a filtragem dos sons de fundo das ligações.

Finalmente após todos esses processos será utilizada a biblioteca de Python LibROSA para extrair as características mais relevantes do áudio, estas características serão armazenadas para servirem de entrada para o sistema de aprendizado de máquina. No sistema de aprendizado de máquina parte dos dados serão utilizado para o treinamento da maquina, e depois por fim com o conjunto final de dados será feita a avaliação da acuratividade do sistema. Não faz parte deste projeto a implementação deste sistema para os pré-vendas pois precisaria ser integrado ao SumForce e esta etapa estaria fora do meu escopo de trabalho dentro da empresa.

Em resumo a Figura 2.3 mostra o sistema final, como entrada do sistema temos os áudios das ligações, e como saída temos a indicação de possível No-Show ou não

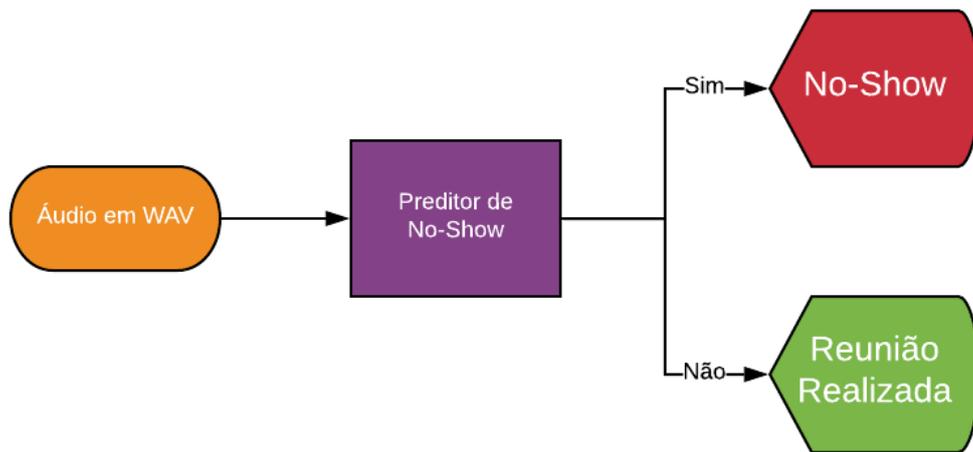


Figura 2.3: Sistema Final

Capítulo 3

Embasamento Teórico e Ferramentas

Já nas primeiras fases de planejamento do projeto foi possível compreender a complexidade do problema e que muitos projetos semelhantes a este utilizam de ferramentas muito avançadas de tratamento de sinais. Além disso uma das escolhas de direcionamento do projeto foi priorizar ferramentas de código livre.

Visto estas escolhas e descobertas, usamos esse conhecimento como parâmetros para escolher as ferramentas que serão apresentadas a seguir. Em seguida, serão apresentados os estudos relevantes que tornaram capaz a compreensão deste projeto.

3.1 Ferramentas

Para desenvolver esse projeto foram utilizadas duas linguagens de programação Python e SQL, mais especificamente PostgreSQL [1] nas etapas iniciais do projeto. Utilizando SQL pude selecionar e filtrar as ligações mais relevantes para o estudo, em resumo existiam 3 tipos de ligações, as de agendamento que geravam No-Shows, as de agendamento que geravam reuniões realizadas e os reagendamentos. Os reagendamentos não foram utilizados para este trabalho, além disso existiam mais dados dentro do sistema que caracterizam os leads, porém estes dados são utilizados na filtragem da área de marketing antes de chegarem aos pré-vendedores o que poderia enviar os estudos.

Para todo o resto do projeto foi utilizada apenas Python, inclusive para interagir com outras ferramentas como a base de dados na Nuvem da AWS. As ferramentas mais relevantes do projeto foram:

- SciPy
- LibROSA
- Scikit-learn

3.1.a SciPy

SciPy é um ecossistema de código aberto baseado em Python, dentro deste ecossistema existem diversas ferramentas, entre elas Pandas, NumPy e SciPy. A biblioteca SciPy é feita para aplicações de rotinas numéricas e dentro desta ferramenta existem métodos de integração, interpolação, otimização, álgebra linear e estatística. Além dessas rotinas o SciPy também possui uma série de funções para tratamento de sinais. [2]

Pandas é uma biblioteca para manipulação e análise de dados conhecida pela sua alta performance. Utilizando esta ferramenta é possível importar vários tipos diferentes de dados como planilhas no Excel, arquivos, json e até mesmo tabelas em bancos de dados relacionais. A alta performance da biblioteca não vem apenas da velocidade de execução, dados com volume muito grande de informações podem ser manipulados por esta biblioteca, além de ter ferramentas para lidar com planilhas não higienizadas, como por exemplo arquivos CSV com vírgulas em campos "errados". Depois de importado os dados ficam armazenados em um Data Frame, este formato é similar a uma tabela mas é formado de vetores.

Através desses vetores o acesso a linhas, colunas ou até mesmo células específicas fica simplificado, funcionando de forma similar ao recurso dicionário que existe no Python. No Pandas também é possível fazer diversos tipos de análises como, cálculos estatísticos e geração de gráficos, e mesmo que não fosse possível realizar essas atividades no Pandas existem muitas bibliotecas de análise de dados como Seaborn e Matplotlib que utilizam a estrutura de Data Frame do Pandas como padrão de entrada.

Como explicado, a estrutura dos Data Frames do Pandas são formados de vetores, contudo na linguagem Python não existe vetores nativos, a forma mais comum de

gerar conjuntos é através de listas. Por isso toda a estrutura do Pandas é construída em cima de uma outra biblioteca chamada de Numpy.[3]

A biblioteca Numpy é a biblioteca mais utilizada como vetores dentro do Python, possui várias funções diferentes e também é conhecida por ter alta performance. Através do NumPy é possível gerar os mais variados tipo de vetores e até mesmo utilizar funções de álgebra linear. Por fim todas as ferramentas seguintes utilizam de NumPy para manipular seus dados, como entrada ou saída de suas funções.[4]

3.1.b LibROSA

LibROSA é uma biblioteca para análise de áudios e sinais musicais. O foco dessa biblioteca é mais voltado a retirar características do áudio que se relacionam com análises de músicas, contudo dentro das comunidades de análise de áudio utilizando Python esta biblioteca é amplamente usada para os mais variados propósitos. Esta biblioteca também tem o código aberto e sua base de cálculos funciona usando SciPy e NumPy como base.

As suas funções tentam estabelecer uma normalização de variáveis, valores padrões e parâmetros para estas características de som, visto que é uma área muito nova muitas dessas características são nomeadas de diferentes formas em variados artigos. Além disso a comunidade de Music Information Retrieval (MIR) e outras comunidades envolvidas em processamento digital de sinais utilizam majoritariamente o Matlab como ferramenta de análise, e por esse motivo o LibROSA tem uma estrutura que foca em ser fácil de ser compreendida e de exportar informações do próprio Matlab.[5]

3.1.c Scikit-learn

Scikit-learn é uma biblioteca em Python dedicada a funções de aprendizado de máquinas, dentro da biblioteca existem várias opções de algoritmos de classificação, regressão e clusterização. A base dessa biblioteca também é feita em cima de NumPy e SciPy. Para esse projeto foi escolhido utilizar o algoritmo de classificação de árvore de decisões.[6]

3.1.d PyAudioAnalysis

Esta biblioteca não está listada acima pois não foi utilizada neste projeto, porém é uma biblioteca relevante à utilização de Python para análise digital de sinais de áudio. O PyAudioAnalysis é uma biblioteca focada em análise de áudios, e diferente do LibROSA não tem o foco dedicado a sinais de músicas. Dentro desta biblioteca também já tem integrado sistemas de classificação por aprendizado de máquina. Contudo as funções de entrada não parecem ser muito estáveis e mesmo tentando converter os áudios para diferentes formatos simplificados, o que já prejudicaria a qualidade das conclusões que seriam tomadas, não fui capaz de abrir arquivos de áudio nesta ferramenta. Isso somado ao fato que a comunidade do LibROSA é mais ativa e atualizada contribui com a decisão de utilizar apenas a biblioteca LibROSA.[7]

Além destas ferramentas foram utilizadas outras bibliotecas que serão explicadas no capítulo seguinte junto com o propósito pela qual elas foram utilizadas, as ferramentas e as respectivas referências são listadas a seguir

- Requests[8]
- Wave[9]
- Soundfile[10]

3.2 Embasamento Teórico

Sinais de áudio podem ser analisados de diversas formas diferentes, e cada forma de análise pode ter variadas funções como classificação de tipos diferentes de som, compreensão de sentimentos, transcrição de diálogos entre outras. A pesquisa realizada para realizar este projeto teve duas vertentes principais que poderiam solucionar o problema de detecção de No-Shows. A primeira abordagem foi transcrever o áudio e utilizar ferramentas de aprendizado de máquina para analisar as palavras para tentar detectar sentimentos. A segunda abordagem foi tentar detectar os sentimentos diretamente do áudio. Ambos os estudos foram fundamentais para compreender as ferramentas da biblioteca LibROSA e tomar a decisão de quais características possíveis de se retirar dos áudios seriam utilizadas. Outra fonte interessante utilizada para compreender melhor o No-Show foi um outro projeto realizado para avaliar a qualidade das ligações usando a metodologia SPIN Selling. Além disso a base de processamento

digital de áudio foi estudada para realizar o pre-processamento devido das ligações e, por fim, foi estudado o funcionamento da técnica de aprendizado de máquina das árvores de decisão.

Dentro da Zygo Tecnologia já foram utilizados softwares de transcrição de áudio, a intenção inicial não era especificamente combater o No-Show, mas sim ter um registro de mais rápidos acessos ao conteúdo das ligações e assim utilizar esse registro para melhorar cada vez mais a abordagem dos pré-vendedores. No começo do projeto foi recomendado a utilização dessas ferramentas para transcrever os áudios e assim usar os resultados da transcrição com ferramentas de aprendizado de máquina, contudo os software que já tinham sido utilizados na empresa ou tinham sido descontinuados, ou eram ferramentas pagas.

Assim foi estabelecido que a empresa disponibilizaria de uma renda para contratar um software de transcrição, mas se esta fosse a melhor opção para realizar a tarefa.

3.2.a Softwares de Transcrição de Áudio

Como início dos estudos em transcrição de áudio eu utilizei um projeto de um colega de trabalho, o `sgbcloud`[11], este projeto possuía muitas referências, mas a principal dela foi a biblioteca em Python `SpeechRecognition`[12]. Esta biblioteca é um intermediário para utilizar APIs de terceiros para realizar o reconhecimento de voz e transcrição. Entre as APIs disponíveis temos `CMUSphinx`[13], `Google Cloud Speech-to-Text`[14], `IBM Watson Speech to Text`[15], entre outros. Destes APIs, poucos possuíam código aberto, e estes de código aberto não eram capazes de trabalhar com Português. Em seguida foram estudados os APIs capazes de transcrever os áudios em português, sua acuratividade e também os seus preços. Um outro API foi encontrado também chamado `Retextoar`[16] pela empresa `Dígito Tecnologia` que tem sua sede aqui em Florianópolis.

O fato destes aplicativos não serem de código aberto impedia com que eu pudesse avaliar o tamanho da perda de informação da transcrição e conseqüentemente prejudicar o trabalho de detecção do No-Show. No que pode ser explorado nas bibliotecas de código aberto, os resultados eram obtidos perante a análise de algumas características que também são usadas na Análise de Sentimento, e os exemplos demonstrados nas documentações facilitaram a compreensão dessas características.

3.2.b Análise de Sentimentos

A análise de sentimentos é uma área de estudos que não possui muitos consensos no meio científico, cada pesquisa busca compreender quais características são mais relevantes para identificar os sentimentos. Um consenso pode ser definido que um sentimento só pode ser detectado por uma composição de diferentes características de um áudio.

A base de dados utilizadas nesta análise de sentimentos é mais uma complexidade deste tipo de pesquisa, pois a definição de que um arquivo de áudio representa o sentimento de felicidade ou raiva é subjetivo e pode ter interpretações diferentes de pessoa para pessoa. Considerando o propósito desse trabalho não é possível estabelecer um sentimento que esteja diretamente relacionado a incidência de No-Shows, em outras palavras, faz parte do estudo compreender os dados e depois de encontrado o caminho utilizado para classificar os áudios fazer comparações com outros estudos e compreender se existem mesmo sentimentos relacionados a esses resultados. O foco dos estudos deste tema foi compreender quais características são mais comumente usadas para não utilizar dados irrelevantes na classificação e ao mesmo tempo facilitar comparações futuras.

Um arquivo de áudio é formado de uma composição de sinais dentro do tempo, a forma padrão que um sinal analógico é representado digitalmente, então no momento de se calcular as características desse áudio isso pode ser feito separando o áudio em amostras ou globalmente. A maioria dos pesquisadores concordam que a análise global das características trazem melhores resultados para a classificação de sentimentos, mas uma vantagem muito relevante para um estudo simplificado como este é que a análise global possui menos variáveis para serem consideradas e diminui o trabalho de processamento desses dados e de suas análises. Em contraponto a análise das características globais é mais eficiente para sentimentos de alta empolgação como raiva, medo e alegria, enquanto sentimentos de menor empolgação como tristeza são melhores detectados considerando a análise por amostragens.

Além da análise global ou local existem vários tipos diferentes de características, uma classificação utilizada é separar as características em características contínuas, características qualitativas, características espectrais e características baseadas no operador de energia de Teager. [17]

- **Características Contínuas:** Estas características são naturais do sinal sem qualquer tipo de transformação para análise, como por exemplo o passo do sinal de áudio, o formato do sinal, a energia, a frequência natural, a duração e vários

outros. Além disso as características de prosódia também são consideradas contínuas, estas características são utilizadas frequentemente em estudos que envolvam análise de sentimentos.

- **Características Qualitativas:** Estas características são relacionadas a qualidade da voz, como por exemplo passo da voz, amplitude da voz, frases, fonemas e estruturas temporais. Acredita-se que estas características sejam de grande valor a análise, contudo pouco se sabe sobre esse tipo de características pois não existe um consenso entre as classificações usadas em relação a esses dados, e ao mesmo tempo, são características com alta complexidade para automatizar sua obtenção.
- **Características Espectrais:** Estas características são complementares as características baseadas em tempo, representam o sinal na frequência. Estes dados podem ser obtidos através de transformações do sinal de áudio como coeficientes lineares de predição(LPC), método de coerência a curto tempo(SMC). Estes são alguns métodos de análise de frequência, mas para explorar melhor estes dados os espectros estimados devem passar por um banco de filtros passa faixa e da saída desses filtros são tiradas as características. Além disso como as percepções humanas de passo não são lineares estes dados precisam ser transformados para uma escala não linear como a Mel-Frequency.
- **Características baseadas no Operador de Energia de Teager:** De acordo com os estudos feitos por Teager, a fala é produzida por um fluxo não-linear no sistema vocal e a tensão nos músculos do sistema vocal geram o som da voz. Seguindo esse raciocínio são necessárias características não lineares para ser analisar corretamente os sentimentos em um áudio de fala. O Teager-energy-operator(TEO) foi desenvolvido por Teager e Kaiser com a hipótese de que a escuta era a capacidade de detectar energia. Assim seguindo este modelo matemático muitas características são possíveis de serem extraídas de um áudio e estas características trazem bons resultados na análise de sentimentos.

A escala de Mel é uma função quasi-logarítmica da frequência acústica projetada de tal forma que os intervalos de tom perceptualmente semelhantes (por exemplo, oitavas) pareçam iguais em largura em toda a faixa de audição completa. Como a definição da escala de mel é condicionada por um número finito de experimentos psicoacústico subjetivos, várias implementações coexistem na literatura de processamento de sinais de áudio [18]. Por padrão, a librosa replica o comportamento do MATLAB Auditory Toolbox of Slaney [19]. De acordo com essa implementação padrão, a conversão de Hertz para mel é linear abaixo de 1 kHz e logarítmica acima de 1 kHz.

Estas são as características mais frequentes nos estudos de análise de sentimentos que podem ser retiradas do arquivo de áudio, contudo para uma análise ainda mais precisa outras características podem ser usadas como conceitos de linguística para analisar a escolha de vocabulário utilizado pelo falante ou até mesmo um vídeo que contenha as expressões faciais do interlocutor.

Por mais que todos estes dados sejam gerados é a composição deles que geram uma análise eficiente, existem vários modelos e várias composições diferentes, cada uma com suas vantagens e desvantagens.[20] Para este projeto serão escolhidos as características mais próximas de uma análise de sentimentos levando em consideração as limitações da biblioteca LibROSA, e futuramente as hipóteses levantadas em outro trabalho realizado dentro da Zygo Tecnologia envolvendo SPIN Selling.

3.2.c SPIN Selling

SPIN Selling é uma metodologia que orienta quais são as perguntas que devem ser feitas para aumentar as taxas de vendas, e dentro da Zygo já foi feita uma pesquisa para compreender a frequência com que esse tipo de pergunta é feita e qual o impacto delas no resultado final. As perguntas são baseadas em 4 palavras que formam os conceitos principais dessa metodologia, Situação, Problema, Implicação e Necessidade como ilustrado na Figura 3.1. A pesquisa foi feita por Guilherme Carneiro e ele manualmente ouviu uma amostragem de 52 ligações de pré-vendas e anotou as frequências com que eram realizadas as perguntas mais relevantes e a incidência delas relativo ao tempo da ligação. Estas informações foram comparadas com a ocorrência ou não do agendamento e além disso a realização ou não desta reunião.

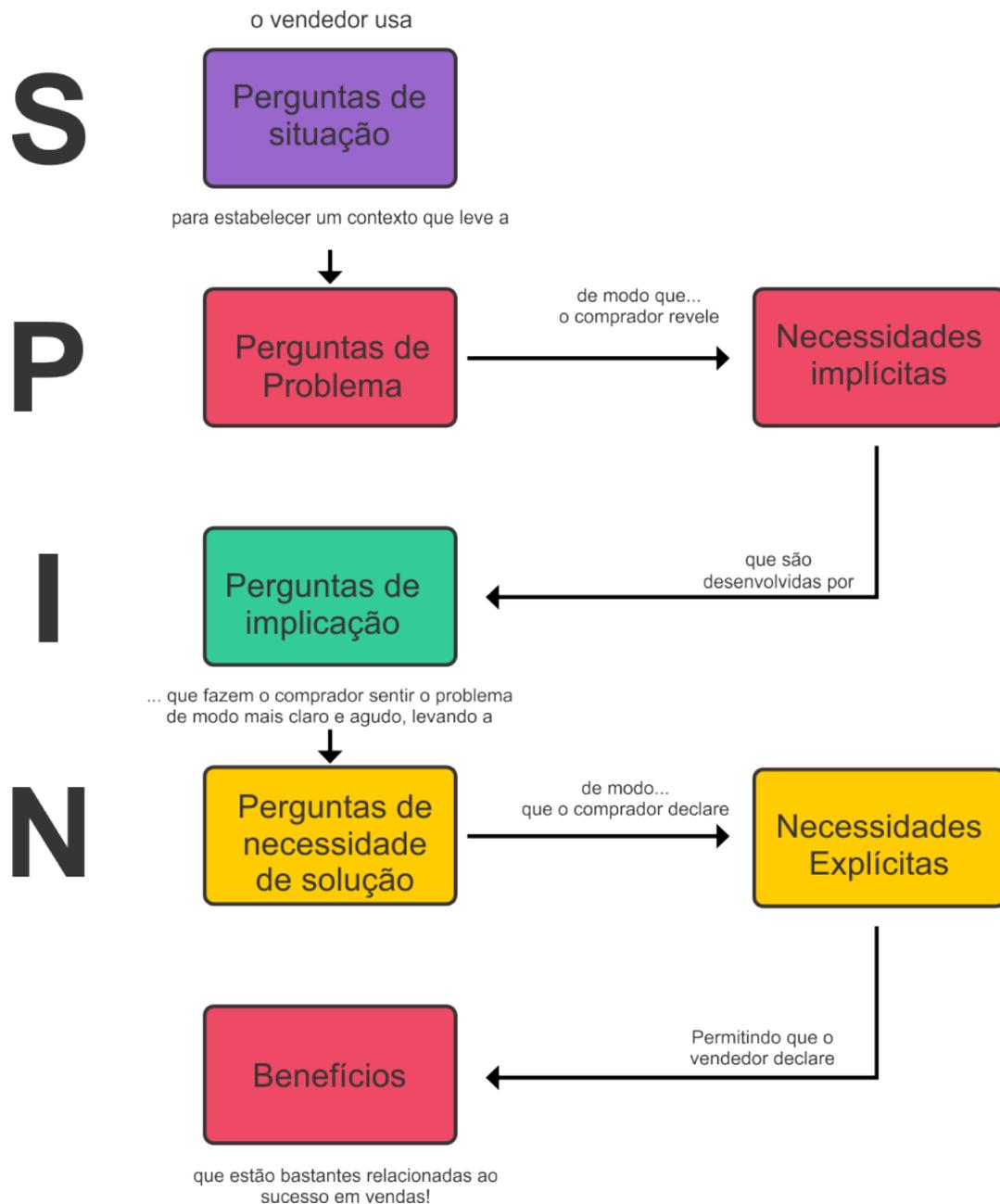


Figura 3.1: Linha Temporal da Zygo

Essa pesquisa trouxe padrões interessantes como por exemplo o simples fato de perguntar se o lead tem interesse em marcar uma reunião o torna 30% mais propenso a comparecer a reunião. Todos estes dados contribuíram para escolha das características

e também como seria o melhor pre-processamento dos dados.

3.2.d Pré-Processamento de Dados

Antes de extrair as características desejadas é preciso ter certeza que os dados estão tratados da melhor forma possível, esse tratamento envolve não só modelar o arquivo de áudio para a melhor qualidade possível, mas também certificar-se que os dados selecionados são relevantes para o estudo.

Para isso foram selecionados apenas os arquivos de áudio com duração superior a 3 minutos, considerando que um agendamento não seria capaz de ser realizado neste tempo. As ligações de agendamento tem em média 15 minutos atualmente, porém dentro do 1 ano de informações adquiridas e usadas nessa pesquisa a abordagem de um pré-vendedor já mudou inúmeras vezes, e não era incomum agendamentos que aconteciam em 5 minutos. Exatamente por essas mudanças frequentes e considerando que as ligações ainda terão mais mudanças futuramente a formação das bases de treinamento e de teste precisa ter arquivos de diferentes épocas da empresa.

Além disso foi garantido que as ligações não são resultados de reagendamento, pois esta seria uma complexidade a mais ao projeto. Esta garantia vem dos registros feitos pelo time de venda através do SumForce.

Por fim os arquivos de áudio também passaram por um tratamento. Foram separados os canais que existem no arquivo de áudio, sendo um dos canais o áudio do nosso pre-vendas e o outro canal sendo o lead. Além disso para alguns tipos de análise a parcela do áudio de silencio pode ser retirada para acelerar a análise, contudo neste caso o silencio também pode conter informações importantes sobre a os sentimentos.[17][21]

3.3 Escolha das Características do LibROSA

Depois de todo o trabalho de pesquisa feito para compreender a melhor maneira de gerar características interessantes o próximo passo foi explorar o LibROSA e escolher dentre os vários recursos da biblioteca aqueles que faziam mais sentido para esta análise. No ponto de vista dos pré-vendedores a capacidade de prever um No-Show esta diretamente envolvida com a compreensão de sentimentos, mesmo assim eles não

conseguem definir exatamente que sentimentos seriam esses, por esse motivo a escolha das características a serem usadas não tem a intenção apenas de serem as entradas mais eficientes para um sistema de aprendizado de máquina, mas também espera-se compreender melhor o que leva a ocorrência de um No-Show.

As seguintes características foram escolhidas:

- **Tempo de duração do áudio:** Esta característica é a mais simples entre as escolhidas, o tempo de duração do áudio é uma característica contínua e pode ser um indicativo do engajamento do lead com a ligação, além disso pode demonstrar o quão rápido o pre-vendedor falou. Segundo a pesquisa realizada sobre o SPIN Selling pré-vendedores que falam muito rápido tem uma maior disposição a receber mais No-Shows.
- **Ritmo do áudio:** Outra característica contínua, o ritmo do áudio foi a adaptação para o LibROSA do passo do áudio. Esta medida é um valor de tempo de ritmo representado por quantas batidas se tem por minuto.
- **Potencia:** A última característica contínua, a potência do áudio é a adaptação do LibROSA para a energia do áudio, e para devida compreensão deste valor sua resposta esta escalada em decibéis. Pelo LibROSA o cálculo desta potencia é feito entre intervalos calculados automaticamente pelo LibROSA para identificar principais mudanças de energia, porém como queremos um ponto de vista global da ligação são utilizadas a média desta potencia e o desvio padrão.
- **Mel-Spectrogram:** O espectrograma mel é a função pura do espectro na escala mel do áudio, o vetor gerado passou pelo banco de filtros da escala mel e gerou um vetor de potencias. É usada a média e desvio padrão deste vetor.

Capítulo 4

Aplicação à Predição de *Shows/No-Shows* e Resultados

Neste capítulo será descrita toda a aplicação dos conceitos estudados e apresentados no capítulo anterior, além de algumas etapas mais operacionais ou cujo desenvolvimento não precisou de uma grande base teórica. Em seguida serão apresentados os resultados obtidos.

4.1 Aplicação à Predição de *No-Shows*

O projeto sofreu modificações no planejamento inicial conforme foi se construindo um conceito mais sólido de como seriam aplicadas as ferramentas e quais eram os objetivos de cada processo. Independente disso o primeiro passo de qualquer trabalho de análise de dados começa com a obtenção de dados qualificados para o estudo.

4.1.a PostgreSQL e o SumForce

A Zygo Tecnologia é uma empresa que tem alguns pontos estruturais para consolidar, mas isso não impede a empresa de se preparar para o futuro. A principal aposta para o crescimento depois de estruturados é um gestão direcionada pela análise de dados, e para serem capazes de atingir este objetivo, muitas informações são armazenadas.

É através do SumForce que muitas destas informações são armazenadas, as principais são as informações referentes aos processos da área de *Inside Sales* e da área de Sucesso do Cliente, já que estas duas áreas tem a maior parte dos seus processos qualitativos e precisam ser armazenadas manualmente pelos funcionários que estão realizando estas tarefas. Todas essas informações vão para os bancos de dados em *PostgreSQL*.

Para obter os dados de ligações corretas foram feitas duas *querrys* no *PostgreSQL*, uma para identificar todas as ligações que geraram Reuniões Realizadas e todas as ligações que geraram *No-Shows*. As duas *querrys* tinham estruturas diferentes. A tabela *Interactions* é a tabela que define as atividades que os pré-vendas realizam com os *leads*, da qual foram selecionados todos os *leads* que tiveram um agendamento de reunião. Dentro da tabela *Calls* existem as informações desejadas das ligações e também uma chave-estrangeira que conecta as ligações aos agendamentos. O próximo passo é identificar destas ligações quais foram os resultados, para isso outra chave-estrangeira que existe tanto na tabela *Interactions* quando na tabela *Calls* foi usada. A esta chave conecta estas tabelas a tabela *Leads*, mas esta conexão não foi necessária, o intuito era saber quais destes *leads* tiveram a interação de *No-Show* ou Reunião Realizada. Com as duas listas geradas individualmente o único processo que faltou foi o de tirar destas listas os *leads* que foram reagendados para não levar a uma complexidade a mais ao estudo. Para retirar estes *leads* foi feita uma terceira lista dos *leads* que possuíam tanto reunião realizada quanto *No-Show* e esta lista foi retirada das outras duas.

4.1.b Amazon S3

Com as listas de ligações o próximo passo foi baixar estes arquivos de áudio do sistema de armazenamento da AWS. *Amazon Simple Storage Service*(Amazon S3) é um sistema de armazenamento de objetos na nuvem, através dele a Zygo armazena vários tipos de arquivos que possam ser relevantes para empresa no futuro. É neste repositório que estão armazenados os arquivos de áudio da empresa, e para retirar eles inicialmente foi pensado em utilizar o próprio sistema de requisição da Amazon S3. O sistema da Amazon S3 separa os arquivos em *Bucket*. Utilizando esse sistema teríamos que encontrar os *buckets* relevantes e depois separar os arquivos relevantes. Porém especificamente os *buckets* relacionados aos arquivos de áudio estavam públicos e não precisariam de certificações para serem acessados, assim utilizando requisições para as páginas web referentes a esses arquivos, foi possível fazer o download em escala apenas dos arquivos relevantes.

Para realizar estas requisições foi feito um *script* em Python com a biblioteca Requests, e para fazer esta escalar estas requisições foi utilizada a biblioteca pandas.

4.1.c Separação de Canais

Todos os arquivos obtidos estão no formato WAV, este formato é foi desenvolvido pela Microsoft e IBM para armazenar digitalmente sinais de áudio. Este formato mede a intensidade(energia) dos sinais sonoros em decibel e modula esta medida usando o método Modulação por código de pulso linear(LPCM). Este método que é usado para representar digitalmente sinais analógicos é padrão para sinais de áudio e é composto de duas propriedades básicas, o período de amostragem que indica a frequência com que o áudio é medido e a profundidade de bits que indica o numero de bits máximos que serão usados para representar uma amostra.

Os arquivos de áudio armazenados pela Zygo Tecnologia armazenam as falas dos pré-vendedores em um canal diferente a fala dos *leads*. Dessa forma caso seja necessário é simples fazer a separação dos canais para analisar separadamente os áudios.

Estes áudios foram gravados em 32 bits, uma ótima qualidade, mas os programas mais comuns de mixagem são capazes de trabalhar com áudios mais simples em 16 bits. Desta forma antes de fazer a separação foi preciso converter estes áudios. Utilizando a biblioteca Soundfile do Python foram convertidos em massa os áudios para 16 bits, e em seguida usando a biblioteca Wave foram separados os canais.

Os áudios originais foram mantidos, e novos áudios com os canais separados foram gerados. Esta preocupação está em não apenas garantir que não se perca dados, mas também pelo fato de não ser totalmente confiável a qualidade destes áudios convertidos uma vez que eles tiveram literalmente sua qualidade reduzida.

4.1.d Retirada das Características

Com os áudios processados e prontos para análise foi iniciado o processo de extrair as características utilizando o LibROSA. Dentro de todos os processos explicados anteriormente as tarefas em escala só eram aplicadas depois de serem feitos testes em alguns áudios de amostragem. Nesta etapa os testes com as características trouxe mais clareza sobre como funcionam os seus processos e quais os resultados esperados.

Ritmo do Áudio

Para analisar o ritmo do áudio foram utilizados os conceitos *Onset Strength* (Força onset) e *Beat* (Batida). *Onset Strength* é um agrupamento dos sinais sonoros que possuem maior energia depois de períodos com baixa energia, ou seja, picos de som entre espaços de silêncio. *Onset Strength* é utilizado pela audição humana para organizar e integrar sons que originam de uma mesma fonte permitindo assim sua compreensão. Depois de determinados os *Onset Strength* é estimado um período entre estes picos, e os valores de energia contidos dentro deste período é chamado de Batida.

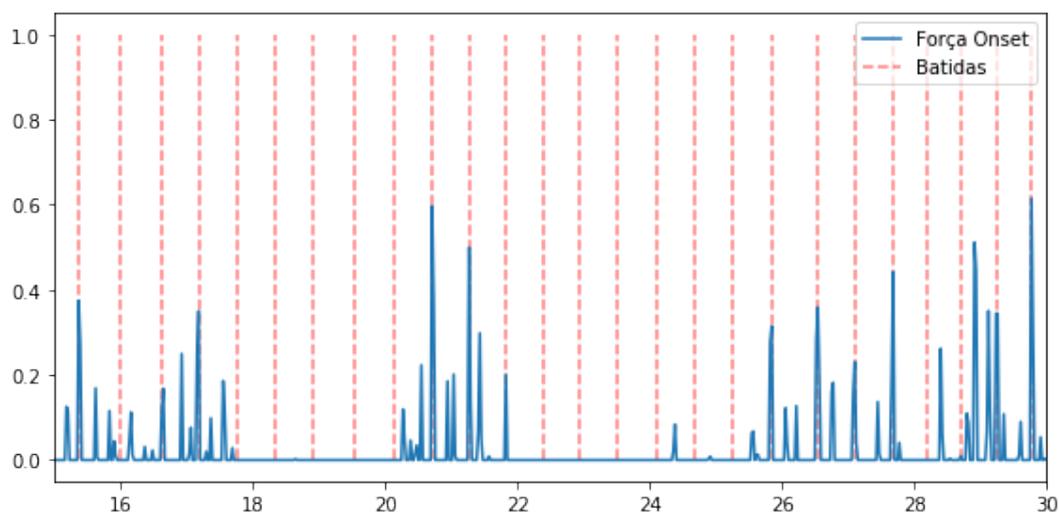


Figura 4.1: Gráfico de ritmo

A Figura 4.1 representa o gráfico de ritmo de um dos sinais de áudio, o eixo X representa o tempo enquanto o eixo Y representa a energia do sinal em decibéis, e pode ser observado que o padrão de início dos sinais *onset* não seguem perfeitamente a demarcação do ritmo calculado pelo LibROSA, mas os sinais mais relevantes estão contidos nesse período.

Onset Strength é calculado em um tempo t como $\max(0, S[f, t+1] - S[f, t])$ onde S é o *Mel-Spectrogram* da potência em escala logarítmica e ref é S depois da filtragem local máxima ao longo da frequência. Se o áudio possuir mais de um canal é tirada a média do *Onset Strength* dos dois canais. Este processo é realizado utilizando a função *beat_track* do LibROSA.

Potencia

Para observar a potencia do arquivo de áudio e demonstrar a importância da mudança de escala para observação foram feitos dois gráficos que aparecem na Figura 4.2.

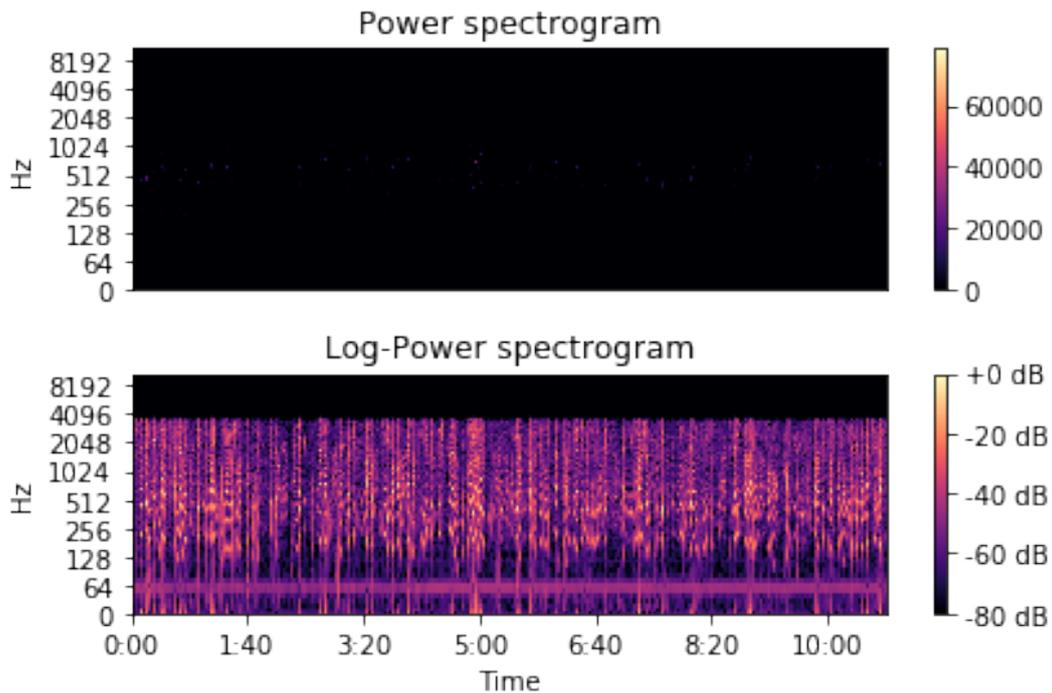


Figura 4.2: Gráfico de potencia

Em ambos os gráficos podemos observar a potência sendo representada pelas cores em função do tempo e da frequência. Para o gráfico sem mudança de escala pouco pode se inferir sobre a potencia do áudio, isso se deve a natureza logarítmica desta característica, e por ter essa natureza é a melhor forma de representar estes valores é com a mudança de escala como mostrado no gráfico de decibéis.

O processo para obter o sinal de frequência é realizar a transformada de Fourier de Curto tempo no sinal de energia padrão do áudio e depois tirar o módulo deste resultado. Este módulo então elevado ao quadrado representa a potência e então é convertido para a escala logarítmica. Este atributo é obtido utilizando a função *stft* do LibROSA para transformada de Furier, a função *abs* do NumPy para obter

o módulo e por fim a função *power_to_db* do LibROSA para obter o sinal na escala logarítmica.

Mel-Spectrogram

A frequência Mel tem um tratamento dedicado para retirar ruídos que possam interferir nessa análise de áudio e apresenta a escala mais ajustada para a natureza logarítmica e não linear das características acústicas. O sinal de potência depois de passar pelo banco de filtros mel é representado pela Figura 4.3

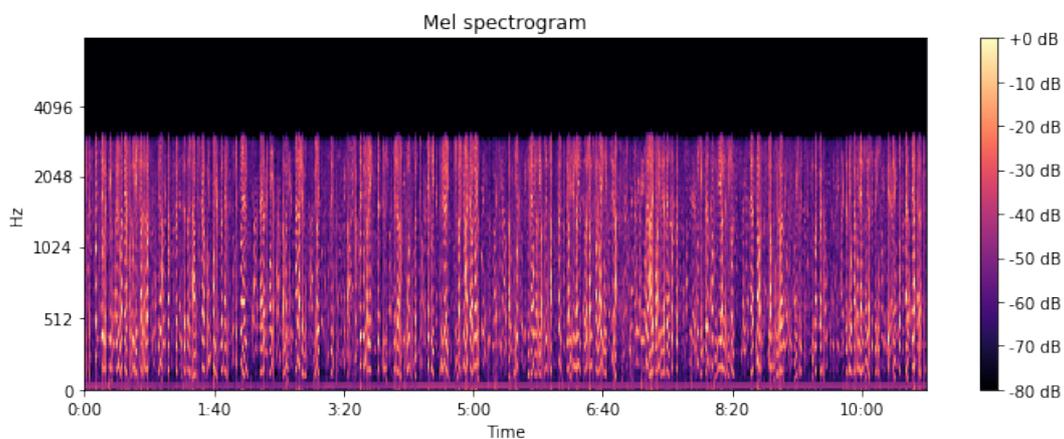


Figura 4.3: Espectrograma na escala Mel

A Figura 4.3 possui, indicada pelas cores, a magnitude da potencia do sinal na escala mel em relação à frequência do sistema e ao tempo. Através do *Mel-Spectrogram* é possível fazer várias inferências sobre o sinal como o ritmo, a qualidade do som e desvios de padrão relacionados a erros de gravação. Além disso o *Mel-Spectrogram* tenta trabalhar como a audição humana que possui filtros não uniformes, possuindo mais filtros em regiões de baixa frequência e menos filtros na região de alta frequência.

O processo para se obter esse espectrograma pode variar, mas em geral é composto dos seguintes passos:

- Tirar a Transformada de Fourier de um sinal ou alguma outra transformação espectral.

- Mapear as potências obtidas na escala Mel usando *triangular overlapping windows*.
- Transformar as potências para escala log em cada uma das frequências Mel
- Fazer a transformada de cosseno discreta da lista de potências em log como se fossem um sinal.
- O resultado são os coeficientes do Espectograma Mel

Para obter estes dados foi utilizado a função *melspectrogram* do LibROSA

O **Tempo de duração do áudio** é uma informação que estava armazenada na tabela *Calls* do SumForce e por esse motivo não foi necessário ser extraída pelo LibROSA

Outro detalhe importante de se trabalhar com em escala é que as funções do LibROSA que geram as características são complexas e isso faz com que o programa demore bastante. Para evitar erros ou perdas de informação a obtenção das características foi realizada de com 500 áudios por vez.

Independente do algoritmo de aprendizado de máquina utilizado, se as características utilizadas como entradas no sistema não forem adequadas o sistema de aprendizado não terá bons resultados. Portanto o enfoque foi analisar os dados de áudio e conseguir os melhores resultados para futuramente concluir o trabalho aplicando um sistema de aprendizado de máquina.

4.2 Análise e Resultados

As características foram escolhidas, em seguida elas foram testadas. Dentro dos estudos envolvendo análises acústicas alguns termos não são totalmente definidos ou possuem diferentes significados dentro de contextos diferentes. Por esse motivo diversos testes foram feitos com características como a Tom do áudio, que depois de mostraram completamente irrelevante para este estudo.

Esta análise teve alguns processos, dos 5500 áudios obtidos da Amazon S3 apenas 4000 foram utilizados, os 1500 retirados tinham uma duração menor do que 3 minutos ou eram arquivos corrompidos. Estes 4000 arquivos de áudio foram separados de

acordo com a data em que foram gravados. Três grupos foram formados, os mais antigos, os mais atuais e os intermediários. O intuito dessa separação é que as amostras que seriam utilizadas no processo de aprendizado de máquina, o grupo de teste e de treinamento seria composto de parcelas aleatórias dos três grupos garantindo que o período não influenciasse na análise.

Para analisar as características primeiramente foram feitos diversos gráficos comparando de dois a dois cada uma das características, como pode ser observado na Figura 4.4. Foi olhado individualmente cada um desses gráficos em busca de padrões que distinguissem a incidência de *NoShow* com a de Reunião Realizada, mas como já era de se esperar nenhum padrão foi encontrado.

Neste gráfico está contido os valores de ritmo em batidas por minuto, de tempo de duração da ligação, média e desvio padrão do Espectograma Mel, e da Potencia em Decibéis, isso tanto para o áudio em 32 bits com os dois canais quanto para o canal separado do lead em 16 bits.

Este resultado era esperado, pois a identificação de um sentimento é feita da composição de várias características e não seria um par que teria uma resposta satisfatória.

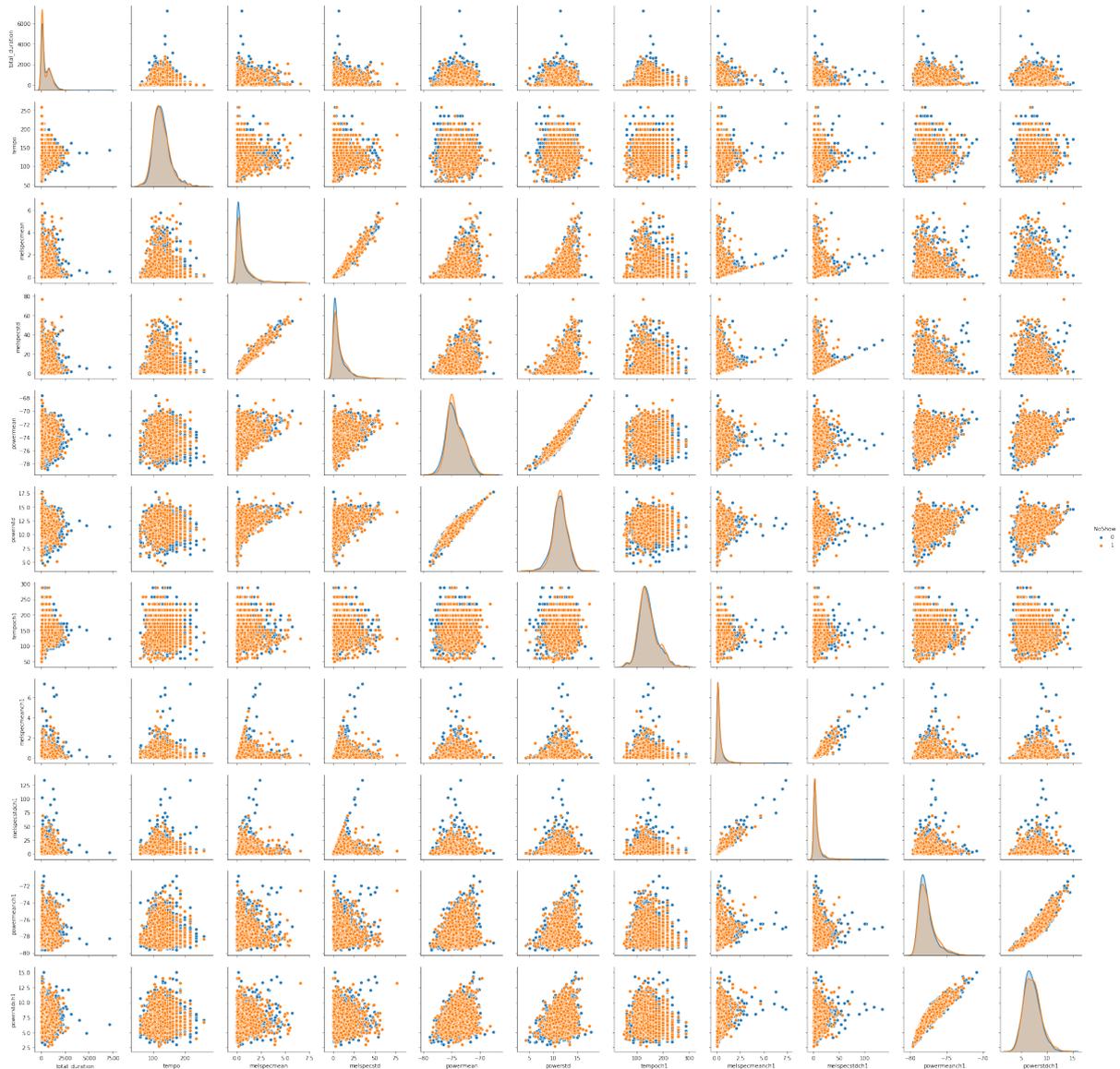


Figura 4.4: Diagramas de dispersão par a par

Outras informações podem ainda ser retiradas destes dados antes de uma análise voltada ao aprendizado de máquina. O tempo de duração de uma ligação sempre foi uma preocupação grande entre os pré-vendedores, eles foram orientados a passar um "roteiro" de ligação de uma forma fluida e a orientação que eles tinham era que ligações muito longas costumavam ter uma menos taxa de sucesso. Essa afirmação não é real, aparentemente a duração da ligação não tem ligação direta e independente

com o sucesso da ligação, mas considerando agora que uma ligação mais longa tem um maior custo para empresa tanto em funcionários, quanto em valor telefônico, isso quer dizer que o time de treinamento dos pré-vencedores pode se dedicar a tornar a ligação cada vez mais curta e assim ter mais eficiência sem perder a eficácia do trabalho.

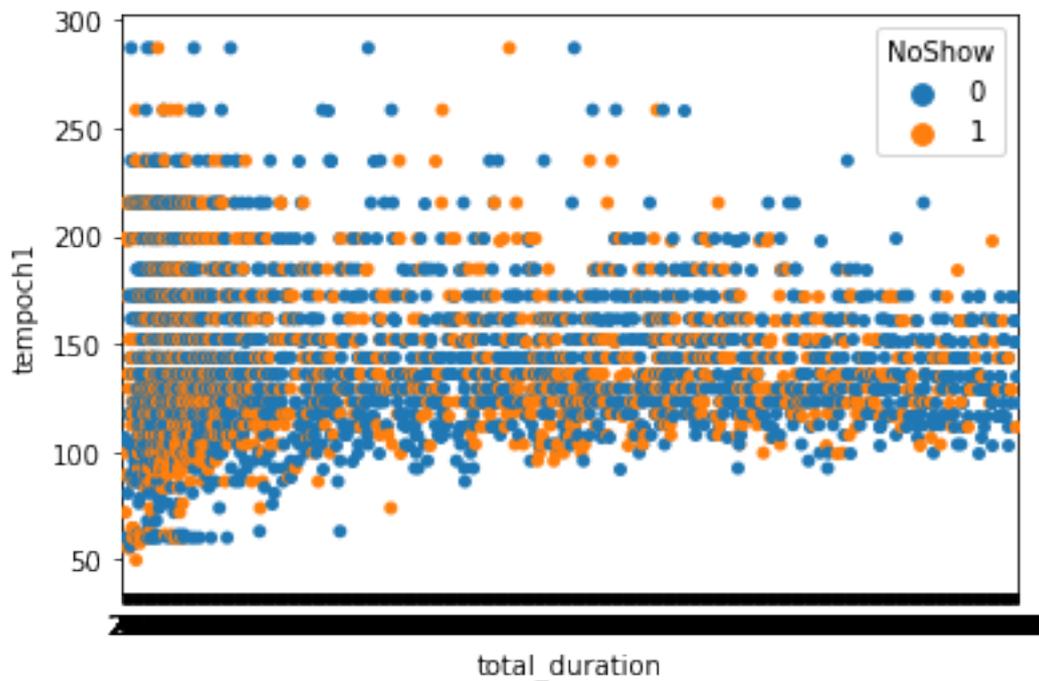


Figura 4.5: Gráfico de tempo de duração pelo ritmo

Para conseguir analisar os dados obtidos a melhor maneira encontrada foi utilizar o tempo de duração como base por ser o menos complexo, ou seja, os gráficos que relacionavam as características coletadas ao tempo tinham uma mais fácil interpretação. Pode se observar, por exemplo, que a faixa de potencia tem uma limitação funcionando entre a faixa de -70 e -78 decibéis. Isso é lógico de se pensar que por mais que umas pessoas falem mais alto que outras dentro de um dialogo sem conflitos as pessoas tendem a falar em uma altura de voz moderada.

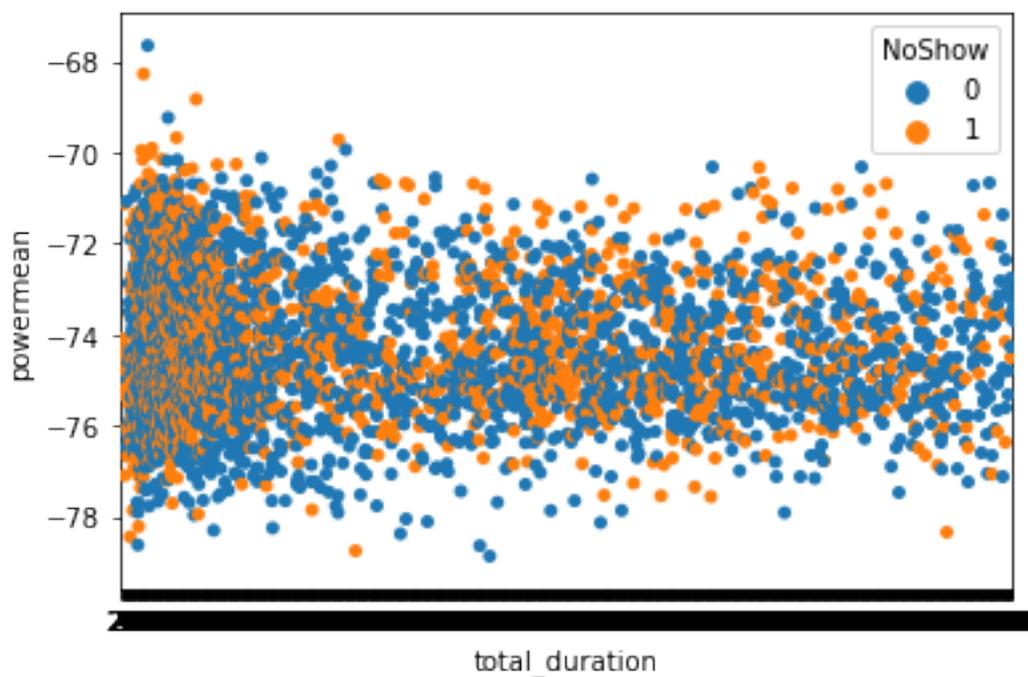


Figura 4.6: Gráfico de tempo de duração pela potência

Na Figura 4.5 além de observar o tempo em si também podemos compreender que os pré-vendedores possuem a capacidade e manter um ritmo específico dentro da sua conversa, isso contribui para a fluidez das informações e para a indicação de No-Show não parece ter muita relevância, mas talvez investigando as ligações que foram incapazes de marcar uma reunião este atributo demonstre um padrão mais relevante.

Capítulo 5

Conclusões

A magnitude e complexidade deste trabalho estava muito além do era imaginado, para alguém que nunca tinha tido contato com análise digital de sinais acústicos, compreender os conceitos básicos do trabalho foi um desenvolvimento longo e dificultoso. Além disso a biblioteca LibROSA foi feita para ser utilizada por pessoas com alto conhecimento sobre análise de áudios o que a torna pouco explicativa e com quase nenhuma indicação de métricas ou escalas. Contudo esse estudo abriu portas para todo um novo espectro de possibilidades de trabalho e desenvolvimento.

Este documento conclui os estudos realizados até o determinado momento mas o projeto seguirá em desenvolvimento. Os dados obtidos agora estão processados e prontos para serem utilizados em um algoritmo de aprendizado de máquina. Além disso outras ferramentas foram encontradas para enriquecer o projeto.

A Linguagem Python é conhecida por sua facilidade em aplicar conceitos de inteligência artificial e aprendizado de máquina, contudo na escolha deste projeto não foi a melhor opção. Como é uma linguagem com a qual eu tenho mais facilidade tornou o trabalho mais fluido e facilitou o meu processo de pesquisa. Mas para continuidade deste projeto será trabalhado em C++ junto com a biblioteca Kaldi. Nesta biblioteca será possível retirar características do áudio mais dedicadas à análise de sentimentos. O Python contudo não será descartado, esta ainda é a melhor ferramenta sobre o meu domínio para geração de gráficos e manipulação dos dados.

Com esta nova ferramenta a diante, os resultados já encontrados, mesmo que não satisfatórios, e por fim toda essa base desenvolvida em conceitos de analise de áudios, este projeto terá agora uma trilha mais planejada para a magnitude de dificuldade

que é conseguir criar um preditor satisfatório. Além disso dentro da Zygo Tecnologia tem um interesse bem grande de que esse projeto traga resultados, com isso tenho a disposição engenheiros sênior para me orientar na continuidade deste trabalho.

Bibliografia

- [1] The PostgreSQL Global Development Group *PostgreSQL 11.4 Documentation*. 1996-2019 <https://www.postgresql.org/files/documentation/pdf/11/postgresql-11-A4.pdf>. [Online; acessado 30-Junho-2019].
- [2] SciPy developers *Scientific Computing Tools for Python*. 2019 <https://www.scipy.org/about.html>. [Online; acessado 30-Junho-2019].
- [3] pandas: powerful Python data analysis toolkit Release 0.24.2 *Wes McKinney and PyData Development Team*. 2019 <http://pandas.pydata.org/pandas-docs/stable/pandas.pdf>. [Online; acessado 30-Junho-2019].
- [4] NumPy User Guide *Scientific Computing Tools for Python*. 2019 <https://docs.scipy.org/doc/numpy/user/index.html>. [Online; acessado 30-Junho-2019].
- [5] LibROSA *librosa development team*. 2018 <http://librosa.github.io/librosa/>. [Online; acessado 30-Junho-2019].
- [6] Documentation of scikit-learn 0.21.2 *scikit-learn developers*. 2019 <https://scikit-learn.org/stable/documentation.html>. [Online; acessado 30-Junho-2019].
- [7] Py Audio Analysis A Python library for audio feature extraction, classification, segmentation and applications *Theodoros Giannakopoulos*. 2019 <https://github.com/tyiannak/pyAudioAnalysis>. [Online; acessado 30-Junho-2019].
- [8] Requests: HTTP for Humans™ *Kenneth Reitz*. 2019 <https://2.python-requests.org/en/master/>. [Online; acessado 30-Junho-2019].
- [9] wave — Read and write WAV files *Python Software Foundation*. 2019 <https://docs.python.org/3/library/wave.html>. [Online; acessado 30-Junho-2019].
- [10] SoundFile 0.10.2 *Python Software Foundation*. 2019 <https://pypi.org/project/SoundFile/>. [Online; acessado 30-Junho-2019].

- [11] sgbcloud *Lucas Pedro Bordignon*. 2018 <https://github.com/lucaspbordignon/sgbcloud>. [Online; acessado 30-Junho-2019].
- [12] SpeechRecognition *Anthony Zhang (Uberi)*. 2017 https://github.com/Uberi/speech_recognition. [Online; acessado 30-Junho-2019].
- [13] CMUSphinx Documentation <https://cmusphinx.github.io/wiki/>. [Online; acessado 30-Junho-2019].
- [14] Cloud Speech-to-Text <https://cloud.google.com/speech-to-text/>. [Online; acessado 30-Junho-2019].
- [15] Speech to Text <https://www.ibm.com/watson/services/speech-to-text/>. [Online; acessado 30-Junho-2019].
- [16] Sobre o Retextoar *Retextoar*. 2019 <https://www.retextoar.com.br/sobre/>. [Online; acessado 30-Junho-2019].
- [17] Survey on speech emotion recognition: Features, classification schemes, and databases *Moataz El Ayadi, Mohamed S. Kamel, Fakhri Karray* 2010
- [18] Fitting the mel scale. In Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP) *Umesh.S, Cohen.L., Nelson.D.* 1998.
- [19] Auditory Toolbox: A MATLAB Toolbox for Auditory Modeling Work. *Slaney, M.* 1998.
- [20] The Relevance of Voice Quality Features in Speaker Independent Emotion Recognition *Marko Lugger, Bin Yang* 2007 <https://ieeexplore.ieee.org/abstract/document/4218026> [Não mais liberado acesso; acessado 30-Junho-2019].
- [21] Digital Processing of Speech Signals *L.R.Rabiner, R.W.Schafer* 1978
- [22] Spin Selling *Neil Rackham* 1988
- [23] Machine Learning *Tom M. Mitchell* 1997
- [24] Speech Technology: A Practical Introduction *Kishore Prahallad* http://www.speech.cs.cmu.edu/15-492/slides/03_mfcc.pdf[Online; acessado 30-Junho-2019]