



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
IFCE
IFCE CAMPUS FORTALEZA
CURSO DE MESTRADO EM ENGENHARIA DE TELECOMUNICAÇÕES

JONES MONTEIRO JACINTO

MODELAGEM DE AMPLIFICAÇÃO RAMAN DISTRIBUÍDA DE PRIMEIRA E
SEGUNDA ORDEM UTILIZANDO APRENDIZADO DE MÁQUINA

FORTALEZA
2024

JONES MONTEIRO JACINTO

MODELAGEM DE AMPLIFICAÇÃO RAMAN DISTRIBUÍDA DE PRIMEIRA E
SEGUNDA ORDEM UTILIZANDO APRENDIZADO DE MÁQUINA

Dissertação apresentada à Coordenação de Pós- Graduação em Engenharia de Telecomunicações do Instituto Federal do Ceará (IFCE) - Campus Fortaleza, como requisito parcial para obtenção do grau de Mestre em Engenharia de Telecomunicações. Área de concentração: Micro-Ondas e Óptica Integrada.

Orientador: Prof. Dr. Glendo de Freitas Guimarães.

Co-orientador: Dr. Yosdan Martínez Camejo.

FORTALEZA

2024

JONES MONTEIRO JACINTO

MODELAGEM DE AMPLIFICAÇÃO RAMAN DISTRIBUÍDA DE PRIMEIRA E
SEGUNDA ORDEM UTILIZANDO APRENDIZADO DE MÁQUINA

Dissertação apresentada à Coordenação de Pós-Graduação em Engenharia de Telecomunicações do Instituto Federal do Ceará (IFCE) - Campus Fortaleza, como requisito parcial para obtenção do grau de Mestre em Engenharia de Telecomunicações. Área de concentração: Micro-Ondas e Óptica Integrada. Aprovado pela comissão formada pelos professores:

Aprovado em ____/____/____

BANCA EXAMINADORA

Prof. Dr. Glendo de Freitas Guimarães
(Orientador)
PPGET, Campus Fortaleza (IFCE)

Prof. Dr. Aderaldo Irineu Levartoski de Araújo
Campus Fortaleza (IFCE)

Prof. Dr. Jose Rubens Rodrigues de Sousa
Universidade de Fortaleza. (UNIFOR)

Dedico este trabalho a todos o PPGETI
que me auxiliaram nessa jornada de
escrita dessa dissertação.

AGRADECIMENTOS

Antes de tudo agradeço a Deus por tudo de maravilhoso que ocorreu e venha a ocorrer na minha vida, além da conclusão desse trabalho.

Meus Agradecimentos a minha mãe, dona Terezinha, que sempre ao meu lado me confortava e a minha família que é o alicerce principal na minha vida.

Agradeço ao meu orientador, Glendo de Freitas Guimarães, depois de tanto tempo que nos conhecemos ainda solícito e prestativo na orientação no mestrado.

Agradeço também ao meu coorientador Yosdan Martinez Camejo, pela força e perseverança que guiou esse trabalho além de compartilhar conhecimento de forma prestativa.

Aos participantes da banca examinadora e a todos os professores que compartilharam o seu conhecimento e ampliaram a minha capacitação profissional.

Aos demais que me auxiliaram neste trabalho que não se sintam menosprezados por não serem citados, mas que estimo um agradecimento enorme pelo auxílio.

"Damos-te graças, ó Deus, damos-te graças, pois perto está o teu nome; todos falam dos teus feitos maravilhosos. Olhe à sua volta, veja quanta vida e beleza Deus criou para a nossa alegria."

(Bíblia Sagrada, Salmo 75, 1)

RESUMO

O presente estudo combina simulações de Inteligência Artificial (IA) com sinais ópticos, utilizando amplificação Raman, com objetivo de aprimorar a transmissão de dados em sistemas de comunicação de fibras ópticas. Para isso, usa-se algoritmos como Decision Tree Regressor (DTR), que é baseado em uma só árvore de decisão onde os nós são recursivamente segmentados de acordo com os valores das features e apesar de ter uso recomendável para datasets de baixa dimensão, tende a um overfitting em datasets de maior dimensão; Random Forest Regressor (RFR), que soluciona o problema de construção de várias árvores em subconjuntos randômicos do dataset e depois faz uma média de previsões de cada árvore, e tem uso recomendável para datasets de baixo ou alto grau de dimensão; Gradient Boosting Regressor (GBR), que constrói árvores de decisão serializadas, sendo que cada árvore de decisão é melhorada, e seu uso é recomendável para datasets de maior grau de dimensão mas o seu processamento pode ser elevado; Histogram Boosting Regressor (HBR), versão melhorada do GBR que usa histogramas para segmentar os valores das variáveis em dois grupos e tem uso recomendável para datasets maiores e seu consumo computacional é pequeno para o processamento das features; e AdaBoost Regressor (ABR), que é um algoritmo de boosting que tenta corrigir os erros de cálculos dos pesos anteriores para recalculer os dados de entrada e seu uso é recomendável para uma abordagem mais leve do processamento das features mas é mais susceptível a outliers. A simulação de fibras ópticas para amplificação do sinal desempenha um papel crucial na regeneração do sinal nesses sistemas. O uso de algoritmos de inteligência artificial oferece uma abordagem inovadora para resolver desafios complexos associados ao ajuste dos parâmetros necessários para obtenção da amplificação Raman em diversos arranjos, com o intuito de obter os respectivos conjuntos de dados. Por meio do desenvolvimento de modelos preditivos, os algoritmos de inteligência artificial possibilitam o ajuste fino das características das fibras ópticas para obter a respectiva amplificação, em que é requerida uma combinação de fatores para que se ocorra a amplificação para o efeito Raman. Essa integração entre simulação de fibras ópticas para amplificação do sinal e algoritmos de inteligência artificial representa um avanço significativo na criação de sistemas de comunicação óptica mais eficazes e adaptáveis. Os resultados obtidos demonstram que, tanto o amplificador Raman de primeira ordem como o de segunda ordem, possuem viabilidade de implementação com o auxílio dos algoritmos de aprendizado de máquina, tendendo a escolha do Decision Tree Regressor (DTR) no amplificador de primeira ordem e os algoritmos Decision Tree Regressor (DTR) e Random Forest Regressor (RFR) no amplificador Raman de segunda ordem.

Palavras-chave: Amplificação Raman. Inteligência Artificial. Fibra Óptica. Machine Learning.

ABSTRACT

This study combines Artificial Intelligence (AI) simulations with optical signals using Raman amplification in order to improve data transmission in optical fiber communication systems. To this end, algorithms such as Decision Tree Regressor (DTR), which is based on a single decision tree where the nodes are recursively segmented according to the values of the features and its use is recommended for low dimensional datasets but tends to overfitting in higher dimensional datasets, Random Forest Regressor (RFR), which solves the problem of building several trees in random subsets of the dataset and then averages the predictions of each tree and its use is recommended for low or high dimensional datasets, Gradient Boosting Regressor (GBR), which builds serialized decision trees with each decision tree being improved and its use is recommended for datasets with a higher degree of dimension but its processing can be high, Histogram Boosting Regressor (HBR), which is an improved version of GBR that uses histograms to segment the values of the variables into two groups and its use is recommended for datasets with a higher degree of dimension. in two groups and its use is recommended for larger datasets and its computational consumption is small for processing the features, and AdaBoost Regressor (ABR) , which is a boosting algorithm that tries to correct the calculation errors of the previous weights to recalculate the input data and its use is recommended for a lighter approach to processing the features but it is more susceptible to outliers. The simulation of optical fibers for signal amplification plays a crucial role in signal regeneration in these systems. The use of artificial intelligence algorithms offers an innovative approach to solving complex challenges associated with adjusting the parameters required to obtain Raman amplification in various arrangements in order to obtain the respective data sets. Through the development of predictive models, artificial intelligence algorithms make it possible to fine-tune the characteristics of optical fibers in order to obtain the respective amplification, in which a combination of factors is required for amplification to occur for the Raman effect. This integration between optical fiber simulation for signal amplification and artificial intelligence algorithms represents a significant advance in the creation of more efficient and adaptable optical communication systems. The results obtained show that both the first-order and second-order Raman amplifiers are feasible to implement with the help of machine learning algorithms, with the Decision Tree Regressor (DTR) algorithm tending to be chosen for the first-order amplifier and the Decision Tree Regressor (DTR) and Random Forest Regressor (RFR) algorithms for the second-order Raman amplifier.

Keywords: Raman amplification. Artificial Intelligence. Machine Learning. Fiber optic.

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivos	14
1.1.2	Específicos	14
2	MATERIAIS E MÉTODOS	15
2.1	Amplificação Raman	15
2.1.1	Não linearidade	15
2.1.2	Espalhamento Raman	23
2.1.3	Propriedades vibracionais dos materiais	25
2.1.4	Espalhamento inelástico da luz e interação com a matéria	26
2.1.5	Espalhamento no contexto de amplificação Raman	29
2.1.6	Amplificadores Raman de primeira ordem	37
2.1.7	Amplificadores Raman de segunda ordem	41
2.1.8	Figuras de mérito	42
2.1.8.1	Relação sinal ruído óptica (OSNR)	43
2.1.8.2	Fator de qualidade de transmissão óptica	43
2.1.8.3	Taxa de erros de bit do enlace óptico	44
2.2	Aprendizado de máquina	44
2.2.1	Decision Tree Regressor	46
2.2.2	AdaBoost Regressor	47
2.2.3	Histogram Gradient Booster Regressor	47
2.2.4	Gradient Boosting Regression	48
2.2.5	Random Forest Regression	49
2.2.6	Análise de correlação	50
2.2.6.1	Coeficiente de correlação de Pearson	51
2.2.6.2	Coeficiente de correlação de postos de Spearman	51
2.2.6.3	Coeficiente de correlação de Kendall	52
2.2.7	Análise de performance do modelo	53
2.2.7.1	Coeficiente de determinação(R^2)	54
2.2.7.2	Erro médio absoluto	55
2.2.7.3	Erro médio quadrático	56
3	METODOLOGIA	58
3.1	Metodologia de simulação do amplificador Raman	58
4	RESULTADOS	60
4.1	Amplificação Raman de primeira ordem	60

4.1.1	Matriz de correlação	63
4.1.2	Performance dos modelos	65
4.2	Amplificação Raman de segunda ordem	79
4.2.1	Matriz de correlação	79
	REFERÊNCIAS	97

1 INTRODUÇÃO

O princípio do efeito Raman, ou espalhamento Raman, se baseia na dispersão inelástica da luz quando interage com moléculas ou materiais cristalinos. Estes efeitos foram descobertos pelo renomado cientista indiano C.V. Raman em 1928 (Raman; Krishnan, 1928), e ocorre quando fótons de luz incidente colidem com as vibrações moleculares ou da rede cristalina, resultando em uma mudança na energia dos fótons emergentes. Essa mudança na energia leva à emissão de luz espalhada com frequência diferente da luz incidente, revelando informações valiosas sobre a estrutura molecular, composição química e propriedades vibracionais das substâncias analisadas. O efeito de amplificação é obtido quando determinada frequência do laser de bombeamento é sintonizado com relação ao sinal óptico incidente.

A amplificação Raman foi descoberta por acidente em 1962, durante experimento com laser de Rubi de comutação-Q (Woodbury; Ng, 1962). Um laser de comutação-Q introduz a perda do sinal óptico na cavidade de ressonância com intuito de obter as inversões de população. Logo após, desliga a perda para produzir picos de intensidade do sinal muito alto. O sistema montado no experimento continha um polarizador e um aparato de birrefringência induzida por campo elétrico em uma célula de nitrobenzeno. O resultado encontrado foi o surgimento de uma intensa saída no comprimento de onda, característico do efeito Raman deslocado da célula de nitrobenzeno.

Os cientistas R. H. Stolen e E. P. Ippen, nos laboratórios Bell Telephone Laboratories, no ano de 1973, realizavam estudos de espalhamento Raman e seus efeitos em fibras ópticas, em que pudesse aplicar a não-linearidade do sinal óptico nas fibras de silício. Os resultados obtidos foram calibrados junto aos obtidos por meio do laser de Rubi de comutação-Q, anteriormente descrito. Chegando à conclusão de que o meio da própria fibra óptica serve para a amplificação do sinal incidente. Dessa forma, descreveu um modelo de amplificação Raman e, com essa configuração, em que a própria fibra é utilizada como meio para amplificação, o chamou de Amplificador Distribuído Raman (DRA) (Stolen; Ippen, 1973).

Até o final da década de 90, os amplificadores ópticos, principalmente os EDFAs (Erbium Doped Fiber Amplifiers), foram objetos de intenso estudo. Com o tempo, a demanda por novas tecnologias de amplificação cresceu, buscando atender diferentes requisitos e superar as limitações dos amplificadores ópticos de estado sólido, tais como a necessidade de repetição de amplificação a altas distâncias.

Porém, o alto custo de aplicação, com a necessidade de instalar repetidores na segmentação de rede óptica, torna a operação financeiramente dispendiosa e esse cenário abre caminho para o desenvolvimento de amplificador usando o espalhamento Raman como principal tecnologia, o que se apresenta como uma alternativa mais econômica.

Uma característica notável do efeito Raman é a possibilidade de amplificar o sinal que é transmitido no meio sem nenhuma necessidade de conversão eletro-óptica, ao contrário da tecnologia do EDFA, tornando a sua magnitude maior e facilitando sua detecção pelo fotodetector na chegada do sistema de comunicação óptica. A amplificação do efeito Raman é frequentemente alcançada por meio de várias técnicas. Esse tipo de amplificação ocorre dentro do meio guiado, no caso, a fibra óptica. Nessa abordagem, as moléculas de interesse são absorvidas pelo meio e a interação com os fótons gera uma amplificação do sinal óptico.

O estudo sobre o espalhamento aplicado na amplificação Raman é exposto sistematicamente em diversos artigos. Em Kuzin (1999), é investigada a respectiva aplicação de um enlace de 100 metros e a análise das componentes necessários para amplificação Raman e seu agrupamento com as fontes de bombeamento, que vão além de uma fonte de bombeamento, que é descrito como amplificador Raman de primeira ordem.

Segundo Smith (1972), o efeito de espalhamento Raman estimulado detalha o manuseio da energia do pulso óptico na fibra de sílica. A baixa perda é um cenário a ser obtido, pois é necessário buscar condições favoráveis para propagação. Uma potência crítica é desejável para operacionalizar o Espalhamento Raman Estimulado propagante e contrapropagante. Tal potência crítica é definida pela área efetiva do núcleo da fibra e o cálculo de diversos parâmetros, para buscar a efetividade do Espalhamento Raman Estimulado.

Decorrente ao exposto por Bromage (2004), a amplificação Raman possibilitou aumentos significativos no alcance e na capacidade de sistemas ópticos. O artigo expõe o motivo, as propriedades fundamentais do ganho do Espalhamento Raman Estimulado e o respectivo acúmulo de ruído da emissão espontânea amplificada. Outras fontes de ruído são catalogadas e expostas. Também são expostos novos esquemas de bombeamento Raman recentemente desenvolvidos.

Atualmente, o que se pode adquirir nas pesquisas é o arranjo ideal para amplificadores Raman de segunda ordem e sua combinação com a primeira ordem,

dispostas na fibra para fibras de longa distância, sem a necessidade de amplificadores de bombeamento no meio de transmissão. Tendo em vista esse desafio de conciliar o respectivo controle da não linearidade do sinal óptico gerado pelas altas potências do feixe, é necessário ter um controle para que o respectivo sistema possa se adaptar para amplificar o respectivo sinal óptico.

Por outro lado, as configurações de amplificação Raman necessitam de constantes ajustes para seu devido funcionamento, tais como um fino controle da potência do sinal e da frequência do laser de bombeamento. Para otimizar esse ajuste, podem ser aplicados algoritmos de inteligência artificial (IA), sendo os métodos usados no aprendizado de máquina (ML- *Machine Learning*) uma alternativa viável para a solução do problema em questão. Vale a pena salientar que o ML é uma ramificação da inteligência artificial baseada na ideia de que padrões e tendências em um conjunto de dados podem ser obtidos por meio de determinado modelo.

O aprendizado de máquina não é uma área nova, uma vez que algoritmos relacionados ao ML existem desde pelo menos a década de 1970. No entanto, um aumento significativo na capacidade computacional, ao longo da última década, tem contribuído para o surgimento da moderna tecnologia de aprendizado profundo (*Deep Learning*), uma classe de abordagens avançadas de ML que demonstra um desempenho superior dos algoritmos de aprendizado convencionais.

Este trabalho está estruturado em cinco (5) capítulos, onde o primeiro é a introdução, no segundo são apresentados os materiais e métodos utilizados, onde serão apresentadas as teorias relacionadas aos efeitos não lineares, de espalhamento Raman e o efeito de amplificação Raman. Também serão descritos os algoritmos de ML que serão utilizados no trabalho. Dentre os algoritmos de ML utilizados temos os seguintes: Decision Tree Regressor (DTR), Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Histogram Boosting Regressor (HBR) e AdaBoost Regressor (ABR).

O terceiro capítulo está dedicado à metodologia, onde serão descritos os arranjos de simulação no software *Optisystem 7* para amplificação Raman. O quarto capítulo trará a apresentação dos resultados, onde serão expostos os resultados da simulação de amplificação Raman, além de expor os gráficos desse conjunto de dados e descrever a matriz de correlação. Ainda nos resultados será exposta a interpretação dos resultados obtidos mediante a aplicação dos algoritmos de ML,

utilizando como parâmetros de performance R^2 , MSE (*Mean Squared Error*) e o MAE (*Mean Absolute Error*).

1.1 Objetivos

1.1.1 Geral

O objetivo geral do estudo é aplicar a inteligência artificial na amplificação do sinal óptico com efeito Raman, visando aprimorar o aumento do alcance do sinal óptico e a eficiência do processo de transmissão do sinal óptico, com o intuito de otimizar a respectiva amplificação do sinal no meio de transmissão.

1.1.2 Específicos

- Analisar o efeito Raman de primeira ordem;
- Analisar o fator de qualidade do efeito Raman de primeira ordem variando a potência do laser de bombeamento e o comprimento da fibra óptica;
- Analisar o efeito Raman de segunda ordem;
- Analisar o fator de qualidade do enlace variando a potência do laser de bombeamento 1 e laser de bombeamento 2 e o comprimento da fibra óptica.
- Treinar e testar o algoritmo de inteligência artificial de primeira ordem;
- Treinar e testar o algoritmo de inteligência artificial de segunda ordem; Além disso, serão calculadas utilizando as seguintes correlações:
 - de Pearson;
 - de Spearman;
 - de Kendall.

Com a produção da base de dados dos cenários a serem simulados de primeira e segunda ordem, posteriormente será abordada a eficácia dos algoritmos de aprendizado de máquina. Além disso, será feito um comparativo entre os algoritmos e o seu desempenho no que tange à obtenção do modelo de treinamento e de sua respectiva validação para os devidos testes.

Por outro lado, o processo de simulação será integrado com algoritmos de IA capazes de otimizar os parâmetros experimentais obtidos das simulações para maximizar a amplificação do sinal Raman, bem como a criação de modelos preditivos que permitam ajustar o comportamento do efeito Raman em diferentes cenários.

Vale ressaltar que o estudo visa auxiliar a utilização a IA para automatizar a análise de grandes conjuntos de dados gerados pelo cenário proposto na simulação de amplificação Raman de primeira e segunda ordem, identificando padrões e tendências que podem ser facilmente perdidos em análises manuais no respectivo ajuste de parâmetros, tais como potência e comprimento de onda do sinal óptico. O objetivo final é aprimorar a aplicação prática do efeito Raman em uma vertente de amplificação do sinal óptico, com o escopo de estruturar sistemas de uso do efeito Raman e automatizar o ajuste dos parâmetros dependentes a eficiente amplificação.

2 MATERIAIS E MÉTODOS

2.1 Amplificação Raman

A amplificação Raman baseia-se no efeito de espalhamento do feixe óptico, um fenômeno em que a luz dispersa por uma amostra apresenta mudanças na frequência em relação à luz incidente. Essas mudanças são diretamente relacionadas às vibrações moleculares da amostra, fornecendo informações valiosas sobre sua composição molecular. No entanto, o sinal resultante da interação junto ao meio possui magnitude pequena em relação ao feixe incidente, dificultando sua detecção em muitas aplicações práticas. Os alicerces da amplificação Raman utilizando espalhamento estão nos estudos dos efeitos não lineares e seu respectivo processo de espalhamento.

2.1.1 Não linearidade

A ampla popularização do laser com diversas aplicações no cotidiano da humanidade, e a restrição do assunto para o setor de telecomunicações foi produto da necessidade do uso de maior intensidade e altas frequências necessárias nos feixes ópticos, em comparação aos produzidos por quaisquer fontes convencionais em que utilizavam sinais elétricos. O estudo das interações não lineares iniciou com a aplicação de cenários que necessitavam de alta potência em seu sinal incidente.

Quando a intensidade do campo correspondente atinge um nível da ordem de 100 kV/m ou mais, materiais que possuem uma resposta linear ao sinal óptico em níveis mais baixos, a resposta do sinal no meio, que na maioria das vezes nas comunicações ópticas é a fibra óptica, os respectivos níveis de intensidade da luz

podem se tornar “não lineares”, no sentido de que as “constantes” ópticas não são mais “constantes” independentes da intensidade da luz. Como consequência, quando o campo não é fraco, a suscetibilidade óptica χ e a constante dielétrica correspondente ϵ do meio podem se tornar funções do campo elétrico, respectivamente.

Tal dependência de campo nos parâmetros ópticos do material pode originar uma ampla gama de fenômenos ópticos não lineares e esses fenômenos serem utilizados para uma grande variedade de novas aplicações (Bass et al., 1995). Segundo Bass et al. (1995), quando a intensidade do pulso óptico não possui grande magnitude, a dependência da polarização com as componentes de ordem superior do campo elétrico pode ser negligenciada.

Segundo Gangopadhyay; Kumbhakar; Mandal (2019) quando esta condição não é cumprida por parte da polarização, cuja variável é P , na qual é gerada no meio dielétrico não linear e pode ser descrita a seguir:

$$P = \epsilon_0 \chi E, \quad (2.1)$$

sendo ϵ_0 a permissividade no espaço livre, χ a suscetibilidade e E o respectivo campo elétrico.

O desenvolvimento da equação leva em conta que χ não é uma constante, mas dependente do campo elétrico E , portanto:

$$\chi(E) = \chi_1 + \chi_2 \cdot E + \chi_3 \cdot E^2 + \dots \quad (2.2)$$

A fórmula de polarização possui a seguinte configuração:

$$P = \epsilon_0 (\chi_1 \cdot E + \chi_2 \cdot E^2 + \chi_3 \cdot E^3 + \dots) \quad (2.3)$$

podendo ser decomposta também assim:

$$P = P^{(1)} + P^{(2)} + P^{(3)} + \dots \quad (2.4)$$

Generalizando os cenários, $\chi_0 \gg \chi_1 \gg \chi_2 \dots$, tendendo que o segundo termo $\epsilon_0 \chi_1 E^2$ e o terceiro termo $\epsilon_0 \chi_2 E^3$ são muito menores que o primeiro termo. Portanto para atender a maioria dos casos tem obtém a seguinte equação:

$$\chi = \chi_0 + \chi_1 E, \quad (2.5)$$

Onde ϵ_0 é a permissibilidade elétrica do vácuo, e podemos observar que a polarização tem componentes dependentes de ordem superior.

Como a fibra óptica é composta de óxido de silício (SiO_2), nesta configuração podem ser desprezados os efeitos de segunda ordem observados na equação da polarização 2.4, tendo como valores predominantes o de primeira e terceira ordem para o cálculo da não linearidade (Boyd; Gaeta; Giese, 2008). Portanto a seguinte fórmula é utilizada na maioria dos casos levando em consideração a dependência da polarização dos efeitos de maiores ordens:

$$P = \epsilon_0 \chi_0 E^3 + \epsilon_0 \chi_1 E^2. \quad (2.6)$$

Os efeitos de terceira ordem compõem os respectivos efeitos não lineares. Também é necessário calcular o índice de refração do meio ($\bar{n}$), a frequência (ω) da luz incidente na presença da não linearidade e o respectivo módulo do campo elétrico, tendo em vista que o estudo do comportamento perante a não linearidade deva ser dimensionado por meio desses parâmetros, que é calculado abaixo:

$$\bar{n}(\omega, |E|^2) = \bar{n}(\omega) + n_2 |E|^2 \quad (2.7)$$

Tal equação define o respectivo coeficiente de não linearidade (n_2) e é dependente do meio em que incide o feixe óptico que define as características. Conforme Agrawal (2019), para calcular o coeficiente de não linearidade será utilizada

a terceira componente da equação 2.2 relativo a $\chi^{(3)}$, que está ligada a não linearidade.

$$n_2 = \frac{3}{8n} \chi^{(3)} \approx 2,3 \times 10^{-22} \frac{m^2}{v^2} \quad (2.8)$$

A listagem de comparação do respetivo coeficiente de não linearidade para o vidro é em torno de $2,3 \times 10^{-22} m^2/v^2$. Tal valor é muito pequeno. Para que o respectivo efeito possa interagir junto com a fibra óptica, será necessário escalar o efeito de não linearidade anteriormente exposto.

A análise do dimensionamento da não linearidade torna-se mais compreensível ao realizar o cálculo de sua figura de mérito. Este cálculo será desenvolvido considerando dois cenários distintos, que ilustram diferentes formas de manifestação da não linearidade: o primeiro cenário avalia a incidência da não linearidade em um meio de propagação no espaço livre, enquanto o segundo considera sua atuação no meio guiado por uma fibra óptica.

A partir dessa abordagem, será possível identificar como os fenômenos não lineares se comportam em diferentes condições de propagação, especialmente destacando as particularidades das interações que ocorrem no interior da fibra óptica. Nessas condições, o sinal óptico sofre modificações específicas que o distinguem da propagação em meios livres. Tais diferenças são relevantes para a compreensão dos efeitos que a não linearidade impõe ao sistema.

Como ponto de partida, o cálculo será iniciado considerando o cenário de propagação no espaço livre, permitindo uma análise inicial das propriedades da não linearidade nesse ambiente antes de avançar para a análise do meio guiado.

$$\eta \sim I \cdot L_{eff} \quad (2.9)$$

sendo η a respectiva figura de mérito entre os dois cenários, I a intensidade da luz incidente na fibra óptica e L_{eff} a extensão da fibra necessária para devida interação. À medida em que a intensidade cresce de forma diretamente proporcional a respectiva figura de mérito do sistema óptico, a figura de mérito vai contornando sendo necessário uma maior intensidade do sinal óptico.

O cenário de espaço livre, Agrawal (2019) define como dimensionamento da respectiva intensidade do feixe a seguir:

$$I = \frac{P}{\pi\omega_0^2} , \quad (2.10)$$

Sendo os parâmetros nessa equação ω_0 o raio da abertura do feixe óptico, considerando uma distribuição gaussiana e P a potência do sinal óptico.

Como forma de cálculo da respectiva figura de mérito, será calculada utilizando a intensidade no espaço livre que será consoante (Agrawal, 2019) com os seguintes parâmetros:

$$L_{eff} = \frac{\pi\omega_0^2}{\lambda}, \quad (2.11)$$

Além disso, o dimensionamento do cenário de figura de mérito no espaço livre conforme (Agrawal, 2019) é calculado como:

$$\eta_{livre} = I \cdot L_{eff} = \frac{P}{\pi\omega_0^2} \cdot \frac{\pi\omega_0^2}{\lambda} = \frac{P}{\lambda} \quad (2.12)$$

O cenário de intensidade para fibra óptica é calculado a seguir, dependendo da potência e de dimensões da fibra óptica é dado como:

$$I = \frac{P}{\pi a^2} \quad (2.13)$$

Sendo a o respectivo raio do núcleo da fibra. Considerando a constante de atenuação (α). Quando o sinal óptico se propaga na fibra óptica a sua potência decai exponencialmente dado uma potência inicial P(0) conforme a seguinte equação, segundo Agrawal (2019):

$$P(z) = P(0)e^{-\alpha z} \quad (2.14)$$

Segundo as condições satisfeitas, a seguinte aproximação numérica pode ser feita conforme Agrawal (2019):

$$L_{eff} \approx \frac{1}{\alpha} \quad (2.15)$$

A outra magnitude da figura de mérito na propagação da luz pelo meio da fibra óptica é regido pela seguinte equação segundo Agrawal (2019):

$$\eta_{fiber} = \frac{P}{\pi a^2} \cdot \frac{1}{\alpha} \quad (2.16)$$

O fator de figura de mérito, levando em conta os parâmetros citados em Agrawal (2019) pode ser numericamente equivalente conforme a equação:

$$\eta \sim I \cdot L_{eff}, \quad (2.17)$$

sendo η a respectiva figura e mérito entre os dois cenários, I a intensidade da luz incidente na fibra óptica e L_{eff} a extensão da fibra necessária para devida interação. A medida que a intensidade é diretamente proporcional a respectiva figura de mérito do sistema óptico. Conforme Agrawal (2019), a razão entre as figuras de mérito pode ser obtida com o arranjo das seguintes variáveis para obter a equação:

$$\frac{\eta_{fiber}}{\eta_{livre}} = \frac{\lambda}{\pi a^2 \alpha} \quad (2.18)$$

Segundo Agrawal (2019), para valores dos seguintes parâmetros: $a \sim 2\mu m$, $\alpha = 8 \cdot 10^{-5}/m$ e $\lambda = 1,55\mu m$ temos o seguinte cálculo da figura de mérito entre os cenários de figura de mérito de propagação dentro da fibra óptica e no espaço livre:

$$\frac{\eta_{fiber}}{\eta_{livre}} = 10^9 \quad (2.19)$$

O cálculo acima prova a seguinte correlação entre os cenários de espaço livre e na fibra óptica. Com a dimensão da figura de mérito, foi obtido um valor na casa de 1 bilhão, o que corresponde para que possa obter o fenômeno da não linearidade no espaço livre seja necessário 1 watt de potência, o mesmo efeito na fibra óptica é obtido com a incidência de 1 nanômetro de watt.

A forma de como a não linearidade possa ser estimada pode ser modelada a partir das equações de Maxwell, que definem as respectivas propriedades, com o objetivo de desenvolver a equação da onda.

Ressaltando que, conforme Agrawal (2019), anteriormente exposto, tanto a intensidade como a distância da fibra óptica influenciam na polarização e de forma reflexa na geração de não linearidade do sistema a ser estudado. O estudo da influência do campo elétrico na permeabilidade, importante para que possa dimensionar a dependência do campo elétrico é dado por:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} = -\mu_0 \frac{\partial \vec{H}}{\partial t} \quad (2.20)$$

Aplicando o operador rotacional no campo elétrico E , obtém-se diferenciando a densidade do campo magnético B em relação ao tempo, resulta na permeabilidade magnética não varia em função do tempo tornando o parâmetro $-\mu_0$ uma constante vezes o campo magnético B . A condição da equação é aplicada conforme a seguinte:

$$\vec{\nabla} \times \vec{H} = -\frac{\partial \vec{D}}{\partial t}, \quad (2.21)$$

o rotacional do campo magnético H é diretamente proporcional à taxa de variação no tempo do deslocamento elétrico D , mas com um sinal negativo. Isso significa que uma variação temporal do campo elétrico gera um campo magnético associado. Segundo Agrawal (2019) aplicando o rotacional na densidade do fluxo elétrico é dado por:

$$\vec{\nabla} \times \vec{D} = 0 \quad (2.22)$$

A equação 2.21 esclarece que o valor de D fornece a condição de que geralmente surge em situações eletrostáticas, onde não há variações temporais do campo elétrico ou da densidade de carga. Em outras palavras, o campo está associado a cargas estacionárias e não a correntes ou campos eletromagnéticos variáveis. Além disso para as condições de propagação na fibra obtém-se a seguinte equação:

$$\vec{\nabla} \times \vec{B} = 0 \quad (2.23)$$

Em termos físicos, essa condição apresentada na equação 2.22 implica a ausência de correntes elétricas livres e campos elétricos variáveis no tempo na região analisada.

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P} = \epsilon_0 \vec{E} + \epsilon_0 \chi^{(1)} \cdot E \quad (2.24)$$

O caso acima descreve o regime linear de operação conforme as equações de Maxwell, de acordo com o vetor de deslocamento conforme a constante do dielétrico, que possui natureza complexa.

Como forma de obter o operador rotacional será obtido o seguinte cálculo conforme Agrawal (2019):

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = -\vec{\nabla} \times \left[\mu_0 \frac{\partial \vec{H}}{\partial t} \right] = -\mu_0 \frac{\partial [\vec{\nabla} \times \vec{H}]}{\partial t} \quad (2.25)$$

Esta equação 2.25 reflete a dinâmica do campo elétrico em função do campo magnético variável no tempo. Em resumo um campo magnético que varia no tempo induz um campo elétrico e essa interação é uma das bases para a propagação de ondas eletromagnéticas, como as descritas pela equação de onda eletromagnética.

Substituindo a equação de Maxwell na equação para o rotacional obtemos a seguinte equação:

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = -\mu_0 \frac{\partial \left[\frac{\partial \vec{D}}{\partial t} \right]}{\partial t} \quad (2.26)$$

No qual está relacionando assim o comportamento do campo elétrico às propriedades do meio através da densidade de deslocamento elétrico (D).

A substituição do termo D pelo parâmetro de polarização calculado em citar o cálculo da polarização, nos fornece a seguinte equação:

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = -\mu_0 \frac{\partial^2 [\epsilon_0 \vec{E} + \vec{P}]}{\partial t^2} = -\mu_0 \frac{\partial^2 (\epsilon_0 \vec{E})}{\partial t^2} - \mu_0 \frac{\partial^2 \vec{P}}{\partial t^2} \quad (2.27)$$

Conforme a equação 2.3 descreve, a polarização possui dois termos resultantes, em que, pela característica da fibra, a resultante de segunda ordem foi desprezada, obtendo como equação da polarização com a seguinte configuração, segundo Agrawal (2019):

$$P = \epsilon_0 (\chi_1 \cdot E + \chi_3 \cdot E^3) \quad (2.28)$$

, observando que a componente de primeira ordem ($\chi_1 \cdot E$) corresponde à característica linear da polarização e a componente de terceira ordem ($\chi_3 \cdot E^3$) corresponde à característica não linear da polarização.

$$\nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\frac{1}{c^2} \cdot \frac{\partial^2 E}{\partial t^2} - \mu_0 \left[\frac{\partial^2 P_L}{\partial t^2} + \frac{\partial^2 P_{NL}}{\partial t^2} \right], \quad (2.29)$$

A equação 2.29 descreve a variação espacial do campo elétrico, onde $\nabla^2 \vec{E}$ é o Laplaciano e $\nabla \cdot \vec{E}$ relaciona-se à divergência do campo. O termo $-\frac{1}{c^2} \cdot \frac{\partial^2 E}{\partial t^2}$ é responsável pela propagação das ondas eletromagnéticas. Ele reflete como o campo elétrico varia no tempo, e a velocidade da onda no vácuo é c , a velocidade da luz. Já o termo $\mu_0 \left[\frac{\partial^2 P_L}{\partial t^2} + \frac{\partial^2 P_{NL}}{\partial t^2} \right]$ incorpora a contribuição da polarização linear e polarização não linear do meio (PL e PLN), que é proporcional ao campo elétrico.

A solução da respectiva equação da onda corresponde, consequentemente descrito em Agrawal (2019), leva à seguinte equação:

$$\nabla^2 \vec{E} - \frac{1}{c^2} \cdot \frac{\partial^2 E}{\partial t^2} = \mu_0 \frac{\partial^2 P_L}{\partial t^2} + \mu_0 \frac{\partial^2 P_{NL}}{\partial t^2}, \quad (2.30)$$

Para resolver a equação acima, é necessário descrever a equação do campo elétrico no domínio do tempo, correspondente aos campos harmônicos:

$$\vec{E} = E_0 e^{j\omega_0 t} \quad (2.31)$$

O estudo do sinal corresponde à largura desse sinal. A não linearidade possui diversos efeitos secundários, porém o objeto deste estudo são os efeitos dentro da fibra óptica na sua propagação. Como descrito na equação acima, ω_0 corresponde a frequência no qual está centrado o pulso, porém, pela forma do pulso, sua envoltória também é importante objeto de estudo no que corresponde ao estudo da não linearidade. Dessa forma, é importante investigar o comportamento da envoltória do pulso óptico.

2.1.2 Espalhamento Raman

O espalhamento Raman é um de vários tipos de espalhamento, e pode ser categorizado como espalhamento Raman Estimulado ou como Espalhamento Raman espontâneo. Quando a luz incide sobre um material, vários efeitos de interação podem ocorrer, dentre eles a polarização da luz incidente. Essa grandeza pode ser utilizada como forma de medição dos modos vibracionais de um determinado material, no qual

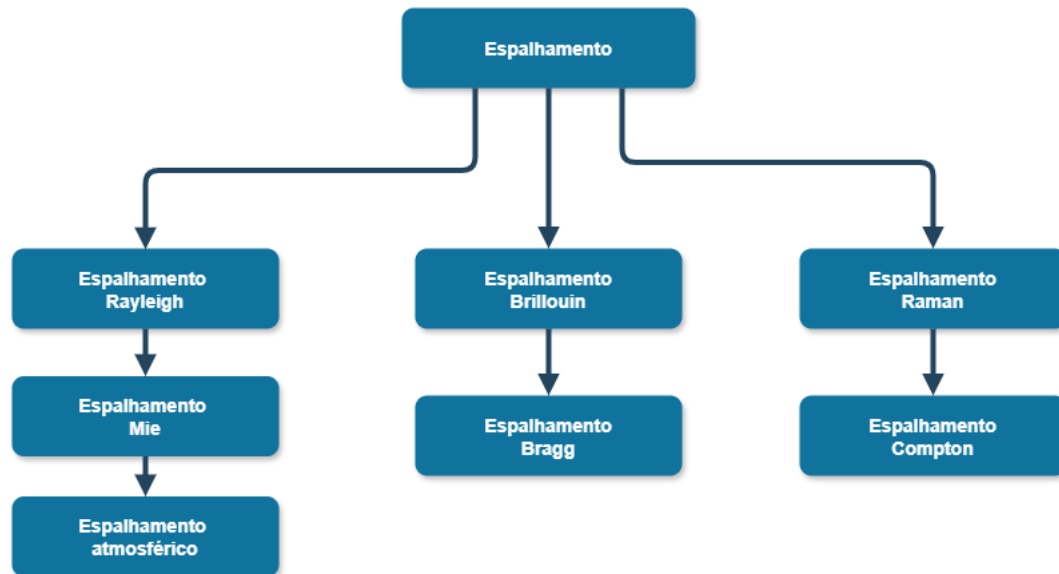
são gerados diferentes tipos de espalhamento, como descrito na Figura 1. A representação de cada ramo é sobre a afinidade do assunto.

O primeiro ramo descreve os espalhamentos que são caracterizados pelos efeitos da incidência dos raios solares com a atmosfera: Rayleigh, Mie e atmosférico. O espalhamento Rayleigh ocorre quando a luz incidente incide sobre uma partícula possui diâmetro menor que o comprimento de onda da luz incidente, por exemplo uma partícula de gás. Já o espalhamento Mie ocorre quando a luz incidente incide sobre uma partícula que possui diâmetro igual ao comprimento de onda da luz incidente, por exemplo uma partícula de poeira. Em relação ao espalhamento atmosférico, este ocorre quando a partícula possui diâmetro muito maior do que o comprimento de onda da luz incidente, que, no caso, pode ser uma partícula de vapor de água.

O ramo central da Figura 1 aponta para os espalhamentos Brillouin e Bragg. O espalhamento Brillouin é a interação de dois campos ópticos incidentes junto com uma vibração mecânica de baixa frequência. Já o espalhamento Bragg fornece assinaturas de materiais analisados conforme a incidência de radiação, que é fornecido com o espalhamento elástico gerado pelo material.

O último ramo da Figura 1 descreve dois tipos de espalhamento: Raman e Compton. O espalhamento Compton é descrito como a emissão de fótons de alta frequência depois da incidência em uma partícula carregada eletricamente. Já o espalhamento Raman possui como característica principal o estudo do comportamento inelástico do feixe óptico por meio da matéria.

Figura 1 – Tipos de espalhamento



Fonte: próprio autor

2.1.3 Propriedades vibracionais dos materiais

O estudo da interação da luz junto ao material a ser analisado pode ser simplificado por meio de um modelo de dois átomos que, hipoteticamente, estão conectados por uma mola com massa tendendo a zero. O sistema, então, descreve os modos normais de vibração do sistema que descreve um oscilador harmônico. Conforme Dresselhaus, (2011), o sistema é modelado conforme a Lei de Hooke, que descreve a atuação de um sistema massa-mola e contém uma respectiva força restauradora:

$$m \frac{d^2}{dt^2} (x - x_0) = -K(x - x_0), \quad (2.32)$$

em que a massa do átomo é descrita como m , x_0 é a posição de equilíbrio do sistema e K é a constante de força de restauração do oscilador. Resolvendo 2.32, chega-se a seguinte solução do sistema:

$$x(t) = x_0 + A \cos(\omega_q t + \varphi), \quad (2.33)$$

sendo que a variável $\omega_q \equiv \sqrt{\frac{K}{m}}$ descreve a frequência angular da oscilação do sistema, A descreve a amplitude e φ descreve a fase que oscila o sistema, conforme condições iniciais em $t = 0$.

A modelagem descrita acima não leva em consideração efeitos quânticos de átomos, conforme Pena (2018), mas é preciso descrever efeitos relacionados à

solução da equação de Schrödinger para um dado oscilador harmônico que é descrito como:

$$\hat{H}\Psi_n = \left[-\frac{\partial^2 \hbar^2}{\partial x^2 2m} + \frac{1}{2}Kx^2 \right] \Psi_n = \varepsilon_n \Psi_n \quad (2.34)$$

no qual $\hat{H}$ descreve o operador hamiltoniano do sistema oscilatório, ε_n corresponde a um autovalor de $\hat{H}$, que se refere à respectiva energia vibracional, e Ψ_n é a respectiva função da onda.

A solução da equação 2.34 encontra-se a respectiva amplitude da função de onda, que é diretamente relacionada à variável n , sendo a descrição de quantas vibracionais que corresponde aos fónons.

O operador Hamiltoniano do sistema de oscilador harmônico pode ser descrito com os operadores de criação e aniquilação, que representam a criação e a destruição do respectivo fónon de energia do sistema Ψ_q . A criação do fónon é descrita na fórmula:

$$\hat{a}^\dagger = \frac{p + i\omega_q m x}{\sqrt{2\hbar\omega_q m}}, \quad (2.35)$$

e o operador de destruição do fónons é descrito a seguir:

$$\hat{a} = \frac{p - i\omega_q m x}{\sqrt{2\hbar\omega_q m}}, \quad (2.36)$$

Segundo Pena (2018), o cenário no qual o comutador entre o comutador do momento (p) e a posição (x) resolverem a um determinado valor não nulo $-i\hbar$ deduz que o operador utilizado para criar e destruir o fónon dado por $[a, a^\dagger] = 1$ e tendo como base que $p = -i\hbar \frac{\partial}{\partial x}$ obtêm-se de 2.33:

$$\hat{H} = \frac{1}{2m} = [(p + i\omega_q m x)(p - i\omega_q m x) + m\hbar\omega_q] = \hbar\omega_q [\hat{a}^\dagger \hat{a} + 1/2]. \quad (2.37)$$

2.1.4 Espalhamento inelástico da luz e interação com a matéria

A figura 2 descreve o comportamento da molécula em relação à transição que é excitada pela incidência do fóton (1), que possui a frequência ω_0 do nível eletrônico fundamental E_1 , para o nível acima, que corresponde ao nível de excitação E_2 , com

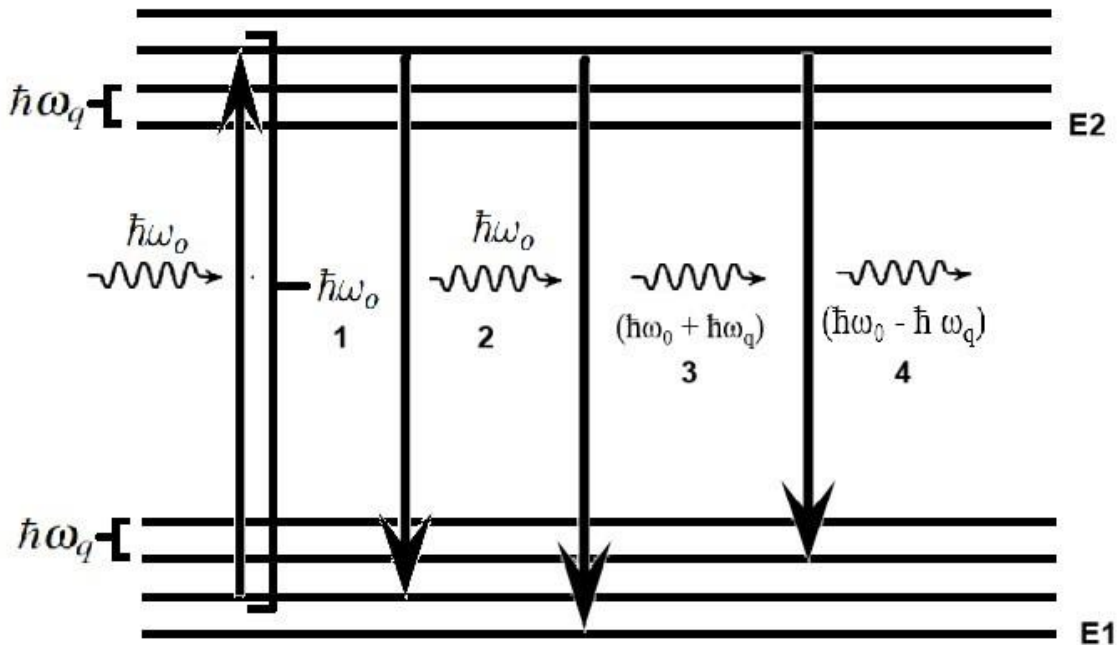
ausência de variação do nível vibracional. Fazendo uso do princípio de conservação de energia, tem-se a seguinte equação:

$$E_{ab} = \hbar\omega_o = \left[E_2 + \left(n + \frac{1}{2} \right) \hbar\omega_q \right] - \left[E_1 + \left(n + \frac{1}{2} \right) \hbar\omega_q \right] = (E_2 - E_1). \quad (2.38)$$

A representação $\hbar\omega_o = (E_2 - E_1)$ corresponde à diferença entre os níveis E_1 e E_2 . Quando ocorre o denominado espalhamento elástico (2), o fóton é emitido com a semelhante energia do fóton incidente. A imagem (3) representa o comportamento da molécula que retorna ao nível fundamental em um nível de energia vibracional menor e é emitido um fóton com energia $\hbar\omega_o + \hbar\omega_q$.

Vale ressaltar quando o nível vibracional é maior (4), com emissão do fóton na forma $\hbar\omega_o - \hbar\omega_q$. Esses processos são descritos como espalhamento Raman Stokes e anti-Stokes.

Figura 2 – Diagrama das transições descrevendo: (1) absorção, (2) emissão elástica (3) emissão inelástica Stokes (4) emissão inelástica anti-Stokes.



Fonte: próprio autor

Tal efeito é demonstrado quando há interação da luz com a matéria, que induz a polarização (Ferraro, 2003), conforme a seguinte expressão:

$$\vec{P} = \overline{\alpha} \vec{E}, \quad (2.39)$$

sendo P a polarização incidente no material, α é o tensor de polarização e E o campo elétrico.

A onda eletromagnética de frequência ω_0 , consoante sua amplitude (E) de acordo com o tempo, pode ser definido como:

$$E = E_0 \cos(\omega - t_0) \quad (2.40)$$

Sendo o material possuidor de estrutura cristalina, os seus modos vibracionais são descritos como:

$$Q = Q_0 \cos(\omega - t_0) \quad (2.41)$$

em que Q_0 descreve a amplitude de vibração e ω_q descreve a frequência de vibração da molécula.

Usando a expansão de Taylor, será possível expandir α , tendo como base uma função linear dependente de Q_0 , que resulta em:

$$\alpha(Q_0) = \alpha_0 + \left(\frac{\partial \alpha}{\partial Q} \right)_{Q=0} Q + O^2, \quad (2.42)$$

como forma de simplificação serão desprezados os termos de ordem maior do que dois, obtendo:

$$P = \alpha_0 E_0 \cos(\omega_0 t) + \left(\frac{\partial \alpha}{\partial Q} \right)_{Q=0} Q_0 E_0 \cos(\omega_0 t) \cos(\omega_q t), \quad (2.43)$$

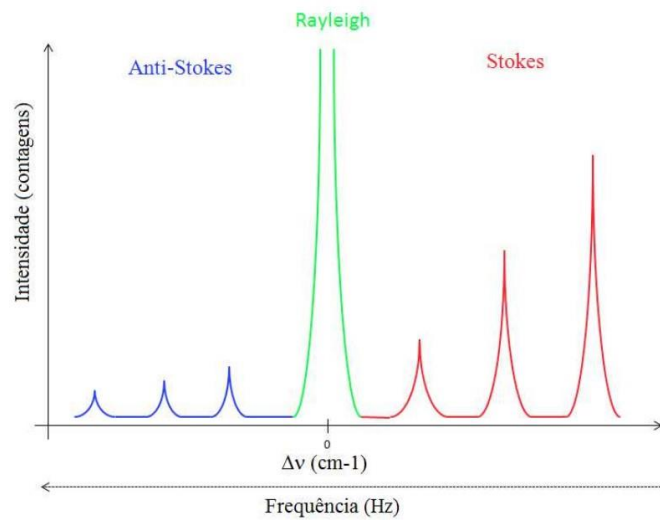
Segundo Pena (2018), convertendo para uma identidade trigonométrica, será possível obter a seguinte equação:

$$P = \alpha_0 E_0 \cos \omega_0 t + \frac{1}{2} \left(\frac{\partial \alpha}{\partial Q} \right)_{Q=0} Q_0 E_0 \{ \cos[(\omega_0 + \omega_q)t] + \cos[(\omega_0 - \omega_q)t] \}. \quad (2.44)$$

que enumera os termos na ordem assim descrita: o primeiro termo é referente ao espalhamento elástico da luz (Rayleigh), os outros termos respectivamente se referem aos espalhamentos anti-Stokes, cuja frequência resultante é $\omega_0 + \omega_q$, e ao espalhamento Raman Stokes de frequência $\omega_0 - \omega_q$.

Segundo o gráfico descrito na Figura 3, o espalhamento Raman pode ser descrito centrado em 0 cm^{-1} . Os espalhamentos Raman representam Stokes à direita e anti-Stokes à esquerda. A figura também descreve um pico de intensidade no espalhamento Stokes e cada pico corresponde a um modo vibracional característico do material a ser analisado.

Figura 3 – Exemplo esquemático de espectro Raman



Fonte: (CARVALHO et al., 2022)

2.1.5 Espalhamento no contexto de amplificação Raman

A figura 4 descreve a passagem do feixe óptico em uma estrutura de óxido de silício, que está disposta na forma de estrutura cristalina. Em um aglomerado de íons de silício e oxigênio, a respectiva polarização do feixe óptico é alterada. Em materiais iônicos (condutores) ou semi-iônicos (semicondutores) a polarização em um dado percurso de um íon vizinho pode inverter a polarização na passagem do feixe óptico pela estrutura cristalina. A respectiva passagem do feixe óptico gera uma vibração atômica e esse efeito é descrito como fônon óptico.

O fenômeno descrito anteriormente é o resultado da interação do feixe óptico com a nuvem de elétrons. O resultado descreve que o fóton incidente é espalhado e que resulta em um fônon óptico. Tal efeito está em concordância com a lei de conservação de energia, que deve obedecer a diferença entre o fóton incidente e o fônon óptico característico do material. A teoria exposta converge com a equação 2.44.

O resultado dessa interação gera um espalhamento inelástico da luz e esse efeito é descrito como Espalhamento Raman Estimulado.

$$E_{\text{fóton espalhado}} = E_{\text{fóton incidente}} - E_{\text{fônon óptico}} \quad (2.45)$$

A convenção mostra que o valor para a combinação $a^\dagger a$ é igual ao operador N_b , e o desenvolvimento leva a:

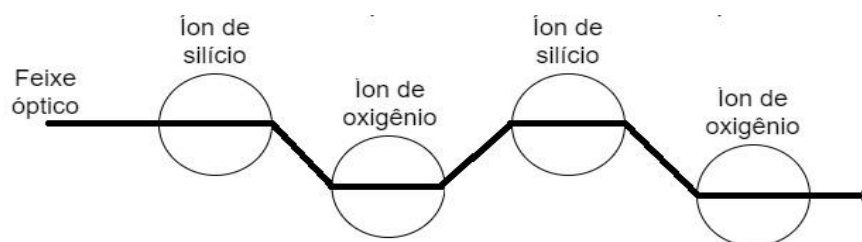
$$\hat{H}|n\rangle = \hbar\omega_q \left[\hat{N} + \frac{1}{2} \right] |n\rangle = \hbar\omega_q \left[n + \frac{1}{2} \right] |n\rangle = \varepsilon_n |n\rangle. \quad (2.46)$$

Conforme um determinado estado $|n\rangle$ que é determinado pelo número quântico n , o estado passará para um estado ($|n+1\rangle$), que corresponde a um nível maior de energia vibracional quando aplicar o operador de criação $a^\dagger$ e caso ele poderá ir a um estado menor ($|n-1\rangle$) caso ele utilize do operador destruição ($\hat{a}$).

A luz resultante é espalhada em direções aleatórias. Ou seja, o Espalhamento Raman Estimulado transfere energia de um componente de comprimento de onda menor para outro componente de maior comprimento de onda por meio da interação com os fônons do material que percorre o feixe óptico.

Além do efeito de espalhamento em que o fóton é atenuado, pode ocorrer a junção de um fóton incidente e um fônon óptico, gerando um fóton espalhado com energia elevada. Esse efeito é o principal da amplificação Raman.

Figura 4 – Propagação do fônon



Fonte: próprio autor

O estudo do espalhamento requer a distinção entre os diferentes tipos de interação da luz com a matéria. A figura 5 descreve o espalhamento espontâneo, que ocorre quando um laser, com uma determinada potência, interage com o material, resultando no espalhamento da luz.

Esse espalhamento caracteriza o comprimento de onda correspondente a esse efeito. Além disso, a luz incidente é espalhada em todas as direções e é gerada de forma espontânea. Tal efeito gera as chamadas ondas de Stokes. A principal característica desse fenômeno é que ele ocorre sem a necessidade de qualquer fonte adicional ou outro estímulo externo para desencadear o espalhamento.

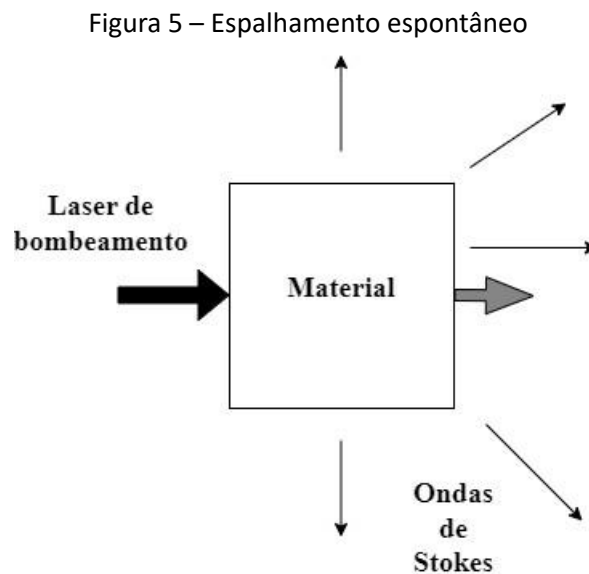
Um outro efeito do espalhamento estimulado envolve tanto um laser de bombeamento quanto a incidência de um laser adicional, conforme está ilustrado na

figura 6, correspondente a uma respectiva fonte de um sinal óptico, que combinam seus efeitos para amplificar o espalhamento.

Esse fenômeno difere do espalhamento espontâneo anterior, pois é coordenado de forma que as interações entre os fótons incidentes e os fônons do material entrem em ressonância, amplificando o sinal e gerando a onda de Stokes, responsável pelo efeito de amplificação.

Levando em consideração que E_f representa o nível de energia do fóton incidente, E_0 denota o nível de energia da molécula em seu estado fundamental (quando a princípio está em repouso), E_1 representa o nível de energia da molécula quando está vibrando em um de seus modos normais, e E_V indica o máximo nível de energia que a molécula poderia alcançar após ser excitada pelo fóton.

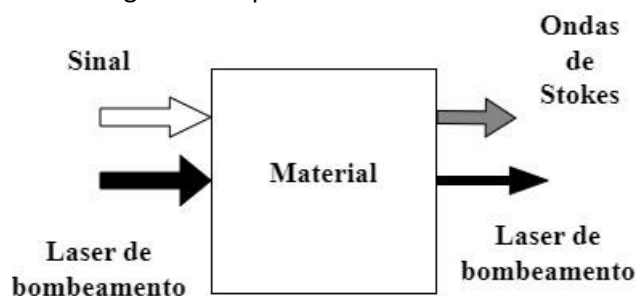
A ocorrência do espalhamento Raman se faz necessário à condição de que a energia do fóton incidente deve ser muito maior do que a energia do estado fundamental E_0 da molécula. O fóton incidindo sobre a molécula pode excitá-la para um modo virtual em que sua respectiva energia é E_V .



Fonte: próprio autor

Há a possibilidade de a molécula voltar ao estado fundamental relativo a E_0 . No caso de o fóton incidente ser somente espelhado, acontece o relativo ao espalhamento Rayleigh conforme descrito nas Figuras 3 e 7. A maioria dos fótons incidentes na matéria são espalhadas dessa forma.

Figura 6 – Espalhamento estimulado



Fonte: próprio autor

Porém em alguns casos, a molécula não retorna ao estado E_0 (inicial), logo após decair, ela atinge um estado E_1 e é emitido novamente um fóton que possui energia $E_f - E_1$, pois parte da sua energia foi repassada para a estrutura cristalina da molécula. Esse saldo foi gerado pela vibração da molécula e esse fenômeno é um tipo de espalhamento Raman.

O fóton depois de ser gerado passa a ter uma frequência menor do que o fóton incidente. Essa nova frequência é chamada de frequência de Stokes. Quando a energia do fóton diminui, alocando parte de sua energia para a molécula, o efeito é o Stokes. Quando ocorre o acréscimo de energia do fóton, pois ele tomou a energia da molécula, esse efeito é o anti-stokes, conforme Figura 3.

A Figura 3 mostra como está disposta a questão do espalhamento, que define a onda de Stokes criada a partir de um acréscimo da sua frequência, para assim obter o fenômeno da amplificação, e descreve o respectivo acréscimo na frequência. Vale destacar que nem toda vibração da matéria ocasiona o espalhamento Raman. Para que haja absorção da luz incidente pela molécula, a vibração deve variar conforme o dipolo elétrico da molécula e a vibração deve fazer esse respectivo momento de o dipolo variar.

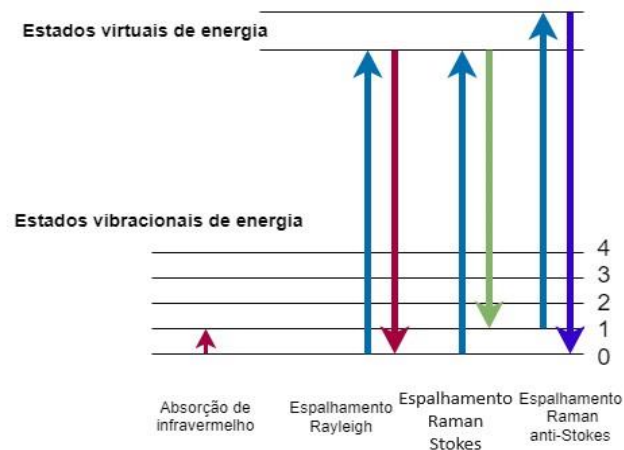
A noção do efeito é de que a luz incidente atua como um bombeamento e, a partir dessa interação com a variação na frequência, é criada uma radiação correspondente, chamada de onda de Stokes que, para outros fins, é utilizada como meio para espectroscopia.

A amplificação Raman, por outro lado, pode ser obtida conforme as disposições da Figura 7, em que todas as frequências de Stokes são geradas a partir do espalhamento Raman junto com a do sinal incidente. A amplificação Raman é

observada quando se quer um feixe de Stokes de alta qualidade, em que normalmente é gerada uma determinada frequência de Stokes, com os requisitos de um respectivo sinal de entrada desejado.

A Figura 7 representa o processo de espalhamento Raman, tendo como representação os níveis de energia da matéria. Os átomos em uma molécula estão sempre em movimento, pois estão em contínuas oscilações. A partir dessas oscilações, os níveis de energia podem variar conforme a interação e a absorção de energia. Tendo como exemplo a Figura 7, o primeiro efeito representa a absorção de uma onda na faixa do infravermelho, no qual a matéria absorve a radiação e não emite nenhum fóton, convertendo a energia absorvida em calor pela matéria.

Figura 7 – Esquema de amplificação Raman de primeira ordem



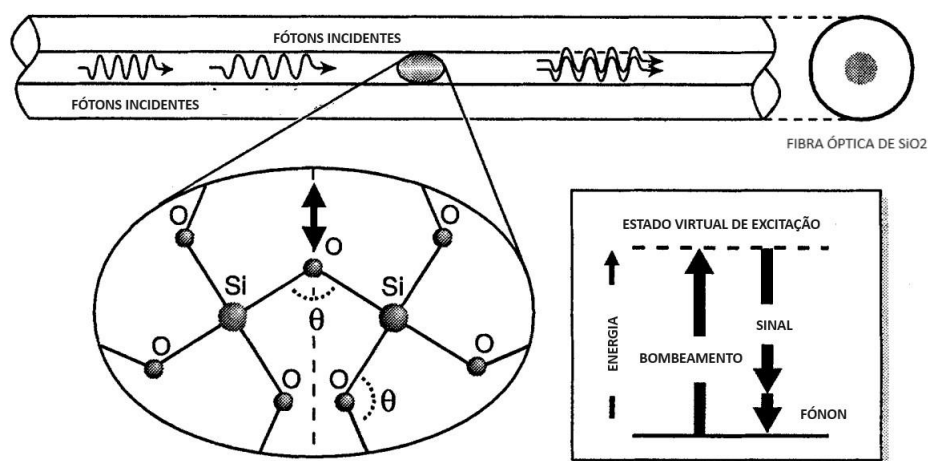
Fonte: próprio autor.

O espalhamento Rayleigh é descrito quando o fóton é espalhado elasticamente em um ângulo θ sem perda de energia. Já nos componentes do espalhamento Raman, parte da energia é absorvida pela estrutura cristalina da matéria, em que o fóton excita a molécula, cedendo à estrutura molecular parte de sua energia, conforme representado na Figura 7. O fóton incidente possui um nível de estado vibracional mais atenuado do que o incidente. Já a componente anti-Stokes é o contrário, o fóton é espalhado com um estado vibracional maior do que o fóton incidente, que tira a energia da molécula para si, aumentando o seu nível de estado vibracional.

O efeito de espalhamento Raman como amplificação possui características em fibras ópticas de óxido de silício (SiO_2). As interações com o fóton incidente possuem dependência com relação ao respectivo ângulo do movimento de flexão da ligação Si-O-Si (ver ângulo de ligação na Figura 8). Para realizar o efeito Raman, convencionou-se a utilizar uma fonte óptica e denominá-la fonte de bombeamento. Na Figura 8, a luz incidente da fonte de bombeamento e a luz incidente do sinal óptico são coerentemente acopladas pelo processo Raman, que é explicado com a adição de energia do estado de excitação do sinal, junto com a energia liberada pela estrutura cristaliza da matéria, que no caso é o óxido de silício.

O efeito é descrito na Figura 8. Segundo o que ocorre baseado na mecânica quântica, apresenta no diagrama de níveis de energia da Figura 8, um fóton da fonte de bombeamento que é convertido no segundo fóton de sinal óptico, e este é uma réplica exata do primeiro, vindo da fonte de bombeamento, e a energia restante produz um fónon óptico. O fóton de sinal óptico inicial, portanto, foi amplificado. Este processo é considerado não ressonante porque o estado superior é um estado virtual de curta duração.

Figura 8 – Esquema da amplificação Raman estimulada em uma fibra óptica.



Fonte: (BROMAGE, 2004)

O efeito explicado anteriormente é característico da incidência da espectroscopia Raman, estimulado em um meio de óxido de silício. As fibras ópticas possuem como composição justamente o óxido de silício, tornando viável a aplicação do efeito Raman como meio de amplificação. O resultado da amplificação Raman é obtido usando a própria fibra óptica como meio. O efeito de amplificação no efeito

Raman (Agrawal, 2019) possui como requisito a alta potência, o que torna todo sistema custoso, pois a tecnologia de laser necessita em torno de 1 watt de potência. Porém, com a devida configuração, em que o feixe de Stokes esteja bem selecionado, o efeito de amplificação Raman pode ocorrer em torno de 100 mW.

Tabela 1 – Parâmetros de entrada no Raman estimulado.

Variável	Bombeamento	Sinal
Intensidade do feixe	I_P	I_S
Frequência	ω_P	ω_S
Coefficiente de atenuação	α_P	α_S

O esquema de bombeamento Raman estimulado é composto por variáveis necessárias composta na tabela 1. Segundo (Agrawal, 2019), a propagação do sinal é modelada conforme a seguinte equação que deriva a intensidade do sinal I_S em relação ao eixo Z :

$$\frac{\partial I_S}{\partial z} = g_R I_P I_S - \alpha_S I_S \quad (2.47)$$

É importante destacar que o mesmo estudo também se aplica ao laser de bombeamento, cuja intensidade (I_P) é derivada em função de sua propagação ao longo do eixo z conforme a equação a seguir:

$$\frac{\partial I_P}{\partial z} = -\frac{\omega_P}{\omega_S} g_R I_P I_S - \alpha_P I_P, \quad (2.48)$$

sendo $\frac{\omega_P}{\omega_S}$ relativo à conservação dos fótons no sistema. Ressaltando que o processo Raman ocorre em certas transições, nas quais o fóton é absorvido pelo material, aumentando sua energia, e, pela conservação de energia, o sistema mantém a quantidade de fótons.

O cálculo do número de fótons do sistema combinando o sinal e o bombeamento pode ser quantizado com o respectivo cálculo diferencial, que deve ser igual a zero ao percorrer a fibra óptica, segundo a equação abaixo:

$$\frac{d \left[\frac{I_P}{\omega_P} + \frac{I_S}{\omega_S} \right]}{dz} = 0 \quad (2.49)$$

A solução da equação 2.28 é proposta com a mensuração da intensidade do laser de bombeamento conforme a equação a seguir:

$$I_P(z) = I_{P0} e^{-\alpha_P z} \quad (2.50)$$

Substituindo o I_P , segundo (Agrawal, 2019) na equação 2.28 fica:

$$\frac{dI_S}{dz} = g_R I_{P0} e^{-\alpha_P z} I_S - \alpha_S I_S \quad (2.51)$$

$$I_S(z) = I_S(0) \exp[g_R I_{P0} L_{eff} - \alpha_S z] \quad (2.52)$$

Sendo que aplicando a simplificação, a equação pode ser descrita como:

$$I_S(L) = I_S(0) \exp(g_R I_{P0} L_{eff} - \alpha_S L), \quad (2.53)$$

Além disso, levando em consideração que o comprimento efetivo da fibra óptica é diretamente proporcional ao efeito gerado pela intensidade do feixe óptico, o efeito pode ser modelado pela seguinte equação:

$$L_{eff} = [1 - \exp(-\alpha_P L / \alpha_P)] \quad (2.54)$$

O efeito do espalhamento Raman induz por meio da emissão térmica, portanto será necessário calcular a potência total gerada, conforme as interações entre o laser de bombeamento e as ondas de Stokes geradas. Para calcular tal efeito, será necessário integrar a frequência com o escopo de calcular a potência da onda de Stokes, segundo a equação abaixo.

$$P_S(z) = \int_{-\infty}^{\infty} h\omega (g_R(\Omega) I_{P0} L_{eff} - \alpha_S z) d\omega, \quad (2.55)$$

sendo Ω a frequência relativa referente à diferença entre a frequência de bombeamento com a frequência de entrada ($\omega_P - \omega$).

A largura de banda efetiva (B_{eff}) da fibra óptica tida como meio de propagação, segundo (Agrawal, 2019):

$$B_{eff} = \left(\frac{2\pi}{I_{P0} L_{eff}} \right)^{\frac{1}{2}} \left| \frac{\partial^2 g_R}{\partial \omega^2} \right|^{\frac{1}{2}} \quad (2.56)$$

Além disso, segundo Agrawal (2019), a potência de entrada é:

(2.57)

$$P_{S0} = h\omega s \cdot B_{eff}$$

e o cálculo da potência de saída é feita conforme a equação:

$$P_S(z) = P_{S0} \exp(g_R(\Omega) I_{P0} L_{eff} - \alpha_S z) \quad (2.58)$$

O cálculo do respectivo limiar Raman é obtido conforme (AGRAWAL, 2019):

$$P_S(z) = P_P(z) = P_{P0} e^{-\alpha_P z} \quad (2.59)$$

Considerando que a mensuração é obtida conforme o seguinte cálculo:

$$P_{P0} = I_{P0} A_{eff}, \quad (2.60)$$

sendo A_{eff} a área efetiva da fibra e I_{P0} a intensidade do sinal óptico de bombeamento. Vale ressaltar a dependência da área efetiva para dimensionar a potência do laser de bombeamento conforme a equação apresentada.

2.1.6 Amplificadores Raman de primeira ordem

Os sistemas de amplificação óptica mais utilizados no atual contexto de mercado focam a implementação na janela de 1550 nanômetros, por meio de tecnologias de amplificação EDFA. Essa faixa concentrou as pesquisas devido à sua baixa atenuação em fibras ópticas e à crescente aplicação da tecnologia do EDFA, que entrega amplificação eficiente e estável para sistemas de amplificação óptica.

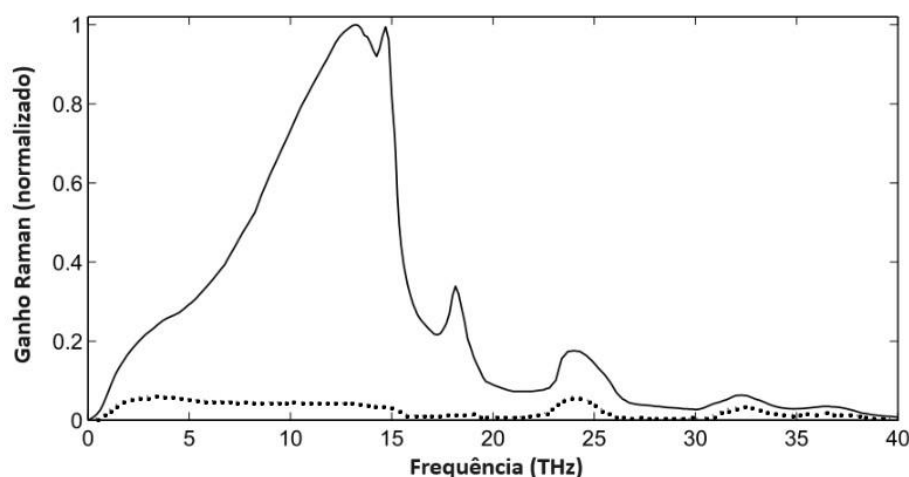
Porém, com a expansão de pesquisas para encontrar amplificação para janelas que suportem maior largura de banda, estão ampliando o escopo de pesquisa para janelas de 1260 nanômetros ou até 1600 nanômetros, nesse aspecto a tecnologia EDFA não consegue obter sucesso desejado, pois tende a diminuir significativamente a eficiência de sua amplificação.

Tal contexto é necessário para amplificadores Raman que se destacam no sentido de serem flexíveis e pela capacidade de amplificação. Esse aspecto leva vantagem em relação ao EDFA, pois justamente onde possui ineficiência, a

amplificação Raman mostra potencial para expansão da capacidade de amplificação de redes ópticas em faixas de transmissão menos convencionais para uso comercial.

A conceituação de um parâmetro importante para o efeito Raman de amplificação é o espectro de ganho Raman $g(\Omega)$, conforme Agrawal (2019), que representa a diferença entre a frequência de bombeamento e a frequência de Stokes. Esse parâmetro é o mais importante para definição do Espalhamento Raman Estimulado, e o seu valor depende da constituição da fibra, em que há considerável alteração de acordo com a dopagem.

Figura 9 – Ganho Raman normalizado para fibra fundida de sílica. A curva pontilhada descreve quando o bombeamento e as ondas de Stokes são polarizadas ortogonalmente.



Fonte: (AGRAWAL, 2019)

A Figura 9 expõe o ganho Raman em função da distribuição de frequência de uma fibra de sílica fundida com uma fonte de bombeamento de $\lambda_p = 1\mu m$. A principal função do ganho Raman é definir o seu valor máximo de determinação em torno de 13 THz, conforme a Figura 9. Tal característica é obtida com a natureza não cristalina da fibra óptica pois, em materiais amorfos, e a sílica é um desses materiais, as frequências vibracionais das moléculas se sobrepõem e criam um contínuo, e esse efeito é essencial para a amplificação Raman.

O feixe de bombeamento pode ter variação de sentido, sendo propagante ou contra propagante, conforme a disposição do arranjo de amplificação Raman. Segundo Agrawal (2019), para obtenção do efeito de amplificação Raman, é necessário obter o respectivo limiar do efeito Raman.

O limiar Raman é necessário para que um feixe de onda contínua em uma fibra óptica possa realizar a função de laser de bombeamento. Isso é condizente com a densidade de potência na fibra que naturalmente vai decaindo, e deve ser levado em consideração a interação não linear entre o laser de bombeamento e as ondas de Stokes do sinal óptico a ser amplificado, conforme as equações 2.60 e 2.61. Vale ressaltar que é necessária a condição das ondas serem contínuas ou quase contínuas. Segundo Agrawal (2019):

$$\frac{dI_s}{dz} = g_R I_p I_s - \alpha_s I_s \quad (2.61)$$

A equação para a intensidade do laser de bombeamento é descrita a seguir:

$$\frac{dI_p}{dz} = g_R I_p I_s - \alpha_p I_p \quad (2.62)$$

na qual α_s e α_p são as perdas na fibra óptica do sinal de Stokes e do laser de bombeamento, respectivamente. Por outro lado, temos a constante do ganho Raman g_R que possui um valor para a fibra de sílica de $g_R = 1 \times 10^{-13}$, assim como, ω_p e ω_s são a frequência do laser de bombeamento e do sinal relativo ao Stokes, respectivamente. Além disso, I_p e I_s representam as intensidades do laser de bombeamento e do sinal relativo ao Stokes, respectivamente.

A descrição completa do efeito Raman estimulado leva em consideração a saturação do ganho, tornando a 2.60 na 2.62, desconsiderando o seu lado direito da equação que representa a saturação. A seguinte equação relaciona a variação da intensidade do feixe do sinal óptico:

$$\frac{dI_s}{dz} = g_R I_0 \exp(-\alpha_p z) I_s - \alpha_s I_s, \quad (2.63)$$

na qual I_0 é a intensidade do bombeamento incidente em $z=0$.

A devida solução da equação 2.63 é 2.53, conforme as condições antes expostas de espalhamento Raman, sendo L o comprimento total da fibra e L_{eff} o seu comprimento efetivo que é calculado por 2.64.

Consequentemente, a extensão efetiva da fibra é descrita como L_{eff} e calculado pela equação 2.64. Assim como, o ganho efetivo da fibra G_A é descrito pela equação 2.65.

$$L_{eff} = [1 - \exp(-\alpha_p L / \alpha_p)] \quad (2.64)$$

$$G_A = \exp(g_R P_0 L_{eff} / A_{eff}) \quad (2.65)$$

A solução trazida em 2.52 mostra que o efeito é decorrente da absorção do bombeio e, para efetivar o efeito de espalhamento, o comprimento efetivo (L_{eff}) deve ser menor do que o comprimento da fibra (L).

Sendo A_{eff} a abertura efetiva da fibra, o cálculo da respectiva potência é obtido através da equação 2.65.

$$P_0 = I_0 A_{eff}, \quad (2.66)$$

Segundo Smith (1972), a potência de bombeamento crítica necessária para o limiar Raman com boa aproximação é definida na equação 2.66.

$$\frac{g_R P_0^{Cr} L_{eff}}{A_{eff}} \approx 16, \quad (2.66)$$

sendo P_0^{Cr} o limiar Raman, A_{eff} a área efetiva da fibra óptica.

O modo de como a forma operacional das fibras ópticas utilizarem no seu cotidiano de instalação de fibras ópticas com potências abaixo de 10 mW, é difícil de ocorrer o espalhamento Raman estimulado com essas condições usuais de operação, já que necessita de altas potências conforme as equações demonstradas.

Para viabilizar o espalhamento Raman estimulado, o uso da região visível em que o A_{eff} é costumeiro utilizar $60\mu m^2$ em fibras monomodo, para usar na equação 2.66, é obtido um valor para o limiar de $P_0^{Cr} \approx 10W$ para uma fibra de comprimento de 10 metros.

A amplificação Raman de primeira ordem, segundo Bromage (2004) é governada pelas equações 2.68 e 2.69.

$$\frac{dP_S}{dz} = -\alpha_S P_S + C_R P_P P_S \quad (2.68)$$

$$\kappa \frac{dP_P}{dz} = -\alpha_P P_P - \frac{\lambda_S}{\lambda_P} C_R P_S P_P \quad (2.69)$$

Sendo P_S a potência do sinal propagante, P_P a potência do laser de bombeamento (pump), α_S e α_P são os coeficientes de perda do sinal e do laser de bombeamento, respectivamente. O parâmetro κ depende da configuração do arranjo, que pode ser (1) no sentido propagante e (-1) no sentido contra propagante e o parâmetro C_R é o coeficiente Raman que depende do sinal e do comprimento de onda do laser de bombeamento.

O contexto de amplificação objeto deste estudo é para que um determinado sinal de entrada possa ter sua amplitude aumentada junto com um feixe de

bombeamento monocromático, e que dessa diferença a frequência esteja dentro da largura de banda do ganho Raman.

Além disso, a configuração experimental do amplificador Raman para contextualização do funcionamento da amplificação Raman, que simula o efeito de amplificação Raman, é diferenciada pelo arranjo de propagante ou contra propagante, com relação aos feixes de bombeamento, todos na unidade de Watts (Bromage, 2004), segundo as equações 2.70 e 2.71.

$$\frac{dP_s}{dz} = -\alpha_s P_s + C_R(\lambda_s, \lambda_p)[P_p^+ + P_p^-]P_s \quad (2.70)$$

$$\pm \frac{dP_p^\pm}{dz} = -\alpha_p P_p^\pm - \frac{\lambda_s}{\lambda_p} C_R(\lambda_s, \lambda_p)P_s P_p^\pm \quad (2.71)$$

onde α_s e α_p são os coeficientes de perda do sinal óptico e do laser de bombeamento, respectivamente na unidade de km^{-1} , λ_s e λ_p são respectivamente o comprimento de onda da fonte do sinal e do laser de bombeamento. A força do acoplamento é calculado por meio da eficiência do ganho Raman na fibra representado por $C_R(\lambda_s, \lambda_p)$ e sua respectiva unidade é $(W \cdot km)^{-1}$.

2.1.7 Amplificadores Raman de segunda ordem

A ampliação de fontes de bombeamento tornou-se objeto de estudo de vários cientistas, conforme Kuzin (1999), em que é representado um estudo de vários arranjos, utilizando emissão estimulada do efeito Raman, com o intuito de otimizar o respectivo efeito. O autor também cita o desenvolvimento da modelagem numérica do amplificador Raman de segunda ordem segundo as equações 2.72, 2.73 e 2.74.

$$\frac{dP_P}{dz} + \alpha_P P_P = \frac{-g P_P (P_{S1} + p)}{A_{eff}} \quad 2.72$$

$$\frac{dP_{S1}}{dz} + \alpha_{S1} P_{S1} = \frac{g(\frac{\lambda_p}{\lambda_{S1}}) P_P (P_{S1} + p)}{A_{eff}} - \frac{g P_{S1} (P_{S2} + p)}{A_{eff}} \quad 2.73$$

$$\frac{dP_{S2}}{dz} + \alpha_{S2} P_{S2} = \frac{g(\frac{\lambda_{S1}}{\lambda_2}) P_{S1} (P_{S2} + p)}{A_{eff}} - \frac{g P_{S2} (P_{S3} + p)}{A_{eff}} \quad 2.74$$

onde:

- P_p , P_{s1} , P_{s2} e P_{s3} representam, respectivamente, os níveis de potência do feixe de bombeamento, da primeira, segunda e terceira componente da emissão estimulada relativo ao feixe de Stokes;
- $(\lambda_p, \lambda_{s1}, \lambda_{s2}$ e $\lambda_{s3})$ representam, respectivamente, os comprimentos de onda do feixe de bombeamento, da primeira, segunda e terceira componente da emissão estimulada relativo ao feixe de Stokes;
- $(\alpha_p, \alpha_{s1}, \alpha_{s2}, \alpha_{s3})$ descrevem os respectivos coeficientes de atenuação do feixe de bombeamento, da primeira, segunda e terceira componente da emissão estimulada relativo ao feixe de Stokes;
- g e A_{eff} são respectivamente o ganho Raman e a área efetiva da fibra óptica.

Vale ressaltar a correlação entre os feixes de Stokes relacionados aos componentes do arranjo de amplificação Raman. No caso das fórmulas apresentadas, os parâmetros que serão utilizados são a frequência do laser de bombeamento 1 e o laser de bombeamento 2, além da potência do laser de bombeamento 1, tendo em vista que o ganho Raman e a área efetiva permanecerão constantes.

2.1.8 Figuras de mérito

As comunicações ópticas tiveram na sua evolução certas métricas de desempenho e eficiência dos sistemas. Tais métricas possuem o foco na comparabilidade quantitativa do uso de diferentes tecnologias, dimensionando quais tecnologias são de maior performance a uma específica aplicação. A seguir estão algumas descrições da importância do uso de figuras de mérito:

- As figuras de mérito, tais como taxa de erro de bits (BER), eficiência espectral e razão sinal-ruído (SNR) são importantes para determinar a eficiência de um dado sistema e sua qualidade na transmissão. Eles indicam quantos bits estão sendo transmitidos errados, a respectiva razão entre o sinal e o ruído na transmissão como forma de quantificar a qualidade de transmissão.
- As soluções de comunicação óptica possuem diversas tecnologias e o uso de ferramentas de figuras de mérito dimensionam devidamente que a devida solução possa ser devidamente mais adequada a um determinado sistema.
- O uso de figuras de mérito pode ser decisivo na definição do custo do projeto com o foco na relação entre o sistema e sua devida viabilidade econômica do projeto.

- A figura de mérito pode ser decisiva na modelagem de soluções de comunicação óptica em ambientes com condições extremas que necessitam de estabilidade na comunicação.

Portanto, o uso de figuras de mérito nas comunicações ópticas garante que a eficiência de determinado sistema possa ser devidamente mensurado e direciona devidos ajustes na perda de performance. A seguir serão descritos os parâmetros de figuras de mérito utilizados na simulação para sua mensuração.

2.1.8.1 Relação sinal ruído óptica (OSNR)

A performance do sistema de comunicação óptico possui vários parâmetros para identificar seu respectivo bom funcionamento. Segundo Gumaste; Antony (2003) e Agrawal (2019), a configuração do enlace óptico é na forma de ponto a ponto, em que possa mensurar de forma eficaz o sistema.

Um desses parâmetros é a relação sinal ruído óptico (OSNR) é uma medida para quantificar a interferência de sinais ópticos. Essa medida é uma dimensão que leva em conta uma efetiva largura de banda do sistema de comunicação. Essa medida é um indicador de performance do sistema óptico. O cálculo é feito na seguinte equação 2.74.

$$OSNR = 10\log_{10}(P_s/P_r) = dB(sinal) - dB(ruido) \quad (2.75)$$

,

sendo P_s a potência do sinal e P_r a potência do ruído.

Além disso, caso existam múltiplos estágios no enlace de comunicação óptica, o cálculo da OSNR pode ser efetuado segundo 2.75.

$$OSNR = \frac{P_{in}}{NF_{stage} h \nu \Delta f} \quad (2.76)$$

sendo NF_{stage} a figura de ruído do estágio a ser analisado, h é a constante de Planck ($6,6260 \times 10^{-34}$), ν é a frequência óptica de 193 THz e Δf é a largura de banda a ser medida, usualmente é 0.1 nanômetros.

2.1.8.2 Fator de qualidade de transmissão óptica

Outro parâmetro para medir a performance do sistema é o fator de qualidade do sistema óptico. O conceito de fator de qualidade é herdado da literatura da

eletrônica, no qual assenta sua teoria nos ressonadores. Para sistema ópticos, aplica-se o conceito de ressonadores ópticos. Segundo Gumaste; Antony (2003) e Agrawal (2000), o cálculo do fator de qualidade é descrito na equação 2.76.

$$Q_{dB} = 20 \log \sqrt{OSNR} \sqrt{\frac{B_0}{B_C}} \quad (2.77)$$

sendo B_0 a largura de banda no fotodetector e B_C a largura de banda do sinal elétrico no filtro do receptor.

A relação entre o fator de qualidade e a relação sinal ruído óptico é evidente. A detecção de ruídos é feita por meio de analisadores de espectro. O fator de qualidade é um importante parâmetro para eficiência do sistema óptico.

2.1.8.3 Taxa de erros de bit do enlace óptico

A principal função do receptor no enlace de comunicação é receber o sinal óptico que é modulado por meio de pulsos e essa sequência é medida por meio de uma taxa igual aos bits transmitidos. Então o receptor decide se o sinal corresponde a zero ou um, tal limiar é definido e dele vem a respectiva decisão.

Porém há casos em que a atenuação da fibra torna o sinal fraco, diferente do transmitido e recebido, o que induzirá a produção de erros por parte do receptor. Para manter a boa comunicação, segundo Ghatak; Thyagarajan (1998) e Agrawal (2019), é desejável manter as taxas abaixo de 10^{-9} ou 10^{-12} na realidade de operação dos receptores.

A medida da qualidade do sistema de comunicação é especificada por meio do cálculo da taxa de erros de bit (em inglês, bit error rate), que é definida na equação 2.77.

$$BER = \frac{\text{Bits errados recebidos em um determinado momento}}{\text{Bits recebidos durante determinado momento}} \quad (2.78)$$

2.2 Aprendizado de máquina

Machine learning, ou Aprendizado de Máquina, é um subcampo da inteligência artificial (IA) que se concentra no desenvolvimento de algoritmos e técnicas que permitem aos computadores aprenderem e melhorarem com base em dados sem serem explicitamente programados.

Em vez de seguir instruções específicas para realizar uma tarefa, os sistemas de ML são capazes de identificar padrões nos dados e fazer previsões ou tomar decisões com base nesses padrões. Isso é feito através do treinamento de modelos em conjuntos de dados, onde o modelo aprende a reconhecer correlações e tendências nos dados de entrada para fazer previsões ou tomar decisões.

Existem várias abordagens de ML, incluindo aprendizado supervisionado, aprendizado não supervisionado, aprendizado por reforço e aprendizado semi-supervisionado. O aprendizado supervisionado envolve treinar um modelo com pares de entrada e saída, enquanto o aprendizado não supervisionado envolve encontrar estrutura em dados sem rótulos.

O aprendizado por reforço envolve aprender a tomar ações em um ambiente para maximizar uma recompensa, e o aprendizado semi-supervisionado envolve treinar um modelo com uma combinação de dados rotulados e não rotulados. Essas abordagens são utilizadas em uma variedade de aplicações, como reconhecimento de padrões, reconhecimento de voz, processamento de linguagem natural, diagnóstico médico, previsão financeira, entre outros.

Métodos de regressão baseiam-se na aplicação do método dos mínimos quadrados. Generalizando, o principal objetivo de uma regressão é obter uma variável aleatória X pertencente aos números reais e um vetor $y = (y_1, y_2, \dots, y_n)$ pertencentes aos números reais (Izbicki; Santos, 2020). Quando a variável X possui uma característica quantitativa, temos um problema de regressão. Caso a variável X possui característica qualitativa é um problema de classificação.

As observações contidas em $y = (y_1, y_2, \dots, y_n)$ são definidas como variáveis explicativas ou independentes. Também pode ser chamada de atributos (features em inglês) e as variáveis de entrada de alvos (target em inglês).

A escolha dos algoritmos para desenvolver o modelo é fundamentada na questão principal que por exemplo, modelos de Decision Tree Regressor (DTR), que é baseado em uma só árvore de decisão onde os nós são recursivamente segmentados de acordo com os valores das features e seu uso é recomendável para datasets de baixa dimensão, mas tende a um overfitting em datasets de maior dimensão.

O modelo baseado em Random Forest Regressor (RFR), que soluciona o problema de construção de várias árvores em subconjuntos randômicos do dataset e depois faz uma média de previsões de cada árvore e seu uso é recomendável para

datasets de baixo ou alto grau de dimensão, Gradient Bosting Regressor (GBR), que constrói arvores de decisão serializadas sendo que cada árvore de decisão é melhorada e seu uso é recomendável para datasets de maior grau de dimensão mas o seu processamento pode ser elevado.

Histogram Boosting Regressor (HBR), que é uma versão melhorada do GBR que usa histogramas para segmentar os valores das variáveis em dois grupos e seu uso é recomendável para datasets maiores e seu consumo computacional é pequeno para o processamento das features, e AdaBoast Regressor (ABR) ,que é uma algoritmo de boosting que tenta corrigir os erros de cálculos dos pesos anteriores para recalcular os dados de entrada e seu uso é recomendável para uma abordagem mais leve do processamento das features mas é mais susceptível a outliers.

De forma geral pela variedade de algoritmos e suas diferenças em determinados cenários será aplicado os algoritmos de modelagem: Decision Tree, Random Forest, Gradient Boosting, Histogram Boosting e Adaboost.

2.2.1 Decision Tree Regressor

Um dos algoritmos mais populares no campo do aprendizado de máquina é o Decision Tree Regressor, também conhecido como Regressor baseado em Árvore de Decisão. Este modelo, além de ser amplamente utilizado de forma independente, serve como base para outros algoritmos mais avançados, como os ensembles de árvores, incluindo Random Forests e Gradient Boosting (Breiman, 2001; Friedman, 2001).

Uma árvore de decisão organiza-se em uma estrutura hierárquica composta por nós de decisão e nós folha. Cada nó de decisão representa uma condição específica ou regra baseada em atributos do conjunto de dados, enquanto os nós folha fornecem a saída do modelo. No caso de tarefas de classificação, o valor atribuído ao alcançar um nó folha é chamado de rótulo, enquanto, para regressão, o valor corresponde a uma estimativa numérica (Breiman, 2017). A análise ocorre de forma iterativa e hierárquica, com a árvore sendo percorrida do nó raiz até um nó folha, de acordo com as condições definidas em cada nó ao longo do caminho.

Um aspecto fundamental na construção e desempenho de uma árvore de decisão é sua profundidade, que corresponde ao número máximo de níveis na estrutura da árvore. Quanto maior a profundidade, mais complexa e detalhada é a árvore, permitindo que ela capture padrões complexos nos dados. No entanto, isso

também pode aumentar o risco de overfitting, ou seja, o modelo torna-se excessivamente ajustado ao conjunto de treinamento e perde capacidade de generalização (Quinlan, 1986). Assim, é essencial equilibrar a profundidade da árvore para garantir que ela processe todas as condições relevantes sem perder desempenho em novos dados.

2.2.2 AdaBoost Regressor

O AdaBoost é um algoritmo de boosting no qual avalia as previsões do modelo e aumenta os seus pesos e que foca em melhorar a acurácia de um modelo de classificação ao combinar vários modelos fracos. Ele funciona ajustando os pesos das instâncias de treinamento de acordo com a dificuldade de classificação. Instâncias mal classificadas recebem pesos maiores nas próximas iterações, forçando o modelo a focar nelas. O AdaBoost é frequentemente usado para problemas de classificação binária, mas também pode ser adaptado para problemas de regressão, como no caso desse estudo em que os dados são obtidos de uma simulação (Drucker, 1997; Hastie et al., 2009).

Conforme Kunapuli (2023), o respectivo algoritmo é capaz de treinar um novo modelo fraco priorizando exemplos com alto peso já categorizados anteriormente. Para dar continuidade é necessário atualizar os pesos para aumentar o valor de cada peso de exemplos mal classificados. Com isso, vai inicializar as iterações com pesos iguais e mostrar como os pesos mudam nas etapas seguintes, tendo como contínua esse ajuste de pesos.

O modelo de Adaboost utiliza como metodologia o uso de árvores de decisão rasas (fracos) para aprendizado rápido, como forma de otimizar a varredura. Nessa varredura é atribuído pesos para reponderar exemplos e criar diversidade no conjunto de dados tido como base de treinamento. Esse método usa pesos para ponderar o impacto de diferentes modelos fracos.

2.2.3 Histogram Gradient Booster Regressor

O Histogram Gradient Booster (HGB) é um algoritmo que utiliza histogramas para otimizar a seleção de features e o processo de divisão em árvores de decisão. O HGB é mais rápido que o Gradient Boosting tradicional e pode lidar com um grande número de features. O HGB é frequentemente utilizado para problemas de

classificação e regressão, especialmente quando há um grande número de features (MAYER Et Al., 2022; FISHER, 1958).

Segundo Kunapuli (2023), para compreender a necessidade da aprendizagem em árvores baseada em histogramas, precisamos lembrar como o algoritmo de árvore de decisão constrói uma árvore de regressão. Ele o faz de cima para baixo, um nó de decisão por vez. A maneira padrão de fazer isso é pré-ordenar os valores dos recursos, enumerar todas as divisões possíveis e, em seguida, avaliá-las para encontrar a melhor.

Uma alternativa é reorganizar os valores dos recursos em um pequeno número de intervalos (bins). Neste exemplo hipotético, suponhamos que agrupamos cada coluna de recursos em 100 bins. Agora, para encontrar a melhor divisão, precisamos pesquisar apenas 10.000 divisões ($100 \times 100 = 10^4$), o que pode acelerar drasticamente o treinamento.

A versão 0.21 do scikit-learn introduziu uma versão do gradient boosting chamada HistogramGradientBoostingClassifier que implementa o gradient boosting baseado em histogramas, o que melhora significativamente o tempo de treinamento.

2.2.4 Gradient Boosting Regression

O Gradient Boosting Regression (GBR) é um algoritmo de boosting que utiliza árvores de decisão para modelar a relação entre features e a variável target. O GBR é um algoritmo robusto e versátil que pode ser usado para diversos tipos de problemas de regressão (FRIEDMAN, 2001; HASTIE Et Al., 2009; FRIEDMAN, 2002). O GBR é frequentemente utilizado para previsão de valores contínuos, como preço de um imóvel ou temperatura. Segundo Kunapuli (2023), o Gradient Boosting é um método de conjunto que treina uma sequência de modelos fracos para aproximar o gradiente a cada iteração. Esse método, junto com seu sucessor, o Newton Boosting, é considerado atualmente o estado da arte em conjuntos de árvores de decisão e amplamente utilizado em diversas aplicações.

Antes de implementarmos nosso próprio modelo, vamos contrastar o Gradient Boosting com outro método de boosting conhecido, o AdaBoost. Esta comparação nos dará uma base para entender melhor o funcionamento do Gradient Boosting.

Um componente chave dos métodos de conjunto sequencial, como AdaBoost e Gradient Boosting, é que eles treinam um novo modelo fraco a cada iteração para corrigir os erros cometidos pelo modelo anterior. No entanto, a forma como esses

métodos identificam quais exemplos focar difere bastante. O AdaBoost usa pesos para priorizar exemplos mal classificados. Exemplos com maior peso recebem mais atenção do modelo fraco na iteração atual. Já o Gradient Boosting usa resíduos (a diferença entre o valor real e o previsto) para identificar quais exemplos precisam de mais atenção na próxima iteração.

Para um exemplo específico, o resíduo é simplesmente o erro entre o valor real e a previsão realizada pelo modelo. Intuitivamente, um exemplo corretamente classificado terá um resíduo pequeno, enquanto um exemplo mal classificado terá um resíduo maior.

O Gradient Boosting usa esses resíduos para direcionar o modelo a aprender com os exemplos onde ele cometeu os maiores erros nas iterações anteriores. Com isso, ele obtém um desempenho geral melhor ao longo do tempo.

2.2.5 Random Forest Regression

O Random Forest Regressor (RFR) é um algoritmo de ensemble que combina várias árvores de decisão para melhorar a acurácia e a robustez do modelo. O RFR é menos propenso a overfitting do que outras técnicas de regressão e pode lidar com um grande número de features. O RFR é frequentemente utilizado para previsão de valores contínuos, como preço de um imóvel ou temperatura (Breiman, 2001; Geurts; Ernst; Wehenkel, 2006).

Segundo Kunapuli (2023), “Floresta aleatória” se refere especificamente a um conjunto de árvores de decisão aleatórias construídas usando a técnica de bagging. Florestas aleatórias realizam amostragem bootstrap para gerar um subconjunto de treinamento (exatamente como o bagging) e, em seguida, utilizam árvores de decisão aleatórias como estimadores base.

Estas árvores são treinadas usando um algoritmo de aprendizado de árvore de decisão modificado, que introduz aleatoriedade durante o crescimento das árvores. Essa fonte adicional de aleatoriedade aumenta a diversidade do conjunto e geralmente leva a um melhor desempenho preditivo.

A principal diferença está em como é construído um nó de decisão. Na construção padrão, todas as features disponíveis são avaliadas exaustivamente para identificar a melhor para a divisão. Como o aprendizado de árvore de decisão é um algoritmo ganancioso, ele escolherá as features com maior pontuação para dividir.

No bagging, essa enumeração exaustiva (combinada com aprendizado ganancioso) significa que muitas vezes é possível que um mesmo pequeno número de features dominantes seja usado repetidamente em diferentes árvores. Isso torna o conjunto menos diverso.

Para superar essa limitação, as florestas aleatórias introduzem um elemento adicional de aleatoriedade no aprendizado da árvore. Em vez de considerar todas as features para identificar a melhor divisão, um subconjunto aleatório de features é avaliado para identificar a melhor para a divisão.

Dessa forma, as florestas aleatórias usam um algoritmo de aprendizado de árvore modificado, que primeiro seleciona aleatoriamente features antes de criar um nó de decisão. A árvore resultante é uma árvore de decisão aleatória, que é um novo tipo de estimador base.

Como você verá, as florestas aleatórias essencialmente estendem o bagging usando árvores de decisão aleatórias como estimadores base. Portanto, as florestas aleatórias contêm dois tipos de randomização: Amostragem aleatória de features: para aprender árvores de decisão aleatórias.

2.2.6 Análise de correlação

As matrizes de correlação são ferramentas estatísticas que medem a força e a direção da relação entre duas variáveis aleatórias, e podem ser causal ou não causal. Pode ser definida como qualquer tipo de associação, na estatística é descrito como linearmente dependentes.

Correlações são muito úteis pois podem indicar uma relação de predição que pode ser explorada no cotidiano do usuário. De certa forma, a correlação é uma dependência entre variáveis de estudo, o que vai de encontro ao princípio de variáveis aleatórias, esse princípio é levado em conta a independência das variáveis.

A correlação será mensurada utilizando três técnicas: coeficiente de correlação de Pearson, spearman e kendall. O coeficiente de correlação de Pearson é adequado quando a relação entre as variáveis é linear, no qual fornece um direcionamento entre força e direção da correlação entre variáveis analisadas.

O coeficiente de Spearman pode ser empregado quando a relação entre variáveis pode ser não linear, mas ainda mantém a monotonicidade, podendo ser útil em datasets que não possuem uma distribuição normal como forma de medição mais maleável entre as variáveis.

O coeficiente de Kendall é utilizado tanto em relação linear quanto não linear, além de datasets robustos modelados em pequenos conjuntos de dados como forma de medição mais precisa das relações monotônicas do que a de Spearman.

Portanto, a escolha das correlações de Pearson, Spearman, e Kendall provê uma ampla análise das correlações, envolvendo relações lineares e não lineares além de diferentes tipos de dados para que qualquer tipo de variável possa ser identificado, não dependendo de suas características de distribuição.

A seguir será apresentado os três métodos de mensuração da correlação entre o respectivo conjunto de dados.

2.2.6.1 Coeficiente de correlação de Pearson

O coeficiente de correlação de Pearson é amplamente utilizado para medir a força e a direção da relação linear entre duas variáveis. Ele é definido como a razão entre a covariância das duas variáveis e o produto de seus desvios padrão, representando, portanto, uma forma de normalização da covariância (Kendall, 1970).

Esse coeficiente assume valores entre -1 e 1, onde +1 indica uma correlação linear perfeita positiva, -1 uma correlação linear perfeita negativa, e 0 a ausência de correlação linear. No entanto, uma limitação importante do coeficiente de Pearson é sua sensibilidade a outliers, que podem distorcer significativamente o valor da correlação, reduzindo sua precisão como medida da relação linear entre as variáveis (Mukaka, 2012).

Assim, o uso do coeficiente exige cuidado especial na análise de dados com pontos extremos, pois esses podem inflar ou reduzir artificialmente a correlação observada. A equação é descrita a seguir:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}} \quad (2.79)$$

,sendo $x_1, x_2, \dots, x_n$ e $y_1, y_2, \dots, y_n$ os valores a serem mensurados. Vale ressaltar que as seguintes equações $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i)$ e $\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n (y_i)$ representam as médias aritméticas de ambas variáveis a serem medidas.

2.2.6.2 Coeficiente de correlação de postos de Spearman

O coeficiente de correlação de Spearman é semelhante ao coeficiente de correlação de Pearson. Enquanto a correlação de Pearson foca nas relações lineares entre os conjuntos, a correlação de Spearman foca nas relações monotônicas, no

qual define se é linear ou não. Essa correlação não assume que as variáveis são distribuídas normalmente. É menos sensível a outliers do que a correlação de Pearson.

A equação que calcula a correlação de Spearman segundo Zar (2014):

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n}, \quad (2.80)$$

sendo r_s coeficiente de correlação de Spearman para uma amostra de tamanho n e $\sum d_i^2$ é a soma do quadrado das variáveis pareadas.

2.2.6.3 Coeficiente de correlação de Kendall

O coeficiente de correlação de Kendall, também conhecido como τ de Kendall, é uma medida estatística que avalia a associação ordinal entre dois conjuntos de dados. Essa ferramenta é amplamente utilizada para identificar o grau de concordância entre classificações ou rankings atribuídos a elementos comuns em dois conjuntos de observações. O coeficiente assume valores entre -1 e +1, onde +1 indica concordância total entre os dois rankings (classificações idênticas), -1 representa completa discordância (um ranking é o inverso do outro), e 0 sugere ausência de correlação ordinal (Kendall, 1970).

O cálculo do coeficiente de Kendall baseia-se na comparação de pares ordenados de elementos em dois conjuntos. Consideremos um conjunto $X=\{a,b,\dots,x,y\}$, onde os elementos estão dispostos de acordo com uma ordem específica. Para ilustrar, tomemos um exemplo com $n=4$: $S=\{a,c,b,d\}$, que possui as classificações $R=\{1,3,2,4\}$. Quando analisamos as relações entre os elementos, podemos decompor S em $2N(N-1)$ pares ordenados distintos associados ao conjunto S . Esses pares representam todas as possíveis combinações de elementos no conjunto.

No contexto de Kendall, a comparação entre dois conjuntos ordenados envolve a contagem de pares concordantes e discordantes:

- Pares concordantes: Para dois elementos i e j , a classificação relativa (ordem) em ambos os conjuntos é a mesma.

- Pares discordantes: Para os mesmos elementos i e j , a ordem relativa difere entre os conjuntos.

Para calcular o coeficiente, introduz-se o conceito de diferença simétrica entre dois pares $P1$ e $P2$, representada como $d(P1,P2)$. A diferença simétrica é utilizada para avaliar discrepâncias entre os dois conjuntos de dados ordenados.

A Fórmula que calcula o coeficiente de Kendall sendo Abdi (2007) é:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (2.81)$$

Sendo:

- n_c é o número de pares concordantes,
- n_d é o número de pares discordantes,
- $n(n-1)/2$ é o número total de pares.

Esse cálculo fornece uma métrica robusta para avaliar a associação ordinal entre os dois conjuntos, especialmente em situações onde os dados não seguem distribuições normais ou contêm outliers (Kendall, 1970).

Portanto o modo de interpretar o coeficiente de Kendall sendo que τ é baseado no montante de pares diferentes entre dois conjuntos ordenados, e podem ser interpretados como a diferença da probabilidade entre objetos estarem na mesma ordem [$P(\text{iguais})$] e a probabilidade de os objetos estarem em ordem diferente [$P(\text{diferente})$], conforme a equação a seguir:

$$\tau = P(\text{igual}) - P(\text{diferente}) \quad (2.81)$$

Comumente o intervalo de -1 até 1 se divide em dois intervalos de (-1,0) e de (0,1) onde as correlações tem um comportamento monotonicamente decrescente e crescente, respectivamente. Por outro lado, se identificam os casos extremos (0) como totalmente descorrelacionados e (1) como totalmente correlacionados, assim como no intervalo de (0,90 - 0,99) se diz que as variáveis envolvidas estão fortemente correlacionadas, no intervalo (0,70 - 0,89) a correlação é forte, de (0,40 - 0,69) moderada, de (0,20 - 0,39) fraca e de (0,01 - 0,19) muito fraca.

2.2.7 Análise de performance do modelo

A análise da performance do modelo gerado é de grande importância, para que possa ser avaliada sua qualidade e eficácia. Tais métricas fornecem uma visão quantitativa sobre como o modelo está se comportando em relação ao problema a ser abordado. Em problemas que envolvam regressão, métricas como Mean Absolute Error (MAE), Mean Squared Error (MSE) e o Coeficiente de Determinação (R^2) medem a capacidade de predição de valores contínuos, tendo em consideração tanto a precisão dos erros quanto à proporção de variabilidade dos dados. A escolha da devida métrica depende do atendimento aos objetivos a serem alcançados e o balanço dentre diversos tipos de erros a serem analisados.

2.2.7.1 Coeficiente de determinação(R^2)

Um dos parâmetros de performance utilizado para caracterizar os modelos de ML é o coeficiente de determinação ou mais conhecido como R^2 . Segundo (Hahn, 1973), o coeficiente de determinação é a proporção da possível variabilidade da variável dependente cujo cálculo é obtido mediante uma regressão das variáveis independentes.

Um valor para R^2 igual a 1 indica que a equação de regressão leva em conta a variabilidade dos dados da respectiva variável dependente contida no espaço amostral. Se o valor for 0, aponta que o respectivo modelo da regressão não leva em conta qualquer variabilidade.

A ausência de informações sobre as variáveis independentes torna a predição do valor para uma variável Y ser $\bar{Y}$ para uma amostra de tamanho n . Sendo $\bar{Y}$ a média das observações da amostra de tamanho n . O cálculo de sua variabilidade total para um tamanho n é dado na seguinte fórmula:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (2.82)$$

Quando os valores associados com as variáveis independentes são conhecidos para um futuro valor Y , então o valor $\hat{Y}$ é obtido com o ajuste da equação de regressão, que nos leva ao erro de predição $Y - \hat{Y}$. Sendo $\hat{Y}$ o valor estimado que vai ser objeto de previsão. A medida da variabilidade depois do ajuste é dada pela equação da soma dos quadrados dos desvios em torno de $\hat{Y}$:

$$\sum_{i=1}^n (Y_i - \hat{Y})^2 \quad (2.83)$$

O respectivo coeficiente de determinação é calculado com o complementar da razão das equações acima descritas conforme a seguinte equação:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.84)$$

As vantagens do uso do coeficiente de determinação são:

- O coeficiente de determinação explica o quão o modelo se ajusta aos dados.
- O intervalo do resultado do coeficiente de determinação ser entre 0 e 1 facilita a comparabilidade entre modelos, não sendo dependente da escala dos dados originários.

Desvantagens do uso do coeficiente de determinação são:

- Pode não informar com precisão se o modelo está impreciso, pois não garante um erro absoluto baixo.
- Em modelos complexos pode ser equivocado, gerando a impressão de que o modelo é bom sem ser realmente.
- É mais adequado em modelos lineares, pois em modelos não lineares pode não refletir a qualidade do modelo.

2.2.7.2 Erro médio absoluto

Outra métrica utilizada para mensurar a efetividade entre o valor predito e o valor obtido pela simulação dos arranjos de amplificação Raman é o erro médio absoluto (MAE — do inglês Mean Absoluto Error).

Essa medida descreve os erros entre observações de eventos semelhantes. O binômio de medição é entre o que é predito com o que é observado. É a soma dos erros absolutos, que em outras literaturas é denominada também como distância de Manhattan.

A fórmula de cálculo é descrita a seguir:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, \quad (2.85)$$

,sendo Y_i o valor obtido pelas simulações e $\hat{Y}_i$ o valor predito pelo modelo de aprendizado de máquina. Esse modelo possui o fator módulo para tratar os chamados

outliers, que são valores muito acima, porém o uso do módulo não aponta para onde se é para mais ou para menos.

As vantagens do uso do erro médio absoluto são:

- Interpretação simples devido à similaridade de valores reais e preditos.
- Em comparação a outras medidas é menos sensível a outliers.
- Possui uma homogeneidade entre os erros dos valores base, útil em cenário que é necessário o devido equilíbrio.

As desvantagens são:

- Dificuldade de destaque de diferenças entre modelos, podendo tornar inviável se os modelos forem semelhantes.
- A aplicação do MAE em algoritmos que utilizam derivadas tal como redes neurais, torna-se pouco eficiente pois sua função de perda não é diferenciável.

2.2.7.3 Erro médio quadrático

Além das métricas acima descritas, outro método de cálculo de correlação é o erro quadrático médio (MSE — do inglês Mean Squared Error) que possui a seguinte fórmula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y})^2, \quad (2.86)$$

,sendo Y_i o valor obtido pelas simulações e $\hat{Y}_i$ o valor predito pelo modelo de aprendizado de máquina. Diferente do MAE, o MSE usa o valor ao quadrado da diferença entre o valor real e o predito. A análise do MSE pode ser simplificada para quanto o maior valor de MSE, o modelo não está performando bem.

Vantagens do uso do erro médio quadrático são:

- Penaliza grandes erros devido ao cálculo do erro ao quadrado, podendo ser desejável em situações que erros devam ser evitados.
- Melhor aplicado em datasets de grande tamanho, pois pode garantir que o modelo foque em minimizar desvios grandes.

As desvantagens são:

- Valores atípicos podem influenciar no erro total, podendo distorcer o desempenho do modelo.

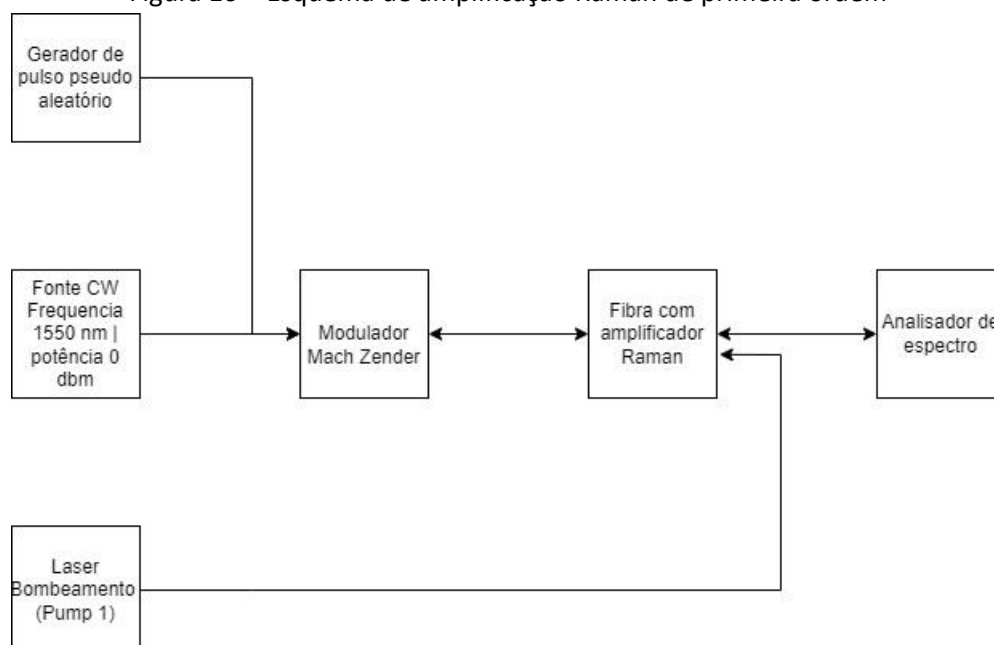
-
- A interpretação dos resultados pode ser prejudicada pois a métrica não está na escala igual aos dos dados originais.

3 METODOLOGIA

3.1 Metodologia de simulação do amplificador Raman

A simulação da amplificação Raman foi feita por meio do software *Optisystem* 7, cuja plataforma foi desenhada com o intuito de simular e testar componentes e sistemas de comunicação óptica com o objetivo de obter resultados semelhantes ao de experimentos físicos. A escalabilidade e a agilidade na adição de componentes tornam o ambiente *Optisystem* robusto e confiável para simulação de experimentos. O arranjo da simulação consiste na seguinte listagem de componentes:

Figura 10 – Esquema de amplificação Raman de primeira ordem



Fonte : próprio autor

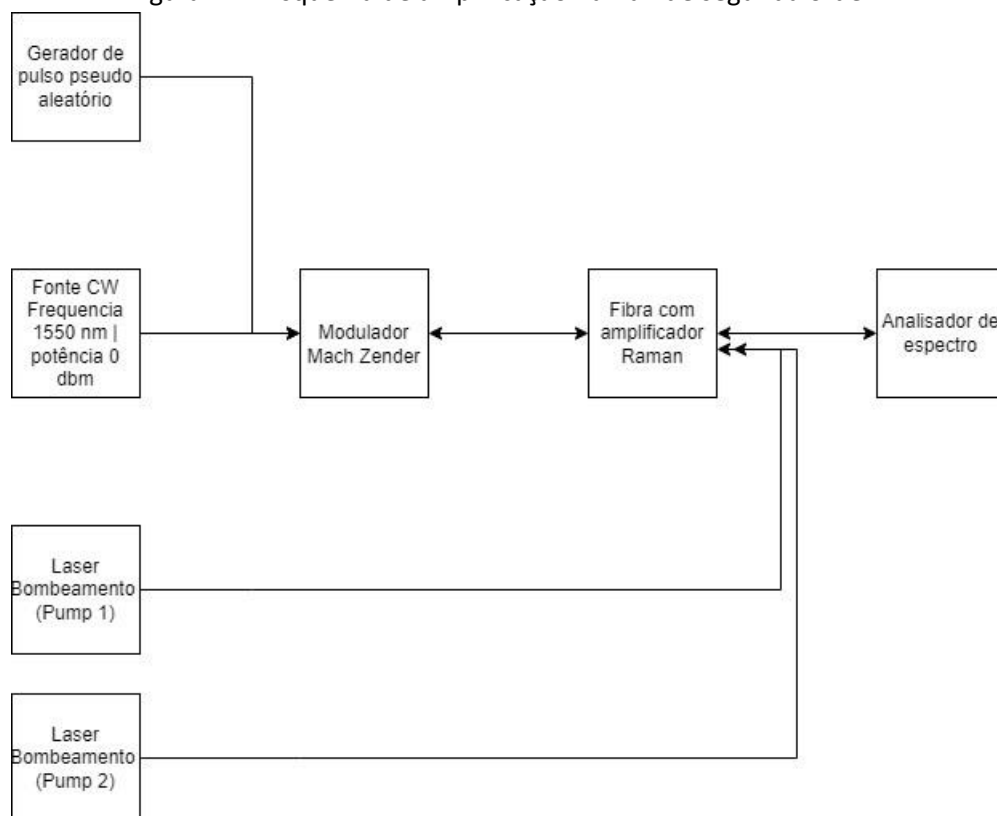
O arranjo mostrado na Figura 10 irá gerar um grande conjunto de dados para ser processado junto aos algoritmos de aprendizado de máquina. Diversos parâmetros serão levados em conta tais como o comprimento de onda e o fator de qualidade da comunicação óptica. Esses parâmetros irão direcionar para uma análise mais fidedigna do modelo a ser levado em conta para cada algoritmo de aprendizado de máquina.

A correlação com o mundo real de como adquirir esses parâmetros utiliza sensores ópticos para capturar dados em tempo real sobre diversos aspectos da comunicação, como intensidade do sinal, frequência, fase e polarização. Essa captura em tempo real permite a análise de eventos dinâmicos e a detecção de anomalias

com alta precisão, o que é crucial para aplicações como monitoramento de redes ópticas, diagnóstico de falhas e segurança de dados.

Para representar o cenário alvo da simulação para gerar os dados de amplificação Raman de primeira e segunda ordem, segundo 2, serão definidos como dados de entrada para o laser de bombeamento 1 (pump 1) de 1400 nanômetros até 1600 nanômetros, além do limite de potência de 200 miliwatts até 400 miliwatts e também variando a extensão do enlace de 10 km até 110 km. para a amplificação Raman de segunda ordem serão variados os parâmetros de 1543 nanômetros até 1553 nanômetros para as duas fontes de laser de bombeamento 1 e 2, fixou a potência de 150 miliwatts para o laser de bombeamento 1 (pump 1) e variou a potência do segundo laser de bombeamento (pump 2) de 10 miliwatts até 110 miliwatts, além de variar o comprimento do enlace de 10 quilômetros até 110 quilômetros.

Figura 11 – Esquema de amplificação Raman de segunda ordem



Fonte: próprio autor

O conjunto de dados gerado pelo esquema de comunicação óptica mostrado nas Figuras 10 e 11 possui vários arranjos e pode ser utilizado para alimentar diversos algoritmos de aprendizado de máquina em no contexto de amplificação Raman. O

esquema de comunicação óptica apresentado, em conjunto com o grande conjunto de dados que ele gera, abre vários campos de estudo para o Machine Learning na área de comunicações ópticas. As aplicações inovadoras que podem ser desenvolvidas com base nesse esquema são vastas e promissoras, com potencial para otimizar o monitoramento e a qualidade da transmissão de comunicações ópticas.

Tabela 2 – Parâmetros utilizados na simulação da amplificação Raman de primeira e segunda ordem.

	Raman de primeira ordem	Raman de segunda ordem contrapropagante
Frequência pump 1	1400 nm - 1600 nm	1543 nm -1553 nm
Frequência pump 2	Não possui	1543 nm -1553 nm
Potência Pump 1	200 mW - 400 mW	150 mW
Potência Pump 2	Não possui	10 mW - 110 mW
Distância	10 km -110 km	10 km -110 km

4 RESULTADOS

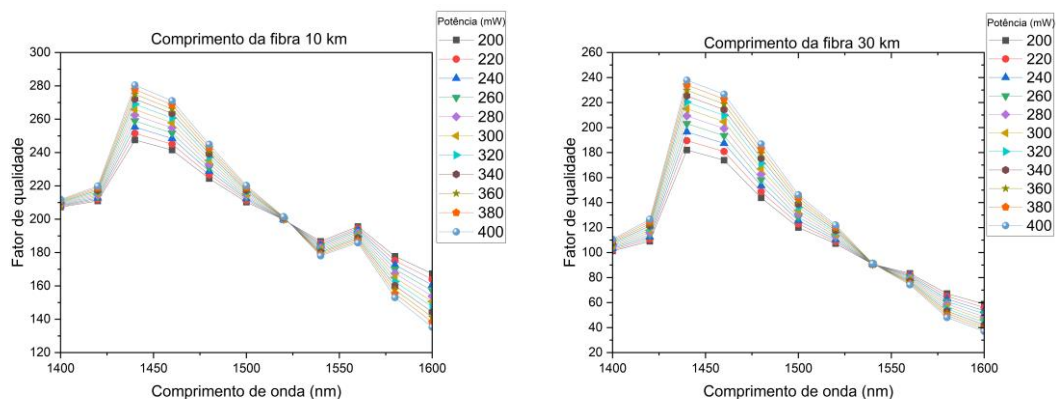
Nesse capítulo serão apresentados os resultados obtidos nas simulações usando o software *Optsystem* de um sistema que descreve um amplificador Raman de primeira e segunda ordem. Também será realizado uma análise de correlação dos dados sintéticos obtidos da simulação e posteriormente mediante algoritmos de aprendizado de máquina serão treinadas as bases de dados obtidas.

4.1 Amplificação Raman de primeira ordem

Como foi descrito na seção de metodologia, foi realizado um estudo da variação da distância e a frequência do laser de bombeamento, e a disposição do sistema de amplificação Raman de primeira ordem e de segunda ordem, no qual são modelados de forma contrapropagante em relação ao sinal incidente (Figura 10).

Nas Figuras 12 ,13 e 14, destaca-se a variação de parâmetros que estão contidos no arranjo de amplificação Raman de primeira ordem. Onde pode se observar as mudanças do fator de qualidade do enlace óptico que é representado pela variável MaxQFactor e a distância do enlace.

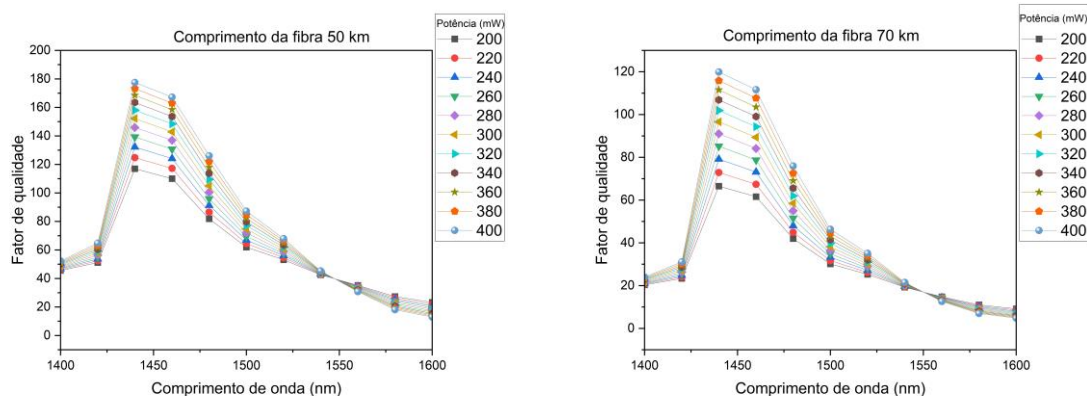
Figura 12 – Amplificação Raman de primeira ordem com fibra óptica de 10 km e 30 km de extensão.



Fonte : próprio autor

Além disso, segundo Stolen (2004), em um determinado arranjo de um sinal óptico e o laser de bombeamento, caso o comprimento de onda do sinal for menor do que o comprimento de onda do laser de bombeamento, o sistema atuará com absorção não-linear em vez de amplificação.

Figura 13 – Amplificação Raman de primeira ordem com fibra óptica de 50 km e 70 km de extensão.



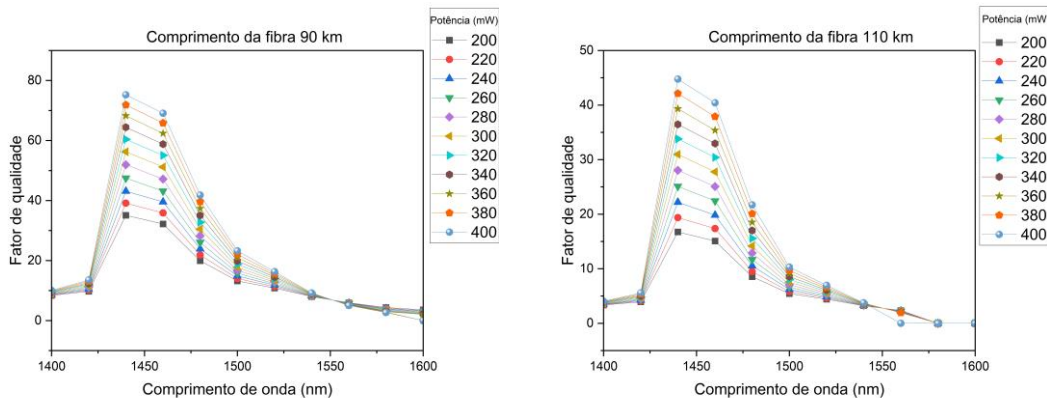
Fonte : próprio autor

A respectiva curva de ganho é ampla e se aproxima da curva característica do vidro, que no caso é o componente principal da fibra óptica. Vale ressaltar que o comprimento de onda do sinal da simulação é de 1550 nanômetros.

Analisando a Figura 12, podemos observar que existe um ponto que possui comprimento de onda de 1520 nm, no qual a potência não influencia no respectivo fator de potência variando de 200 miliwatts até 400 miliwatts. Esse ponto inicia um padrão no qual coincide com o descumprimento do padrão para que ocorra o devido

espalhamento Raman, que em vez da amplificação da fonte do sinal de 1550 nanômetros, tal efeito é indiferente em relação ao espalhamento do sistema.

Figura 14 – Amplificação Raman de primeira ordem com fibra óptica de 90 km e 110 km de extensão.



Fonte : próprio autor

Devido ao efeito Raman e suas componentes Stokes e anti-Stokes terem o seu funcionamento explicado por meio da Figuras 3, nota-se que a variação de potência no ponto de 1520 nanômetros sobre o efeito que ocorreu é pertencente a dependência do cálculo do respectivo fator de qualidade do sinal óptico (MaxQFactor) com a relação sinal-ruído óptico (OSNR) segundo as equações 2.74 e 2.76. As equações indicam uma função da potência do sinal de entrada que influencia no efeito Rayleigh no qual não há interação no espalhamento dentro da fibra óptica, ocasionando somente a passagem do sinal óptico sem a interação junto ao material, indo de encontro ao efeito de amplificação desejado.

Em um contexto de amplificação Raman, inicialmente é notado que a maior frequência de incidência do laser de bombeamento implica em uma amplificação Raman de maior efetividade, e conseqüentemente menor comprimento de onda, há uma região que o efeito anti-Stokes(atenuação), Rayleigh (estagnado) e Stokes(amplificação) segundo a Figura 3. Vale ressaltar que isso leva a explicação da interação do respectivo laser do sinal de 1550 nm e do laser de bombeamento, influenciando o fator de qualidade diretamente, pois é calculado por meio da relação sinal-ruído. O efeito de amplificação Raman não tem com a geração do respectivo efeito Rayleigh que foi notado na faixa entre 1500 e 1540 nanômetros, Stokes na faixa de até 1500 nanômetros e anti-Stokes além de 1540 nanômetros, em que há um decréscimo do fator de qualidade do sinal nessa faixa descrita como anti-Stokes.

Já na análise do enlace com 30 quilômetros de distância da Figura 12 (direita), é observado que o respectivo ponto descrito como relacionado ao efeito Rayleigh fica no comprimento de onda do laser de bombeamento de 1540 nanômetros e o relacionado ao efeito Stokes em 1520 nanômetros e o anti-Stokes em 1560 nanômetros, permanecendo o padrão de disposição das três faixas de amplificação, estagnação e atenuação do sinal óptico incidente.

Desde que o comprimento do enlace cresce além dos 50 km, o efeito anti-Stokes é atenuado, não gerando magnitudes contrárias de cada potência do laser de bombeamento até o comprimento de 110 km, no qual o padrão é quebrado e tal efeito não é mais percebido pela análise gráfica como se pode observar na Figura 14.

Além disso vale ressaltar a evolução entre o respectivo comprimento de onda de 1440 nm e na potência de 400 mW que fornece o máximo fator de qualidade em todos os cenários. A tabela 3 fornece a razão relativa ao link de 10 km, que será tomado como base e o respectivo decaimento do fator de qualidade do enlace até o valor de 110 quilômetros.

Tabela 3 – Razão do valor do fator de qualidade da amplificação Raman de primeira ordem.

Distância (km)	10	30	50	70	90	110
Porcentagem de fator de qualidade em relação a 10 km	100 %	73,53 %	47,26 %	26,86 %	14,17 %	6,76 %

Em todos os casos investigados observamos inicialmente um aumento do fator de qualidade com o comprimento de onda do pump, até um valor máximo para comprimentos de onda próximos a 1440 nm. Posteriormente é observado um decaimento ou diminuição do fator de qualidade com uma inversão entorno de 1540 nm em relação a potência do sinal, o que pode ser atribuído ao início da zona de transição ao efeito Rayleigh e a componente anti-Stokes. Vale ressaltar que ao passo do aumento da distância do link para o arranjo do Raman de primeira ordem, o máximo valor observado do fator de qualidade do enlace diminui.

4.1.1 Matriz de correlação

A noção de correlação é representada por meio de um mapa de calor, no qual é uma forma de visualização de dados que associa cores a respectiva intensidade ou frequência da variável objeto. No caso em estudo a correlação entre variáveis de estudo. O objeto de estudo é distinguir quais valores estão mais próximos de 0 (variáveis totalmente descorrelacionadas) e os que estão mais próximos de 1/-1

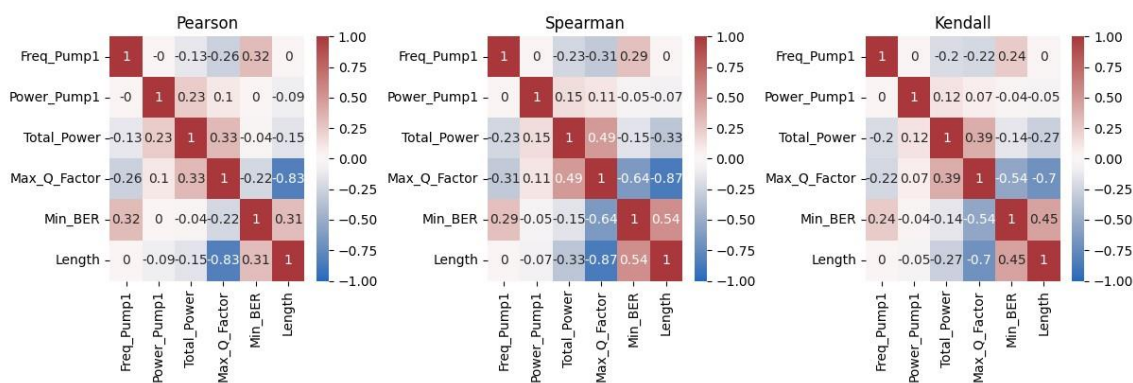
(variáveis totalmente correlacionadas de forma monotonicamente crescente e decrescente, respectivamente).

A Figura 15 representa a relação dos coeficientes de correlação para o arranjo de primeira ordem. Os atributos levados em conta para o processamento são as entradas: Frequency (Frequência), Power (potência), Length (Distância) e as saídas: MaxQFactor (Fator de qualidade) e MinBer.

Quando analisamos o arranjo de amplificação Raman de primeira ordem em relação a variável de saída MaxQFactor e os parâmetros de entrada, podemos observar que a variável (i-) Distância: possui uma correlação moderada negativa; (ii-) Frequência: correlação moderada negativa; (iii-) Potência: correlação moderada positiva, para as correlações de Pearson (ρ), Spearman (r_s) e Kendall (τ).

O mapa de calor fornece uma visão sobre qual modelo de correlação deve ser levado em conta, o foco principal de qual modelo a ser escolhido é o maior valor para os valores da correlação.

Figura 15 – Matriz de correlação para o arranjo com um pump de bombeamento.



Fonte : próprio autor

Pela análise dos mapas nota-se que o modelo Spearman possui maior incidência de relacionamentos monotonicamente crescentes, representados pelas cores mais vermelhas e por isso será tomado como base na análise.

Tendo como a correlação de Spearman como base, pois os dados não possuem características de distribuição normal e a relação complexa dos dados com a tendência da não linearidade, a escolha do Spearman é mais vantajosa.

4.1.2 Performance dos modelos

Os modelos Decision Tree (DTR) e Random Forest (RFR) mostraram um desempenho bastante sólido ao longo das diversas proporções de dados de treino, com valores de R^2 extremamente próximos de 1 em todos os cenários. Isso significa que eles conseguem explicar praticamente toda a variação presente nos dados de treino, o que é um excelente indicativo de que o modelo está conseguindo capturar bem os padrões do dataset.

Entre os dois, o Random Forest (RFR) teve uma leve vantagem sobre o Decision Tree, apresentando valores de R^2 ligeiramente mais altos. Essa diferença ocorre devido à natureza do Random Forest, que é uma técnica de ensemble, onde diversas árvores de decisão são combinadas, o que aumenta sua robustez e reduz a chance de ele cometer erros devido ao viés de uma única árvore.

Tabela 4 – Performance do modelo para 1 pump usando o R^2 para os algoritmos Decision Tree e Random Forest.

Porcentagem	R^2 DTR	R^2 RFR
0,1	0,9672045265648102	0,9701218553562915
0,2	0,9849904791840066	0,9929318088238813
0,3	0,9880431328249591	0,9952479947541101
0,4	0,9947252968033242	0,9972466394247157
0,5	0,9969135285642996	0,9974833015641209
0,6	0,9980840696819597	0,9982290842495706
0,7	0,9981715712902653	0,9982802340406640
0,8	0,9986578482701366	0,9990235069985891
0,9	0,9989991038032925	0,9992456829390056

Tabela 5 – Performance do modelo para 1 pump usando o R^2 para os algoritmos Gradient Boosting e Histogram Boosting.

Porcentagem	R^2 GBR	R^2 HBR
0,1	0,9828039754372725	0,9485086162576118
0,2	0,9944838116361633	0,9656162196017647
0,3	0,9949424697879509	0,9974032474815374
0,4	0,9971265962627144	0,9979254001704363
0,5	0,9971835954120785	0,9988251460035741
0,6	0,9975136156524650	0,9986349927203029
0,7	0,9974955649575942	0,9987761735300889
0,8	0,9976165309435869	0,9993410607846293
0,9	0,9982652638119929	0,9989850575368355

Apesar do ótimo desempenho, valores de R^2 próximos de 1 podem ser um sinal de alerta para sobreajuste (overfitting). Isso quer dizer que o modelo pode estar aprendendo os detalhes e ruídos específicos dos dados de treino, e não apenas os padrões gerais. Embora isso seja desejável durante o treino, pode ser problemático na hora de generalizar e fazer previsões em novos dados.

Tabela 6 – Performance do modelo para 1 pump usando o R^2 para o algoritmo Adaboost.

Porcentagem	R^2 ABR
0,1	0,9399982195157092
0,2	0,9259864081395256
0,3	0,9406690343345586
0,4	0,9478830785682990
0,5	0,9326773570118032
0,6	0,9322833599741827
0,7	0,9411845100757552
0,8	0,9384678805529710
0,9	0,9371150221825989

O Adaboost (ABR), por outro lado, teve um desempenho ligeiramente inferior comparado aos outros modelos. Os valores de R^2 para o Adaboost ficaram em torno de 0,92 a 0,94, o que, embora ainda seja um bom desempenho, não se compara ao nível de ajuste apresentado pelos outros modelos.

O fato de o Adaboost ter um desempenho um pouco inferior pode ser um reflexo de sua maior sensibilidade a ruídos nos dados, o que torna mais difícil a tarefa de generalização quando o volume de dados de treino aumenta. O Adaboost é projetado para se concentrar em exemplos difíceis durante o treinamento, mas isso também pode levar o modelo a dar muita importância a pontos fora da curva, o que pode ser prejudicial para a generalização.

A análise de desempenho é uma principal ferramenta para julgar se o modelo está tendendo e convergindo à devida predição. vale ressaltar que para mensurar a performance do modelo irá variar a base de dados de 10 até 90 por cento entre a proporção de treinamento e teste. Os gráficos 16 e 17 estão representados a porcentagem da base de dados do treinamento.

A análise do desempenho usando o MAE como parâmetro é descrita a seguir. A análise de desempenho dos modelos utilizando o erro médio absoluto (MAE) é uma ferramenta fundamental para avaliar se os modelos estão convergindo para uma predição consistente e confiável. Com o objetivo de medir essa performance, foi feita uma análise considerando diferentes proporções de treino e teste.

Os modelos Decision Tree Regressor (DTR) e Random Forest Regressor (RFR) apresentaram valores de MAE relativamente baixos ao longo de todas as proporções de treino, conforme podemos observar na Tabela 7. Esses valores indicam que ambos os modelos estão cometendo erros menores em termos de previsões absolutas.

- DTR: Começou com um MAE de aproximadamente 7,36 quando 10% dos dados eram usados para treinamento, e esse valor foi progressivamente reduzido até 1,52 quando a proporção de treino chegou a 90%. Isso demonstra que o DTR é bastante beneficiado por um maior volume de dados de treinamento, o que reduz seus erros de predição. No entanto, o modelo se mostra sensível à quantidade de dados, apresentando erros maiores com menores proporções de treino.

- RFR: O Random Forest apresentou um desempenho mais consistente e ligeiramente melhor do que o DTR, com MAE inicial de 7,69 e caindo para 1,18 com 90% de treino.

Tabela 7 – Erro Médio Absoluto (MAE) para 1 pump para os algoritmos Decision Tree e Random Forest.

Porcentagem	MAE DTR	MAE RFR
0,1	7,3613134732888135	7,6948229237729350
0,2	4,5520661558685450	3,5339412107981207
0,3	4,2197984753218885	2,7147700652467870
0,4	2,9983725231539430	2,2663490593366764
0,5	2,4433118498498500	2,0419512510660707
0,6	1,9280214896810506	1,6640024717073176
0,7	1,9328735500000003	1,6435319765749992
0,8	1,7429654456928838	1,3259871512359556
0,9	1,5175995074626870	1,1866162487313485

Isso se deve à natureza de ensemble do RFR, onde múltiplas árvores são combinadas para reduzir o viés e a variância, resultando em maior precisão.

Os modelos Gradient Boosting Regressor (GBR) e Histogram Boosting Regressor (HBR) também mostraram bons resultados, mas com algumas diferenças interessantes, conforme evidenciado na Tabela 8.

- GBR: O Gradient Boosting começou com um MAE de aproximadamente 5,70 e reduziu para 2,12 com 90% de treino. Esse comportamento sugere que o GBR é capaz de aprender de forma consistente conforme mais dados são fornecidos, embora os erros ainda sejam um pouco maiores do que os do RFR.
- HBR: O Histogram Boosting, por outro lado, começou com um MAE relativamente alto de 11,60 e apresentou uma queda significativa, alcançando 1,27 com 90% dos dados de treino. Isso indica que o HBR tem uma sensibilidade elevada à quantidade de dados de treino, ajustando-se melhor quando um volume maior de informações está disponível.

Ambos os modelos de boosting mostraram ser promissores, mas foram menos eficientes com proporções de treino menores, especialmente o HBR. Eles precisam de um volume maior de dados para poder fazer predições precisas.

Tabela 8 – Erro Médio Absoluto (MAE) para 1 pump para os algoritmos Gradient Boosting e Histogram Boosting.

Porcentagem	MAE GBR	MAE HBR
0,1	5,7082344150838170	11,604767285682410
0,2	3,2019847043835270	8,6039124565943420
0,3	3,0972683761598234	2,4306042420290317
0,4	2,5411086554800120	2,0515917412243270
0,5	2,5637397386561530	1,5042771998386208
0,6	2,3020477994863160	1,3624742105621594
0,7	2,4598086611440966	1,3527412745185208
0,8	2,3583538908674564	1,1029835479020458
0,9	2,1279712712636420	1,2737903378033029

Tabela 9 – Erro Médio Absoluto (MAE) para 1 pump para o algoritmo Adaboost.

Porcentagem	MAE ABR
0,1	15,490862485887058
0,2	15,203142920864440
0,3	15,514893095894825
0,4	14,969217793946063
0,5	17,197095587732154
0,6	16,767254653259005
0,7	15,577021174056524
0,8	16,170344145893220
0,9	17,826414659405113

Tabela 10 – Erro Quadrático Médio (MSE) para 1 pump para os algoritmos Decision Tree e Random Forest.

Porcentagem	MSE DTR	MSE RFR
0,1	180,58395721083957	164,52007026339462
0,2	82,94576625967392	39,060309810045624
0,3	65,71772515711552	26,118126941014270
0,4	29,497244133103777	15,397368543278102
0,5	16,997895033747305	13,860026485215770
0,6	9,614983793474295	8,887236701491336
0,7	9,459568946720665	8,897390736619515
0,8	7,273256760435009	5,291714912703862
0,9	5,848226130694900	4,4074667897109245

Tabela 11 – Erro Quadrático Médio (MSE) para 1 pump para os algoritmos Gradient Boosting e Histogram Boosting.

Porcentagem	MSE GBR	MSE HBR
0,1	94,68764553647446	283,53052615136045
0,2	30,483616118083884	190,01199618558468
0,3	27,797363271129374	14,272356279544136
0,4	16,068675027080960	11,601596406353880
0,5	15,510575929626500	6,470186206889194
0,6	12,477773842187464	6,850208876952654
0,7	12,956959070971568	6,331595434544519
0,8	12,916261285606234	3,5708586415927837
0,9	10,136045613858690	5,930298320400790

O Adaboost Regressor (ABR), apresentado na Tabela 9, teve um desempenho inferior em comparação aos outros modelos. Seus valores de MAE variaram entre 15,49 e 17,82, e não mostraram uma tendência clara de diminuição conforme mais dados eram utilizados no treino. Em alguns casos, o erro até aumentou conforme a proporção de treino aumentou.

Tabela 12 – Erro Quadrático Médio (MSE) para 1 pump para o algoritmo Adaboost.

Porcentagem	MSE ABR
0,1	330,3919054854355
0,2	409,0146624771854
0,3	326,09679758313433
0,4	291,44873135336303
0,5	370,760994471388
0,6	339,8319815006332
0,7	304,28814594269375
0,8	333,44881490973387
0,9	367,43627531971407

Este comportamento sugere que o Adaboost tem dificuldades em lidar com o aumento da complexidade dos dados de treino e pode estar dando muito peso a exemplos não representativos, prejudicando a capacidade de generalização do modelo. Dessa forma, o ABR apresenta o pior desempenho entre os modelos analisados, mostrando que talvez não seja a melhor escolha para este dataset.

Os modelos Decision Tree Regressor (DTR) e Random Forest Regressor (RFR) apresentaram uma queda significativa no MSE conforme a quantidade de dados de treino aumentou, como pode ser visto na Tabela 10.

- DTR: Começou com um MSE de aproximadamente 180,58 quando 10% dos dados foram usados para treinamento, reduzindo para 5,84 quando a proporção de treino alcançou 90%.

Isso demonstra que o modelo Decision Tree se beneficia bastante de um maior volume de dados de treino.

- RFR: O Random Forest teve um desempenho melhor do que o DTR, começando com um MSE de 164,52 e caindo para 4,40 com 90% de treino. Essa melhoria constante é resultado da natureza de ensemble do Random Forest, onde várias árvores trabalham em conjunto para melhorar a precisão das previsões e reduzir a variância dos erros.

Os modelos Gradient Boosting Regressor (GBR) e Histogram Boosting Regressor (HBR) também apresentaram uma boa redução nos valores de MSE à medida que a quantidade de dados de treino aumentou, como visto na Tabela 11.

- GBR: O Gradient Boosting começou com um MSE de aproximadamente 94,69, caindo para 10,13 quando a proporção de treino chegou a 90%. Essa diminuição consistente no erro demonstra que o GBR é capaz de aprender com mais eficiência à medida que mais dados são fornecidos, capturando melhor os padrões do dataset.
- HBR: O Histogram Boosting apresentou o maior valor de MSE inicial, com 283,53 para a proporção de 10% de treino, reduzindo consideravelmente para 5,93 com 90% de treino. Isso indica que o HBR é extremamente sensível à quantidade de dados de treino, necessitando de um volume maior para obter bons resultados.

O Adaboost Regressor (ABR), conforme mostrado na Tabela 10, teve um desempenho consideravelmente inferior em relação aos outros modelos analisados. Os valores de MSE para o Adaboost variaram entre 330,39 e 367,43, mostrando pouca consistência e não apresentando uma clara diminuição conforme a proporção de treino aumentou.

Este comportamento sugere que o Adaboost tem dificuldades para lidar com o aumento da complexidade dos dados de treino e pode estar dando muita importância a exemplos que não são representativos, o que acaba prejudicando a capacidade de generalização do modelo. Dessa forma, o ABR apresentou o pior desempenho entre os modelos analisados, sendo uma opção menos recomendada para este dataset específico.

Ambos os modelos de boosting (GBR e HBR) apresentaram uma boa melhora no desempenho, mas são mais sensíveis em proporções de treino menores, especialmente o HBR. Com mais dados, esses modelos se ajustam melhor, o que os torna boas opções quando há um volume suficiente de dados disponíveis.

Com a exposição dos dados, será feita uma análise detalhada dos gráficos apresentados nas Figuras 16 e 17, que ilustram o desempenho dos modelos de aprendizado de máquina em termos de Erro Médio Absoluto (MAE) e Erro

Quadrático Médio (MSE), respectivamente. O objetivo é entender como os diferentes modelos se comportam à medida que a proporção de dados de treinamento aumenta, variando de 10% até 90%.

A Figura 16 apresenta a performance dos modelos em relação ao Erro Médio Absoluto (MAE), que mede a média dos erros absolutos cometidos nas previsões. Abaixo, discutimos cada um dos modelos:

- Decision Tree Regressor (DTR): O comportamento do MAE para o DTR demonstra uma grande redução no erro absoluto médio à medida que a base de dados de treino aumenta. Inicialmente, o DTR apresenta um erro mais elevado, mas o gráfico indica uma queda significativa e contínua até estabilizar com uma proporção de dados maior, em torno de 70% a 90%.
- Random Forest Regressor (RFR): O RFR apresenta menores valores de MAE quando comparado ao DTR, especialmente a partir de 40% dos dados de treino. Esse comportamento indica que o modelo Random Forest possui uma robustez maior ao lidar com a variabilidade dos dados e consegue manter um erro baixo com um aumento moderado na proporção de dados.
- Gradient Boosting Regressor (GBR): O desempenho do GBR também é positivo, mostrando um comportamento semelhante ao RFR. A partir de 40% da base de dados, o MAE do GBR estabiliza em valores relativamente baixos.
- Histogram Boosting Regressor (HBR): O HBR apresenta uma tendência interessante de estabilidade, com o erro médio absoluto diminuindo de forma significativa até se estabilizar quando a base de dados atinge uma proporção maior.
- Adaboost Regressor (ABR): Entre todos os modelos, o ABR se destacou negativamente, pois apresenta a maior magnitude de MAE e é menos responsivo ao aumento da base de dados de treino.

A Figura 17 mostra o desempenho dos modelos em termos de Erro Quadrático Médio (MSE), que penaliza mais severamente erros maiores e é uma boa métrica para identificar grandes desvios nas previsões.

- Decision Tree Regressor (DTR): O MSE do DTR é inicialmente elevado, mas apresenta uma queda repentina com o aumento da base de dados, até atingir uma estabilização em torno de 30% a 40%. Essa queda rápida indica que o modelo Decision Tree é bastante sensível à quantidade de dados de treino, melhorando significativamente quando há um aumento na base de dados.
- Random Forest Regressor (RFR): O MSE do RFR começa um pouco elevado, mas reduz de forma consistente, alcançando valores bastante baixos a partir de 40% da base de dados e permanecendo estável.
- Gradient Boosting Regressor (GBR): O GBR também apresenta um comportamento muito parecido com o RFR, onde o MSE diminui progressivamente e se estabiliza em valores baixos.
- Histogram Boosting Regressor (HBR): O HBR começa com um valor de MSE muito alto, o que é uma característica negativa inicial, mas ao longo do aumento dos dados de treino, o erro diminui de maneira significativa. Quando a proporção de treino é alta, o MSE do HBR se estabiliza, sugerindo que o modelo é eficaz para grandes volumes de dados, embora seja menos eficiente nas proporções menores.
- Adaboost Regressor (ABR): Assim como no MAE, o Adaboost apresentou os maiores valores de MSE entre todos os modelos e não demonstrou uma tendência clara de diminuição conforme a base de dados aumentou.

A partir das análises feitas com base nas Figuras 16 e 17, poderá tirar algumas conclusões importantes sobre o desempenho dos modelos:

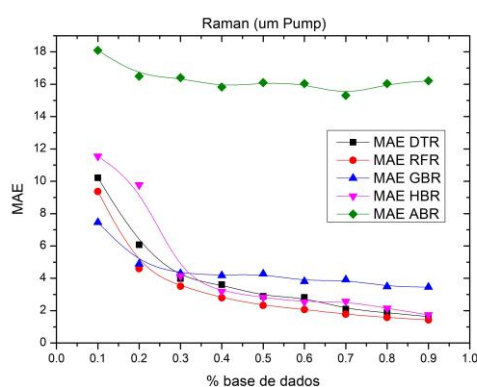
- Os modelos Random Forest (RFR) e Gradient Boosting (GBR) mostraram-se particularmente eficazes em termos de MAE e MSE, atingindo valores baixos e estáveis quando a proporção de dados de treino foi aumentada.
- O Decision Tree Regressor (DTR) apresentou uma melhora significativa à medida que a base de dados aumentou, reduzindo tanto o MAE quanto o MSE, mas ainda ficou ligeiramente atrás dos modelos RFR e GBR em termos de consistência.

- O Histogram Boosting Regressor (HBR) teve um desempenho mediano, começando com valores elevados de erro, mas mostrando uma boa redução e estabilização conforme a proporção de dados aumentava.
- O Adaboost Regressor (ABR) apresentou os piores desempenhos em ambas as métricas (MAE e MSE), mostrando uma grande dificuldade em generalizar os padrões dos dados, especialmente quando o volume de treino foi aumentado.

Portanto, as evidências sugerem que o Random Forest e o Gradient Boosting são as escolhas mais adequadas para esse problema, considerando a boa capacidade de previsão e a consistência de desempenho em diferentes proporções de dados de treino.

A análise do Erro Médio Absoluto (MAE) e do Erro Quadrático Médio (MSE) é fundamental para avaliar a capacidade dos modelos de previsão em termos de precisão e robustez. As Figuras 16 e 17 mostram o comportamento desses erros para cinco diferentes modelos conforme a proporção de dados de treino varia de 10% até 90%. Abaixo, faremos uma análise detalhada do comportamento de cada modelo ao longo deste processo.

Figura 16 – Performance dos modelos usando MAE



Fonte : próprio autor

A Figura 16 apresenta o comportamento do MAE para os modelos estudados:

- Decision Tree Regressor (DTR): O DTR apresentou uma grande redução do MAE conforme a proporção da base de dados utilizada para treino aumentou.

Nos estágios iniciais, o erro era relativamente alto, mas ele diminuiu de forma consistente até estabilizar em valores baixos a partir de cerca de 50% dos dados de treino.

- Random Forest Regressor (RFR): O RFR teve uma performance robusta ao longo de todas as proporções de treino, com um MAE menor do que o DTR, principalmente a partir de 40% dos dados de treino.
- Gradient Boosting Regressor (GBR): Assim como o RFR, o GBR apresentou um desempenho muito bom, com uma redução consistente no MAE conforme mais dados eram adicionados ao treino. A partir de 40% da base de dados, o GBR estabiliza em um valor de erro semelhante ao RFR, mostrando ser uma excelente opção para a tarefa, especialmente quando há dados suficientes disponíveis.
- Histogram Boosting Regressor (HBR): O HBR começou com um MAE mais elevado, mas à medida que a base de dados aumentava, o erro diminuiu e se estabilizou em valores baixos.
- Adaboost Regressor (ABR): Diferente dos outros modelos, o ABR apresentou um MAE muito maior e com pouca redução ao longo do aumento da base de dados de treino.

A Figura 17 mostra a análise do Erro Quadrático Médio (MSE) para os modelos. O MSE é uma métrica importante pois penaliza de maneira mais severa os erros grandes, tornando-a ideal para identificar grandes desvios nas previsões.

- Decision Tree Regressor (DTR): O MSE do DTR começou com um valor relativamente elevado, mas, com o aumento da base de dados de treino, houve uma queda significativa, até que o erro se estabilizou em torno de 30% a 40% de treino.
- Random Forest Regressor (RFR): O MSE do RFR apresentou valores iniciais altos, mas caiu rapidamente e estabilizou em valores baixos após 40% dos dados de treino.

- Gradient Boosting Regressor (GBR): O GBR demonstrou um comportamento semelhante ao RFR, com uma queda acentuada do MSE e uma estabilização em valores baixos.
- Histogram Boosting Regressor (HBR): O MSE do HBR iniciou com um valor muito elevado, mas diminuiu de maneira significativa conforme a quantidade de dados aumentou.
- Adaboost Regressor (ABR): O ABR apresentou os maiores valores de MSE entre todos os modelos, com pouca melhora ao longo do aumento dos dados de treino. O erro permaneceu alto e sem uma tendência clara de redução significativa, indicando um desempenho inferior em comparação aos demais modelos.

Com base nas análises realizadas para o Erro Médio Absoluto (MAE) e o Erro Quadrático Médio (MSE), podemos concluir que os modelos Decision Tree Regressor (DTR), Random Forest Regressor (RFR) e Gradient Boosting Regressor (GBR) apresentaram os melhores desempenhos.

- Decision Tree Regressor (DTR), Random Forest Regressor (RFR) e Gradient Boosting Regressor (GBR): Estes três modelos apresentaram curvas de MAE e MSE muito próximas, com uma tendência clara de redução dos erros conforme o volume de dados de treino aumentou, e atingiram um nível de estabilidade em valores baixos. O modelo GBR se destacou como sendo ligeiramente superior, demonstrando menor sensibilidade a erros grandes e uma estabilização mais eficiente.
- Histogram Boosting Regressor (HBR): Embora tenha começado com valores elevados de erro, o HBR apresentou uma boa capacidade de estabilização e se mostrou promissor quando há um volume maior de dados de treino. Seu comportamento foi muito semelhante ao GBR, mas teve uma resposta inicial menos eficiente.
- Adaboost Regressor (ABR): O Adaboost, por outro lado, apresentou um desempenho inferior, tanto em termos de MAE quanto de MSE. Os valores altos e a falta de uma redução significativa do erro ao longo do aumento da base de

dados indicam que o modelo teve dificuldades em se ajustar bem ao problema, tornando-o a escolha menos favorável.

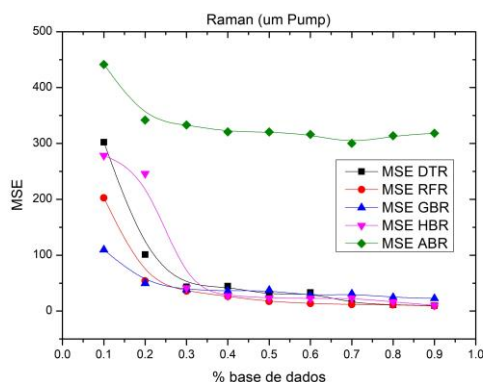
Portanto, de acordo com as métricas de desempenho analisadas, o Gradient Boosting Regressor (GBR) e o Random Forest Regressor (RFR) se mostraram as melhores opções para este problema, proporcionando uma boa combinação de precisão e estabilidade, principalmente à medida que a proporção de dados de treino aumentou. Já o Adaboost se destacou negativamente, indicando que suas limitações o tornaram menos eficiente para o contexto analisado.

Vale ressaltar que somente o Adaboost não obteve o desempenho esperado pela natureza do conjunto de dados, pois por ser dados não muito numeroso, tal efeito influenciou no dimensionamento de aplicação do algoritmo Adaboost na amplificação de primeira ordem.

O MSE HBR inicialmente possui um MSE elevado, porém diminui com o aumento da base de dados, possuindo estabilidade em valores baixos. Já o MSE ABR possui um MSE elevado, com pouca redução ao passo que aumenta a base de dados, tendendo a possuir desempenho inferior aos outros modelos.

Portanto com a análise do MAE e MSE indicam que Decision Tree Regressor (DTR), Random Forest Regressor (RFR) e Gradient Boosting Regressor têm curvas de MSE e MAE muito próximas e o modelo GBR tende a ser ligeiramente melhor em relação a performance do modelo.

Figura 17 – Performance dos modelos usando MSE



Fonte : próprio autor

4.2 Amplificação Raman de segunda ordem

O arranjo de amplificação Raman de segunda ordem com o sinal no sentido propagante teve variação da frequência do pump 1 e pump 2 no qual teve como medida de qualidade do enlace óptico também o fator de qualidade do sinal para medir a efetividade do sistema.

A análise gráfica do sistema de amplificação Raman de segunda ordem possui na distância do enlace uma variação de 10 a 110 quilômetros com passo de 20 quilômetros como ilustrado na tabela 4. Esse cenário resulta no pico de amplificação quando o laser de bombeamento número 1 possui o valor de 1550 nanômetros e o laser de bombeamento número 2 possui o valor de 1553 nanômetros como valor ótimo de fator de qualidade do sinal.

Tabela 4 – Razão do valor do fator de qualidade da amplificação Raman de segunda ordem.

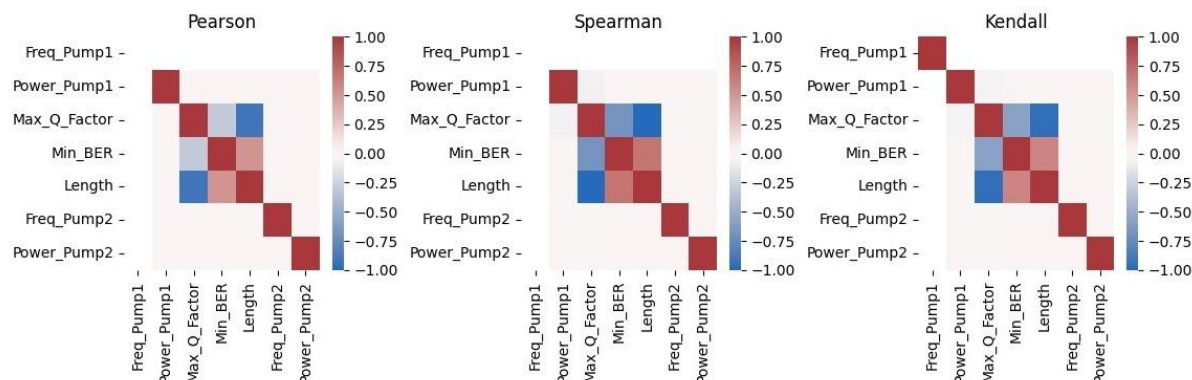
Distância (km)	10	30	50	70	90	110
Porcentagem de fator de qualidade em relação a 10 km	100 %	68,29 %	45,36 %	9,26 %	3,39 %	1,34 %

Vale ressaltar com o aumento da distância do enlace óptico, há um decréscimo significativo com relação a efetividade do fator de qualidade do sistema em questão conforme 4. Com a evolução para o cenário de 30 quilômetros, foi perdido mais da metade do fator de qualidade do sistema de amplificação Raman de segunda ordem e no cenário de maior distância um pouco mais de 1 por cento em relação a distância de 10 quilômetros.

4.2.1 Matriz de correlação

O estudo é baseado na análise da matriz de correlação das variáveis de frequência do laser de bombeamento 1 e 2, potência do laser de bombeamento 1 e 2, além do fator de qualidade representado por MaxQFactor, o tamanho do comprimento da fibra e a taxa de erro de bits representado por MinBER.

Figura 18 – Matriz de correlação para o arranjo com dois pumps de bombeamento co propagantes.



Fonte : próprio autor

Já o estudo para o arranjo de amplificação Raman de segunda ordem a relação entre o MAXQFactor e os parâmetros de entrada como mostrado na figura 18 tem o seguinte comportamento:

- Pearson: mede a correlação entre duas variáveis. Observa-se que “Length” têm correlações bem fracas em relação a “MaxQFactor”. “MinBER”. As variáveis “FreqPump2” e “PowerPump2” possuem correlação negativa bem evidente.
- Spearman: mede a correlação baseando-se a ordem entre dois valores. A correlação de Spearman possui magnitude semelhante a de Pearson e possuem relações de ranking entre as variáveis de entrada.
- Kendall: Possui magnitude semelhante a de Pearson porém evidencia a ordem de classificação dos dados, possuindo o mister de detectar outliers.

Tabela 5 – Performance do modelo cujo dataset do MaxQFactor usando o coeficiente de determinação para os algoritmos de Decision Tree, Random Forest, Gradient Boosting, Histogram Boosting e Adaboost

Porcentagem	R ² DTR	R ² RFR	R ² GBR	R ² HBR	R ² ABR
0.1	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹
0.2	1,0000	1,0000	1,0000	1,0000	9,998.10 ⁻¹
0.3	1,0000	1,0000	1,0000	1,0000	9,998.10 ⁻¹
0.4	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹
0.5	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹
0.6	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹
0.7	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹
0.8	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹
0.9	1,0000	1,0000	1,0000	1,0000	9,992.10 ⁻¹

Como pode ser observado os parâmetros de entrada e saídas possuem o mesmo comportamento indistintamente da correlação utilizada, além disso, vale a pena salientar que o Max_Q_Factor está totalmente descorrelacionado das variáveis de entradas Frequência do Pump1 (F1), Pump2 (F2) e da potência do Pump2. Este fato pode influir diretamente no treinamento dos modelos de ML.

4.2.2.1 Resultados dos algoritmos de aprendizado de máquina

Logo depois do estudo sobre a questão do pré-processamento dos dados, os dados serão processados para obter o modelo utilizando os algoritmos de aprendizado de máquina: Decision Tree Regressor (DTR), Random Forest Regressor (RFR), Gradient Bosting Regressor (GBR), Histogram Boosting Regressor (HBR) e AdaBoast Regressor (ABR).

A análise será feita conforme o estudo prévio da influência de duas features no modelo que são: MaxQFactor que representa o fator que qualidade descrito no tópico 2.1.8.2 e MinBER que representa a taxa de erro de bits na comunicação descrito no tópico 2.1.8.3. Fundamentado nesses dois tópicos, a análise será pormenorizada avaliando os modelos de aprendizado de máquina.

4.2.2.2 Dataset com uso do MaxQfactor

Ao analisar os dados utilizando a variável MaxQFactor, que indica a qualidade do sistema óptico, observamos que os modelos Decision Tree (DTR) e Random Forest (RFR) apresentaram valores de R^2 extremamente próximos de 1. Isso indica que esses modelos conseguem explicar quase toda a variação nos dados de treino, o que sugere um excelente ajuste. No entanto, essa precisão elevada pode também ser um indicativo de overfitting, pois os modelos podem estar capturando não apenas os padrões gerais dos dados, mas também detalhes e ruídos específicos, o que compromete a capacidade de generalização para novos dados.

De forma semelhante, os modelos Gradient Boosting (GBR) e Histogram Boosting (HBR) também atingiram valores de R^2 muito próximos de 1, mostrando que eles estão explicando praticamente toda a variação dos dados de treino. Isso aponta para um ótimo ajuste, embora exista o risco de que essa precisão esteja sendo alcançada por um ajuste excessivo aos detalhes específicos dos dados, em vez de se focar nos padrões mais gerais e úteis.

O modelo AdaBoost (ABR) também apresentou coeficientes de determinação (R^2) consistentemente altos, variando entre 0,9991 e 0,9998, o que demonstra uma boa capacidade de explicação da variabilidade dos dados de treino. Essa alta precisão é semelhante à dos outros modelos mencionados, sugerindo que o AdaBoost possui um excelente ajuste, embora um pouco menos preciso em relação aos demais.

Entretanto, esse comportamento dos modelos — valores de R^2 muito próximos de 1 — sugere que eles podem estar sendo treinados para capturar detalhes específicos demais dos dados. Isso é problemático quando se pensa em previsões futuras, uma vez que modelos sobreajustados não conseguem generalizar bem para dados novos. Para melhorar a robustez dos modelos e evitar o risco de overfitting, é importante adotar técnicas como a validação cruzada e ajustar hiperparâmetros para garantir um equilíbrio entre ajuste e generalização.

Em resumo, os algoritmos DTR, RFR, GBR, e HBR apresentaram um ajuste muito bom ao dataset, com uma capacidade significativa de explicar a variabilidade dos dados de treino. O AdaBoost também teve um desempenho consistente, embora com uma ligeira perda de precisão em comparação aos outros modelos. Para garantir um uso eficaz desses modelos em cenários reais, é crucial aplicar estratégias que assegurem que o modelo possa se adaptar bem a dados não vistos.

Analisando as Tabelas 14, 15 e 16, podemos observar as performances dos modelos de regressão Decision Tree (DTR), Random Forest (RFR), Gradient Boosting (GBR), Histogram Boosting (HBR) e Adaboost com relação ao erro médio absoluto (MAE).

Na Tabela 14, tanto o modelo Decision Tree Regressor (DTR) quanto o Random Forest Regressor (RFR) apresentam valores de MAE muito baixos, na ordem de 10^{-6} , o que indica um ótimo desempenho desses modelos ao estimar os valores reais. Embora haja uma pequena variação nos valores de MAE ao longo das diferentes porcentagens analisadas, o comportamento de ambos os modelos é bastante estável, sugerindo uma boa adaptação aos dados de treino. O Random Forest apresenta resultados ligeiramente melhores do que o Decision Tree, o que significa que ele consegue capturar um pouco mais de precisão no ajuste, mas ambos ainda são altamente eficientes.

De forma geral, os modelos DTR e RFR apresentaram os melhores desempenhos, com MAE muito baixos e estáveis. O Histogram Boosting Regressor também se destaca, especialmente por apresentar um desempenho competitivo, mantendo um erro médio relativamente baixo e estável, demonstrando que é uma boa opção para esse tipo de problema.

Quando observamos a Tabela 15, os modelos Gradient Boosting Regressor (GBR) e Histogram Boosting Regressor (HBR) têm valores de MAE um pouco maiores, situando-se na ordem de 10^{-3} . Esses valores indicam que, embora os modelos ainda apresentem uma boa capacidade de previsão, eles possuem um erro um pouco mais elevado do que os modelos anteriores (DTR e RFR). No entanto, os valores mantêm-se estáveis ao longo das diferentes porcentagens, mostrando que os modelos possuem uma boa capacidade de generalização e são

Tabela 14 – Performance do modelo cujo dataset do MaxQFactor usando o erro médio absoluto (MAE) para os algoritmos de Decision Tree e Random Forest.

Porcentagem	MAE DTR	MAE RFR
0.1	$1,7575040775724257 \times 10^{-6}$	$1,2345911534649713 \times 10^{-6}$
0.2	$2,575502283068501 \times 10^{-6}$	$1,4734256820493978 \times 10^{-6}$
0.3	$2,0284390025024635 \times 10^{-6}$	$1,565724206064726 \times 10^{-6}$
0.4	$1,922269857524031 \times 10^{-6}$	$1,5274213441901342 \times 10^{-6}$
0.5	$2,115678819900925 \times 10^{-6}$	$1,4156680234100844 \times 10^{-6}$
0.6	$2,004475346007703 \times 10^{-6}$	$1,5244161884377013 \times 10^{-6}$
0.7	$1,960264946909851 \times 10^{-6}$	$1,5722929711639746 \times 10^{-6}$
0.8	$2,0013377317058457 \times 10^{-6}$	$1,6685361807779847 \times 10^{-6}$
0.9	$1,9672066473556213 \times 10^{-6}$	$1,661689555712821 \times 10^{-6}$

Tabela 15 – Performance do modelo cujo dataset do MaxQFactor usando o erro médio absoluto (MAE) para os algoritmos de Gradient Boosting e Histogram Boosting.

Porcentagem	MAE GBR	MAE HBR
0.1	0,0016213642755296693	0,001521649722658171
0.2	0,0016581863016463115	0,001523233901782516
0.3	0,0016550348115849582	0,0015285142102828015
0.4	0,001655748762736986	0,0015294094234557497
0.5	0,0016645624433417956	0,0015305150152759002

0.6	0,001650403871518269	0,001530449169522812
0.7	0,001641873032422812	0,0015262982244024406
0.8	0,0016323996811016064	0,0015197825217898914
0.9	0,001600711408746486	0,0014944772147232202

consistentes ao longo das bases de dados.

Finalmente, na Tabela 16, o modelo Adaboost (ABR) apresenta valores de MAE significativamente maiores, variando de 1,0198 para 0,9941 dependendo da porcentagem analisada. Isso indica que o Adaboost possui um desempenho inferior em comparação aos outros modelos, já que o erro médio absoluto é consideravelmente mais alto. Isso pode ser um indicativo de que o modelo não está conseguindo capturar os padrões dos dados tão bem quanto os demais, resultando em um ajuste menos preciso.

O modelo AdaBoost Regressor (ABR) apresentou valores de MAE mais altos em comparação aos outros modelos, variando entre 10^{-1} e 10^0 . Esse aumento pode estar relacionado ao comportamento característico do Adaboost, que tende a se ajustar fortemente aos dados mal classificados, especialmente em situações de treino difícil. Esse ajuste mais agressivo pode resultar em um maior erro médio absoluto quando o modelo tenta aprender de forma intensiva sobre dados complexos ou ruidosos.

Além disso, os valores de MAE do Adaboost mostram uma variação significativa, o que sugere uma certa dependência do tamanho da amostra de treino. Essa dependência é especialmente notável quando a quantidade de dados de treino é menor, como indicado pelas porcentagens menores de treino.

Por outro lado, os modelos Decision Tree, Random Forest, Gradient Boosting e Histogram Boosting apresentam erros médios absolutos bem menores, indicando uma boa capacidade preditiva e uma maior estabilidade ao lidar com diferentes tamanhos de amostras. Entre eles, o Random Forest (RFR) e o Decision Tree (DTR) são os que mostram os menores valores de MAE, o que aponta para um desempenho preditivo mais preciso e consistente.

Tabela 16 – Performance do modelo cujo dataset do MaxQFactor usando o erro médio absoluto (MAE) para os algoritmos de Adaboost.

Porcentagem	MAE ABR
0.1	1,0198030998739114
0.2	0,3975518539722742
0.3	0,4027506449577905
0.4	1,0161803295563518
0.5	1,0314976525302098
0.6	1,0055400329482111
0.7	0,994727790980497
0.8	0,9987102091159754
0.9	0,9941060516721221

Tabela 17 – Performance do modelo cujo dataset do MaxQfactor usando o Erro Quadrático Médio (MSE) para os algoritmos de Decision Tree e Random Forest

Porcentagem	MSE DTR	MSE RFR
0.1	$1,0975841307585416 \times 10^{-11}$	$5,868280993212789 \times 10^{-12}$
0.2	$2,8147698277605497 \times 10^{-11}$	$8,318569685470361 \times 10^{-12}$
0.3	$1,558883606256019 \times 10^{-11}$	$9,525815642369868 \times 10^{-12}$
0.4	$1,387574874274725 \times 10^{-11}$	$9,076019812255381 \times 10^{-12}$
0.5	$1,6710573372704543 \times 10^{-11}$	$7,53252792414729 \times 10^{-12}$
0.6	$1,5132244524856985 \times 10^{-11}$	$8,218773785707737 \times 10^{-12}$
0.7	$1,4860712648522326 \times 10^{-11}$	$8,732133810411987 \times 10^{-12}$
0.8	$1,4027744320581603 \times 10^{-11}$	$9,693060724089474 \times 10^{-12}$
0.9	$1,3711648688895227 \times 10^{-11}$	$9,751656025754397 \times 10^{-12}$

As Tabelas 17, 18 e 19 mostram o Erro Quadrático Médio (MSE) dos diferentes modelos de regressão aplicados ao dataset do MaxQFactor. Estes modelos incluem Decision Tree Regressor (DTR), Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Histogram Boosting Regressor (HBR) e AdaBoost Regressor (ABR), com diferentes proporções de dados usados para treino, variando de 0,1 a 0,9. Cada tabela destaca o desempenho de cada modelo em termos de MSE, o que nos permite entender a precisão do ajuste dos modelos e

como eles se comportam ao longo das diferentes proporções de dados de treino e teste.

O Random Forest Regressor (RFR), também representado na Tabela 17, apresenta valores de MSE ainda menores, na ordem de 10^{-12} . Isso indica que o Random Forest é mais preciso do que o Decision Tree, graças à sua estrutura que combina várias árvores para reduzir o erro geral. A estabilidade desse modelo também é evidente, já que ele mantém o desempenho mesmo quando mudamos a quantidade de dados de treino.

Já o modelo Gradient Boosting Regressor (GBR), descrito na Tabela 18, apresenta um MSE consistente na ordem de 10^{-6} . Embora este valor seja maior do que o de DTR e RFR, ainda se trata de uma boa precisão e demonstra uma estabilidade aceitável em relação à variação da porcentagem de treino. O Histogram Boosting Regressor (HBR) mostra resultados similares ao GBR, com MSE também na ordem de 10^{-6} , indicando que ambos têm um desempenho comparável, mas estão um nível acima em termos de erro se comparados aos modelos anteriores.

Tabela 18 – Performance do modelo cujo dataset do MaxQfactor usando o Erro Quadrático Médio (MSE) para os algoritmos de Gradient Boosting e Histogram Boosting.

Porcentagem	MSE GBR	MSE HBR
0.1	$3,6586490452951896 \times 10^{-6}$	$3,153969659901903 \times 10^{-6}$
0.2	$3,7111663929293105 \times 10^{-6}$	$3,1552707249298446 \times 10^{-6}$
0.3	$3,6911946730238473 \times 10^{-6}$	$3,1368622645371336 \times 10^{-6}$
0.4	$3,683897489535087 \times 10^{-6}$	$3,1657551394489938 \times 10^{-6}$
0.5	$3,6975084684976187 \times 10^{-6}$	$3,1396699260135116 \times 10^{-6}$
0.6	$3,6824454939869995 \times 10^{-6}$	$3,158480776431273 \times 10^{-6}$
0.7	$3,6555962343330554 \times 10^{-6}$	$3,1444195332104516 \times 10^{-6}$
0.8	$3,6057531339424847 \times 10^{-6}$	$3,1183119367855603 \times 10^{-6}$
0.9	$3,4629392668668584 \times 10^{-6}$	$2,9794632408874695 \times 10^{-6}$

A Tabela 19 mostra os resultados do modelo AdaBoost Regressor (ABR), que apresenta valores de MSE significativamente mais altos, em torno de 10^0 . Isso indica que o AdaBoost tem um desempenho inferior em comparação aos outros modelos,

com erros mais elevados. Como resultado, o modelo pode acabar se tornando enviesado e ter dificuldades em generalizar quando aplicado a novos dados.

Em resumo, os modelos DTR e RFR apresentaram os menores valores de MSE, mostrando uma alta precisão e capacidade de generalização. Já os modelos GBR e HBR também apresentam boa precisão, mas com um nível de erro um pouco maior. O AdaBoost, por sua vez, mostrou um desempenho menos favorável devido ao seu maior valor de MSE, sugerindo a necessidade de um cuidado maior para evitar sobreajuste em certas situações.

Os valores do MAE variam de acordo com o comportamento de cada modelo. O Decision Tree Regressor (DTR) apresentou os menores valores de MAE, o que indica que ele comete menos erros ao ponderar os valores absolutos, mostrando um desempenho muito bom em termos de precisão. O Random Forest Regressor (RFR) tem um MAE ligeiramente maior do que o DTR, mas ainda é bastante competitivo. Já os modelos Gradient Boosting (GBR), Histogram Boosting (HBR) e AdaBoost (ABR) apresentaram valores de MAE na ordem de 0,002, sugerindo uma precisão razoável e consistente, embora um pouco inferior em relação ao DTR e RFR.

Ao analisarmos o MSE como métrica de desempenho, percebemos que o DTR apresenta valores mais elevados. Isso indica que, embora cometa poucos erros em termos absolutos (conforme visto pelo MAE), esses erros podem ser de grande magnitude, afetando a precisão global do modelo. Por outro lado, o RFR mostra valores de MSE significativamente menores, indicando que os erros cometidos são menores em magnitude, o que o torna um modelo mais consistente e confiável para previsões.

Os valores de MSE dos modelos GBR e HBR são baixos e estáveis, o que é um bom indicativo de que esses modelos conseguem manter um erro controlado ao longo das diferentes proporções de treino. No entanto, quando olhamos para o modelo GBR, percebe-se que ele tem os valores mais altos de MSE, o que sugere que suas previsões são menos precisas.

A análise do conjunto de dados visa ajustar os parâmetros de potência do sinal juntamente com a avaliação do respectivo modelo, usando como referência o coeficiente de determinação (R^2). O objetivo principal é desenvolver um modelo de aprendizado de máquina capaz de ajustar os parâmetros de forma eficiente,

mantendo a qualidade do enlace dentro de níveis aceitáveis, garantindo assim um bom desempenho geral do sistema.

Tabela 19 – Performance do modelo cujo dataset do MaxQfactor usando o Erro Quadrático Médio (MSE) para os algoritmos de Adaboost.

Porcentagem	MSE ABR
0.1	3,6714816341405307
0.2	0,8424400916158032
0.3	0,8859438443789145
0.4	3,6630509117361374
0.5	3,730106312375756
0.6	3,6069486115193063
0.7	3,5641182924846935
0.8	3,5997450458067934
0.9	3,5849706493237057

A análise do conjunto de dados visa ajustar os parâmetros de potência do sinal juntamente com a avaliação do respectivo modelo, usando como referência o coeficiente de determinação (R^2). O objetivo principal é desenvolver um modelo de aprendizado de máquina capaz de ajustar os parâmetros de forma eficiente, mantendo a qualidade do enlace dentro de níveis aceitáveis, garantindo assim um bom desempenho geral do sistema.

4.2.2.3 Dataset com uso do MinBER

Ao analisar os dados utilizando a variável MaxQFactor, percebemos que os modelos Decision Tree (DTR) e Random Forest (RFR) apresentaram coeficientes de determinação (R^2) extremamente próximos de 1. Isso significa que esses modelos conseguem explicar praticamente toda a variação dos dados de treino, o que é um ótimo sinal de ajuste. Porém, essa precisão tão elevada também pode ser um indicativo de que os modelos estão sobreajustados aos dados, ou seja, estão aprendendo não apenas os padrões gerais, mas também as particularidades e os ruídos, o que pode acabar prejudicando a capacidade deles de se adaptar bem a novos dados.

De forma similar, os modelos Gradient Boosting (GBR) e Histogram Boosting (HBR) também alcançaram valores de R2 próximos de 1. Isso nos mostra que eles têm uma excelente capacidade de capturar a variação dos dados durante o treino, sugerindo um ajuste muito bom. Entretanto, é importante ter em mente que esse ajuste pode ser excessivo, levando ao risco de o modelo não funcionar bem quando for aplicado em dados novos.

O modelo AdaBoost (ABR) também apresentou valores de R2 consistentemente altos, o que indica que ele tem uma boa capacidade de explicar a variabilidade nos dados de treino, embora tenha uma leve perda de precisão quando comparado aos outros modelos. Esse comportamento indica que o AdaBoost é eficaz, mas pode ser um pouco mais suscetível a erros quando se trata de generalizar para novos dados.

No geral, embora todos esses modelos tenham mostrado excelentes resultados, com valores de R2 bem próximos de 1, o que indica um ajuste quase perfeito, há também o alerta de que eles podem estar se ajustando demais aos dados de treino. Para evitar problemas de sobreajuste e garantir que os modelos sejam robustos e capazes de generalizar bem, é recomendável utilizar técnicas de validação cruzada e aplicar ajustes adequados para conseguir um equilíbrio saudável entre ajuste e generalização.

Em resumo, todos os modelos analisados — DTR, RFR, GBR, HBR, e até o AdaBoost — mostraram um ótimo desempenho em termos de ajuste aos dados, mas precisamos ter cautela com o sobreajuste. A chave para garantir bons resultados é encontrar um equilíbrio que permita que o modelo tenha um bom desempenho tanto nos dados de treino quanto em dados novos.

A análise dos valores de Erro Absoluto Médio (MAE) apresentados nas Tabelas 20, 21 e 22 mostra que, de maneira geral, todos os modelos apresentaram uma boa performance em relação ao conjunto de dados da variável MinBER. Esses valores de MAE são bastante baixos, sugerindo uma capacidade preditiva sólida dos modelos, porém, à medida que a quantidade de dados de treino aumenta, alguns modelos parecem ser mais suscetíveis ao sobreajuste.

Tabela 20 – Performance do modelo cujo dataset do MinBER usando o Erro Absoluto Médio (MAE) para os algoritmos de Decision Tree e Random Forest.

Porcentagem	MAE DTR	MAE RFR
0.1	0,0000017575040775724257	0,0000012345911534649713

0.2	0,000002575502283068501	0,0000014734256820493978
0.3	0,0000020284390025024635	0,000001565724206064726
0.4	0,000001922269857524031	0,0000015274213441901342
0.5	0,000002115678819900925	0,0000014156680234100844
0.6	0,000002004475346007703	0,0000015244161884377013
0.7	0,000001960264946909851	0,0000015722929711639746
0.8	0,0000020013377317058457	0,0000016685361807779847
0.9	0,0000019672066473556213	0,000001661689555712821

Tabela 21 – Performance do modelo cujo dataset do MinBER usando o Erro Absoluto Médio (MAE) para os algoritmos de Gradient Boosting e Histogram Boosting.

Porcentagem	MAE GBR	MAE HBR
0.1	0,0016213642755296693	0,001521649722658171
0.2	0,0016581863016463115	0,001523233901782516
0.3	0,0016550348115849582	0,0015285142102828015
0.4	0,001655748762736986	0,0015294094234557497
0.5	0,0016645624433417956	0,0015305150152759002
0.6	0,001650403871518269	0,001530449169522812
0.7	0,001641873032422812	0,0015262982244024406
0.8	0,0016323996811016064	0,0015197825217898914
0.9	0,001600711408746486	0,0014944772147232202

O modelo Decision Tree Regressor (DTR) apresentou valores de MAE estáveis, com uma leve tendência de aumento conforme a proporção de treino cresce. Isso significa que, embora o desempenho seja geralmente bom, há um pequeno impacto no erro absoluto à medida que o modelo tem mais dados para ajustar.

O Random Forest Regressor (RFR), por outro lado, teve um desempenho excelente nas primeiras porcentagens de treino, com MAE menor que o do DTR. Contudo, notamos um leve aumento no erro conforme mais dados são usados para treino, embora o erro continue em um nível muito baixo, indicando uma boa estabilidade.

Os modelos Gradient Boosting Regressor (GBR) e Histogram Boosting Regressor (HBR), conforme mostrado na Tabela 20, começam com erros um pouco mais altos do que os modelos RFR e DTR, mas mantêm valores estáveis ao longo de

todas as porcentagens de treino. O MAE do GBR fica em torno de 10⁻³, enquanto o HBR apresenta valores muito semelhantes, mostrando que ambos os modelos têm um desempenho consistente e conseguem manter uma boa capacidade de generalização.

Tabela 22 – Performance do modelo cujo dataset do MinBER usando o Erro Absoluto Médio (MAE) para os algoritmos de Adaboost.

Porcentagem	MAE ABR
0.1	1,0198030998739114
0.2	0,3975518539722742
0.3	0,4027506449577905
0.4	1,0161803295563518
0.5	1,0314976525302098
0.6	1,0055400329482111
0.7	0,994727790980497
0.8	0,9987102091159754
0.9	0,9941060516721221

O Adaboost Regressor (ABR), mostrado na Tabela 22, apresenta um comportamento interessante: ele começa com um MAE significativamente maior em comparação aos outros modelos, mas esse erro se mantém relativamente estável à medida que a porcentagem de dados de treino aumenta. Isso indica que o Adaboost tende a se ajustar de maneira mais agressiva aos dados, mas não consegue melhorar substancialmente com mais dados de treino, o que pode torná-lo mais suscetível ao sobreajuste.

Por outro lado, à medida que mais dados de treino são adicionados, os modelos mais simples, como o Decision Tree e o Adaboost, parecem ser os mais impactados, apresentando um aumento mais evidente no erro em comparação aos modelos baseados em boosting, como o GBR e o HBR. Isso sugere que os modelos mais complexos são mais capazes de se beneficiar de maiores volumes de dados, mantendo um equilíbrio entre ajuste e generalização.

Tabela 23 – Performance do modelo cujo dataset do MinBER usando o Erro Quadrático Médio (MSE) para os algoritmos de Decision Tree e Random Forest.

Porcentagem	MSE DTR	MSE RFR
0.1	0,000000000010975841307585416	0,000000000005868280993212789
0.2	0,000000000028147698277605497	0,000000000008318569685470361
0.3	0,00000000001558883606256019	0,000000000009525815642369868
0.4	0,00000000001387574874274725	0,000000000009076019812255381
0.5	0,000000000016710573372704543	0,00000000000753252792414729
0.6	0,000000000015132244524856985	0,000000000008218773785707737
0.7	0,000000000014860712648522326	0,000000000008732133810411987
0.8	0,000000000014027744320581603	0,000000000009693060724089474
0.9	0,000000000013711648688895227	0,000000000009751656025754397

O Erro Quadrático Médio (MSE), como mostrado nas Tabelas 23, 24 e 25, é uma métrica que penaliza fortemente os erros maiores, ou seja, desviações grandes dos valores reais. Se o MSE é baixo, isso indica que o modelo está tendo um bom desempenho, com pouca diferença entre os valores previstos e os reais.

O Decision Tree Regressor (DTR) começa com um MSE bem pequeno, mas, conforme o tamanho do dataset de treino cresce, o MSE também cresce. Isso sugere que o DTR é sensível ao aumento dos dados de treino, e seu erro tende a aumentar com mais informações, talvez indicando uma dificuldade em ajustar os dados adicionais.

O Random Forest Regressor (RFR) também começa com um MSE muito baixo, o que é ótimo. À medida que a porcentagem de treino aumenta, o erro também aumenta, mas o crescimento é bem controlado, o que sugere que o modelo tem um desempenho estável. Mesmo com mais dados, ele mantém uma consistência que é favorável.

O Gradient Boosting Regressor (GBR), como mostrado na Tabela 24, apresenta um MSE um pouco maior comparado ao DTR e ao RFR. No entanto, ele mantém um nível de estabilidade muito interessante conforme a base de treino cresce. O MSE não apresenta um aumento significativo, o que é um ponto positivo, mostrando que o algoritmo consegue lidar bem com mais dados.

Já o Histogram Boosting Regressor (HBR) tem um comportamento parecido com o GBR. Ele começa com um MSE um pouco alto e vai aumentando levemente conforme o tamanho do dataset de treino cresce, mas esse aumento é suave, indicando também uma boa estabilidade.

Tabela 24 – Performance do modelo cujo dataset do MinBER usando o Erro Quadrático Médio (MSE) para os algoritmos de Gradient Boosting e Histogram Boosting.

Porcentagem	MSE GBR	MSE HBR
0.1	0,0000036586490452951896	0,000003153969659901903
0.2	0,0000037111663929293105	0,0000031552707249298446
0.3	0,0000036911946730238473	0,0000031368622645371336
0.4	0,000003683897489535087	0,0000031657551394489938
0.5	0,0000036975084684976187	0,0000031396699260135116
0.6	0,0000036824454939869995	0,000003158480776431273
0.7	0,0000036555962343330554	0,0000031444195332104516
0.8	0,0000036057531339424847	0,0000031183119367855603
0.9	0,0000034629392668668584	0,0000029794632408874695

O AdaBoost Regressor (ABR), na Tabela 25, começa com um dos menores valores de MSE inicial, o que sugere um excelente desempenho para os primeiros conjuntos de dados de treino. Contudo, à medida que o dataset aumenta, o MSE também cresce de forma notável. Isso indica que o Adaboost, assim como o DTR e o RFR, tem dificuldades em manter o desempenho com a adição de muitos dados, o que pode ser sinal de que o modelo está sobreajustando ou que não está conseguindo generalizar bem.

De modo geral, a análise das tabelas mostra que os modelos Random Forest (RFR) e Gradient Boosting (GBR), que são baseados em técnicas de ensemble, tendem a ser mais estáveis em relação ao aumento dos dados de treino. O modelo HBR também mostra um comportamento semelhante ao GBR. Por outro lado, o AdaBoost (ABR) e o Decision Tree Regressor (DTR) começam com valores de MSE baixos, mas esses valores aumentam de forma significativa, sugerindo dificuldades de generalização conforme mais dados são adicionados ao treino.

Tabela 25 – Performance do modelo cujo dataset do MinBER usando o Erro Quadrático Médio (MSE) para os algoritmos de Adaboost.

Porcentagem	MSE ABR
0.1	3,6714816341405307
0.2	0,8424400916158032
0.3	0,8859438443789145
0.4	3,6630509117361374
0.5	3,730106312375756
0.6	3,6069486115193063
0.7	3,5641182924846935
0.8	3,5997450458067934
0.9	3,5849706493237057

Com base na análise realizada, podemos concluir que os melhores algoritmos para a aplicação no arranjo de amplificação são o Decision Tree Regressor (DTR) e o Random Forest Regressor (RFR). Isso ocorre porque tanto as métricas de Erro Absoluto Médio (MAE) quanto o Erro Quadrático Médio (MSE) apresentam valores muito próximos de zero em todas as porcentagens de divisão do dataset.

CONCLUSÃO

Nesse trabalho foi tomado como estudo de como a integração entre a amplificação Raman e diversos algoritmos de aprendizado de máquina. A junção dessas duas áreas tornou a obtenção de resultados mostrados tanto para a amplificação Raman de primeira ordem como de segunda ordem. Foi variados diversos parâmetros para provar o arranjo ótimo de amplificação.

A conclusão das simulações é a efetiva dependência do sinal de bombeamento e a distância do enlace como forma de manter o respectivo fator de qualidade do sistema. Porém o uso de diferentes comprimentos de onda lançou um dado para que possa abrir pesquisas em relação ao uso de diferentes comprimentos de onda e seu devido ajuste utilizando algoritmos de inteligência artificial.

A correlação feita com os dados obtidos da simulação reforça a dependência do fator de qualidade do sistema com a distância do enlace, sendo essa correlação descrita pela análise dos dados obtidos das simulações na parte de pré-processamento dos dados e seu devido tratamento em relação ao uso de algoritmos de aprendizagem de máquina.

A simulação do Raman de primeira ordem indica que Random Forest Regressor (RFR) e Gradient Boosting Regressor (GBR) têm curvas de MSE e MAE muito próximas e o modelo GBR tende a ser ligeiramente melhor. Nota-se também que com o aumento da quantidade de dados de treino, o MSE diminui consideravelmente, o que sugere que esses modelos estão generalizando melhor com mais dados.

A simulação de amplificação Raman de segunda ordem, parece que, olhando os valores de MAE, os algoritmos Random Forest (RFR) e Decision Tree (DTR) são os que se saem melhor. Eles conseguem minimizar o erro médio absoluto de forma mais consistente em comparação com os outros modelos, enquanto o AdaBoost (ABR) foi o que teve o pior desempenho.

Quando analisamos os valores de MSE, novamente o Random Forest (RFR) e o Decision Tree (DTR) são as opções mais adequadas para minimizar o erro quadrático com o dataset MaxQFactor. Os modelos de boosting, como Gradient Boosting (GBR) e Histogram Boosting (HBR), também apresentaram bons resultados, mas ainda com erros maiores do que os do RFR e DTR. Já o AdaBoost (ABR) realmente se destacou pelo lado negativo, mostrando um desempenho bem abaixo

dos demais. Isso indica que talvez ele precise de ajustes mais específicos ou, quem sabe, nem seja a melhor escolha para este problema.

No geral, todos os modelos apresentam uma performance satisfatória, com valores baixos de MAE. Os métodos Random Forest (RFR) e Gradient Boosting (GBR) parecem ter um equilíbrio melhor entre erro médio absoluto e estabilidade ao lidar com diferentes proporções de treino. O AdaBoost, por outro lado, até começou bem, com um erro inicial pequeno, mas piorou bastante conforme mais dados foram incluídos no treino. O Histogram Boosting (HBR) também foi um pouco problemático, pois se mostrou o mais sensível ao aumento da quantidade de treino, apresentando mais variações no MAE.

Já quando olhamos especificamente o MSE, o Gradient Boosting Regressor (GBR) e o Histogram Boosting Regressor (HBR) acabaram se saindo bem, mantendo erros baixos e bastante estáveis, não importando o quanto o conjunto de treino aumentava. Esses modelos parecem ser mais resistentes ao aumento da quantidade de dados e conseguem capturar padrões de forma bem eficaz. Isso os torna uma boa escolha se o objetivo for manter erros quadráticos baixos e garantir um bom ajuste. O Adaboost e Decision Tree, embora apresentem os menores erros iniciais, têm maior dificuldade em manter o desempenho com mais dados, mostrando um aumento significativo no MSE. O Random Forest apresenta um comportamento intermediário, com erros baixos e uma leve tendência de aumento, mas sem grandes flutuações.

Com os resultados obtidos, ficou definido que a amplificação Raman se tornou efetiva com a análise contínua dos dados, mesmo com a diminuição do fator de qualidade com o aumento do comprimento da fibra óptica, os algoritmos de aprendizado de máquina foram efetivos tendo como bons indicadores de validação e treinamento mesmo com a adversa base de dados.

Com perspectivas futuras ao projeto podemos ampliar o escopo de aplicação de algoritmos de aprendizado de máquina para uma massiva quantidade de dados utilizando técnicas componentes de Deep Learning que podem dar resultados promissores a diversos arranjos de amplificação desejados.

REFERÊNCIAS

- ABDI, H. **The kendall rank correlation coefficient**. *Encyclopedia of measurement and statistics*, Sage Thousand Oaks, CA, v. 2, p. 508–510, 2007.
- AGRAWAL, G. **Nonlinear Fiber Optics**. [S.I.]: Elsevier Science, 2019.
- AGRAWAL, G. P. **Nonlinear fiber optics**. [S.I.]: Springer, 2000.
- BASS, M. et al. **Handbook of optics**. [S.I.]: McGraw-Hill New York, 1995. v. 2.
- BOYD, R. W.; GAETA, A. L.; GIESE, E. **Nonlinear optics**. [S.I.]: Springer, 2008. 1097–1110 p.
- BREIMAN, L. **Random forests**. *Machine learning*, Springer, v. 45, p. 5–32, 2001.
- _____. **Classification and regression trees**. [S.I.]: Routledge, 2017.
- BROMAGE, J. **Raman amplification for fiber communications systems**. *Journal of lightwave technology*, IEEE, v. 22, n. 1, p. 79, 2004.
- CARVALHO, L. V. de et al. **O espalhamento Raman correlacionado Stokes-anti-Stokes: um estudo da influência da largura temporal do pulso de excitação**. Dissertação (Mestrado) — UFMG, 2022.
- DRESSELHAUS, M. S. **Raman spectroscopy in graphene related systems**. Wiley Online Library.
- DRUCKER, H. **Improving regressors using boosting techniques**. In: CITESEER. *lcm1*. [S.I.], 1997. v. 97, p. 107–115.
- FERRARO, J. R. **Introductory raman spectroscopy**. [S.I.]: Elsevier, 2003.
- FISHER, W. D. **On grouping for maximum homogeneity**. *Journal of the American statistical Association*, JSTOR, p. 789–798, 1958.
- FRIEDMAN, J. H. **Greedy function approximation: a gradient boosting machine**. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.
- _____. **Stochastic gradient boosting**. *Computational statistics & data analysis*, Elsevier, v. 38, n. 4, p. 367–378, 2002.
- GANGOPADHYAY, T. K.; KUMBHAKAR, P.; MANDAL, M. K. **Photonics and fiber optics: foundations and applications**. [S.I.]: CRC Press, 2019.
- GEURTS, P.; ERNST, D.; WEHENKEL, L. **Extremely randomized trees**. *Machine learning*, Springer, v. 63, p. 3–42, 2006.

GHATAK, A. K.; THYAGARAJAN, K. **An introduction to fiber optics**. [S.l.]: Cambridge university press, 1998.

GUMASTE, A.; ANTONY, T. **DWDM network designs and engineering solutions**. [S.l.]: Cisco press, 2003.

HAHN, G. J. **The coefficient of determination exposed**. Chemtech, v. 3, n. 10, p. 609–612, 1973.

HASTIE, T. et al. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer, 2009. v. 2.

IZBICKI, R.; SANTOS, T. M. dos. **Aprendizado de máquina: uma abordagem estatística**. [S.l.]: Rafael Izbicki, 2020.

KENDALL, M. G. **Rank correlation methods**. 4th ed. ed. London: Griffin, 1970. OCLC: 136868. ISBN 9780852641996.

KUNAPULI, G. **Ensemble methods for machine learning**. [S.l.]: Simon and Schuster, 2023.

KUZIN, E. A. e. a. **Stimulated raman scattering in a fiber with bending loss**. Optics communications, Elsevier, v. 169, n. 1-6, p. 87–91, 1999.

MAYER, M. et al. **Machine learning applications to land and structure valuation**. Journal of Risk and Financial Management, MDPI, v. 15, n. 5, p. 193, 2022.

PENA, A. P. **Correlação stokes-anti-stokes no espalhamento raman de líquidos transparentes**. Universidade Federal de Minas Gerais, 2018.

RAMAN, C. V.; KRISHNAN, K. S. **A new type of secondary radiation**. Nature, Nature Publishing Group UK London, v. 121, n. 3048, p. 501–502, 1928.

SMITH, R. G. **Optical power handling capacity of low loss optical fibers as determined by stimulated raman and brillouin scattering**. Applied optics, Optica Publishing Group, v. 11, n. 11, p. 2489–2494, 1972.

STOLEN, R.; IPPEN, E. **Raman gain in glass optical waveguides**. Applied Physics Letters, American Institute of Physics, v. 22, n. 6, p. 276–278, 1973.

STOLEN, R. H. **Fundamentals of raman amplification in fibers**. In: . Raman Amplifiers for Telecommunications 1: Physical Principles. New York, NY: Springer New York, 2004. p. 35–59.

WOODBURY, E.; NG, W. **Ruby laser operation in the near ir**. proc. IRE, v. 50, n. 11, p. 2347–2348, 1962.

ZAR, J. H. **Spearman rank correlation: overview**. Wiley StatsRef: Statistics Reference Online, Wiley Online Library, 2014.