



**MARCELO RODRIGO DOS SANTOS**

**NOVOS INDICADORES DE QoE PARA REDES DE SERVIÇOS  
MÓVEIS POR MEIO DA EXTRAÇÃO DE OPINIÕES DAS  
REDES SOCIAIS**

**LAVRAS – MG  
2023**

**MARCELO RODRIGO DOS SANTOS**

**NOVOS INDICADORES DE QoE PARA REDES DE SERVIÇOS MÓVEIS POR MEIO DA  
EXTRAÇÃO DE OPINIÕES DAS REDES SOCIAIS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para a obtenção do título de Mestre.

Prof. DSc. Renata Lopes Rosa  
Orientadora

Prof. DSc. Demóstenes Zegarra Rodríguez  
Coorientador

**LAVRAS – MG  
2023**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Santos, Marcelo Rodrigo dos.

Novos Indicadores de QoE para Redes de Serviços Móveis por Meio da Extração de Opiniões das Redes Sociais / MARCELO RODRIGO DOS SANTOS. - 2023.. – Lavras : UFLA, 2023.

113 p. :

Dissertação (Mestrado) – Universidade Federal de Lavras, 2023.

Orientadora: Prof. DSc. Renata Lopes Rosa .

Bibliografia.

1. Qualidade de Experiência (QoE). 2. Redes móveis. 3. Classificação de Textos (PLN). I. Rosa, Renata Lopes. II. Rodríguez, Demóstenes Zegarra. III. Título.

**MARCELO RODRIGO DOS SANTOS**

**NOVOS INDICADORES DE QoE PARA REDES DE SERVIÇOS MÓVEIS POR MEIO DA  
EXTRAÇÃO DE OPINIÕES DAS REDES SOCIAIS**

**NEW QoE INDICATORS FOR MOBILE NETWORKS THROUGH THE EXTRACTION OF  
SOCIAL MEDIA OPINIONS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para a obtenção do título de Mestre.

APROVADA em 30 de Junho de 2023.

Prof. DSc. Renata Lopes Rosa                      UFLA  
Prof. DSc. Luiz Henrique Andrade Correia      UFLA  
Prof. DSc. Darlinton Babosa Feres Carvalho    UFSJ

Prof. DSc. Renata Lopes Rosa  
Orientadora

Prof. DSc. Demóstenes Zegarra Rodríguez  
Coorientador

**LAVRAS – MG  
2023**

*Dedico esta pesquisa à minha esposa Eveline Cristina da Silva Santos e aos meus filhos João Pedro da Silva Santos e João Gabriel da Silva Santos, bênçãos de Deus em minha vida.*

## **AGRADECIMENTOS**

Agradeço a DEUS PAI, criador dos céus e da terra.

Agradeço a DEUS JESUS CRISTO, autor e consumidor da minha Salvação.

Agradeço a DEUS ESPIRITO SANTOS, consolador presente em todo tempo.

Agradeço aos meus pais pelo incentivo incessante aos meus estudos.

Agradeço à minha esposa por ser companheira em todas as circunstâncias.

Agradeço à minha orientadora, Prof. DSc Renata Lopes Rosa, pelas preciosas orientações e direção durante todas as fases do mestrado.

Agradeço ao meu coorientador, Prof. DSc. Demóstenes Zegarra Rodriguez, pelas contribuições nesta pesquisa.

Agradeço à Universidade Federal de Lavras - UFLA, pela oportunidade de integrar como discente essa instituição.

Agradeço à CAPES por proporcionar acesso ao acervo que foi fundamental na construção desta pesquisa.

Agradeço aos membros da banca, Prof. DSc. Luiz Henrique Andrade Correia e Prof. DSc. Darlington Barbosa Feres Carvalho.

Agradeço aos demais professores com os quais tive a oportunidade de cursar disciplinas no curso de pós-graduação em Ciência da Computação.

## RESUMO

As redes móveis desempenharam um papel importante na sociedade, permitindo que os usuários realizem uma ampla variedade de atividades, como chamadas de vídeo e voz, conexão com a internet, operações bancárias, comunicação via redes sociais, entre outras. Isso tem impulsionado o mercado de telecomunicações, que busca constantemente atender a clientes cada vez mais acostumados com as facilidades oferecidas por essas redes. Diante dessa realidade, torna-se essencial que as operadoras prestadoras de serviços de redes móveis busquem compreender a qualidade de experiências vivenciadas por seus usuários. Nesse contexto, estudos sobre a qualidade de experiência (QoE) têm contribuído para a avaliação com base na experiência dos usuários em relação a determinados produtos ou serviços, sendo aplicados especialmente no campo das telecomunicações. Nesta pesquisa, nosso objetivo foi identificar novos indicadores de QoE para as principais operadoras de redes móveis no Brasil (Oi, Tim, Vivo e Claro). Para isso, foram utilizadas opiniões dos usuários nas redes sociais como base de análise. Coletamos essas opiniões no Twitter e aplicamos técnicas de aprendizado de máquina com o modelo BERT (*Bidirectional Encoder Representations from Transformers*), nas versões BERT multilíngue e BERTimbau, pois ambos os modelos possuem conhecimento linguístico para o português brasileiro. O modelo BERTimbau obteve uma precisão superior, alcançando 0,80. Na fase de predição automática, os resultados demonstraram uma paridade nos indicadores de QoE entre as operadoras nas redes sociais. Comparamos esses resultados com informações da Anatel, agência reguladora das redes móveis no país. Foi observado que as redes sociais são utilizadas principalmente pelos usuários para expressar insatisfação com os serviços, enquanto as pesquisas da Anatel revelaram diferenças na satisfação do usuário. Além disso, observou-se uma correlação entre o número de reclamações registradas na Anatel e as manifestações nas redes sociais.

**Palavras-chave:** Redes móveis. Qualidade de Experiência. BERT. Anatel.

## ABSTRACT

Mobile networks have played an important role in society, allowing users to perform a wide variety of activities, such as video and voice calls, internet connection, banking operations, social media communication, and others. This has boosted the telecommunications market, which is constantly seeking to meet customers who are increasingly accustomed to the facilities offered by these networks. In this reality, it becomes essential that mobile network service providers seek to understand the quality of experiences experienced by their users. In this context, studies on quality of experience (QoE) have contributed to the assessment based on the user experience in relation to certain products or services, being applied especially in the field of telecommunications. In this research, our objective was to identify new QoE indicators for the main mobile network operators in Brazil (Oi, Tim, Vivo and Claro). For this, user opinions on social media were used as a basis for analysis. They were collected opinions on Twitter and applied machine learning techniques with the BERT model (Bidirectional Encoder Representations from Transformers), in the BERT multilingual and BERTimbau versions, because both models have linguistic knowledge for Brazilian Portuguese. The BERTimbau model obtained a superior accuracy, reaching 0.80. In the automatic prediction phase, the results showed parity in QoE indicators among operators on social media. They were compared these results with information from Anatel, the country's mobile network regulatory agency, revealing that social media are mainly used by users to express dissatisfaction with services, while Anatel surveys revealed differences in user satisfaction. In addition, a correlation was observed between the number of complaints registered with Anatel and the demonstrations on social media.

**Keywords:** Mobile networks. Quality of Experience. BERT. Anatel.

## LISTA DE FIGURAS

Figura 2.1 – Redes Celulares . . . . .	22
Figura 2.2 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil . . . . .	25
Figura 2.3 – Processo <i>knowledge discovery in databases</i> - KDD . . . . .	30
Figura 2.4 – Grafo - RNA . . . . .	31
Figura 2.5 – Arquitetura Transformes . . . . .	37
Figura 2.6 – Pré-treinamento BERT . . . . .	39
Figura 3.1 – Fluxograma da Primeira fase - Classificador . . . . .	46
Figura 3.2 – Fluxograma da Segunda fase - Comparação de Resultados . . . . .	47
Figura 3.3 – Resumo do modelo de classificação . . . . .	55
Figura 4.1 – Matriz de confusão - BERTimbau . . . . .	62
Figura 4.2 – Precisão, <i>Recall</i> e <i>F1-Score</i> - BERTimbau . . . . .	63
Figura 4.3 – Matriz de confusão - BERT Multilíngue . . . . .	64
Figura 4.4 – Precisão, <i>Recall</i> e <i>F1-Score</i> - BERT Multilíngue . . . . .	65
Figura 4.5 – Mapa panorâmico das reclamações . . . . .	75
Figura 4.6 – Reclamações por operadora . . . . .	75
Figura 4.7 – Reclamações por operadora (São Paulo) . . . . .	82
Figura 4.8 – Reclamações por operadora (Rio de Janeiro) . . . . .	89
Figura 4.9 – Reclamações por operadora (Belo Horizonte) . . . . .	94
Figura 1 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil - Operadora TIM . . .	109
Figura 2 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil - Operadora Vivo . . .	110
Figura 3 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil - Operadora Claro . . .	111

## LISTA DE GRÁFICOS

Gráfico 2.1 – Presença das Operadoras por Municípios no Brasil . . . . .	26
Gráfico 3.1 – Definição de classes para o processos de treino e testes . . . . .	53
Gráfico 3.2 – Balanceamento de classes . . . . .	54
Gráfico 3.3 – Consulta Anatel . . . . .	58
Gráfico 3.4 – Serviços Moveis Anatel . . . . .	60
Gráfico 4.1 – Treinamento com BERTimbau . . . . .	61
Gráfico 4.2 – Treinamento com BERT Multilíngue . . . . .	62
Gráfico 4.3 – Classificação Geral . . . . .	66
Gráfico 4.4 – Classificação por operadora . . . . .	67
Gráfico 4.5 – Classificação geral por localidade . . . . .	69
Gráfico 4.6 – Classificação por operadora - São Paulo . . . . .	70
Gráfico 4.7 – Classificação por operadora - Rio de Janeiro . . . . .	71
Gráfico 4.8 – Classificação por operadora - Belo Horizonte . . . . .	73
Gráfico 4.9 – Media de reclamações recebidas na Anatel . . . . .	74
Gráfico 4.10 – Reclamações recebidas por operadora - ANATEL . . . . .	76
Gráfico 4.11 – <i>Ranking</i> de acessos aos serviços de telecomunicação por operadora . . . . .	77
Gráfico 4.12 – Pesquisa de satisfação geral - Banda Larga . . . . .	77
Gráfico 4.13 – Pesquisa de satisfação geral- Pós Pago . . . . .	78
Gráfico 4.14 – Pesquisa de satisfação geral - Pré pago . . . . .	78
Gráfico 4.15 – Pesquisa de satisfação com funcionamento - Banda Larga . . . . .	79
Gráfico 4.16 – Pesquisa de satisfação com funcionamento - Pós pago . . . . .	80
Gráfico 4.17 – Pesquisa de satisfação com funcionamento - Pré pago . . . . .	80
Gráfico 4.18 – Qualidade de atendimento das operadoras . . . . .	81
Gráfico 4.19 – Média de reclamações recebidas na Anatel (São Paulo) . . . . .	82
Gráfico 4.20 – Reclamações recebidas por operadora - Anatel (São Paulo) . . . . .	83
Gráfico 4.21 – <i>Ranking</i> de acessos aos serviços de telecomunicação por operadora - São Paulo . . . . .	84
Gráfico 4.22 – Pesquisa de satisfação geral - Banda Larga (São Paulo) . . . . .	84
Gráfico 4.23 – Pesquisa de satisfação geral - Pós pago (São Paulo) . . . . .	85
Gráfico 4.24 – Pesquisa de satisfação geral - Pré pago (São Paulo) . . . . .	85
Gráfico 4.25 – Pesquisa de satisfação com funcionamento - Banda Larga (São Paulo) . . . . .	86
Gráfico 4.26 – Pesquisa de satisfação com funcionamento - Pós pago (São Paulo) . . . . .	86
Gráfico 4.27 – Pesquisa de satisfação com funcionamento - Pré pago (São Paulo) . . . . .	87

Gráfico 4.28 – Qualidade de atendimento das operadoras (São Paulo) . . . . .	88
Gráfico 4.29 – Média de reclamações recebidas na Anatel (Rio de Janeiro) . . . . .	89
Gráfico 4.30 – Reclamações recebidas por operadora - ANATEL (Rio de Janeiro) . . . . .	89
Gráfico 4.31 – <i>Ranking</i> de acessos aos serviços de telecomunicação por operadora - Rio de Janeiro	90
Gráfico 4.32 – Pesquisa de satisfação geral - Banda Larga (Rio de Janeiro) . . . . .	90
Gráfico 4.33 – Pesquisa de satisfação geral - Pós pago (Rio de Janeiro) . . . . .	91
Gráfico 4.34 – Pesquisa de satisfação geral - Pré pago (Rio de Janeiro) . . . . .	91
Gráfico 4.35 – Pesquisa de satisfação com funcionamento - Banda Larga (Rio de Janeiro) . . . . .	92
Gráfico 4.36 – Pesquisa de satisfação com funcionamento - Pós pago (Rio de Janeiro) . . . . .	92
Gráfico 4.37 – Pesquisa de satisfação com funcionamento - Pré pago (Rio de Janeiro) . . . . .	93
Gráfico 4.38 – Qualidade de atendimento das operadoras (Rio de Janeiro) . . . . .	94
Gráfico 4.39 – Média de reclamações recebidas na Anatel (Belo Horizonte) . . . . .	94
Gráfico 4.40 – Reclamações recebidas por operadora - ANATEL (Belo Horizonte) . . . . .	95
Gráfico 4.41 – <i>Ranking</i> de acessos aos serviços de telecomunicação por operadora - Belo Horizonte	95
Gráfico 4.42 – Pesquisa de satisfação geral - Banda Larga (Minas Gerais) . . . . .	96
Gráfico 4.43 – Pesquisa de satisfação geral - Pós pago (Minas Gerais) . . . . .	96
Gráfico 4.44 – Pesquisa de satisfação geral - Pré pago (Minas Gerais) . . . . .	97
Gráfico 4.45 – Pesquisa de satisfação com funcionamento - Banda Larga (Minas Gerais) . . . . .	97
Gráfico 4.46 – Pesquisa de satisfação com funcionamento - Pós pago (Minas Gerais) . . . . .	98
Gráfico 4.47 – Pesquisa de satisfação com funcionamento - Pré pago (Minas Gerais) . . . . .	98
Gráfico 4.48 – Qualidade de atendimento das operadoras (Minas Gerais) . . . . .	99

## LISTA DE TABELAS

Tabela 2.1 – Escala de pontuação MOS . . . . .	28
Tabela 4.1 – Avaliação Geral . . . . .	66
Tabela 4.2 – Classificação por Operadora . . . . .	68
Tabela 4.3 – Classificação por localidade . . . . .	69
Tabela 4.4 – Classificação por localidade - São Paulo . . . . .	70
Tabela 4.5 – Classificação por localidade - Rio de Janeiro . . . . .	72
Tabela 4.6 – Classificação por localidade - Belo Horizonte . . . . .	73
Tabela 4.7 – Tecnologias por operadora (São Paulo) . . . . .	88
Tabela 4.8 – Tecnologias por operadora (Rio de Janeiro) . . . . .	93
Tabela 4.9 – Tecnologias por operadora (Belo Horizonte) . . . . .	98

## LISTA DE QUADROS

Quadro 2.1 – Disponibilidade de dados para pesquisa acadêmica - Twitter . . . . .	45
Quadro 3.1 – Disponibilidade de dados para pesquisa acadêmica - Twitter . . . . .	49
Quadro 3.2 – Parâmetros de busca de <i>Tweets</i> . . . . .	50
Quadro 3.3 – Disponibilidade de dados para pesquisa acadêmica - Twitter . . . . .	51
Quadro 3.4 – Classificação das Expressões . . . . .	51

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	16
1.1	Contextualização e Motivação	16
1.2	Justificativa	18
1.3	Objetivo	19
1.4	Objetivos Específicos	20
1.5	Organização do Trabalho	20
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	21
2.1	Redes Móveis	21
2.1.1	A Evolução das Redes Móveis	22
2.1.2	Redes Móveis no Brasil	24
2.2	QoE - Definição	27
2.2.1	Pontuação Subjetiva MOS	28
2.3	Redes Sociais	29
2.4	Mineração de Dados	29
2.5	Redes Neurais e Aprendizagem de Máquina	31
2.5.1	PLN - Processamento de Linguagem Natural	33
2.5.2	Análise de Sentimentos	34
2.5.3	<i>Transformer</i>	35
2.5.4	BERT	38
2.6	Trabalhos Relacionados	40
<b>3</b>	<b>METODOLOGIA</b>	46
3.1	Classificador e coleta de dados	47
3.1.1	Coleta de Opiniões	48
3.1.2	Definições de Classes	50
3.1.3	Pré-processamento	51
3.1.4	Divisão do conjunto de dados para treino e teste	53
3.1.5	Modelo de Classificação	54
3.1.6	Métricas de avaliação do modelo	56
3.1.7	Predição e Geração de QoE	57
3.2	Dados da Anatel	57
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	61
4.1	Avaliação do modelo de classificação	61

<b>4.2</b>	<b>Classificação automática</b>	65
<b>4.2.1</b>	<b>Classificação geral</b>	65
<b>4.2.2</b>	<b>Classificação por operadora</b>	66
<b>4.2.3</b>	<b>Classificação por localidade</b>	68
<b>4.3</b>	<b>Dados Anatel</b>	74
<b>4.4</b>	<b>Dados Anatel - Geral e por operadora</b>	74
<b>4.4.1</b>	<b>Dados da Anatel por localidade</b>	81
<b>4.5</b>	<b>Comparações entre os dados da predição automática com a Anatel</b>	100
<b>5</b>	<b>CONCLUSÃO</b>	102
	<b>REFERÊNCIAS</b>	104
	<b>APENDICE A – Mapas Panorâmicos de cobertura por operadora</b>	109
	<b>APENDICE B – Pseudocondigos</b>	112

## 1 INTRODUÇÃO

Este trabalho busca novos indicadores de QoE (*Quality of Experience*) ou, em português, qualidade de experiência, de usuários de redes móveis, com base em opiniões publicadas em redes sociais e comparação de resultados com dados públicos, disponibilizados pela Anatel sobre a satisfação dos consumidores e tecnologias empregadas pelas principais operadoras que atuam em território brasileiro. Nesse contexto, neste capítulo, apresenta-se a contextualização e motivação da pesquisa aqui, bem como a justificativa, objetivo geral, objetivos específicos e a organização do trabalho.

### 1.1 Contextualização e Motivação

As redes móveis foram inicialmente projetadas para a comunicação de voz, mas com o passar do tempo, evoluíram para oferecer uma variedade de serviços, incluindo mensagens instantâneas, conexão com a internet, operações bancárias, comunicação por meio de redes sociais, entre outros. Essa evolução pode ser observada em conformidade com MARTINS JUNIOR (2020), que relata a progressão das redes móveis desde a primeira geração, que oferecia apenas serviços de voz, até a quinta geração, que busca englobar mais serviços para atender à IoT (Internet das Coisas) e oferecer ainda mais conectividade aos consumidores. Essa evolução exige que as operadoras busquem constantemente novas tecnologias para atender às demandas do mercado de telecomunicações e se manterem competitivas, oferecendo serviços de qualidade aos seus clientes.

No Brasil, as redes móveis são monitoradas pela agência governamental "Anatel" (Agência Nacional de Telecomunicações) e os serviços prestados para os consumidores são disponibilizados por empresas privadas, sendo as principais: Oi, Tim, Vivo e Claro, que ofertam serviços em 2G, 3G, 4G e, mais recentemente, o 5G. Com base nos serviços oferecidos pela Anatel, os usuários de telefonia móvel no Brasil têm acesso a informações precisas sobre pesquisas de satisfação dos serviços prestados à sociedade brasileira, bem como reclamações e tecnologias empregadas pelas operadoras por municípios. Essas informações podem, também, contribuir para as empresas no relacionamento com seus clientes, no entendimento das necessidades dos consumidores, bem como auxiliá-las na evolução dos serviços prestados (BRASIL, 2023).

A internet e, especialmente, as redes sociais, também têm aproximado prestadores de serviços e consumidores. Com o crescimento exponencial das redes sociais e a quantidade cada vez maior de usuários interagindo entre si, empresas e organizações têm utilizado ferramentas de coleta de dados para monitorar a reputação da marca e as opiniões dos consumidores online. Um estudo da Maia (2022), indica que as redes sociais tornaram-se uma ferramenta fundamental para as empresas alcançarem os

consumidores, assim, estar nas redes sociais pode ser considerado uma estratégia para melhorar o relacionamento empresa/cliente. As redes sociais, também, apresentam-se como um grande campo de pesquisa, em decorrência da geração de informações espontâneas compartilhadas entre seus usuários, como afirmam Costa (2018). Portanto, é possível observar que diferentes pesquisadores estão se dedicando a explorar esse campo em expansão, utilizados técnicas de Inteligência Artificial (IA), Redes Neurais (RN), aprendizagem de máquina, ou do inglês *Machine Learning* (ML), e à análise de sentimentos (SOUZA et al., 2017).

A análise de sentimentos é uma área crucial da ciência de dados, que tem como objetivo extrair informações valiosas sobre as emoções humanas expressas em textos, como avaliações de produtos, críticas de filmes e posts em mídias sociais. Para realizar essa tarefa desafiadora, as redes neurais de processamento de linguagem natural (PLN) emergiram como uma ferramenta poderosa e versátil, capaz de entender a complexidade e a sutileza do contexto linguístico. O modelo de rede neural *transformer*, proposto por Vaswani et al. (2017), apresentou-se como um grande avanço na área de PLN, atraindo o interesse de pesquisadores e tornando-se uma arquitetura de referência e base para projetos como o BERT (*Bidirectional Encoder Representations from Transformers*) de Devlin et al. (2018). Esse modelo usa técnicas de transformadores e aprendizado de máquina bidirecional para a análise de sentimentos, alcançando novos patamares de precisão e eficiência e tornando-se uma técnica amplamente utilizada em diferentes setores, como marketing, finanças e pesquisa acadêmica (KHAN et al., 2021; LIN et al., 2021). O BERT é um modelo pré-treinado com dados não rotulados, que utiliza modelagem de aprendizado profundo. Embora o modelo de Devlin et al. (2018) tenha sido pré-treinado em inglês, outros pesquisadores têm pré-treinado o modelo, adaptando-o para outros idiomas, como o modelo BERTimbau treinado em português brasileiro e o BERTmultilíngue que reconhece mais de 100 idiomas.

Esta pesquisa buscou explorar a relação cliente/prestador de serviços das principais operadoras de redes móveis que atuam no território brasileiro. Para isso, utilizou a análise de QoE, coletando opiniões publicadas na rede social Twitter. As opiniões foram classificadas através de uma Rede Neural (RN) com BERT e os resultados foram comparados com as informações disponibilizadas pela Anatel. O termo QoE, ou qualidade de experiência, surgiu como uma evolução dos conceitos de qualidade de serviço (QoS) e usabilidade, enfatizando a importância da percepção subjetiva do usuário no processo de avaliação e melhoria de produtos e serviços. Assim, a QoE tem como objetivo contribuir com soluções que ajudem na análise de redes móveis, como proposto neste trabalho, buscando uma melhora na prestação de serviços com base na perspectiva dos usuários finais (GOUVEIA; ROSA; RODRÍGUEZ, 2018).

A metodologia empregada nesta pesquisa utiliza os conceitos de análise de sentimentos com o método supervisionado. Essa técnica de aprendizado de máquina tem como base o treinamento da rede neural com um conjunto de dados rotulados, conhecidos como conjunto de treinamento, que incluem tanto as entradas quanto as saídas desejadas correspondentes. Assim, inicialmente foi criado um *DataSet* com *tweets* rotulados manualmente e dividido em três classes: péssimo, regular e excelente. Após esta fase, a classificação passou a ser automatizada e a classificação final convertida para uma pontuação (1 - péssimo, 3 - regular e 5 - excelente), com a finalidade de gerar uma pontuação média por operadora. Isso permite obter um indicador de QoE e compará-lo com os dados disponíveis pela Anatel.

A classificação das opiniões nas redes sociais sobre os serviços de redes móveis apresenta uma abordagem diferenciada em relação às pesquisas que usam métodos tradicionais, especialmente aquelas que analisam a QoE. A maioria dos estudos abordando QoE realiza testes e simulações de serviços que trafegam na rede, proporcionando uma perspectiva interna. No entanto, a abordagem adotada nesta pesquisa busca uma visão externa da rede, com foco no ponto de vista do consumidor. A principal diferença é que essa abordagem se concentra na análise da subjetividade das opiniões expressas pelos usuários nas redes sociais. Em vez de simular e testar serviços específicos, buscamos entender a percepção e a satisfação de consumidores reais, considerando seu comportamento espontâneo e opiniões subjetivamente expressas.

É importante destacar a escassez de métodos automáticos e abordagens na literatura que analisam essa subjetividade do ponto de vista do consumidor externo. Este é um problema que nossa pesquisa visa atender, com novos indicadores vindos das redes sociais.

Nos resultados, foi possível observar que há uma tendência dos usuários de redes móveis brasileiras utilizarem as redes sociais para expressar reclamações e indicar insatisfação com os serviços. Essa tendência foi observada de forma geral, tanto por operadora quanto por localidade. Além disso, notou-se que há uma paridade em relação à satisfação dos usuários em relação às operadoras, com destaque para a operadora Tim, que obteve o índice geral mais elevado.

## **1.2 Justificativa**

O uso das redes móveis vem ganhando cada vez mais adeptos, movimentando o mercado para atender a clientes que se acostumaram com as facilidades oferecidas por esse tipo de rede. Com o avanço das redes móveis, surgem novas aplicações, e segundo Faghihi e Behdadfar (2020), a demanda por elas se deve a fatores vitais, como a disponibilidade de internet acessível ou gratuita, telefonia nacional ou internacional, e o rápido aumento do uso de dispositivos portáteis e inteligentes. Isso tem levado a um subsequente crescimento do mercado.

De acordo com Guimarães (1998), a variedade de serviços oferecidos pelos sistemas de comunicação móvel é um reflexo da necessidade de acesso, tornando-se um fator determinante do sucesso das organizações prestadoras de serviços móveis. As organizações precisam antecipar-se às necessidades dos clientes, estar próximas deles e deslocar-se para fora de seus escritórios para ver de perto a satisfação dos seus clientes com os serviços prestados.

Nesse contexto, a qualidade de experiência (QoE) tem sido objeto de estudos com o objetivo de contribuir com soluções que ajudem na análise de redes e busquem melhorias na prestação de serviços. MARTINS JUNIOR (2020) aponta que as operadoras de telecomunicação buscam entender a qualidade da experiência vivenciada pelos seus clientes, o que pode contribuir para que elas demandem infraestruturas que atendam melhor seus consumidores. Segundo Ciambone et al. (2018), uma infraestrutura de rede bem elaborada contribui para a prestação de serviços com qualidade e aumenta a satisfação do consumidor em relação à QoE.

Nessa perspectiva, o objetivo desta pesquisa é propor novos indicadores de QoE para as operadoras de redes móveis que atuam no Brasil. Para alcançar esse objetivo, o campo exploratório escolhido é o das redes sociais, por causa da interatividade que ocorre nesse ambiente. De acordo com Costa (2018), esse ambiente possibilita análises com diferentes perspectivas, pois é possível realizar agrupamentos de informações com base no comportamento dos usuários.

Para coletar e classificar as opiniões, esta pesquisa utilizará classificadores automatizados, como o BERT na versão BERTimbau e BERTmutiligue, que têm como base o *transformer*, uma tecnologia apontada como revolucionária na área de Processamento de Linguagem Natural (PLN) por Khan et al. (2021) e Lin et al. (2021). Os resultados obtidos serão comparados com as informações de reclamações sobre serviços prestados e tecnologias utilizadas pelas operadoras, disponíveis para consulta na base da Anatel.

Diante desse contexto, acredita-se que a manutenção de novos indicadores de QoE para redes móveis, se justifica, pois pode auxiliar tanto os prestadores de serviços, a melhorar suas atividades, quanto os consumidores na escolha de sua operadora.

### 1.3 Objetivo

O objetivo deste trabalho é propor novos indicadores de QoE para redes móveis das principais operadoras de telecomunicação que atuam no Brasil. Para isso, será realizada uma análise das opiniões dos consumidores manifestadas espontaneamente em redes sociais. Essas opiniões serão classificadas automaticamente utilizando uma Rede Neural BERT. Em seguida, os resultados serão comparados com os dados de reclamações sobre serviços prestados e tecnologias utilizadas pelas operadoras, disponíveis

na base da Anatel, e assim, obter *insights* sobre a percepção dos consumidores em relação à qualidade dos serviços oferecidos pelas operadoras e sua relação com as reclamações registradas junto à Anatel.

#### 1.4 Objetivos Específicos

Os objetivos específicos deste trabalho, são:

- a) coletar opiniões manifestadas por meio de redes sociais sobre as principais operadoras que atuam no Brasil e criar um conjunto de dados (*dataset*) para treinamento e teste do modelo de classificação supervisionado;
- b) treinar e testar uma Rede Neural (RN) utilizando as variantes BERTimbau e BERTmultiligue, e comparar o desempenho de ambos os modelos;
- c) classificar automaticamente as opiniões sobre as principais operadoras de forma geral e de forma específica, ou seja, por operadoras e por localidade;
- d) comparar os resultados obtidos pela classificação automática com as informações de reclamações, pesquisa de satisfação e *rank* de acessos por operadoras disponibilizadas pela Anatel, a fim de verificar possíveis correlações;
- e) analisar a correlação entre os resultados da classificação automática e as tecnologias utilizadas pelas operadoras, identificando se existe alguma relação entre a percepção dos consumidores e as tecnologias empregadas pelas empresas.

#### 1.5 Organização do Trabalho

Este trabalho divide-se em cinco capítulos. O primeiro capítulo é a Introdução, que já foi apresentada acima. No segundo capítulo, intitulado Referencial Teórico, abordam-se os temas que fundamentam esta dissertação, bem como os Trabalhos Relacionados, onde são citadas pesquisas publicadas por diferentes pesquisadores e instituições com temas correlacionados e abordados nesta pesquisa.

No terceiro capítulo, Metodologia, é apresentada a metodologia utilizada e as ferramentas empregadas para as pretensões desta pesquisa. No quarto capítulo, Resultados e Discussões, são apresentadas discussões sobre o processo para obtenção dos resultados, bem como os próprios resultados obtidos. Por fim, o quinto capítulo, Conclusão, apresenta as conclusões desta pesquisa e sugestões de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo, são abordados temas correlacionado com esta pesquisa. O objetivo é apresentar temas com base literária de referência e trabalhos publicados sobre a proposta. Assim, os assuntos tratados são: Redes Móveis, A Evolução das Redes Móveis, Redes Móveis no Brasil, QoE - Definição, Pontuação Subjetiva MOS, Redes Sociais, Mineração de Dados, Redes Neurais e Aprendizagem de Máquina, PLN - Processamento de Linguagem Natural, Análises de Sentimentos, *Transformer*, BERT e trabalhos relacionados à QoE e Redes Sociais.

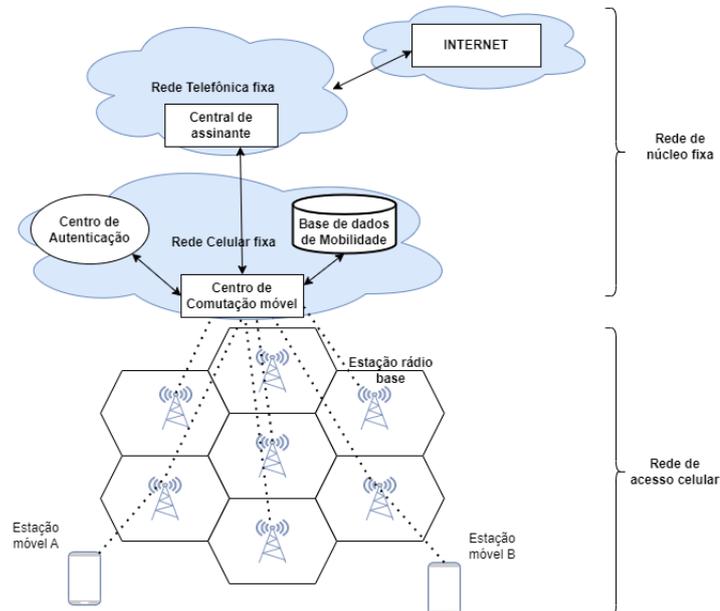
### 2.1 Redes Móveis

As redes móveis, de telecomunicação celular, têm assumido um papel cada vez mais crucial no mundo atual, permitindo que as pessoas se conectem e se comuniquem em qualquer lugar e a qualquer momento. Embora inicialmente projetadas para comunicação de voz em dispositivos móveis, a evolução dessas redes, especialmente em sua capacidade de transmissão e recepção de dados, possibilitou a interação de diferentes tipos de tecnologia e aplicativos por meio do seu uso. Com o tempo, esse tipo de redes móveis têm permitido uma ampla gama de atividades para seus usuários, desde transmissão de voz e vídeo até armazenamento de dados em nuvem, conectividade via internet, interação com outros usuários por meio de diferentes aplicativos, como mensagens instantâneas e redes sociais, e até mesmo serviços de transporte e transações bancárias. Como resultado, as operadoras de redes móveis estão buscando formas de compreender as expectativas de seus usuários e desenvolver soluções que possam torná-las mais competitivas (OLIVEIRA; ALENCAR; LOPES, 2018; MARTINS JUNIOR, 2020).

As redes móveis celulares são baseadas na comunicação sem fio, ou seja, não é necessária a utilização de fios, para estabelecer a comunicação entre dois pontos, possuindo assim transmissão, através de canais de rádio frequência. O termo "móvel" se deve principalmente pela possibilidade de comunicação constate sem interrupções e em movimento através da passagem da comunicação de uma "célula" para outra (PIROTTI; ZUCCOLOTTO, 2009).

O funcionamento das Redes Móveis também é abordado por ROCHOL (2018) onde diz que as redes de sistemas celulares possuem comunicação, por meio de redes de comunicação por rádio frequência, permitindo assim mobilidade contínua, através de muitas células. Na Figura 2.1, apresenta-se um cenário que ilustra a infraestrutura de redes de telefonia móvel.

Figura 2.1 – Redes Celulares



Fonte: ROCHOL (2018)

Na Figura 2.1, apresenta-se um modelo de comunicação, as estações móvel (EM) estão representando dispositivos que vão comunicar-se, as estações rádio base (ERB) são representadas pelas torres, o centro de comutação móvel é responsável, por fazer interligação de uma chamada telefônica entre o pontos de células diferentes ou interligações com redes de telefonia física ou mesmo interligações com a internet, a base de dados de mobilidade é responsável por verificar o sinal do dispositivo e sua localização, a interface aérea são os enlace de rádio entre EM e a ERB e o centro de autenticação é responsável por autenticar o usuário na sua operadora e células que referenciam a amplitude das ERB (ROCHOL, 2018).

De acordo com Ciambone et al. (2018) uma infraestrutura de redes bem elaborada contribui para a prestação de serviços com qualidade, aumentado assim seu QoE. Assim, nesta pesquisa, acredita-se que, manter indicadores de QoE atualizados, colaboram com as operadoras, na elaboração de projetos, sejam estes de melhoria ou de novas soluções para seus serviços prestados.

### 2.1.1 A Evolução das Redes Móveis

Ao longo do tempo, a comunicação móvel vem agregando serviços para atender a um público cada vez mais acostumado com os benefícios oferecidos pelos sistemas de redes celulares. A evolução da tecnologia para redes móveis é marcada pela letra "G" indicando a geração.

De acordo com Santos (2008), a chamada primeira geração (1G) trouxe, consigo uma mudança de panorama, uma vez que a comunicação telefônica era física e sem mobilidade, a 1G teve sucesso, ao permitir realizar chamadas telefônicas longe de casa, em um aparelho sem fio e em movimento.

SENA (2012) aponta para o fato de os aparelhos celulares desta geração serem de custos elevados, de tamanhos considerados grandes e utilizavam frequências analógicas para as chamadas telefônicas. Segundo ROCHOL (2018), a 1G utilizava as tecnologias AMPS (*Advanced Mobile Phone System*) e a FDMA (*Frequency Division Multiple Access*) para a conversão em frequência analógica. No FDMA, a largura de banda é dividida em *slots* de frequência que são alocados para cada usuário.

Já a segunda geração (2G), Santos (2008) diz que surgiu da necessidade de oferecer o que a 1G deixava a desejar. Observa-se aqui a necessidade de sempre evoluir e oferecer melhores serviços aos consumidores. Ainda em conformidade com o autor já citado, os EUA estavam à procura de maior capacidade, enquanto que, na Europa, havia um esforço para uniformizar os sistemas para o MCE (Mercado Comum Europeu), dando surgimento ao GSM (*Global System for Mobile Communications*) que se tornou a principal tecnologia da 2G. De acordo com SENNA (2012), os celulares da segunda geração passaram a ter tamanhos menores, a ser mais acessíveis quanto ao custo e passaram a permitir o envio de SMS. ROCHOL (2018) aponta para o GSM, TDMA (*Time Division Multiple Access*) e CDMA (*Code Division Multiple Access*) como tecnologias utilizadas no 2G. O TDMA tem como característica ser um tipo de tecnologia com a capacidade do canal digital de RF (Rádio Frequência) dividida em fatias de tempo (*time-slots*) que formam um quadro básico que se repete ciclicamente ao longo do tempo. No CDMA, a sequência de espalhamento é definida, por meio de um código próprio de cada usuário, assim, todos os usuários podem compartilhar a mesma frequência, sendo que cada usuário tem a sua própria sequência de “*chipping*” (ou seja, código) para codificar dados. Entre a 2G e a 3G teve-se a 2.5G conhecida como segunda geração e meia. A tecnologia 2,5G foi uma evolução da tecnologia 2G e permitiu uma melhoria significativa na capacidade de transmissão de dados em dispositivos móveis. Essa tecnologia foi desenvolvida para atender às crescentes demandas por serviços de dados móveis e ofereceu taxas de transferência mais rápidas em comparação com a tecnologia 2G, sendo marcada como processo de transição entre a 2G e a 3G (SANTOS, 2008).

A terceira geração, o (3G), segundo Santos (2008) é marcada pelo advento da internet, que passa a integrar os serviços de conectividade aos dispositivos móveis, junto com os já tradicionais serviços de Voz e o SMS do 2G. SENNA (2012) aponta para fato de que, no 3G, os celulares passaram a ser conhecidos com *smartphones*, pois os aparelhos passaram a vir com sistemas operacionais. ROCHOL (2018) diz que no 3G utiliza-se o UMTS (*Universal Mobile Telecommunications Service*) que representa uma evolução do padrão GSM, com a sua implementação, podendo, inclusive aproveitar a estrutura do GSM. O UMTS utiliza a tecnologia do CDMA já mencionada acima no 2G, passando para a versão CDMA-2000.

A quarta geração, o (4G), é marcada pelas redes baseadas em IP. Essa tecnologia permite que os dispositivos móveis se conectem a uma rede de alta velocidade, com capacidade de transmissão podendo

passar de 1Gbps, o que significa uma conexão mais rápida. O 4G permite que os usuários realizem chamadas de voz tanto da rede de telefonia tradicional, quanto de serviços de comunicação sem fio atuais, como o VoIP (*Voice over Internet Protocol*), e outros serviços, como, o envio de mensagens de texto por meio de uma conexão de dados, acesse a *streaming* e jogos de vídeo de alta definição, entre outros (MARTINS JUNIOR, 2020). ROCHOL (2018) diz que as tecnologias que marcaram o 4G são as redes LTE (*Long Tem Evolution*).

A quinta geração, o (5G), foi apontada por MARTINS JUNIOR (2020) como embrionária e segundo Brasil (2023) recentemente passou por processo de leilão espectral no Brasil. A tecnologia 5G oferece taxas de transferência de dados ainda mais rápidas em relação à tecnologia 4G, além de uma latência significativamente menor. O 5G vem com várias expectativas em questão dos serviços propostos, principalmente na integração com o IoT (internet das coisas). MARTINS JUNIOR (2020) aborda algumas das evoluções esperadas, como uma ampliação na transmissão de dados, interações multimídia como voz e vídeos, suporte a serviços de carros autônomos, serviços para cidades inteligentes entre outros.

Assim pode se observar com a evolução a constata busca por melhora dos serviços prestados.

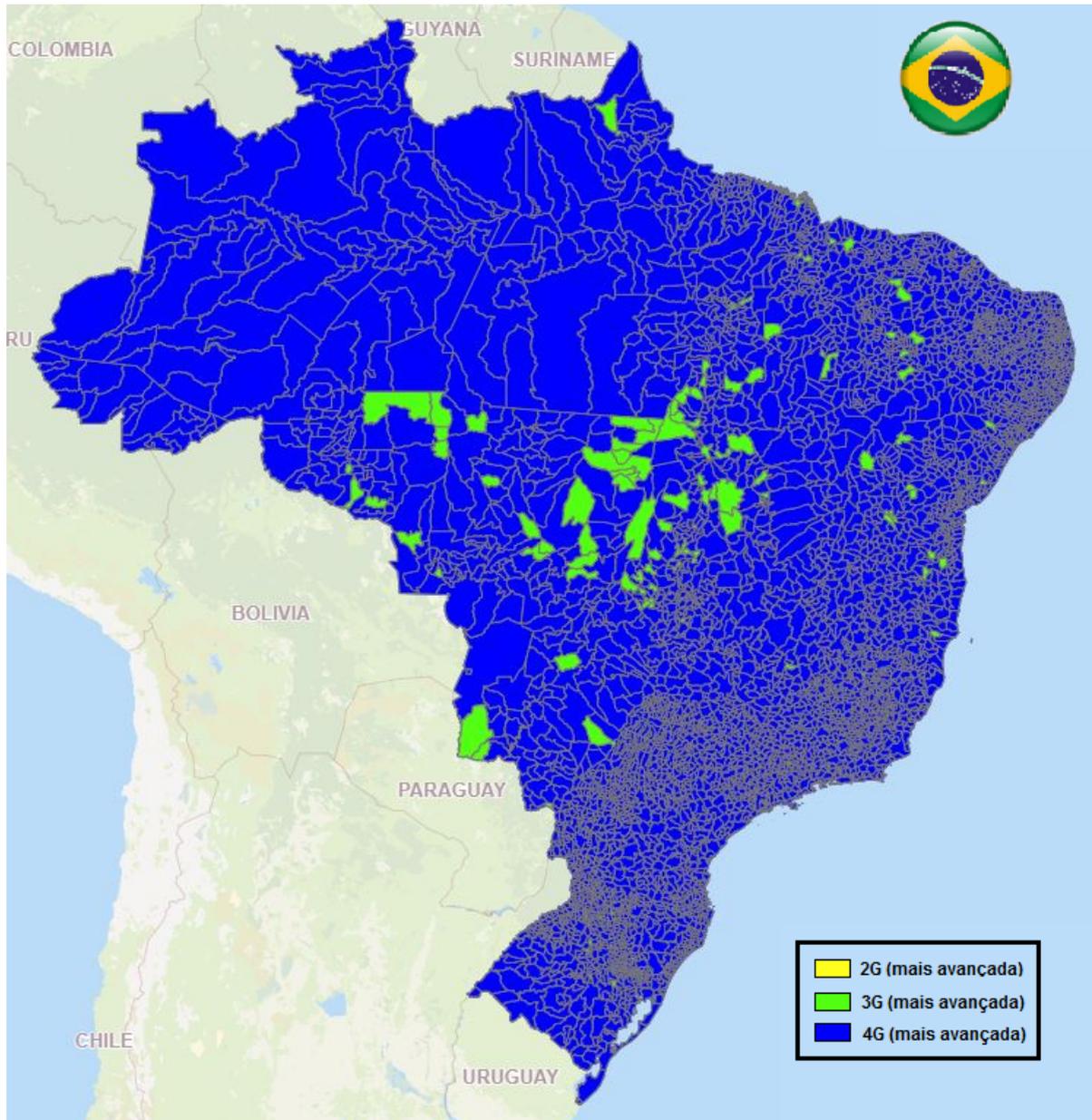
### **2.1.2 Redes Móveis no Brasil**

O Brasil possui redes de telefonia Móvel (Serviço Móvel Pessoal - SMP) com tecnologias 2G, 3G e 4G, além de ter leiloado, recentemente, o 5G. O órgão regulamentador do espectro de frequência ao qual as redes de telefonia móvel operam é a Anatel (Agência Nacional de Telecomunicações), que, por sua vez, é vinculada ao Ministério das Comunicações, e foi criada pela Lei Geral de Telecomunicações Brasil Lei nº 9.472 (1997) sendo a primeira agência reguladora a ser instalada no Brasil. A Anatel é responsável pelo desenvolvimento das telecomunicações, realização espectral dos leilões, para as operações de telecomunicação, pela monitoria da eficiência em relação à infraestrutura de telecomunicações e de averiguar a capacidade dos fornecedores de oferecer à sociedade serviços adequados, diversificados e a preços justos, tendo um papel fundamental na supervisão das empresas de telecomunicações que prestam serviços móveis em território brasileiro, buscando, garantir que tais empresas, cumpram as normas e regulamentações em vigor, além de proteger os interesses dos consumidores. Assim, todas as operadoras que desejam oferecer seus serviços de telecomunicação, no Brasil, necessitam da concessão emitida pela Anatel (BRASIL, 2023).

Apesar de as operadoras necessitarem de concessão, os serviços de redes móveis são prestados em regime privado e, de acordo com Brasil (2023), é baseada nos princípios constitucionais da atividade econômica, conforme o art. 126 da Lei Geral de Telecomunicações. Assim sendo, a oferta de serviços,

por regiões do Brasil, depende do interesse comercial das prestadoras. Na Figura 2.2, apresenta-se um mapa panorâmico das principais tecnologias empregadas no Brasil.

Figura 2.2 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil



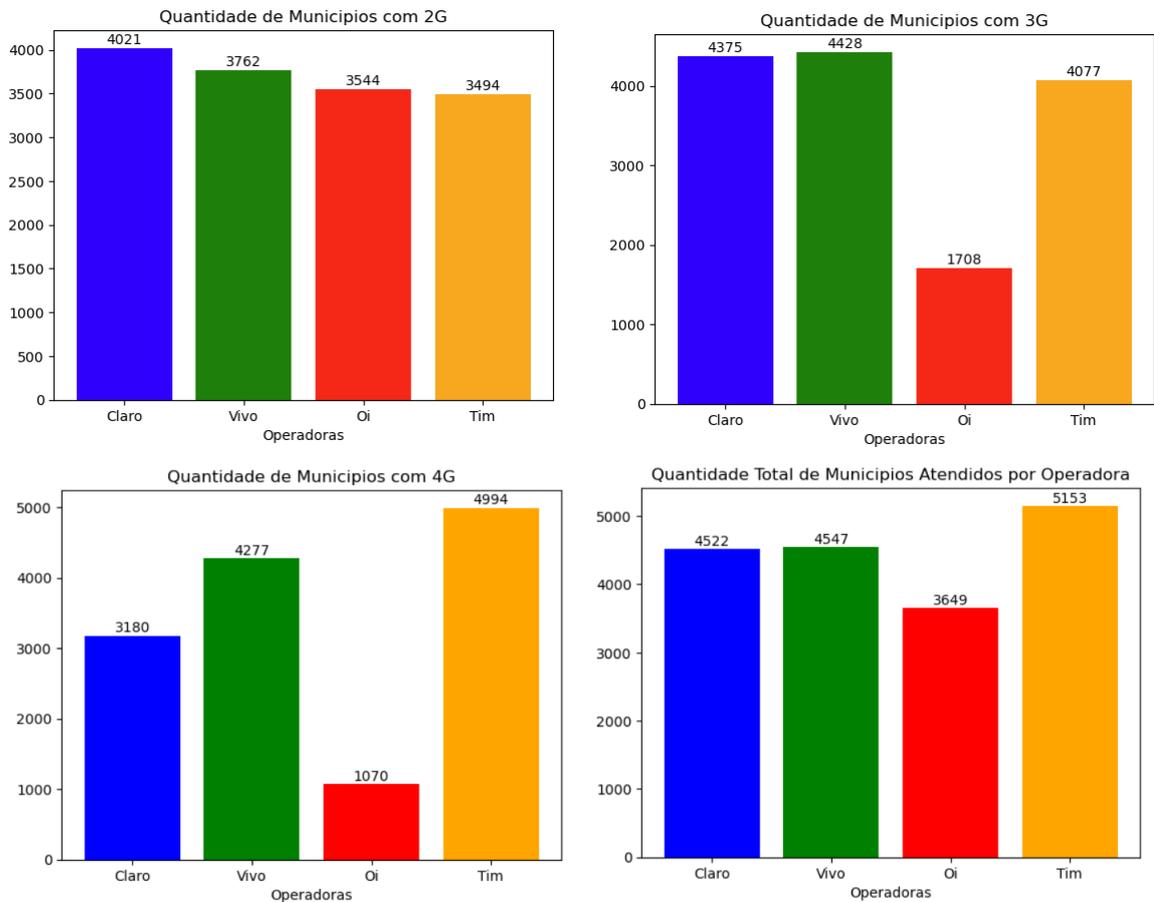
Fonte: Brasil (2023)

Brasil (2023) ressalta que a Figura 2.2 não retrata a cobertura do serviço móvel de forma real, ou seja, onde o usuário vai encontrar o serviço disponível, apenas indica a presença do serviço. No mapa é possível observar que, o 4G está disponível em grande parte do território brasileiro, mas a qualidade do serviço pode variar dependendo da região. No apêndice A, é apresentado os mapas panorâmicos de

cobertura por operadora, possibilitando uma visão mais clara das tecnologia utilizada pelas operadoras por região.

As principais operadoras que atuam no Brasil, são a Tim, Vivo, Claro e Oi, com atuações em diferentes áreas regionais. Nos gráficos a seguir, realça-se a representatividade das tecnologias utilizadas por cada operadora no Brasil e a presença de cada uma delas nos municípios.

Gráfico 2.1 – Presença das Operadoras por Municípios no Brasil



Fonte: Adaptado de Brasil (2023)

É importante ressaltar que a operadora Oi foi incorporada recentemente pelas operadoras Tim, Vivo e Claro, com divisões da área em que presta serviços. Na pesquisa aqui proposta, buscam-se novos indicadores de QoE dos serviços de redes móveis prestados pelas operadoras que atuam no mercado brasileiro, devido a divisão, os resultados apresentados no Capítulo 4, aborda todas as operadoras, mas quando comparados os resultados por tecnologia, atenta-se apenas as operadoras Tim, Vivo e Claro.

## 2.2 QoE - Definição

A Qualidade de Experiência (QoE) é uma medida que busca avaliar o quanto um usuário está satisfeito com um serviço ou aplicativo com base em sua experiência pessoal, podendo ser utilizada para avaliar a qualidade de serviços como vídeo, voz, jogos, *streaming* de música e outros. A QoE surgiu como uma evolução da já conhecida, qualidade de serviço (QoS). A diferença entre QoS e QoE, é que QoS foca nos aspectos técnicos de serviços de telecomunicações, como por exemplo, perda de pacote, latência, *jitter*, entre outros, já a QoE, foca na perspectiva do usuário final, ou seja, a qualidade percebida pelo usuário é o que realmente importa. Além disso, a QoE é bastante subjetiva inerentemente e mais realista do que a QoS. Os parâmetros de QoS, tais como produção, atraso de ponta a ponta, tremor e taxa de perda de pacotes não avaliam, com sucesso, os serviços como percebidos pelos usuários, mas há uma correlação entre as questões de QoS e QoE (CIAMBRONE et al., 2018; LASELVA et al., 2018; FAGHIHI; BEHDADFAR, 2020).

A QoE pode ser medida por meio de métodos quantitativos e qualitativos, podendo ser usada para melhorar continuamente os serviços e desenvolver produtos e serviços melhores e mais adaptados às necessidades do usuário. Os métodos quantitativos são baseados em parâmetros objetivos e mensuráveis, como a taxa de bits, a latência, a largura de banda e outros. Esses parâmetros são medidos por equipamentos de medição e a execução de algoritmos específicos e são usados para avaliar a qualidade técnica do serviço ou aplicativo. Por exemplo, a taxa de bits é uma medida da quantidade de dados transmitidos por segundo, na avaliação da qualidade de vídeos ou áudios. Já os métodos qualitativos são baseados na experiência subjetiva do usuário. Eles são usados para avaliar a satisfação do usuário com o serviço ou aplicativo e leva em consideração fatores como a facilidade de uso, a clareza das informações fornecidas e a estética do produto. Esses métodos geralmente envolvem pesquisas de opinião, entrevistas e questionários para avaliar a experiência do usuário. Este método tende a ser mais limitado principalmente pelo custo e tempo necessário para a realização de testes (SU et al., 2016; JELASSI et al., 2012; GOUVEIA; ROSA; RODRÍGUEZ, 2018).

Os métodos quantitativos e qualitativos são complementares e podem ser usados em conjunto para obter uma visão mais completa da QoE. Por exemplo, os parâmetros técnicos podem ser usados para avaliar a qualidade de vídeo ou áudio, enquanto as pesquisas de opinião podem ser usadas para avaliar a satisfação geral do usuário com o serviço ou aplicativo. A combinação desses métodos permite que os provedores de serviços obtenham uma visão mais precisa e completa da qualidade percebida pelo usuário.

Nesta pesquisa, a classificação de opiniões para se chegar a novos indicadores de QoE, é baseada no método subjetivo, assim acreditamos que a coleta automática de opiniões espontâneas manifestadas em redes sociais seja um diferencial quanto a esse tipo de método.

### 2.2.1 Pontuação Subjetiva MOS

O MOS (*Mean Opinion Score*) é uma técnica padronizada pelo grupo de estudos em redes de telecomunicações ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*), de avaliação subjetiva, e que tem como objetivo medir a qualidade perceptiva de serviços de telecomunicações, como por exemplo, serviços de transmissão de voz e vídeo. A escala MOS é uma escala numérica de 1 a 5, em que 1 representa qualidade muito ruim e 5 representa qualidade excelente. Os valores do escalar MOS buscam representação da qualidade percebida pelo usuário final podendo avaliar em diferentes aspectos, tais como, por exemplo, a qualidade do áudio, qualidade do vídeo, clareza da imagem, nitidez da cor, entre outros. A escala MOS é amplamente utilizada em pesquisas de satisfação do cliente e em testes de qualidade de serviços de telecomunicações. É importante destacar que a qualidade percebida pode variar de acordo com diversos fatores, tais como as expectativas do usuário, a qualidade da rede, a disponibilidade de recursos, entre outros (XU et al., 2011; HU et al., 2020).

O MOS, tem sido utilizado na indústria de telecomunicações e é acompanhado de recomendações técnicas do ITU-T, como a recomendação ITU-T P.800, que busca representar com o escalar de 5 níveis a satisfação dos usuários finais conforme apresentado na Tabela 2.1.

Tabela 2.1 – Escala de pontuação MOS

Nível (MOS)	COMPREENSÃO
5	Excelente
4	Boa
3	Regular
2	Ruim
1	Péssimo

Fonte: Adaptado de INTERNATIONAL TELECOMMUNICATION UNION ITU-T P.800 (1996)

Como já mencionado nas definições de QoE, esta pesquisa utilizará o processo subjetivo para chegar a novos indicadores de QoE para as operadoras de redes móveis que atuam no mercado brasileiro. Assim, a partir da coleta automática de opiniões expressadas, espontaneamente, nas redes sociais, é feita uma analogia entre a opinião coletada e o escalar MOS para pontuá-las.

### 2.3 Redes Sociais

As redes sociais se tornaram uma das principais fontes de dados do mundo contemporâneo. Com a popularização do acesso à internet e a facilidade de interação proporcionada pelas redes sociais, milhões de pessoas utilizam diariamente plataformas como *WhatsApp*, *Telegram*, *YouTube*, *Instagram*, *Facebook*, *Twitter*, *Linkedin*, *TikTok*, entre outras, para compartilhar informações e se comunicar com outras pessoas. Essas informações podem ser de grande valor para empresas, pesquisadores e governos, permitindo a identificação de tendências, opiniões e comportamentos que antes eram difíceis de obter. Apesar de as redes sociais terem o mesmo princípio, ou seja, a conexão de pessoas, elas atuam de formas distintas e com características diferentes. Redes sociais como *WhatsApp* e *Telegram*, por exemplo, utilizam métodos de troca de mensagens instantâneas de usuário para usuário. Já redes sociais como *YouTube* e *TikTok* têm como método de comunicação a publicação de vídeos que podem ser vistos por milhares de pessoas. As redes sociais como *Instagram*, *Facebook* e *Twitter* têm como característica a interação entre perfis, seja pelo laço de amizade, familiar, grupos de interesses em comum, entre outros. (MARQUES, 2018).

As redes sociais podem ser usadas para diferentes funções, como divertimento, procura de amigos com o mesmo perfil, procura de empregos, publicações de opiniões, aprendizagem, entretenimento entre outros. Em razão dessas características, as redes sociais têm se tornado um campo para pesquisadores, pois possui uma interação contínua dos usuários com geração constante de informações que permitem análises em diferentes perspectivas e agrupamento de informações com bases no comportamento de seus usuários (COSTA, 2018; ARAÚJO; CRUZ, 2017).

Por suas características, a rede social utilizada na busca de opiniões, nesta pesquisa, será o *Twitter*. De acordo com Al-Makhadmeh e Tolba (2020), essa rede social é utilizada pelas pessoas com o intuito de seguir outras pessoas, participar de atividades sociais e transmitir suas opiniões, por meio de comentários.

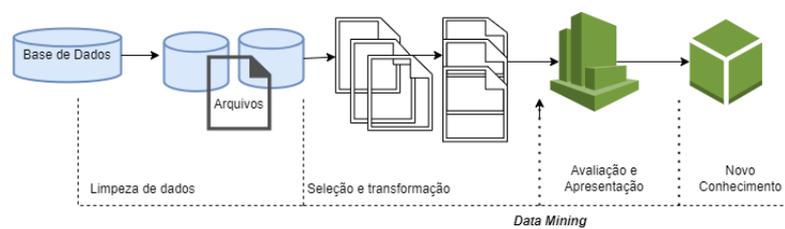
### 2.4 Mineração de Dados

A Mineração de dados é amplamente estudada no meio computacional e, de acordo com Amo (2004), teve o seu início, na década de 80, pela necessidade vista pelos profissionais das empresas e organizações de analisar dados e informações estocadas e inutilizadas nas grandes bases de dados da época. A mineração de dados também é conhecida como *Data Mining* e é utilizada para processos analíticos projetados para explorar grandes quantidades de dados que podem estar relacionados a diferentes áreas como negócios, mercado ou pesquisas científicas, na busca de padrões (ou perfis) consistentes e com relacionamentos sistemáticos.

Araújo e Cruz (2017) apontam para o fato de a mineração de dados ser utilizada nas áreas científicas com aplicações em banco de dados, estatística e IA e que a sua finalidade está em ajudar na compreensão da natureza distribuída e conexa dos dados.

A mineração de dados possibilita análise a partir do processo de busca de conhecimento em base de dados, abordado na literatura com KDD *knowledge discovery in databases*. Mosafi (2022) define o KDD como um processo não trivial, interativo com o objetivo de identificar padrões compreensíveis, válidos, novos e potencialmente úteis a partir de uma grande quantidade de dados. A Figura 2.3 a seguir apresenta o processo de KDD e as etapas para a mineração.

Figura 2.3 – Processo *knowledge discovery in databases* - KDD



Fonte: Fernandes e Filho (2019)

O KDD é composto por etapas, conforme mostrado, na Figura 2.3 sendo elas:

- a) limpeza de dados: Etapa onde são removidos dados inconsistentes e fora de padrões;
- b) integração: Etapa que possibilita a integração com diferentes fontes de dados, mantendo a consistência e coerência;
- c) seleção: Etapa da mineração de dados onde são selecionados os dados relevantes;
- d) transformação: Etapa que atua na transformação ou consolidação dos dados em um formato apropriado e se utiliza de operações do tipo de sumarização ou agregação entre outras;
- e) avaliação: Etapa que busca identificar padrões e correlação dos dados para a análise de interesse;
- f) apresentação: Etapa que utiliza técnicas que permitem visualizar os dados em padrões, relacionamentos e descoberta de novos fatos;
- g) novo conhecimento: Nesta etapa, permite-se o conhecimento dos fatos, bem como introduzi-los no âmbito estudado (FERNANDES; FILHO, 2019; CÔRTEZ; PORCARO; LIFSCHITZ, 2002).

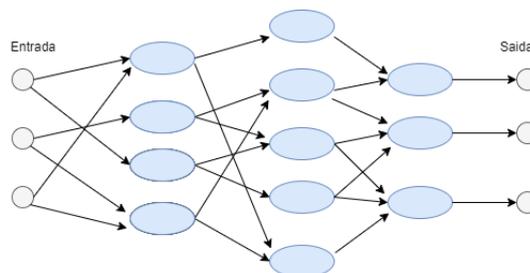
As redes sociais têm se tornado uma grande base de dados com informações, decorrentes da interação social entre os usuários que ocorre, diariamente, nesse ambiente. Assim, a mineração de dados aplicada nas redes sociais, busca obter resultados de comportamentos ou tendência social agrupando e analisando informações por grupos de interesse, opiniões correlacionadas, manifestação sobre diferentes assuntos do cotidiano das pessoas (COSTA, 2018; ARAÚJO; CRUZ, 2017).

## 2.5 Redes Neurais e Aprendizagem de Máquina

Redes Neurais Artificiais (RNAs) são modelos matemáticos inspirados no funcionamento do cérebro humano e estão diretamente ligadas a Inteligência Artificial (IA), sendo motivo de estudos constantes nas áreas de sistemas computacionais. Elas são compostas por neurônios artificiais interconectados, que processam informações e aprendem a partir de exemplos, assim, como ocorre em uma rede neural biológica, ou seja humana, as RNAs, atuam na busca de identificações, por intermédio de estratégias baseadas em meios observados. O objetivo de uma RNA não é replicar, mas sim servir de modelo para o aprendizado e resoluções de problemas complexos. O uso de RNAs tem se mostrado efetivo em diversas áreas, como processamento de imagem, reconhecimento de voz e, especialmente, em classificação de textos (COLLE; RODRIGUES; SILVA, 2022).

Exemplificando o funcionamento de uma RNA, basicamente seu desenvolvimento busca se assemelhar ao cérebro humano, buscando conhecimento obtido, por meio de etapas de aprendizagem e atualização de pesos sinápticos que são usados para armazenar o conhecimento. A sinapse é o nome dado à conexão existente entre neurônios. Assim, uma RNA pode ser representada através de um grafo onde os nós são os neurônios as ligações fazem a função das sinapses (FERNEDA, 2006). A Figura 2.4 traz essa representação.

Figura 2.4 – Grafo - RNA



Fonte: Ferneda (2006)

Na Figura 2.4, é possível observar as camadas de entrada, saída e as camadas do meio encontradas também na literatura como camadas ocultas. A camada de entrada é responsável por receber os dados a serem processados pela RNA as arestas de entrada apontam para a próxima camada. As camadas ocultas buscam transformar os dados de entrada e ajustá-los corretamente até que se atinja o formato e valor esperado para a camada de saída, assim como, na camada de entrada as arestas representam as conexões entre suas camadas. A camada de saída representa os dados resultantes depois do processamento realizado nas camadas ocultas.

As RNAs são utilizadas para processo de aprendizagem de máquina. A aprendizagem de máquina se concentra em desenvolver algoritmos e modelos computacionais capazes de aprender a partir de dados, sem a necessidade de serem explicitamente programados, ou seja, com o mínimo de intervenção humana, os algoritmos podem identificar padrões e tomar decisões. Os métodos de aprendizagem de máquina podem ser divididos em três categorias principais: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço (MONARD; BARANAUSKAS, 2003; FERNEDA, 2006).

- a) supervisionado - O algoritmo é treinado com um conjunto de dados rotulados, onde cada exemplo é associado a um rótulo ou classe. O objetivo do algoritmo é aprender uma função que possa prever corretamente a classe de novos exemplos. Por exemplo, um algoritmo de classificação de imagens pode ser treinado com um conjunto de imagens rotuladas, previamente definidos, o mesmo pode acontecer com classificação de texto. Nessa forma, um agente externo apresenta à RNA alguns conjuntos de padrões de entrada previamente definidos e seus correspondentes padrões de saída, ou seja, os dados para treino e teste possuem valores rotulados para que a RNA conheça o comportamento esperado. Assim, para cada entrada pré-definida explicitamente se espera uma resposta calculada. A resposta fornecida pela RNA é comparada à resposta esperada. Os valores das previsões, permite o ajuste com base nos erros verificados e são informados à RNA, assim são feitos ajustes a fim de melhorar suas futuras respostas;
- b) não supervisionado - O algoritmo é treinado com um conjunto de dados não rotulados, e o objetivo é encontrar padrões ou estruturas ocultas nos dados. Por exemplo, um algoritmo de clusterização pode ser usado para agrupar clientes com base em seus comportamentos de compra, sem que haja uma rotulagem prévia dos clientes. Nessa forma, não existe um agente externo para acompanhar o processo de aprendizado, ou seja, não existem resultados pré-definidos para utilizar como referência na aprendizagem, somente os padrões de entrada estão disponíveis para a RNA. A RNA processa as entradas e tenta encontrar semelhanças e diferenças entre os dados progressivamente, estabelecendo representações internas para codificar características e classificá-las automaticamente;
- c) aprendizagem por Reforço - O algoritmo aprende a partir de interações com um ambiente, recebendo recompensas ou penalidades com base em suas ações. O objetivo do algoritmo é aprender a tomar ações que maximizem a recompensa ao longo do tempo. Por exemplo, um algoritmo de jogos pode ser treinado para jogar um jogo de xadrez, recebendo recompensas por cada jogada bem-sucedida e penalidades por jogadas ruins, aprendendo assim a jogar melhor (MONARD; BARANAUSKAS, 2003; FERNEDA, 2006).

O método de aprendizagem de máquina supervisionado, será utilizada no processo de classificação de opiniões oriundas das redes sociais. No Capítulo 3, são apresentados mais detalhes.

### 2.5.1 PLN - Processamento de Linguagem Natural

O processamento de linguagem natural (PLN), também encontrada na literatura pela sua nomenclatura em Inglês NLP (*Natural Language Processing*), é um campo da inteligência artificial que se concentra em permitir que máquinas compreendam e processem a linguagem humana de forma semelhante aos seres humanos. Com PLN é possível realizar automatizações para medir semelhanças de requisitos, analisar similaridades e assim adotar critérios para classificar/filtrar, tais requisitos de acordo com a probabilidade de serem equivalentes (PAIS; CORDEIRO; JAMIL, 2022; FALCÃO; LOPES; SOUZA, 2022).

As redes neurais artificiais de PNL podem ajudar a resolver muitos problemas interessantes, como identificação de geolocalização, mineração de opinião públicas, análise de sentimentos/emoções, análise de tendências, extração de eventos, detecção de controvérsias, monitoramento de multidões, monitoramento de saúde pública, gerenciamento de desastres, etc (AL-AYYOUB et al., 2018). Existem diversos tipos de redes neurais usadas em PLN, como:

- a) redes Neurais Recorrentes (RNNs) - As RNNs são uma classe de redes neurais capazes de processar sequências de dados, como frases ou textos inteiros. Elas conseguem manter uma memória de longo prazo dos dados de entrada anteriores, o que as torna úteis para tarefas como tradução automática, geração de texto e análise de sentimento;
- b) redes Neurais LSTM (*Long Short-Term Memory*) são um tipo especial de rede neural recorrente (RNN) que foram projetadas para lidar com o desafio de aprender dependências de longo prazo em sequências de dados. A arquitetura LSTM é capaz de armazenar informações relevantes por um período prolongado, permitindo que a rede neural faça conexões entre eventos distantes no tempo;
- c) as redes neurais BiLSTM (*Bidirectional Long Short-Term Memory*) são uma extensão das redes LSTM, nas quais duas LSTMs são empregadas: uma processando a sequência na direção direta (do início ao fim) e outra processando a sequência na direção reversa (do fim ao início). Isso permite que a rede capture informações contextuais tanto do passado quanto do futuro em relação a um determinado ponto na sequência;
- d) redes Neurais Convolucionais (CNNs) - As CNNs são outra classe de redes neurais amplamente utilizadas em PLN. Elas foram desenvolvidas para extrair recursos importantes de dados de entrada bidimensionais, como imagens ou texto formatado em colunas, por meio de convoluções

em camadas consecutivas. Em PLN, as CNNs são usadas principalmente para classificação de texto, tais como tarefas de classificação de spam e detecção de sentimentos;

- e) redes Neurais Bayesianas (BNs) - As BNs são um tipo de rede neural que usa inferência probabilística para realizar análises de linguagem natural. Elas conseguem lidar com problemas de ambiguidade e incerteza, comuns em tarefas de PLN, e são usadas principalmente em análise semântica de texto, classificação de texto e tradução automática;
- f) redes Neurais de Transformadores (*Transformer*) - As *Transformer Networks* são uma classe relativamente nova de redes neurais, introduzida em 2017, e foram criadas especificamente para lidar com tarefas de processamento de linguagem natural. Elas são consideradas as melhores abordagens para muitas dessas tarefas, pois usam um mecanismo de atenção para identificar as palavras ou partes do texto mais importantes, permitindo uma maior precisão nas análises semânticas (SCHMIDT, 2019; KHAN et al., 2021; ANCHIÊTA et al., 2021).

Neste contexto, observa-se que existem diferentes implementações de redes neurais que são amplamente utilizadas no processamento de PLN devido à sua capacidade de lidar com a complexidade dos dados linguísticos e produzir resultados precisos em diversas tarefas. Cada tipo de rede neural apresenta suas próprias vantagens e limitações, e a escolha da abordagem mais adequada dependerá da tarefa específica em questão. Nesta pesquisa, será utilizado a RNA BERT, que é baseada na RNA *transformes*, para a classificação de opiniões coletadas das redes sociais, sobre as operadoras de redes móveis brasileiras. A escolha por este tipo de RNA é com base na literatura, que tem apontado *transformes* com uma revolução no PLN e apresentado melhor resultados em relação aos outros tipos de RNA para essas tarefas (KHAN et al., 2021).

### 2.5.2 Análise de Sentimentos

Ao longo dos últimos anos, a análise de sentimentos tem evoluído significativamente em termos de precisão e eficiência. Inicialmente, as abordagens de análise de sentimentos eram baseadas em regras heurísticas, que requeriam uma lista de palavras positivas e negativas para classificar o sentimento de um texto. Essa abordagem tinha limitações significativas, pois era difícil criar listas de palavras que cobrissem todas as nuances e contextos em que as palavras poderiam ser usadas (MATOS; MAGALHÃES; SOUZA, 2020; ZHANG; WANG; LIU, 2018).

Com o advento do aprendizado de máquina, as abordagens baseadas em modelos tornaram-se cada vez mais populares na análise de sentimentos. Essas abordagens permitem que o modelo aprenda automaticamente a partir de dados rotulados, sem a necessidade de criar regras específicas para cada

cenário. Esses modelos geralmente utilizam técnicas, que examina combinações de palavras adjacentes em um texto para identificar padrões de sentimento (OBEICA; MARTINS, 2022).

Com o surgimento de modelos de linguagem pré-treinados, como o BERT, a análise de sentimentos tornou-se ainda mais precisa e eficiente. Esses modelos são treinados em grandes conjuntos de dados, permitindo que eles capturem nuances e contextos mais complexos na linguagem natural. Eles também podem ser adaptados para tarefas específicas, como a análise de sentimentos em textos de mídias sociais, que muitas vezes apresentam desafios únicos, como o uso de gírias e abreviações (ZHANG; WANG; LIU, 2018).

Além disso, abordagens mais recentes, como a Rede Neural (*transformer*), têm sido aplicada com sucesso na análise de textos, com por exemplo em Yang et al. (2021) e Khan et al. (2022). Essas redes utilizam camadas de autoatendimento para aprender representações de texto de alta qualidade, permitindo que elas capturem informações contextuais mais complexas do que as arquiteturas mais tradicionais (KHAN et al., 2021).

Em resumo, a evolução na análise de sentimentos tem sido impulsionada pela evolução das técnicas de PLN e do aprendizado de máquina. Modelos cada vez mais sofisticados têm sido desenvolvidos, permitindo que eles capturem nuances e contextos mais complexos na linguagem natural. A tendência é que essas abordagens continuem evoluindo e se tornando cada vez mais poderosas, permitindo que as empresas e organizações usem a análise de sentimentos para tomar decisões mais informadas e efetivas.

Esta pesquisa utiliza a análise de sentimento para classificar as opiniões em três níveis, sendo eles, excelente, regular e péssimo, que posteriormente são associados ao escalar MOS.

### **2.5.3 Transformer**

A rede neural transformadora (ou *transformer*), revolucionou o processamento de linguagem natural, com sua arquitetura baseado no mecanismo de atenção, desde seu surgimento. Ela foi introduzida em 2017 por Vaswani et al. (2017), e se destacou por sua capacidade de lidar efetivamente com tarefas de PLN, como tradução automática, sumarização de texto, resposta a perguntas e muito mais, atraindo, assim, o interesse de pesquisadores acadêmicos, além da indústria, tornando-se uma arquitetura de referência no PNL (KHAN et al., 2021).

Ao contrário das arquiteturas de rede neural convolucionais e recorrentes, que eram amplamente utilizadas no PLN antes do surgimento dos *transformers*, essa nova abordagem trouxe uma inovação fundamental: a atenção. A atenção é um mecanismo que permite que o modelo dê mais peso a certas partes da entrada, com base em sua relevância para a tarefa em questão. Com a atenção, o *transformer*

pode capturar relações entre palavras distantes em um texto, sem a necessidade de sequências fixas de operações (VASWANI et al., 2017).

Por exemplo, enquanto as redes BiLSTM, que são baseadas em células recorrentes das RNR, processam sequencialmente a entrada palavra por palavra, os *transformers* utilizam mecanismos de atenção para calcular a importância de cada palavra em relação a todas as outras palavras do texto. Essa atenção permite que os *transformers* capturem relações contextuais complexas, mesmo entre palavras distantes, sem a necessidade de dependência sequencial. Dessa forma, eles são capazes de processar informações de forma mais eficiente e lidar com sequências de comprimentos variáveis (KHAN et al., 2021; VASWANI et al., 2017).

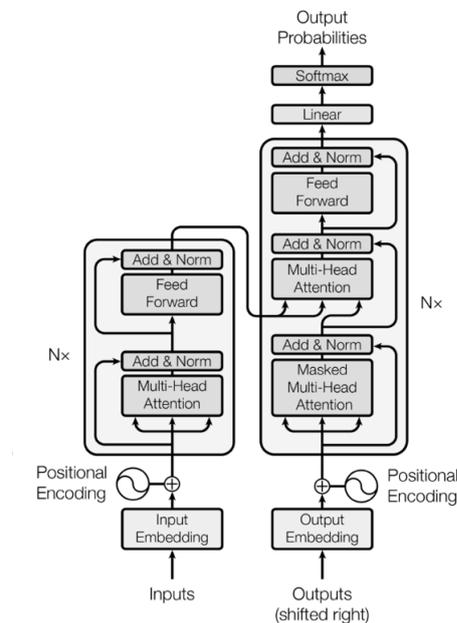
Uma outra diferença significativa entre os *transformers* e as redes BiLSTM é a capacidade de processamento paralelo. Enquanto as redes recorrentes processam sequencialmente as palavras, os *transformers* podem calcular a atenção para todas as palavras simultaneamente, o que permite um processamento mais rápido e eficiente em GPUs e acelera a convergência do treinamento (KHAN et al., 2021; VASWANI et al., 2017).

Além disso, os *transformers* introduzem o conceito de posicionamento relativo, que permite que o modelo capture informações de posição relativa entre as palavras sem depender exclusivamente da ordem sequencial. Essa capacidade é particularmente útil em idiomas com ordens de palavras flexíveis ou em tarefas que requerem uma compreensão robusta das relações de posição entre as palavras (KHAN et al., 2021; VASWANI et al., 2017).

O *transformer* possui arquitetura *encoder-decoder* ou, em português, codificador e decodificador, nessa arquitetura, a RN recebe como entrada uma sequência de palavras, codificá-as em representações nas chamadas "camadas de atenção" e as decodifica em palavras novamente. À medida que o modelo processa cada palavra em posição de sequência de entrada, um mecanismo de atenção computa a importância de palavras em outras posições para codificar a palavra atual. Exemplificando, o *transformer* atua no entendimento das relações entre as palavras de uma frase, em vez de computar uma por uma em ordem. No processamento dos textos tanto o codificador quanto o decodificador são compostos por uma pilha de seis camadas idênticas. No codificador as camadas são divididas em duas subcamadas, onde a primeira é denominada "*Multi-head Self-Attention*" e a segunda camada é uma RNA *feedforward*. O decodificador, possui as mesma subcamadas acrescido de uma terceira subcamada de atenção que processa os dados vindos da camadas do codificador (VASWANI et al., 2017).

Na Figura 2.5, ilustra-se a arquitetura *transformer* de Vaswani et al. (2017):

Figura 2.5 – Arquitetura Transformes



Fonte: Vaswani et al. (2017)

Na Figura 2.5, apresenta-se a arquitetura do *transformer* onde, no lado esquerdo, encontra-se o processo de *encoder* (codificador), que processa sua entrada, ou seja, um texto/frase recebida e gera codificações que contenham informações sobre quais partes das entradas são relevantes entre si, já, do lado direito, esta *decoder* (decodificador) que pega todas as codificações e as processa usando suas informações contextuais incorporadas para gerar uma sequência de saída. Demais ações são:

- input - output* - Recebe os *embedding*, ou seja, um texto/frase completa para o processamento, a diferença e que *input* para processo de codificação e *output* para o processo de saída no decodificador;
- positional encoding* - Atua na adição de números depois da camada de *embedding* indicado importância para as ordens das palavras;
- self-attention* - Atua no codificador e no decodificador e busca a recomposição de uma sequência para saber como cada elemento está relacionado com os outros dentro de um texto/frase;
- multi-head Self-Attention* - A Autoatenção de múltiplas cabeças, está relacionada ao principal mecanismo da arquitetura *transformer*, ou seja, o conceito de atenção. Nessa etapa, são calculadas, paralelamente, as pontuações das palavras. Assim, todos os cálculos de atenção semelhantes são combinados para produzir uma pontuação de atenção final;
- add & norm* - Atua no não esquecimento de informações da etapa anterior, realizando a propagação, ou seja, o *backpropagation*, no processo de aprendizagem;
- feed forward* - Camadas densas, compostas por transformações lineares;

- g) *linear* - Trata-se da saída do decodificador que após, passa por uma camada densa de acordo com o tamanho do vocabulário e com aplicação da função *softmax*, gera a probabilidade para cada palavra;
- h) *softMax* - É a ultima camada da arquitetura onde são atribuídos, para cada palavra, no vocabulário, uma probabilidade.

Segundo Khan et al. (2021) o *transformer* tem apresentando um desempenho exemplar para diferentes tarefas de PLN, como, por exemplo, a tradução de textos automaticamente ou criação de *chats* para perguntas e respostas, além de se tornar base de diferentes modelos de classificação como o BERT que será utilizado nesta pesquisa, o GPT (*Generative Pre-trained Transformer*), RoBERTa (*Robustly Optimized BERT Pretraining Approach*) e o T5 (*text-to-Text Transfer Transformer*).

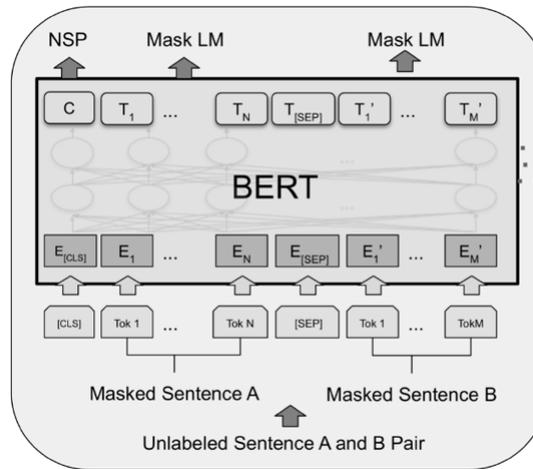
#### 2.5.4 BERT

BERT (*Bidirectional Encoder Representations from Transformers*) é modelo pré-treinado para PLN, que foi introduzido por Devlin et al. (2018), sendo, posteriormente, utilizado pela empresa Google em seu buscador *web*. Desde o seu seguimento, o BERT tem sido alvo de contantes estudos e experimentos de PLN, tornando-se uma base onipresente e inspirando vários estudos com base em diferentes modelos e propostas (ROGERS; KOVALEVA; RUMSHISKY, 2020).

O BERT é um modelo pré-treinado com dados não rotulados, que utiliza modelagem de aprendizado profundo, é baseado em *transformer* com múltiplas camadas e atua com representações bidirecionais de palavras, baseadas no contexto à esquerda e à direita. O BERT foi disponibilizado em dois tamanhos, sendo eles o BERT-BASE (contendo 12 camadas transformadoras e 768 camadas ocultas) e o BERT-LARGE (contendo 24 camadas transformadoras e 1024 camadas ocultas). Ambos os modelos foram treinados nos mesmos conjuntos de dados, com o *BookCorpus* (800 milhões de palavras) e o *Wikipedia* (2,500 milhões de palavras) (DEVLIN et al., 2018).

O pré-treinamento do BERT, consiste em dois modos, sendo eles o modo MLM (*Masked Language Model*) e o modo NSP (*Next Sentence Prediction*). Durante o processo de pré-treinamento ambos são combinados com a finalidade de melhorar o desempenho de aprendizado. Na Figura 2.6, apresenta-se o processo de pré-treinamento.

Figura 2.6 – Pré-treinamento BERT



Fonte: Devlin et al. (2018)

No processo MLM o modelo mascara aleatoriamente 15% das palavras, substituindo-as pelo (*token*) [MASK] na entrada, em seguida, executa toda a frase com o objetivo de prever as palavras mascaradas, já o processo NSP o modelo realiza a concatenação de duas sentenças mascaradas e recebidas nas entradas, durante o pré-treinamento. Nesse contexto, pode ser que as frases estivessem próximas umas das outras, no texto original, mas também ocorre de não estarem. O objetivo do modelo é prever se as duas sentenças estavam seguindo uma à outra ou não (DEVLIN et al., 2018);

Inicialmente, o BERT foi desenvolvido e aplicado à língua inglesa, mas, diferentes pesquisadores têm desenvolvido modelos baseados em outros idiomas, assim novos pré-treinamentos têm ocorrido. Um exemplo, de novo pré-treinamento é o BERT Multilíngue, que é uma versão do BERT que foi pré-treinada em um conjunto de dados que abrange mais de 100 idiomas. Ele foi projetado para ser capaz de lidar com o processamento de linguagem natural em diferentes línguas, permitindo uma aplicação mais ampla em contextos multilíngues. Com o BERT Multilíngue, é possível realizar tarefas de tradução automática, análise de sentimentos, classificação de texto e outras tarefas de PLN em diversos idiomas, aproveitando o conhecimento prévio adquirido durante o treinamento em vários corpus multilíngues (KHAN et al., 2022; SOUZA; NOGUEIRA; LOTUFO, 2020).

Outro exemplo é BERTimbau, que é uma versão do BERT específica para o idioma português brasileiro. Ele foi treinado em um grande conjunto de dados em língua portuguesa, com o objetivo de capturar as particularidades e nuances do idioma. O BERTimbau, permite que os modelos de PLN tenham um melhor desempenho e uma compreensão mais precisa do português brasileiro e, assim como BERT Multilíngue, pode ser aplicado em várias tarefas, como análise de sentimento, sumarização de texto, identificação de entidades nomeadas, entre outras (SOUZA; NOGUEIRA; LOTUFO, 2020).

Tanto o BERT Multilíngue quanto o BERTimbau são ferramentas valiosas para pesquisadores, desenvolvedores e profissionais que trabalham com processamento de linguagem natural em ambientes multilíngues ou no contexto específico da língua portuguesa. Eles permitem uma abordagem mais eficiente e precisa para o processamento de texto em vários idiomas, aproveitando os benefícios da arquitetura do BERT e o conhecimento prévio adquirido durante o treinamento em larga escala. Essas extensões do BERT representam avanços importantes no campo do PLN, impulsionando a capacidade de compreensão e análise de texto em diferentes idiomas e contribuindo para a evolução das tecnologias linguísticas.

Devido à extensa literatura existente sobre o BERT, conforme discutido por Rogers, Kovaleva e Rumshisky (2020), bem como sua comprovada eficácia no processo de classificação de textos, conforme demonstrado por Silva (2020) e Lee, Lee e Ahn (2022), esta pesquisa adotou o BERT como base para a classificação de opiniões provenientes de redes sociais. No Capítulo 3, apresentamos o modelo de classificação utilizado, no qual comparamos tanto o BERT Multilíngue quanto o BERTimbau para o processo de classificação das opiniões.

## 2.6 Trabalhos Relacionados

Pesquisas que abordam o QoE têm sido amplamente publicadas com métricas e medições que colaboram, principalmente, em serviços prestados via redes de telecomunicação. Igualmente a classificação de texto e análise de dados provenientes das redes sociais têm se tornado um grande campo para pesquisadores, isso pode ser observado nas buscas em diferentes bases de publicações de artigos e periódicos. A seguir são citadas algumas publicações que abordam QoE e classificação de textos.

O estudo de Cui et al. (2018) buscou otimizar a alocação de recursos em redes de acesso múltiplo não ortogonal multicelula multiportadora (MC-NOMA) para melhorar a qualidade de experiência (QoE). Propuseram um algoritmo de duas etapas com baixa complexidade, baseado em correspondência 2D, e estratégias de alocação de energia ótima usando *branch and bound*. As redes NOMA superaram as convencionais em termos de QoE, e os algoritmos propostos alcançaram melhoria significativa na equidade em relação à taxa de soma. Pesquisadores acreditam que esses avanços podem direcionar melhorias na QoE para serviços como navegação na web, vozes, *streaming* de áudio e vídeo, entre outros.

O estudo de Bermudez et al. (2019) focou na avaliação da qualidade de experiência (QoE) em transmissões de vídeo online usando o modelo ITU-T P.1203 em redes LTE. Medições de QoE são cruciais para provedores de serviços, gerentes de rede e produtores de conteúdo. A implementação funcional do modelo e o estudo de caso em *Live vídeo Streaming* (LVS) mostraram que o protocolo

DASH teve maior QoE (MOS) em relação ao RTMP. O método proporciona estimativa objetiva de QoE de forma econômica e com baixa complexidade, podendo ser usado em cenários mais extensos.

O estudo de Zhong et al. (2019) propôs uma estrutura analítica que combina geometria estocástica e teoria de filas para avaliar a relação entre o tráfego espaço-temporal perplexo e o desempenho da rede com mobilidade. Definiram matematicamente a qualidade da experiência e o custo do sistema, explorando o equilíbrio entre eles. Os resultados mostram que a QoE nem sempre aumenta com o aumento do custo do sistema, mas estabiliza após um certo ponto. Além disso, demonstram que a mobilidade em redes sem fio pode reduzir o atraso e melhorar a QoE. Essa pesquisa fornece uma referência útil para o projeto de redes sem fio com mobilidade espaço-temporal, sendo significativa para o projeto das redes de próxima geração que atendam aos requisitos de QoE sob restrições de custo do sistema.

O estudo de He, Jiang e Wei (2021) propôs um algoritmo dinâmico de seleção de rede com eficiência energética, considerando diferentes tipos de negócios e status de rede. O algoritmo *Q-learning* foi usado para otimizar a QoE e o consumo de energia da rede, com aprendizado online e análise dos tempos de comutação e consumo de energia. Os resultados mostraram que o algoritmo de seleção dinâmica de rede e destino é ideal para melhorar a QoE do usuário, maximizar o valor da QoE, otimizar o consumo de energia da rede e reduzir o número de comutações.

O trabalho de Park e Chung (2021) propôs um esquema de colaboração de borda baseado em previsão de recursos para melhorar a QoE. A pesquisa aborda o crescimento de aplicativos e dispositivos IoT, que não conseguem atender às crescentes demandas de recursos computacionais. Os servidores de borda podem fornecer recursos suficientes, mas a eficiência diminui com o aumento do número de dispositivos conectados. Para resolver isso, propõe-se um processo de colaboração entre os servidores de borda. Os esquemas de colaboração de borda existentes têm baixa QoE, pois não consideram os recursos de computação ou o tempo de comunicação de outros servidores. O esquema proposto estima o uso de recursos de computação com base nas tarefas dos dispositivos e colabora probabilisticamente com outros servidores de borda. Os resultados mostraram que o esquema proposto alcançou baixo tempo de conclusão e alta taxa de sucesso, superando os esquemas existentes para várias tarefas.

O estudo de Ciambone et al. (2018) avaliou políticas de gestão orientadas por QoE para VoIP e *streaming* de vídeo em redes móveis. Utilizaram o simulador *SimuLTE/OMNeT++* para reproduzir redes sem fio e analisar o *tradeoff* entre taxa de perda de pacotes e atraso, visando a QoE dos usuários finais. Os resultados destacaram a importância de políticas de gestão que reconhecem e mudam o caminho do tráfego para manter a qualidade dos serviços. Escolhas precisas de parâmetros de projeto foram consideradas fundamentais para alcançar níveis de QoE desejáveis

O estudo de Daengsi e Wuttidittachotti (2019) avaliou a qualidade VoIP do *codec* G.729, aprimorando o E-Model simplificado com o modelo MOS de previsão subjetiva. A pesquisa foi baseada em usuários tailandeses nativos que utilizam o idioma tailandês-tonal. Os resultados mostraram que a modificação do E-Model simplificado apresentou 58% de melhoria em relação ao modelo original, reduzindo erros. Os pesquisadores acreditam que essa pesquisa pode contribuir para novas investigações em países com idiomas tonais, como na Tailândia, e realizar testes subjetivos com seus falantes nativos, utilizando os métodos simplificados de aprimoramento do E-Model.

O trabalho de Hu et al. (2020) avaliou a relação entre Qualidade de Serviço (QoS) e Qualidade de Experiência (QoE) em redes VoIP. Diferentemente de estudos anteriores que analisaram parâmetros de QoS de forma independente, os pesquisadores investigaram o impacto das deficiências de rede na QoE e estabeleceram limites para diferentes parâmetros em relação à QoE. Utilizaram a pontuação MOS como métrica de QoE e realizaram análises em um banco de testes VoIP. Para gerar conjuntos de dados experimentais, usaram ferramentas como o *NIST Net* para emulação de rede e o aplicativo *OpenPhone* para VoIP. Os resultados mostraram a relação não linear entre parâmetros como o *jitter* e a QoE, utilizando algoritmos de correlação de distância e aprendizado de máquina. O estudo contribuiu para uma compreensão mais abrangente dos impactos da QoS na QoE, aprimorando a avaliação da QoE no VoIP e fornecendo requisitos típicos de desempenho de rede para alcançar uma QoE desejada em aplicativos VoIP.

O estudo de Santos et al. (2023) aborda a qualidade de serviços de *streaming* de áudio e a preferência do usuário em relação ao conteúdo. Foram conduzidos testes subjetivos para determinar os parâmetros-chave nesse tipo de serviço, relacionados com as interrupções no *streaming*. A métrica AsQM foi proposta, considerando esses fatores e a preferência do usuário. Os resultados mostraram que sua implementação tem baixo impacto em consumo de energia, processamento e memória. O estudo destaca a importância de considerar a preferência do usuário e os parâmetros das interrupções para garantir uma boa qualidade de *streaming* de áudio.

O estudo de Morente-Molinera et al. (2018) buscou extrair informações úteis de debates em redes sociais, especialmente no *Twitter*. Eles propuseram uma nova maneira de analisar o progresso do debate, consenso entre os participantes e preferências. Utilizaram a análise de sentimento para medir a aceitação dos usuários em relação a tópicos e obter valores de preferência para tomada de decisão em grupo. Os resultados indicaram que o método proposto é útil para analisar decisões em grupo e a organização das discussões. A ferramenta desenvolvida pode ser útil para analisar grandes quantidades de informações em redes sociais e auxiliar outras pesquisas.

O estudo de Chang, Ku e Chen (2019) analisou avaliações de hotéis feitas por consumidores no *TripAdvisor*, utilizando a extração de dados de redes sociais. A pesquisa propôs uma estrutura integrada para análise, incluindo rastreamento de dados, pré-processamento, árvore sensível ao sentimento, classificação do *kernel* de convolução, extração de aspectos e detecção de categorias, com análises visuais. Os resultados mostraram que a abordagem superou algoritmos de linha de base e métodos de classificação de sentimentos conhecidos, alcançando alta precisão (0,95) e *recall* (0,96). Além disso, observou-se que viajantes a negócios tendem a dar classificações mais baixas, enquanto casais tendem a dar classificações mais altas. Houve também variação nas avaliações de acordo com o período de hospedagem. Os pesquisadores acreditam que o novo método proposto, automatizado e baseado na detecção de sentimentos com aspectos de avaliações *on-line*, pode contribuir para pesquisas em domínios multidisciplinares como turismo, sistemas de informação e ciência da computação.

A pesquisa de Sundararaj e Rejeesh (2021) teve como objetivo analisar a mudança de comportamento do cliente em relação a diversos produtos por meio de Redes Sociais. A análise incluiu decisões de compra, comportamento da marca, influência do preço nas compras e opiniões dos clientes. Utilizando o método *Chi-Square*, os autores avaliaram o perfil e a avaliação do cliente em etapas, calculando a qualidade da avaliação e comportamento do cliente, bem como comparando as avaliações. Os resultados mostraram que a qualidade do produto variou de acordo com o perfil do cliente, com alta eficiência e precisão na coleta de informações via perfil e avaliação. Os pesquisadores acreditam que essa análise a partir de dados de Redes Sociais contribui para empresas analisarem seus públicos-alvo, além de melhorar a sociedade ao considerar o aspecto psicológico do comportamento de compra dos consumidores.

A pesquisa de Li (2022) aborda métodos de Processamento de Linguagem Natural (PLN) baseados em Aprendizado Profundo para análise de sentimentos em Redes Sociais. Os pesquisadores utilizaram a aprendizagem profunda para projetar um modelo aplicável no contexto social real, visando a compreensão e representação semântica da linguagem natural com informações contextuais. Foram analisados algoritmos de análise de sentimento em dois níveis: unimodal e multimodal, desenvolvendo um modelo para análise de sentimento de texto e outro para análise multimodal, considerando imagens e texto de redes sociais. Em experimentos comparativos com modelos existentes em diferentes conjuntos de dados, a precisão dos dois modelos propostos superou os modelos de referência em 4,45% e 5,2%, respectivamente, validando a eficácia dos mesmos. Os pesquisadores também concluíram que dados de texto com sentimento negativo foram mais discriminativos na detecção de rumores online em comparação com dados de texto com sentimento positivo e neutro.

A pesquisa de Santos e Lopes (2022) aborda QoE em redes móveis no Brasil, usando extração de opiniões do Twitter. O objetivo foi desenvolver indicadores automatizados de QoE para operadoras

móveis, atribuindo pontuações com base na MOS. As contribuições incluem analisar a correlação entre a qualidade do serviço e as reclamações dos usuários no Twitter, e propor um novo indicador baseado na correlação entre qualidade e dados da plataforma. Os resultados mostram que os usuários tendem a usar o Twitter para reclamações, em vez de expressar satisfação.

A pesquisa de Liu, Shin e Burns (2021) examinou o impacto do marketing de mídia social de marcas de luxo no envolvimento do cliente. Usando análise de big data e processamento de linguagem natural no Twitter, analisaram 3,78 milhões de tweets de 15 principais marcas de luxo. Os resultados mostram que focar nas dimensões de entretenimento, interação e tendência do marketing de mídia social aumenta o envolvimento do cliente, enquanto a personalização não teve impacto significativo. Essas descobertas têm implicações importantes para o design e gerenciamento do marketing de mídia social para marcas de luxo.

A pesquisa de Yang et al. (2021) desenvolveu e avaliou um sistema automático para classificar conversas sobre serviços de saúde no Twitter, com foco na Medicaid, a maior fonte de cobertura de saúde nos Estados Unidos. Eles coletaram e rotularam manualmente dados em 5 categorias para treinamento e teste de algoritmos de classificação de texto. Os resultados mostraram que o classificador BERT obteve o melhor desempenho, alcançando acurácias de 81,7% e 80,7% para os dois conjuntos de dados.

A pesquisa de Khan et al. (2022), teve como objetivo a criação de um conjunto de dados e anotado, manualmente, para análise de sentimento com base na língua Urdu. Para atingir os seus objetivos foi utilizado o aprendizado de máquina, baseado em regras (SVM, NB, *Adabboost*, MLP, LR e RF) e aprendizado profundo (CNN-1D, LSTM, Bi-LSTM, GRU e Bi-GRU), além de ajustar o BERT multilíngue para análise de sentimento em Urdu. Em seus resultados, os autores apontam que o modelo BERT multilíngue proposto com incorporação de palavras pré-treinadas, superou o aprendizado profundo, aprendizado de máquina e classificadores baseados em regras e alcançou uma pontuação *F1-Score* de 81,49%.

A pesquisa aqui apresentada tem como objetivo realizar uma análise abrangente da Qualidade de Experiência (QoE), levando em consideração os trabalhos mencionados anteriormente sobre esse tema, além de explorar o potencial dos textos das redes sociais. O diferencial dessa pesquisa reside na metodologia empregada, que busca estabelecer correlações entre a QoE e a classificação de textos provenientes das redes sociais.

O Quadro 2.1, identifica os trabalhos citados e suas características distintas. No entanto, observamos uma lacuna na literatura em relação à abordagem subjetiva e automática na avaliação da QoE, o que representa uma oportunidade para esta pesquisa contribuir significativamente.

Quadro 2.1 – Disponibilidade de dados para pesquisa acadêmica - Twitter

Tema	Autores e pesquisa
QoE sobre redes móveis	(CUI et al., 2018) - "Alocação de recursos baseada em QoE para redes multi-células NOMA" (ZHONG et al., 2019) - "QoE e custos para redes sem fio com mobilidade sob tráfego espaço-temporal" (HE; JIANG; WEI, 2021) - "Um algoritmo dinâmico de seleção de rede com eficiência energética baseado em QoE" (PARK; CHUNG, 2021) - "Esquema de colaboração de borda baseado em previsão de recursos para melhorar a QoE"
QoE sobre serviços específicos	(BERMUDEZ et al., 2019) - "Avaliação de <i>streaming</i> de vídeo ao vivo usando o modelo ITU-T P.1203 QoE em redes LTE" (CIAMBRONE et al., 2018) - "Avaliação de políticas de gestão orientadas por QoE para Prestação de serviços de VoIP e <i>streaming</i> de vídeo" (DAENGS; WUTTIDITTACHOTTI, 2019) - "Modelagem QoE para Voz sobre IP: Aprimoramento Simplificado do E-Model Utilizando o Modelo Subjetivo de Previsão MOS: Um Caso de Usuários G.729 e Tailandeses" (HU et al., 2020) - "Avaliando QoE em redes VoIP com mapeamento de QoS e algoritmos de aprendizado de máquina" Santos et al. (2023) - "AsQM: Métrica de qualidade de <i>streaming</i> de áudio baseada em deficiências de rede e preferências do usuário"
Classificação de textos das redes sociais	(MORENTE-MOLINERA et al., 2018) - "Analisando discussões em Redes Sociais usando métodos de tomada de decisão em grupo e análise de sentimentos" (LI, 2022) - "Métodos de Processamento de Linguagem Natural Baseados em <i>Deep Learning</i> para Análise de Sentimentos em Redes Sociais" (KHAN et al., 2022), "Análise de sentimento multiclasse de texto urdu usando BERT multilíngue"
Classificação de texto das redes sociais sobre serviço ou produto	(CHANG; KU; CHEN, 2019) - "Análise de Redes Sociais: Extraindo e visualizando classificações e comentários de hotéis <i>Hilton</i> do <i>TripAdvisor</i> " (SUNDARARAJ; REJEESH, 2021) - "Uma análise comportamental detalhada sobre o comportamento e mudança do consumidor e do cliente em relação aos sistemas de Redes Sociais" (SANTOS; LOPES, 2022) - "Análise de QoE para serviços de rede móvel usando extração de opinião do Twitter" (LIU; SHIN; BURNS, 2021) - "Examinando o impacto do marketing de mídia social de marcas de luxo no envolvimento do cliente: Usando análise de <i>big data</i> e processamento de linguagem natural" (YANG et al., 2021) - "Desenvolvendo um sistema automático para classificar conversas sobre serviços de saúde no <i>Twitter</i> : estudo de caso para <i>Medicaid</i> "

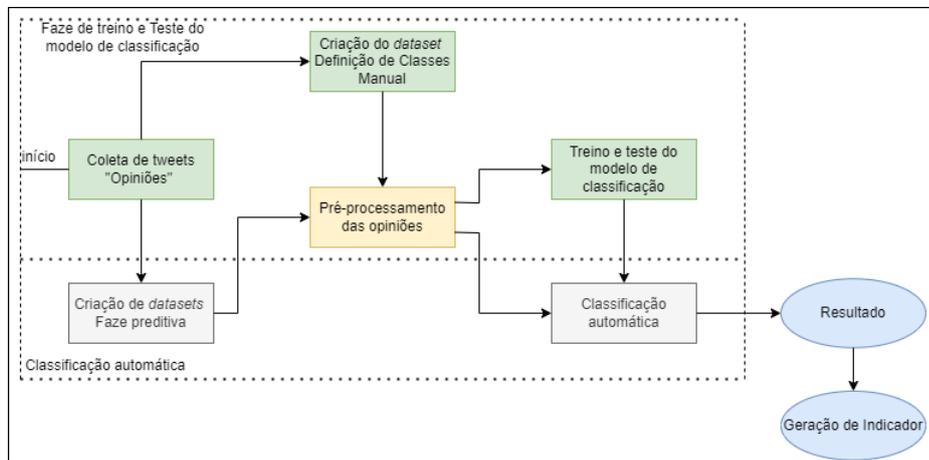
Fonte: Do Autor (2023)

### 3 METODOLOGIA

Nos capítulos anteriores, esta pesquisa foi desenvolvida mediante a utilização de acervos teóricos, possibilitando o desenvolvimento do referencial teórico, bem como, a verificação de trabalhos relacionados. Foram utilizados livros de referência na área de pesquisa, além das bases *Web of Science*, Google Acadêmico e *IEEE Xplore*.

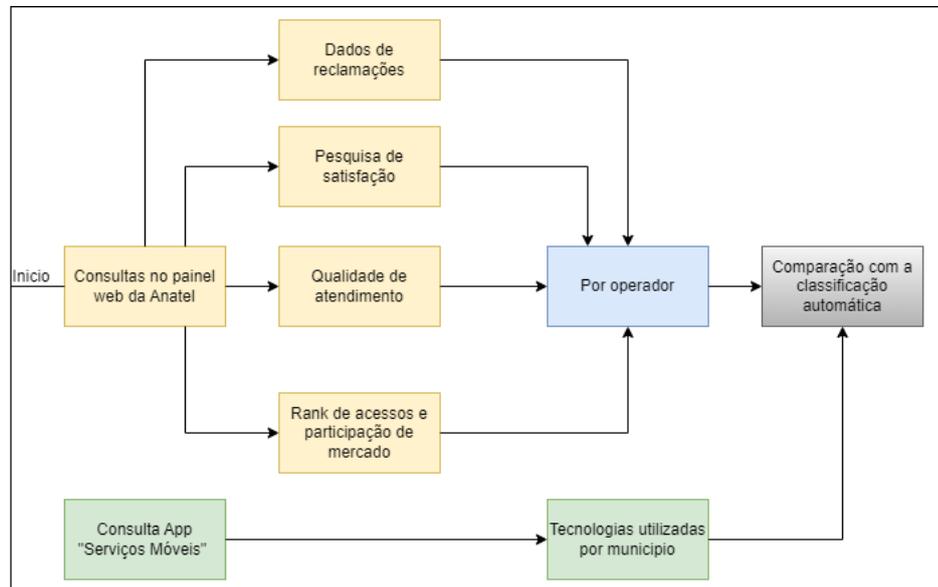
No capítulo atual, é abordada a metodologia empregada nesta pesquisa, que consiste em duas fases. Na primeira fase, são coletados *tweets* (Opiniões) com o objetivo de classificá-los automaticamente. Nessa fase, é criado um conjunto de dados (*DataSet*) para treinar e testar o modelo classificador, contendo todas as opiniões, bem como conjuntos de dados individuais por operadora e cidade para a classificação automática. Na segunda fase, os resultados obtidos serão comparados com os dados disponibilizados pela Anatel. Nessa fase, é utilizada a plataforma *web* e o aplicativo "Serviços Móveis" da Anatel para gerar gráficos sobre as reclamações dos usuários e consultar as tecnologias utilizadas pelas empresas para atender os municípios, a fim de compará-las com os resultados obtidos na primeira fase. As figuras 3.1 e 3.2 apresentam os fluxogramas da primeira fase e da segunda fase, respectivamente.

Figura 3.1 – Fluxograma da Primeira fase - Classificador



Fonte: Do Autor (2023)

Figura 3.2 – Fluxograma da Segunda fase - Comparação de Resultados



Fonte: Do Autor (2023)

Na seqüência, do capítulo é abordado os processo de cada etapa dos digramas apresentados acima.

### 3.1 Classificador e coleta de dados

Para a coleta de opiniões do *Twitter*, e a classificação preditiva das opiniões, nesta pesquisa, utilizou-se os recursos disponibilizados pela linguagem de programação *Python*. De acordo com Millman e Aivazis (2011), desde o surgimento dessa linguagem, ela vem se tornando um padrão, de fato, para pesquisas científicas exploratórias, interativas e baseadas em computação. Apesar dessa linguagem não ter sido projetada para a comunidade científica, ela tem atraído o interesse científico com uma sintaxe expressiva e uma rica coleção de tipos de dados internos (como strings, listas, etc..), além de diferentes bibliotecas que contribuem para a exploração de dados. Algumas características do *Python* é que, se trata de uma linguagem de alto nível orientada a objetos, de tipagem dinâmica e forte, interpretativa e interativa, possui sintaxe clara que favorece no desenvolvimento de código-fonte além de ser de código aberto com licença compatível a GPL (*General Public License*) (BORGES, 2014).

A fase, de coleta e classificação das opiniões, foram executadas no *Jupyter Notebook*. O *Jupyter Notebook* é uma plataforma interativa e de código aberto que permite criar e compartilhar documentos que contêm código, visualizações, texto e equações matemáticas. Ele se tornou uma ferramenta popular entre cientistas de dados, pesquisadores e programadores, devido à sua flexibilidade e facilidade de uso (MCKINNEY, 2022).

Uma das principais características do *Jupyter Notebook* é a sua capacidade de combinar células de código, texto e visualizações em um único documento. Isso permite que os usuários executem e visualizem os resultados do código em tempo real, facilitando a análise de dados, a experimentação e a criação de relatórios interativos (MCKINNEY, 2022).

O computador utilizado possui as seguintes especificações técnicas: processador 11th Gen Intel(R) Core(TM) i7-11390H @ 3.40GHz, 16GB de memória RAM, SSD M.2 de 48GB e sistema operacional Windows 11 Pro.

A seguir são apresentadas as etapas realizadas na primeira fase: sendo elas o processo de coleta de *tweets*, criação do conjunto de dados para treino e teste do modelo de classificação, análise geral, por operadora e localidade, balanceamento do conjunto de dados, modelo de classificação, métricas de avaliação do modelo, processo de predição automática e conversão para escalar MOS.

### 3.1.1 Coleta de Opiniões

A mineração de opiniões em redes sociais tem se mostrado um campo promissor para análise e extração de informações subjetivas expressas em textos, como sentimentos, emoções e opiniões, devido ao uso exponencial de diferentes usuários nesse ambiente. Entre as redes sociais mais utilizadas para expressar opiniões, o Twitter se destaca como uma plataforma popular e acessível. A mineração de *tweets* pode fornecer *insights* valiosos sobre a percepção do público em relação a produtos, serviços e eventos, auxiliando empresas e pesquisadores na tomada de decisões estratégicas (COSTA, 2018).

A coleta de opiniões, desta pesquisa, é via API (*Application Programming Interface*) ou em português “Interface de Programação de Aplicativos”. API é um conjunto de rotinas e padrões de programação, para acesso a um aplicativo (*software*) e é baseada em plataformas de desenvolvimento *web*. Geralmente as APIs são disponibilizadas por empresas que têm a intenção de permitir que diferentes *softwares* interajam com sua aplicação. A rede social *Twitter* disponibiliza para desenvolvedores e pesquisadores acadêmicos acessos via API, possibilitando a recuperação e análise de dados de forma programática. Este método é utilizado por pesquisas como a de Batra et al. (2021) e Khan et al. (2020).

Para o acesso a API do *Twitter* é necessária a criação de uma conta "*developer*", onde é indicado por quais motivos estará utilizado a API, pois o *Twitter* possui diferentes modalidades de acessos. No contexto desta pesquisa foi utilizado a modalidade estudante, sendo necessário se inscrever no programa estudante. O processo de inscrição envolve preencher um formulário e enviar um comprovante de matrícula em uma instituição de ensino superior. Após a aprovação, foi possível criar um projeto na plataforma e obter as credenciais de acesso à API. O Quadro 3.1, a seguir, apresenta detalhes desta modalidade.

Quadro 3.1 – Disponibilidade de dados para pesquisa acadêmica - Twitter

Especificações	Busca
Limite do	10 milhões de <i>tweets</i> / mês
Regras de consulta	1024 caracteres, 1000 regras de <i>streaming</i>
Taxas de transmissão	50 solicitações / 15 minutos, por aplicativo

Fonte: Adaptado de Twitter (2022)

A biblioteca *Python* utilizada para o acesso foi a *Tweepy*. A *Tweepy* é uma biblioteca de código aberto para *Python* que permite acessar e interagir com a API do Twitter de forma fácil e intuitiva. Com o *Tweepy*, os desenvolvedores podem escrever programas para coletar dados do Twitter, publicar *tweets*, buscar por palavras-chave, obter informações de usuários entre outras utilidades. Uma das principais vantagens do *Tweepy* é sua simplicidade de uso. Ele fornece uma interface simples e consistente para acessar os recursos da API do Twitter. Os desenvolvedores podem autenticar suas aplicações com suas credenciais do Twitter e começar a realizar solicitações à API em questão de minutos (BATRA et al., 2021; KHAN et al., 2020).

Este processo de coleta de *tweets* contendo opiniões sobre as operadoras Oi, Tim, Vivo e Claro apresentou um desafio significativo devido à vasta quantidade de dados disponíveis nas redes sociais. Para melhorar a eficiência da busca e garantir a relevância das informações coletadas, utilizamos a técnica de vetores com termos-chave específicos, como por exemplo, "#Oi", "@Oi\_oficial", "#Tim", "@Tim-Barisl", "#Vivo", "@vivobr", "#Claro" e "@ClaroBrasil". Essa abordagem permitiu direcionar de forma mais precisa as buscas, otimizando o processo de coleta e possibilitando análises mais acuradas posteriormente.

O processo de coleta foi realizado em duas etapas. Na primeira etapa buscou-se opiniões para a criação de um *dataSet* de treino e teste. Nessa etapa a busca foi por opiniões para todas as operadoras, e os *tweets* foram armazenado em um arquivo (.CSV). Na segunda etapa, a busca por opiniões foi realizada de forma individual, ou seja, por operadora. Nesta etapa criou-se *dataSets* por operadora e por localidade, esses *dataSets* foram os utilizados para a predição automática. Em ambas as etapas, a biblioteca *Pandas* foi utilizada para a criação e armazenamento dos *dataSets*. A biblioteca *Pandas* possui o código aberto e é utilizada para análise de dados, dando a capacidade de o *Python* trabalhar com dados em formato de planilha. O Quadro 3.2, a seguir, apresenta os principais parâmetros de busca para a criação dos *dataSets*.

Quadro 3.2 – Parâmetros de busca de *Tweets*

Parâmetros	Definição
<i>ID</i>	O ID único do tweet.
<i>User</i>	O nome do usuário que postou o tweet
<i>Date</i>	A data e hora em que o tweet foi criado
<i>Text</i>	O texto completo do tweet
<i>Location</i>	A localização geográfica associada ao tweet (se o usuário permitir que essa informação fosse compartilhada)

Fonte: Adaptado de Twitter (2022)

O pseudocódigo com o processo de autenticação no Twitter, bem como a *query* de busca e o processo até salvar os *dataSets* encontra-se no apêndice B.

No total, foram criados vinte conjuntos de dados. Isso inclui um conjunto geral, construído por meio de uma junção entre os conjuntos de dados por operadora, quatro conjuntos de dados, um para cada operadora, contendo opiniões por operadora em território brasileiro, e quinze conjuntos de dados por localidade, sendo cinco para a cidade de São Paulo, cinco para a cidade do Rio de Janeiro e cinco para a cidade de Belo Horizonte.

### 3.1.2 Definições de Classes

Conforme apresentado no capítulo anterior, a metodologia adotada nesta pesquisa para classificação de opiniões é baseada no aprendizado de máquina supervisionado. Esse método requer um conjunto de dados rotulados para o treinamento do modelo, a fim de que ele possa aprender a classificar novas instâncias de forma precisa.

No caso específico desta pesquisa, será utilizada a técnica de classificação multiclasse, que permite classificar instâncias em três ou mais classes distintas. Como explicado por Anchiêta et al. (2021), na classificação multiclasse, cada instância é atribuída a uma classe específica e é associada a um valor que indica a probabilidade dessa classe ser a correta. Nesta pesquisa utilizamos a classificação manual.

Selecionamos 4.233 (quatro mil e duzentos e trinta e três) opiniões e as classificamos manualmente em três categorias: excelente, regular e péssimo. Essas categorias foram definidas com base nos predicados do objeto da “Opinião”, presentes nos *tweets*. O Quadro 3.3 ilustra o modelo de classificação multiclasse adotado neste estudo, com as classes definidas e os respectivos valores associados a cada uma delas.

Quadro 3.3 – Disponibilidade de dados para pesquisa acadêmica - Twitter

Classe	Instância - (opinião)
excelente	Melhor operadora da minha região Aqui o sinal está bom
regular	O serviço está funcionando, mas poderia ser melhor
péssimo	Péssimo serviço, não consigo fazer nada

Fonte: Do Autor (2023)

Para indicar as classes de sentenças de opinião manualmente, é necessário seguir critérios específicos, uma vez que não existe um padrão definido nas expressões usadas pelos usuários. Assim essa pesquisa tomou como critério a identificação de expressões que indique o predicado do objeto na frase que forma a opinião. Em português o predicado do objeto é o elemento da frase que atribui uma característica ao objeto. De forma mais clara, o predicado do objeto é o complemento verbal que se liga diretamente ao objeto direto, podendo ser um adjetivo, substantivo ou locução adjetiva. Ele indica uma qualidade, estado ou atributo do objeto direto, que é o elemento da frase que recebe a ação do verbo. Por exemplo, na frase "Eu achei o filme ótimo", o objeto direto é "o filme", enquanto o predicado do objeto é "ótimo", que expressa a opinião do sujeito sobre o objeto direto. A classificação manual foi realizada por uma equipe de anotadores treinados que seguiram tais critérios pré-definidos para atribuir as categorias às opiniões. Este processo visa garantir a consistência e qualidade da classificação (HICKMAN et al., 2022; IŞIK; DAĞ, 2020).

No Quadro 3.4, apresentam-se exemplos de expressões que podem ser classificadas como predicado do objeto e suas respectivas compreensões, e que traz o entendimento de como as classes de sentenças de opinião foram indicadas.

Quadro 3.4 – Classificação das Expressões

Nível (MOS)	COMPREENSÃO	Exemplos de Expressões
5	Excelente	sinal ótimo, a melhor operadora funcionando, está bom.
3	Regular	poderia ser melhor, pelo menos pega
1	Péssimo	odeio, vou trocar, xingamentos, Sem sinal, falhando, não funciona.

Fonte: Do Autor (2023)

### 3.1.3 Pré-processamento

A qualidade e a integridade dos dados são fatores essenciais que afetam diretamente o desempenho e a eficácia dos modelos de aprendizagem de máquina. Portanto, a fase de pré-processamento

desempenha um papel crucial na garantia da confiabilidade e na maximização do potencial dos modelos de aprendizado. Nesta etapa, o tratamento do conjunto de dados é realizado para preparar os textos de entrada, com o objetivo de padronizá-los para as tarefas subsequentes, eliminando elementos desnecessários para o processo de treinamento, teste e classificação preditiva (ANCHIÊTA et al., 2021).

Esta pesquisa utilizou as bibliotecas:

- a) *spaCy* - É uma biblioteca de código aberto para PLN em *Python*. Ela oferece recursos avançados para processamento de texto, como extração de informações, sistemas de compreensão de linguagem natural e pré-processamento de texto para aprendizado profundo;
- b) *re* - É uma biblioteca utilizada para tratar expressões regulares em *Python*. Ela fornece recursos poderosos para busca e manipulação de padrões em textos, sendo útil para remoção de elementos indesejados, como *links*, *hashtags*, *stopwords* e números (ANCHIÊTA et al., 2021).

Nesta etapa de pré-processamento, foram realizados os seguintes processos:

- a) exclusão de opiniões (*tweets*) duplicados: Durante a busca por opiniões em redes sociais, é comum encontrar sentenças duplicadas, coletadas mais de uma vez. Nessa etapa, essas opiniões duplicadas serão removidas, garantindo uma base de dados mais limpa e evitando viés nos resultados;
- b) remoção de *links*, *hashtags*, *stopwords* e números: Esses elementos não possuem valores semânticos relevantes para o processo de classificação e podem introduzir ruído nos dados. Portanto, serão removidos para simplificar e otimizar o processamento;
- c) conversão de maiúsculo para minúsculo: Para unificar os termos e evitar discrepâncias causadas por letras maiúsculas e minúsculas, todas as letras dos textos serão convertidas para minúsculas. Isso ajudará a obter uma representação consistente dos *tokens* obtidos;
- d) *tokenização*: Nesta etapa, os textos serão divididos em uma lista de *tokens*, ou seja, segmentados em unidades menores, como palavras individuais. A *tokenização* é importante para análises posteriores e permite um processamento mais eficiente dos textos.

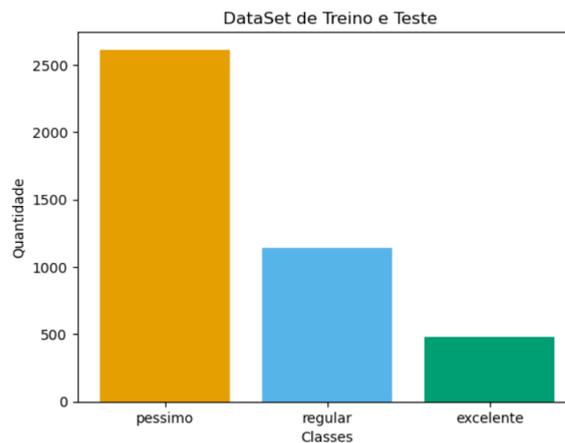
Essa fase de pré-processamento é essencial para preparar os dados de forma adequada, removendo informações irrelevantes, padronizando o formato e criando representações mais adequadas para a etapa de aprendizagem de máquina. O uso das bibliotecas mencionadas permite uma implementação eficiente e robusta do processo de pré-processamento, contribuindo para a obtenção de resultados mais precisos e confiáveis em tarefas de classificação e análise de texto.

### 3.1.4 Divisão do conjunto de dados para treino e teste

Uma etapa importante, para se testar o modelo de classificação é a divisão do conjunto de dados para o treinamento e teste. Essa divisão é feita para avaliar o desempenho do modelo em dados não vistos durante o treinamento (ANCHIÊTA et al., 2021). Dividimos em 70% para treinamento e 30% para teste.

Devido à maioria das opiniões classificadas manualmente serem rotuladas como "péssimas" e terem poucas opiniões classificadas como "excelentes", precisávamos lidar com um desequilíbrio de classes no conjunto de dados. Esse desequilíbrio pode levar a um viés no treinamento do modelo, pois o algoritmo tende a se inclinar para a classe majoritária, resultando em desempenho insatisfatório na classificação das classes minoritárias (BRANCO; TORGO; RIBEIRO, 2017). O Gráfico 3.1 ilustra o desequilíbrio do conjunto de dados.

Gráfico 3.1 – Definição de classes para o processos de treino e testes



Fonte: Do autor (2023)

Assim fez-se necessário realizar o balanceamento do conjunto de dados. O balanceamento do conjunto de dados é uma prática importante no treinamento de modelos de aprendizado de máquina quando as classes alvo estão desproporcionalmente representadas nos dados. Em muitos casos, um conjunto de dados pode conter uma classe majoritária com uma frequência muito maior do que as classes minoritárias. O desequilíbrio nas classes de destino pode levar a problemas de aprendizado enviesados e resultados subótimos. Quando um modelo é treinado em um conjunto de dados desbalanceado, ele tende a favorecer a classe majoritária e ter desempenho inferior nas classes minoritárias (HICKMAN et al., 2022; IŞIK; DAĞ, 2020; BRANCO; TORGO; RIBEIRO, 2017).

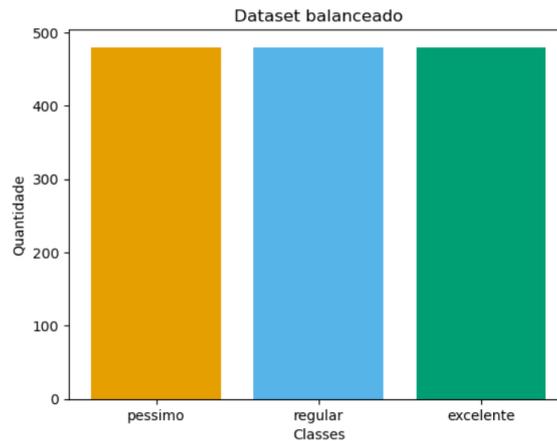
Para mitigar esse problema, é comum realizar o balanceamento do conjunto de dados. Duas abordagens comuns são abordadas neste caso, sendo elas, a subamostragem e a sobreamostragem:

- a) subamostragem: Consiste em reduzir a quantidade de exemplos da classe majoritária para torná-la mais equilibrada em relação às classes minoritárias;

- b) *sobreamostragem*: Envolve a replicação ou geração sintética de exemplos das classes minoritárias para aumentar sua representação no conjunto de dados (HICKMAN et al., 2022; IŞIK; DAĞ, 2020; BRANCO; TORGO; RIBEIRO, 2017).

No desenvolvimento desta pesquisa utilizamos a abordagem de *subamostragem*. O Gráfico 3.2 ilustra como ficou a divisão.

Gráfico 3.2 – Balanceamento de classes



Fonte: Do autor (2023)

Ao realizar o balanceamento do modelo, garantimos que o algoritmo receba informações igualmente representativas de todas as classes, permitindo que ele aprenda padrões relevantes em cada categoria. Isso resulta em um desempenho mais equilibrado e preciso na classificação das opiniões, independentemente da sua classificação original.

### 3.1.5 Modelo de Classificação

O modelo de classificação utilizado nesta pesquisa foi construído com base na arquitetura BERT. Para isso utilizamos as bibliotecas em *Python Tensorflow, Keras e Sklearn*.

- a) *Tensorflow* - Trata-se de um *framework open-source* disponível em *Python* e utilizado no processamento de dados, para o aprendizado de máquina, por meio de um modelo baseado em fluxo de dados. Ele foi desenvolvido em 2015 pelo time da Google responsável por pesquisas na área de IA e *Deep Learning*;
- b) *Keras* - É uma biblioteca para RNA de alto nível, escrita em *Python* e utilizada no processo de criação das RNAs para resolução de várias tarefas, como classificação, detecção de objetos e regressão;
- c) *Sklearn* - É uma biblioteca de aprendizado de máquina em *Python*, que possui vários algoritmos de classificação, regressão e *clustering*.

Exemplificando a arquitetura, o BERT é dividido em duas partes para o processo de classificação de texto, como já mencionado no referencial teórico desta pesquisa. A primeira parte é o processo de *encoder*, onde os textos são *tokenizados* e as palavras são quebradas em subpalavras para produzir diferentes representações do texto de entrada. Nesta etapa, o BERT é treinado em dados não rotulados. No contexto desta pesquisa, são utilizados os BERT pré-treinados na versão BERT Multilíngue e BERTimbau, disponibilizados pelo *huggingface*, pois ambos possuem um amplo conhecimento linguístico adquirido a partir de grandes quantidades de dados textuais e podem ser trabalhados com vocabulários da língua portuguesa brasileira.

Nessa tarefa, o BERT recebe um texto e, de forma aleatória, substitui algumas palavras por “[MASK]” a fim de posteriormente predizê-las. A partir desse treinamento em dados não rotulados, é possível fazer o treinamento em dados rotulados, ou seja, a segunda parte, conhecida como *decoder*. No *decoder*, o BERT será ajustado para alguma tarefa específica, como, por exemplo, tradução automática, sumarização automática ou, como nesta proposta, a previsão de classes para as opiniões coletadas do Twitter (ANCHIÊTA et al., 2021). A Figura 3.3 apresenta o resumo do modelo utilizado por esta pesquisa.

Figura 3.3 – Resumo do modelo de classificação

```
Model: "model"
```

Layer (type)	Output Shape	Param #
input_word_ids (InputLayer)	[(None, 72)]	0
tf_bert_model (TFBertModel)	TFBaseModelOutputWithPoolingAndCrossAttentions(1 ast_hidden_state=(None, 72, 768), pooler_output=(None, 76 8), past_key_values=None, h idden_states=None, atten tions=None, cross_attent ions=None)	108923136
tf.__operators__.getitem (S licingOpLambda)	(None, 768)	0
dropout_37 (Dropout)	(None, 768)	0
dense (Dense)	(None, 256)	196864
dense_1 (Dense)	(None, 3)	771

```
=====  
Total params: 109,120,771  
Trainable params: 109,120,771  
Non-trainable params: 0
```

Fonte: Do Autor (2023)

O modelo é composto por uma camada de entrada para receber os *input\_ids* dos textos, seguida de uma camada que realiza o processamento do modelo BERT sobre os *input\_ids*. O token [CLS], que representa o início da sequência, é extraído da saída do BERT e passa por uma camada de *dropout* para evitar *overfitting*. Em seguida, são adicionadas camadas densas intermediárias para aumentar a capacidade de aprendizado do modelo. Por fim, uma camada densa final com ativação *softmax* é utilizada para produzir as probabilidades de classificação para cada uma das categorias.

### 3.1.6 Métricas de avaliação do modelo

A avaliação de modelos de classificação de texto é de suma importância para determinar o desempenho e a eficácia desses modelos em tarefas específicas, permitindo a comparação entre diferentes abordagens e a seleção do modelo mais adequado para a aplicação em questão (ANCHIÊTA et al., 2021). O método de avaliação, do modelo de classificação, utilizado nesta pesquisa, foi feito a partir da análise das seguintes métricas:

- a) matriz de confusão - Trata-se de uma tabela que permite a visualização do desempenho de um algoritmo de classificação através dos seguintes elementos:
  - TP (*true positive*) - Os verdadeiros positivos, são representados na matriz de confusão pelo conjunto de classe que estamos buscando, e que foi preditos corretamente;
  - FP (*false positive*) - Os falsos positivos, são representados na matriz de confusão pelos conjuntos de classes que foram preditos incorretamente;
  - TN (*true negative*) - Os falsos positivos, são representados na matriz de confusão pelos conjuntos de classes que não estamos buscando prever, mas, foram preditos corretamente;
  - FN (*false negative*) - Os falso negativos, são representados na matriz de confusão pelos conjuntos de classes que não estamos buscando prever, mas, foram preditos incorretamente;
- b) acurácia - Utilizada para indicar a performance geral do modelo, calculando a media de acertos no processo de classificação, pela seguinte formula;

$$Acurcia = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

- c) *precision* - Precisão indica o quanto dentre todas as classificações de classe preditas como positivas estão correta, seu calculo utiliza a seguinte formula;

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

- d) *recall* - Revocação indica o quanto dentre todas as situações de classe preditas como positivas obteve-se o valor esperado, ou seja, quantas estão corretas, seu calculo utiliza a seguinte formula;

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

- e) *f1 score* - É a média harmônica entre *precision* e *recall*, seu calculo utiliza a seguinte formula;

$$f1score = 2 * \frac{precision * recall}{precision + recall} \quad (3.4)$$

Nessa etapa são utilizadas bibliotecas como *Numpy*, *Seaborn* e *Matplotlib*.

### 3.1.7 Predição e Geração de QoE

Esta etapa visa classificar automaticamente as opiniões coletas por operada. Assim, após coletar as opiniões e armazená-las em arquivos (.CSV), as opiniões serão carregadas na memória, para a realização dos processos de pré-processamento, classificação preditiva, contagem dos pontos MOS e, novamente, armazenadas com as predições feitas.

Para a geração do QoE foi considerado os seguintes parâmetros:

- a) Q = total de recorrências das classes;
- b) P = total de pontos.

O cálculo possui a seguinte formula:

$$MOS = \frac{\sum P}{Q} \quad (3.5)$$

Os resultados são apresentados em formato de gráficos e tabelas para cada operadora.

## 3.2 Dados da Anatel

A segunda fase desta pesquisa busca comparar os resultados obtidos com a classificação automática das opiniões coletadas do Twitter com os dados sobre as operadoras Tim, Vivo, Claro e Oi disponibilizados pela Anatel. Para isso, utilizamos o Painel de Informações *Web* e o aplicativo de "Serviço Móvel" da Anatel.

O Painel de Informações é uma ferramenta online fornecida pela agência governamental, que oferece acesso aos dados abertos da base de informações da Anatel. Através desse painel, é possível visualizar gráficos, mapas e indicadores relacionados ao setor de telecomunicações. Neste estudo, analisamos especificamente os dados referentes aos consumidores, como reclamações, *ranking* de acesso aos

serviços das operadoras, pesquisas de satisfação e níveis de atendimento. A Figura 3.3 mostra os menus que são utilizados para consultar as informações sobre os consumidores.

Gráfico 3.3 – Consulta Anatel



Fonte: Brasil (2023)

No menu "Reclamações", é possível consultar informações sobre as solicitações registradas no sistema de suporte ao atendimento aos usuários desde o ano de 2015. Os registros são divididos em quatro tipos: denúncia, reclamação, sugestão e pedido de informação, e é possível filtrá-los por submenus. Esta pesquisa está focada nas reclamações, portanto foram selecionadas, nos filtros, as reclamações, as marcas (ou seja, operadoras), os anos, onde verificamos a proporção de reclamações dos últimos cinco anos (de 2018 a junho de 2023) e por cidade, para comparação por localidade.

No menu "Pesquisa de Satisfação", é possível consultar informações do painel de dados das pesquisas de satisfação e qualidade percebida realizadas pela Anatel nos anos de 2021 e 2022. De acordo com o estudo de Brasil (2023), cerca de 90 mil consumidores em todo o país foram entrevistados. Os entrevistados responderam utilizando uma escala de 0 (zero) a 10 (dez), em que 0 representa a avaliação mínima do atributo e 10 a avaliação máxima.

Esta pesquisa selecionou a média de satisfação em relação à "satisfação geral" dos serviços pré-pagos, pós-pagos e de banda larga. É importante ressaltar que as operadoras possuíam mais serviços do que os selecionados. Utilizamos esses serviços para verificar a correlação entre os serviços prestados e as opiniões manifestadas nas redes sociais. A pesquisa de "satisfação geral" avalia os seguintes quesitos: consideração da intenção de continuidade do cliente, a satisfação geral e a recomendação do serviço. Além disso, a pesquisa também aborda a "satisfação com funcionamento" dos serviços pré-pagos, pós-pagos e de banda larga. A pesquisa de "satisfação com funcionamento" analisa aspectos como a rapidez no envio de fotos/vídeos por aplicativos de mensagens/redes sociais, a rapidez na exibição de conteúdo em sites e redes sociais, a capacidade de assistir filmes sem interrupções ou queda de qualidade, a qualidade das chamadas de vídeo e a experiência de jogar online sem interrupções ou queda de qualidade.

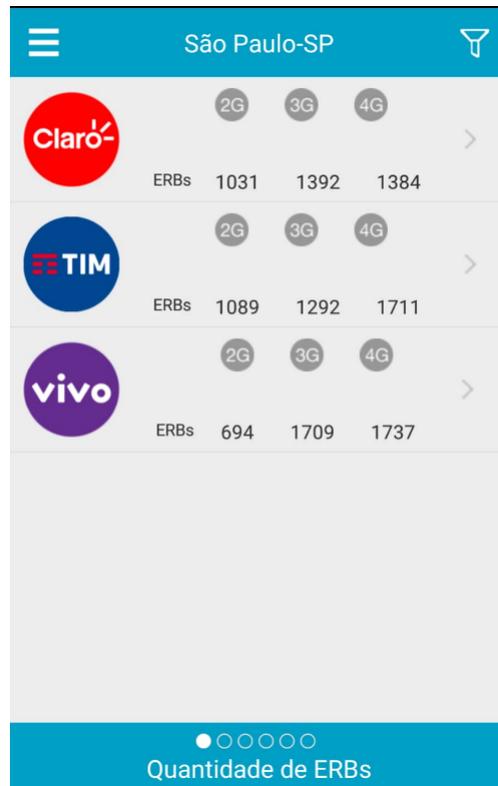
No menu "Indicadores de Qualidade de Atendimento", é possível consultar informações sobre o desempenho das operadoras no atendimento às reclamações registradas por usuários de telecomunicações no sistema Anatel Consumidor. Neste menu, a pesquisa concentra-se principalmente na satisfação do cliente em relação ao atendimento prestado pelas operadoras, apresentando uma média das notas fornecidas pelos consumidores. As notas variam de 1 (muito insatisfeito) a 5 (muito satisfeito), e os atendimentos não avaliados não são considerados no cálculo do índice.

Para a avaliação por localidade da pesquisa de Pesquisa de Satisfação e Indicadores de Qualidade de Atendimento, foram considerados os estados das cidades de São Paulo, Rio de Janeiro e Belo Horizonte, uma vez que não há filtros específicos para cidades.

Para o ranking de acesso a serviços, utilizamos o menu "Acessos". De acordo com Brasil (2023), no painel de informações são publicados mensalmente a quantidade de acessos relativos aos principais serviços de telecomunicações. Os dados são fornecidos à Anatel pelas próprias prestadoras, mostrando a situação dos números de acessos.

Já o aplicativo de "Serviço Móvel" está disponível para uso em *smartphones Android* e iOS. Ele exibe as antenas (estações) de telefonia móvel em cada município, juntamente com as tecnologias utilizadas por cada operadora de redes móveis. Utilizamos essas informações para avaliar a correlação entre os índices de satisfação e as tecnologias empregadas pelas operadoras. Nesta etapa, avaliamos as operadoras Tim, Vivo e Claro, assim como as tecnologias 2G, 3G e 4G, uma vez que o aplicativo não traz informações sobre a operadora Oi e nem sobre a mais recente tecnologia 5G. A Figura 3.4 ilustra o aplicativo.

Gráfico 3.4 – Serviços Moveis Anatel



Fonte: Brasil (2023)

No próximo capítulo, "Resultados e Discussões", apresentamos os resultados obtidos nesta pesquisa e analisamos a correlação entre as fases de predição automática com a coleta de informações realizada da base de informações da Anatel.

## 4 RESULTADOS E DISCUSSÕES

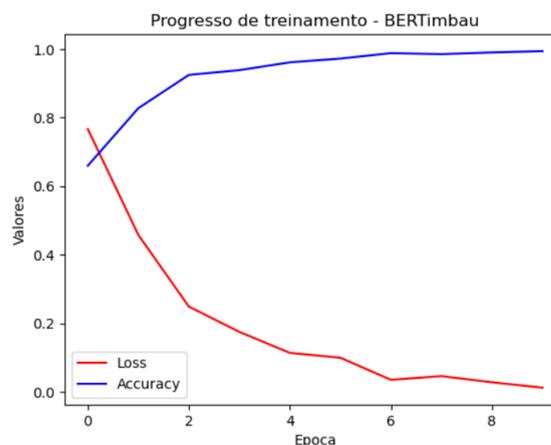
Neste capítulo, são apresentados os resultados do modelo de classificação em testes realizados com o uso do BERT multilíngue e BERTimbau. Além disso, são mostradas as classificações automáticas utilizando o modelo de melhor desempenho, o BERTimbau, e são comparados os resultados com os dados sobre as operadoras disponibilizados pela Anatel.

### 4.1 Avaliação do modelo de classificação

Realizamos testes com os modelos BERT multilíngue e BERTimbau, utilizando uma divisão de 70% dos dados para treinamento e 30% para teste, conforme mencionado no Capítulo 3. Essa divisão permitiu avaliar o desempenho dos modelos e verificar sua capacidade de generalização.

Durante o processo de treinamento, submetemos ambos os modelos a 20 épocas de execução. Após diversos testes, notamos que, ao atingir a época 10, ambos os modelos demonstraram uma estabilidade em seu desempenho. Com base nesses resultados, optamos por treiná-los com 10 épocas cada. Observamos resultados comparáveis entre os modelos, com uma leve vantagem para o modelo BERTimbau. Os gráficos a seguir apresentam a perda (*loss*) e a acurácia durante essa etapa.

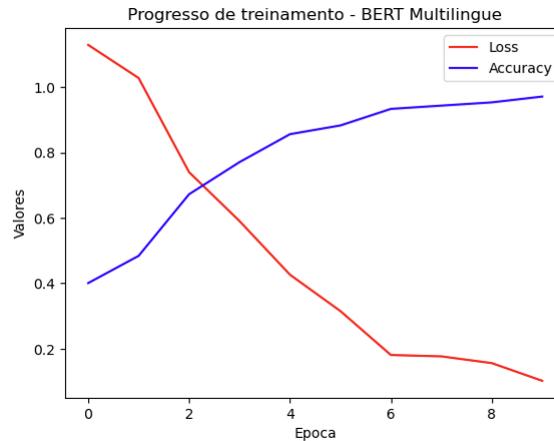
Gráfico 4.1 – Treinamento com BERTimbau



Fonte: Do autor (2023)

No Gráfico 4.1, observa-se que o modelo utilizando o BERTimbau inicia com um valor de perda (*loss*) abaixo de 0.8. Conforme as épocas de treinamento avançam, a perda diminui. Por outro lado, a acurácia, que começa em torno de 0.6 a 0.7, aumenta ao longo do processo de treinamento, alcançando uma acurácia de 0.99 ao final.

Gráfico 4.2 – Treinamento com BERT Multilíngue

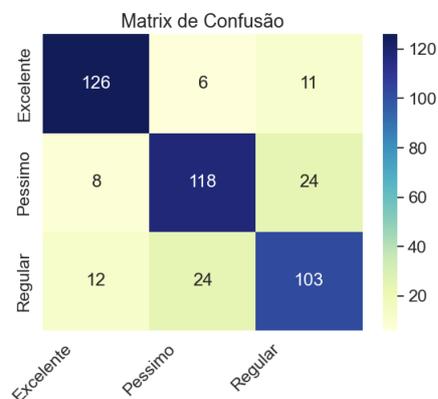


Fonte: Do autor (2023)

No Gráfico 4.2, observa-se que o modelo utilizando o BERT Multilíngue inicia com um valor de perda (*loss*) acima de 1.0, consideravelmente maior do que o modelo com BERTimbau. Conforme as épocas de treinamento avançam, a perda diminui gradualmente. Por outro lado, a acurácia, que começa em torno de 0.4, aumenta ao longo do processo de treinamento, alcançando uma acurácia semelhante à do modelo com o BERTimbau, ou seja, em torno de 0.9.

Na fase de testes, assim como na fase de treinamento, o modelo BERTimbau apresentou um melhor desempenho. Nessa etapa, o modelo recebe os dados que ainda não são conhecidos por ele, a fim de realizar as predições e, posteriormente, permitir a análise do desempenho. Os resultados dessa etapa são avaliados por meio da matriz de confusão, precisão, *F1 score*, *recall* e acurácia. As Figuras 4.1, 4.2, 4.3 e 4.4 a seguir apresentam os resultados obtidos nessa etapa.

Figura 4.1 – Matriz de confusão - BERTimbau



Fonte: Do autor (2023)

A Figura 4.1 apresenta a matriz de confusão do modelo BERTimbau. No eixo vertical da matriz, temos os dados preditos pelo modelo, enquanto no eixo horizontal temos os dados reais, ou seja, as definições atribuídas manualmente. Podemos observar o seguinte:

- a) para a classe "Excelente", o modelo realizou a predição correta de 126 (cento e vinte e seis) opiniões. Além disso, 17 (dezesete) opiniões que pertenciam a essa classe foram preditas como pertencentes a outras classes, e 20 (vinte) opiniões foram preditas como "Excelente", mas na verdade pertenciam a outras classes;
- b) para a classe "Péssimo", o modelo fez a predição correta de 118 (cento e dezoito) opiniões. Houve também 32 (trinta e duas) opiniões que pertenciam a essa classe, mas foram preditas como pertencentes a outras classes, e 30 (trinta) opiniões foram preditas como "Péssimo", mas na verdade pertenciam a outras classes;
- c) para a classe "Regular", o modelo acertou a predição de 103 (cento e três) opiniões. No entanto, 36 (trinta e seis) opiniões que pertenciam a essa classe foram preditas como pertencentes a outras classes, e 35 (trinta e cinco) opiniões foram preditas como "Regular", mas pertenciam a outras classes.

A Figura 4.2 apresenta os demais resultados, incluindo a precisão de cada classe, o *recall*, o *f1-score* e a acurácia obtida nesta etapa.

Figura 4.2 – Precisão, *Recall* e *F1-Score* - BERTimbau

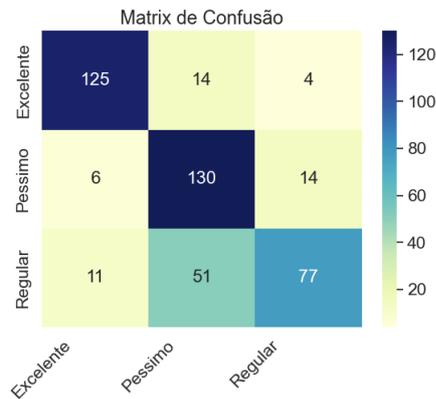
	precision	recall	f1-score	support
Excelente	0.86	0.88	0.87	143
Pessimo	0.80	0.79	0.79	150
Regular	0.75	0.74	0.74	139
accuracy			0.80	432
macro avg	0.80	0.80	0.80	432
weighted avg	0.80	0.80	0.80	432

Fonte: Do autor (2023)

O modelo obteve uma acurácia final de 0.8, o que indica que, ao ser avaliado em dados não vistos durante o treinamento, ele foi capaz de realizar classificações corretas em 80% dos casos. Esses resultados são semelhantes aos obtidos por Yang et al. (2021) e Khan et al. (2022), que também utilizaram modelos BERT para classificação de opiniões em pesquisas semelhantes. Isso sugere que o modelo proposto neste estudo demonstrou um desempenho satisfatório e a capacidade de generalizar e aplicar o conhecimento adquirido durante o treinamento em novos dados.

Os resultados utilizando o modelo com o BERT Multilíngue apresentaram um desempenho um pouco inferior em relação ao BERTimbau. As Figuras 4.3 e 4.4 a seguir apresentam os resultados detalhados.

Figura 4.3 – Matriz de confusão - BERT Multilíngue



Fonte: Do autor (2023)

A Figura 4.3 apresenta a matriz de confusão com o modelo BERT Multilíngue. É possível observar que:

- para a classe "Excelente", o modelo realizou a predição correta de 125 (cento e vinte e cinco) opiniões. No entanto, 18 (dezoito) opiniões que pertenciam a essa classe foram preditas como outras classes, e 17 (dezessete) opiniões foram preditas como "Excelente", mas pertenciam a outras classes;
- para a classe "Péssimo", 130 (cento e trinta) opiniões foram preditas corretamente. No entanto, 20 (vinte) opiniões que pertenciam a essa classe foram preditas como outras classes, e 65 (sessenta e cinco) opiniões foram preditas como "Péssimo", mas pertenciam a outras classes;
- para a classe "Regular", 77 (setenta e sete) opiniões foram preditas corretamente. No entanto, 62 (sessenta e duas) opiniões que pertenciam a essa classe foram preditas como outras classes, e 18 (dezoito) opiniões foram preditas como "Regular", mas pertenciam a outras classes.

A Figura 4.4 apresenta os demais resultados, incluindo a precisão de cada classe, o *recall*, o *f1-score* e a acurácia obtidos pelo modelo utilizando o BERT Multilíngue.

Figura 4.4 – Precisão, *Recall* e *F1-Score* - BERT Multilíngue

	precision	recall	f1-score	support
Excelente	0.88	0.87	0.88	143
Pessimo	0.67	0.87	0.75	150
Regular	0.81	0.55	0.66	139
accuracy			0.77	432
macro avg	0.79	0.76	0.76	432
weighted avg	0.78	0.77	0.76	432

Fonte: Do autor (2023)

Utilizando o modelo com o BERT Multilíngue, o desempenho nos testes obteve uma acurácia de 0.77. A diferença em relação ao BERTimbau ocorreu principalmente devido ao baixo desempenho na predição da classe "Regular", como mostrado nas Figuras 4.3 e 4.4.

Devido aos resultados melhores obtidos com o BERTimbau, as predições dos conjuntos de dados das próximas etapas foram feitas exclusivamente com este modelo.

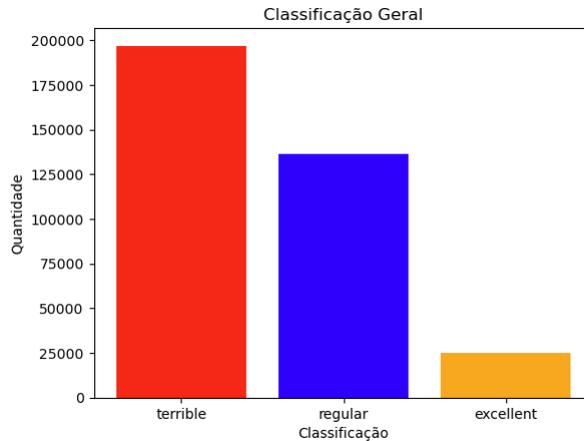
## 4.2 Classificação automática

Após aplicarmos técnicas de classificação e realizar o treinamento e teste do modelo de rede neural com o vocabulário do BERTimbau, projetado especificamente para a língua portuguesa, procedemos ao pré-processamento dos conjuntos de dados. Esses conjuntos incluíam um conjunto de dados contendo todas as opiniões de todas as operadoras que atuam no Brasil, quatro conjuntos de dados coletados individualmente para cada operadora e quinze conjuntos de dados distribuídos por localidade. Esses conjuntos de dados foram coletados no período de novembro de 2022 a maio de 2023, com coletas semanais das opiniões.

### 4.2.1 Classificação geral

Inicialmente, realizamos a predição automática utilizando opiniões de todas as operadoras de redes móveis em um único conjunto de dados, que contava com 358.408 (trezentos e cinquenta e oito mil, quatrocentos e oito) registros. Os resultados são apresentados no Gráfico 4.3 e na Tabela 4.1.

Gráfico 4.3 – Classificação Geral



Fonte: Do autor (2023)

Pelo Gráfico 4.3, pode-se observar que há uma tendência dos usuários de redes móveis brasileiras utilizarem as redes sociais, em especial, para reclamar ou fazer indicações de que o serviço não está satisfatório. Isso ocorre porque, neste conjunto de dados, a classe "Péssimo" obteve 197.022 (cento e noventa e sete mil e vinte e dois) registros, enquanto a classe "Regular" obteve 136.233 (cento e trinta e seis mil e duzentos e trinta e três) registros e apenas 25.153 (vinte e cinco mil e cento e cinquenta e três) registros indicaram satisfação, ou seja, a classe "Excelente". A geração do MOS é apresentada na Tabela 4.1.

Tabela 4.1 – Avaliação Geral

<b>Classes</b>	<b>Predição Total</b>	<b>Score MOS</b>	<b>Total de pontuação</b>
excelente	25153	5	125765
péssimo	197022	1	197022
regular	136233	3	408699
<b>MOS</b>		<b>2,04</b>	

Fonte: Do autor (2023)

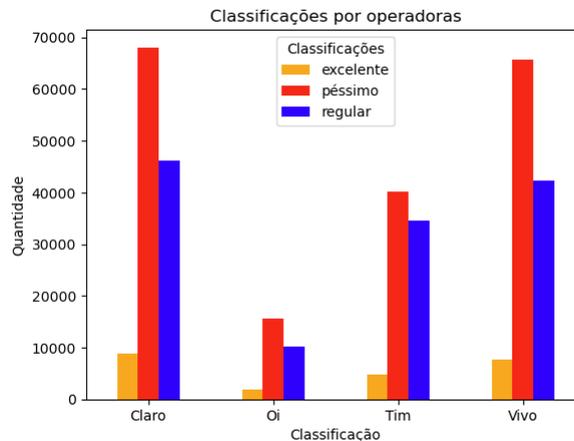
O MOS geral referente à satisfação geral dos usuários de redes móveis ficou em 2,04 (dois vírgula zero quatro).

#### 4.2.2 Classificação por operadora

Na classificação por operadora considerando todo o território nacional, foram previstas 27.739 (vinte e sete mil setecentos e trinta e nove) opiniões para a operadora Oi, 79.603 (setenta e nove mil seiscentos e três) para a operadora TIM, 115.696 (cento e quinze mil seiscentos e noventa e seis) para a operadora Vivo e 123.131 (cento e vinte e três mil cento e trinta e um) para a operadora Claro. Assim,

é possível observar que os usuários da rede móvel Claro utilizam com mais frequência as redes sociais para expressar suas opiniões. O Gráfico 4.4 e a Tabela 4.2 apresentam a classificação nesta etapa.

Gráfico 4.4 – Classificação por operadora



Fonte: Do autor (2023)

Assim como na previsão geral, a previsão das operadoras aponta para uma tendência entre os usuários da rede móvel brasileira de utilizar as redes sociais, principalmente para reclamações ou recomendações quando estão insatisfeitos com os serviços. Nos conjuntos de dados, a classe "péssima" teve o maior número de registros para a Oi, com 15.663 (quinze mil seiscentos e sessenta e três), para a Tim, com 40.151 (quarenta mil cento e cinquenta e um), para a Vivo, com 65.690 (sessenta e cinco mil seiscentos e noventa), e para a Claro, com 68.055 (sessenta e oito mil e cinquenta e cinco). Enquanto isso, a classe "regular" teve 10.241 (dez mil duzentos e quarenta e um) registros para a Oi, 34.599 (trinta e quatro mil quinhentos e noventa e nove) para a Tim, 42.290 (quarenta e dois mil duzentos e noventa) para a Vivo e 46.230 (quarenta e seis mil duzentos e trinta) para a Claro. Já a classe "excelente" teve 1.835 (um mil oitocentos e trinta e cinco) registros para a Oi, 4.853 (quatro mil oitocentos e cinquenta e três) para a Tim, 7.716 (sete mil setecentos e dezesseis) para a Vivo e 8.846 (oito mil oitocentos e quarenta e seis) para a Claro. A Tabela 4.2, abaixo apresenta a pontuação e MOS de satisfação por operadora.

Tabela 4.2 – Classificação por Operadora

<b>Operadora</b>	<b>Classes</b>	<b>Predição total</b>	<b>Score MOS</b>	<b>Total de pontuação</b>
Oi	excelente	1835	5	9175
	péssimo	15663	1	15663
	regular	10241	3	30723
<b>MOS</b>			<b>2,04</b>	
Tim	excelente	4853	5	24265
	péssimo	40151	1	40151
	regular	34599	3	103797
<b>MOS</b>			<b>2,11</b>	
Vivo	excelente	7716	5	38580
	péssimo	65690	1	65690
	regular	42290	3	126870
<b>MOS</b>			<b>1,99</b>	
Claro	excelente	8846	5	44230
	péssimo	68055	1	68055
	regular	8846	3	138690
<b>MOS</b>			<b>2,03</b>	

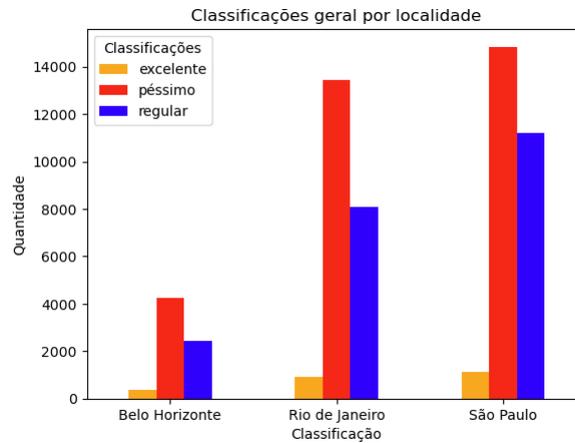
Fonte: Do autor (2023)

A operadora Oi obteve um MOS de 2,04 (dois vírgula zero quatro), a Tim obteve um MOS de 2,11 (dois vírgula onze), a Vivo obteve um MOS de 1,99 (um vírgula noventa e nove) e a Claro obteve um MOS de 2,03 (dois vírgula zero três). Assim, observa-se, que há uma paridade em termos de satisfação do cliente, com uma pequena diferença entre aquela que obteve a maior média (Tim) e aquela que obteve a menor média (Vivo).

#### 4.2.3 Classificação por localidade

Os dados previstos automaticamente por localidades, ou seja, por São Paulo, Rio de Janeiro e Belo Horizonte, apresentaram um menor número de registros. Isso ocorre devido à exigência de que os usuários permitam que seu "tweet" ou "opinião" exiba dados de localização. O Gráfico 4.5 e a Tabela 4.3 apresentam os resultados gerais.

Gráfico 4.5 – Classificação geral por localidade



Fonte: Do autor (2023)

Como visto na classificação geral, a classificação por localidade segue a mesma tendência de os usuários utilizarem as redes sociais, especialmente para reclamar ou indicar sua insatisfação com os serviços. A Tabela 4.3 apresenta o MOS por localidade.

Tabela 4.3 – Classificação por localidade

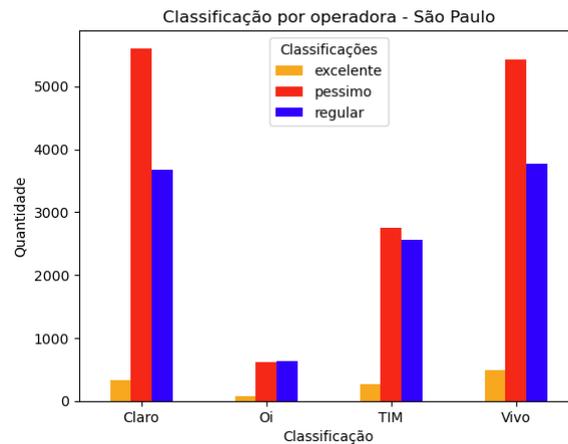
Localidade	Classes	Predição total	Score MOS	Total de pontuação
São Paulo	excelente	1138	5	5690
	péssimo	14845	1	14842
	regular	11227	3	33681
MOS			1,80	
Rio de Janeiro	excelente	937	5	4685
	péssimo	13460	1	13460
	regular	8090	3	24270
MOS			1,88	
Belo Horizonte	excelente	364	5	1820
	péssimo	4243	1	4243
	regular	2452	3	7356
MOS			1,90	

Fonte: Do autor (2023)

Pelos resultados apresentados na Tabela 4.3, observa-se que os usuários de São Paulo estão mais insatisfeitos com os serviços prestados pelas operadoras, apresentando um MOS de 1,80 (um vírgula oitenta). No entanto, as demais localidades, como o Rio de Janeiro com MOS de 1,88 (um vírgula oitenta e oito) e Belo Horizonte com MOS de 1,90 (um vírgula noventa), também apresentaram valores abaixo do resultado geral apresentado na seção 4.2.1.

Na avaliação por operadora para São Paulo, foram processados 1.329 (um mil trezentos e vinte e nove) registros para a operadora Oi, 5.563 (cinco mil quinhentos e sessenta e três) registros para a Tim, 9.677 (nove mil seiscentos e setenta e sete) registros para a Vivo e 9.605 (nove mil seiscentos e cinco) registros para a Claro. O gráfico 4.6 e a tabela 4.4 apresentam os resultados.

Gráfico 4.6 – Classificação por operadora - São Paulo



Fonte: Do autor (2023)

A tendência já observada se repete em São Paulo, ou seja, as reclamações e insatisfações com o serviço superam os elogios. A Tabela 4.4 apresenta o MOS por operadora em São Paulo.

Tabela 4.4 – Classificação por localidade - São Paulo

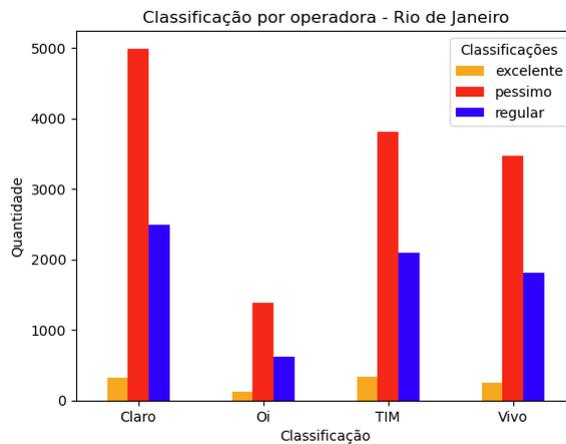
Operadora	Classes	Predição Total	Score MOS	Total de Pontuação
Oi	excelente	80	5	400
	péssimo	618	1	618
	regular	631	3	1893
MOS			2,19	
Tim	excelente	263	5	1315
	péssimo	2744	1	2744
	regular	2556	3	7668
MOS			2,10	
Vivo	excelente	7716	5	38580
	péssimo	65690	1	65690
	regular	42290	3	126870
MOS			1,97	
Claro	excelente	332	5	1660
	péssimo	5606	1	5606
	regular	3667	3	11001
MOS			1,90	

Fonte: Do autor (2023)

O MOS para a operadora Oi ficou em 2,19 (dois vírgula dezenove), a Tim obteve um MOS de 2,10 (dois vírgula dez), a Vivo obteve um MOS de 1,97 (um vírgula noventa e sete) e a Claro obteve um MOS de 1,9 (um vírgula nove). No contexto de São Paulo, a operadora Oi obteve pontuações mais altas em relação às demais.

Para o processamento preditivo do Rio de Janeiro, foram processados 2.124 (dois mil cento e vinte e quatro) registros para a operadora Oi, 6.229 (seis mil duzentos e vinte e nove) para a Tim, 5.536 (cinco mil quinhentos e trinta e seis) para a Vivo e 7.803 (sete mil oitocentos e três) para a Claro. O Gráfico 4.7 e a Tabela 4.5 apresentam os resultados.

Gráfico 4.7 – Classificação por operadora - Rio de Janeiro



Fonte: Do autor (2023)

Da mesma forma, nos demais testes realizados, persiste a tendência de insatisfação. A Tabela 4.5 apresenta o MOS por operadora para o Rio de Janeiro.

Tabela 4.5 – Classificação por localidade - Rio de Janeiro

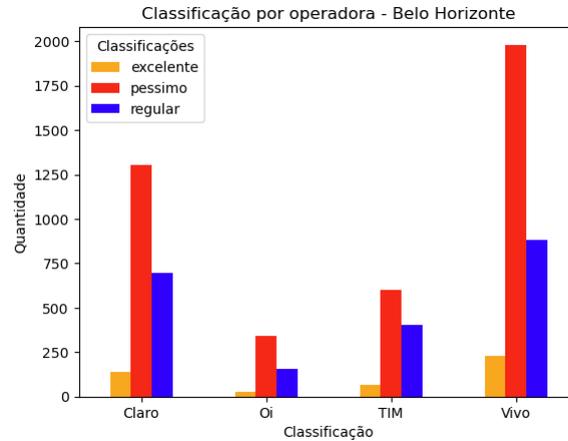
<b>Classes</b>	<b>Classes</b>	<b>Predição Total</b>	<b>Score MOS</b>	<b>Total de Pontuação</b>
Oi	excelente	117	5	585
	péssimo	1381	1	1381
	regular	629	3	1887
<b>MOS</b>			<b>1,81</b>	
Tim	excelente	334	5	1670
	péssimo	3807	1	3807
	regular	2088	3	10440
<b>MOS</b>			<b>2,55</b>	
Vivo	excelente	251	5	1255
	péssimo	3474	1	3474
	regular	1811	3	5433
<b>MOS</b>			<b>1,83</b>	
Claro	excelente	324	5	1620
	péssimo	4992	1	4992
	regular	2487	3	7461
<b>MOS</b>			<b>1,80</b>	

Fonte: Do autor (2023)

A operadora Oi obteve um MOS de 1,81 (um vírgula oitenta e um), a Tim obteve um MOS de 2,55 (dois vírgula cinquenta e cinco), a Vivo obteve um MOS de 1,83 (um vírgula oitenta e três) e a Claro obteve um MOS de 1,80 (um vírgula oitenta). No contexto do Rio de Janeiro, a operadora TIM obteve pontuações mais altas em relação às demais.

Para o processamento preditivo de Belo Horizonte, foram processados 520 (quinhentos e vinte) registros para a operadora Oi, 1073 (um mil e setenta e três) para a Tim, 3093 (três mil e noventa e três) para a Vivo e 2140 (dois mil cento e quarenta) para a Claro. O Gráfico 4.8 e a Tabela 4.6 apresentam os resultados.

Gráfico 4.8 – Classificação por operadora - Belo Horizonte



Fonte: Do autor (2023)

Assim como em São Paulo e Rio de Janeiro, a insatisfação também é a maioria das opiniões em Belo Horizonte. A Tabela 4.6 apresenta o MOS por operadora para Belo Horizonte.

Tabela 4.6 – Classificação por localidade - Belo Horizonte

Operadoras	Classes	Predição Total	Score MOS	Total de Pontuação
Oi	excelente	26	5	130
	péssimo	339	1	339
	regular	155	3	465
MOS			1.79	
Tim	excelente	68	5	340
	péssimo	603	1	603
	regular	402	3	1206
MOS			2.0	
Vivo	excelente	229	5	1145
	péssimo	1980	1	1980
	regular	884	3	2652
MOS			1.86	
Claro	excelente	141	5	705
	péssimo	1303	1	1303
	regular	696	3	2088
MOS			1.91	

Fonte: Do autor (2023)

A operadora Oi obteve um MOS de 1,79 (um vírgula setenta e nove), a Tim obteve um MOS de 2,0 (dois vírgula zero), a Vivo obteve um MOS de 1,86 (um vírgula oitenta e seis) e a Claro obteve um MOS de 1,91 (um vírgula noventa e um). No contexto de Belo Horizonte, a operadora Tim foi a melhor em relação às demais.

Comparando todos os resultados, observa-se que, assim como na etapa por operadoras, a operadora Tim obteve melhor desempenho, ficando atrás da operadora Oi apenas na cidade de São Paulo. E mesmo em São Paulo, a Tim ficou à frente das operadoras Vivo e Claro. Nas demais localidades, Rio de Janeiro e Belo Horizonte, a operadora Tim ficou à frente de todas as outras.

### 4.3 Dados Anatel

Usando o painel de informações da Anatel, analisamos os registros de reclamações dos últimos cinco anos, ou seja, de 2018 a 2023, por operadora em todo o território nacional, bem como nas localidades de São Paulo, Rio de Janeiro e Belo Horizonte. Além disso, examinamos a cobertura das tecnologias utilizadas por cada uma delas. Também levamos em consideração as últimas pesquisas de satisfação realizadas nos anos de 2021 e 2022, bem como a satisfação em relação ao atendimento e o *ranking* de acesso por operadora.

### 4.4 Dados Anatel - Geral e por operadora

Inicialmente, apresentamos a média de reclamações das operadoras Oi, Tim, Vivo e Claro, bem como a área em que essas reclamações estão mais concentradas. O Gráfico 4.9 e a Figura 4.5 apresentam esses dados.

Gráfico 4.9 – Média de reclamações recebidas na Anatel



Fonte: Brasil (2023)

No gráfico acima, é possível observar que a Anatel recebe de 100.000 (cem mil) a 150.000 (cento e cinquenta mil) reclamações por mês em relação às operadoras, com uma concentração destacada na região Sudeste do país, como mostrado no mapa da Figura 4.5.

Figura 4.5 – Mapa panorâmico das reclamações



Fonte: Brasil (2023)

No mapa, é possível observar que a região Sudeste, onde estão localizadas São Paulo, Rio de Janeiro e Belo Horizonte, é a área com o maior número de registros de reclamações, com destaque para São Paulo. A Figura 4.6 apresenta o total de reclamações por operadora em todo o território nacional até maio de 2023.

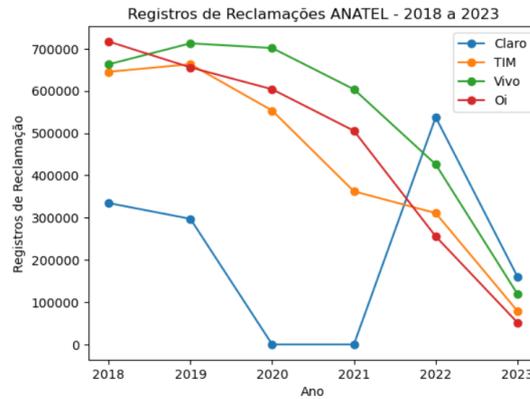
Figura 4.6 – Reclamações por operadora

	⊕ 2018	⊕ 2019	⊕ 2020	⊕ 2021	⊕ 2022	⊕ 2023
⊕ CLARO	334.496	297.136	-	-	537.468	160.230
⊕ OI	716.782	655.418	603.755	505.494	255.593	51.649
⊕ TIM	645.021	663.043	553.358	362.224	310.744	78.963
⊕ VIVO	662.782	712.425	701.324	603.683	426.069	119.205

Fonte: Brasil (2023)

No Gráfico 4.10, é apresentada uma análise mais detalhada da progressão das reclamações por operadoras ao longo do tempo. Essa visualização permite uma melhor compreensão da evolução das queixas e do desempenho de cada operadora ao lidar com as demandas dos usuários.

Gráfico 4.10 – Reclamações recebidas por operadora - ANATEL

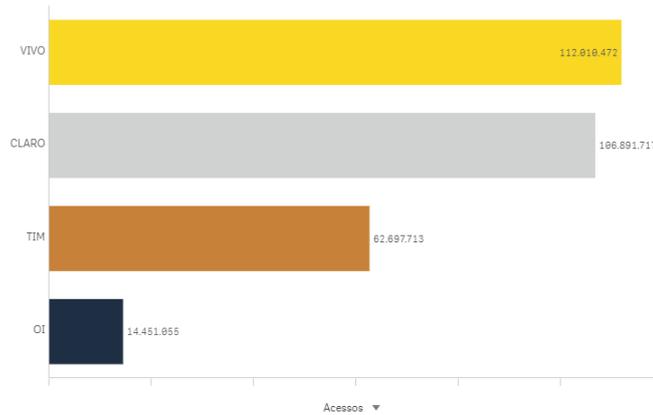


Fonte: Do Autor (2023)

Pela Figura 4.6 e pelo Gráfico 4.10, observa-se que nos últimos anos as reclamações em relação aos serviços prestados pelas operadoras têm apresentado uma queda considerável. Especificamente nos últimos dois anos, 2022 e 2023, período em que esta pesquisa iniciou a coleta de dados do Twitter, nota-se que a operadora Claro registrou o maior número de reclamações, o que está em consonância com os resultados desta pesquisa, na qual a operadora Claro também foi a mais mencionada pelos usuários.

Outro ponto relevante é que a operadora Oi apresentou a maior redução no número de reclamações. No entanto, é importante levar em consideração que a operadora Oi passou recentemente por processos de incorporação em diferentes partes pelas outras três operadoras. Além disso, é interessante destacar a correlação entre as reclamações recebidas pela Anatel e o desempenho da operadora Tim. A Tim obteve o melhor MOS entre as operadoras e registrou o menor número de reclamações.

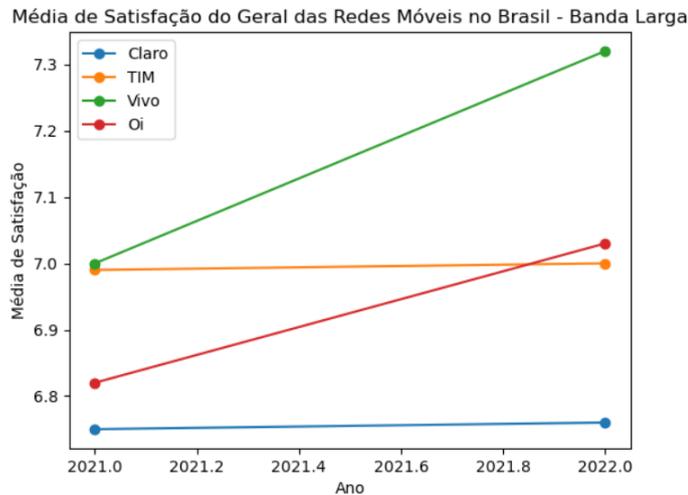
Ao comparar as reclamações com a utilização dos serviços de telecomunicação, nota-se uma diferença interessante: embora a Claro seja a operadora com o maior número de reclamações, a Vivo a supera em termos de acessos aos serviços, com uma diferença mínima. Isso sugere que, por terem maiores números de acessos, essas operadoras também possuem maiores números de reclamações e menções nas redes sociais. Em seguida, a operadora Tim e o Oi vêm nessa sequência, atrás da operadora Claro. Essa análise ressalta a importância de entender a relação entre o desempenho do serviço, a satisfação do cliente e o engajamento nas redes sociais para aprimorar a qualidade e a eficiência dos serviços prestados pelas operadoras. O Gráfico 4.11 apresenta o *ranking* de acessos por operadora.

Gráfico 4.11 – *Ranking* de acessos aos serviços de telecomunicação por operadora

Fonte: Brasil (2023)

Já na pesquisa de satisfação, analisamos a satisfação geral e a satisfação de funcionamento para banda larga, telefonia móvel pós-paga e pré-paga das operadoras. Os dados são referentes aos anos de 2021 e 2022. Os gráficos a seguir apresentam os resultados.

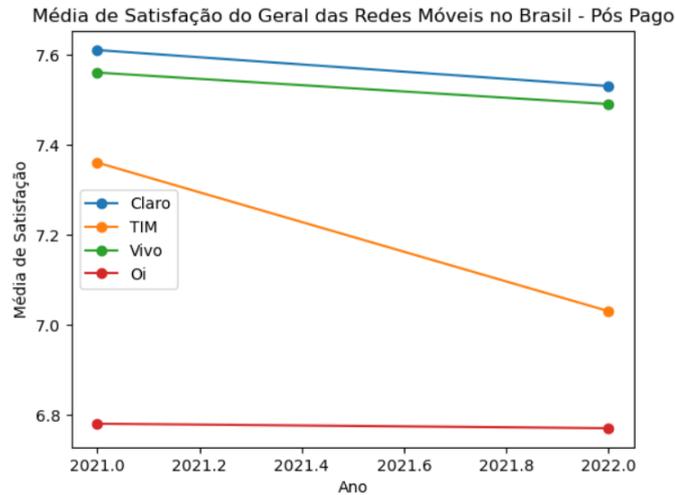
Gráfico 4.12 – Pesquisa de satisfação geral - Banda Larga



Fonte: Do Autor (2023)

Pelo Gráfico 4.12 apresentado acima, é possível observar que a operadora Vivo possui a melhor satisfação em relação aos serviços de banda larga, enquanto a operadora Claro possui a menor satisfação.

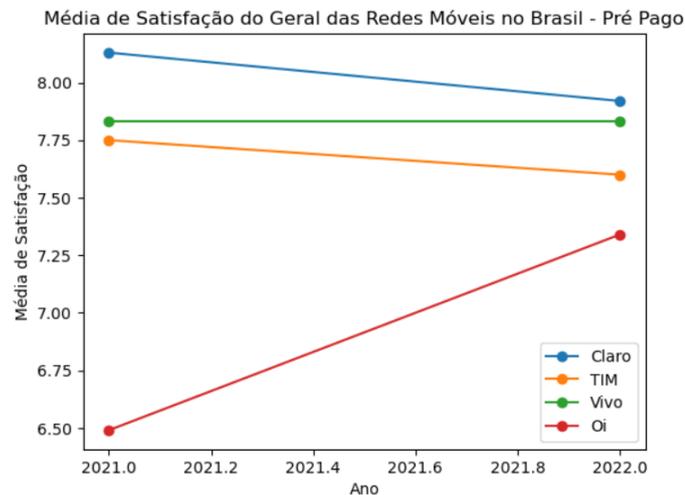
Gráfico 4.13 – Pesquisa de satisfação geral- Pós Pago



Fonte: Do Autor (2023)

Quando se trata de serviços pré-pagos, observe-se que a operadora Claro possui o melhor índice de satisfação, como visto no Gráfico 4.13, enquanto a operadora Oi apresenta o menor índice nesse quesito.

Gráfico 4.14 – Pesquisa de satisfação geral - Pré pago



Fonte: Do Autor (2023)

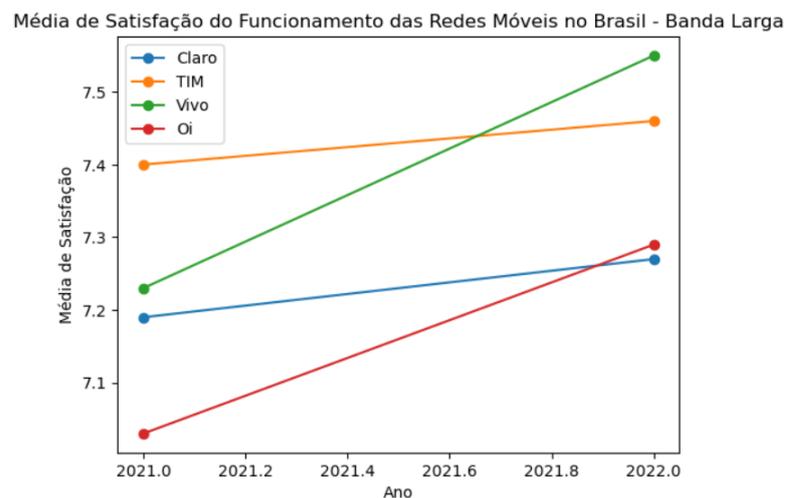
Assim como nos serviços pós-pagos, no pré-pago também é possível observar, pelo Gráfico 4.14, que a operadora Claro possui a melhor satisfação, enquanto a operadora Oi mantém o menor índice.

Pelos gráficos apresentados acima, é possível observar que a operadora Vivo se destaca no serviço de banda larga. No entanto, quando se trata de serviços pós-pagos e pré-pagos, fica atrás da operadora Claro, que, por sua vez, possui a menor média em relação ao serviço de banda larga. Por outro

lado, a operadora TIM, que obteve o melhor desempenho na classificação automatizada, apresenta uma satisfação inferior em relação à Vivo e Claro nos serviços de pós-pagos e pré-pagos, superando apenas a operadora Oi nesse quesito. É interessante notar que há uma variação de um ano para o outro e uma correlação entre os serviços e a satisfação dos usuários, demonstrando a importância de se avaliar continuamente a qualidade dos serviços oferecidos pelas operadoras.

Quanto à "satisfação com o funcionamento", o cenário se mantém no ano de 2022. A operadora Vivo possui o melhor índice de satisfação para o serviço de banda larga, enquanto a operadora Claro obtém melhores índices em relação às outras operadoras nos serviços de pós-pago e pré-pago. Esses dados são apresentados nos gráficos a seguir.

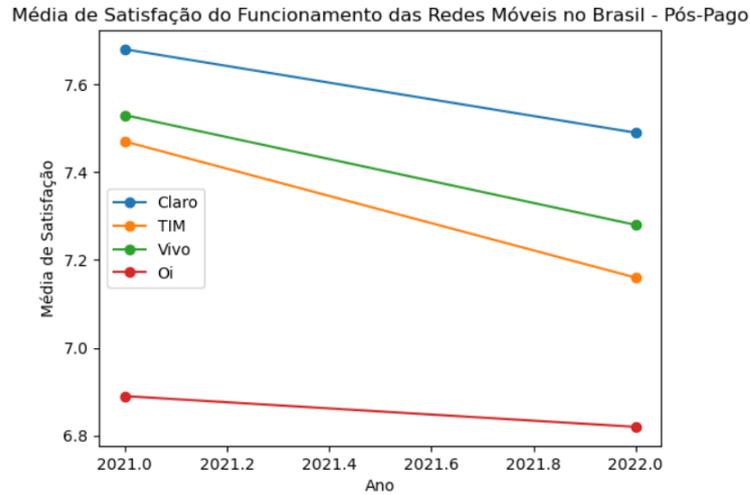
Gráfico 4.15 – Pesquisa de satisfação com funcionamento - Banda Larga



Fonte: Do Autor (2023)

Diferentemente do que ocorreu na satisfação geral, no ano de 2021, a operadora Tim obteve o melhor desempenho no serviço de banda larga. Isso reforça a ideia de variação na satisfação ao longo dos anos, uma vez que, como mencionado anteriormente, no ano de 2022 a Vivo possui o melhor índice.

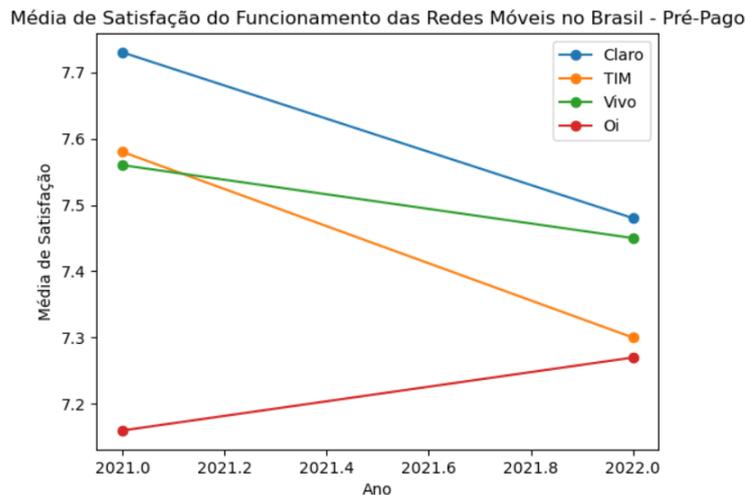
Gráfico 4.16 – Pesquisa de satisfação com funcionamento - Pós pago



Fonte: Do Autor (2023)

Como já observado na satisfação geral, a operadora Claro também possui o melhor índice na satisfação com o funcionamento, no que diz respeito aos serviços pós-pagos e a operadora Oi permanece o menor índice nesse quesito. O mesmo ocorre com o pré-pago como aprestado no Gráfico 4.17.

Gráfico 4.17 – Pesquisa de satisfação com funcionamento - Pré pago



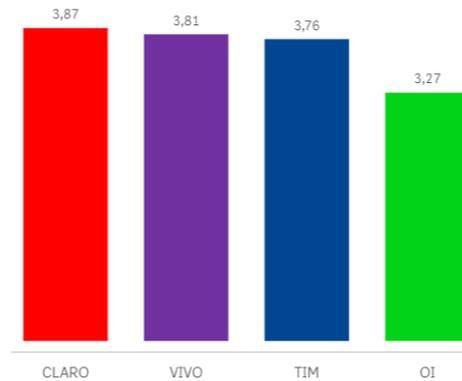
Fonte: Do Autor (2023)

Nos gráficos de "satisfação por funcionalidades", é possível observar que a operadora Vivo se destaca no ano de 2022 quanto aos serviços de banda larga, assim como foi nos gráficos de "satisfação geral". No entanto, assim como na "satisfação geral", ainda fica atrás da operadora Claro, no que diz respeito aos serviços pós-pagos e pré-pagos. Por outro lado, a operadora Claro possui uma satisfação melhor em relação ao serviço de banda larga, se comparado com o resultado apresentado na "satisfação geral", mas ainda está atrás da operadora Vivo e Tim. Já a operadora Oi apresenta o menor índice na

"satisfação por funcionalidade", ficando apenas um pouco acima da operadora Claro no serviço de banda larga no ano 2022.

Por fim, analisamos os indicadores de qualidade de atendimento das operadoras, o Gráfico 4.18 abaixo apresenta o resultados.

Gráfico 4.18 – Qualidade de atendimento das operadoras



Fonte: Brasil (2023)

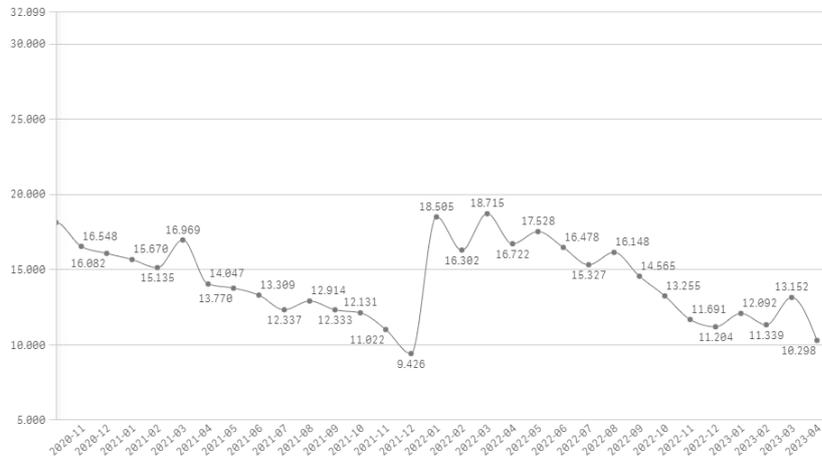
Pelo Gráfico 4,18, é possível observar que a operadora Claro possui o melhor desempenho no quesito atendimento, com uma pequena diferença em relação às operadoras Vivo e Tim. Já a operadora Oi apresenta o menor índice nesse aspecto. Esses dados sugerem que a avaliação do atendimento não está diretamente correlacionada à insatisfação gerada nas redes sociais ou ao número de reclamações recebidas pela Anatel, pois a operadora Claro, embora seja a mais mencionada em reclamações na Anatel, possui um MOS que supera apenas a operadora Vivo, na classificação por operadora. Por outro lado, isso pode indicar uma correlação com a satisfação geral, uma vez que a Claro obteve um melhor índice nos serviços de pós-pago e pré-pago, enquanto a Vivo, que é a segunda melhor em atendimento, obteve bons índices no serviço de banda larga.

#### 4.4.1 Dados da Anatel por localidade

Nesta etapa, buscamos analisar as médias de reclamações recebidas pela Anatel sobre as operadoras Oi, Tim, Vivo e Claro nas cidades de São Paulo, Rio de Janeiro e Belo Horizonte. Além disso, verificamos o *ranking* de acessos e também a correlação entre as reclamações e as tecnologias utilizadas por cada operadora. Os gráficos a seguir apresentam os dados coletados.

Essa análise nos permite compreender melhor o panorama das reclamações nessas localidades específicas e identificar possíveis tendências de cada operadora.

Gráfico 4.19 – Média de reclamações recebidas na Anatel (São Paulo)



Fonte: Brasil (2023)

O Gráfico 4.19 apresenta a média de reclamações mensais recebidas pela Anatel provenientes de São Paulo. Ao analisar o gráfico, é possível observar que a Anatel recebe uma média de reclamações mensais, que varia entre 10.000 (dez mil) e 20.000 (vinte mil). A Figura 4.7 apresenta os dados por operadora.

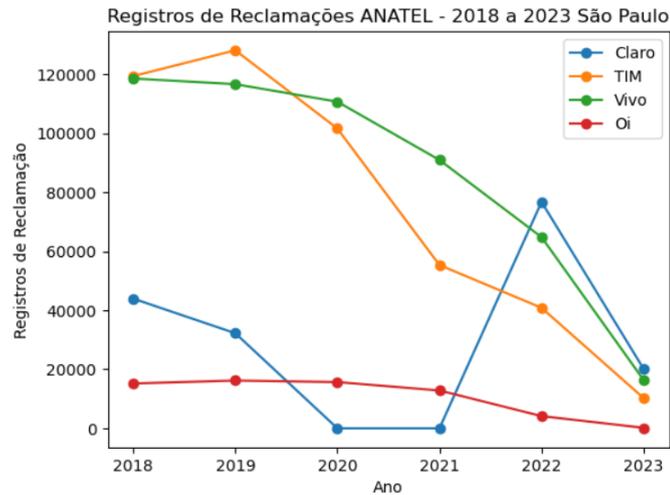
Figura 4.7 – Reclamações por operadora (São Paulo)

	➕ 2018	➕ 2019	➕ 2020	➕ 2021	➕ 2022	➕ 2023
➕ CLARO	43.945	39.223	-	-	76.702	20.132
➕ OI	15.185	16.186	15.663	12.812	4.132	158
➕ TIM	119.348	128.090	101.585	55.324	40.832	10.231
➕ VIVO	118.537	116.608	110.671	90.927	64.774	16.360

Fonte: Brasil (2023)

O Gráfico 4.20 ilustra a evolução das reclamações recebidas pela Anatel provenientes de São Paulo nos últimos cinco anos.

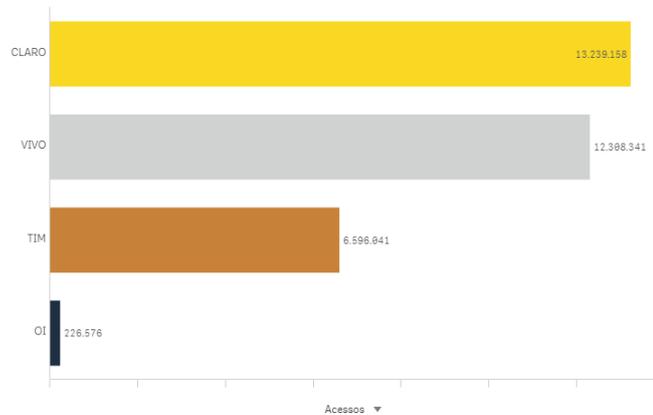
Gráfico 4.20 – Reclamações recebidas por operadora - Anatel (São Paulo)



Fonte: Do Autor (2023)

Assim como já mencionado nas reclamações recebidas pela Anatel em âmbito nacional, o Gráfico 4.20 das reclamações registradas em São Paulo demonstra uma queda no número de registros nos últimos anos. Ao analisarmos os últimos dois anos, 2022 e 2023, período em que a coleta de opiniões no Twitter foi realizada, observa-se que as operadoras Claro e Vivo obtiveram o maior número de reclamações. Por outro lado, a operadora Oi, que apresentou o melhor índice MOS para São Paulo, mantém uma regularidade nas reclamações, ficando abaixo das demais operadoras. Vale destacar que a operadora TIM, que obteve o segundo melhor índice MOS, registrou uma maior redução no número de reclamações quando atendidas nos últimos cinco anos.

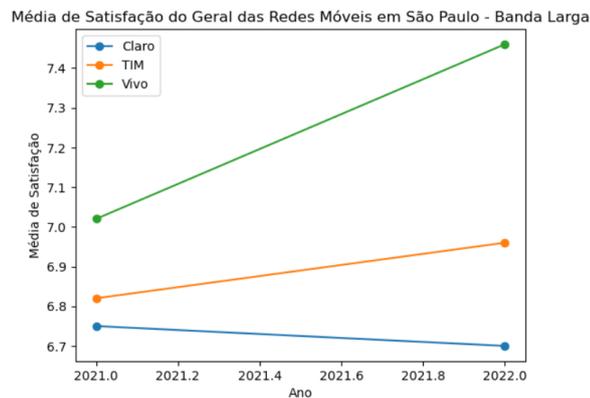
Ao analisarmos o *ranking* de acesso aos serviços de telecomunicação, observamos que, na avaliação específica para São Paulo, a operadora Claro possui o maior número de acessos, assim como tem o maior número de reclamações, enquanto a operadora Oi tem o menor número de reclamações e também possui poucos acessos. O Gráfico 4.21 apresenta o ranking de acessos para São Paulo.

Gráfico 4.21 – *Ranking* de acessos aos serviços de telecomunicação por operadora - São Paulo

Fonte: Brasil (2023)

Na pesquisa de satisfação por localidade, foram considerados os estados de São Paulo, Rio de Janeiro e Minas Gerais, assim como na avaliação por operadora. É importante destacar que, no Gráfico referente ao serviço de banda larga, para o estado de São Paulo, a operadora Oi não foi contabilizada devido à falta de informações disponíveis no painel de informações da Anatel. A seguir, são apresentados os resultados por meio dos gráficos.

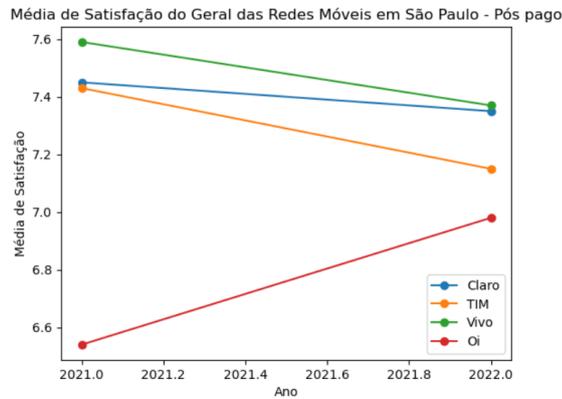
Gráfico 4.22 – Pesquisa de satisfação geral - Banda Larga (São Paulo)



Fonte: Do Autor (2023)

No Gráfico 4.22, é possível observar que a operadora Vivo possui a melhor satisfação geral para o serviço de banda larga no estado de São Paulo e a Claro o menor índice.

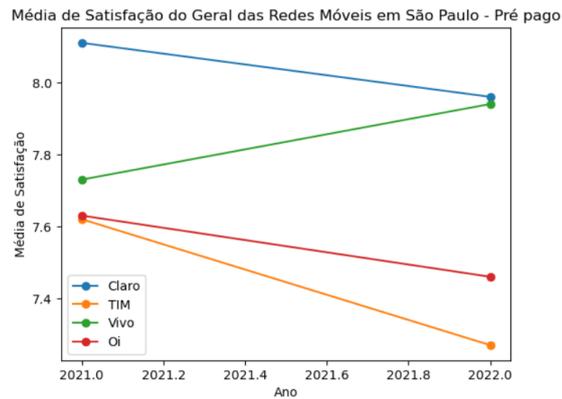
Gráfico 4.23 – Pesquisa de satisfação geral - Pós pago (São Paulo)



Fonte: Do Autor (2023)

No Gráfico 4.23, a operadora Vivo é destacada com o melhor índice de satisfação geral para o serviço de pós-pago no estado de São Paulo, e a operadora Oi possui o menor índice.

Gráfico 4.24 – Pesquisa de satisfação geral - Pré pago (São Paulo)

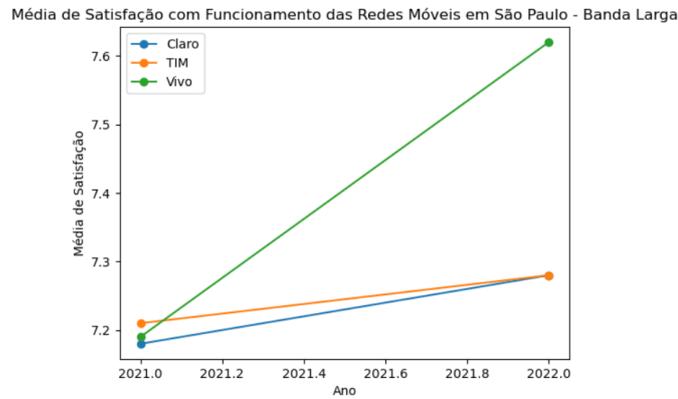


Fonte: Do Autor

No Gráfico 4.24, a operadora Claro é destacada com o melhor índice de satisfação geral para o serviço de pré-pago no estado de São Paulo, e a operadora Tim obteve o menor índice.

Os gráficos a seguir apresentam a "satisfação com funcionamento", no estado de São Paulo.

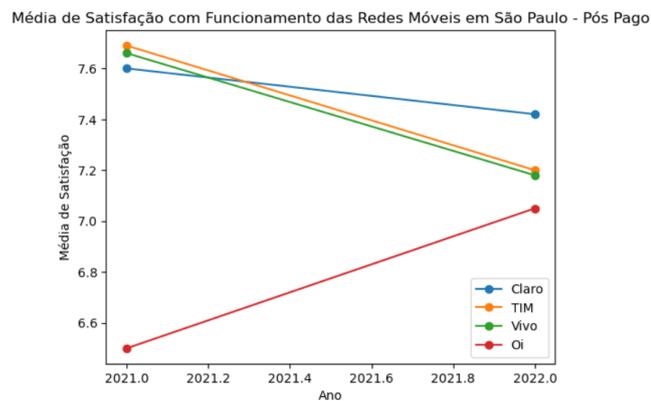
Gráfico 4.25 – Pesquisa de satisfação com funcionamento - Banda Larga (São Paulo)



Fonte: Do Autor (2023)

Pelo Gráfico 4.25, observa-se que a operadora Vivo possui o melhor índice de satisfação no quesito banda larga no ano de 2022, mas fica atrás da TIM no ano de 2021 e a operadora Claro possui o menor índice.

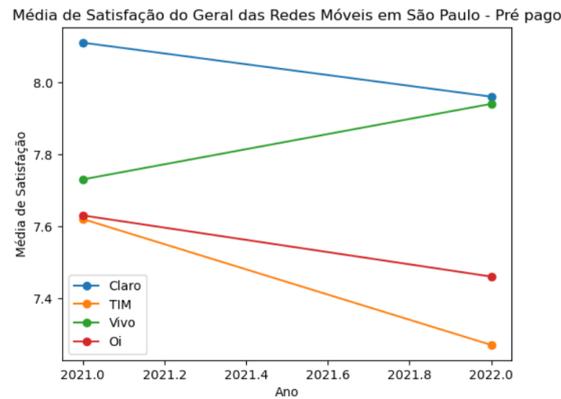
Gráfico 4.26 – Pesquisa de satisfação com funcionamento - Pós pago (São Paulo)



Fonte: Do Autor (2023)

No Gráfico 4.26, observa-se que, para os serviços de pós-pago, a operadora Claro possui o melhor índice, no estado de São Paulo, no ano de 2022, porém fica atrás das operadoras Tim e Vivo no ano de 2021, e a operadora Oi possui o menor índice.

Gráfico 4.27 – Pesquisa de satisfação com funcionamento - Pré pago (São Paulo)



Fonte: Do Autor (2023)

No Gráfico 4.27 a operadora Claro possui o melhor índice para o serviço, pré pago, e a operadora Tim o menor índice.

Comparando os resultados para o estado de São Paulo, observa-se que a operadora Vivo possui os melhores índices quando se trata de banda larga, com resultados semelhantes tanto na pesquisa de satisfação geral quanto na pesquisa de satisfação com o funcionamento. No que diz respeito ao serviço de pós-pago, verificou-se que a operadora Vivo possui o melhor índice na satisfação geral, mas, na pesquisa de satisfação com o funcionamento, ficou atrás da operadora Claro no ano de 2022. Isso indica que o funcionamento dos serviços pode afetar a satisfação do usuário.

Outro ponto que pode ser observado é que as reclamações podem ser devido a dificuldades momentâneas enfrentadas pelo usuário, já que as operadoras Claro e Vivo apresentam o maior número de ocorrências de reclamações e, na predição automática, obtiveram MOS inferiores em relação à Oi e Tim. Essas duas últimas operadoras também demonstraram menor satisfação tanto na pesquisa de satisfação geral quanto na pesquisa de satisfação com o funcionamento, tanto nos serviços de pós-pago quanto nos serviços pré-pago. A exceção ocorre no serviço de banda larga, onde a operadora Claro obteve o menor índice.

Nesta etapa também realizamos comparação entre as reclamações, pesquisa de satisfação e as tecnologias utilizadas pelas operadoras, e conforme abordado na metodologia, serão consideradas apenas as operadoras Tim, Vivo e Claro. Não será levada em conta a tecnologia 5G, uma vez que o aplicativo "Serviços Móveis" disponibilizado pela Anatel não apresenta essa informação no momento da redação desta pesquisa. A Tabela 4.7, apresentam as tecnologias por operadora, utilizada na cidade de São Paulo.

Tabela 4.7 – Tecnologias por operadora (São Paulo)

<b>Operadora</b>	<b>2G</b>	<b>3G</b>	<b>4G</b>
Tim	1089	1292	1711
Vivo	694	1709	1737
Claro	1031	1392	1384

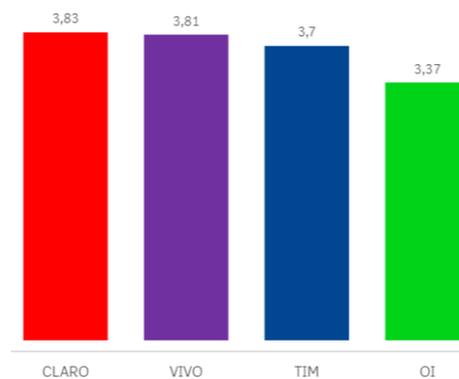
Fonte: Do Autor (2023)

Pela Tabela 4.7, observa-se que a operadora Vivo possui o maior número de antenas 4G em São Paulo, seguida de perto pela operadora Tim. Isso pode indicar o motivo pelo qual a operadora Vivo obteve os melhores índices para o serviço de banda larga, uma vez que a rede 4G possui uma tecnologia mais avançada para a transmissão de dados em comparação com o 3G e 2G.

Outro ponto que reforça a tese de que as reclamações podem ser momentâneas ou decorrentes de dificuldades específicas do usuário é o fato de a operadora Vivo possuir uma quantidade superior de antenas nas tecnologias 4G e 3G, ficando atrás apenas da Tim e Claro na tecnologia 2G, apresentando índices de reclamações e MOS próximos aos da operadora Claro, que registra os maiores índices de reclamação nos últimos dois anos e menor MOS para São Paulo.

Quanto ao atendimento das operadoras em São Paulo, a operadora Claro possui os melhores índices. Esses dados são apresentados no Gráfico 4.28.

Gráfico 4.28 – Qualidade de atendimento das operadoras (São Paulo)

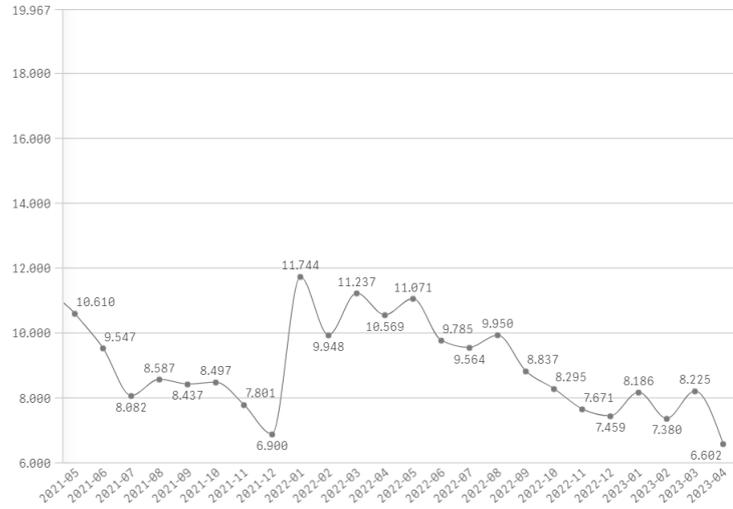


Fonte: Brasil (2023)

Pelos índices apresentados no Gráfico 4.28, pressupõe-se que não há uma relação direta entre as reclamações e o MOS com o atendimento, pois as operadoras Claro e Vivo, que apresentam os melhores índices de para atendimento, obtiveram um maior número de registros de reclamação e os menores MOS para São Paulo.

Para o Rio de Janeiro, a Anatel recebe de 6.000 (seis mil) a 12.000 (doze mil) reclamações por mês, conforme apresentado no Gráfico 4.29.

Gráfico 4.29 – Média de reclamações recebidas na Anatel (Rio de Janeiro)



Fonte: Brasil (2023)

A Figura 4.8 apresenta o total de reclamações de 2018 a 2023, provenientes do Rio de Janeiro e por operadora.

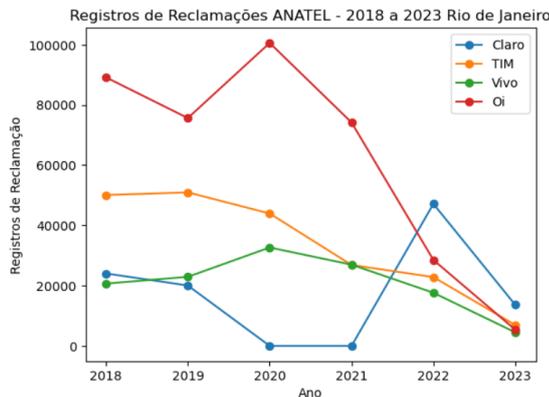
Figura 4.8 – Reclamações por operadora (Rio de Janeiro)

	⊕ 2018	⊕ 2019	⊕ 2020	⊕ 2021	⊕ 2022	⊕ 2023
⊕ CLARO	24.040	20.044	-	-	47.145	13.611
⊕ OI	89.154	75.636	100.575	74.174	28.524	5.362
⊕ TIM	50.080	50.966	43.954	26.869	22.830	6.928
⊕ VIVO	20.672	22.923	32.668	26.963	17.631	4.492

Fonte: Brasil (2023)

O Gráfico 4.30 mostra a evolução das reclamações recebidas pela Anatel do Rio de Janeiro nos últimos cinco anos.

Gráfico 4.30 – Reclamações recebidas por operadora - ANATEL (Rio de Janeiro)



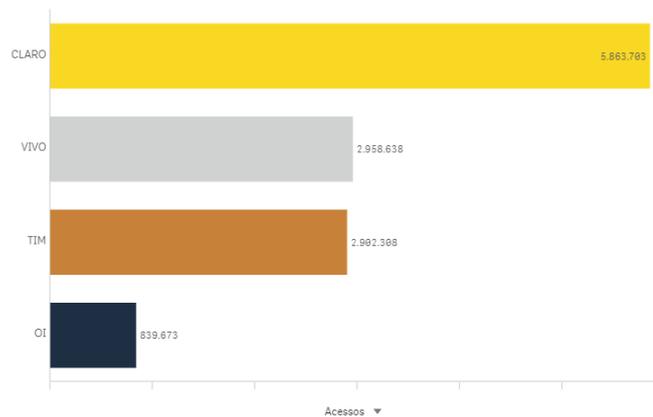
Fonte: Do Autor (2023)

Assim como nos dados gerais e nos dados referentes a São Paulo, no Rio de Janeiro também se observa uma queda em relação às reclamações recebidas pela Anatel. Nesse cenário, a operadora Claro se destaca como aquela com o maior número de reclamações nos últimos dois anos.

No entanto, diferente da situação em São Paulo, a operadora Oi é a que apresenta a maior queda no número de reclamações no Rio de Janeiro. A operadora Tim também possui baixos índices de reclamações. Vale ressaltar que, de acordo com a fase de predição automática, a operadora Tim obteve o melhor MOS em relação às demais operadoras para o Rio de Janeiro.

No que diz respeito ao *ranking* de acessos, verifica-se uma semelhança com o cenário de São Paulo: a operadora Claro é líder tanto em número de acessos quanto em reclamações no Rio de Janeiro. O Gráfico 4.31 apresenta o ranking de acessos para o Rio de Janeiro.

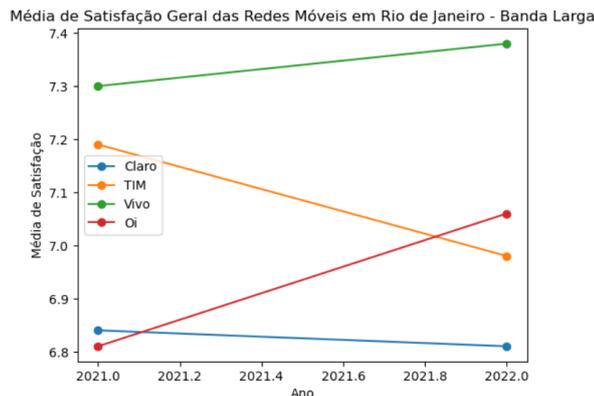
Gráfico 4.31 – *Ranking* de acessos aos serviços de telecomunicação por operadora - Rio de Janeiro



Fonte: Brasil (2023)

Os gráficos a seguir apresentam a pesquisa de satisfação geral e de satisfação com o funcionamento para o estado do Rio de Janeiro.

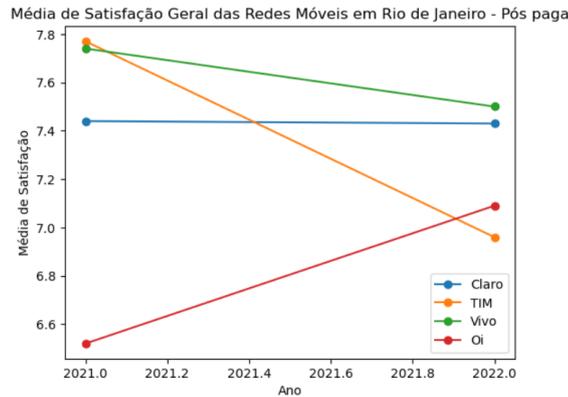
Gráfico 4.32 – Pesquisa de satisfação geral - Banda Larga (Rio de Janeiro)



Fonte: Do Autor (2023)

Pelo Gráfico 4.32, observa-se que a operadora Vivo possui o melhor índice de satisfação no quesito banda larga, e Claro o menor índice no ano de 2022.

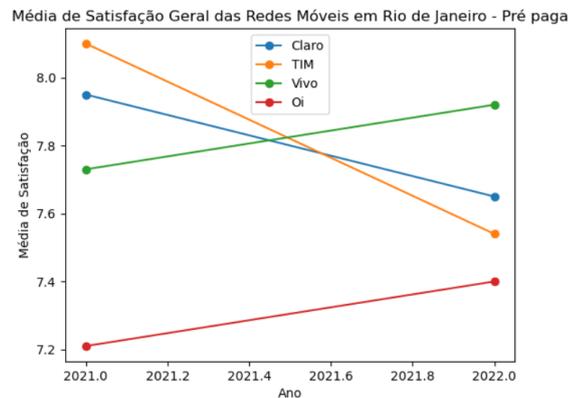
Gráfico 4.33 – Pesquisa de satisfação geral - Pós pago (Rio de Janeiro)



Fonte: Do Autor (2023)

No Gráfico 4.33, observa-se que, para os serviços de pós-pago, a operadora Vivo possui o melhor índice, no ano de 2022, porém fica atrás das operadoras Tim no ano de 2021, que por sua vez possui o menor índice no ano de 2022, o que indica uma variação acentuada da satisfação para operadora Tim.

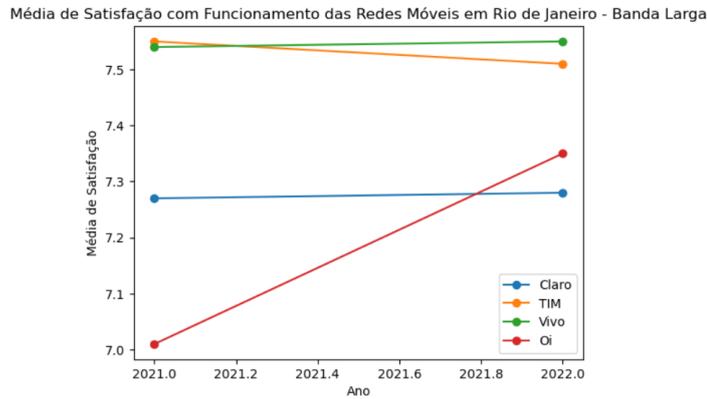
Gráfico 4.34 – Pesquisa de satisfação geral - Pré pago (Rio de Janeiro)



Fonte: Do Autor (2023)

No Gráfico 4.34, observa-se que a operadora Vivo possui o melhor índice para o serviço pré-pago no ano de 2022. No entanto, no ano de 2021, estava atrás das operadoras Tim e Claro. A operadora Oi, por sua vez, apresenta o menor índice de satisfação nesse serviço. Isso reforça que pode haver uma variação na satisfação dos usuários ao longo do tempo, como já mencionado, mostrando que as operadoras podem ter um desempenho diferente em diferentes períodos.

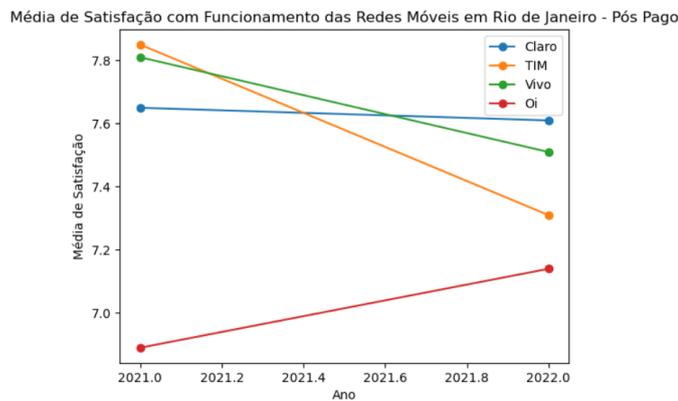
Gráfico 4.35 – Pesquisa de satisfação com funcionamento - Banda Larga (Rio de Janeiro)



Fonte: Do Autor (2023)

De acordo com o Gráfico 4.35, é possível observar que a operadora Vivo possui o melhor índice de satisfação no quesito banda larga no ano de 2022. No entanto, ela está muito próxima da Tim, que foi a melhor no ano de 2021. Por outro lado, a operadora Claro apresenta o menor índice de satisfação no ano de 2022.

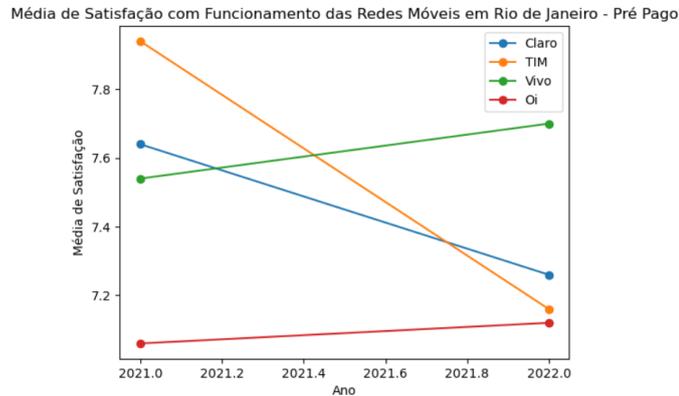
Gráfico 4.36 – Pesquisa de satisfação com funcionamento - Pós pago (Rio de Janeiro)



Fonte: Do Autor (2023)

No Gráfico 4.36, observa-se que, para os serviços de pós-pago, a operadora Claro possui o melhor índice, no ano de 2022, porém fica atrás das operadoras Tim e Vivo no ano de 2021, e a operadora Oi possui o menor índice.

Gráfico 4.37 – Pesquisa de satisfação com funcionamento - Pré pago (Rio de Janeiro)



Fonte: Do Autor (2023)

No Gráfico 4.37, verifica-se que a operadora Vivo apresenta o melhor índice de satisfação para o serviço pré-pago no ano de 2022. No entanto, no ano de 2021, ela ficava atrás das operadoras Tim e Claro, que sofreram uma queda acentuada em seus índices. Já a operadora Oi registra o menor índice de satisfação nesse período.

As tecnologias empregada pelas operadoras no Rio de Janeiro, são apresentadas na Tabela 4.8.

Tabela 4.8 – Tecnologias por operadora (Rio de Janeiro)

Operadora	2G	3G	4G
Tim	648	1241	1490
Vivo	585	1113	1196
Claro	816	1193	1183

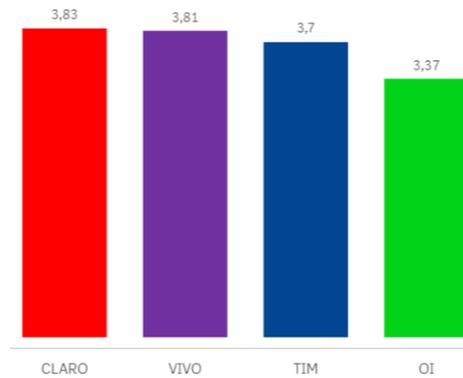
Fonte: Do Autor (2023)

Diferentemente de São Paulo, a operadora Tim possui o maior número de antenas 4G e 3G no Rio de Janeiro, ficando atrás da Claro apenas no quesito 2G. No entanto, assim como em São Paulo, a Vivo conquistou o maior índice de satisfação geral e de funcionamento para os serviços de banda larga, especialmente no ano de 2022. A Tim, por outro lado, apresentou uma queda acentuada nesse aspecto.

A Vivo também mantém o melhor índice de satisfação geral nos serviços de pós-pago e pré-pago, ficando atrás da Claro somente na satisfação com o funcionamento no serviço pós-pago no ano de 2022. Observa-se aqui que o panorama em relação ao maior número de antenas das tecnologias mais recentes, ou seja, 4G e 3G, não indica necessariamente uma maior satisfação, como ocorreu em São Paulo, onde a Vivo possuía a maior quantidade de antenas 4G e obteve os melhores índices de satisfação. No entanto, em relação às reclamações e ao MOS, o panorama permanece semelhante ao de São Paulo, pois as operadoras Claro e Vivo apresentam um maior número de registros de reclamações e menor MOS, mas possuem um desempenho melhor nas pesquisas de satisfação.

Já em relação à qualidade de atendimento, o panorama observado em São Paulo também se mantém no Rio de Janeiro. Ou seja, a operadora Claro possui o melhor atendimento, mas, conforme visto na fase de predição automática, apresenta o menor MOS. O Gráfico 4.38 apresenta esses dados.

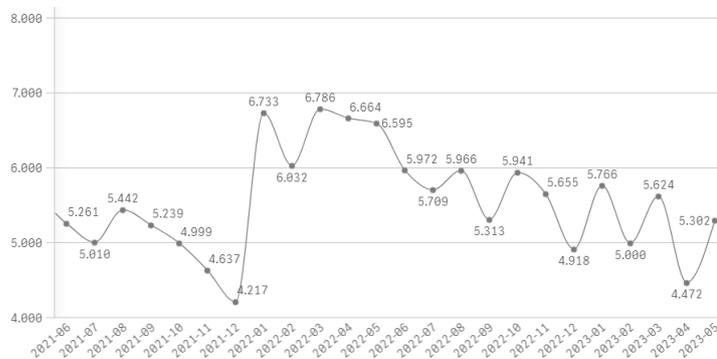
Gráfico 4.38 – Qualidade de atendimento das operadoras (Rio de Janeiro)



Fonte: Brasil (2023)

Para Belo Horizonte, a Anatel recebe em média de 4.000 (quatro mil) a 7.000 (sete mil) reclamações por mês, conforme apresentado no Gráfico 4.39.

Gráfico 4.39 – Media de reclamações recebidas na Anatel (Belo Horizonte)



Fonte: Brasil (2023)

A Figura 4.10 apresenta o total de reclamações de 2018 a 2023, de Belo Horizonte

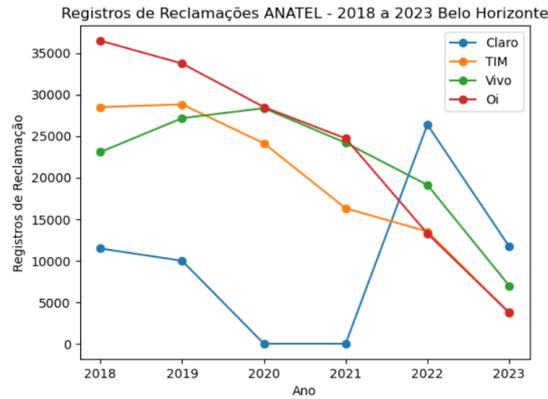
Figura 4.9 – Reclamações por operadora (Belo Horizonte)

	⊕ 2018	⊕ 2019	⊕ 2020	⊕ 2021	⊕ 2022	⊕ 2023
⊕ CLARO	11.466	9.983	-	-	26.381	11.712
⊕ OI	36.493	33.743	28.498	24.727	13.269	3.742
⊕ TIM	28.497	28.820	24.141	16.306	13.512	3.719
⊕ VIVO	23.081	27.166	28.365	24.197	19.122	6.991

Fonte: Brasil (2023)

O gráfico 4.40 mostra a evolução das reclamações recebidas pela ANATEL de Belo Horizonte nos últimos cinco anos.

Gráfico 4.40 – Reclamações recebidas por operadora - ANATEL (Belo Horizonte)

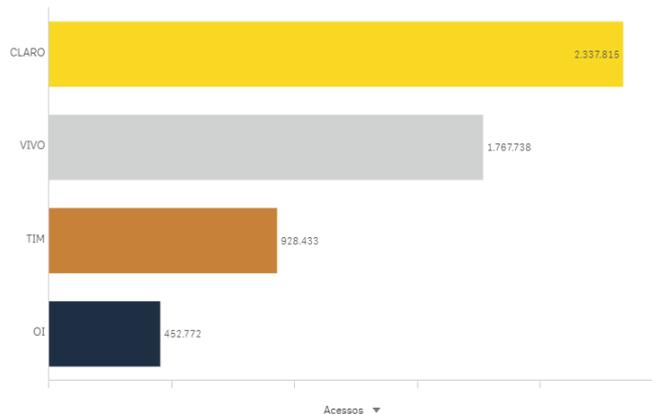


Fonte: Do Autor (2023)

Assim como nos dados gerais, nos dados de São Paulo e do Rio de Janeiro, também é observada uma queda nas reclamações recebidas pela Anatel em Belo Horizonte, sendo a operadora Claro destacada como aquela com o maior número de reclamações nos últimos dois anos. No entanto, em contraste com a situação em São Paulo e em concordância com o Rio de Janeiro, a operadora Oi é aquela que apresenta a maior redução no número de reclamações. As operadoras Tim e Oi possuem os índices mais baixos de reclamações nos anos de 2022 e 2023. É importante ressaltar que, de acordo com a fase de predição automática, a operadora Tim obteve o melhor MOS em relação às demais operadoras em Belo Horizonte.

Quanto ao *ranking* de acessos, observe-se que assim como ocorreu em São Paulo e no Rio de Janeiro, a operadora Claro possui o maior número de acessos e reclamações, seguida pela operadora Vivo. O Gráfico 4.41 apresenta o ranking de acessos para Belo Horizonte.

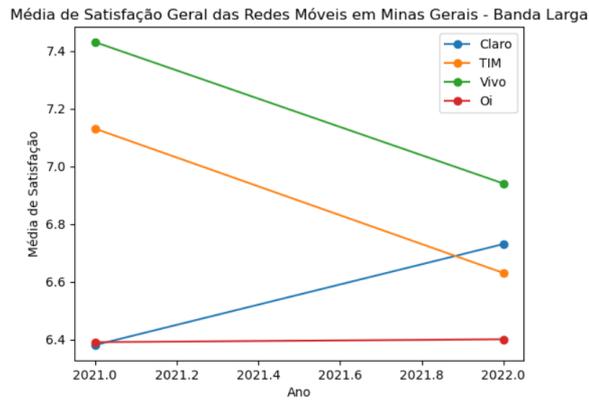
Gráfico 4.41 – Ranking de acessos aos serviços de telecomunicação por operadora - Belo Horizonte



Fonte: Brasil (2023)

Os gráficos a seguir apresentam a pesquisa de satisfação geral e de satisfação com o funcionamento para o estado do Minas Gerais.

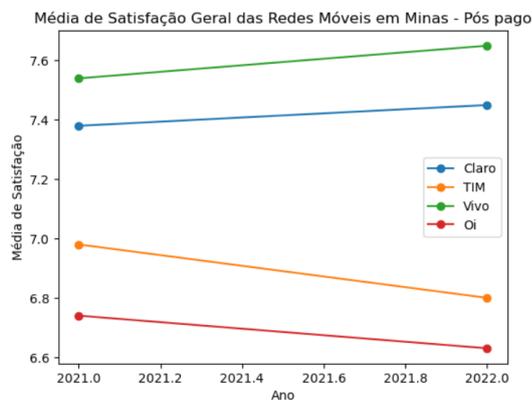
Gráfico 4.42 – Pesquisa de satisfação geral - Banda Larga (Minas Gerais)



Fonte: Do Autor (2023)

Pelo Gráfico 4.42, observa-se que a operadora Vivo possui o melhor índice de satisfação no quesito banda larga, em Minas Gerais, e a Oi menor índice no ano de 2022.

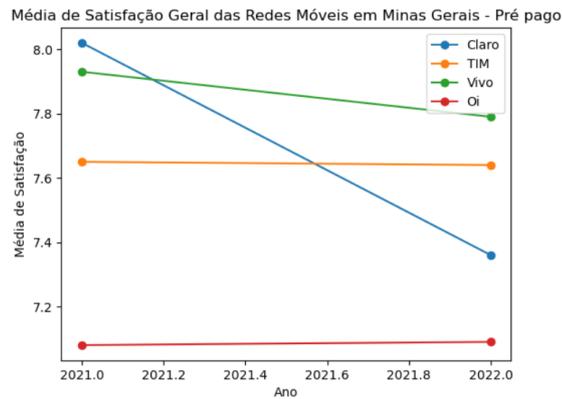
Gráfico 4.43 – Pesquisa de satisfação geral - Pós pago (Minas Gerais)



Fonte: Do Autor 92023)

No Gráfico 4.43, observa-se que, para os serviços de pós-pago, a operadora Vivo possui o melhor índice, em Minas Gerais, e que a operadora Oi possui o menor índice.

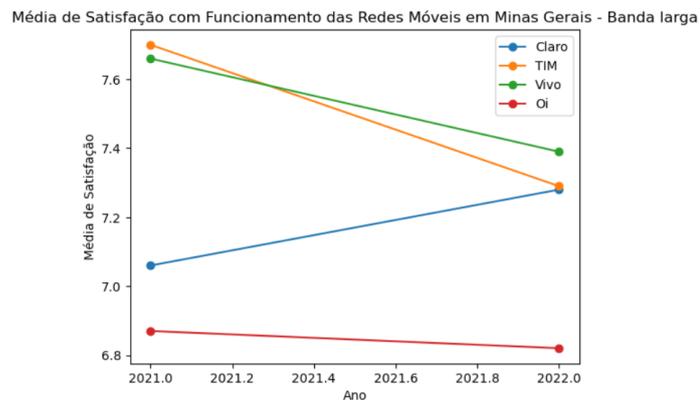
Gráfico 4.44 – Pesquisa de satisfação geral - Pré pago (Minas Gerais)



Fonte: Do Autor (2023)

No Gráfico 4.44, observa-se que a operadora Vivo possui o melhor índice para o serviço pré-pago no ano de 2022. No entanto, no ano de 2021, a Claro possuía o melhor índice. A operadora Oi, por sua vez, apresenta o menor índice de satisfação nesse serviço.

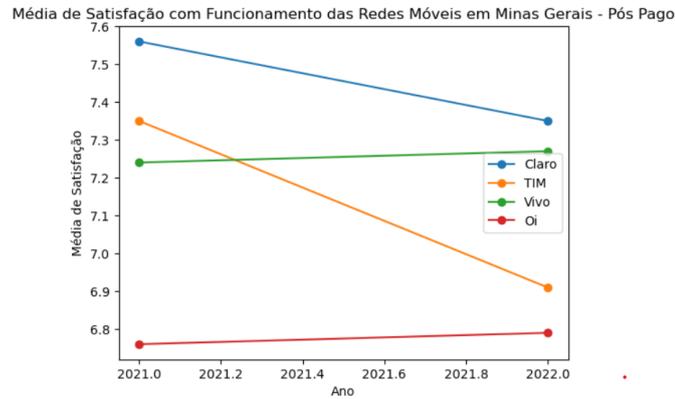
Gráfico 4.45 – Pesquisa de satisfação com funcionamento - Banda Larga (Minas Gerais)



Fonte: Do Autor (2023)

No Gráfico 4.45, é possível observar que a operadora Vivo possui o melhor índice de satisfação no quesito banda larga no ano de 2022. No entanto, a Tim foi a melhor no ano de 2021, e que a operadora Oi apresenta o menor índice.

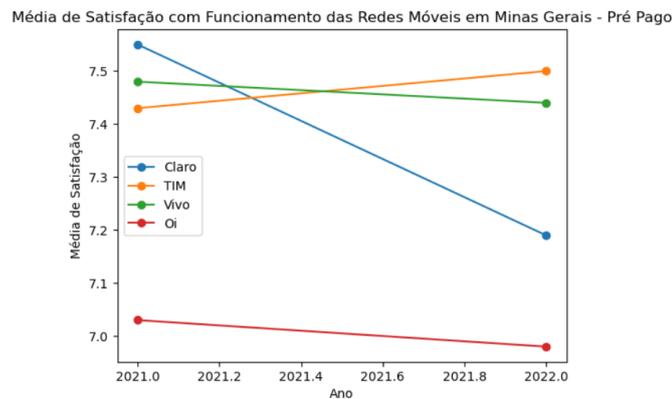
Gráfico 4.46 – Pesquisa de satisfação com funcionamento - Pós pago (Minas Gerais)



Fonte: Do Autor (2023)

No Gráfico 4.46, observa-se que, para os serviços de pós-pago, a operadora Claro possui o melhor índice, e a operadora Oi possui o menor índice.

Gráfico 4.47 – Pesquisa de satisfação com funcionamento - Pré pago (Minas Gerais)



Fonte: Do Autor (2023)

No Gráfico 4.47, verifica-se que a operadora Tim apresenta o melhor índice de satisfação para o serviço pré-pago no ano de 2022. No entanto, no ano de 2021, ela ficava atrás das operadoras Claro e Vivo, e a operadora Oi registra o menor índice.

As tecnologias empregada pelas operadoras em Belo Horizonte, são apresentadas na Tabela 4.8.

Tabela 4.9 – Tecnologias por operadora (Belo Horizonte)

Operadora	2G	3G	4G
Tim	213	262	357
Vivo	287	352	374
Claro	337	451	455

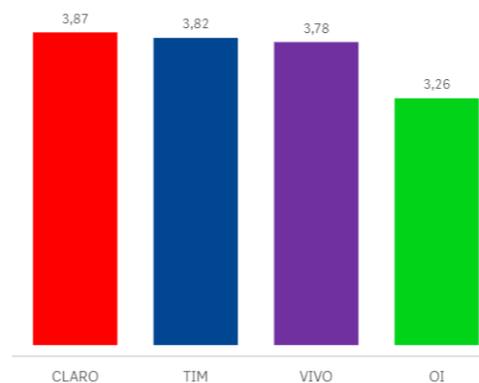
Fonte: Do Autor (2023)

Diferentemente de São Paulo e do Rio de Janeiro, a operadora Claro possui o maior número de antenas 4G, 3G e 2G em Belo Horizonte. A Vivo, apesar de ter números menores em comparação com a Claro, apresenta uma cobertura superior à Tim. No entanto, mesmo com a ampla infraestrutura da Claro, isso não se reflete na satisfação geral ou no funcionamento dos serviços. Assim como em São Paulo e no Rio de Janeiro, a Vivo conquistou o maior índice de satisfação geral e de funcionamento para os serviços de banda larga, especialmente no ano de 2022. A Claro ainda foi superada pela Vivo na satisfação geral dos serviços de pós-pago e pré-pago, e também ficou atrás da Tim, sendo superior apenas na pesquisa de satisfação com o funcionamento do serviço pós-pago.

Outra observação importante, já vista em São Paulo e no Rio de Janeiro, é que o MOS obtido pela Tim está mais relacionado à quantidade de registros de reclamação do que necessariamente com a tecnologia ou com a satisfação do serviço. Isso é evidenciado pelo fato de a Vivo, que possui um MOS menor, apresentar um melhor índice de satisfação, tanto no geral quanto em relação ao funcionamento dos serviços.

Quanto à qualidade no atendimento, a operadora Claro obteve o melhor índice em Minas Gerais, São Paulo e Rio de Janeiro. No entanto, é importante mencionar que a operadora Tim possui o segundo melhor índice em Minas Gerais, mas não obteve o mesmo desempenho em São Paulo e Rio de Janeiro, onde a Vivo ocupa essa posição. Ao compararmos com a fase de predição, percebe-se que não há uma correlação exata entre o atendimento e as reclamações, uma vez que a Tim obteve o melhor MOS, seguida pela Claro, essa situação é oposta. O Gráfico 4.48, mostra esse dados.

Gráfico 4.48 – Qualidade de atendimento das operadoras (Minas Gerais)



Fonte: Brasil (2023)

#### 4.5 Comparações entre os dados da predição automática com a Anatel

Na fase de predição automática, constatou-se que os usuários utilizam principalmente as redes sociais para expressar sua insatisfação com os serviços das operadoras. Essa tendência pode ser observada tanto na classificação geral quanto na classificação por operadora e localidade.

Na análise dos dados coletados pela Anatel revelou-se uma correspondência interessante entre as reclamações registradas e a quantidade de menções nas redes sociais. As operadoras Claro e Vivo, que se destacam por possuir o maior número de reclamações tanto em nível geral quanto por localidade, também são as mais mencionadas nas redes sociais. Essa relação sugere que a percepção negativa dos clientes em relação a essas operadoras é refletida tanto nos registros formais de reclamações quanto nas informações informais nas redes sociais.

Um ponto relevante a ser observado é o *ranking* de acessos por operadoras. A nível nacional, a operadora Vivo lidera com o maior número de acessos, seguida pela Claro, TIM e, por último, a Oi. No entanto, ao analisar por localidade, a operadora Claro se destaca com o maior número de acessos, seguido pela Vivo.

Esses dados do *ranking* de acessos, nacional e por localidade, nos leva a uma importante reflexão: o maior número de acessos pode acarretar em mais reclamações dos clientes, o que pode explicar as pontuações mais baixas de MOS atribuídas às operadoras Claro e Vivo.

No que diz respeito às pesquisas de satisfação, é importante destacar que os resultados podem variar ao longo do tempo e diferir entre os diferentes serviços oferecidos e localidades. Por exemplo, no serviço de banda larga, a operadora Vivo obteve índices mais altos tanto nacionalmente quanto em nível local. No entanto, essa mesma tendência não se repetiu nos serviços pós-pagos e pré-pagos, nos quais a operadora Claro obteve melhores índices em nível nacional. Além disso, ao analisar por localidade, houve regiões em que a Vivo superou a Claro em termos de satisfação. Esses resultados indicam que a satisfação dos usuários pode variar de acordo com a região e o tipo de serviço, o que não necessariamente se reflete na satisfação expressa nas redes sociais.

Outro aspecto relevante é que nem sempre o maior número de antenas de uma operadora se traduz em melhores índices de satisfação. Isso pôde ser observado em cidades como Rio de Janeiro e Belo Horizonte, onde a quantidade de antenas não apresentou uma correlação direta com os melhores índices de satisfação. Da mesma forma, ter um bom desempenho no atendimento ao cliente não necessariamente resulta em menos reclamações registradas ou em uma melhora na satisfação expressa nas redes sociais.

Essas análises demonstram a complexidade e a variedade de fatores que influenciam a satisfação dos usuários e as reclamações registradas. Compreender essa discrepância é fundamental para

que as operadoras possam adotar medidas visando melhorar a qualidade de seus serviços e atender às expectativas dos consumidores.

## 5 CONCLUSÃO

Nesta pesquisa, avaliou-se a Qualidade de Experiência dos consumidores dos serviços de rede móvel no Brasil, coletando opiniões na plataforma de rede social Twitter. Utilizou-se o BERT, um modelo de rede neural baseado em *transformers*, conhecido por sua capacidade de entender o contexto das palavras em um texto. Após a classificação dos dados, eles foram correlacionados com o *Mean Opinion Score* (MOS), uma métrica que mede a satisfação do usuário com as redes de comunicação. Os resultados foram apresentados em termos de satisfação geral dos usuários, satisfação por operadora e satisfação por região, comparando-os com informações fornecidas pela Anatel, agência reguladora de telecomunicações do Brasil.

Dado o contexto da pesquisa no território brasileiro, realizamos uma comparação entre os modelos BERT Multilíngue e BERTimbau, ambos com conhecimento linguístico voltado para o português brasileiro. Os resultados indicaram uma precisão de 0.77 nos testes utilizando o BERT Multilíngue e 0.80 utilizando o BERTimbau. Devido aos resultados superiores obtidos com o BERTimbau, optamos por utilizar esse modelo para as predições nos conjuntos de dados, visando obter novos indicadores de QoE por meio das redes sociais.

Na fase de predição automática, observou-se uma tendência dos usuários de redes móveis brasileiras utilizarem as redes sociais para expressar reclamações e indicar insatisfação com os serviços. Essa tendência foi observada de forma geral, tanto por operadora quanto por localidade. Além disso, notou-se que há uma paridade em relação à satisfação dos usuários em relação às operadoras, com destaque para a operadora Tim, que obteve o índice MOS geral mais elevado.

Durante o levantamento de dados da base da Anatel, foi observada uma discrepância nas informações, uma vez que o número de reclamações registradas pela Anatel não reflete diretamente na pesquisa de satisfação geral, funcionamento ou avaliação das tecnologias utilizadas, assim como na satisfação com o atendimento aos consumidores.

Ao comparar as fases, notou-se que a correlação mais evidente entre a satisfação manifestada nas redes sociais e as informações da Anatel é o registro de reclamações e o *ranking* de acesso por operadora. Conforme mencionado anteriormente, há uma tendência dos usuários utilizarem as redes sociais, especialmente para expressar suas queixas. Ao analisar os registros, verificou-se que a operadora Claro, que possui o maior número de menções nas redes sociais nos últimos anos, também apresenta o maior número de registros de reclamações e o maior número de acessos.

Sugerimos algumas indicações para trabalhos futuros, como analisar as razões por trás das diferenças nas informações disponibilizadas pela Anatel, além da criação de conjuntos de dados com entidades nomeadas para identificar os serviços que recebem mais reclamações nas redes sociais. Também

é sugerido realizar testes com novos modelos pré-treinados para o mercado brasileiro, como o "*gpt2-small-portuguese*", visando aprimorar as análises e predições no contexto específico do país.

Em suma, acreditamos que a análise conjunta dos dados das redes sociais e da base da Anatel permite uma compreensão mais completa dos desafios e pontos fortes de cada operadora, auxiliando no aprimoramento contínuo dos serviços oferecidos.

## REFERÊNCIAS

- AL-AYYOUB, M. et al. Deep learning for arabic nlp: A survey. **Journal of computational science**, New York, v. 26, p. 522–531, 2018.
- AL-MAKHADMEH, Z.; TOLBA, A. Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach. **Computing**, Wien, v. 102, n. 2, p. 501–522, 2020.
- AMO, S. D. Técnicas de mineração de dados. **Jornada de Atualização em Informatica**, Porto Alegre, p. 26, 2004. Disponível em: <[https://www.researchgate.net/profile/Sandra-Amo/publication/260300816\\_Tecnicas\\_de\\_Minerao\\_de\\_Dados/links/54230bd80cf290c9e3ae25e3/Tecnicas-de-Minerao-de-Dados.pdf](https://www.researchgate.net/profile/Sandra-Amo/publication/260300816_Tecnicas_de_Minerao_de_Dados/links/54230bd80cf290c9e3ae25e3/Tecnicas-de-Minerao-de-Dados.pdf)>. Acesso em: 10 jun. 2022.
- ANCHIÊTA, R. et al. Pln: Das técnicas tradicionais aos modelos de deep learning. **Sociedade Brasileira de Computação**, 2021.
- ARAÚJO, M. M.; CRUZ, E. A. Análise e mineração de dados sociais com a ferramenta weka. **FAMETRO**, Manaus, v. 2017, p. 12, 2017.
- BATRA, R. et al. Evaluating polarity trend amidst the coronavirus crisis in peoples' attitudes toward the vaccination drive. **Sustainability**, MDPI, v. 13, n. 10, p. 5344, 2021.
- BERMUDEZ, H.-F. et al. Live video-streaming evaluation using the itu-t p. 1203 qoe model in lte networks. **Computer Networks**, Amsterdam, v. 165, p. 106967, 2019.
- BORGES, L. E. **Python para desenvolvedores: aborda Python 3.3**. São Paulo: Novatec Editora, 2014.
- BRANCO, P.; TORGO, L.; RIBEIRO, R. P. Relevance-based evaluation metrics for multi-class imbalanced domains. In: SPRINGER. **Advances in Knowledge Discovery and Data Mining: 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part I 21**. [S.l.], 2017. p. 698–710.
- BRASIL LEI Nº 9.472, . **Dispõe sobre a organização dos serviços de telecomunicações, a criação e funcionamento de um órgão regulador e outros aspectos institucionais, nos termos da Emenda Constitucional nº 8, de 1995. Diário Oficial da União**. Brasília, 1997. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/19472.htm#:~:text=LEI%20N%2%BA%209.472%2C%20DE%2016%20DE%20JULHO%20DE%201997.&text=Disp%3%B5e%20sobre%20a%20organiza%3%A7%3%A3o%20dos,Constitucional%20n%2%BA%208%2C%20de%201995](http://www.planalto.gov.br/ccivil_03/leis/19472.htm#:~:text=LEI%20N%2%BA%209.472%2C%20DE%2016%20DE%20JULHO%20DE%201997.&text=Disp%3%B5e%20sobre%20a%20organiza%3%A7%3%A3o%20dos,Constitucional%20n%2%BA%208%2C%20de%201995)>. Acesso em: 20 out. 2022.
- BRASIL, M. d. C. **Agência Nacional de Telecomunicações**. Brasília, 2023. Disponível em: <<https://www.gov.br/anatel/pt-br>>. Acesso em: 20 out. 2022.
- CHANG, Y.-C.; KU, C.-H.; CHEN, C.-H. Social media analytics: Extracting and visualizing hilton hotel ratings and reviews from tripadvisor. **International Journal of Information Management**, Guildford, v. 48, p. 263–279, 2019.
- CIAMBRONE, D. et al. Assessing qoe-driven management policies for voip and video streaming service provisioning. In: **IEEE INTERNATIONAL WORKSHOP ON COMPUTER AIDED MODELING AND DESIGN OF COMMUNICATION LINKS AND NETWORKS**. Barcelona: [s.n.], 2018. p. 1–6.
- COLLE, J. E.; RODRIGUES, W. M.; SILVA, W. R. da. Redes neurais artificiais: ideias básicas no ensino médio. **Brazilian Journal of Development**, São José dos Pinhais, v. 8, p. 21795–21810, 2022.

CÔRTEZ, S. da C.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados-funcionalidades, técnicas e abordagens**. [S.l.]: PUC, 2002.

COSTA, B. R. L. Bola de neve virtual: o uso das redes sociais virtuais no processo de coleta de dados de uma pesquisa científica. **Revista interdisciplinar de gestão social**, Salvador, v. 7, n. 1, 2018.

CUI, J. et al. Qoe-based resource allocation for multi-cell noma networks. **IEEE Transactions on Wireless Communications**, New York, v. 17, n. 9, p. 6160–6176, 2018.

DAENGSI, T.; WUTTIDITTACHOTTI, P. Qoe modeling for voice over ip: Simplified e-model enhancement utilizing the subjective mos prediction model: A case of g. 729 and thai users. **Journal of Network and Systems Management**, Porto Alegre, v. 27, n. 4, p. 837–859, 2019.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv**, Ithaca, Oct. 2019, 2018. Disponível em: <<https://arxiv.org/abs/1810.04805>>. Acesso em: 20 ago. 2022.

FAGHIHI, E.; BEHDADFAR, M. Adaptive resource utilization and quality management in voip networks with quantitative relationship between qoe and qos using hmm. In: IEEE. **International Conference on Information Networking**, 34., 2020. Barcelona, 2020. p. 493–498.

FALCÃO, L. C. de J.; LOPES, B.; SOUZA, R. R. Absorção das tarefas de processamento de linguagem natural (nlp) pela ciência da informação (ci): uma revisão da literatura para tangibilização do uso de nlp pela ci. **Em Questão**, v. 28, n. 1, p. 13–34, 2022.

FERNANDES, F. T.; FILHO, A. D. P. C. Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. **Revista Brasileira de Saúde Ocupacional**, São Paulo, v. 44, p. 1–12, 2019.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, DF, v. 35, p. 25–30, 2006.

GOUVEIA, D. J. de S.; ROSA, R. L.; RODRÍGUEZ, D. Z. Avaliação da qualidade da voz em serviços de comunicação usando deep learning. In: SBC. **Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, Anais**. Campos do Jordão, 2018. p. 183–196.

GUIMARÃES, D. A. Introdução às comunicações móveis. **Revista INATEL Telecomunicações**, Santa Rita do Sapucaí, v. 1, n. 01, 1998. Disponível em: <<http://www.wirelessbrasil.org/wirelessbr/colaboradores/dayani/intro01.html>>. Acesso em: 20 out. 2021.

HE, L.; JIANG, D.; WEI, C. A qoe-based dynamic energy-efficient network selection algorithm. **Wireless Networks**, Amsterdam, v. 27, n. 5, p. 3585–3595, 2021.

HICKMAN, L. et al. Text preprocessing for text mining in organizational research: Review and recommendations. **Organizational Research Methods**, Sage Publications Sage CA: Los Angeles, CA, v. 25, n. 1, p. 114–146, 2022.

HU, Z. et al. Evaluating qoe in voip networks with qos mapping and machine learning algorithms. **Neurocomputing**, New York, v. 386, p. 63–83, 2020.

INTERNATIONAL TELECOMMUNICATION UNION ITU-T P.800. **SERIES P: TELEPHONE TRANSMISSION QUALITY Methods for objective and subjective assessment of quality**. 1996. Disponível em: <<https://www.itu.int/rec/T-REC-P.800-199608-I>>. Acesso em: 21 out. 2021.

- IŞIK, M.; DAĞ, H. The impact of text preprocessing on the prediction of review ratings. **Turkish Journal of Electrical Engineering and Computer Sciences**, v. 28, n. 3, p. 1405–1421, 2020.
- JELASSI, S. et al. Quality of experience of voip service: A survey of assessment approaches and open issues. **IEEE Communications surveys & tutorials**, New York, v. 14, n. 2, p. 491–513, 2012.
- KHAN, A. et al. Predicting politician's supporters' network on twitter using social network analysis and semantic analysis. **Scientific Programming**, Hindawi Limited, v. 2020, p. 1–17, 2020.
- KHAN, L. et al. Multi-class sentiment analysis of urdu text using multilingual bert. **Scientific Reports**, London, v. 12, n. 1, p. 1–17, 2022.
- KHAN, S. et al. Transformers in vision: A survey. **ArXiv**, Ithaca, 2021. Disponível em: <<https://arxiv.org/abs/2101.01169>>. Acesso em: 10 ago. 2022.
- LASELVA, D. et al. Advancements of qoe assessment and optimization in mobile networks in the machine era. In: **IEEE wireless communications and networking conference workshops, 2018**. Barcelona: [s.n.], 2018. p. 101–106.
- LEE, E.; LEE, C.; AHN, S. Comparative study of multiclass text classification in research proposals using pretrained language models. **Applied Sciences**, Islamabad, v. 12, n. 9, p. 4522, 2022.
- LI, Y. Deep learning-based natural language processing methods for sentiment analysis in social networks. **Mathematical Problems in Engineering**, Cairo, v. 7, July 2022.
- LIN, T. et al. A survey of transformers. **Machine Learning**, Ithaca, June 2021. Disponível em: <<https://arxiv.org/abs/2106.04554>>. Acesso em: 20 ago. 2022.
- LIU, X.; SHIN, H.; BURNS, A. C. Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. **Journal of Business Research**, Athens, v. 125, p. 815–826, 2021.
- MAIA, G. Marketing digital e redes sociais: a importância para as empresas durante a pandemia. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 8, n. 2, p. 929–950, 2022.
- MARQUES, V. Redes sociais 360. Alfragide: Leya, 2018.
- MARTINS JUNIOR, C. A. **Classificação de índices de satisfação de usuários de telefonia móvel usando aprendizado de máquinas**. Dissertação (Mestrado) — Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2020.
- MATOS, F. F.; MAGALHÃES, L. H. d.; SOUZA, R. R. Recuperação e classificação de sentimentos de usuários do twitter em período eleitoral. **Informação Informação**, v. 25, n. 1, p. 92–114, 2020.
- MCKINNEY, W. **Python for Data Analysis**. [S.l.]: "O'Reilly Media, Inc.", 2022.
- MILLMAN, K. J.; AIVAZIS, M. Python for scientists and engineers. **Computing in Science & Engineering**, IEEE, New York, v. 13, n. 2, p. 9–12, 2011.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, São Paulo, v. 1, n. 1, p. 32, 2003.
- MORENTE-MOLINERA, J. A. et al. Analysing discussions in social networks using group decision making methods and sentiment analysis. **Information Sciences**, New York, v. 447, p. 157–168, 2018.

- MOSAFI, F. **Machine Learning: metodologia de mineração automatizada com dados das redes sociais e processamento de linguagem natural**. São Paulo: Dialética, 2022.
- OBEICA, I. C. de O.; MARTINS, D. M. S. Análise de sentimentos em tweets: Um estudo de caso sobre a opinião das pessoas em relação a vacina em tempos da pandemia do covid-19. **Caderno de Estudos em Engenharia de Software**, v. 4, n. 1, 2022.
- OLIVEIRA, L. A.; ALENCAR, M. S.; LOPES, W. T. A. Evolução da arquitetura de redes móveis rumo ao 5g. **Revista de Tecnologia da Informação e Comunicação**, Campina Grande, v. 8, n. 2, p. 43–50, 2018.
- PAIS, S.; CORDEIRO, J.; JAMIL, M. L. Nlp-based platform as a service: a brief review. **Journal of Big Data**, London, v. 9, n. 1, p. 1–26, 2022.
- PARK, J.; CHUNG, K. Resource prediction-based edge collaboration scheme for improving qoe. **Sensors**, New York, v. 21, n. 24, p. 8500, 2021.
- PIROTTI, R.; ZUCCOLOTTO, M. Transmissão de dados através de telefonia celular: arquitetura das redes gsm e gprs. **Revista Liberato**, Nova Hamburgo, v. 10, n. 13, p. 81–90, 2009.
- ROCHOL, J. **Sistemas de comunicação sem fio: conceitos e aplicações**. Bookman. Porto Alegre: [s.n.], 2018.
- ROGERS, A.; KOVALEVA, O.; RUMSHISKY, A. A primer in bertology: What we know about how bert works. **Transactions of the Association for Computational Linguistics**, Ithaca, v. 8, p. 842–866, 2020.
- SANTOS, M. dos; LOPES, R. Qoe analysis for mobile network services using twitter opinion extraction. **INFOCOMP Journal of Computer Science**, v. 21, n. 1, 2022.
- SANTOS, M. R. dos et al. Asqm: Audio streaming quality metric based on network impairments and user preferences. **IEEE Transactions on Consumer Electronics**, IEEE, 2023.
- SANTOS, R. D. L. **EREDES GSM, GPRS, EDGE E UMTS**. Rio de Janeiro, 2008. Disponível em: <[https://www.gta.ufrj.br/ensino/eel879/trabalhos\\_vf\\_2008\\_2/ricardo/1\\_1.html#celula](https://www.gta.ufrj.br/ensino/eel879/trabalhos_vf_2008_2/ricardo/1_1.html#celula)>. Acesso em: 21 out. 2021.
- SCHMIDT, R. M. Recurrent neural networks (rnns): A gentle introduction and overview. **arXiv preprint arXiv:1912.05911**, 2019.
- SENA, F. R. **Evolução da Tecnologia Móvel Celular e o Impacto nos Resíduos de Eletroeletrônicos**. Rio de Janeiro: PUC-RIO, 2012.
- SILVA, A. V. **Um modelo de classificação para o Reconhecimento de Entidades Nomeadas**. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2020.
- SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. Bertimbau: Pretrained bert models for brazilian portuguese. In: CERRI, R.; PRATI, R. C. (Ed.). **Intelligent Systems**. Cham: Springer International, 2020. p. 403–417.
- SOUZA, O. d. et al. Um método de sumarização automática de textos através de dados estatísticos e processamento de linguagem natural. **Informação & Sociedade: Estudos**, 2017.
- SU, G.-M. et al. Qoe in video streaming over wireless networks: perspectives and research challenges. **Wireless networks**, Amsterdam, v. 22, n. 5, p. 1571–1593, 2016.

SUNDARARAJ, V.; REJEESH, M. A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites. **Journal of Retailing and Consumer Services**, Amsterdam, v. 58, p. 102190, 2021.

Twitter. **Developer platform**. 2022. Disponível em: <<https://developer.twitter.com/en/docs>>. Acesso em: 24 de março 2022.

VASWANI, A. et al. Attention is all you need. **ArXiv**, Ithaca, v. 30, 2017. Disponível em: <<https://arxiv.org/abs/1706.03762>>. Acesso em: 20 ago. 2022.

XU, J. et al. On the properties of mean opinion scores for quality of experience management. In: **IEEE international symposium on multimedia 2011**. Dana Point IEEE: [s.n.], 2011. p. 500–505.

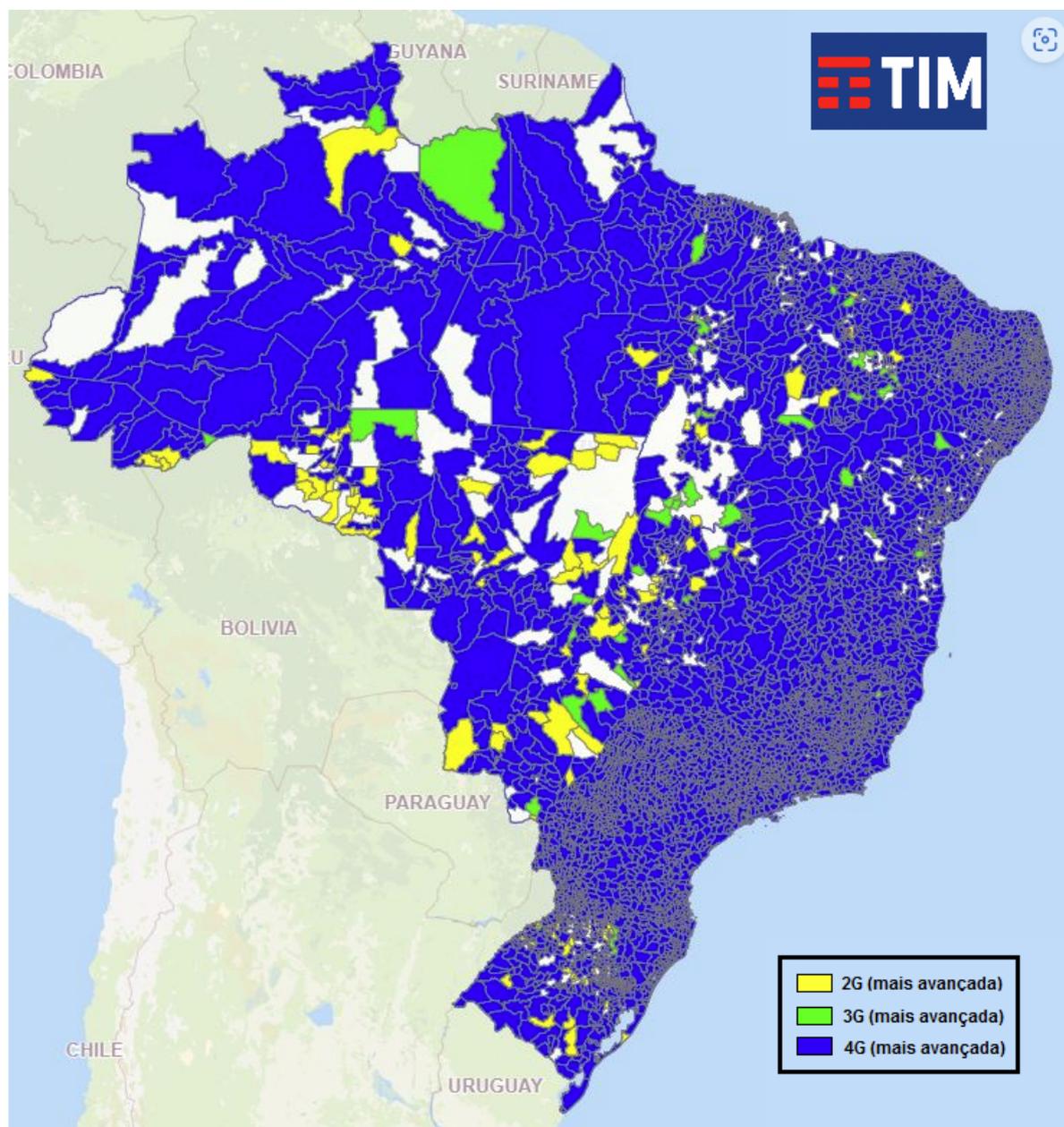
YANG, Y.-C. et al. Developing an automatic system for classifying chatter about health services on twitter: Case study for medicaid. **Journal of medical Internet research**, Toronto, v. 23, n. 5, p. e26616, 2021.

ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, p. e1253, 2018.

ZHONG, Y. et al. Qoe and cost for wireless networks with mobility under spatio-temporal traffic. **IEEE Access**, New York, v. 7, p. 47206–47220, 2019.

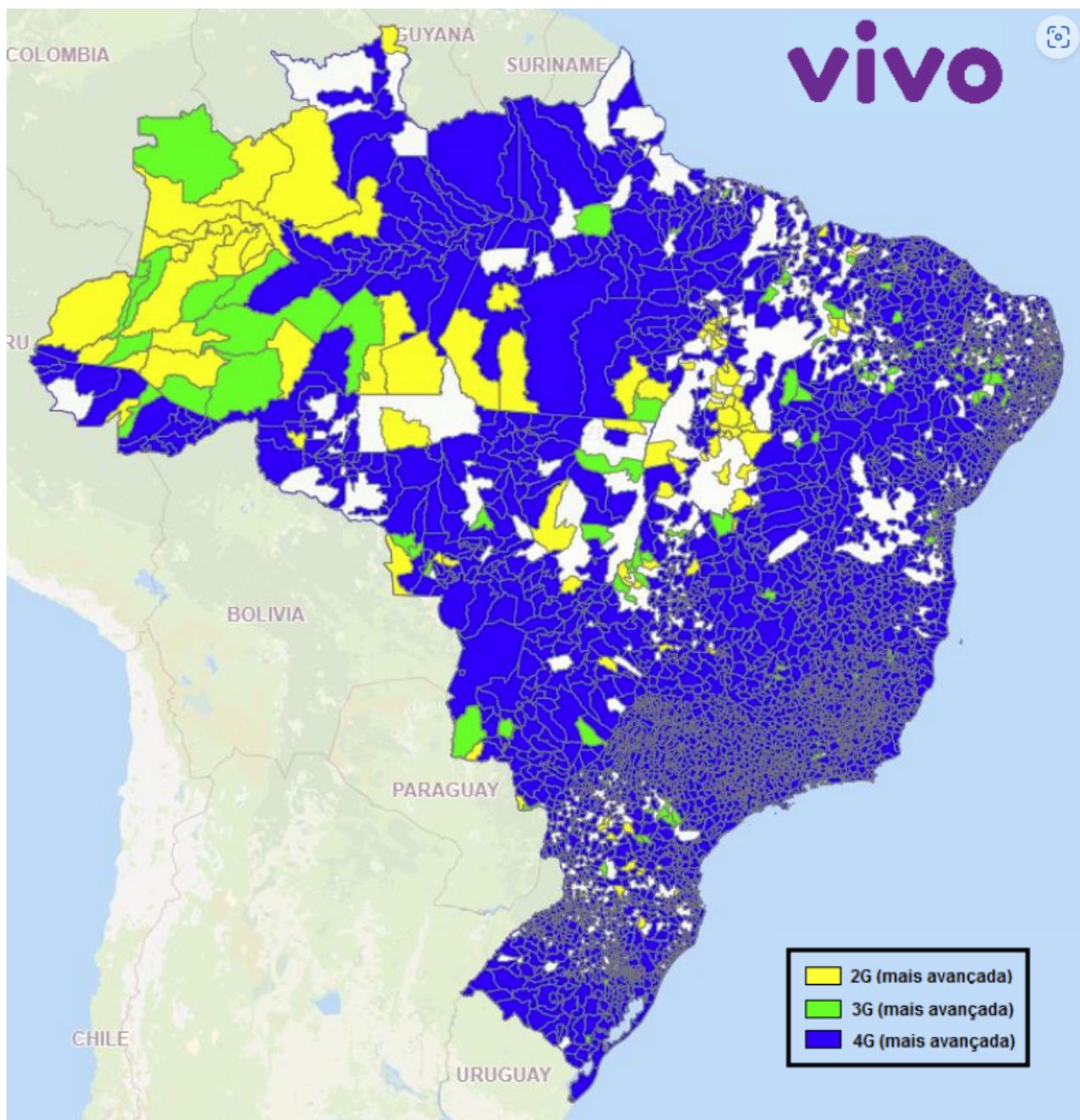
**APÊNDICE A – Mapas Panorâmicos de cobertura por operadora**

Figura 1 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil - Operadora TIM



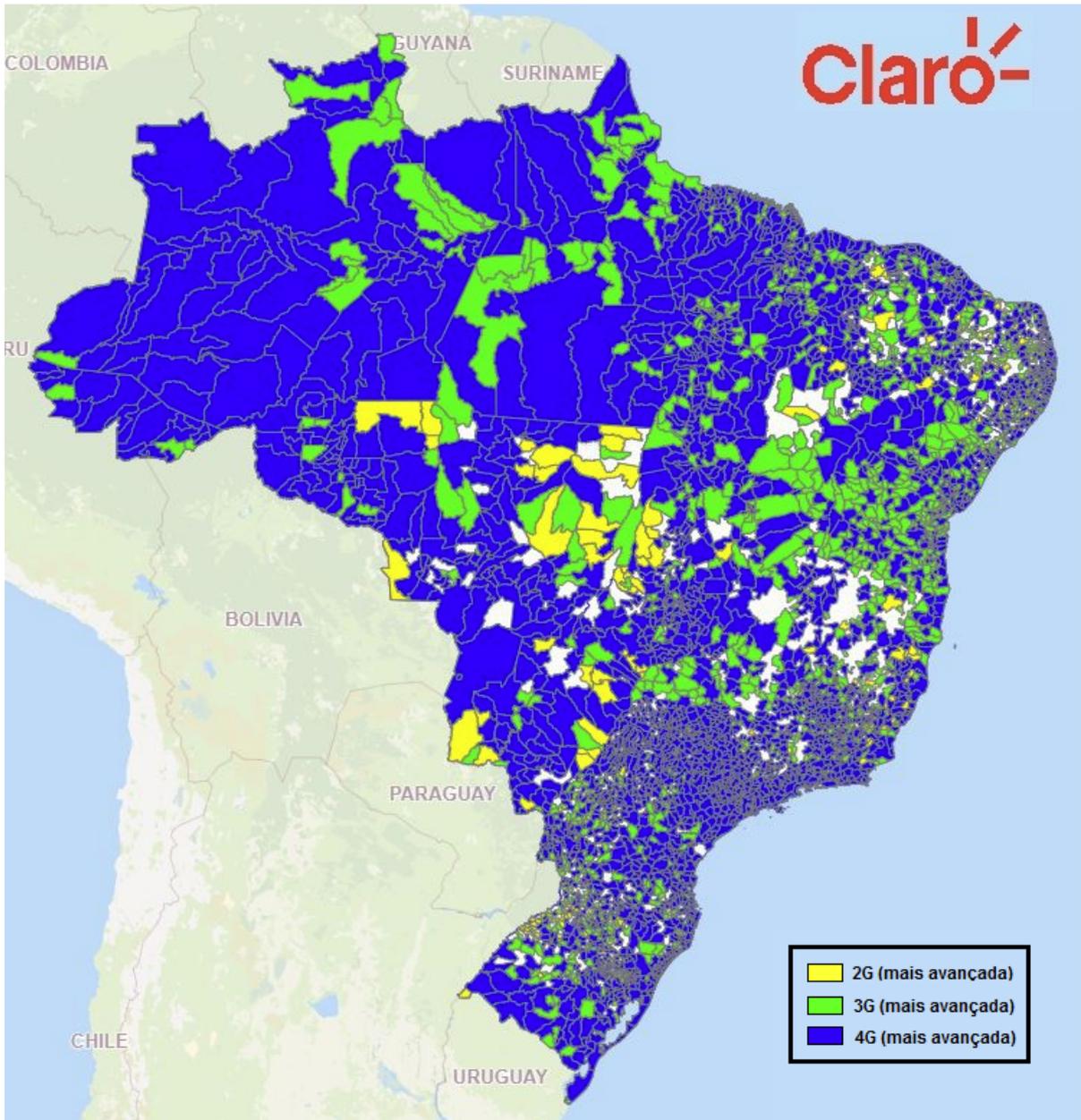
Fonte: Brasil (2023)

Figura 2 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil - Operadora Vivo



Fonte: Brasil (2023)

Figura 3 – Mapa Panorâmico das Tecnologias de Redes Moveis no Brasil - Operadora Claro



Fonte: Brasil (2023)

### APÊNDICE B – Pseudocondigos

**Input:** query - a consulta de busca, filename - o nome do arquivo para salvar os tweets

```

consumer_key = "SUA_CONSUMER_KEY";
consumer_secret = "SUA_CONSUMER_SECRET";
access_token = "SEU_ACCESS_TOKEN";
access_token_secret = "SEU_ACCESS_TOKEN_SECRET";

auth = criar_instância_do_Auth(consumer_key, consumer_secret);
definir_acesso_token(auth, access_token, access_token_secret);
api = criar_instância_do_API(auth, wait_on_rate_limit=True);
tweets = criar_lista_vazia();

for cada tweet em obter_tweets(api, query, "pt", "extended") do
    tweet_info = { };
    tweet_info['id'] = obter_id(tweet);
    tweet_info['user'] = obter_nome_do_usuario(tweet);
    tweet_info['date'] = obter_data(tweet);
    tweet_info['text'] = obter_texto_completo(tweet);
    tweet_info['favorite_count'] = obter_contagem_de_favoritos(tweet);
    tweet_info['retweet_count'] = obter_contagem_de_retweets(tweet);
    tweet_info['location'] = obter_localização_do_usuario(tweet) se
        obter_localização_do_usuario(tweet) senão "N/A";
    adicionar_elemento_na_lista(tweets, tweet_info);

end

df = criar_dataframe(tweets);
salvar_dataframe_em_csv(df, filename, index=False);

exibir_mensagem("Dados salvos com sucesso!");
Algorithm 1: busca_tweets(query, filename)

```

**Entrada:** Dataframe

**Saída** : modelo treinado

```
/* Definir tamanho máximo para os textos */
```

```
msize ← 2 × o comprimento máximo do texto
```

```
/* Tokenização utilizando o modelo BERT */
```

**Function** *regular\_encode*:

```
  Função que recebe um conjunto de textos e um tokenizador com o comprimento
  máximo e retorna os textos tokenizados
```

```
  return array de ['input_ids']
```

**end**

```
/* Inicializar o tokenizador BERT */
```

```
tokenizer ← Instancia o tokenizador BERT pré-treinado
```

```
/* Divisão do dataset em dados de treino e teste */
```

Divide o conjunto de dados balanceado em conjuntos de treinamento e teste usando a função `train_test_split` do scikit-learn.

```
/* Tokenizar os textos em one-hot vectors usando
```

```
tf.keras.utils.to_categorical */
```

Codifica os textos de treinamento usando a função `regular_encode` e o tokenizador BERT

Codifica as classes de treinamento em vetores one-hot usando a função `to_categorical` do Keras.

Codifica os textos de teste usando a função `regular_encode` e o tokenizador BERT.

Codifica as classes de teste em vetores one-hot usando a função `to_categorical` do Keras.

**Algorithm 2:** Divisão de Treino e Teste

```
/* Construir o modelo BERT */
```

**Function** *build\_model*:

```
  Função que cria e compila o modelo BERT. Recebe o transformer (a camada do modelo
  BERT), a função de perda e o comprimento máximo.
```

```
  model.compile(tf.keras.optimizers.Adam(lr=3e-5), loss=loss, metrics=['accuracy'])
```

```
  return model
```

**end**

```
/* Inicializar o modelo BERT */
```

```
transformer_layer ← Carrega o modelo BERT pré-treinado
```

```
model ← Cria o modelo utilizando a função build_model com a camada do modelo BERT e
o tamanho máximo.
```

**Algorithm 3:** Modelo de Classificação