



RELATÓRIO FINAL DE INICIAÇÃO CIENTÍFICA

CAMPUS GUARULHOS

MODELO DE RELATÓRIO DE INICIAÇÃO CIENTÍFICA

ALUNO: HÁLAN WILLIAM DA COSTA SOUSA

ORIENTADOR: RODRIGO CAMPOS BORTOLETTO

AGOSTO DE 2025.

RESUMO

Este estudo analisou dados de portabilidade numérica no Brasil, disponibilizados pela ANATEL, para identificar padrões de comportamento de usuários usando análise exploratória de dados e algoritmos de aprendizado de máquina. Para isso, foram utilizados dados do Portal de Dados Abertos do Governo Federal, processados com Python e bibliotecas como *pandas*, *sklearn* e *matplotlib*. A análise revelou que os modelos de regressão linear tradicionais tiveram um desempenho baixo, com uma taxa de acerto de aproximadamente 4,7%. Mesmo após o tratamento dos dados, a taxa de acerto subiu apenas para 6,49%. A mudança para um modelo classificador não supervisionado, que dividiu os dados em três grupos, foi muito mais eficaz, atingindo taxas de acerto acima de 85%, e até 100% em uma das classes. O estudo conclui que a portabilidade é fortemente influenciada pelo tipo de serviço (SMP ou STFC) e pela localização do usuário, e que o modelo classificador pode auxiliar na tomada de decisões por operadoras e órgãos reguladores.

PALAVRAS-CHAVE: *Portabilidade numérica; Telecomunicações; Serviço Móvel Pessoal; Serviço de Telefonia Fixa Comutada; Aprendizado de Máquina.*

ABSTRACT

This study analyzed data on number portability in Brazil, made available by ANATEL, to identify user behavior patterns through exploratory data analysis and machine learning algorithms. For this purpose, data from the Federal Government's Open Data Portal were processed using Python and libraries such as pandas, scikit-learn, and matplotlib. The analysis revealed that traditional linear regression models performed poorly, with an accuracy rate of approximately 4.7%. Even after data preprocessing, the accuracy increased only to 6.49%. The transition to an unsupervised classification model, which divided the data into three groups, proved to be much more effective, achieving accuracy rates above 85% and even 100% in one of the classes. The study concludes that number portability is strongly influenced by the type of service (SMP or STFC) and by the user's location, and that the classification model can support decision-making by service providers and regulatory agencies.

KEYWORDS: *Number portability; Telecommunications; Personal Mobile Service; Switched Fixed Telephone Service; Machine Learning.*

OBJETIVO GERAL

Desenvolver algoritmos de aprendizado de máquina capazes de prever quantidade de portabilidades que serão efetuadas.

Objetivos específicos

Estudar a preparação de uma base de dados para algoritmos de AM, discorrer sobre a matriz telefônica no Brasil.

SUMÁRIO

OBJETIVO GERAL	4
INTRODUÇÃO	6
MATERIAIS E MÉTODOS	7
ANÁLISE DE DADOS E RESULTADOS OBTIDOS	12
CONCLUSÕES E SUGESTÕES	19

INTRODUÇÃO

A portabilidade numérica foi implementada no Brasil no ano de 2008, permitindo que usuários de operadoras telefônicas, tanto do Serviço Móvel Pessoal (SMP) quanto do Serviço de Telefonia Fixa Comutada (STFC), trocassem de ofertante do serviço sem ter o seu número alterado. O STFC refere-se a um serviço destinado à transmissão de sinais entre pontos fixos determinados, podendo abranger distâncias locais ou internacionais de longa distância. Já o SMP é caracterizado como um serviço de comunicação móvel terrestre coletivo, que possibilita a comunicação entre estações móveis, bem como entre estas e outras estações (TELECO, s.d.).

O Aprendizado de máquina (AM) é uma tecnologia poderosa que permite trabalhar uma grande quantidade de dados que seria inviável de serem analisados manualmente. A criação de algoritmos de AM vem sendo amplamente realizada no ramo das telecomunicações, tanto para medir o índice de satisfação de clientes (MARTINS JUNIOR, 2020), quanto para alocação de recursos e escalonamento de usuários (LEITE, 2014).

Nesse contexto, o presente trabalho tem como objetivo analisar essas portabilidades entre operadoras com base nos dados disponibilizados pela Agência Nacional de Telecomunicações (ANATEL), por meio de uma Análise Exploratória de Dados (AED) bem como algoritmos de AM. O propósito deste estudo é identificar padrões de comportamento entre os usuários, contribuindo para uma melhor compreensão, por parte das operadoras e órgãos reguladores, dos mecanismos e dinâmicas envolvidos nesse processo. Para que assim, medidas reguladoras possam ser tomadas para fomentar a competitividade das empresas entrantes nesse mercado.

MATERIAIS E MÉTODOS

Neste estudo foram utilizados dados de tabelas disponibilizados no Portal de Dados Abertos do Governo Federal (BRASIL, s.d.) sobre as portabilidades numéricas que ocorrem entre duas empresas e o saldo de clientes que determinada prestadora obtém em um determinado período de tempo, os quais foram processados por meio da linguagem de programação Python, utilizando o ambiente de desenvolvimento integrado (IDE) Jupyter Notebook. É válido destacar que a linguagem Python foi utilizada tanto por facilitar o manejo de dados armazenados em formato de tabela, pois são convertidos em bancos de dados, quanto por possuir uma vasta gama de bibliotecas empregadas no campo da ciência de dados. Segundo a Data Science Academy, “existem bibliotecas para praticamente todos os aspectos do aprendizado de máquina aplicado” (SILVA, 2021).

Os dados, originalmente em formato de arquivo *Comma Separated Values* (CSV), foram transformados em bancos de dados, a partir da biblioteca pandas, sendo submetidos, posteriormente, à um processo de limpeza e verificação de possíveis dados faltantes, que são etapas extremamente importantes para melhorar o resultado do algoritmo de aprendizado, dado que “*dados mais precisos levam a generalizações mais precisas*” (TERESA, 2021). Após isso, foram necessárias mais algumas adequações, tais como a transformação do campo referente à data no tipo *datetime64[ns]*, para que análises temporais pudessem ser realizadas, a substituição do nome da operadora Nextel pela Claro S.A., em virtude da aquisição de uma pela outra (G1, 2019), a renomeação dos campos para nomes mais curtos sem perder seu significado. Posto isso, os dois bancos de dados estavam prontos para serem analisados.

Na Análise Exploratória de Dados, buscou-se o comportamento de correlações entre os campos desses dados, para a identificação de qual seria mais adequado para ser o alvo em um algoritmo de AM. Dessa maneira, foram feitas análises visando as relações entre variáveis numéricas, como o saldo, diferença entre o número de clientes recebidos e doados por cada operadora em um dado período de tempo, e cada operadora telefônica. Nessa parte, as análises foram feitas utilizando-se as bibliotecas python *matplotlib* e *seaborn*, ideais para o esboço de gráficos.

Os modelos de AM, sejam eles de regressão ou classificação utilizam apenas dados numéricos, o que não se traduz na realidade dos dados trabalhados nesse estudo, visto que a unidade federativa (UF), prestadora e tipo de serviço prestado são informações categóricas ou qualitativas destes dados. Por conta disso, é necessário aplicar a esse conjunto de dados uma codificação de dados alfanuméricos para dados puramente numéricos, o que é realizado por

meio da biblioteca *scikit-learn*, em Python. Com essa ferramenta, é possível criar essa codificação dos dados, para dados com menos informações, como o tipo de serviço, o mais recomendado é o uso do método de *OneHotEncoder* que categoriza esses dados a partir de uma matriz esparsa, ou seja, para o caso da codificação dos serviços, divididos em STFC e SMP, uma assume o valor zero, enquanto a outra assume o valor um. Da mesma forma, os campos que representam a unidade federativa e a operadora também precisam ser codificados, porém, pela quantidade de valores distintos que esses campos possuem, torna-se inviável de utilizar o *OneHotEncoder*, sendo codificados a partir do método *LabelEncoder*, que associa para cada valor textual, um valor numérico. Assim, para as informações sobre a UF e prestadora, foram criados bancos de dados que funcionam como tabela de consulta sobre o valor chave que cada um assume.

Para a definição de qual campo seria o alvo buscado em um modelo de regressão, aplicou-se um método da biblioteca pandas que busca a correlações entre variáveis, que podem variar no intervalo [-1,1], ou seja, correlação negativa, sendo duas variáveis inversamente proporcionais, ou correlação positiva, de modo que sejam diretamente proporcionais.

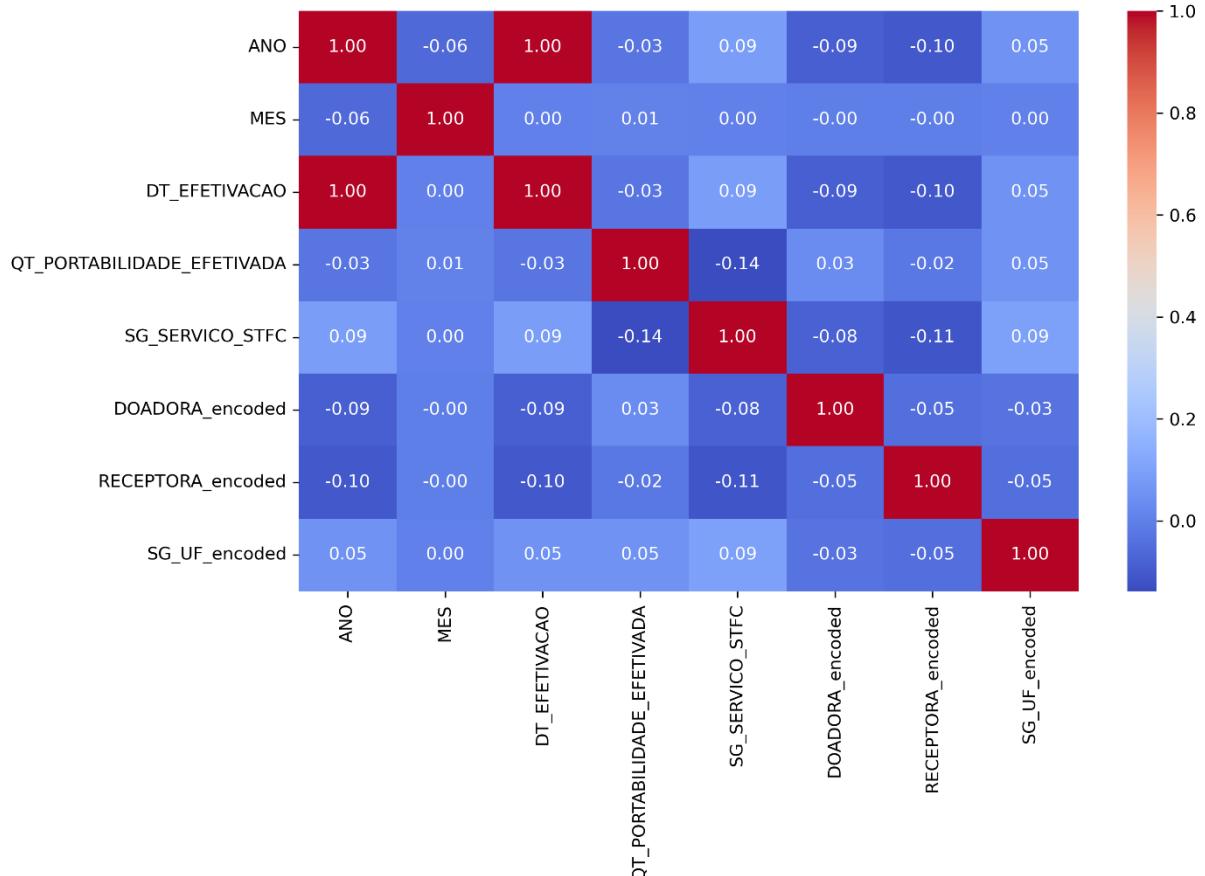


Figura 1 – correlação entre os campos do banco de dados de portabilidades
Fonte: própria do autor (2025)

Para o banco de dados sobre as portabilidades, nenhuma variável teve nenhuma correlação, positiva ou negativa, com valor expressivo para ser utilizada num modelo de regressão, como se pode ver na figura 1.

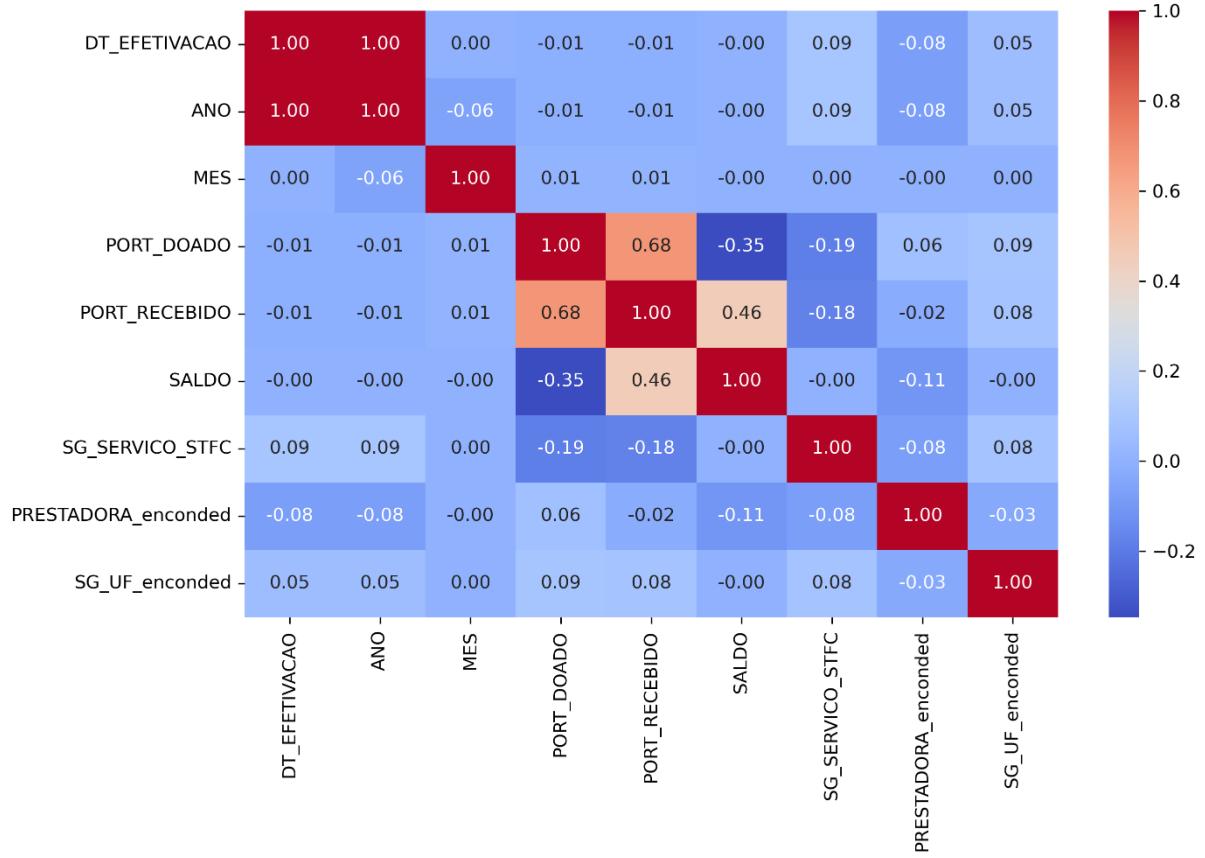


Figura 2 – correlação entre os campos do banco de dados de saldo
Fonte: própria do autor (2025)

Já na figura 2, observa-se variáveis com fortes correlações entre si, como a relação positiva entre a portabilidade recebida e a doadas, de 0,68, entre o saldo e a portabilidade recebida, de 0,46, a relação negativa entre saldo e portabilidade doadas, de 0,35. Entretanto, a portabilidade doadas é calculada da seguinte maneira.

$$\text{SALDO} = \text{portabilidade recebida} - \text{portabilidade doadas}$$

Portanto, como a variável saldo é obtida pela diferença das outras duas, podemos dizer que elas são linearmente dependentes (SANTOS, s.d.), logo, em um modelo de regressão a utilização dessas variáveis em conjunto resultaria em um modelo viciado, que acertaria em todas as previsões de dados que ele já conhece.

Sendo assim, a relação de maior expressividade e sem viés é entre o tipo de serviço e a quantidade de portabilidade doadas, de 0,19. Então, após essa verificação, iniciou-se a sondagem de modelos de AM por regressão. Mesmo com a baixa correlação positiva, as realizações de

testes em modelos de regressão mais simples são ideais para avaliar o desempenho desses modelos com a base de dados. Desse modo, foram feitos testes com os modelos de regressão *LinearRegression*, *Lasso* e *Ridge* da biblioteca *scikit-learn*. Assim, por conta dos resultados de baixo índice de acerto nesses modelos de regressão, foi realizado o tratamento entre os intervalos interquartis do conjunto (IQR), para diminuir a quantidade de valores extremos, que atrapalham o desempenho de modelos de AM. Mesmo assim, a taxa de acerto continuou muito inviável, então optou-se por outro modelo de método de conjuntos, “combinam as previsões de vários estimadores base criados com um determinado algoritmo de aprendizado para melhorar a generalização em relação a um único estimador” (PEDREGOSA, 2024), sendo o modelo *XGBoostRegressor* utilizado para essa finalidade.

Porém, mesmo com as adequações do método de conjuntos, a regressão mostrou-se ineficaz para essa base de dados, por conta da baixa correlação entre as variáveis. A partir disso, trocou-se o modelo de aprendizado supervisionado de regressão, para um modelo de AM não supervisionado de classificação. No modelo classificador, foram utilizadas apenas as variáveis numéricas de portabilidade dada, e portabilidade recebida, para evitar o problema da dependência linear, e as variáveis categóricas de UF, prestadora e tipo de serviço. Então, essas variáveis foram normalizadas com média igual a zero e desvio padrão igual a um. Isso é imprescindível para modelos classificadores pois “variáveis medidas em escalas diferentes não contribuem igualmente para a análise e podem acabar criando um viés” (RAY, 2020). Então, para continuar com o modelo foi utilizada a técnica do cotovelo, utilizando, novamente a biblioteca *scikit-learn*, com o método *KMeans*, conforme destaca Bengfort.

O *KElbowVisualizer* implementa o método “cotovelo” para ajudar os cientistas de dados a selecionar o número ideal de clusters ajustando o modelo com uma gama de valores para K. Se o gráfico de linhas se assemelhar a um braço, o “cotovelo” (o ponto de inflexão na curva) é uma boa indicação de que o modelo subjacente se ajusta melhor naquele ponto.
(BENGFORT, 2024)

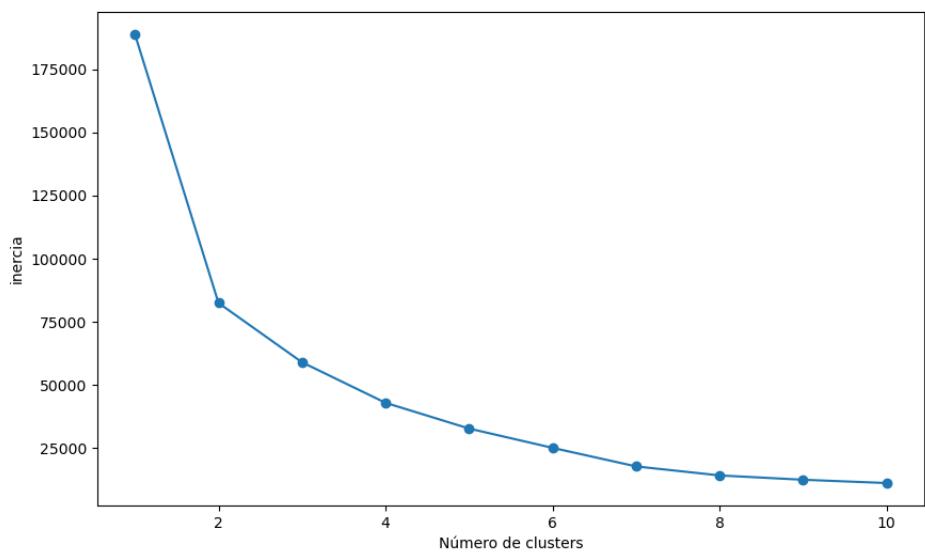


Figura 3 – técnica do cotovelo empregada na base de dados de saldo
Fonte: própria do autor (2025)

Como se pode observar na figura 3, o ponto de inflexão ocorre quando o número de agrupamentos for igual a 3. Portanto, esse é o número ideal de classes a serem criadas e passadas como parâmetro para o *KMeans*, que divide a base de dados em três grupos, numerados de 0 a 2. Dessa maneira, cada linha da base de dados tem seus valores analisados e associados a algum dos três grupos. Na última etapa, essa nova classificação é utilizada como alvo no modelo classificador *XGBoostClassifier*.

ANÁLISE DE DADOS E RESULTADOS OBTIDOS

Os resultados obtidos neste estudo se dividem nos resultados da análise exploratória de dados, e nos resultados de modelos de AM, ambos podem ajudar a determinar a dinâmica comportamental dos clientes.

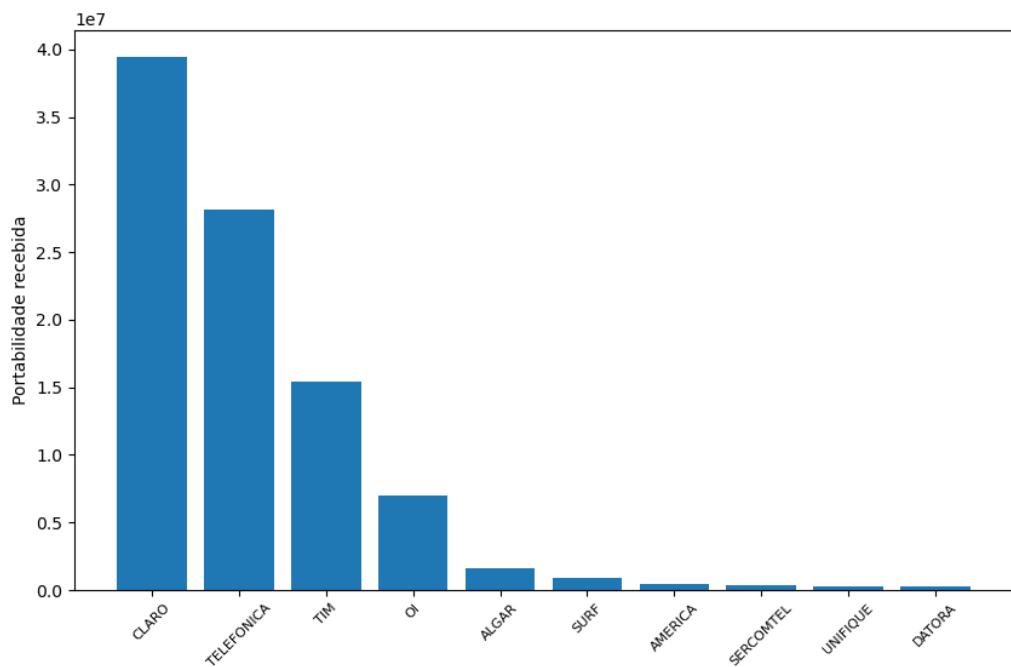


Figura 4 – operadoras que mais recebem clientes, 2008 a 2025
Fonte: própria do autor (2025)

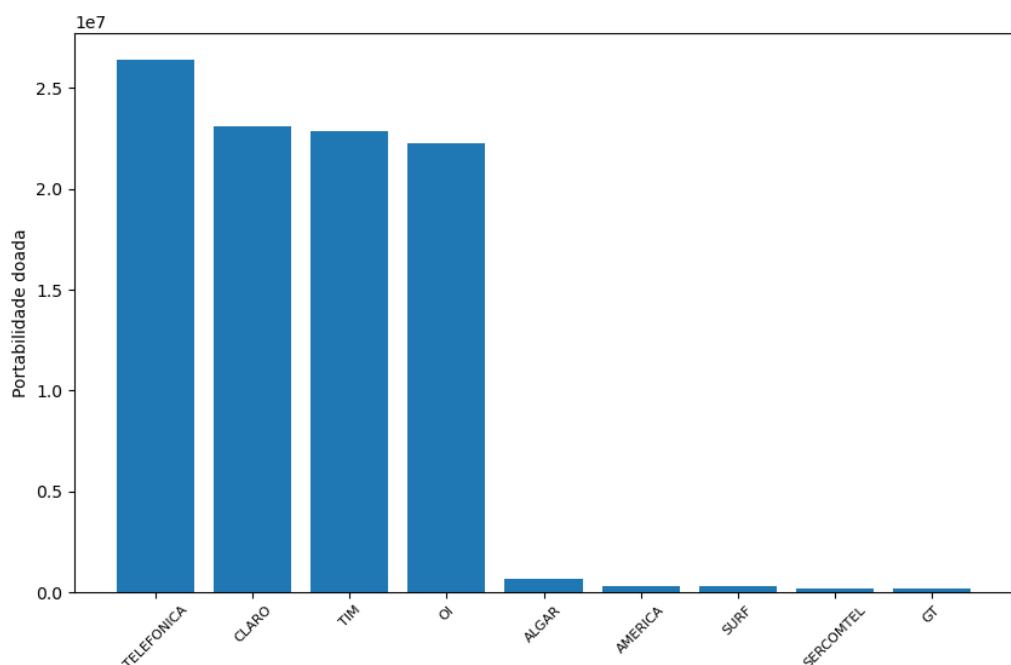


Figura 5 – operadoras que mais cedem cliente, 2008 a 2025
Fonte: própria do autor (2025)

As figuras 4 e 5 ajudam a compreender a dinâmica concentrada do mercado brasileiro de telecomunicações, que estão concentradas em quatro grandes empresas, Telefônica Brasil (Vivo), Claro S.A., Tim S.A. e Oi S.A. Por conta dessa concentração, foi realizado um estudo entre os canais de relação entre essas quatro operadoras telefônicas, ou seja, a quantidade de clientes doado e recebido entre elas.

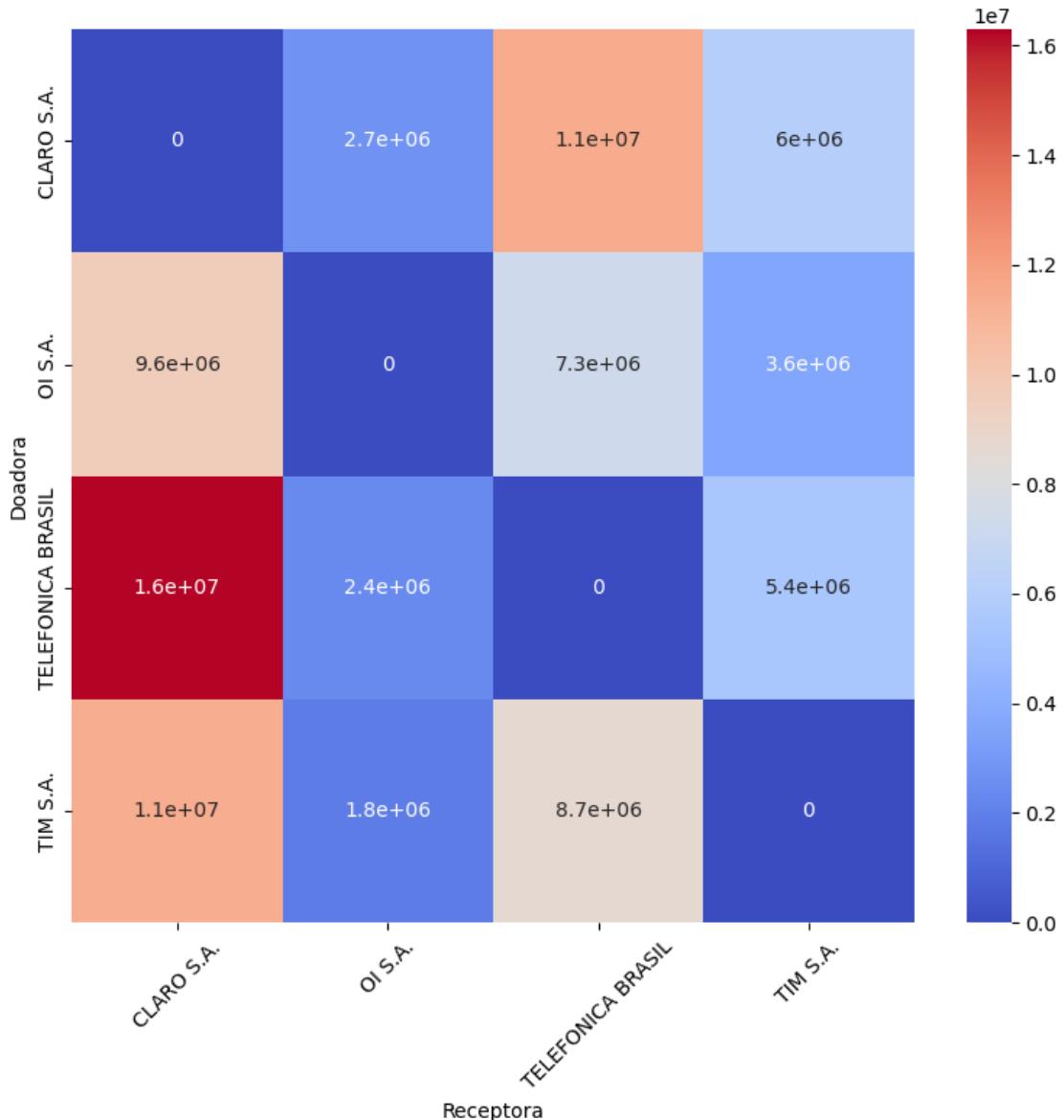


Figura 6 – mapa de calor das portabilidades entre maiores operadoras, 2008 a 2025
Fonte: própria do autor (2025)

A figura 6 demonstra que a Claro S.A. tem vantagem sobre as outras maiores operadoras brasileiras, ao passo que os maiores valores do mapa de calor estão concentrados nela, recebendo a maior parte das portabilidades das outras 3 operadoras.

Para buscar a explicação para essa diferença, foi analisado o conjunto num período menor, de 2020 a 2025, divididos na classificação de STFC e SMP.

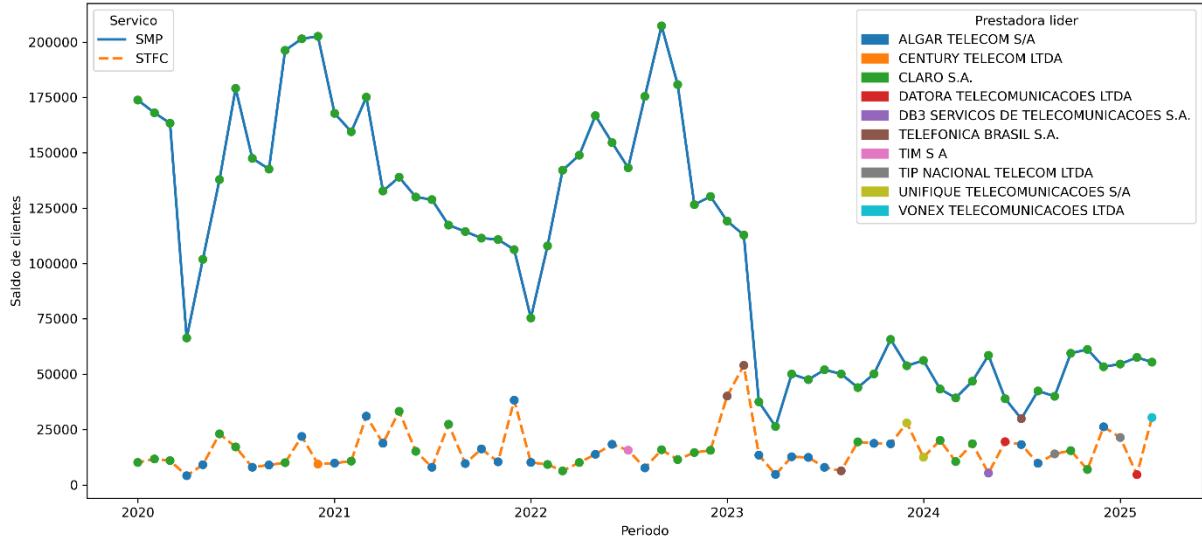


Figura 7 – prestadoras com maiores saldos no Brasil, STFC e SMP
Fonte: própria do autor (2025)

A figura 7 demonstrou como a dominância da Claro S.A. está consolidada, concentrada, principalmente no SMP, enquanto no STFC houve uma variabilidade de operadoras com maiores saldos mensais, o que destaca que esse mercado é mais competitivo, evidenciando, então, que o tipo de serviço influencia, diretamente, na operadora escolhida pelo cliente.

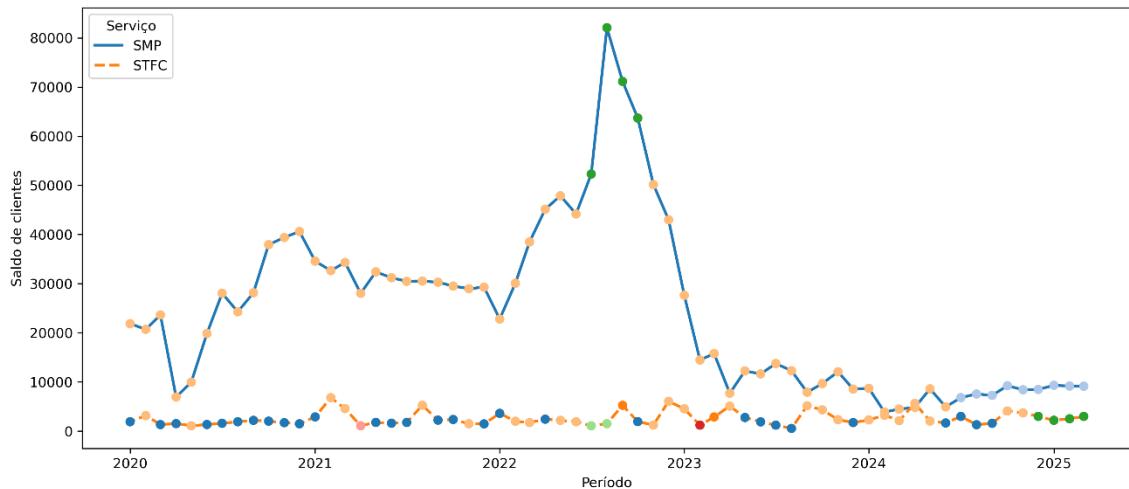


Figura 8 – prestadoras com maiores saldos na região Nordeste, STFC e SMP
Fonte: própria do autor (2025)

Outro fator que influencia a escolha da prestadora de serviço, certamente, é a regionalidade, o que é destacado pela figura 8, visto que a variabilidade de prestadoras no STFC

é maior, e apresenta uma dinâmica de concentração no ano de 2020 em uma operadora que não tem tanta expressividade como as quatro maiores, a Algar Telecom S.A. Já no SMP, a partir do segundo semestre de 2025 o domínio da Claro S.A. é contestado pela Brisanet Serviços de Telecomunicações, que ampliou sua cobertura 5G, tendo um alcance mais abrangente que outras operadoras em regiões mais distantes dos grandes centros urbanos (TELESINTESE, 2024). Em 2021, a Algar já estava com 90% de sua banda larga fixa ofertada por meio de fibra óptica (AMARAL, 2021), o que pode explicar a seus índices de portabilidade nesse período.

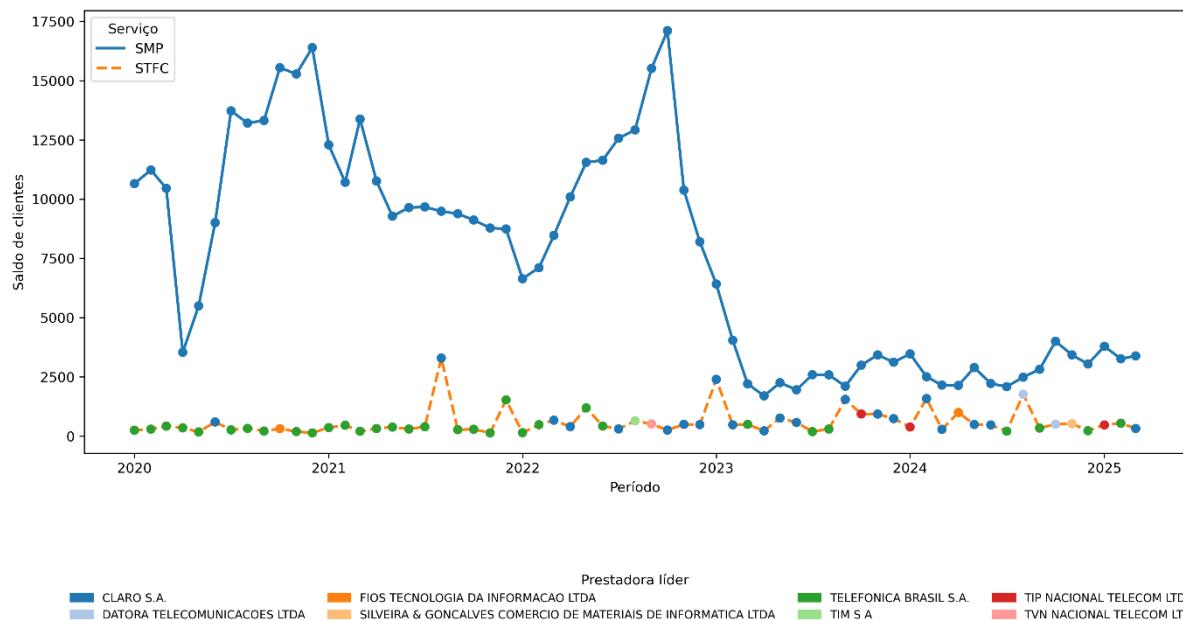


Figura 9 – prestadoras com maiores saldos na região Norte, STFC e SMP
Fonte: própria do autor (2025)

Na região Norte do país, a Claro S.A. continua com a maior parte das portabilidades no que tange ao SMP, porém, diferentemente do panorama nacional, no STFC, a ela não consegue ter uma grande quantidade de portabilidades, superada, na maioria dos meses no período analisado, pela Vivo.

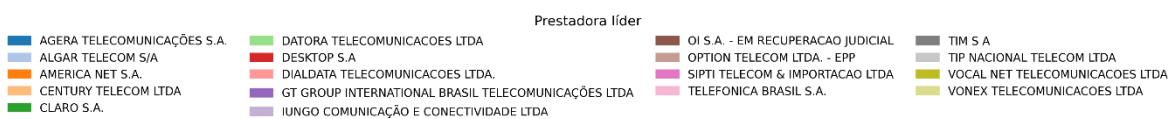
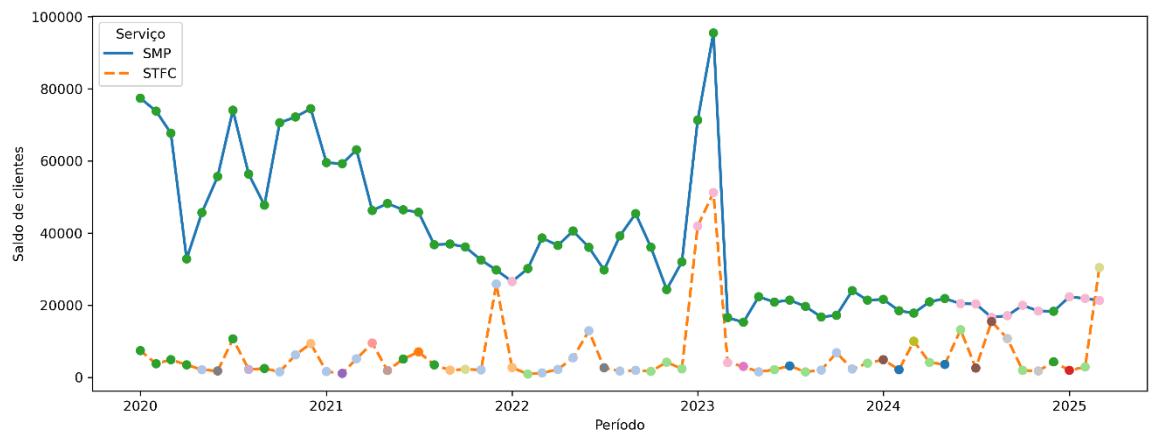


Figura 10 – prestadoras com maiores saldos no estado de São Paulo, STFC e SMP
Fonte: própria do autor (2025)

O mercado de telecomunicações no estado de São Paulo concentra cerca de 31,77% dos acessos à serviços STFC e 30,09% dos acessos de SMP, logo é imprescindível estudar o comportamento das portabilidades ocorridas no estado. A figura 10 demonstra, assim como as demais, o SMP concentrado na operadora Claro S.A., entretanto, o STFC é bem diversificado, o que demonstra uma alta competitividade regional.

Todas as análises demonstraram diferenciações no comportamento dos clientes, que são afetados pela operadora, tipo de serviço, e UF. Então os modelos de AM foram desenvolvidos visando uma tentativa de previsão comportamental da quantidade de portabilidade dada, que foi a variável que teve a maior correlação com as variáveis classificatórias. Os primeiros modelos, por serem simples modelos de regressão linear, tiveram uma efetividade muito baixa, avaliada pelo método pelo coeficiente de determinação.

Modelo	LinearRegression	Ridge	Lasso
Desempenho (%)	4,7072	4,7072	4,7099

Tabela 1 – avaliação dos modelos de regressão pelo método r-quadrado.
Fonte: própria do autor

Os modelos avaliados na tabela 1 têm uma medida muito baixa, visto que apenas 4,7% da variabilidade no alvo está sendo explicada pelo modelo. Após isso, foi realizado o tratamento de intervalo interquartil, retirando os valores extremos que podiam estar influindo negativamente a previsibilidade do modelo.

Modelo	LinearRegression	Ridge	Lasso
Desempenho (%)	6,49	6,49	6,49

Tabela 2 – avaliação dos modelos após o tratamento de IQR, pelo método r-quadrado
Fonte: própria do autor

Após o tratamento, o índice melhorou de maneira sutil, como se pode observar na tabela 2. Por conta disso, um modelo que trabalha com método de conjuntos foi construído visando aumentar a eficiência das previsões realizadas. Utilizando o *XGBoostRegressor* foi possível alcançar uma previsibilidade de 25%, sendo consideravelmente mais alta que os modelos tradicionais, mas ainda inviável para previsão de fenômenos reais.

Nesse cenário, foi com a mudança para um modelo classificador, agora espera-se não mais a previsão de uma possível quantidade de portabilidade dada, mas sim um comportamento. Pelo método do cotovelo utilizando *KMeans*, foi possível avaliar que a quantidade de classes ideais para serem trabalhadas nesse modelo era igual à três. Desse modo, a cada linha que constava no banco de dados de saldo de portabilidade atribui-se um número, variando de zero à dois, que divide esses dados nesses três grupos.

Agrupamento	portabilidade doada (média)	portabilidade recebida (média)	Serviço (moda)	Operadora (moda)	Unidade Federativa (moda)
0	-0,992353	-0,992422	1	0,705	0,923077
1	-0,971436	-0,971012	-1	-0,670	0,923077
2	-0,993105	-0,989532	1	-0,670	0,923077

Tabela 3 – valores estatísticos de cada agrupamento
Fonte: própria do autor (2025)

A tabela 3 representa os dados estatísticos que definem cada grupo, para as variáveis numéricas, foi utilizado a média dos valores, enquanto para as categóricas utilizou-se a moda. O novo campo, denominado de agrupamento, foi utilizado, então, como a variável alvo de um modelo classificador, com os parâmetros de usar 80% dos dados para treino e 20% para teste, para que assim, o modelo tentasse aprender a como classificar portabilidade em um dos agrupamentos.

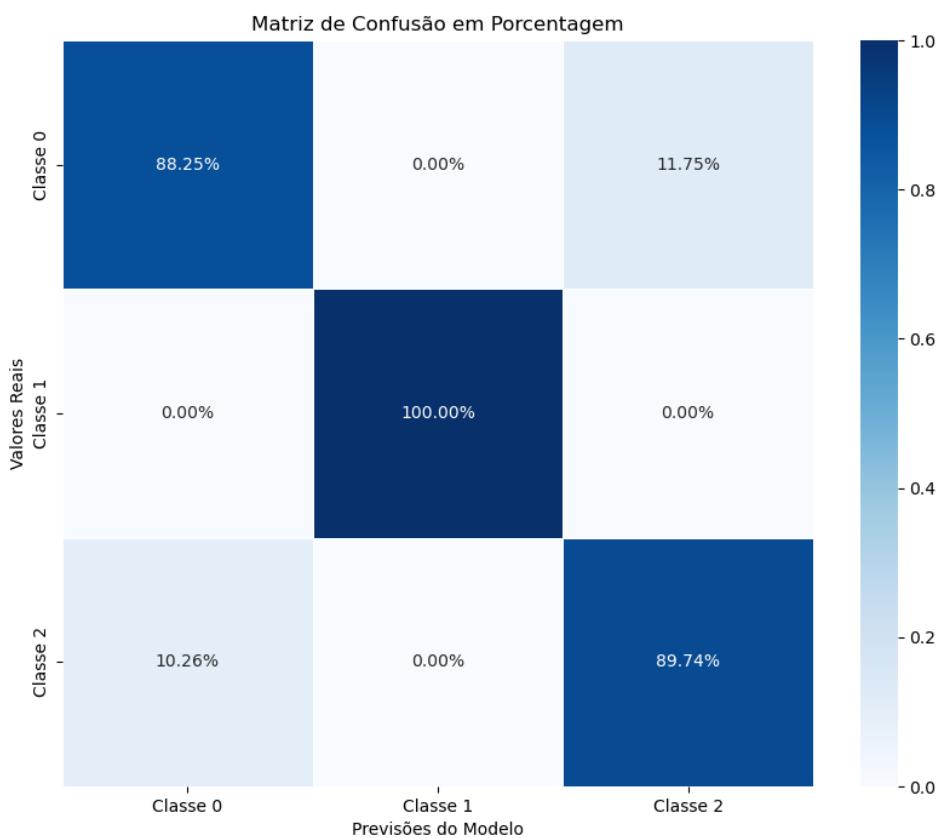


Figura 11 – Matriz de confusão dos agrupamentos
Fonte: própria do autor (2025)

A partir da construção desse modelo de AM não supervisionado, chegou-se em um resultado definitivamente mais assertivo que os modelos de regressão, que atingiram acerto máximo em 25%. Na figura 11 pode-se perceber que para as classes zero e dois foi obtido ótima faixa de acerto, acima de 85%, e para a classe um, o acerto foi de 100%, o que demonstra que para essa base de dados, o modelo classificador se ajustou de maneira melhor que o de regressão.

CONCLUSÕES E SUGESTÕES

Então, é visto que as portabilidades podem ser alteradas por diversos fatores externos aos trazidos pela base de dados. Outrossim, este presente estudo permitiu concluir que um fator determinante no número de portabilidades realizadas são o tipo do serviço, SMP e STFC, e a localidade onde ele está sendo prestado, de tal modo que foram construídos diversos modelos de aprendizado de máquina, buscando uma previsibilidade nos números de portabilidades doadas. Porém, o conjunto de dados analisado não permitiu tal abordagem, sendo, assim, possível contornar o problema da falta de correlações fortes na montagem de um algoritmo. Tal classificação, possibilita a divisão de novos dados em grupos, que podem ter seus fatores socioeconômicos estudados, para que assim, possam ser utilizados na tomada de decisões das prestadoras de serviço, bem como os órgãos reguladores.

REFERÊNCIAS BIBLIOGRÁFICAS

BENGFORT, B.; BILOGUR, A.; et al. **Elbow Visualizer**. In: _____. **Yellowbrick: Machine Learning Visualization**. [S. l.], 2024. Disponível em: <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>.

BRASIL. Ministério das Comunicações. **Portabilidade numérica**. Dados.gov.br, [s.d.]. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/portabilidade-numerica>.

AMARAL, Bruno do. **Algar Telecom já tem 90% da base de banda larga fixa em fibra**. *Teletime News*, São Paulo, 9 nov. 2021. Disponível em: <https://teletime.com.br/09/11/2021/algar-telecom-ja-tem-90-da-base-de-banda-larga-fixa-em-fibra/>.

G1. Dona da Claro fecha acordo para comprar Nextel Brasil. G1 Economia, 18 mar. 2019. Disponível em: <https://g1.globo.com/economia/noticia/2019/03/18/dona-da-claro-fecha-acordo-para-comprar-nextel-brasil.ghtml>.

INVESTE SP – Agência Paulista de Promoção de Investimentos e Competitividade. **Telecomunicações**. *Investe SP*, São Paulo, 2020. Disponível em: <https://www.investe.sp.gov.br/por-que-sp/infraestrutura/telecomunicacoes/>.

LEITE, João Paulo. **Aplicação de Técnicas de Aprendizado por Reforço à Alocação de Recursos e ao Escalonamento de Usuários em Sistemas de Telecomunicações**. 2014. 131 f. Tese (Doutorado em Engenharia Elétrica) – Universidade de Brasília, Faculdade de Tecnologia, Departamento de Engenharia Elétrica, Brasília, 2014. Disponível em: <https://core.ac.uk/download/pdf/33549108.pdf>.

MARTINS JÚNIOR, Carlos Alberto. **Classificação de índices de satisfação de usuários de telefonia móvel usando aprendizado de máquinas**. 2020. 62 f. Dissertação (Mestrado em Engenharia Eletrônica) - Universidade do Estado do Rio de Janeiro, Faculdade de Engenharia, Rio de Janeiro, 2020. Disponível em: https://www.pel.uerj.br/wp-content/uploads/2021/04/Dissertacao_Carlos_Alberto_Martins_Junior.pdf.

PEDREGOSA, F. et al. **Ensemble methods**. In: _____. **Scikit-learn: Machine Learning in Python**. [S. l.], 2024. Disponível em: <https://scikit-learn.org/stable/modules/ensemble.html>.

RAY, S. **How, When, and Why Should You Normalize / Standardize / Rescale Your Data? Towards AI**, [S. l.], 2020. Disponível em: <https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff>.

SANTOS, M. M. **Dependência Linear**. Campinas: Instituto de Matemática, Estatística e Computação Científica (IMECC) - UNICAMP, [s.d.]. Disponível em: https://www.ime.unicamp.br/~marcia/AlgebraLinear/dependencia_linear.html.

SILVA, Jones Granatyr. **Por que a linguagem Python é tão popular em Machine Learning e Inteligência Artificial?** DS Academy, 2021. Disponível em: <https://blog.dsacademy.com.br/por-que-a-linguagem-python-e-tao-popular-em-machine-learning-e-inteligencia-artificial/>.

TELECO. **Legislação de Telecomunicações – Regulamentação e Normas**. Teleco, [s.d.]. Disponível em: https://www.teleco.com.br/tutoriais/tutorialleg/pagina_4.asp.

TELESÍNTESE. **Em lançamento oficial, Brisane mira ampla cobertura 5G no Nordeste**. [S. l.], 28 abr. 2022. Disponível em: <https://telesintese.com.br/em-lancamento-oficial-brisanet-mira-ampla-cobertura-5g-no-nordeste/>.

TERESA, B. L. **Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências**. *Estudos Avançados*, 2021. Disponível em: <https://doi.org/10.1590/s0103-4014.2021.35101.007>.