Universidade Federal do Rio Grande do Norte

Centro de Tecnologia - CT

Programa de Pós-Graduação em Engenharia Mecatrônica

Assistente Virtual para Centros de Operação

Camila Barbosa Gomes de Araújo

Natal-RN, Brasil

2023



Universidade Federal do Rio Grande do Norte

Centro de Tecnologia - CT

Programa de Pós-Graduação em Engenharia Mecatrônica

Assistente Virtual para Centros de Operação

Camila Barbosa Gomes de Araújo

Orientador: Samaherni Morais Dias

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Mecatrônica da UFRN como parte dos requisitos para obtenção do título de Mestre em Engenharia Mecatrônica.

Natal-RN, Brasil

2023

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Barbosa, Camila.
Assistente Virtual para Centros de Operação/ Camila Barbosa Gomes de Araújo. - 2023.
81f.: il.

Dissertação (mestrado) - Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecatrônica. Natal, RN, 2023.
Orientador: Prof. Dr. Samaherni Morais Dias

1. Transcrição - Dissertação. 2. Setor Elétrico - Dissertação. 3. Pós-operação - Dissertação. 4. Correlação - Dissertação. 5. Classificação de áudios - Dissertação. 6. Processamento de Linguagem Natural - Dissertação. I. Dias, Samaherni Morais. II. Título

RN/UF/BCZM CDU 004.934(043.3)

Elaborado por Fernanda de Medeiros Ferreira Aquino - CRB-15/301

Assistente Virtual para Centros de Operação

Camila Barbosa Gomes de Araújo

Dissertação de Mestrado aprovada em 13 de abril de 2023 pela banca examinadora composta pelos seguintes membros:

Prof. Dr. Samaherni Morais Dias (orientador) ..... DEE/UFRN

Prof. Dr. Kurios Iuri Pinheiro de Melo Queiroz ..... DEE/UFRN

Prof. Dr. Ivanovitch Medeiros Dantas da Silva ..... DCA/UFRN

Prof. Dr. Orivaldo Vieira de Santana Junior ..... ECT/UFRN

Prof. Dr. Pedro Thiago Valério de Souza ..... DETEC/UFERSA

Natal-RN, Brasil
2023

Resumo

A coordenação entre as atividades de operação do setor elétrico é realizada por um conjunto de ferramentas e tecnologias, dentre as quais destaca-se a comunicação via áudio entre as equipes de tempo real dos agentes e técnicos em campo. Essas chamadas são gravadas, armazenadas e consultadas, sob demanda, para fins de auditoria e rastreabilidade, e para subsidiar as análises pós-operatórias. Contudo, essas gravações, que contém informações relevantes tanto para a operação do sistema quanto para o faturamento dos agentes, acabam sendo subutilizadas devido à dificuldade de acesso ao seu conteúdo. Assim, ao transcrevê-las, torna-se possível a rastreabilidade e a associação desses áudios entre si e com outros registros de diferentes sistemas operativos, viabilizando a otimização de atividades que demandam de um tempo excessivo dos engenheiros de pós-operação. Esta dissertação apresenta um assistente virtual para centros de operação para geradoras, transmissoras e distribuidoras de energia, que transcreve todas as comunicações via áudio desses centros e aplica técnicas de processamento de linguagem natural para extrair informação dos áudios, classificá-los em temas de interesse da operação e correlacionar as gravações entre si e com sistemas operativos. São apresentados a arquitetura do sistema, os modelos de transcrição de áudio para texto utilizados, o módulo de classificação de ligações e a correlação de ligações entre si e com o SCADA. O projeto aqui descrito se propõe a ajudar na tomada de decisão e contribui para maior eficiência, segurança e confiabilidade operacional da operação e pós-operação do sistema elétrico.

Palavras-chave: transcrição, setor elétrico, pós-operação, correlação, classificação de áudios, processamento de linguagem natural.

Abstract

The coordination of the electricity sector's operating activities is carried out by a set of tools and technologies, the most notable of which is audio communication between real-time teams of agents and technicians in the field. These calls are recorded, stored, and consulted on demand for auditing and traceability purposes, as well as to support postoperative analyses. However, because of the difficulty in accessing their content, these recordings, which contain relevant information for both system operation and agent billing, are underutilized. Thus, by transcribing them, it is possible to trace and associate these audios with each other and with other records from different operating systems, allowing for the optimization of activities that require a significant amount of time from post-operation engineers. This dissertation describes a virtual assistant for energy generators, transmitters, and distributors' operation centers that transcribes all audio communications from these centers and uses natural language processing techniques to extract information from the audios, classify them into topics of interest to the operation, and correlate the recordings with each other and with operating systems. The system architecture is presented, as well as the audio-to-text transcription models used, the classification module, and the correlation between audios and with SCADA module. The project described here is intended to aid decision-making and contribute to greater efficiency, safety, and operational reliability in electrical system operation and post-operation.

Keywords: speech-to-text, energy sector, post-operation, correlation, audio classification, natural language processing.

Lista de Ilustrações

Figura 1 – Arquitetura do sistema na *AWS* 24
Figura 2 – Arquitetura do sistema no GCP 26
Figura 3 – Arquitetura do sistema na *Azure* 30
Figura 4 – *Print* da ferramenta na tela de configuração de vocabulário 34
Figura 5 – Diagrama dos testes de vocabulário no *GCP* 35
Figura 6 – *Print* da ferramenta na tela de ligações com os áudios classificados . . 38
Figura 7 – *Print* da tela de configurações dos parâmetros de retreinamento 39
Figura 8 – Gráfico de estabilização da porcentagem de gravações correlacionáveis . 47
Figura 9 – Diagrama do algoritmo completo da correlação de áudios com eventos SCADA 49
Figura 10 – Tela com pop-up da correlação de áudios com SCADA 51
Figura 11 – Diagrama do baseline da correlação de histórias PABX 52
Figura 12 – Diagrama da versão final da correlação de histórias PABX 54
Figura 13 – Diagrama da versão final da correlação de histórias PABX 55
Figura 14 – Diagrama do *baseline* da correlação de áudios de rádio 55
Figura 15 – Diagrama da versão final da correlação de áudios de rádio 57
Figura 16 – WER para os transcritores avaliados 62
Figura 17 – Métrica REC para o transcritor GCP bem como para os baseilines treinados com a adaptação do vocabulário personalizado 63
Figura 18 – Evolução da acurácia do *baseline* (primeiros testes) 64
Figura 19 – Evolução da acurácia do modelo 64
Figura 20 – *Learning curve analysis* do modelo completo 65
Figura 21 – *Learning curve analysis* de cada classe 65
Figura 22 – Gráfico da evolução das métricas do algoritmo 70
Figura 23 – Tela de busca da ferramenta com as classificações 72
Figura 24 – Tela de busca da ferramenta com exemplo de “Indaiatuba” 73
Figura 25 – Tela de busca da ferramenta com exemplo de correlação aberto 73

Lista de Quadros

Quadro 1 – Exemplo de correlação de áudio com SCADA 48
Quadro 2 – Exemplo de correlação de áudio com SCADA 52
Quadro 3 – Exemplo de correlação de áudio com SCADA 53
Quadro 4 – Levantamento do tempo de busca por ligações 71

Lista de Tabelas

Tabela 1 – Distribuição das classes no *dataset* 40
Tabela 2 – REC e WER para os vocabulários testados com a AWS 59
Tabela 3 – FP para os vocabulários testados com a AWS 59
Tabela 4 – Resultado da otimização de vocabulários e *boosts* 60
Tabela 5 – Métricas por classe do modelo final 66
Tabela 6 – Comparação das métricas entre os modelos de *baseline* para o *pipeline* com TF-IDF 66
Tabela 7 – Comparação das métricas dos melhores pipelines com TF-IDF 67
Tabela 8 – Métricas do melhor modelo com TF-IDF 67
Tabela 9 – Métricas do modelo de classificação com BERT 68
Tabela 10 – Comparação de F1-Score das diferentes abordagens para cada classe . . 68
Tabela 11 – Métricas por classe do baseline 69
Tabela 12 – Métricas do *baseline* da correlação de histórias baseadas no ADMS . . 70
Tabela 13 – Métricas da correlação de histórias baseadas no ADMS 71
Tabela 14 – Métricas do *baseline* da correlação de histórias de rádio 71

Lista de Abreviaturas e Siglas

ASR	Reconhecimento de Fala Automático, do inglês *Automatic Speech Recognition*
TF-IDF	do inglês, *Term Frequency - Inverse Frequency*
BERT	do inglês, *Bidirectional Encoder Representations from Transformers*
SCADA	do inglês, *Supervisory Control and Data Acquisition*
ANEEL	Agência Nacional de Energia Elétrica
ONS	Operador Nacional do Sistema
MAE	Mercado Atacadista de Energia
SIN	Sistema Interligado Nacional
PMO	Programa Mensal de Operação Energética
PDO	Programa Diário de Operação
PDP	Programa Diário de Produção
PDFC	Programa Diário de Carga e Frequência
CAG	Controle Automático de Geração
COS	Centros de Operação do Sistema
COG	Centro de Operação da Geração
COI	Centro de Operação Integrado
SAU	Sistema de Acompanhamento de Usinas
SAMUG	Sistema de Apuração de Mudanças de Estados Operativos de Unidades Geradoras, Usinas e Interligações Internacionais
GCP	*Google Cloud Platform*
AWS	*Amazon Web Services*
SFTP	Protocolo de Transferência Segura de Arquivos, do inglês *Secure File Transfer Protocol*
DAG	do inglês, *Directed Acyclic Graph*

API	do inglês, *Application Programming Interface*
ADF	*Azure Data Factory*
LSTM	do inglês, *Long Short-Term Memory*
RNN	do inglês, *Recurrent Neural Networks*
CNN	do inglês, *Convolutional Neural Networks*
GRU	do inglês, *Gated Recurrent Unit*
CTC	do inglês, *Connectionist Temporal Classification*
LAS	do inglês, *Listen, Attend and Spell*
WER	do inglês, *Word Error Rate*
REC	Percentual de Reconhecimento de Palavras-chave
FP	Falso Positivo
STT	do inglês, *Speech-to-Text*
PABX	Ramal telefônico, do inglês *Private Automatic Brach Exchange*
IA	Inteligência Artificial
O&M	Operação & Manutenção
NLP	Processamento de Linguagem Natural, do inglês *Natural Language Processing*
SVM	do inglês, *Support Vector Machine*
TPOT	do inglês, *Tree-based Pipeline Optimization Tool*
LCA	do inglês, *Learning Curve Analysis*
MLM	do inglês, *Masked Language Modeling*
NSP	do inglês, *Next Sentence Prediction*
ADMS	do inglês, *Advanced Distribution Management System*
OS	Ordem de Serviço
SP	Plano de Manobra, do inglês *Switch Plan*
UFRN	Universidade Federal do Rio Grande do Norte

Sumário

1 Introdução . . . 13
1.1 Organização do Trabalho . . . 15

2 Fundamentação teórica . . . 17
2.1 Arquiteturas em Nuvem . . . 17
2.2 Transcrição automática de áudio . . . 17
2.3 Processamento de Linguagem Natural para Classificação de Ligações . . . 19
2.4 Correlação de Ligações . . . 22

3 Assistente Virtual para Centros de Operação . . . 23
3.1 Arquitetura proposta . . . 23
3.1.1 Arquitetura na *Amazon Web Services* . . . 24
3.1.2 Arquitetura no *Google Cloud Platform* . . . 26
3.1.3 Arquitetura na *Azure Cloud Services* . . . 30
3.2 Transcrição automática de áudio . . . 31
3.2.1 Implementação na *Amazon Web Services* . . . 32
3.2.2 Implementação no *Google Cloud Platform* . . . 34
3.2.3 Implementação na *Azure Cloud Services* . . . 36
3.3 Classificação de Ligações . . . 37
3.3.1 Classificação com palavras-chave e informações das ligações como *features* 40
3.3.2 Classificação utilizando a técnica TF-IDF . . . 42
3.3.3 Classificação utilizando o modelo BERT . . . 43
3.4 Correlação de Ligações . . . 44
3.4.1 Correlação de Áudios com Eventos SCADA . . . 46
3.4.2 Correlação de Histórias . . . 50
3.4.2.1 Histórias Baseadas em Eventos SCADA . . . 51
3.4.2.2 Histórias com Áudios de Rádio . . . 55

4 Resultados . . . 58
4.1 Modelo *Speech-to-Text* . . . 58
4.1.1 Modelo da *AWS* . . . 58
4.1.2 Modelo do *GCP* . . . 60
4.1.3 Modelo da *Azure* . . . 61
4.2 Modelo de Classificação de Ligações . . . 63
4.2.1 *Baseline* e Avanços . . . 63
4.2.2 Abordagem com TF-IDF . . . 66
4.2.3 Abordagem com BERT . . . 67

4.2.4 Modelo Adotado em Produção 67
4.3 Correlação de Ligações 68
4.3.1 Correlação de Áudios com Eventos SCADA 68
4.3.2 Correlação de Histórias 69
4.3.2.1 Histórias baseadas em Eventos SCADA 70
4.3.2.2 Histórias com Áudios de Rádio 71
4.4 Assistente Virtual para Centros de Operação 71

5 Considerações Finais e Trabalhos Futuros 74

Referências 75

1 Introdução

O sistema elétrico brasileiro possui as atividades do setor desverticalizadas em segmentos: geração, transmissão e distribuição. Para administrar o setor, são necessárias três entidades: a Agência Nacional de Energia Elétrica (ANEEL), o Operador Nacional do Sistema (ONS) e o Mercado Atacadista de Energia (MAE). A ANEEL é responsável pela regulamentação e controle do setor; O ONS coordena, planeja e opera o sistema elétrico interligado; e o MAE administra o mercado atacadista de energia (FILGUEIRAS; SILVA, 2003).

O segmento de geração de energia elétrica inclui a prospecção, implantação, operação e manutenção das usinas (GANIM, 2008). A matriz elétrica brasileira possui mais de 50% de geração de origem hidrelétrica (EPE, 2022), dada a grande disponibilidade de recursos hídricos no país.

A programação da geração leva em consideração o estado atual de armazenamento dos reservatórios, previsões atualizadas de carga de energia por patamar, a análise das condições meteorológicas verificadas e previstas nas principais bacias do Sistema Interligado Nacional (SIN) e previsões de afluências aos aproveitamentos hidrelétricos. O Programa Mensal de Operação Energética (PMO) é elaborado pelo ONS com a participação dos agentes (FILGUEIRAS; SILVA, 2003; CASTRO et al., 2020).

O segmento de transmissão constitui o elo entre a geração e a distribuição de energia elétrica e compreende a implantação, operação e manutenção das linhas de instalações de transmissão. A Rede Básica do SIN é constituída por todas as subestações e linhas de transmissão em tensões iguais ou superiores a 230 kV (GANIM, 2008).

O segmento de distribuição de energia elétrica distribui a energia recebida do sistema de transmissão aos grandes, médios e pequenos consumidores. A operação dos três segmentos é coordenada por centros de operação, sendo eles, Centros de Operação do Sistema (COS), Centro de Operação da Geração (COG) e Centro de Operação Integrado (COI). Entre as principais responsabilidades desses centros estão a coordenação, supervisão e controle de ações para garantir a constante otimização energética e a segurança operacional do sistema (ONS, 2003).

A coordenação entre as atividades de operação desses segmentos do setor elétrico é realizada por um conjunto de ferramentas e tecnologias, dentre as quais destacam-se comunicações via áudio entre centros de operação e técnicos em campo. Essas chamadas são gravadas, armazenadas e consultadas, sob demanda, para fins de auditoria e rastreabilidade, e para subsidiar as análises pós-operatórias. Contudo, essas gravações, que contém informações relevantes tanto para a operação do sistema quanto para o faturamento dos

agentes, acabam sendo subutilizadas devido à dificuldade de acesso ao seu conteúdo. Assim, ao transcrevê-las, torna-se possível a rastreabilidade e a associação desses áudios entre si e com outros registros de diferentes sistemas operativos, viabilizando a otimização de atividades que demandam de um tempo excessivo dos engenheiros de pós-operação.

A tecnologia de Reconhecimento de Fala Automático (ASR, do inglês *Automatic Speech Recognition*) existe há décadas e está incorporada em muitos produtos, desde chamadas automatizadas, tradução, brinquedos inteligentes, assistentes pessoais, entre outros (MCLEAN; OSEI-FRIMPONG, 2019; TULSHAN; DHAGE, 2019; LIN et al., 2016; WU et al., 2019; AIZAWA; KATO; KOSAKA, 2016). No entanto, os sistemas ASR estão longe de serem perfeitos: seu desempenho degrada rapidamente com ruídos de fundo ou falas distantes, e eles tendem a ter uma taxa de erro mais alta se não possuírem dados de treinamento anotados suficientes (QUINTANILHA; NETTO; BISCAINHO, 2020).

Atualmente, os modelos de estado-da-arte de reconhecimento de fala se baseiam em técnicas de *deep learning* (GOODFELLOW; BENGIO; COURVILLE, 2016). Todavia, esses modelos são muito dependentes de dados para terem um treinamento eficiente (MUSTAFA; ALLEN; APPIAH, 2019). A construção de um grande *corpus* é demorada e não é uma tarefa trivial. Esforços de código aberto, como o *Common Voice* (ARDILA et al., 2019) liderado pela Mozilla, conseguiram reunir mais de 1.000 horas de fala em inglês, muito longe de sua meta original de 10.000 (QUINTANILHA; NETTO; BISCAINHO, 2020). Esses números são ordens de grandeza inferiores aos que empresas privadas, como Amazon e Google, vêm relatando com seus resultados (AMODEI et al., 2016; HE et al., 2019; MCLEAN; OSEI-FRIMPONG, 2019).

Ao permitir uma busca indexada das transcrições, já nota-se a redução na quantidade de trabalho manual envolvido nas atividades pós-operatórias dos centros de operação. Os *call centers* têm utilizado a mesma estratégia, com as chamadas transcritas para várias aplicações: auditoria de desempenho dos operadores (MEHTA et al., 2020), redistribuição de chamadas baseada no reconhecimento de emoções (BOJANIć et al., 2020), assistente virtual em tempo real (VASILATEANU; ENE, 2018), mineração informações do cliente (TAKEUCHI et al., 2008; MEHTA et al., 2020) e classificação de áudios.

No contexto da operação de geração do setor elétrico no Brasil, é possível classificar os áudios em: solicitações de mudança de estado de máquinas, de modulação de geração e ajuste de tensão, notificações de chuva e nível de reservatório e solicitações e atualizações de manutenção de unidades geradoras. Uma ligação não se limita a apenas um desses temas, tornando o problema de classificação numa abordagem *multi-label*.

Nesse trabalho, a classificação de áudio recorre à classificação de texto. A literatura propõe diferentes abordagens para lidar com esse problema. Imane e Mohamed (2017a) propuseram uma abordagem de categorização multi-rótulo para um problema de certidão de óbito na França com algoritmos de classificação tradicionais, Clercq et al. (2019) escolheu

um método semelhante para uma aplicação distinta. Savyanavar et al. (2016) usa *Term Frequency - Inverse Document Frequency* (TF-IDF) para gerenciar e resumir documentos no contexto de organizações que lidam com grandes cargas de arquivos, como escritórios de advocacia. Al-Talib e Hassan (2013) usa estatísticas TF-IDF para ponderação de palavras para classificação de SMS. Liu et al. (2021) propõe uma estrutura que usa um modelo BERT (do inglês, *Bidirectional Encoder Representations from Transformers*) pré-treinado para extração de informações de textos de ensaios clínicos.

Além da classificação de áudios, outra aplicação da transcrição de áudios em centros de operação também é apresentada neste trabalho. A correlação de áudios entre si e com sistemas operativos da transmissão e distribuição (como o SCADA (do inglês, *Supervisory Control and Data Acquisition*), por exemplo) traz para a pós-operação a centralização de todas as informações relacionadas a um evento de campo, independentemente de sua origem, além de diminuir o tempo gasto em análises de ocorrências que impactam diretamente na operação e no faturamento dos agentes.

Na literatura, não são muitos os exemplos de correlação de transcrições de áudios com sistemas operativos, todavia, ao particionar o problema, é possível identificar exemplos como em Sager (1978) e Tekli (2016), onde há a conversão automática de textos para uma base de dados estruturada; Kirk et al. (2012) e Zhang, Zhuang e Wu (2007) aplicam clusterização para associação de dados não estruturados e dados estruturados. Parte da solução aqui apresentada consiste em extrair dados estruturados da transcrição.

Isso posto, o projeto apresentado nesta dissertação objetivou à criação de um assistente virtual para centros de operação para geradoras, transmissoras e distribuidoras, que transcreva todas as comunicações via áudio desses centros e as correlacione com dados que diversos sistemas operativos, de forma a ajudar na tomada de decisão e na análise dessas informações de uma forma automatizada, contribuindo para maior eficiência, segurança e confiabilidade operacional da operação e pós-operação do sistema elétrico.

Esta dissertação engloba um projeto P&D ANEEL cooperado entre a CPFL Energia e Engie Brasil Energia que tem a Radix Engenharia & Software como executora. O código de referência do projeto é PD-00063-3071/2020.

1.1 Organização do Trabalho

O texto está dividido em cinco capítulos, sendo que no primeiro é apresentado o contexto do setor elétrico no Brasil, como é a operação na geração, na transmissão e na distribuição e quais as responsabilidades de cada entidade do sistema. Além disso, também apresenta o P&D ANEEL como fomento de pesquisa e desenvolvimento no setor. O Capítulo 2 tem como propósito apresentar a fundamentação teórica requerida para uma melhor compreensão do texto. Já no Capítulo 3 é apresentada a proposta para a

assistente virtual, com detalhes de sua arquitetura interna, implementações em nuvem e detalhamento de seus serviços internos. No Capítulo 4, são expostos todos os resultados obtidos e, finalmente, no Capítulo 5 são apresentadas as considerações finais do trabalho.

2 Fundamentação teórica

Neste capítulo serão apresentados os principais elementos utilizados na proposta do assistente virtual para centros de operações desenvolvido neste trabalho.

2.1 Arquiteturas em Nuvem

A computação em nuvem representa uma mudança orientada a serviços na forma como os aplicativos de software são projetados, construídos e entregues. O termo computação em nuvem, que se popularizou no início de 2007, refere-se a servidores virtuais, hospedagem distribuída em grandes *datacenters* e a recursos compartilhados disponíveis pela Internet (BIBI; KATSAROS; BOZANIS, 2012; KANT, 2009).

Na computação em nuvem, não há despesas com manutenção e segurança do hardware; é possível fazer *backup* dos dados regularmente; apenas se paga pelos recursos que são utilizados e atualizações de memória e capacidade de processamento tornam-se menos burocráticas - e quase instantânea (DIKAIAKOS et al., 2009).

Já um servidor *on-premise* é caracterizado por possuir toda a infraestrutura, customização, configuração e atualização feita internamente pela empresa responsável. A principal desvantagem de hospedar servidores *on-premise* são os custos associados ao gerenciamento e manutenção de toda a infraestrutura (espaço físico adequado - servidor de incêndios, inundações, desabamentos e roubos) (NARAYANAN; MEENAKUMARI, 2021). Todavia, muitas empresas de setores críticos, como o de energia, mantêm seus dados armazenados localmente, para garantir a segurança e integridade dos mesmos (PETERSEN; SEYMOUR, 2021).

2.2 Transcrição automática de áudio

A tecnologia de Reconhecimento de Fala Automático (ASR, do inglês *Automatic Speech Recognition*) existe há algumas décadas e está incorporada em muitos produtos, desde chamadas automatizadas, tradução, brinquedos inteligentes, assistentes pessoais, entre outros (MCLEAN; OSEI-FRIMPONG, 2019; TULSHAN; DHAGE, 2019; LIN et al., 2016; WU et al., 2019; AIZAWA; KATO; KOSAKA, 2016). No entanto, os sistemas ASR estão longe de serem perfeitos: seu desempenho degrada rapidamente com ruídos de fundo ou falas distantes, e eles tendem a ter uma taxa de erro mais alta se não possuírem dados de treinamento anotados suficientes (QUINTANILHA; NETTO; BISCAINHO, 2020).

Atualmente, os modelos de estado-da-arte de reconhecimento de fala se baseiam em

técnicas de *deep learning* (GOODFELLOW; BENGIO; COURVILLE, 2016). Todavia, esses modelos são muito dependentes de dados para terem um treinamento eficiente (MUSTAFA; ALLEN; APPIAH, 2019). A construção de um grande *corpus* é demorada e não é uma tarefa trivial. Esforços de código aberto, como o *Common Voice* (ARDILA et al., 2019) liderado pela Mozilla, conseguiram reunir mais de 1.000 horas de fala em inglês, muito longe de sua meta original de 10.000 (QUINTANILHA; NETTO; BISCAINHO, 2020). Para idiomas diferentes do inglês, como o holandês, apenas menos de 31 horas de fala anotada estão disponíveis (de uma meta de 1.200 horas). Esses números são ordens de grandeza inferiores aos que empresas privadas, como AWS e Google, vêm relatando com seus resultados (AMODEI et al., 2016; HE et al., 2019; MCLEAN; OSEI-FRIMPONG, 2019).

A falta de áudios anotados e de *datasets* públicos dificulta a avaliação de sistemas ASR para outros idiomas (que não inglês ou mandarim) (QUINTANILHA; NETTO; BISCAINHO, 2020). Isso é especialmente relevante para idiomas com apenas algumas dezenas de horas de dados públicos disponíveis, como o português brasileiro (PT-BR).

A arquitetura dos modelos ponta-a-ponta de reconhecimento de fala baseados em *deep learning* pode ser dividida nos seguintes componentes:

- *Speech processing*: primeiro responsável pelo processamento dos sinais. O trabalho de Jaitly e Hinton (2011) apresenta uma abordagem para modelagem das ondas sonoras em máquina restritas de Boltzmann com *encondings* de alta dimensionalidade, ultrapassando a performance dos modelos estado-da-arte da época;

- Modelos acústicos: fazem a correspondência entre as características acústicas e os fonemas. O trabalho de Hinton et al. (2012) faz uma comparação entre os modelos acústicos baseados em cadeias de Markov e modelos de mistura gaussiana com abordagens mais modernas utilizando redes neurais *feed-forward* e recorrentes, com estas apresentando melhor performance;

- Modelos de pronúncia: fazem a correspondência entre fonemas e partes de palavras. O trabalho de Rao et al. (2015) propõe um modelo "*grapheme-to-phoneme*" utilizando uma rede neural recorrente *Long Short-Term Memory* (LSTM), que apresentou significativa melhora de performance em relação aos modelos da época por considerar o contexto da sequência.

- Modelos de Linguagem: trata-se da distribuição de probabilidade ao longo de uma sequência de palavras. Permitem que as transcrições gerem sentenças que tenham sentido e sejam coerentes. O trabalho de Shewalkar (2019) apresenta uma avaliação da performance de redes neurais profundas aplicadas a reconhecimento de fala,

tratando de RNN (do inglês, *Recurrent Neural Networks*), LSTM (do inglês, *Long Short-Term Memory*) e GRU (do inglês, *Gated Recurrent Unit*).

A fim de treinar o modelo com todas essas componentes de maneira unificada, adota-se a camada *Connectionist Temporal Classification* (CTC) (GRAVES; JAITLY, 2014; ZHANG et al., 2019). O trabalho de Chan et al. (2015) propõe a arquitetura *Listen, Attend and Spell* (LAS), que é formada por uma rede neural recorrente piramidal e outra rede neural recorrente com o mecanismo de *Attention*. Tal modelo apresenta significativa melhora frente a modelos CTC anteriores.

Na literatura recente, algumas técnicas envolvendo dicionário fonético são aplicadas para melhoria do modelo de *Speech Recognition*. Çişman, Li e Tan (2017), por exemplo, propõem um *framework* de conversão por voz que utiliza informação fonética para construir um sub dicionário específico por indivíduo, melhorando, assim, o reconhecimento por voz.

Outras abordagens para modelos de ASR consistem em modelos de mistura gaussiana e sistemas de modelos ocultos de Markov (HE; DONG, 2020; TANAKA et al., 2018; BHABLE; KAYTE, 2020; DIAS et al., 2020).

2.3 Processamento de Linguagem Natural para Classificação de Ligações

No processamento de linguagem natural (NLP, do inglês *Natural Language Processing*), são aplicadas inicialmente técnicas de pré-processamento antes de utilizar os métodos de aprendizado de máquina. Essas técnicas têm o objetivo de homogeneizar o texto de forma a privilegiar performance e memória utilizados em futuras técnicas de pré-processamento.

O trabalho de Vijayarani et al. (2015) faz uma revisão geral de diversas técnicas pré-processamento, destacando-se as seguintes:

- *Tokenization*: processo de dividir um texto (*string*) numa lista de *tokens*. Essa divisão pode ser de texto para sentenças, sentenças em palavras ou até mesmo palavras em caracteres;
- Remoção de *stop words*: artigos, preposições são exemplos de *stop words* que podem ser eliminadas a fim de se manter apenas as palavras com maior relevância semântica. Elas podem ser removidas a partir de uma lista pré-definida ou segundo alguma lógica matemática;
- *Lemmatization* e *stemming*: são técnicas de normalização de textos, que permite extrair raiz das palavras, isto é, a forma não flexionada delas. O algoritmo de *stemming*

funciona retirando o início ou sufixo das palavras, enquanto a *lemmatization* leva em consideração a análise morfológica; e

- *Term Frequency-Inverse Document Frequency* (TF-IDF): estatística que visa quantificar a importância de uma palavra em um texto. O valor de TF-IDF aumenta proporcionalmente ao número de vezes que uma palavra aparece no documento, porém a frequência em que a palavra aparece no *corpus* é descontada. Isso evita que palavras de pouca relevância semântica, tais como *stop words*, apresentem altos valores.

Ao realizar a tokenização, uma representação vetorial dos dados é criada. Essa representação pode ser feita atribuindo-se valores a cada palavra, da qual obtêm-se uma matriz densa, ou binária. Nesses casos, trabalha-se com uma matriz esparsa. A categorização de palavras utilizando TF-IDF também resulta numa representação esparsa. A escolha entre esses dois métodos define como os dados serão armazenados e manuseados.

Arquiteturas de *Deep Learning* como CNN, RNN, LSTM foram formuladas para trabalhar com matrizes densas, dessa forma o *dataset* precisaria ser processado antes de sua aplicação, enquanto técnicas simples de classificação em árvore mostram uma melhora no desempenho com matrizes esparsas (SURESH; SARASWATHI; SUNDARARAJAN, 2010; LORENA et al., 2019). Logo, a montagem do *dataset*, é um ponto crucial para problemas de classificação de texto.

Schmitt et al. (2019) apresentaram uma análise na qual comparam diferentes bibliotecas disponíveis para o pré-processamento dos textos. No trabalho, os seguintes softwares foram testados: StanfordNLP, NLTK, OpenNLP, SpaCy e Gate. Dessas, apenas NLTK e SpaCy já possuem funções próprias para a língua portuguesa.

Na literatura, também foram encontrados trabalhos que tratam da engenharia de *features* aplicada a NLP. Bonin, Pedro e Oliver (2014) utilizaram conversas telefônicas transcritas para determinar quais palavras da transcrição representam as informações relevantes da ligação. Eles incluíram no modelo um *bag-of-words*, a frequência de nomes, partes do discurso e informações de contexto da ligação, como duração, data, localização. Para classificação, foi aplicado uma *Support Vector Machine* (SVM).

Logumanov, Klenin e Botov (2018) propuseram outra abordagem: além do texto transcrito, também utilizaram o áudio como *feature*. Dele foram extraídos dados como sequências de amplitudes do sinal, espectograma e características acústicas. Para o tratamento do áudio, o trabalho utilizou a biblioteca pydub, que está sendo considerada para o projeto.

Quanto a modelos, Imane e Mohamed (2017b) propõem uma solução para uma categorização *multi-label* de certificados de morte na França. Após o pré-processamento

composto pelas etapas citadas acima, foram comparados os resultados com três técnicas de categorização (*Decision Trees*, SVM e um Adaboost). Chen et al. (2018), por outro lado, aplica uma variedade de redes neurais para classificar os conjuntos de textos em um *dataset* em chinês. No trabalho, eles originalmente propuseram uma arquitetura para extrair *features* do n-grama. Também modificaram levemente modelos de RNN para extrair mais informação.

Da mesma forma, com uma abordagem *multi-label*, Chen et al. (2017) utilizou de redes neurais tanto convolucionais quanto recorrentes. Para o pré-processamento foi apresentado um tratamento com o modelo de word2vec, que considera as palavras contextuais como predições e utiliza desse contexto para extrair a semântica.

Uma abordagem que tem se tornado popular na academia é o *Bidirectional Encoder Representations from Transformers* (BERT). BERT é atualmente um dos modelos mais robustos na área de NLP, sendo capaz de obter resultados compatíveis com o estado da arte em diversas tarefas, como reconhecer entidades, classificar textos, responder perguntas, comparar frases e identificar palavras faltantes (ESMAEILZADEH; TAGHVA, 2022; DEVLIN et al., 2018; NUGROHO; SUKMADEWA; YUDISTIRA, 2021).

BERT é um modelo de redes neurais para processamento de linguagem que foi projetado para pré-treinar representações bidirecionais profundas de texto não rotulado e, depois disso, ser ajustado usando texto rotulado para uma ampla gama de tarefas, como responder a perguntas, inferência de idioma e classificação de texto (DEVLIN et al., 2018). Por ser baseado em *Transformer* (um complexo modelo de *Deep Learning*), o método de treinamento do BERT é mais facilmente paralelizável, tornando possível treiná-lo com uma massa de dados muito maior em muito menos tempo do que outros tipos de modelo dessa complexidade (SUN et al., 2019).

O aprendizado de transferência mudou a pesquisa de NLP ao permitir que um modelo treinado pela primeira vez em uma tarefa de origem fosse ajustado em outra. BERTimbau (SOUZA et al., 2020) é um exemplo: é uma versão do BERT para português brasileiro que também foi capaz de alcançar o estado-da-arte em diversas tarefas de NLP. O modelo foi pré-treinado com uma grande base textual conhecida como BrWaC (Brazilian Web as Corpus) (FILHO et al., 2018), o maior *corpus* aberto de português até hoje, com 2,68 bilhões de tokens e que contém textos retirados de mais de 3,5 milhões de páginas da web. Desse modo, o BERTimbau se caracteriza como um potente modelo de linguagem que consegue identificar o contexto em que as palavras se relacionam dentro de uma frase.

Além disso, esse modelo permite que seja facilmente realizado um *fine-tuning*, que é um processo de aprimoramento no qual o modelo é retreinado de uma maneira menos intensa, porém com dados mais específicos do cenário em que se deseja utilizá-lo. Assim, é possível fazer com que o BERTimbau seja ainda mais efetivo na solução de problemas nos quais o vocabulário utilizado seja mais característico, como é o caso do setor elétrico, e do

problema de classificação aqui abordado.

2.4 Correlação de Ligações

Chamadas telefônicas entre centros de operação e agentes de campo, no contexto de transmissão e distribuição de energia, podem fornecer informações úteis sobre as atividades executadas. Esta seção apresenta um sistema capaz de extrair e correlacionar informações de chamadas de voz e dados de sistemas supervisórios (SCADA, do inglês *Supervisory control and data acquisition*), com o objetivo de centralizar todas as informações relacionadas a um evento de campo, independentemente de sua origem, e diminuir o tempo gasto em análises de ocorrências que impactam diretamente na operação e no faturamento das transmissoras e distribuidoras de energia.

Na literatura, não são muitos os exemplos de correlação de transcrições de áudios com sistemas operativos, todavia, ao particionar o problema, é possível identificar exemplos como em Sager (1978) e Tekli (2016), onde há a conversão automática de textos para uma base de dados estruturada. Parte da solução aqui apresentada consiste em extrair dados estruturados da transcrição.

Ao abstrair a correlação de ligações em uma associação de dados não estruturados e dados estruturados, têm-se em Kirk et al. (2012) uma abordagem com métodos de clusterização aplicado à bioinformática, e em Zhang, Zhuang e Wu (2007) um exemplo de clusterização em um *dataset* com áudios e imagens.

3 Assistente Virtual para Centros de Operação

Este projeto foi desenvolvido objetivando a criação de uma ferramenta de análise operacional da comunicação, a partir da correlação inteligente entre dados e comandos por voz. A principal entrada de dados para a realização dessa tarefa são as gravações das conversas telefônicas e de rádio da equipe de operação do setor elétrico, juntamente com seus metadados. Aliados a isso, dados de sistemas SCADA, do Sistema de Acompanhamento de Usinas (SAU) e do Sistema de Apuração de Mudanças de Estados Operativos de Unidades Geradoras, Usinas e Interligações Internacionais (SAMUG) são utilizados para conectar gravações, mensagens e eventos.

Nesse capítulo serão apresentadas a arquitetura básica do assistente virtual proposto e a aplicação da mesma à três diferentes provedores de nuvem.

3.1 Arquitetura proposta

A arquitetura do projeto foi desenvolvida para ser agnóstica, isto é, independente do provedor de serviços de nuvem do usuário. Todavia, neste projeto, essa arquitetura agnóstica foi implantada em três provedores de serviço de nuvem, a saber *Google Cloud Platform* (GCP), *Amazon Web Services* (AWS) e *Azure Cloud Services*, o que validou a versatilidade da arquitetura.

Para implementação desse sistema foram propostas, inicialmente, duas arquiteturas em nuvem, com algumas partes do pipeline hospedadas *on-premise*.

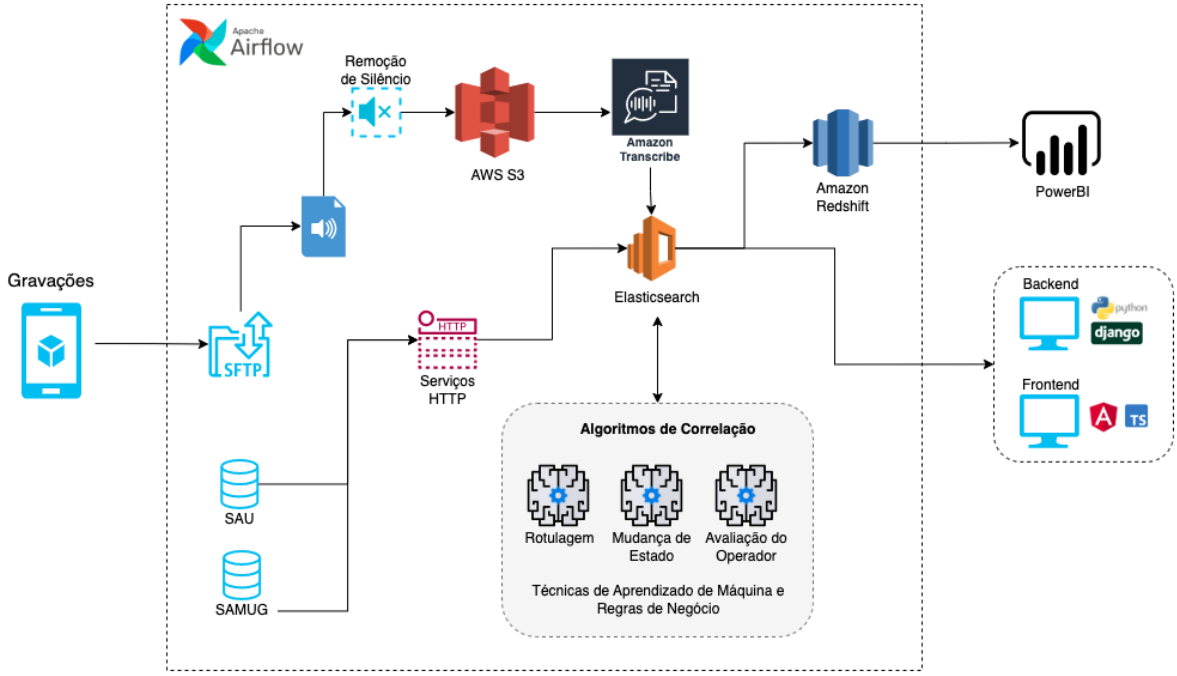
No projeto, com duas empresas clientes do ramo de energia, a formatação da arquitetura se deu de maneiras diferentes. Para a ENGIE, que tem como provedor de nuvem a AWS e possui serviços internos para exportar dados de sistemas operativos dos servidores locais para a *cloud*, foi desenvolvida uma arquitetura inteiramente na AWS.

Para CPFL, por questões de segurança e normas internas da equipe de Tecnologia da Informação (TI), o sistema tem uma arquitetura mista. Parte *on-premise*, parte *on-cloud*. Na formatação do projeto, o servidor de nuvem da empresa era o GCP, porém ao final do projeto foi realizada uma migração para a *Azure Cloud Services*, serviço de nuvem da Microsoft. As três abordagens são apresentadas a seguir.

3.1.1 Arquitetura na *Amazon Web Services*

Para a ENGIE o objetivo é o desenvolvimento do assistente virtual para centros de operação de geração de energia, que enquadra a solução para transcrição dos áudios, que compreende todo o processo de coleta, tratamento, transcrição dos áudios, classificação e correlação das transcrições com outros sistemas, busca através de filtros estruturados e analytics.

Figura 1 – Arquitetura do sistema na *AWS*



Fonte: Autora

Para adequar a arquitetura proposta à necessidade do cliente, deve-se propor um *pipeline* adequado: ingestão dos áudios, a identificação, tratamentos, transcrição, registro, correlações de dados, serviços web e *analytics*. A visão geral da arquitetura implementada na AWS é apresentada no diagrama da Figura 1.

De acordo com a segurança e boas práticas de desenvolvimento, foi criada uma comunicação segura entre o ambiente da ENGIE e o *Amazon S3* para ingestão dos áudios. Essa parte do *pipeline* se deu através da liberação de um servidor de SFTP (do inglês, *Secure File Transfer Protocol* – Protocolo de Transferência Segura de Arquivos) para a ingestão dos áudios.

Nessa etapa, os áudios passarão por um procedimento de remoção de silêncio, visando a redução dos custos de transcrição, associados a duração total do áudio, em segundos. Essa etapa transforma áudios com períodos de espera em conversações rápidas e remove chamadas que vão diretamente para caixa postal.

Após os tratamentos nos arquivos de áudio, é realizada a transcrição através do *Amazon Transcribe*. As transcrições são então indexadas no banco de dados não-relacional *Elastisearch* junto aos seus respectivos metadados (data, hora, ramais da ligação, usina da ligação e duração do áudio).

Após a indexação da transcrição, são desenvolvidos os seguintes módulos:

- Classificação automática das ligações em rótulos pré-definidos;
- Módulo de Mudança de Estado, que correlaciona ligações com os sistemas SAU e SAMUG;
- Módulo de Avaliação do Operador, que dá notas ligação por ligação, baseadas em normas do ONS, para os operadores;

O reslutado dos processamentos de dados nos módulos listados acima também são indexados no *Elasticsearch* e tem duas saídas para serviços clientes: a aplicação web e o *Amazon Redshift*, que conecta a base de dados com o Power BI.

Em resumo, os serviços da AWS que são utilizados na implementação da arquitetura são:

- EC2 – Serviço de aluguel de máquinas virtuais, que está sendo utilizado para manter os três servidores configurados: o de *Front-End*, de *Back-End* e do *Airflow*;
- S3 – Serviço de armazenamento de dados, que está sendo utilizado para armazenar os áudios das ligações gravadas pela ENGIE e os vocabulários personalizados, utilizados no processo de transcrição;
- Transcribe – Serviço de reconhecimento automático de fala (ASR) baseado em um modelo de aprendizado de máquina, que está sendo utilizado para realizar as transcrições dos áudios gravados;
- Redshift - Um *data warehouse* em nuvem totalmente gerenciado e escalável que acelera seu tempo para obter *insights* com análises rápidas, fáceis e seguras em escala; faz o meio entre o *Elasticsearch* e o Power BI;

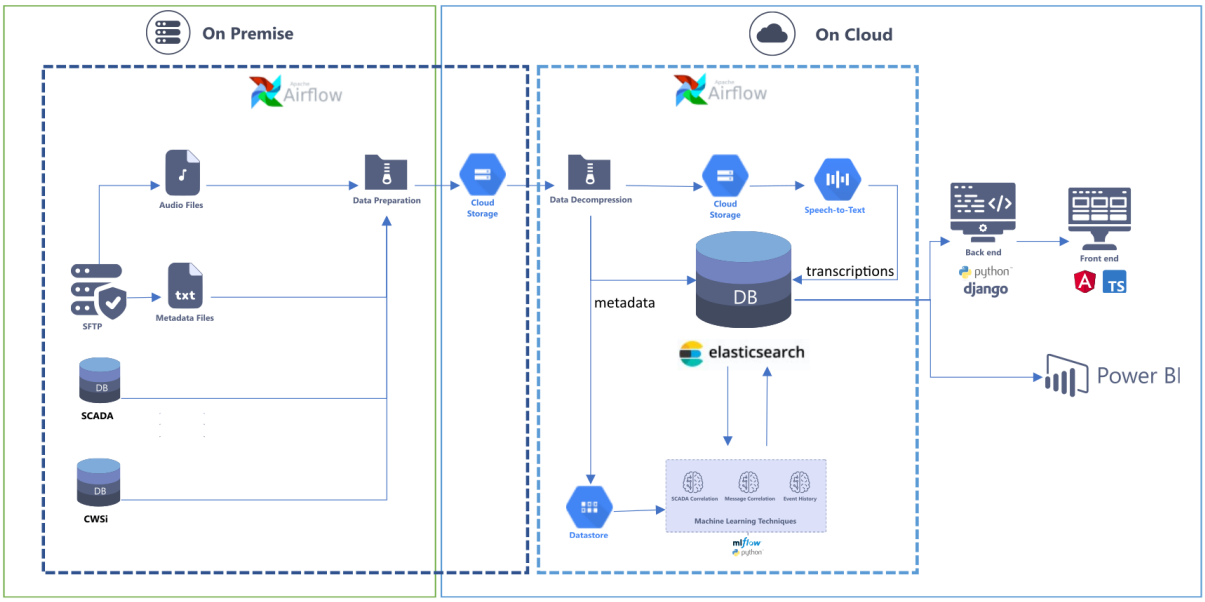
Todas as atividades descritas acima são orquestradas com o *Apache Airflow*, que é o orquestrador das tarefas desempenhadas pelo software a nível de dados. As tarefas são organizadas em DAGs (do inglês, *Directed Acyclic Graph*), que contêm fluxos de trabalho (sequência de realização de tarefas) (HAINES, 2022). Essas tarefas podem ser programadas para serem executadas periodicamente, de acordo com a necessidade da atividade. As DAGs implementadas para o funcionamento do *pipeline* descrito acima são:

- s3_to_es: recolhe os áudios do S3, aplica os tratamentos necessários e indexa os áudios e seus metadados no *Elasticsearch*;
- es_to_transcribe: envia os áudios tratados para a transcrição e indexa os resultados no *Elasticsearch*;
- classify: a partir do modelo de classificação previamente treinado, classifica os áudios nos rótulos pré-definidos; também realiza a avaliação do operador por ligação;
- mudanca_estado: extrai os dados do SAU e SAMUG através dos serviços HTTP e realiza a correlação de informações entre os dados dos dois sistemas e as ligações;
- retrain: realiza o retreino do modelo de classificação utilizado na DAG *classify*;

3.1.2 Arquitetura no *Google Cloud Platform*

Para a CPFL o objetivo é o desenvolvimento de um assistente virtual para auxílio à análise e tomada de decisão da equipe de operação a partir da utilização de gravações de chamadas telefônicas e de rádio e informações vindas de sistemas operativos, sendo esses os principais o sistema de mensageria/mobilidade e o sistema SCADA. O diagrama geral da arquitetura implementada na GCP é apresentado na Figura 2.

Figura 2 – Arquitetura do sistema no GCP



Fonte: Autora

Inicialmente, as gravações e os metadados são carregados na madrugada do dia seguinte a sua gravação em um servidor SFTP para que se tornem acessíveis ao *Airflow* instalado *on-premise*. Se tornando acessível ao *Airflow*, o processo se ramifica em dois ramos. No primeiro, relacionado às gravações, acontece o processo de separação de canais,

remoção de silêncio e compactação. No segundo, relacionada aos metadados, ocorre o processo apenas de compactação. Os dados do CWSi (serviço de mensageria) e do SCADA são acessados via query pelo *Airflow* e são compactados.

Uma vez compactados os dados, eles são carregados no *Google Cloud Storage* pelo *Airflow on-premise*. As tarefas realizadas pelo *Airflow on-premise* podem ser resumidas nos pontos abaixo:

- Remover o silêncio das gravações
- Realizar a separação de canais;
- Compactar os dados para enviá-los ao *Google Cloud Storage*;
- Realizar as consultas ao CWSI e ao SCADA;
- Enviar os dados (gravações compactadas com silêncio removido, metadados compactados, dados do CWSI e do SCADA compactados) ao *Google Cloud Storage*;

A remoção de silêncio se faz necessária devido a dois motivos principais: custos com transcrição e alocação de banda. O custo de transcrição é diretamente proporcional ao tamanho do áudio em minutos. Dessa forma, quanto maior o tamanho de um áudio, maior o seu custo de transcrição. Porém, existem diversos áudios tanto de ramal quanto de rádio que possuem grandes intervalos de silêncio entre as falas dos interlocutores. Esses intervalos de silêncio são também cobrados pelo transcritor. Dessa forma, removê-los diminui o custo de transcrição e diminui o tempo de *payback* do projeto. Com base nos dados atuais, a remoção de silêncio reduz em até 35% o custo com transcrição. O valor estimado da economia obtida através da aplicação de remoção de silêncio é de, aproximadamente, 80 mil reais por ano. Além disso, uma vez que a remoção de silêncio diminui o tamanho do áudio, ela também auxilia na economia de alocação de banda para enviar os dados do servidor *on-premise* para a *cloud*. Para a realização da remoção de silêncio, foi desenvolvida uma biblioteca própria, devido a escassez de bibliotecas para tal fim no setor.

A separação de interlocutores fez-se necessária por diversos motivos, dos quais os principais são fornecer uma maior compreensão do áudio e também uma melhora da qualidade da transcrição devido a característica do transcritor de áudios do Google. A separação de interlocutores pode-se dar de dois modos: via separação de canais (quando a gravação possui característica estéreo) e via separação de fala. A separação de canais é possível apenas para os áudios de ramal, onde a gravação é realizada de modo estéreo. Para as gravações do rádio, é necessário realizar a separação dos interlocutores via algoritmo de separação de fala. O algoritmo utilizado faz uso de técnicas estatísticas, de ciência de dados e de aprendizagem de máquina para dividir os áudios por interlocutores.

Após a separação de canais, os áudios são comprimidos e enviados para o *Cloud Storage*. Essa compressão é realizada visando poupar a banda da empresa durante o envio. Tanto os áudios quanto os dados do SCADA são comprimidos para envio.

A partir desse momento, todas as ações nos dados são performadas pelo *Airflow on-cloud*, instalado em uma máquina virtual dedicada. O *Airflow on-cloud* é responsável pelas tarefas descritas abaixo:

- Descompactar os dados e os armazenar em um outro *bucket* no GCP;
- Descompactar os dados do CWSI e do SCADA e os armazenar no *ElaticSearch* e *Datastore* do *Google*, respectivamente;
- Carregar metadados e transcrições no *Elasticsearch*;
- Realizar a reconstrução do áudio a partir de suas metades com silêncio removido;
- Enviar os áudios para transcrição;
- Executar os algoritmos de correlação (SCADA, CWSI e histórias) e carregar seus resultados no *Elasticsearch*;
- O *back-end* dos serviços clientes acessa então os dados armazenados no *Elasticsearch* e os fornece ao *front-end*, onde a autenticação de usuário é realizada. Para geração dos *dashboards*, os dados são acessados pelo PowerBI pela BigQuery;
- Carregar os dados do *Elasticsearch* no Big Query.

A transcrição dos áudios da operação é realizada através da API de transcrição *Speech-To-Text*, fornecida pela suíte de ferramentas do GCP. Para a realização da transcrição dos áudios, um vocabulário personalizado é inserido com fim de alterar a probabilidade de identificação de palavras-chave da operação, estando estas relacionadas à operação e ao contexto de distribuição e transmissão de energia.

Em seguida, a transcrição é indexada e armazenada no *Elasticsearch*. Esse banco de dados permite que a indexação das transcrições e os filtros funcionem de maneira adequada. Uma vez armazenadas as transcrições, os algoritmos as correlacionam com eventos reportados no sistema de mensageria/mobilidade e com eventos registrados no sistema SCADA. Esse algoritmo é conhecido como Correlação. Também são realizadas a classificação automática dos áudios e a avaliação da comunicação do operador. Para isso, a partir das transcrições e um conjunto de regras definidas pelo ONS, o operador é avaliado e uma nota é conferida à sua comunicação. Por fim, KPI's da aplicação são apresentados no PowerBI, como duração das gravações, quantidade de palavras chaves e outros.

Para a implementação desse *pipeline*, a ferramenta possui uma arquitetura mista, sendo parte *on-premise* e parte *on-cloud*. A parte *on-premise* recebe os dados (áudios, metadados, dados dos sistemas operativos) e aplica diversos tratamentos antes de enviá-los à nuvem. Na parte *on-cloud* do sistema é onde toda a parte de ciência de dados ocorre. O serviço de nuvem utilizado pela ferramenta até a dada da escrita deste documento é o *Google Cloud Plataform*. Porém, a arquitetura foi pensada de modo a ser portável para qualquer provedor de nuvem desde que este possua, pelo menos, a possibilidade de criação de máquinas virtuais com escalamento vertical e horizontal. Para a realização dessa portabilidade é necessário que testes sejam realizados de modo a garantir o correto funcionamento do transcritor e para garantir que as métricas dos módulos da ferramenta continuem dentro dos patamares aceitáveis.

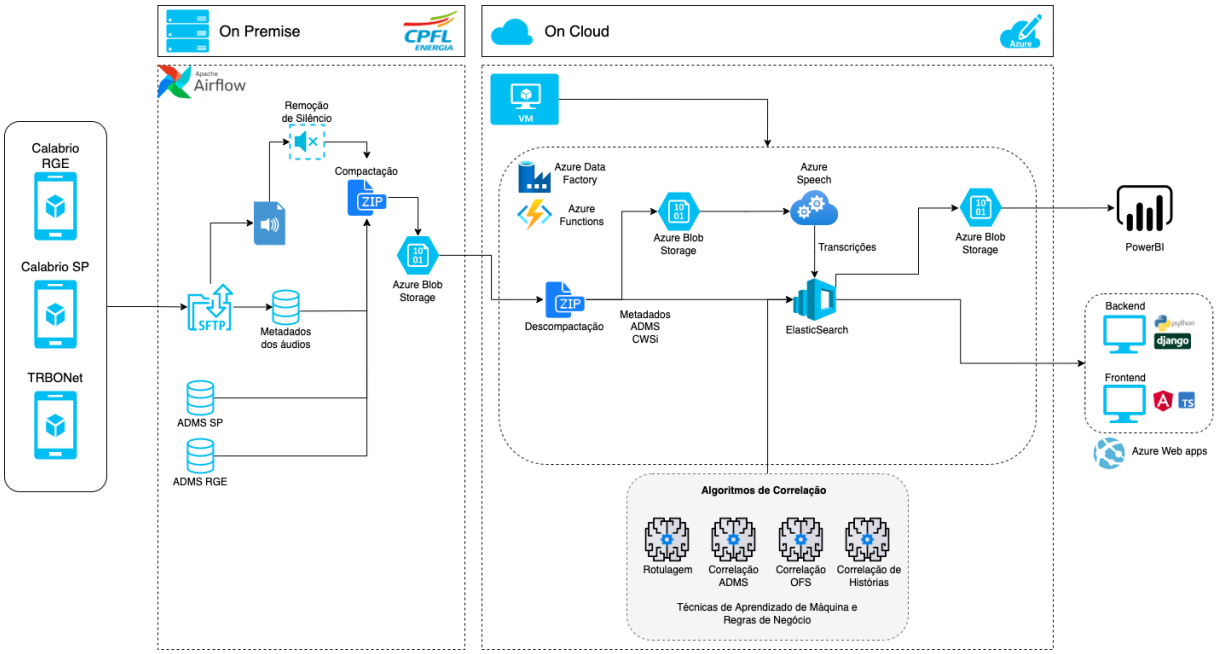
Os principais serviços e tecnologias utilizados são listados abaixo:

- *Apache Airflow*: é o orquestrador das tarefas desempenhadas pelo software a nível de dados. As tarefas são organizadas em DAGs, que contém fluxos de trabalho (sequência de realização de tarefas). Existem duas aplicações *Airflow* sendo utilizadas. Uma *on-premise* no servidor fornecido pela CPFL. Outra *on-cloud* hospedada em uma máquina virtual no GCP.
- *ElasticSearch*: Banco de dados utilizado para armazenar dados semiestruturados, como dados textuais; possui estrutura distribuída e armazena dados semelhantes sob o mesmo índice; Está entre os bancos de dados mais rápidos já conhecidos para busca textual;
- *Google Datastore*: banco de dados NoSQL altamente escalonável. Controla automaticamente a fragmentação e a replicação para fornecer um banco de dados altamente disponível e com muita durabilidade, que é escalonado de forma automática. Utilizado para armazenar metadados e informações do SCADA e do sistema de mensageria;
- *Google Cloud Storage*: é um serviço da Web de armazenamento de arquivos online RESTful para armazenar e acessar dados na infraestrutura do GCP. Utilizado para armazenar os arquivos compactados e as gravações descompactadas que serão transcritas.
- *Google Speech-to-Text*: ferramenta de transcrição de áudios do *Google* utilizada para a realização da transcrição das gravações.
- Django: *framework* para desenvolvimento web em python; utilizado para desenvolvimento do *back-end* da aplicação.
- Typescript e Angular: plataforma de aplicações web de código-fonte aberto e *front-end* baseado em TypeScript; utilizada para desenvolvimento do *front-end* da aplicação.

3.1.3 Arquitetura na *Azure Cloud Services*

Como mencionado, a implementação da arquitetura na Azure foi resultado de uma necessidade de migração do GCP para a provedora de nuvem da Microsoft. Logo, ambas se assemelham bastante. A primeira etapa do pipeline, que é realizada *on-premise* permanece a mesma, por exemplo, com a diferença de que agora os dados são enviados para um *Azure Blob Storage*. O diagrama completo da arquitetura na *Azure* é exibido na Figura 3.

Figura 3 – Arquitetura do sistema na *Azure*



Fonte: Autora

A parcela do pipeline que é realizada em nuvem sofreu alterações que tornaram a arquitetura mais orientada a serviços. Na *Azure*, o *Elasticsearch* e o *front* e *back-end* passaram a ser serviços. O *Elasticsearch*, que antes era hospedado num servidor, passa a ser um serviço; e as aplicações web, previamente também em servidores, se transformaram em *Azure Web Apps*. O orquestrador de tarefas, antes Airflow, é agora substituído pelo conjunto do *Azure Data Factory* (ADF) e *Azure Functions*.

Os principais serviços e tecnologias diferentes dos utilizados na arquitetura GCP são listados abaixo:

- *ElasticSearch*: Banco de dados utilizado para armazenar dados semiestruturados, como dados textuais; possui estrutura distribuída e armazena dados semelhantes sob o mesmo índice; está entre os bancos de dados mais rápidos já conhecidos para busca textual; a instância utilizada é a disponibilizada na Azure;
- *Azure Data Factory* (ADF): funciona como o orquestrador para as rotinas realizadas em nuvem; a partir dos pipelines gerados pelo ADF é que são executadas as operações

com dados da ferramenta, do recebimento e descompressão dos dados, a transcrição e os algoritmos de correlação;

- *Azure App Services*: Serviço totalmente gerenciado com manutenção de infraestrutura integrada, correção de segurança e dimensionamento. Tanto o *front-end* quanto o *back-end* estão hospedados no *App Service*;
- *Azure Function App*: O *Azure Functions* é um serviço de nuvem disponível sob demanda que fornece toda a infraestrutura e os recursos continuamente atualizados necessários para executar os aplicativos. Os pipelines gerenciados pelo ADF são executados através do *Azure Function*; seriam o análogo as DAGs no *Airflow*;
- *Azure Blob Sotrage*: armazenamento de dados do sistema; aqui são armazenados áudios, modelos da rotulagem, dados para o BI e auxiliam no recebimento de dados SCADA e CWSi, que posteriormente são descartados e armazenados no *Elasticsearch*;
- *Azure Speech Services*: serviço responsável pela transcrição dos áudios; permite retreinamento e criação de *endpoint* próprio;
- *Azure Container Registry*: permite a criação e gerenciamento de imagens e artefatos de contêiner com uma instância de distribuição OCI totalmente gerenciada e georreplicada; no *container registry* estão as imagens utilizadas no *App Service Plan* e que executam as *Azure Functions*;
- *Azure App Service Plan*: define um *set* de recursos computacionais para aplicações utilizarem; é utilizado tanto para os *Web Apps*, quanto para as *Azure Functions*;

3.2 Transcrição automática de áudio

Todavia, para o português brasileiro, as abordagens com *deep learning* acabam por encontrar o desafio da quantidade de dados de treinamento. E no contexto desse projeto, limitação de tempo para levantamento de um *corpus* e limitação de recursos para treinamento do modelo foram fatores que balizaram a escolha para os provedores do serviço de *speech-to-text* em nuvem: *AWS*, *Google* e *Microsoft Azure*.

Essas plataformas acabam por ter uma imensa vantagem quanto aos modelos estado-da-arte citados acima: a disponibilidade de dados e a fácil configuração do modelo na arquitetura do projeto. Com auxílio de celulares e assistentes pessoais (*Alexa, Siri, Google Assistant*), os modelos em nuvem têm uma grande variedade de dados utilizados para treinamento; e alguns ainda permitem o retreino com *corpus* customizados, vocabulários customizados e vocabulário fonético.

Nessa seção serão apresentados os modelos de transcrição em nuvem avaliados: *Amazon Transcribe*, *Google Speech-to-Text* e *Azure Speech*.

A quantidade de áudios utilizados para otimização de desempenho dos modelos varia de acordo com a natureza dos áudios. A quantidade dos áudios avaliados na AWS engloba o universo de ligações transcritas manualmente para ENGIE, no contexto de geração; enquanto que para *Azure* e GCP, são consideradas as ligações no contexto da CPFL.

3.2.1 Implementação na *Amazon Web Services*

O *Amazon Transcribe* é o serviço de reconhecimento automático de fala da AWS (do inglês, *Amazon Web Services*). Os recursos do *Transcribe* para português permitem a transcrição de áudios com um vocabulário personalizado. Nesse vocabulário são listadas as palavras-chave de importância para operação, os nomes de usina e os nomes de operadores.

Essa customização funciona como um pós-processamento da transcrição, já que o modelo de transcrição da AWS não prioriza palavras ou expressões de acordo com esse vocabulário. Todavia, após a transcrição, são substituídas as palavras, siglas e expressões equivalentes (AMAZON, 2022).

No formato de vocabulário da AWS, é possível discriminar as palavras-chave de 4 formas:

1. *Phase*: palavra-chave, sigla ou expressão;
2. *SoundsLike*: descrição sílaba-a-sílaba da pronúncia da palavra;
3. *DisplayAs*: como o usuário deseja que a palavra seja transcrita;
4. *IPA*: descrição da palavra de acordo com o alfabeto fonético internacional;

Nos vocabulários testados e apresentados nessa sessão foram utilizados os três primeiros campos descritos acima.

Para as avaliações, foi utilizada uma amostra de 1000 gravações. Esses áudios foram, inicialmente, transcritos manualmente e em seguida transcritos repetidas vezes com o *AWS Transcribe*, cada uma com uma modificação no vocabulário personalizado. O objetivo dessa análise era comparar as métricas das transcrições alimentadas com cada vocabulário com o *groundtruth*, e a cada modificação, acompanhar o desempenho do transcritor, de forma a otimizar o vocabulário para o um REC máximo e WER mínimo.

Duas abordagens foram testadas na customização do *AWS Transcribe*: um único vocabulário para todas as ligações, o V0, e múltiplos vocabulários (um para cada tipo de usina/centro de operação), do V1 ao V5.

Dada a diferença de vocabulário nas ligações de diferentes tipos de usina e visando diminuir a incidência de falsos positivos, as ligações foram associadas a vocabulários

distintos de acordo com tipos de usinas e centros de operação, a partir dos ramais envolvidos nas gravações. Dessa forma, ao invés de um único vocabulário para todos os áudios, são propostos 5 vocabulários, sendo eles:

1. Hidrelétricas;
2. Termelétricas e biomassas;
3. COG;
4. Fotovoltaica e eólica; e
5. Geral (ONS e ligações não-classificadas).

Como os termos utilizados nas usinas termelétricas e de biomassa são compatíveis, ou seja, muitas palavras intersectam os dois conjuntos, as duas entram no mesmo vocabulário customizado, a mesma regra foi aplicada nas usinas fotovoltaicas e eólicas. A lista de palavras em "Geral" é a união de todos os vocabulários.

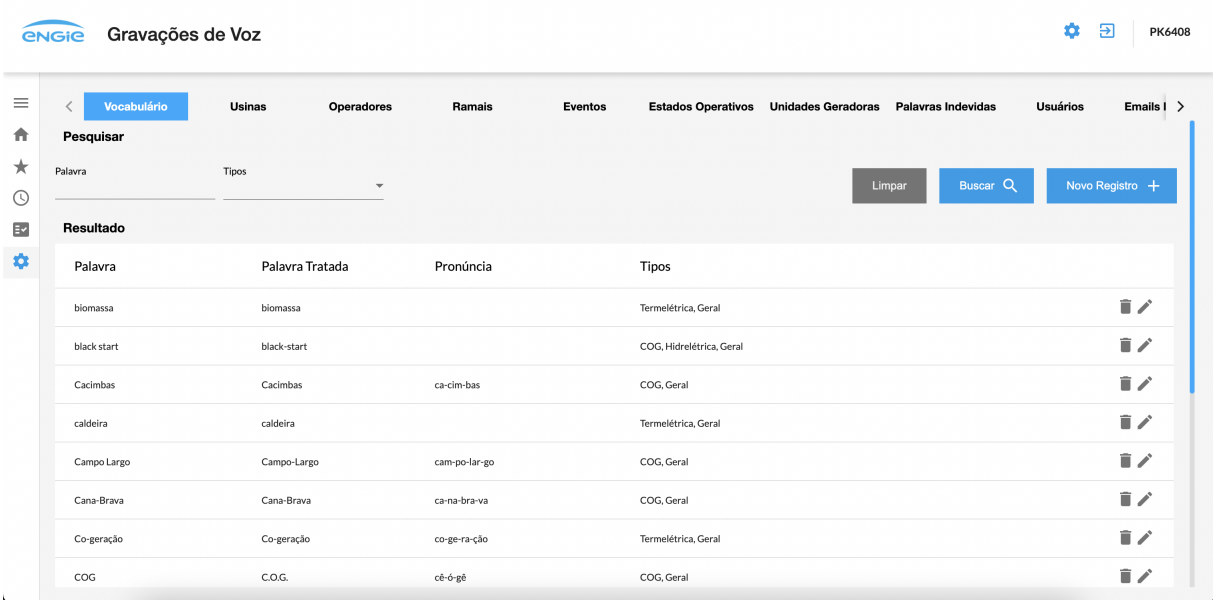
Dessa forma, "Lacerda" (usina eólica) está presente no vocabulário customizado apenas das ligações entre eólicas e COS ou entre ONS e COS, por exemplo.

A partir dos resultados encontrados nas análises, foram propostas novas versões do vocabulário. Visto que a porcentagem de reconhecimento de nomes de operadores estava muito abaixo da expectativa, chegando a ter operadores com 0% de reconhecimento, foram incluídos no vocabulário fonético da AWS os nomes dos operadores do COS. Assim temos:

- Vocabulário V0: um único vocabulário para ligações;
- Vocabulário V1: cinco vocabulários, separados por ramal;
- Vocabulário V2: cinco vocabulários, separados por ramal, com nomes de operadores inclusos;
- Vocabulário V3: cinco vocabulários, separados por ramal, com nomes de operadores inclusos e testes com 'COS Engie, [nome de operador]';
- Vocabulário V4: cinco vocabulários, separados por ramal, com nomes de operadores inclusos e testes com pronúncias distintas de nomes de operadores;
- Vocabulário V5: cinco vocabulários, separados por ramal, com nomes de operadores inclusos e mais testes com pronúncias distintas de nomes de operadores;

Os resultados dos experimentos com cada vocabulário e a comparação da porcentagem de reconhecimento de palavras-chave e de falsos positivos é apresentada na seção 4.1.1. A Figura 4 exibe a tela de configuração de vocabulário.

Figura 4 – *Print* da ferramenta na tela de configuração de vocabulário



Fonte: Autora

3.2.2 Implementação no *Google Cloud Platform*

As gravações e suas respectivas transcrições de áudio desempenham um papel importante para o projeto e para determinar a assertividade da correlação. Na operação da CPFL, as ligações dos diversos Centros de Operações e Empresas possuem peculiaridades diferentes.

O Google Cloud Platform (GCP), no serviço de Speech-to-Text (STT), permite uma customização de vocabulário voltada para adaptação, visando aumentar a assertividade da transcrição. A ferramenta chamada de *boost*, consiste em aplicar um peso numérico a palavras e/ou frases. Dadas as regras de negócio e a presença de palavras-chave, foram realizadas análises para compreender a influência do *boost* na assertividade da transcrição (GOOGLE, 2022).

Para compreender a influência da ferramenta na transcrição dos áudios, foram comparadas transcrições manuais (*groundtruths*) com transcrições realizadas pelo GCP com diferentes níveis de *boost*. O vocabulário personalizado para os testes, assim como os áudios, foi separado por Centros de Operação:

- COS SP/COI Paulista/ COI Santa Cruz;
- COI/COS RGE;
- COI Piratininga.

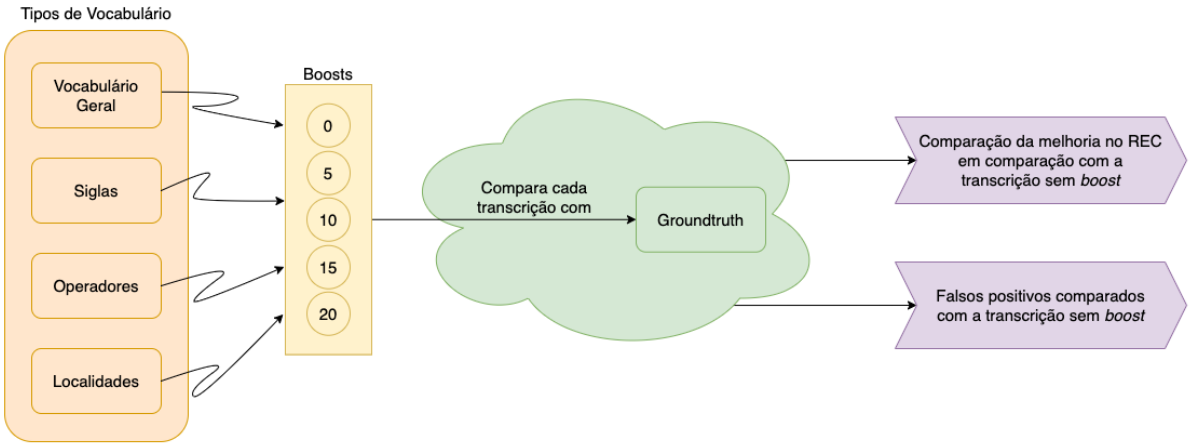
No conjunto, cada centro de operação, portanto, possui seu vocabulário específico.

As métricas de WER, FP e REC, já definidas anteriormente, foram os principais parâmetros para, a partir dos resultados, interpretar qual o melhor *boost* para cada tipo de vocabulário. Vocabulários esses que poderão ser:

- Vocabulário Geral: equipamentos e dispositivos elétricos e eletrônicos e jargões;
- Localidades: cidades, regiões e nomes de subestações;
- Operadores: nomes dos operadores de tempo real;
- Siglas: abreviações comumente citadas nas ligações.

A análise consistiu, como mostra a Figura 5, em transcrever, cada vocabulário - e fonte de áudio - por vez, com diferentes *boosts*. Como a API do GCP permitia valores entre 0 e 20 (PLATFORM, 2021), as palavras-chave foram transcritas com cinco valores distintos: 0, 5, 10, 15 e 20. Apenas esses *boosts* foram testados devido a limitações de custos no projeto. Já que para cada teste, um universo de 2500 ligações eram transcritas.

Figura 5 – Diagrama dos testes de vocabulário no *GCP*



Fonte: Autora

A CPFL é responsável por 14% da distribuição de energia do Brasil e está presente em quatro estados, logo tornou-se necessário reconhecer a extensão da empresa, com sotaques, expressões e vocabulários diferentes. Assim, o vocabulário *Locations* foi dividido em regiões, a saber: *Piratininga*, *Paulista*, *Santa Cruz* e *RGE*. Esses vocabulários também levaram em consideração qual a fonte de áudio da chamada de voz, seja rádio ou ramal telefônico (PABX).

Como a correlação entre as transcrições de áudio e eventos SCADA é construída a partir de palavras-chave, regiões e subestações identificadas no texto, a melhoria de reconhecimento obtida neste trabalho desempenhou um papel importante para o sucesso do projeto.

3.2.3 Implementação na *Azure Cloud Services*

Os serviços cognitivos da *Azure* são um conjunto de modelos de IA pré-criados e personalizáveis para adição de inteligência computacional a aplicações. Fala, decisão, linguagem e visão são alguns dos domínios disponíveis na plataforma.

O serviço de conversão de voz em texto da plataforma é capaz de analisar áudios em tempo real ou em lotes para converter as palavras faladas em texto. Essa conversão utiliza um modelo de linguagem universal como seu modelo base, que foi extensivamente treinado em dados de propriedade da *Microsoft* e reflete a linguagem comumente utilizada em diversos idiomas.

Este modelo base foi pré-treinado com dialetos e fonéticas que representam uma ampla gama de domínios comuns, assim sendo adequado para diversas aplicações sem necessidade de adequação de contexto.

O *Speech-to-Text* da *Azure* possui algumas formas de customização:

- Treinamento de modelo com áudio + *groundtruth* (até 1k horas de áudio);
- Treinamento com texto sem formatação (para aproveitamento de vocabulário);
- Treinamento com dados de pronúncia (útil para siglas, por exemplo)
- Treinamento com lista de frases (localidades, nomes, homônimos, palavras-chave e acrônimos)

Essas customizações podem ser utilizadas de forma isolada ou em conjunto para obtenção de um melhor resultado. As tarefas voltadas para customização consistem em testar a ferramenta e as possíveis customizações da maneira mais exaustiva possível, de forma a compreender melhor o funcionamento de cada funcionalidade e otimizar o resultado (AZURE, 2022).

Apesar da quantidade de recursos disponibilizados pela Azure, para essa aplicação foram realizados poucos testes por limitações de tempo e recursos. A *feature* utilizada foi o treinamento com lista de frases. Como a implementação do sistema com o transcritor Azure foi uma adaptação da já realizada no GCP, foi aproveitado todo o vocabulário especificado anteriormente. As análises realizadas consistiram em comparar o desempenho do transcritor final do GCP com: (i) transcritor padrão da Azure e (ii) transcritor treinado com vocabulário da Azure.

A plataforma de transcrição oferece a possibilidade de customização de vários modelos *baseline* exaustivamente treinados com dados proprietários da *Microsoft*. A análise aqui apresentada se baseou no treinamento de quatro modelos *baselines* com as palavras do vocabulário adaptado. Os resultados são descritos na secção 4.1.3.

3.3 Classificação de Ligações

Encontrar ligações que são relacionadas a determinados tipos de eventos é um trabalho difícil, uma vez que faz necessária a leitura das transcrições e a busca por palavras chaves relacionadas ao dado tipo de evento. Para facilitar essa busca, foi proposto um módulo capaz de classificar as transcrições em tipos de eventos. Esses tipos de eventos representam o dia a dia dos centros de operação de geração. De tal modo, o módulo da rotulagem consiste na classificação das ligações em grupos por meio de modelos de inteligência artificial.

As gravações e suas respectivas transcrições de áudio desempenham um papel importante para o projeto. Na operação da ENGIE (geradora), as ligações entre os centros de operação, as usinas e o Operador Nacional do Sistema Elétrico (ONS) podem tratar de diferentes processos, assuntos e atividades. Nesse contexto, as ligações foram classificadas em seis categorias:

1. Modulação de Geração;
2. Mudança de Estado;
3. Ajuste de Tensão;
4. Nível de Reservatório;
5. Chuva;
6. Operação & Manutenção (O&M Local).

Áudios classificados como *Mudança de Estado* devem conter uma solicitação de mudança de estado, a confirmação/comunicação da mesma (com o horário da manobra) ou a confirmação de programação. Citações na transcrição de Documentos de Intervenção (DI) também é uma indicação desse rótulo. Palavras-chave que sugerem uma Mudança de Estado são: “compensar”, “reverter para gerador/gerando”, “parado”, “desligamento”, “comutação”, entre outras.

Áudios classificados como *Ajuste de Tensão* são aqueles que indicam solicitação de ajuste de tensão ou ajuste de reativo, sempre associados a unidade *Volts*, seja citando kV, volts ou mega-volts. Palavras-chave que sugerem um Ajuste de Tensão são: “aumentar”; “reduzir”; “tensão”; “geração”; “limitar”; referências a unidade de tensão (Volts); entre outros.

Já os áudios classificados como *Modulação de Geração* englobam as solicitações de ajuste de potência. Semelhante ao Ajuste de Tensão, entretanto trata com outra unidade de medida, o Watt, podendo ser kilo-watt ou mega-watt.

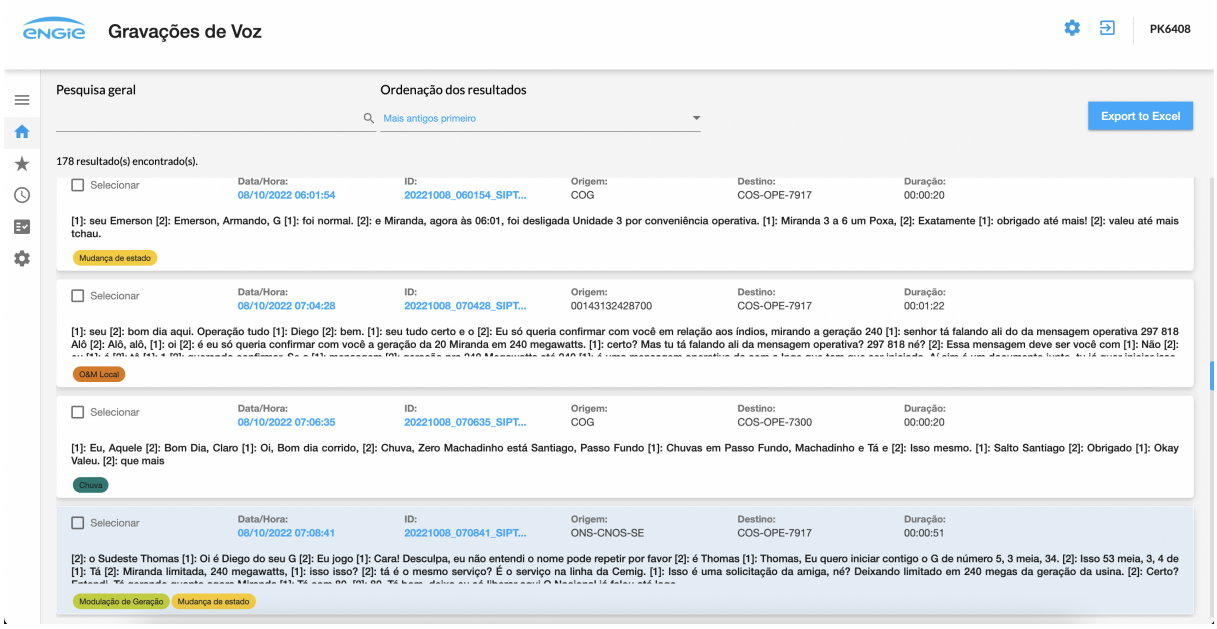
Os rótulos *Chuva* e *Nível de Reservatório* indicam situações relacionadas ao clima e a reservatórios de usinas hidrelétricas, tratando de chuvas, vazão, nível e vertimento nas ligações.

Em O&M estão as ligações menos recorrentes, trata-se de gravações que aconteçam dentro da operação da ENGIE, seja entre a usina, COG ou mesmo COS. Usualmente, são casos entre o centro de operação da usina e o campo/operação com requisição de informação, manutenção e testes. Não se enquadrando em nenhum dos rótulos, não é atribuída nenhuma classe ao áudio.

Por se tratar de um problema em que mais de um rótulo pode ser atribuído a um mesmo áudio, foi necessária uma abordagem *Multilabel*.

Para o problema descrito, foi adotada uma abordagem supervisionada, logo foi necessário a rotulagem manual de ligações para fornecer informação para o modelo. Foram rotuladas cerca de 3 mil ligações. Na Figura 6, é exibida a ferramenta com as ligações rotuladas automaticamente.

Figura 6 – *Print* da ferramenta na tela de ligações com os áudios classificados



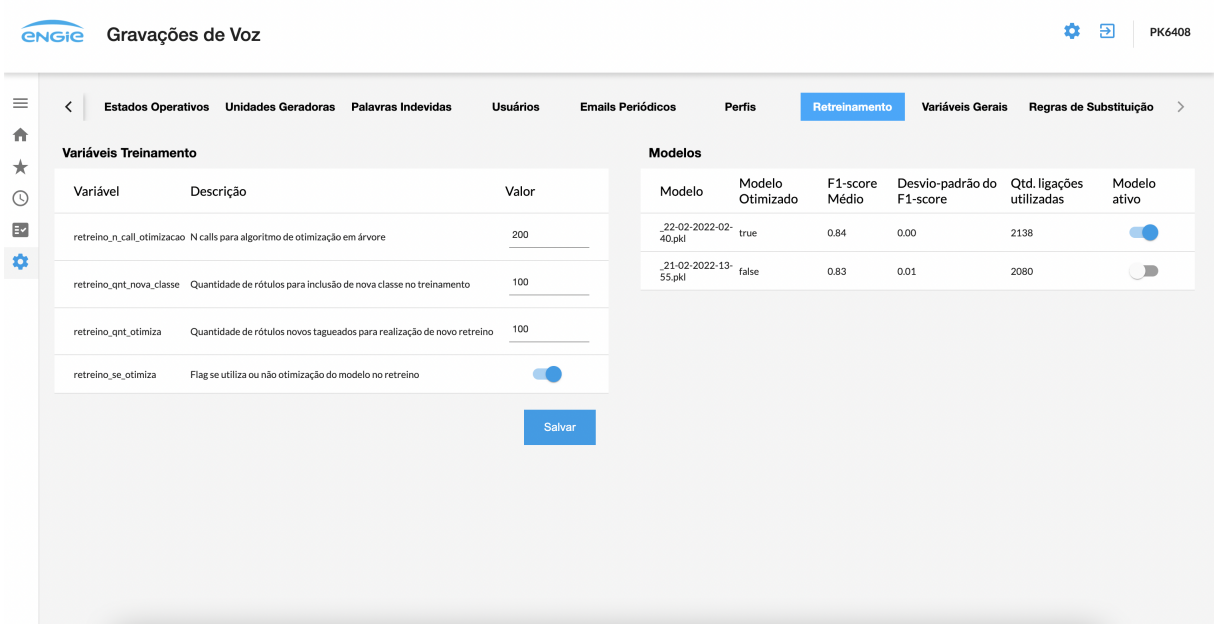
Fonte: Autora

Na plataforma, o usuário também é capaz de corrigir uma classificação incorreta do algoritmo, criar e rotular novas classes. Dada essa possibilidade, a plataforma também conta com o retreinamento do modelo de aprendizado nesse módulo. O algoritmo de retreinamento permite a inserção de aprendizado nos modelos, na ocorrência de novas rotulagens manuais. Além disso, ele permite corrigir a degradação do modelo devido a mudanças no padrão de comunicação e nas descrições dos eventos. Para realizar o retreinamento automático, é utilizada uma regra baseada em número de novas rotulagens.

Quando determinado número de novas rotulagens é atingido, o modelo é retreinado. Durante o retreinamento, o modelo passa por um processo de otimização de parâmetros, de forma a garantir o melhor desempenho na validação cruzada. Após o retreinamento, o desempenho do modelo atual é comparado ao desempenho do modelo antigo. Caso o desempenho do modelo atual seja melhor, ele substitui o modelo antigo que é então armazenado no banco de dados. Caso o desempenho seja pior, o modelo anterior é mantido e o modelo recém treinado é apenas armazenado no banco de dados.

Além disso, é permitido que o usuário configure tanto os parâmetros do treinamento e retreinamento do modelo, como também que escolha manualmente qual modelo que deve estar ativo. Na Figura 7, é exibida a tela de configuração do modelo. Os parâmetros serão descritos nas seções seguintes.

Figura 7 – *Print* da tela de configurações dos parâmetros de retreinamento



Fonte: Autora

Logo, o modelo de classificação de ligações consiste na rotulagem automática de ligações a partir de um modelo de aprendizado de máquina supervisionado. A proposta ainda permite o retreinamento do modelo à medida que novos áudios são rotulados manualmente pelo usuário.

Na sequência será apresentada a comparação entre três abordagens para o problema de classificação de textos: (i) *Baseline* de modelos com palavras-chave e informações das ligações como *features*; (ii) modelos com TF-IDF; (iii) modelos baseados em BERT. O *pipeline* e os desempenhos de cada uma das abordagens serão comparados e apresentados a seguir.

3.3.1 Classificação com palavras-chave e informações das ligações como *features*

O estabelecimento de um *baseline* é importante para que se tenha noção do quanto (e o quê) é possível melhorar num modelo. Portanto, foi utilizado um modelo simples, com o mínimo de tratamento de dados e sem otimização de parâmetros. Para esse modelo foi utilizado uma amostra com 2278 *labels*. A distribuição das classes no *dataset* utilizado é apresentada na Tabela 1.

Tabela 1 – Distribuição das classes no *dataset*

Ajuste de Tensão	Chuva	Modulação de Geração	Mudança de Estado	Nível de Reservatório	O&M Local
616	139	436	682	122	342

Fonte: Autora

As variáveis de entrada do *baseline* consistem em metadados associados ao áudio e recursos extraídos das transcrições. As informações disponíveis são:

- Transcrição dos áudios;
- Data/hora da ligação;
- Duração da Ligação;
- Ramal de Origem (podendo ser associado a um centro de operação ou usina);
- Ramal de Destino (podendo ser associado a um centro de operação ou usina);

A partir das transcrições das ligações foram extraídas citações de palavras-chave e usinas. Nesse contexto, foram avaliadas duas situações: recursos binários, considerando a presença ou não no texto, e recursos de contagem, considerando a frequência com que aparecem na transcrição. No *baseline*, não foram utilizados recursos de texto *bag-of-words*, como TF-IDF, testes com o TF-IDF serão descritos posteriormente. As informações que se tornaram *features* do *baseline* foram:

- Metadados temporais: dia da semana, hora do dia e duração da chamada em segundos;
- Metadados relacionados à origem e ao destino: pelo número do ramal, é possível identificar os centros de operação ou usinas envolvidos na chamada;
- Palavras-chave citadas (contagem ou binário): foi definido um conjunto de palavras que muitas vezes são faladas para comunicar as ações definidas pelos rótulos;

- Usinas citadas (contagem ou binário): foi definido uma lista de plantas relacionadas aos centros de operação analisados.

Como o problema consiste em uma classificação multi-classe, ou seja, o resultado esperado na saída dos modelos é um vetor binário de 6 posições (0 ou 1 para ocorrência de cada classe), algumas abordagens podiam ser testadas, entre elas:

- Treinamento com múltiplos modelos (um para cada classe), e depois a união dos resultados: cada um dos 6 modelos traduz a ocorrência de uma única classe e os resultados são então concatenados num vetor de 6 posições;
- Treinamento com um único modelo para as seis classes: o modelo já tem como saída um vetor binário de 6 posições;

Ambas as abordagens têm vantagens e desvantagens, o treinamento de múltiplos modelos, no qual é treinado um modelo para cada rótulo, se destaca por permitir a melhora de uma classe em específico que esteja desempenhando mal. O treinamento único, por outro lado, pode utilizar a saída de um modelo em outro, permitindo que a classificação aprenda, por exemplo, que uma modulação de geração costuma ser relatada em ligação com uma mudança de estado.

A partir do *baseline*, foram realizados testes com AutoML (HE; ZHAO; CHU, 2021). Experimentos com TPOT (*Tree-based Pipeline Optimization Tool for Automating Data Science*) (LE; FU; MOORE, 2020) em conjunto com modelos com *ClassifierChain* (LEARN, 2020a) obtiveram bons resultados. O *ClassifierChain* permite que os resultados dos primeiros modelos treinados sejam *features* de entrada para os modelos seguintes.

A partir das melhorias alcançadas com o *ClassifierChain*, foi proposta a otimização dos hiperparâmetros utilizados no mesmo. Tanto no próprio classificador em cadeia como no estimador base utilizado, o *RandomForest*.

A otimização utilizada consiste num modelo sequencial em árvore. Esse modelo trata o problema de otimização de parâmetros como um problema de regressão, onde o objetivo é encontrar o valor ótimo da função de custo utilizada para um conjunto de parâmetros escolhidos.

Para a implementação dessa otimização, foi utilizado o framework *scikit-optimize* (LOUPPE, 2017), composto por um conjunto de modelos para otimização sequencial.

A otimização de parâmetros proposta foi combinada com a abordagem do classificador em cadeia com a adição de um novo parâmetro: a ordem das classes a serem treinadas no *ClassifierChain*. Assim, permitindo que o processo de otimização encontre a sequência ótima para o treinamento dos rótulos. Os hiperparâmetros otimizados foram:

- *n_estimators*: Número de árvores no *Random Forest*;
- *max_depth*: Profundidade máxima da árvore;
- *max_features*: Número de *features* a serem consideradas para o melhor *split*;
- *min_samples_leaf*: Mínimo de amostras requeridas para um nó;
- *min_samples_split*: Mínimo de amostras requeridas para um *split* em um nó;
- *bootstrap*: Se amostras *bootstrap* serão utilizadas na árvore;
- *class_weight*: Ou 'balanced' ou 'balanced_subsample';
- *order_classes*: Ordem das classes na cadeia do classificador;

Por fim, foi executado o pipeline descrito acima com 500 *n_calls* no otimizador em árvore. O resultado final é apresentado na Secção 4.2.1.

Na ferramenta, como demonstrado na Figura 7, é permitida a customização de algumas variáveis do treinamento pelo usuário:

- *retreino_n_call_otimizacao*: Quantidade chamadas na otimização em árvore;
- *retreino_qnt_nova_classe*: Quantidade de ligações rotuladas com uma classe para inclusão da mesma no treinamento;
- *retreino_qnt_otimiza*: Quantidade de novas ligações classificadas manualmente para realização de novo retreino;
- *retreino_se_otimiza*: Escola se utiliza ou não a otimização do modelo no retreinamento;

Nessa tela de configuração, o usuário também visualiza os modelos já treinados e as respectivas métricas.

3.3.2 Classificação utilizando a técnica TF-IDF

Além da alternativa de seleção de *features* proposta na subseção anterior, na literatura, *Term Frequency - Inverse Document Frequency* (TF-IDF) é comumente utilizado para criar vetores de atributos para esse tipo de problema (GHOSH; DESARKAR, 2018) - extração de informações e classificação de textos.

O *pipeline* proposto nessa abordagem inclui TF-IDF para seleção de *features* e pode ser dividido em 4 tarefas:

1. Pré-processamento de texto;
2. Vetorização com TF-IDF;
3. Seleção do modelo;
4. Otimização de hiperparâmetros;

A etapa de pré-processamento do texto consiste em preparar o texto para a vetorização do TF-IDF. Essa etapa inclui: uniformizar o texto em letras minúsculas, remover pontuação e *stop words*, e *stemming* (LOVINS, 1968). Como o TF-IDF é uma estatística numérica que reflete a importância de uma palavra para um documento levando em consideração sua aparição no *corpus*, ele resulta na valorização de palavras citadas em contextos específicos e na desvalorização de palavras faladas com grande frequência (QAISER; ALI, 2018). Os vetores gerados para cada transcrição são as entradas para o algoritmo de classificação.

Como o problema requer uma abordagem multiclasse, um classificador de cadeia (*ClassifierChain*) foi aplicado com 5 algoritmos de classificação: *Random Forest* (BREIMAN, 2001), *Decision Tree* (ROKACH; MAIMON, 2005), *K-Nearest Neighbors* (MUCHERINO et al., 2009), *Multi-layer Perceptron* (MLP) (HAYKIN, 1994) e *Ridge* (HILT; SEEGRIST, 1977). Depois disso, aqueles com melhor desempenho tiveram seus hiperparâmetros otimizados.

Dos cinco algoritmos testados, os dois com o melhor desempenho foram selecionados para otimização de hiperparâmetros: *Random Forest* e *Ridge Classifier*.

A otimização do modelo com *Random Forest* seguiu a metodologia apresentada no modelo de otimização em árvore, com o *framework scikit-optimize* tratando o problema como uma regressão. Os hiperparâmetros escolhidos foram os mesmos.

Para o *Ridge Classifier*, o valor de *alpha*, que representa a força de regularização do algoritmo, foi otimizado. Um espaço de log (números espaçados uniformemente em uma escala de log) foi escolhido para determinar a melhor configuração.

Os resultados dos modelos são apresentados na Secção 4.2.2.

3.3.3 Classificação utilizando o modelo BERT

O *BERTimbau* foi retreinado para a classificação de áudio multiclasse aqui apresentada como um problema de aprendizado supervisionado. Dado que o modelo já possui 24 camadas, 1024 dimensões ocultas, 16 cabeças de atenção e 330M de parâmetros (SOUZA et al., 2020), o alto custo computacional deve ser considerado nesse modelo de estado-da-arte de NLP na restrita força computacional disponível para o projeto (quando em produção).

Foi utilizado o modelo pré-treinado de Souza et al. (2020), disponibilido na plataforma *Huggingface*. O modelo, que utiliza *PyTorch* e *Huggingface Transformers* foi modificado para atender ao problema de classificação aqui abordado. As alterações consistiram em adicionar uma última camada e treiná-la para reconhecer os rótulos de ligações. Essas modificações são permitidas através dos *Transformers* (WOLF et al., 2020).

Para permitir a previsão de apenas um ou dois rótulos com o modelo baseado em BERT, cada classe foi treinada individualmente. Dessa forma, havia 6 modelos de BERT, um para cada classe: (1) Mudança de estado, (2) Modulação de geração, (3) Mudança de tensão, (4) Nível do reservatório, (5) Chuva e (6) O&M Local.

Os dados de entrada consistiram em um conjunto de dados balanceados para cada classe em um universo de 2278 gravações rotuladas. A saída do modelo seria binária, indicando se a chamada estava relacionada ou não ao rótulo em questão. Os dados que não estavam relacionados a cada modelo de treinamento foram selecionados aleatoriamente.

É importante considerar o alto custo computacional para treinar e prever ligações com o BERT. Utilizando um sistema com processador Intel i7 de 10ª geração e 16 GB de RAM, foram necessários cerca de 14 horas para treinar e 2 horas para prever um dia de gravações para cada classe.

Os resultados dos modelos são apresentados na secção 4.2.3.

3.4 Correlação de Ligações

A correlação é uma das entregas mais importantes do projeto do assistente virtual e agrega valor e informação aos áudios, visto que a partir das transcrições e de seus metadados, é possível associá-los tanto entre si, como também a eventos SCADA. A partir dessas associações, a pós-operação é capaz de gerar relatórios e realizar análises com maior facilidade.

As gravações possuem informações cruciais para a operação, entre elas pedidos de ações em campo, execuções de ordens de serviço e de documentos de segurança. Essas ações são acompanhadas pelo sistema SCADA das distribuidoras e transmissoras, no caso da CPFL Energia, o ADMS. Uma abertura de chave, por exemplo, requisitada via ligação, é registrada pela rede num evento e confirmada por ligação telefônica.

As gravações de áudio da CPFL podem ter origem em duas fontes principais: Rádio ou Central de Atendimento Automático Privado (PABX, do inglês *Private Automatic Brach Exchange*). Além das fontes de áudio, as gravações da CPFL também são separadas por meio de Centros Operacionais (CO) da seguinte forma:

- COI Piratininga

- COS SP/COI Paulista/ COI Santa Cruz
- COI/COS RGE

O rádio pertence às duas primeiras operações de COs, ambas com áudios PABX em seu processo. A RGE, por outro lado, opera apenas com PABX. Além da transcrição de áudio, todas as fontes de áudio possuem metadados associados. Esses metadados incluem a data e hora da chamada, o nome do operador que fez a chamada, o grupo de rádio e outras informações das gravações.

Para cada transmissão de rádio, os metadados associados são compostos por:

- Remetente: Número do rádio ou nome do operador que iniciou a comunicação;
- Destinatário: Grupo de rádio em que ocorreu a comunicação;
- Data: Data e hora da chamada;
- Região: Região a qual o rádio pertence;
- Empresa: Empresa a qual o rádio pertence;
- CO: Centro de Operação ao qual o rádio pertence;

Os metadados do PABX diferem. Os da COI Piratinga apresentam informações diferentes dos da RGE. Para o COI Piratinga, há apenas a data e hora da ligação e o número dos ramais de telefone envolvidos na ligação. Em oposição, os áudios da RGE apresentam os seguintes dados para cada chamada:

- Data: Data da chamada;
- Hora: Hora da chamada;
- Remetente: Ramal que fez a ligação;
- Destinatário: Ramal que recebeu a chamada;
- ID do Agente: Operador do CO envolvido na chamada;
- Região: Região da qual o telefone pertence;
- Empresa: Empresa da qual o rádio pertence;
- CO: Centro de Operação do qual o rádio pertence;

Além das informações associadas à transcrição, esses metadados alimentam os algoritmos descritos nas seções a seguir.

Um sistema de controle de supervisão e aquisição de dados (SCADA, do inglês *Supervisory control and data acquisition*) é um sistema de comunicação e controle usado para monitoramento, operação e manutenção de redes de infraestrutura de energia. A CPFL utiliza o ADMS (do inglês, *Advanced Distribution Management System*) como seu principal sistema para adquirir dados em tempo real, supervisionar e controlar equipamentos de rede e subestações de distribuição de energia elétrica.

Para os algoritmos propostos neste trabalho, foram extraídos os seguintes recursos do ADMS:

- Data e hora do evento;
- Nome do Dispositivo (*DeviceName*);
- Nome da Subestação (*SubstationName*);
- Nome do Alimentador (*FeederName*);
- Região (*RegionName*);
- Documento (*ReasonId*);
- Mensagem (*Message*);

Nas subseções a seguir serão apresentados dois algoritmos: a correlação de ligações com eventos SCADA e a correlação de histórias (correlação de ligações entre si). Serão descritos a metodologia, a evolução dos resultados e os modelos em execução na ferramenta.

3.4.1 Correlação de Áudios com Eventos SCADA

Primeiramente, um *baseline* consiste em uma primeira versão do algoritmo, a partir da qual foram realizados avanços e melhorias. A ideia é sempre comparar novas versões do algoritmo com o *baseline*.

Antes da proposta do algoritmo inicial que é chamado de *baseline*, foi necessário a definição das métricas para avaliação da correlação das ligações com os eventos SCADA. Para isso, tornou-se necessário a classificação de áudios manualmente. Primeiramente, foram classificadas 200 gravações do COI Piratininga em *Correlacionável* ou *Não Correlacionável*.

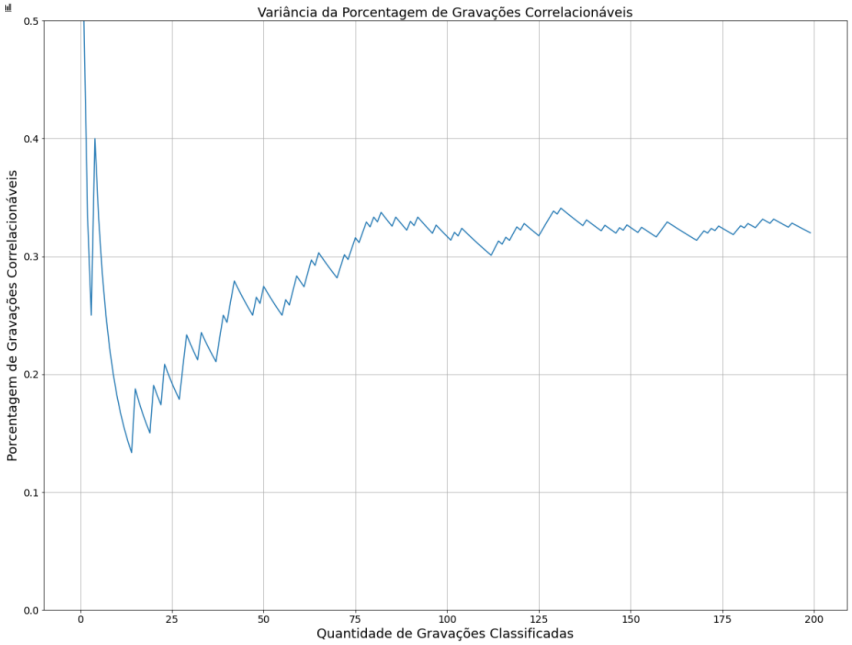
Para classificação dos áudios entre correlacionáveis e não correlacionáveis foram utilizados os seguintes critérios:

- Presença de números que identificam um equipamento;

- Citação de alimentador (nome de subestação + número);
- Citação de número de OS ou SP;
- Citação de equipamentos como trafo, barra/barramento.

Os 200 áudios foram selecionados aleatoriamente da amostra completa de ligações do mês de dezembro de 2020 do COI Piratininga. Essa quantidade de áudio foi escolhida com base na curva de estabilidade de variação da amostra, também conhecida como Lei dos Grandes Números (ETEMADI, 1981), representada na Figura 8.

Figura 8 – Gráfico de estabilização da porcentagem de gravações correlacionáveis



Fonte: Autora

Amostrados um a um, foi calculada a porcentagem de gravações que são consideradas correlacionáveis com o ADMS, estabilizando num total de 32%. No gráfico, é possível perceber que a porcentagem estabiliza a partir de 150 gravações, todavia, foram realizadas mais classificações para mais ligações objetivando garantir uma maior quantidade de exemplos para os testes.

Essas 200 gravações foram utilizadas para gerar as métricas para o algoritmo de correlação SCADA, possibilitando inferir e comparar a acurácia, *recall* (sensibilidade), precisão e F1-score de cada modelo.

Para o *baseline* da correlação de áudios com eventos ADMS, foram realizadas as seguintes ações:

1. Filtro de um intervalo de 30 minutos antes e depois do horário da ligação nos dados do ADMS;

2. Se na transcrição foi identificado algum nome de subestação, filtra-se pela coluna *SubstationName* do ADMS; e

3. Se a transcrição identificou números de 4 ou mais dígitos na transcrição, filtra-se na coluna *DeviceName* do ADMS pelo número identificado na transcrição;

No Quadro 1 está um exemplo de correlação realizada em uma ligação de ramal do COI Piratininga referente ao dia 01/12/2020. Na primeira linha, está a transcrição realizada com a ferramenta *Speech-to-Text* do GCP, e nas linhas abaixo, as associações com o ADMS encontradas pelo algoritmo. A partir do número do religador citado (939064), foram encontrados os eventos ADMS relacionados à ligação.

Quadro 1 – Exemplo de correlação de áudio com SCADA

"... time 3434, só um minuto, por favor vai para o **religador**, **religador 939064**, ok?"				
Datetime	Message	DeviceName	FeederName	SubstationName
01/12/2020 08:18	COPER	939064	REP03	SOROCABA 2-REPRESA
01/12/2020 08:18	CEXEC	939064	REP03	SOROCABA 2-REPRESA
01/12/2020 08:18	ALESTS	939064	REP03	SOROCABA 2-REPRESA

Fonte: Autora

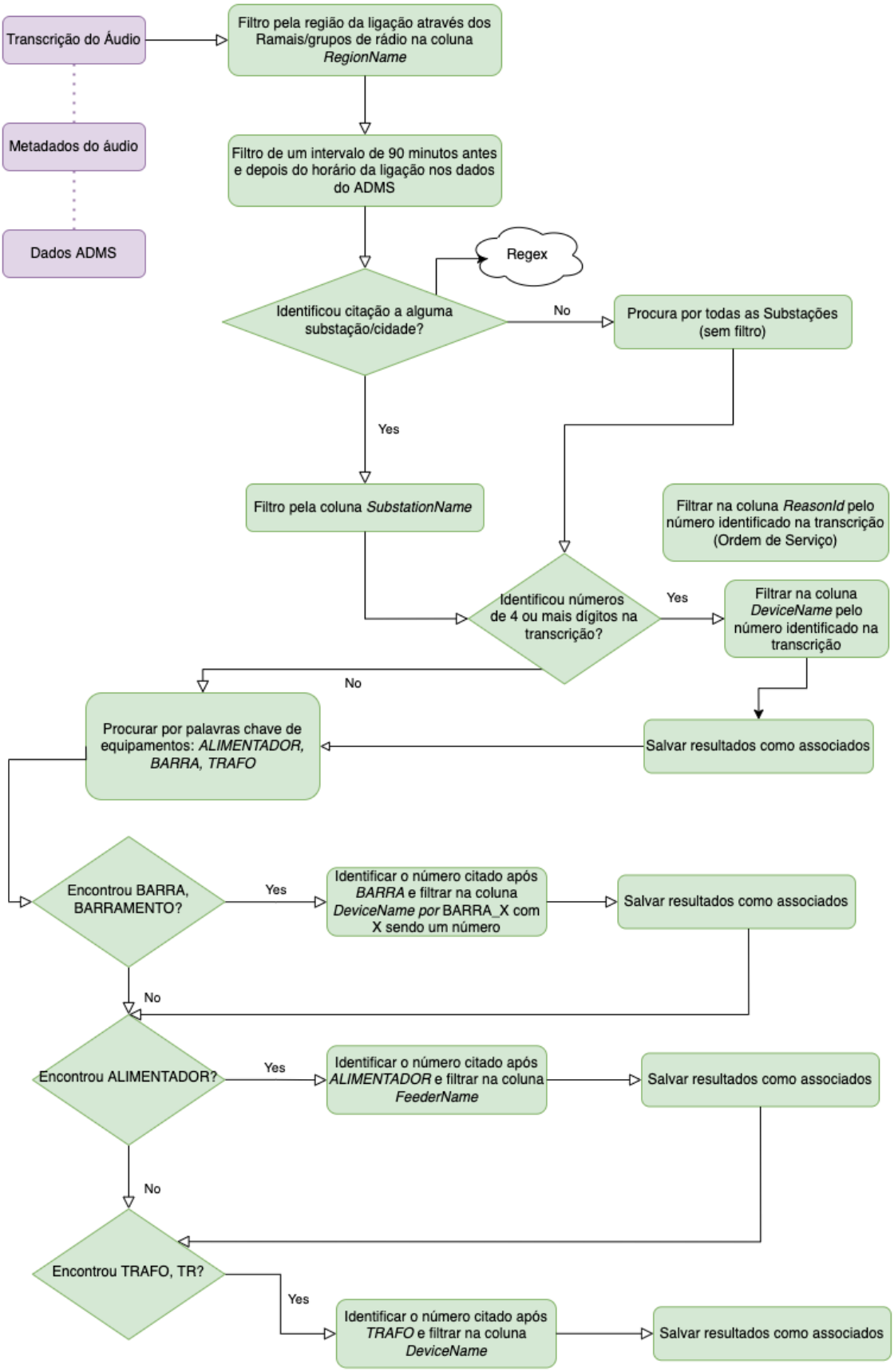
Partindo das métricas do *baseline*, foi sendo aumentada a complexidade do algoritmo. Na seção seguinte serão apresentadas as mudanças realizadas na implementação do algoritmo, de forma a mitigar essa deficiência.

A partir do *baseline* e de análises com os 200 áudios de PBX Piratininga utilizados para as métricas, novas regras de correlação foram implementadas. Para isso, foi preciso o conhecimento da operação adquirido através de sessões de contextualização junto a operadores da CPFL e o contato com um número considerável de gravações e suas respectivas associações no ADMS.

À medida que a complexidade do algoritmo era aumentada, ou mesmo quando algum parâmetro era alterado, eram computadas as métricas da correlação. O objetivo principal dos incrementos era aumentas o reconhecimento das ligações correlacionáveis (*recall* Corr), alcançando pelo menos 60%. A partir do *baseline* (o algoritmo final completo está descrito no diagrama da Figura 9), as seguintes modificações foram realizadas:

1. Inclusão de tratamento para correlação com trafos e barramentos: a partir da identificação na transcrição de trafo, tr, barra ou barramento + número de até dois dígitos e busca de eventos desses dispositivos na coluna DeviceName do ADMS;

Figura 9 – Diagrama do algoritmo completo da correlação de áudios com eventos SCADA



Fonte: Autora

2. Mudança no método de extração de números: números a partir de 8 dígitos partidos em 4, 5 e 6 algarismos (tentativa de mitigar erros de transcrição devido a baixa qualidade da gravação);

3. Aumento do intervalo de tempo para seleção de eventos ADMS para 1 hora antes e depois da ligação;

4. Não filtrar por subestação citada na transcrição caso o resultado final não tenha selecionado nenhum ADMS;

5. Nova mudança no método de extração de números: números a partir de 6 dígitos, partidos em diferentes versões com 4 algarismos;

6. Aumento do intervalo de tempo para seleção de eventos ADMS para 90 minutos antes e depois da ligação;

7. Ajuste da identificação de trafos na transcrição: identificação não estava ajustada para TR04, inclusão de diferentes tipos de citações a trafos no algoritmo;

8. Identificação de itens que não são correlacionáveis na amostra base: após análise de operadores, foram identificados que alguns áudios, antes considerados correlacionáveis com o ADMS, na verdade não eram;

9. Inclusão do tratamento de alimentadores (primeira versão): a partir da identificação pela transcrição do padrão nome de subestação + número de até 2 dígitos, associação do nome à sigla equivalente e seleção de eventos na coluna *FeederName* com sigla do alimentador e número.

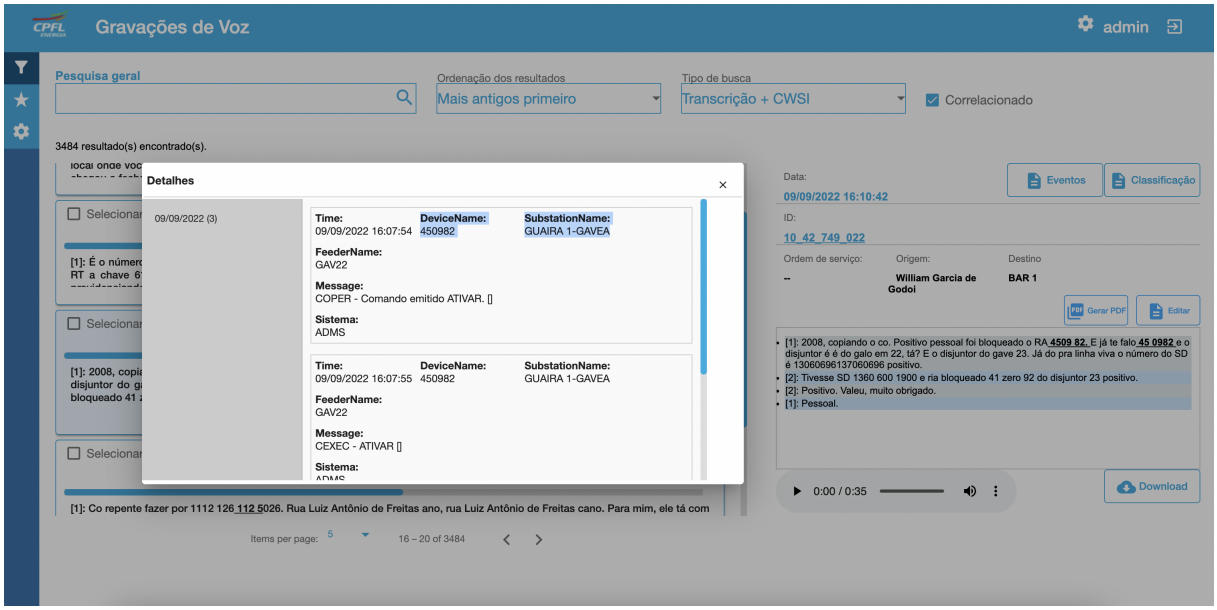
Na ferramenta, a correlação com o ADMS é exibida tal qual na Figura 10. A transcrição separa por interlocutor e com o número do dispositivo grifado está do lado direito da tela e os itens ADMS correlacionados são listados no *pop-up* à esquerda.

Na Figura 10, é apresentado uma ligação que ocorreu em 09/09/2022 às 16:10:42, com o operador William, do centro de operação, comunicando para uma equipe de campo no grupo de rádio BAR1 que o RA450982 havia sido bloqueado. Os comandos, executados remotamente, são registrados pelo ADMS e associados à ligação pelo algoritmo de correlação, e ocorreram às 16:07:54 e 16:07:55.

3.4.2 Correlação de Histórias

A correlação de histórias é essencial para a pós-operação. Através dela, o sistema é capaz de prover aos engenheiros os áudios que se associam entre si, facilitando assim, a análise de uma ocorrência.

Figura 10 – Tela com pop-up da correlação de áudios com SCADA



Fonte: Autora

A correlação entre áudios poderá ocorrer de duas formas, para os áudios de rádio (TurboNet) e para os áudios PABX (baseadas em eventos SCADA). Na primeira, a ideia é unir em uma só associação a sequência de curtas comunicações entre CO e campo realizadas via rádio. Já a segunda, associa entre si ligações que tratem do mesmo evento, equipamento ou alimentador. A seguir ambas as correlações são apresentadas, com avanços de algoritmos, métricas e exemplos.

3.4.2.1 Histórias Baseadas em Eventos SCADA

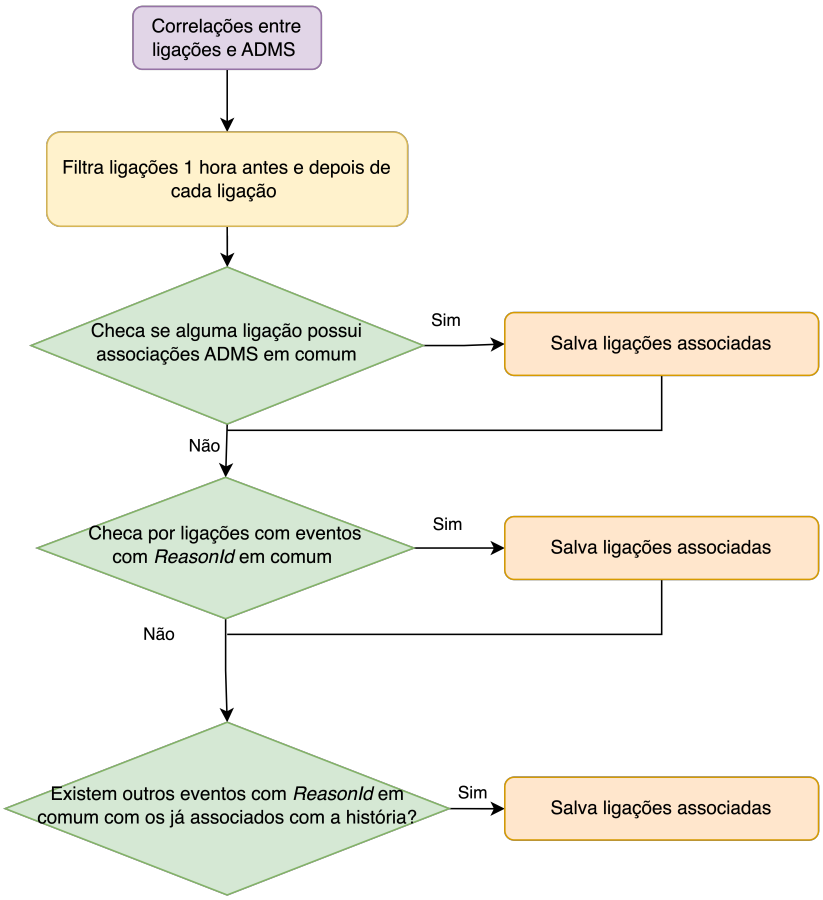
A primeira versão proposta (*baseline*) do algoritmo de correlação de histórias baseia-se no resultado das correlações de histórias do ADMS. Os testes foram realizados com áudios do PBX Piratininga.

A ideia é que sejam correlacionados áudios com eventos ADMS, Ordens de Serviço (OS) e Planos de Manobra (SP, do inglês *Switch Plan*) em comum. Dessa forma, áudios que tratam do mesmo equipamento/evento num determinado período (1 hora) são correlacionados em uma única história. A Figura 11 apresenta o diagrama do *baseline* proposto para a correlação.

Dessa forma, áudios como os do exemplo dos Quadros 2 e 3 são considerados parte de uma mesma história.

O primeiro áudio do Quadro 2 ocorreu às 22:35, enquanto o segundo (Quadro 3), ocorreu às 22:44. No primeiro áudio, é feito o pedido da execução da manobra em anel na chave LB 428461, comando executado às 22:39. No segundo áudio, há a confirmação da execução da manobra e eo pedido para execução de outro procedimento, dessa vez

Figura 11 – Diagrama do baseline da correlação de histórias PABX



Fonte: Autora

Quadro 2 – Exemplo de correlação de áudio com SCADA

"[...] pode fazer anel na conta **428461**. Beleza então [...]"

Datetime	Message	DeviceName	FeederName	SubstationName
13/12/2020 22:39	CEXSIM ABERTA	SC_LB_A 428461	JUN13	JUNDIAI 1
13/12/2020 22:39	CEXSIM ABERTA	SC_LB_B 428461	JUN13	JUNDIAI 1
13/12/2020 22:39	CEXSIM ABERTA	SC_LB_C 428461	JUN13	JUNDIAI 1

Fonte: Autora

na chave 340707. Ambas as ações na rede do mesmo alimentador e subestação: JUN13 e Jundiaí 1, respectivamente. A associação dos dois áudios entre si como uma história acontece pois ambos mencionam eventos na LB 428461.

Quadro 3 – Exemplo de correlação de áudio com SCADA

"[...] eu tô na **42 8461**, mas pediu para abrir ela aqui abrir a **428461** tem anel lá, né? Não eu puxei ele se ninguém pode abrir está em anel, né? Pode abrir a 4461 tem anel aí bloqueado alimentador, né? [...] tá joia **428461** aberto [...] agora é **340707**, nós já mandei para você, beleza? [...]"				
Datetime	Message	DeviceName	FeederName	SubstationName
13/12/2020 22:39	CEXSIM ABERTA	SC_LB_A 428461	JUN13	JUNDIAI 1
13/12/2020 22:39	CEXSIM ABERTA	SC_LB_B 428461	JUN13	JUNDIAI 1
13/12/2020 22:39	CEXSIM ABERTA	SC_LB_C 428461	JUN13	JUNDIAI 1
13/12/2020 22:52	CEXSIM ABERTA	SC_LB_A 340707	JUN13	JUNDIAI 1
13/12/2020 22:52	CEXSIM ABERTA	SC_LB_B 340707	JUN13	JUNDIAI 1
13/12/2020 22:52	CEXSIM ABERTA	SC_LB_C 340707	JUN13	JUNDIAI 1

Fonte: Autora

A validação do *baseline* se deu a partir de avaliações história a história junto a engenheiros de pós-operação. Foram selecionadas 50 histórias (conjuntos de áudios, somando 150 ligações) e 50 áudios sem história para avaliação.

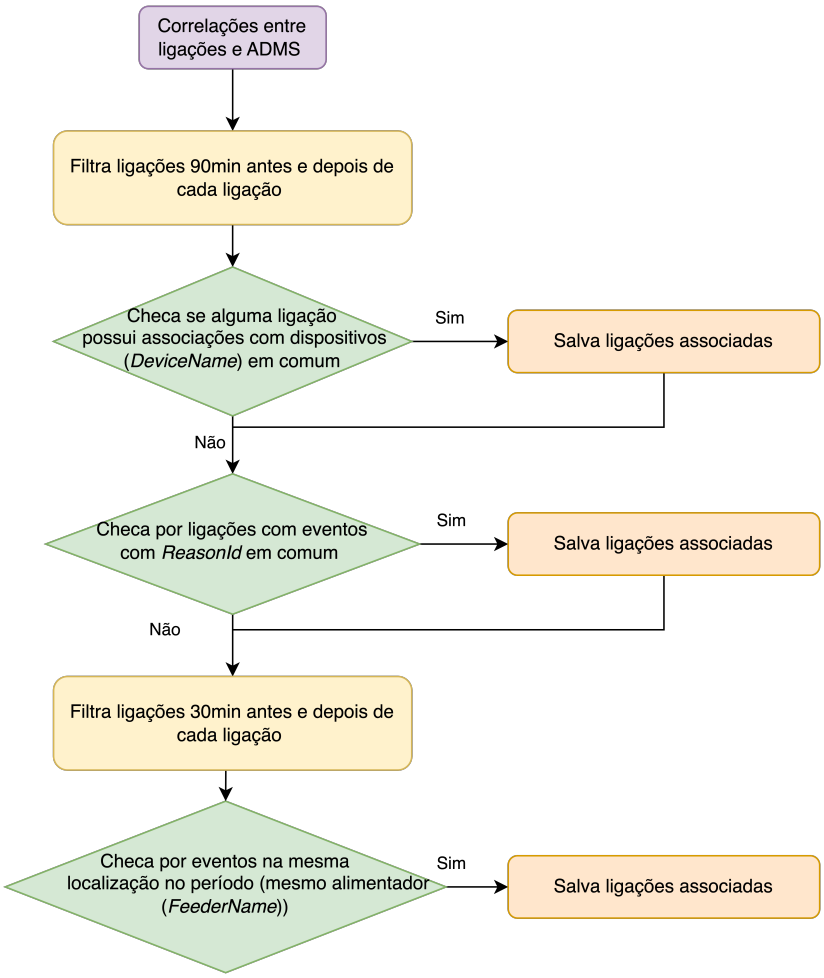
Durante a análise dos resultados, o *baseline* indicou que o principal ponto de melhoria deveria ser na identificação de uma história, logo foram implementadas modificações no algoritmo e a versão final é apresentada no diagrama da Figura 12.

As alterações priorizaram abordagens de diferentes de intervalos de tempo para tipos de associações, levando em consideração a natureza diferente das operações de distribuição e transmissão.

As alterações incorporadas incluem um intervalo inicial maior para a busca de ligações que possam fazer parte da história passando de 60 para 90 minutos, o aumento pelo intervalo de busca se deu pela necessidade de acomodar a ocorrência de eventos que levem muito tempo para serem concluídos, pois a comunicação de sua conclusão deve estar presente na mesma história da comunicação de seu início.

Dentro do novo intervalo de busca, passa-se então a verificação da existência de ligações que possuam em seus registros ADMS associações com dispositivos em comum, ou seja, ligações na qual a informação no campo DeviceName é a mesma, na sequência, é

Figura 12 – Diagrama da versão final da correlação de histórias PABX



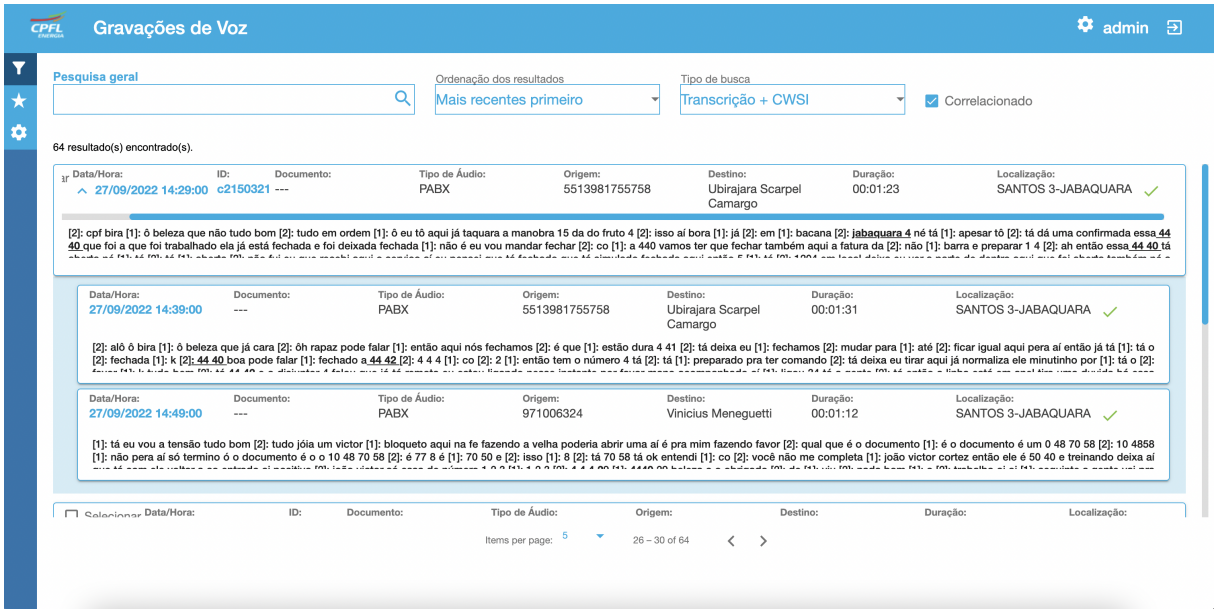
Fonte: Autora

feito uma busca por ligações que possuam o campo ReasonId em comum (evento/OS/SP em comum). Caso se identificada a igualdade de alguns desses campos a ligação já está qualificada para fazer parte da história não sendo necessário mais testes.

Outra adição presente nessa revisão está no fluxo adicional de verificação de alimentadores, foi observado pelos especialistas que para intervalos pequenos, aqui adotado como 30 minutos, eventos ADMS que tratam de um mesmo alimentador devem pertencer a mesma história. Sendo assim, foi realizada a adição de um novo filtro para as ligações que acontecem nesse intervalo, pôde então ser adicionado à história da ligação itens no qual registros ADMS possuíssem ocorrências de mesmo alimentador no campo FeederName. O diagrama atualizado com o novo algoritmo é exibido na Figura 12.

Na ferramenta, a correlação de histórias é apresenta tal qual na Figura 13. No exemplo são associadas ligações que tratam de uma ocorrência da subestação 'SANTOS 3 - Jabaquara', no alimentador Jabaquara 4.

Figura 13 – Diagrama da versão final da correlação de histórias PABX



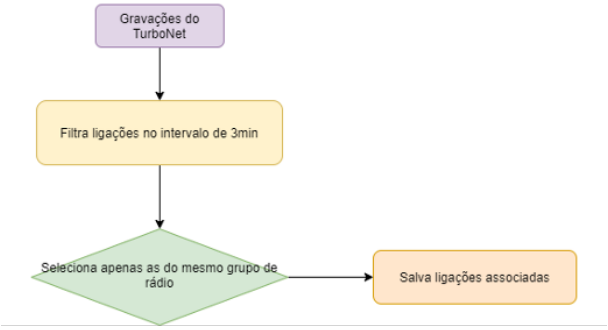
Fonte: Autora

3.4.2.2 Histórias com Áudios de Rádio

Os áudios de rádio possuem características diferentes dos áudios de ramais. Eles são mais curtos e costumam ter conversações cortadas em gravações diferentes. Por isso, tornou-se necessário um estudo à parte para as comunicações via TurboNet.

Foram iniciados os estudos com esses áudios partindo do pressuposto que as comunicações recorrentes acontecem de um mesmo grupo de rádio. O algoritmo inicial proposto está no diagrama da Figura 14.

Figura 14 – Diagrama do *baseline* da correlação de áudios de rádio



Fonte: Autora

Os testes iniciais, todavia, apresentaram resultados inesperados. A comunicação não necessariamente acontece no mesmo grupo de rádio. Foi notado, inclusive, que quando a comunicação parte do COI/COS, o mesmo áudio é gravado em diferentes fontes (grupos de rádio). Logo, outra hipótese foi levantada:

- Quando a comunicação parte do COI/COS, vários grupos de rádio são acionados;
- Quando a comunicação parte do campo, apenas um grupo de rádio é acionado.

A partir dessa nova hipótese, foram realizados novos testes e estudos para compreender como relacionar os áudios do TurboNet. Tais estudos resultaram numa proposta de algoritmo recursivo que trata os casos citados anteriormente.

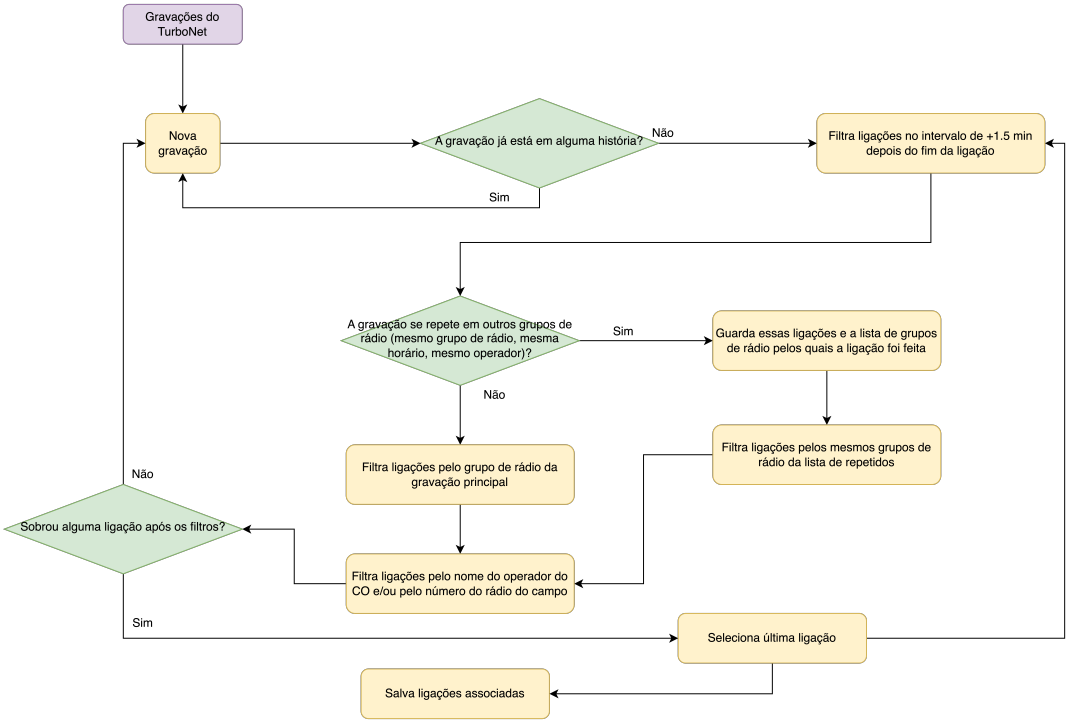
A partir dos problemas encontrados na primeira abordagem com a correlação de histórias para TurboNet, foi proposto uma alternativa com recursividade. Na Figura 15, está descrito o algoritmo implementado.

Dado um conjunto de ligações, o algoritmo irá iterar de ligação em ligação. Ao selecionar uma ligação, são executadas as seguintes etapas:

1. Testar se a ligação já pertence à uma história:
   (a) Se sim, vai para a próxima ligação;
   (b) Caso contrário, segue o algoritmo;
2. Filtrar ligações no intervalo de 1,5 minutos depois do fim da ligação selecionada;
3. Checa se a gravação se repete em outros grupos de rádio (mesmo grupo de rádio, mesmo horário, mesmo operador):
   (a) Se sim, guarda essas ligações e a lista de grupos de rádio pelos quais a ligação foi realizada;
      - Filtra ligações pelos grupos de rádio da lista de repetidos;
   (b) Caso contrário, filtra ligações pelo grupo de rádio da gravação principal;
4. Filtra ligações pelo nome do operador do CO e/ou pelo número do rádio do campo;

A subsecção 4.3.2.2 exibe as métricas do algoritmo.

Figura 15 – Diagrama da versão final da correlação de áudios de rádio



Fonte: Autora

4 Resultados

Nessa secção serão apresentados os resultados de todos os algoritmos e otimizações descritos nesta dissertação.

4.1 Modelo *Speech-to-Text*

Nessa seção serão apresentados os resultados dos modelos de transcrição em nuvem avaliados: *Amazon Transcribe*, *Google Speech-to-Text* e *Azure Speech*.

Na avaliação dos modelos, foram utilizados os recursos de retreinamento disponíveis em cada uma das plataformas, diferentes entre si.

As métricas utilizadas para avaliar a qualidade das transcrições foram: WER (do inglês, *Word Error Rate*), REC (Percentual de Reconhecimento de Palavras-chave) e FP (Percentual de Falsos Positivos de Palavras-chave).

Para avaliação de desempenho dos transcritores com variação nos vocabulários, foram definidas as seguintes métricas:

- Percentual de Reconhecimento (REC): porcentagem de áudios em que a palavra X foi identificada pelo transcritor;
- Percentual de Falsos Positivos (FP): porcentagem de áudios em que a palavra X foi identificada erroneamente pelo transcritor.
- WER (do inglês, *Word Error Rate*): obtido através da contagem de palavras inseridas, substituídas e removidas da transcrição gerada pelo modelo em relação ao *groundtruth*;

4.1.1 Modelo da *AWS*

Como mencionado, para o transcritor da *AWS* foram realizados experimentos com cinco vocabulários diferentes. Na primeira abordagem (V0), foi utilizado um único vocabulário para todas as ligações. Esse vocabulário apresentou uma melhora de 3,85% no REC de palavras e expressões do vocabulário geral com relação a transcrição da AWS bruta, 6,39% para usinas e não houve alteração no reconhecimento dos nomes dos operadores.

A usina que teve o pior resultado de FP foi Ouro-Verde (EOVE), todavia tanto nas transcrições sem vocabulário quanto com o vocabulário V0, a usina teve um percentual de reconhecimento de 100% e um de falso positivo de 15,25%. A usina “Lacerda” (UTLA) apresentou métricas abaixo do esperado, nas duas situações apresentou REC e FP de

16,13% e 7,04% respectivamente. Essa usina exemplifica casos onde o uso da ferramenta de vocabulário customizado precisa ser avaliado cautelosamente, para definir se o aumento de falsos positivos compensa a diferença no percentual de reconhecimento.

Além disso, alguns termos tiveram um aumento de falsos positivos sem que houvesse uma melhora no REC. Foi notado que a maior parte desses casos ocorreu em ligações em que esses termos não pertenciam ao vocabulário, como um pedido para compensar uma unidade geradora numa ligação com uma usina eólica. Por isso, foi proposta a segunda abordagem, ao diferenciar os vocabulários por tipo de usinas tratadas na ligação, era esperado uma diminuição dos falsos positivos.

Na transcrição com vocabulários diferentes para diferentes grupos de usinas, foi possível comprovar que houve melhoria no reconhecimento com o vocabulário separado a partir dos ramais. Enquanto na primeira versão, a média do reconhecimento de palavras-chave chegou a 78,87% para palavras referentes à operação, a segunda apresentou uma melhora para 79,19%. Quanto à incidência de falsos positivos, houve uma alteração significativa, diferente do esperado, saindo de 6,56% para 4,23% na primeira versão do vocabulário separado. A comparação entre as métricas de todos os vocabulários é apresentada nas Tabelas 2 e 3.

Tabela 2 – REC e WER para os vocabulários testados com a AWS

	%REC Sem Vocab	%REC V0	%REC V1	%REC V2	%REC V3	%REC V4	%REC V5
Geral	75,02%	78,87%	79,19%	79,06%	59,52%	79,06%	79,06%
Usinas	49,49%	55,88%	59,37%	58,91%	52,26%	58,94%	58,94%
Operadores	32,85%	32,92%	40,87%	42,93%	38,86%	45,55%	47,21%
WER	0,68	0,43	0,58	0,63	0,51	0,56	0,52

Apesar das tentativas de aumentar a porcentagem de reconhecimento de nomes de operadores, todas as variações do vocabulário após a V1 resultaram no aumento de falsos positivos, consequentemente, também resultaram no aumento do WER, que obteve o melhor resultado com o Vocabulário V1.

Tabela 3 – FP para os vocabulários testados com a AWS

	%FP Sem Vocab	%FP V0	%FP V1	%FP V2	%FP V3	%FP V4	%FP V5
Geral	5,81%	6,56%	4,23%	5,93%	6,57%	5,93%	5,93%
Usinas	3,20%	12,63%	2,55%	8,91%	7,71%	8,28%	8,28%
Operadores	5,73%	8,73%	4,21%	13,41%	17,95%	17,55%	16,73%

Logo, apesar da dificuldade de identificação dos nomes dos operadores, foi mantido a V1 do vocabulário personalizado. A ferramenta desenvolvida ainda permite que o usuário edite o vocabulário utilizado na transcrição, identificando os campos *Phase* (Palavra), *SoundsLike* (Pronúncia) e *DisplayAs* (Palavra Tratada) e em qual tipo de vocabulário a palavra/expressão deve pertencer (Tipos).

4.1.2 Modelo do *GCP*

Para cada tipo de vocabulário e fonte de áudio, todos apresentados na Tabela 4, foram calculadas a taxa de reconhecimento de palavras-chave (REC) e a taxa de falsos positivos (FP). Os resultados apresentados na Tabela 4 exibem o valor de *boost* otimizado para cada vocabulário, de forma que equilibre o reconhecimento das palavras-chave e a taxa de falsos positivos.

Tabela 4 – Resultado da otimização de vocabulários e *boosts*

Vocabulário	%REC	MR	Quantidade de Ligações	*Boost*
Vocabulário Geral	70%	14,61%	1058 de todas os COs e fontes	10
Siglas	51,56%	35,35%	1058 de todas os COs e fontes	Mixed
Operadores	77,55%	23,35%	947 de todas os COs e fontes	15
Localidades (*Paulista* rádio)	84,89%	0%	322 gravações	0
Localidades (*Paulista* PABX)	72,72%	11,11%	106 gravações	10
Localidades (*Piratininga* rádio)	-	-	206 gravações	0
Localidades (*Piratininga* PABX)	79,86%	39,58%	79 gravações	10
Localidades (*Santa Cruz* rádio)	100%	0%	103 gravações	0
Localidades (*Santa Cruz* PABX)	64,04%	23,35%	307 gravações	10
Localidades (*RGE* PABX)	78,52%	12,90%	325 gravações	15

A taxa de falsos positivos para cada palavra nos vocabulários também permitiu filtrar palavras que o GCP estava transcrevendo incorretamente. Os critérios aplicados para remover essas palavras/expressões foram:

1. Palavra/expressão apresentou mais de 10% de FP;

2. Palavra/expressão apresentou mais de 5% de FP e nenhuma melhora do REC;

Com exceção da lista de siglas, os demais vocabulários não tiveram mais de 15 palavras/expressões retiradas devido ao índice de falsos positivos. O vocabulário de siglas, todavia, finalizou a análise com apenas cinco acrônimos. O estudo constatou que algumas siglas apesentavam mais de 60% de FP quando com *boost*. De todos os vocabulários propostos, o de siglas exibiu o menor índice de reconhecimento (51,56%), mas com melhoria de REC de 35%.

O vocabulários geral e de operadores, ambos considerados em gravações de todas as regiões e fontes, apresentaram uma taxa de reconhecimento de palavras-chave de pelo menos 70% e uma melhora de mais de 23% e 14% em relação à transcrição bruta, respectivamente. O vocabulário geral apresentou apenas 1% de falsos positivos e teve nove palavras removidas devido às altas taxas de FP quando com *boost*.

O vocabulário de localizações, conforme mostrado na Tabela 4, são separados por região e fonte de áudio. É característico que gravações de rádio não citem tantas subestações e cidades quanto as de ramal. Isso acontece devido à necessidade de comunicação imediata ao falar pelo rádio. Devido à sua baixa incidência de palavras-chave, os *boosts* desses vocabulários (Localidades + Rádio) são todos zero. Em *Piratininga*, por exemplo, nenhuma subestação ou cidade foi mencionada em 206 registros.

Como a correlação entre as transcrições de áudio e eventos SCADA é construída a partir de palavras-chave, regiões e subestações identificadas no texto, a melhoria de reconhecimento obtida neste trabalho desempenhou um papel importante para o sucesso do projeto.

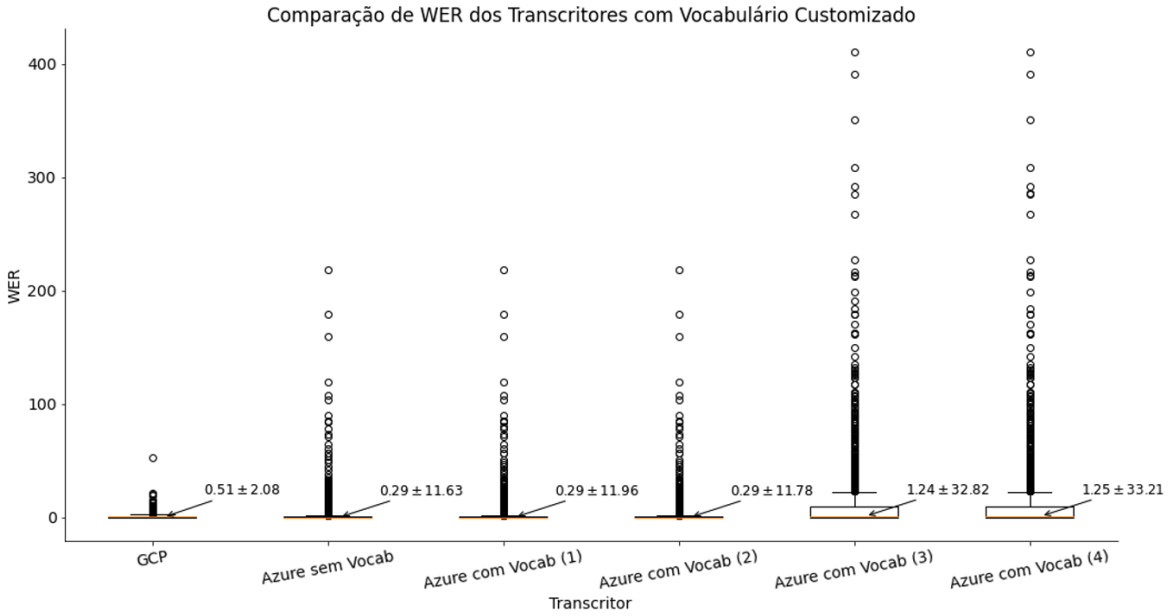
4.1.3 Modelo da *Azure*

As Figuras 16 e 17 apresentam as métricas calculadas para a comparação dos transcritores padrão da plataforma de transcrição da Azure, o transcritor final do GCP e os transcritores Azure com treinamento de vocabulário. A análise consistiu em transcrever todos os áudios que possuem *groundtruth*, ou seja, aqueles áudios para os quais uma transcrição manual foi cadastrada e avaliar WER e reconhecimento de palavras-chave.

A métrica REC é uma medida de grande proeminência na literatura de sistemas de reconhecimento de fala para analisar a capacidade dos modelos em reconhecer palavras de interesse específicas. A métrica trata-se da porcentagem de palavras de interesse que foram corretamente transcritas de um áudio dentre a quantidade de palavras que eram esperadas.

Os resultados da análise da métrica REC para o transcritor do GCP, o transcritor padrão da plataforma (sem vocabulário) e para os quatro modelos *baseline* que foram adaptados utilizando a migração do vocabulário personalizado são apresentados na Figura 17.

Figura 16 – WER para os transcritores avaliados



Fonte: Autora

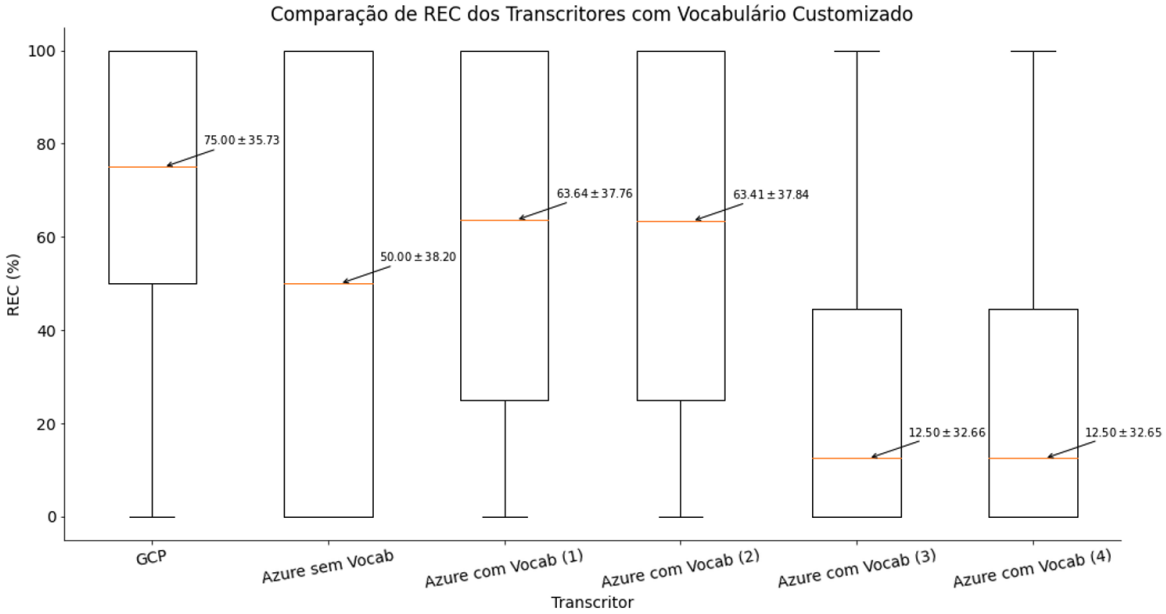
Como pode ser visto na Figura 17 o transcritor do GCP apresentou um reconhecimento mediano de 75% das palavras presentes no vocabulário customizado, já o transcritor padrão ofertado pelo pela Azure obteve apenas 50% de reconhecimento das palavras do vocabulário, vale ressaltar que ambos os transcritores apresentaram altos valores para o desvio padrão para a métrica indicando uma alta variância na performance de reconhecimento.

Ainda na Figura 17, pode-se verificar a análise da métrica de reconhecimento para os transcritores baselines da Azure treinados com a adaptação do vocabulário de palavras específicas para as operações da distribuição e transmissão. Dois transcritores apresentaram desempenhos competitivos após a etapa de treinamento nomeadamente os modelos *Azure com Vocab (1)* e *Azure com Vocab (2)* obtendo respectivamente 63,64% e 63,41%, com esses resultados sendo aproximadamente 11% piores que os obtidos com o melhor modelo de transcrição da plataforma GCP.

Tendo certificado a competência de reconhecimento dos modelos treinados no reconhecimento das palavras presentes no vocabulário deve-se analisar sua performance no reconhecimento geral de palavras para que a qualidade do modelo seja de fato atestada.

Para tanto a Figura 16 mostra a comparação dos resultados obtidos para a métrica

Figura 17 – Métrica REC para o transcritor GCP bem como para os baseilines treinados com a adaptação do vocabulário personalizado



Fonte: Autora

WER para todos os transcritores avaliados. Pode-se perceber a melhor performance mediana do transcritor padrão da Azure obtendo um WER 0,22 melhor que o transcritor GCP, no entanto, apresentou um desvio padrão maior.

Já para os transcritores Azure treinados com a adaptação de vocabulário dois deles apresentaram um desempenho competitivo, nomeadamente o *Azure com Vocab (1)* e *Azure com Vocab (2)*, ambos apresentaram WER mediano de 0,29 com desvios padrões muito próximos entre si e com o do modelo sem adição de vocabulário.

Para os dois últimos modelos da Azure testado a análise de WER confirma o desempenho ruim antes visto no reconhecimento de palavras do vocabulário e agravando ainda mais suas performances os desvios padrões para ambos os modelos foram aproximadamente três vezes maiores que as contrapartidas com melhor performance.

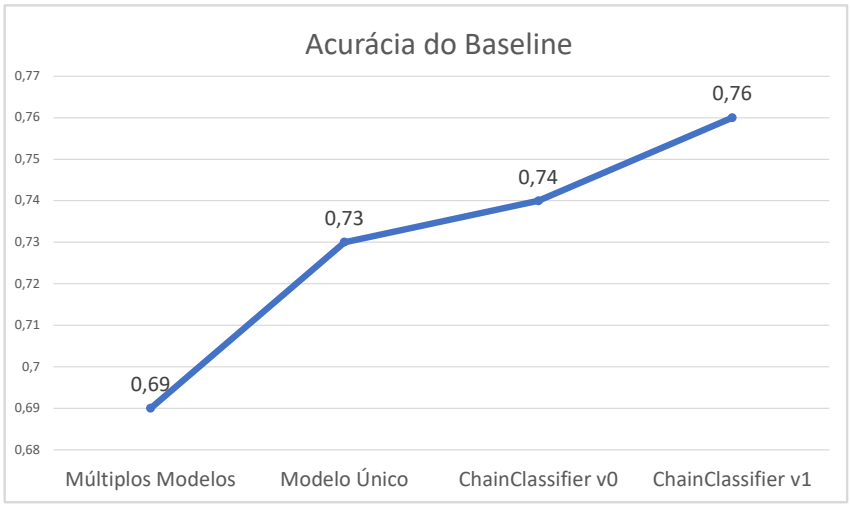
4.2 Modelo de Classificação de Ligações

4.2.1 *Baseline* e Avanços

No gráfico da Figura 18, é possível visualizar o avanço da acurácia nos testes realizados. Para o modelo único, foi inicialmente utilizado o *MultiOutputClassifier* (LEARN, 2020b) com o algoritmo *RandomForest* (BIAU; SCORNET, 2016) e para os modelos múltiplos, foram aplicados 6 modelos *RandomForest*.

A diferença entre as versões v0 e v1 listadas no gráfico é a ordem do treinamento dos

Figura 18 – Evolução da acurácia do *baseline* (primeiros testes)

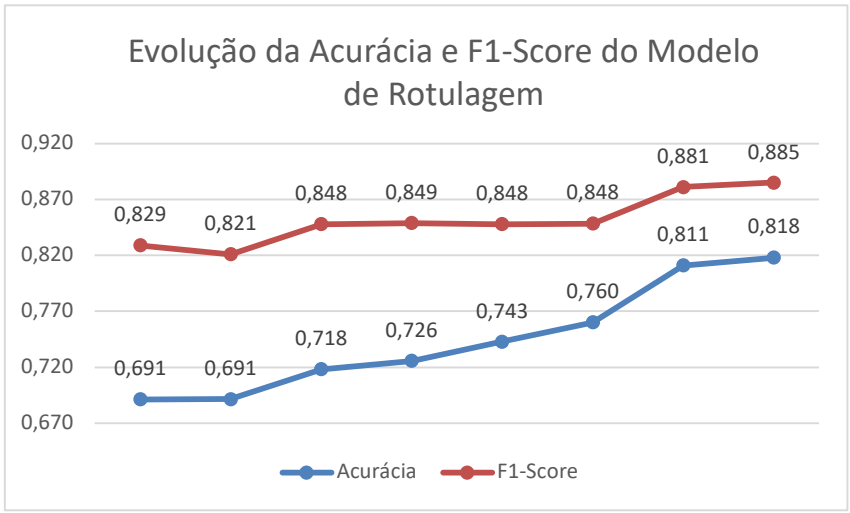


Fonte: Autora

rótulos. Na v1, Mudança de Estado e Modulação de Geração, que não estavam performando bem, foram os últimos na ordem de treinamento, influenciando na acurácia, que aumentou de 74% para 76%. O TPOT utilizado teve como parâmetros 100 gerações e população inicial de 100 modelos.

Foram testados alguns limites de parâmetros e a evolução das métricas é apresentada na Figura 19.

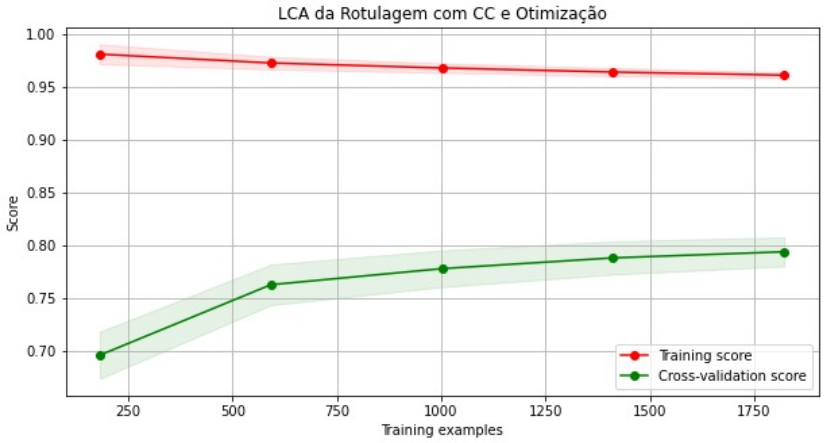
Figura 19 – Evolução da acurácia do modelo



Fonte: Autora

Ao alcançar uma acurácia de 81,8% e um F1-score de 88,5%, também foram analisadas as métricas individuais para cada rótulo. Nos experimentos prévios, houve uma dificuldade com 'Modulação de Geração' e 'Mudança de Estado', rótulos esses que alcançaram acurácia maior que 90%, e F1-score (positivo) de 79% e 83,6% respectivamente.

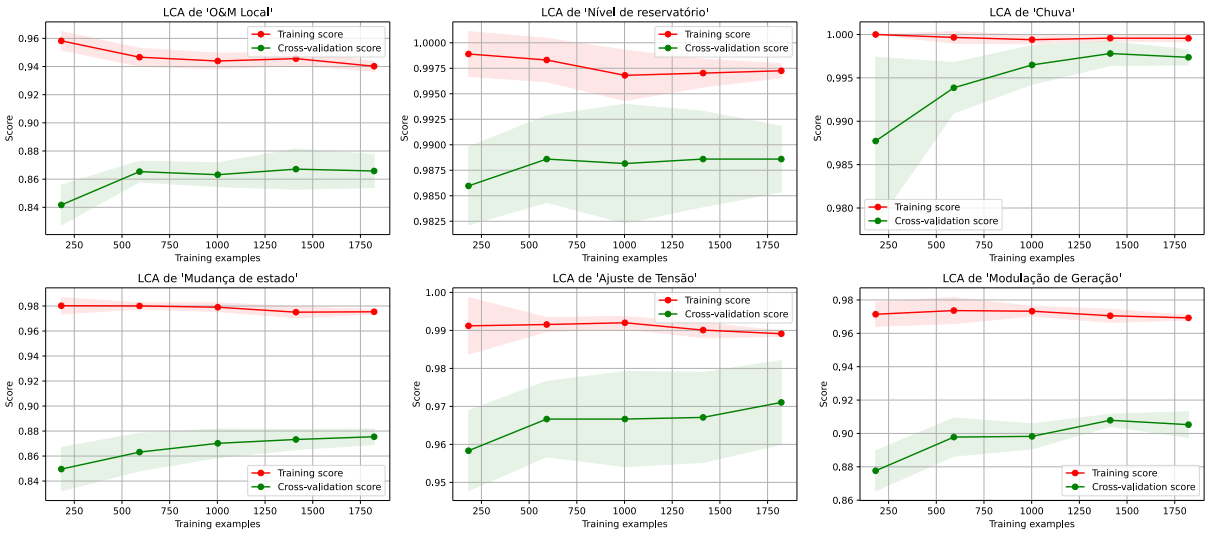
Figura 20 – *Learning curve analysis* do modelo completo



Fonte: Autora

Para auxiliar na análise do comportamento do novo modelo, foi utilizada a *Learning Curve Analysis* (LCA). A LCA é um gráfico que compara a performance de um modelo com dados de treinamento e teste a partir da variação do número de instâncias de treinamento. A partir do gráfico, é possível avaliar *bias*, variância, generalização do modelo, entre outras características. As LCAs para o modelo completo e por rótulo são exibidas nas Figuras 20 e 21.

Figura 21 – *Learning curve analysis* de cada classe



Fonte: Autora

A partir do gráfico LCA do algoritmo completo (Figura 20), é possível identificar que a curva com o *score* de *cross-validation* ainda possui uma leve inclinação. Esse comportamento indica que o tagueamento de mais áudios pode melhorar o desempenho do modelo. O mesmo não ocorre para os rótulos 'Nível de Reservatório', 'Chuva', 'Ajuste

de Tensão' e 'Mudança de Estado', apresentados separadamente na Figura 21. Os dois rótulos restantes apresentam uma leve inclinação.

Por fim, foi executado o pipeline descrito acima com 500 *n_calls* no otimizador em árvore. O resultado final é apresentado na Tabela 5.

Tabela 5 – Métricas por classe do modelo final

Classe	Precisão	*Recall*	*F1-Score*
Ajuste de Tensão	0,98	0,96	0,97
Chuva	0,98	1,00	0,99
Modulação de Geração	0,85	0,85	0,85
Mudança de Estado	0,90	0,85	0,87
Nível de Reservatório	0,98	0,93	0,95
O&M Local	0,72	0,82	0,75
Média	**0,90**	**0,90**	**0,90**

Fonte: Autora

4.2.2 Abordagem com TF-IDF

Na abordagem com TF-IDF foram testados cinco algoritmos, os dois com o melhor desempenho foram selecionados para otimização de hiperparâmetros: *Random Forest* e *Ridge Classifier*. As métricas dos *baselines* para cada modelo estão dispostas na Tabela 6.

Tabela 6 – Comparação das métricas entre os modelos de *baseline* para o *pipeline* com TF-IDF

Método	Acurácia	Precisão	*Recall*	*F1-Score*
Random Forest	**0,72**	**0,81**	**0,75**	**0,78**
Decision Tree	0,62	0,69	0,62	0,59
K-Nearest Neighbors	0,61	0,81	0,48	0,61
Ridge Classifier	**0,68**	**0,82**	**0,65**	**0,74**
Multi-layer Perceptron (MLP)	0,54	0,62	0,47	0,52

Fonte: Autora

A otimização do modelo com *Random Forest* seguiu a metodologia apresentada no modelo de otimização em árvore, com o *framework scikit-optimize* tratando o problema como uma regressão. Os hiperparâmetros escolhidos foram os mesmos. A Tabela 7 apresenta as métricas de ambos os modelos otimizados, onde é possível determinar que o *pipeline* com *Random Forest* apresentou os melhores resultados.

As métricas detalhadas por rótulo para o melhor modelo são mostradas na Tabela 8. Em geral, o modelo TF-IDF teve um desempenho pior do que o *pipeline* padrão. Embora a maioria das classes tenha apresentado F1-Scores 20% menor que o primeiro *pipeline*, *O&M Local* foi 40% pior.

Tabela 7 – Comparação das métricas dos melhores pipelines com TF-IDF

Método	**Acurácia**	**Precisão**	**Recall**	**F1-Score**
Random Forest	**0,74**	0,87	**0,78**	**0,81**
Ridge Classifier	0,70	**0,89**	0,67	0,76

Fonte: Autora

Tabela 8 – Métricas do melhor modelo com TF-IDF

Classe	Precisão	Recall	F1-Score
Ajuste de Tensão	0,96	0,94	0,95
Chuva	1,00	0,73	0,84
Modulação de Geração	0,86	0,86	0,86
Mudança de Estado	0,84	0,84	0,84
Nível de Reservatório	1,00	0,65	0,79
O&M Local	0,53	0,69	0,53
Média	**0,87**	**0,78**	**0,81**

Fonte: Autora

Isso provavelmente ocorreu devido à complexidade da classe *O&M Local*. Enquanto os outros rótulos são baseados em regras, com palavras-chave estabelecidas para cada classe, *O&M* não é. Ligações nesse contexto conterá palavras e frases que o algoritmo aprendeu como relevantes para descrever outras classes, mas com contexto, significaria uma operação ou manutenção na usina e não uma modulação de geração, por exemplo.

Dito isso, BERT pode ajudar a mitigar essa situação.

4.2.3 Abordagem com BERT

Por fim, foram realizados experimentos com a abordagem utilizando BERT. Para permitir a previsão de apenas um ou dois rótulos com o modelo baseado em BERT, cada classe foi treinada individualmente. Dessa forma, havia 6 modelos de BERT, um para cada classe: (1) Mudança de estado, (2) Modulação de geração, (3) Mudança de tensão, (4) Nível do reservatório, (5) Chuva e (6) O&M Local.

A Tabela 9 apresenta os resultados de cada modelo (cada classe) para o BERT. Com a exceção de 'O&M Local', todos os modelos performaram acima de 0,95 de F1-Score.

4.2.4 Modelo Adotado em Produção

Ao comparar os resultados das três abordagens propostas nas seções passadas (Tabela 10), 4 das 6 classes apresentaram melhores resultados com o BERT. 'Modulação de Geração', 'Mudança de Estado', 'Nível de Reservatório' e 'O&M Local', respectivamente, melhoraram 6,89%, 2,11%, 15,29% e 16% quando comparadas com o pipeline padrão. Quando comparado ao TF-IDF, a melhora variou de 10% a 64%.

Tabela 9 – Métricas do modelo de classificação com BERT

Rótulo	Precisão	*Recall*	*F1-Score*
Ajuste de Tensão	0,96	0,96	0,96
Chuva	0,98	0,98	0,98
Modulação de Geração	0,93	0,92	0,93
Mudança de Estado	0,97	0,97	0,97
Nível de Reservatório	0,98	0,98	0,98
O&M Local	0,87	0,87	0,87
Média	**0,95**	**0,95**	**0,95**

Fonte: Autora

Tabela 10 – Comparação de F1-Score das diferentes abordagens para cada classe

	Model's F1-Score		
Label	***Baseline***	**TF-IDF**	**BERT**
Voltage Adjustment	**0,97**	0,95	0,96
Rain	**0,99**	0,84	0,98
Generation Modulation	0,87	0,84	**0,93**
State Change	0,95	0,79	**0,97**
Reservoir Levels	0,85	0,86	**0,98**
Local O&M	0,75	0,53	**0,87**

Fonte: Autora

Como os modelos retreinados com BERT são computacionalmente caros, a solução ótima para esse problema multiclasse seria um modelo misto: uma combinação entre o pipeline padrão e o BERT retreinado.

Visando não comprometer a operação em tempo real, pode ser proposta a utilização de apenas um dos seis modelos de BERT, *O&M Local*. Essa classe foi escolhida por apresentar o pior resultado entre as classes, assim como também apresentou a maior melhora no BERT, com 16%. O pipeline ideal alcançou uma média F1-Score de 0,92.

Todavia, em produção, são transcritos áudios a cada hora, em tempo real, e até mesmo a predição com BERT tem um custo computacional inconveniente (2 horas para classicar 1 dia de ligações), logo, no sistema foi adotado o modelo apelidado de pipeline padrão, descrito na Seção 3.3.1.

4.3 Correlação de Ligações

4.3.1 Correlação de Áudios com Eventos SCADA

Para o desenvolvimento da correlação de áudios com eventos SCADA, foi implementada uma primeira versão do algoritmo, a partir da qual foram realizados avanços e

melhorias. A ideia é sempre comparar novas versões do algoritmo com o *baseline*.

As métricas alcançadas pelo *baseline* estão apresentadas na Tabela 11. A partir delas, é possível identificar os pontos de melhoria do algoritmo. O principal deles, no caso, é o Recall para os áudios *Correlacionáveis*. Isso significa que a primeira versão do algoritmo tinha dificuldade em identificar as ligações correlacionáveis.

Tabela 11 – Métricas por classe do baseline

Métrica	Valor
Acurácia	0,73
Recall (*Correlacionáveis*)	0,277
Recall (*Não Correlacionáveis*)	0,948
Precisão (*Correlacionáveis*)	0,72
Precisão (*Não Correlacionáveis*)	0,762
F1-Score (*Correlacionáveis*)	0,4
F1-Score (*Não Correlacionáveis*)	0,845

Fonte: Autora

A partir do *baseline* e de análises com os 200 áudios de PBX Piratininga utilizados para as métricas, novas regras de correlação foram implementadas. O avanço das métricas a partir de cada uma das modificações é exibido na Figura 22. Entre o baseline e a versão mais recente do algoritmo, houve um aumento de 9,5% de acurácia e de 32,9% de Recall para correlacionáveis. É importante mencionar que essas métricas se referem à classificação entre *Correlacionáveis* e *Não Correlacionáveis*. Uma análise da acurácia do algoritmo com relação à associação correta de itens do ADMS foi realizada separadamente.

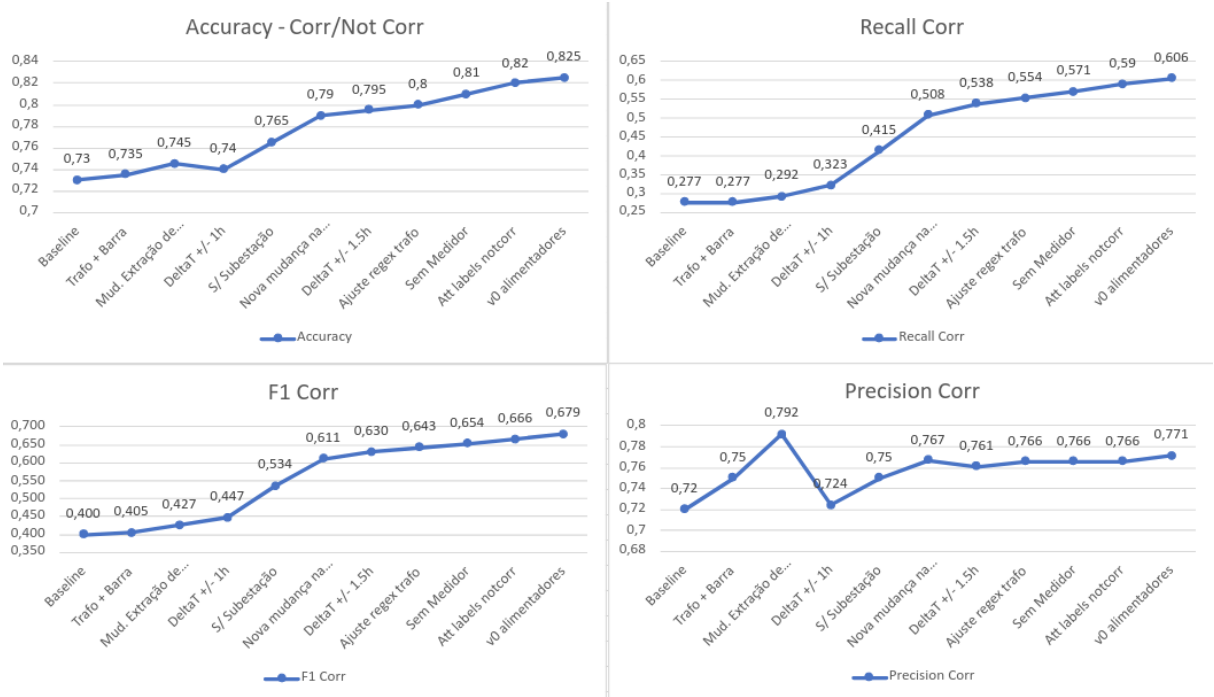
Para avaliar a acurácia, F1, *recall* e precisão do algoritmo com relação as associações, foram escutados todos os áudios classificados como correlacionáveis e analisados um a um os eventos ADMS correlacionados pelo algoritmo. Feito isso, foi encontrada uma acurácia de 86,14% e um *recall* de itens correlacionáveis de 60,6%. Considerando que o algoritmo está suscetível a erros de transcrição do *Speech to Text* do GCP, também foi realizada uma análise considerando o *groundtruth* das transcrições, com a acurácia chegando a 93,11%.

4.3.2 Correlação de Histórias

Para validação das correlações de histórias, tanto a baseada em eventos SCADA, quanto a de áudios de rádio, foram utilizados dois critérios:

- Acurácia do Pertencimento em uma História: Classificar corretamente um áudio como parte de uma história (com/sem história);
- Acurácia dos áudios pertencentes à História: Dada uma história, quantos dos áudios presentes foram associados corretamente.

Figura 22 – Gráfico da evolução das métricas do algoritmo



Fonte: Autora

4.3.2.1 Histórias baseadas em Eventos SCADA

Tal qual outros algoritmos apresentados nesta dissertação, a implementação passou pelo desenvolvimento de um *baseline* para normalizar todo os incrementos e melhorias seguintes.

A Tabela 12 apresenta as métricas do *baseline*. Ter 64% de acurácia do pertencimento em uma história significa que das 200 ligações, 112 foram classificadas erroneamente com/sem história. Já os 88% de acurácia dos áudios pertencentes à história indica que das 150 ligações pertencentes a uma história 132 foram clusterizadas corretamente.

Tabela 12 – Métricas do *baseline* da correlação de histórias baseadas no ADMS

Métrica	Valor
Acurácia do Pertencimento em uma História	0,64
Acurácia dos áudios pertencentes à História	0,88

Fonte: Autora

A partir da análise das métricas e dos dados no *baseline*, como descrito na metodologia, foram realizadas modificações no algoritmo. Na Tabela 13 é apresentada a evolução das métricas do *baseline* para versão final do algoritmo. Na primeira acurácia houve um aumento de 11% e na segunda, de 1%.

Tabela 13 – Métricas da correlação de histórias baseadas no ADMS

Métrica	Baseline	Final
Acurácia do Pertencimento em uma História	0,64	0,75
Acurácia dos áudios pertencentes à História	0,88	0,89

Fonte: Autora

4.3.2.2 Histórias com Áudios de Rádio

Para a validação do algoritmo de histórias com áudios de rádio foram selecionados 100 áudios sem história e 100 áudios com história. Logo, 72% de Acurácia do Pertencimento em uma História implica que 144 dos 200 áudios foram identificados corretamente como com/sem história. E dos que devem pertencer a uma história, 94% foram inseridos corretamente no grupo de áudios que relata a mesma ocorrência. As métricas são exibidas na Tabela 14.

Tabela 14 – Métricas do *baseline* da correlação de histórias de rádio

Métrica	Valor
Acurácia do Pertencimento em uma História	0,72
Acurácia dos áudios pertencentes à História	0,94

Fonte: Autora

4.4 Assistente Virtual para Centros de Operação

Além dos resultados dos modelos de classificação e correlações, que foram apresentados nas seções anteriores, é importante destacar os impactos da ferramenta nos centros de operação em que ela foi implantada. Essa análise, juntamente com a apresentação da solução *online*, será apresentada nesse capítulo.

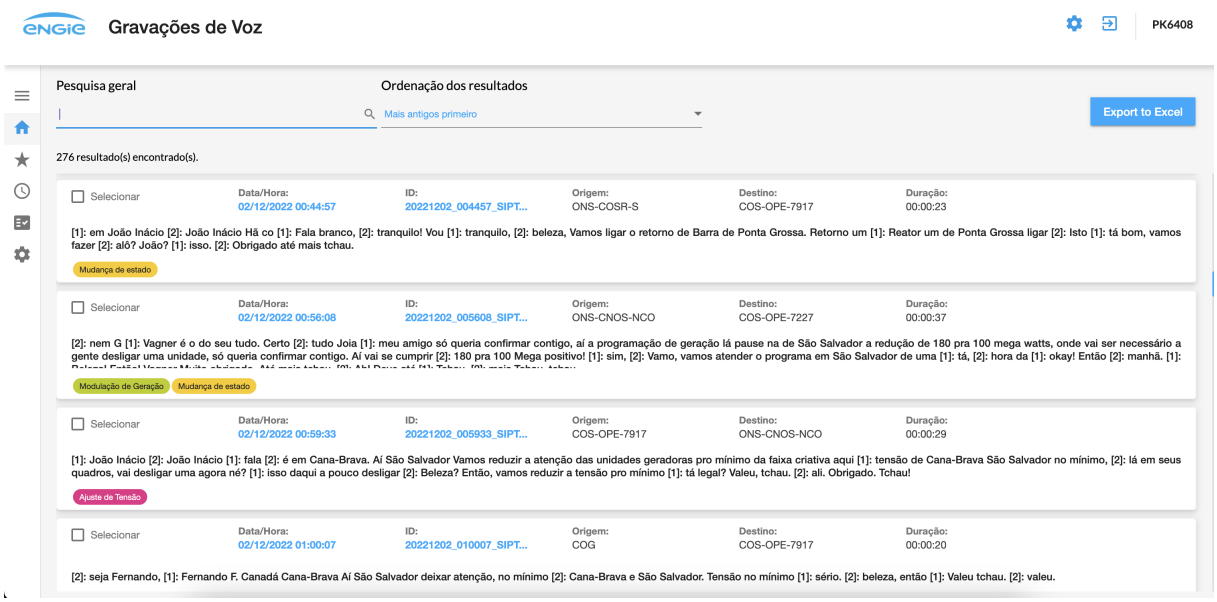
A ferramenta está *online* e em uso pelos centros de operação da Engie e CPFL desde setembro de 2021. Nesse período, foram levantadas estatísticas de duração de atividades da pós-operação com e sem a ferramenta. Um operador levantou a média do tempo de duração da atividade "busca por ligações (relacionadas a uma determinada ocorrência)" e comparou com a média de tempo da mesma busca com a ferramenta. O levantamento considerou 20 ocorrências e está disponível na Tabela 4.

Quadro 4 – Levantamento do tempo de busca por ligações

Sem a Ferramenta	Entre 20 minutos e 2 horas
Com a Ferramenta	Entre 2 e 5 minutos

Fonte: Autora

Figura 23 – Tela de busca da ferramenta com as classificações



Fonte: Autora

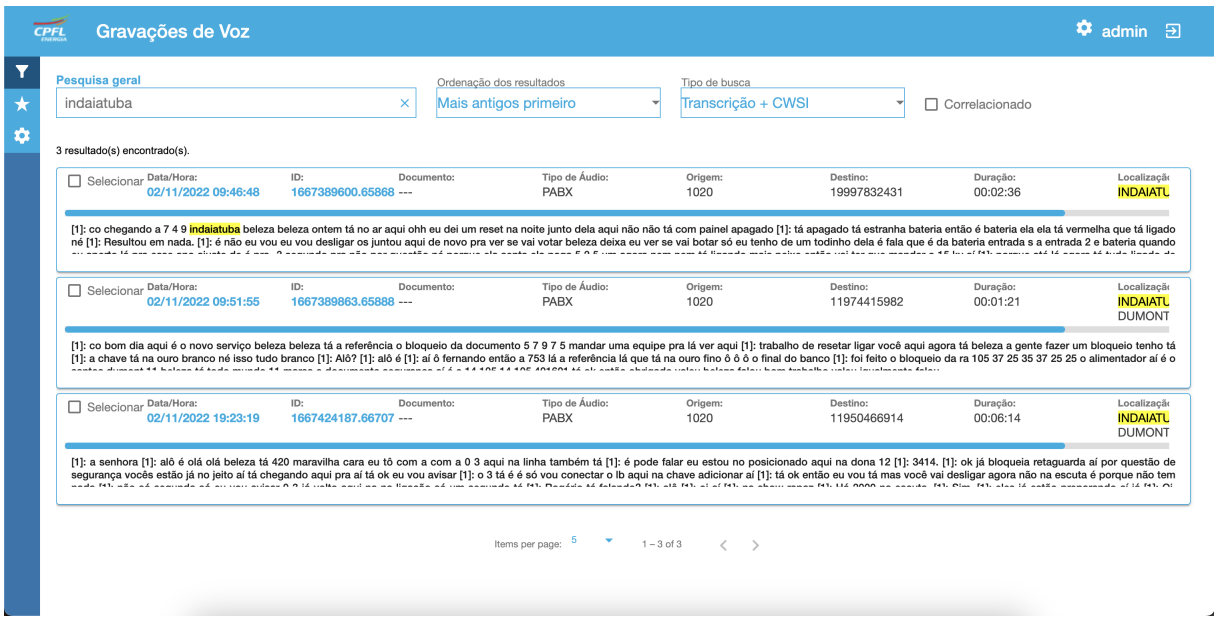
Logo, no universo de testes realizados, foi possível identificar uma diminuição de até 95% do tempo gasto apenas com a busca dos áudios, ainda sem considerar as análises e correlações.

Entre os principais benefícios trazidos pela ferramenta para a pós-operação estão a busca na transcrição por palavras-chave, localidade, data, ramais, usinas e equipamentos, além da classificação das ligações pela atividade que elas tratam. Nas Figuras 23 e 24 são apresentados exemplos.

Na Figura 24, há um exemplo de uma ocorrência na subestação de Indaiatuba que aconteceu no dia 02 de novembro de 2022. Sem a ferramenta, o engenheiro de pós-operação seria forçado a escutar uma quantidade excessiva de áudios para encontrar apenas esses três que se referem a subestação. A ferramenta, além de facilitar a busca, traz os eventos SCADA associados a ocorrência no local, melhorando a qualidade da análise realizada pelo engenheiro e reduzindo a possibilidade de falha humana.

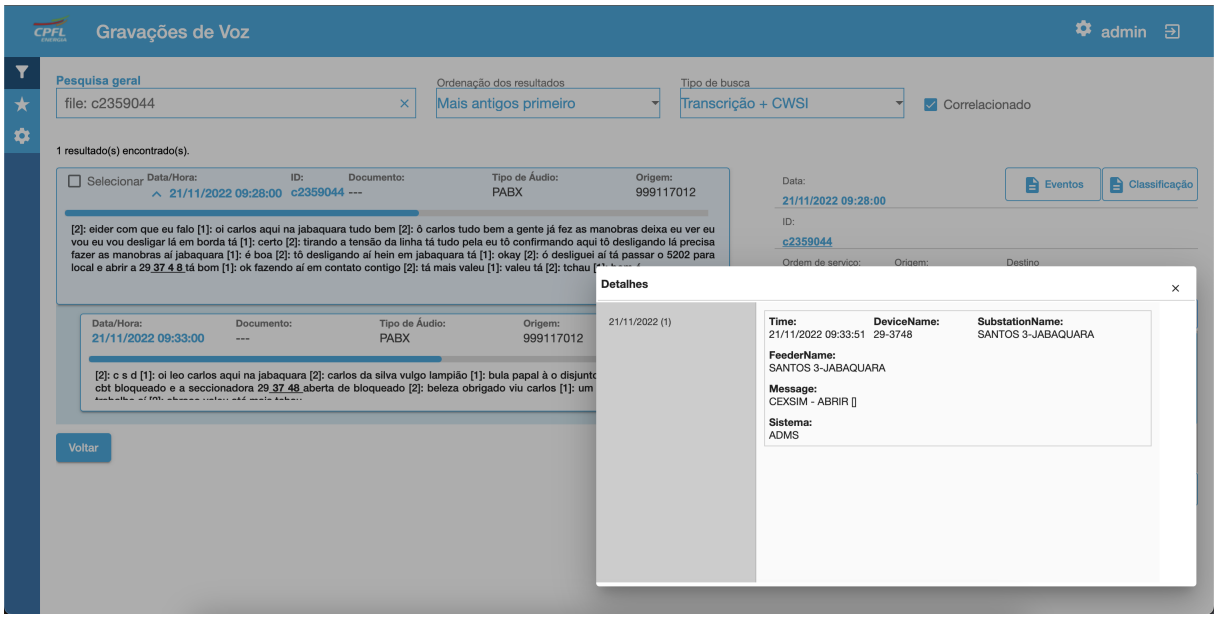
Outro exemplo de correlação SCADA com ligações pode ser visto na Figura 25. Duas ligações tratando da chave seccionadora (29-3748) foram associadas entre si e com o evento SCADA que ambas se referem. Na primeira ligação uma ação é atribuída (abrir a chave) para o operador de campo, enquanto na segunda ligação, há a confirmação da realização da tarefa. O sistema além de associar as ligações entre si, também as correlaciona com o evento SCADA citado durante a ligação.

Figura 24 – Tela de busca da ferramenta com exemplo de “Indaiatuba”



Fonte: Autora

Figura 25 – Tela de busca da ferramenta com exemplo de correlação aberto



Fonte: Autora

5 Considerações Finais e Trabalhos Futuros

A pós-operação do setor elétrico desempenha um papel crucial em auditoria, prevenção de erros e gerenciamento de ocorrências. O setor elétrico brasileiro, especificamente a transmissão e geração, é coordenado e controlado pelo ONS. As comunicações relacionadas às atividades de tempo real gerenciadas pelo ONS devem ocorrer, obrigatoriamente, via serviços telefônicos. Em outras palavras, a comunicação entre Centros de Operação do Sistema (COS), ONS e as usinas de geração devem ocorrer via ligação telefônica. E, apesar dessas ligações serem gravadas, o trabalho de analisá-las e entendê-las para aplicações de pós-operação é extremamente manual. Visando ajudar na tomada de decisão e na análise das transcrições de forma automatizada, este trabalho apresentou um assistente virtual para centros de operação.

O sistema foi desenvolvido e apresentou resultados em três provedores de nuvem: *AWS*, *GCP* e *Azure*. Dessa forma, além de toda a arquitetura discriminada nas diferentes plataformas, também foi comparado o treinamento do modelo de transcrição de áudios nos três provedores.

A partir das transcrições, foram apresentadas duas principais aplicações baseadas em processamento de linguagem natural: a classificação *multi-label* de áudios e a correlação de áudios entre si e com sistemas operativos. Na classificação de áudios, foram comparadas três metodologias diferentes, uma delas estado-da-arte e apresentados os resultados de cada abordagem. O modelo utilizado na ferramenta, apesar de não ser o de melhor desempenho devido ao custo computacional do BERT, apresenta *F1-Score* médio de 0,90.

A ferramenta está *online* e em uso pelos centros de operação da ENGIE Brasil Energia e CPFL Energia desde setembro de 2021. Nesse período, foram levantadas estatísticas de duração de atividades da pós-operação com e sem a ferramenta, podendo uma análise cair de 2 horas para 5 minutos de duração.

Dessa forma, o projeto contribuiu para maior eficiência, segurança e confiabilidade operacional da operação e pós-operação do sistema elétrico.

Para trabalhos futuros, estão sendo desenvolvidas integrações com novos sistemas operativos e melhorias na transcrição e nos algoritmos de classificação e de correlação de ligações. Principalmente, considerando que o campo de NLP avançou muito nos últimos meses com o advento de *Large Language Models* (LLMs).

Referências

AIZAWA, Y.; KATO, M.; KOSAKA, T. Many-to-many voice conversion using hidden markov model-based speech recognition and synthesis. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 140, n. 4, p. 2964–2965, 2016. Citado 2 vezes nas páginas 14 e 17.

AL-TALIB, G. A.; HASSAN, H. S. A study on analysis of sms classification using tf-idf weighting. *International Journal of Computer Networks and Communications Security*, v. 1, n. 5, p. 189–194, 2013. Citado na página 15.

AMAZON, A. *PT*. 2022. Disponível em: <https://aws.amazon.com/pt/transcribe/features/>. Citado na página 32.

AMODEI, D. et al. Deep speech 2 : End-to-end speech recognition in english and mandarin. In: BALCAN, M. F.; WEINBERGER, K. Q. (Ed.). *Proceedings of The 33rd International Conference on Machine Learning*. New York, New York, USA: PMLR, 2016. (Proceedings of Machine Learning Research, v. 48), p. 173–182. Disponível em: <https://proceedings.mlr.press/v48/amodei16.html>. Citado 2 vezes nas páginas 14 e 18.

ARDILA, R. et al. Common voice: A massively-multilingual speech corpus. *CoRR*, abs/1912.06670, 2019. Disponível em: <http://arxiv.org/abs/1912.06670>. Citado 2 vezes nas páginas 14 e 18.

AZURE, C. S. *Speech-to-text Docs*. 2022. Disponível em: <https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/speech-to-text>. Citado na página 36.

BHABLE, S. G.; KAYTE, C. N. Review: Multilingual acoustic modeling of automatic speech recognition(asr) for low resource languages. In: *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*. [S.l.: s.n.], 2020. p. 1–4. Citado na página 19.

BIAU, G.; SCORNET, E. A random forest guided tour. *Test*, Springer, v. 25, p. 197–227, 2016. Citado na página 63.

BIBI, S.; KATSAROS, D.; BOZANIS, P. Business application acquisition: On-premise or saas-based solutions? *IEEE Software*, v. 29, n. 3, p. 86–93, 2012. Citado na página 17.

BOJANIć, M. et al. Call redistribution for a call center based on speech emotion recognition. *Applied Sciences*, MDPI AG, v. 10, n. 13, p. 4653, Jul 2020. ISSN 2076-3417. Disponível em: <http://dx.doi.org/10.3390/app10134653>. Citado na página 14.

BONIN, F.; PEDRO, J. S.; OLIVER, N. A context-aware nlp approach for noteworthiness detection in cellphone conversations. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. [S.l.: s.n.], 2014. p. 25–36. Citado na página 20.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 43.

CASTRO, N. José de et al. *Programa de P&D da ANEEL: Avaliação & Perspectivas*. [S.l.]: Publit, 2020. Citado na página 13.

CHAN, W. et al. *Listen, Attend and Spell*. arXiv, 2015. Disponível em: <https://arxiv.org/abs/1508.01211>. Citado na página 19.

CHEN, G. et al. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In: IEEE. *2017 International joint conference on neural networks (IJCNN)*. [S.l.], 2017. p. 2377–2383. Citado na página 21.

CHEN, Y. et al. Multi-label text classification with deep neural networks. In: IEEE. *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*. [S.l.], 2018. p. 409–413. Citado na página 21.

CLERCQ, D. D. et al. Multi-label classification and interactive nlp-based visualization of electric vehicle patent data. *World Patent Information*, v. 58, p. 101903, 2019. ISSN 0172-2190. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0172219018300851>. Citado na página 14.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. Citado na página 21.

DIAS, A. L. et al. Towards a free, forced phonetic aligner for brazilian portuguese using kaldi tools. In: CERRI, R.; PRATI, R. C. (Ed.). *Intelligent Systems*. Cham: Springer International Publishing, 2020. p. 621–635. ISBN 978-3-030-61377-8. Citado na página 19.

DIKAIAKOS, M. D. et al. Cloud computing: Distributed internet computing for it and scientific research. *IEEE Internet Computing*, v. 13, n. 5, p. 10–13, 2009. Citado na página 17.

EPE, E. d. P. E. *Balanço Energético Nacional de 2021*. 1. ed. [S.l.], 2022. Citado na página 13.

ESMAEILZADEH, A.; TAGHVA, K. Text classification using neural network language model (nnlm) and bert: An empirical comparison. In: ARAI, K. (Ed.). *Intelligent Systems and Applications*. Cham: Springer International Publishing, 2022. p. 175–189. ISBN 978-3-030-82199-9. Citado na página 21.

ETEMADI, N. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Springer, v. 55, n. 1, p. 119–122, 1981. Citado na página 47.

FILGUEIRAS, A.; SILVA, T. M. V. Wind energy in brazil—present and future. *Renewable and Sustainable Energy Reviews*, v. 7, n. 5, p. 439–451, 2003. ISSN 1364-0321. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364032103000686>. Citado na página 13.

FILHO, J. A. W. et al. The brwac corpus: A new open resource for brazilian portuguese. In: *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. [S.l.: s.n.], 2018. Citado na página 21.

GANIM, A. *Setor elétrico brasileiro: aspectos regulamentares, tributários e contábeis*. [S.l.]: Synergia, 2008. v. 1. ISBN 978-85-61325-13-8. Citado na página 13.

GHOSH, S.; DESARKAR, M. S. Class specific tf-idf boosting for short-text classification: Application to short-texts generated during disasters. In: *Companion Proceedings of the The Web Conference 2018*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018. (WWW '18), p. 1629–1637. ISBN 9781450356404. Disponível em: <https://doi.org/10.1145/3184558.3191621>. Citado na página 42.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado 2 vezes nas páginas 14 e 18.

GOOGLE, G. C. P. *Speech-to-text documentation*. Google, 2022. Disponível em: <https://cloud.google.com/speech-to-text/docs>. Citado na página 34.

GRAVES, A.; JAITLY, N. Towards end-to-end speech recognition with recurrent neural networks. In: XING, E. P.; JEBARA, T. (Ed.). *Proceedings of the 31st International Conference on Machine Learning*. Bejing, China: PMLR, 2014. (Proceedings of Machine Learning Research, 2), p. 1764–1772. Disponível em: <https://proceedings.mlr.press/v32/graves14.html>. Citado na página 19.

HAINES, S. Workflow orchestration with apache airflow. In: _____. *Modern Data Engineering with Apache Spark: A Hands-On Guide for Building Mission-Critical Streaming Applications*. Berkeley, CA: Apress, 2022. p. 255–295. ISBN 978-1-4842-7452-1. Disponível em: <https://doi.org/10.1007/978-1-4842-7452-1_8>. Citado na página 25.

HAYKIN, S. *Neural networks: a comprehensive foundation*. [S.l.]: Prentice Hall PTR, 1994. Citado na página 43.

HE, X.; ZHAO, K.; CHU, X. Automl: A survey of the state-of-the-art. *Knowledge-Based Systems*, Elsevier, v. 212, p. 106622, 2021. Citado na página 41.

HE, Y.; DONG, X. Real time speech recognition algorithm on embedded system based on continuous markov model. *Microprocessors and Microsystems*, v. 75, p. 103058, 2020. ISSN 0141-9331. Disponível em: <https://www.sciencedirect.com/science/article/pii/S014193312030123X>. Citado na página 19.

HE, Y. et al. Streaming end-to-end speech recognition for mobile devices. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2019. p. 6381–6385. Citado 2 vezes nas páginas 14 e 18.

HILT, D. E.; SEEGRIST, D. W. *Ridge, a computer program for calculating ridge regression estimates*. [S.l.]: Department of Agriculture, Forest Service, Northeastern Forest Experiment . . . , 1977. v. 236. Citado na página 43.

HINTON, G. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, v. 29, n. 6, p. 82–97, 2012. Citado na página 18.

IMANE, A.; MOHAMED, B. A. Multi-label categorization of french death certificates using nlp and machine learning. In: *Proceedings of the 2nd International Conference on Big Data, Cloud and Applications*. New York, NY, USA: Association for Computing Machinery, 2017. (BDCA'17). ISBN 9781450348522. Disponível em: <https://doi.org/10.1145/3090354.3090384>. Citado na página 14.

IMANE, A.; MOHAMED, B. A. Multi-label categorization of french death certificates using nlp and machine learning. In: *Proceedings of the 2nd international conference on big data, cloud and applications*. [S.l.: s.n.], 2017. p. 1–4. Citado na página 20.

JAITLY, N.; HINTON, G. Learning a better representation of speech soundwaves using restricted boltzmann machines. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2011. p. 5884–5887. Citado na página 18.

KANT, K. Data center evolution: A tutorial on state of the art, issues, and challenges. *Computer Networks*, v. 53, n. 17, p. 2939–2965, 2009. ISSN 1389-1286. Virtualized Data Centers. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128609003090>. Citado na página 17.

KIRK, P. et al. Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics*, v. 28, n. 24, p. 3290–3297, 10 2012. ISSN 1367-4803. Disponível em: <https://doi.org/10.1093/bioinformatics/bts595>. Citado 2 vezes nas páginas 15 e 22.

LE, T. T.; FU, W.; MOORE, J. H. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*, Oxford University Press, v. 36, n. 1, p. 250–256, 2020. Citado na página 41.

LEARN, S. *Sklearn.multioutput.ClassifierChain*. 2020. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.ClassifierChain.html#sklearn.multioutput.ClassifierChain>. Citado na página 41.

LEARN, S. *Sklearn.multioutput.multioutputclassifier*. 2020. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.MultiOutputClassifier.html>. Citado na página 63.

LIN, S.-Y. et al. Validation of energy expenditure prediction models using real-time shoe-based motion detectors. *IEEE transactions on biomedical engineering*, IEEE, v. 64, n. 9, p. 2152–2162, 2016. Citado 2 vezes nas páginas 14 e 17.

LIU, X. et al. Clinical trial information extraction with bert. In: IEEE. *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*. [S.l.], 2021. p. 505–506. Citado na página 15.

LOGUMANOV, A. G.; KLENIN, J. D.; BOTOV, D. S. Sentiment analysis of telephone conversations using multimodal data. In: SPRINGER. *International Conference on Analysis of Images, Social Networks and Texts*. [S.l.], 2018. p. 88–98. Citado na página 20.

LORENA, A. C. et al. How complex is your classification problem? a survey on measuring classification complexity. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 52, n. 5, sep 2019. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/3347711>. Citado na página 20.

LOUPPE, G. Bayesian optimisation with scikit-optimize. In: *PyData Amsterdam*. [S.l.: s.n.], 2017. Citado na página 41.

LOVINS, J. B. Development of a stemming algorithm. *Mech. Transl. Comput. Linguistics*, v. 11, n. 1-2, p. 22–31, 1968. Citado na página 43.

MCLEAN, G.; OSEI-FRIMPONG, K. Hey alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, v. 99, p. 28–37, 2019. ISSN 0747-5632. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0747563219301840>. Citado 3 vezes nas páginas 14, 17 e 18.

MEHTA, S. et al. Birdeview: Advance version of call monitoring system by using mining techniques. In: *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. [S.l.: s.n.], 2020. p. 561–566. Citado na página 14.

MUCHERINO, A. et al. K-nearest neighbor classification. In: *Data mining in agriculture*. [S.l.]: Springer, 2009. p. 83–106. Citado na página 43.

MUSTAFA, M. K.; ALLEN, T.; APPIAH, K. A comparative review of dynamic neural networks and hidden markov model methods for mobile on-device speech recognition. *Neural Computing and Applications*, v. 31, n. 2, p. 891–899, Feb 2019. ISSN 1433-3058. Disponível em: <https://doi.org/10.1007/s00521-017-3028-2>. Citado 2 vezes nas páginas 14 e 18.

NARAYANAN, A.; MEENAKUMARI. Enterprise systems and emerging technologies - a futuristic perspective and recommendations for a paradigm shift and sustainability. In: VENUGOPAL, K. R. et al. (Ed.). *Data Science and Computational Intelligence*. Cham: Springer International Publishing, 2021. p. 105–117. ISBN 978-3-030-91244-4. Citado na página 17.

NUGROHO, K. S.; SUKMADEWA, A. Y.; YUDISTIRA, N. Large-scale news classification using bert language model: Spark nlp approach. In: *6th International Conference on Sustainable Information Engineering and Technology 2021*. New York, NY, USA: Association for Computing Machinery, 2021. (SIET '21), p. 240–246. ISBN 9781450384070. Disponível em: <https://doi.org/10.1145/3479645.3479658>. Citado na página 21.

ONS. *Submódulo 2.10: Norma de Operação - Hierarquia Operacional*. 2012.12. ed. [S.l.], 2003. Disponível em: <https://bit.ly/3OZS8N9>. Citado na página 13.

PETERSEN, C.; SEYMOUR, L. F. Explaining the network of factors that influence the timing of and decision to upgrade enterprise systems. In: DENNEHY, D. et al. (Ed.). *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*. Cham: Springer International Publishing, 2021. p. 506–518. ISBN 978-3-030-85447-8. Citado na página 17.

PLATFORM, G. C. *Improve transcription accuracy with speech adaptation boost*. [S.l.], 2021. Disponível em: <https://cloud.google.com/speech-to-text/docs/boost>. Citado na página 35.

QAISER, S.; ALI, R. Text mining: use of tf-idf to examine the relevance of words to documents. *International Journal of Computer Applications*, v. 181, n. 1, p. 25–29, 2018. Citado na página 43.

QUINTANILHA, I. M.; NETTO, S. L.; BISCAINHO, L. P. An open-source end-to-end asr system for brazilian portuguese using dnns built from newly assembled corpora. *Journal of Communication and Information Systems*, v. 35, n. 1, p. 230–242, Sep. 2020. Disponível em: <https://jcis.sbrt.org.br/jcis/article/view/721>. Citado 3 vezes nas páginas 14, 17 e 18.

RAO, K. et al. Grapheme-to-phoneme conversion using long short-term memory recurrent neural networks. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2015. p. 4225–4229. Citado na página 18.

ROKACH, L.; MAIMON, O. Decision trees. In: *Data mining and knowledge discovery handbook*. [S.l.]: Springer, 2005. p. 165–192. Citado na página 43.

SAGER, N. Natural language information formatting: The automatic conversion of texts to a structured data base. In: YOVITS, M. C. (Ed.). Elsevier, 1978, (Advances in Computers, v. 17). p. 89–162. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0065245808603915>. Citado 2 vezes nas páginas 15 e 22.

SAVYANAVAR, P. et al. Multi-document summarization using tf-idf algorithm. *International Journal Of Engineering And Computer Science*, v. 5, n. 4, p. 16253–16256, 2016. Citado na página 15.

SCHMITT, X. et al. A replicable comparison study of ner software: Stanfordnlp, nltk, opennlp, spacy, gate. In: IEEE. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. [S.l.], 2019. p. 338–343. Citado na página 20.

SHEWALKAR, A. Performance evaluation of deep neural networks applied to speech recognition : Rnn, lstm and gru. *Journal of Artificial Intelligence and Soft Computing Research*, Vol. 9, No. 4, p. 235–245, 2019. Opracowanie rekordu w ramach umowy 509/P-DUN/2018 ze środków MNiSW przeznaczonych na działalność upowszechniającą naukę (2019). Citado na página 18.

SOUZA, F. et al. Bertimbau: pretrained bert models for brazilian portuguese. In: SPRINGER. *Brazilian Conference on Intelligent Systems*. [S.l.], 2020. p. 403–417. Citado 3 vezes nas páginas 21, 43 e 44.

SUN, F. et al. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2019. (CIKM '19), p. 1441–1450. ISBN 9781450369763. Disponível em: <https://doi.org/10.1145/3357384.3357895>. Citado na página 21.

SURESH, S.; SARASWATHI, S.; SUNDARARAJAN, N. Performance enhancement of extreme learning machine for multi-category sparse data classification problems. *Engineering Applications of Artificial Intelligence*, v. 23, n. 7, p. 1149–1157, 2010. ISSN 0952-1976. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0952197610001326>. Citado na página 20.

TAKEUCHI, H. et al. Mining of business-oriented conversations at a call center. *Transactions of the Japanese Society for Artificial Intelligence*, v. 23, n. 6, p. 384–391, 2008. Citado na página 14.

TANAKA, T. et al. Neural speech-to-text language models for rescoring hypotheses of dnn-hmm hybrid automatic speech recognition systems. In: *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. [S.l.: s.n.], 2018. p. 196–200. Citado na página 19.

TEKLI, J. An overview on xml semantic disambiguation from unstructured text to semi-structured data: Background, applications, and ongoing challenges. *IEEE Transactions on Knowledge and Data Engineering*, v. 28, n. 6, p. 1383–1407, 2016. Citado 2 vezes nas páginas 15 e 22.

TULSHAN, A. S.; DHAGE, S. N. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In: THAMPI, S. M. et al. (Ed.). *Advances in Signal Processing and Intelligent Recognition Systems*. Singapore: Springer Singapore, 2019. p. 190–201. ISBN 978-981-13-5758-9. Citado 2 vezes nas páginas 14 e 17.

VASILATEANU, A.; ENE, R. Call-center virtual assistant using natural language processing and speech recognition. *Journal of ICT, Design, Engineering and Technological Science*, v. 2, p. 40–46, 12 2018. Citado na página 14.

VIJAYARANI, S. et al. Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, v. 5, n. 1, p. 7–16, 2015. Citado na página 19.

WOLF, T. et al. Transformers: State-of-the-art natural language processing. In: *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*. [S.l.: s.n.], 2020. p. 38–45. Citado na página 44.

WU, S. et al. Research on trend prediction of internet user intention understanding and public intelligence mining based on fractional differential method. *Chaos, Solitons & Fractals*, Elsevier, v. 128, p. 331–338, 2019. Citado 2 vezes nas páginas 14 e 17.

ZHANG, H.; ZHUANG, Y.; WU, F. Cross-modal correlation learning for clustering on image-audio dataset. In: *Proceedings of the 15th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2007. (MM '07), p. 273–276. ISBN 9781595937025. Disponível em: <https://doi.org/10.1145/1291233.1291290>. Citado 2 vezes nas páginas 15 e 22.

ZHANG, W. et al. Towards end-to-end speech recognition with deep multipath convolutional neural networks. In: SPRINGER. *International Conference on Intelligent Robotics and Applications*. [S.l.], 2019. p. 332–341. Citado na página 19.

ÇIŞMAN, B.; LI, H.; TAN, K. C. Sparse representation of phonetic features for voice conversion with and without parallel data. In: *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. [S.l.: s.n.], 2017. p. 677–684. Citado na página 19.