

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Instituto de Ciências Exatas**  
**Programa de Pós-Graduação em Ciência da Computação**

Henrique Nicolas dos Santos Pereira

**Ceramics Predictor System (CPS): Sistema de Análise e Previsão de  
Propriedades Físicas de Materiais Cerâmicos**

Belo Horizonte  
2021

Henrique Nicolas dos Santos Pereira

**Ceramics Predictor System (CPS): Sistema de Análise e Previsão de  
Propriedades Físicas de Materiais Cerâmicos**

**Versão Final**

Dissertação apresentada ao Programa de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Minas  
Gerais, como requisito parcial à obtenção do título de Mestre  
em Ciência da Computação.

Orientador: Omar Paranaíba Vilela Neto

Belo Horizonte  
2021

2021, Henrique Nicolas dos Santos Pereira.  
Todos os direitos reservados

Pereira, Henrique Nicolas dos Santos.

P436c      Ceramics Predictor System (CPS): [recurso eletrônico] /  
sistema de Análise e previsão de propriedades físicas de  
materiais cerâmicos. Henrique Nicolas dos Santos Pereira.  
– 2021.  
1 recurso online (96 f. il, color.) : pdf.

Orientador: Omar Paranaíba Vilela Neto.

Dissertação (Mestrado) - Universidade Federal de Minas  
Gerais, Instituto de Ciências Exatas, Departamento de  
Ciências da Computação.

Referências: f. 85-89

1. Computação – Teses. 2. Aprendizado do computador –  
Teses. 3. Nanotecnologia computacional – Teses. 4. Materiais  
cerâmicos – Teses - I. Vilela Neto, Omar Paranaíba.

II. Universidade Federal de Minas Gerais, Instituto de Ciências  
Exatas, Departamento de Computação. III. Título.

CDU 519.6\*82(043)

Ficha catalográfica elaborada pela bibliotecária Irenquer Vismeg Lucas Cruz  
CRB 6/819 - Universidade Federal de Minas Gerais - ICEx



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

## FOLHA DE APROVAÇÃO

Ceramics Predictor System (CPS): Sistema de Análise e Previsão de  
Propriedades Físicas de Materiais Cerâmicos

**HENRIQUE NICOLAS DOS SANTOS PEREIRA**

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. OMAR PARANAÍBA VILELA NETO - Orientador  
Departamento de Ciência da Computação - UFMG

PROFA. GISELE LOBO PAPP  
Departamento de Ciência da Computação - UFMG

PROF. GILBERTO MEDEIROS RIBEIRO  
Departamento de Ciência da Computação - UFMG

PROF. ANDERSON DIAS  
Departamento de Química do ICEX - UFMG

Belo Horizonte, 9 de setembro de 2021.

*Eu dedico este trabalho a Deus, minha família e aqueles que amo.*

# Agradecimentos

Agradeço primeiramente a Deus, por ter me guiado pelos melhores caminhos, ter me dado forças e colocado pessoas incríveis na minha vida. Aos meus pais, Maria e Milton, que sempre me amaram, apoiaram e estiveram ao meu lado em todos os momentos. A minha irmã Danielle, por toda a amizade e companheirismo. Vocês são a minha inspiração para continuar crescendo cada vez mais. As minhas avós Maria Aparecida (in memoriam) e Maria José, meus avôs Cleres (in memoriam) e Antônio (in memoriam). A todos os meus familiares e amigos, que participaram da minha jornada e torceram pelo meu sucesso. Ao Éverton, pelo apoio no meu desenvolvimento pessoal. Sem todos vocês, esse trabalho não poderia ter sido feito.

Gostaria de expressar a minha gratidão ao meu orientador Professor Omar Paranaíba, com quem aprendi muito ao longo deste trabalho. Ele sempre esteve disponível, me auxiliou e direcionou com grande maestria. Ao João Paulo pelo apoio com os testes e treinamentos dos modelos, Pablo pelo apoio com a arquitetura de sistemas e Marina pela ajuda na análise de resultados. Aos professores Anderson, Gilberto, Gisele e Roberto, por compartilharem conhecimentos fundamentais de química, física e aprendizado de máquina.

Agradeço às empresas Algar Tech, dti Digital e Construtora Barbosa Mello, que me incentivaram na jornada acadêmica e me possibilitaram conciliar o trabalho com o mestrado. Ao laboratório Nanocomp, às instituições de apoio a pesquisa CAPES, CNPq e FAPEMIG, funcionários do Departamento de Ciência da Computação e da UFMG. Aos demais professores por todo o ensinamento dado. A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

*“Nessa estrada, não nos cabe conhecer ou ver o que virá.  
O fim dela ninguém sabe bem ao certo onde vai dar.  
Vamos todos numa linda passarela, de uma aquarela, que um dia enfim...  
Descolorirá.”*

(Toquinho - Aquarela)

# Resumo

Materiais cerâmicos são popularmente conhecidos por serem utilizados em vidros, louças, refratários e cimento. Além das aplicações mais comuns, existem uma gama de materiais cerâmicos aplicados em soluções mais tecnológicas, tais como térmicas, eletroeletrônicos, ópticas, químicas e biomédicas. O estudo e a análise a respeito dessa classe de materiais são de suma importância para avanços tecnológicos. Neste sentido, diversas pesquisas vêm sendo realizadas com o intuito de encontrar novos materiais com propriedades distintas, abrindo possibilidades inovadoras de aplicações em diferentes segmentos. Neste trabalho exploramos uma base de dados existente e focamos em duas propriedades (Constante Dielétrica e Fator de Qualidade), visando aplicações eletromagnéticas.

Com esse objetivo, primeiramente apresentamos uma ferramenta interativa de análise e exploração de materiais cerâmicos, mostrando informações sobre a estrutura e suas respectivas propriedades físicas. Em seguida, utilizando algoritmos de aprendizado de máquina, desenvolvemos um módulo para previsão das propriedades físicas de interesse para novos materiais. Os resultados obtidos demonstram boa capacidade de previsão da Constante Dielétrica e Fator de Qualidade do composto, cuja acurácia dos modelos preditivos é 86% e 75%, respectivamente. Todos os módulos aqui desenvolvidos foram integrados em uma ferramenta web interativa, permitindo aos usuários analisar e explorar de maneira simples a proposição de novos materiais cerâmicos. Acreditamos que esse trabalho é um passo fundamental para acelerar a descoberta eficiente de novos materiais cerâmicos, focados em aplicações específicas.

**Palavras-chave:** nanotecnologia computacional; aprendizado de máquina; cerâmicos, AutoML.

# Abstract

Ceramic materials are popularly known for being used in glass, crockery, refractories and cement. In addition to the most common applications, there is a range of ceramic materials given in more technological solutions, such as thermal, electro-electronic, optical, alloys and bio-medical. The study and analysis regarding this class of materials are of paramount importance for technological advances. In this regard, several complementary researches are being carried out in order to find new materials with different properties, opening up innovative possibilities for applications in different segments. In this work we explore an existing database and focus on two properties (Dielectric Constant and Quality Factor), electromagnetic layers.

For this purpose, first we present an interactive tool for analysis and exploration of ceramics, showing information on the structure and their physical properties. Then, applying machine learning algorithms, we developed a model to predict the physical properties of interest for new materials. The results show good predictability of the Dielectric Constant and Quality Factor of the compound, the accuracy of predictive models is 86 % and 75 % respectively. All modules developed here were integrated into an interactive web tool, allowing users to analyze and explore in a simple way the proposition of new ceramic materials. We believe that this work is a fundamental step to accelerate the efficient discovery of new ceramic materials, focused on specific applications.

**Keywords:** computational nanotechnology, machine learning, ceramics, AutoML.

# Lista de Figuras

2.1	Diversas aplicações de cerâmicos no cotidiano. . . . .	22
2.2	Exemplos de aplicação de cerâmicos em tecnologia. . . . .	22
2.3	Evolução da tecnologia de materiais aplicada na construção de capacitores. Atualmente o material cerâmico substitui o ar como camada isolante, aumentando a eficiência. . . . .	24
2.4	Circuito exemplificando a utilização de um capacitor. . . . .	24
2.5	Gráfico com uma onda cuja frequência é 2 Hertz, na variação de tempo de 0 a 1 segundo. . . . .	27
2.6	Exemplo de ondas com baixo e alto Fator de Qualidade. . . . .	28
2.7	A Largura a Meia Altura (FWHM) é o valor que corresponde variação que vai de $x_1$ a $x_2$ , marcados pela metade da altura da onda, representado por $\frac{f_{max}}{2}$ . . . . .	29
2.8	A rede social Instagram é uma das plataformas que utilizam a biblioteca React. . . . .	30
2.9	Gráfico da biblioteca Canvas.js com dados de exemplo, exibindo valores de vendas, expectativa e lucro ao longo do tempo. . . . .	31
2.10	O sistema desenvolvido possui a Interface Gráfica de Usuário criada utilizando React, Material-UI, canvas.js e pacotes npm auxiliares. Ela se conecta via requisição API com o Microsserviço responsável pela previsão das propriedades físicas, desenvolvido utilizando Docker, Flask e AutoGluon. . . . .	36
2.11	Exemplo de modelo com valor da métrica R2 próximo a 1. . . . .	39
4.1	Atualização de estados nos componentes do DOM do navegador e do virtual. . . . .	48
4.2	Esquematização dos componentes da interface de usuário desenvolvida. . . . .	49
4.3	Tela inicial da interface de usuário. . . . .	51
4.4	Tipos de gráficos presentes na interface de usuário. . . . .	53
4.5	Menu contendo os detalhes da lista de materiais da Frente de Pareto. . . . .	54
4.6	Tipos de filtros disponibilizados para alterar o conteúdo dos gráficos. . . . .	55
4.7	Materiais filtrados pelos elementos Al e Ba, via Lógica "E" e utilizando o Filtro por Elementos. . . . .	56
4.8	Tabela original contendo todos os materiais presentes na interface. . . . .	57
4.9	Módulo previsor que permite a criação de uma nova fórmula química e submissão para análise. . . . .	59
4.10	Fórmula química inserida na lista de elementos químicos e pronta para ser enviada ao previsor. . . . .	60
4.11	Resultados de duas fórmulas químicas enviadas para análise. . . . .	61

---

4.12	Comparação entre o material real e o material previsto pela ferramenta. . . . .	61
5.1	Exemplo de arquitetura de rede neural artificial. . . . .	65
5.2	Validação cruzada com 4 <i>folds</i> . . . . .	69
5.3	Estratégia de empilhamento multicamadas do AutoGluon, exemplificada com duas camadas de empilhamento e $n$ tipos de modelos de aprendizado básicos. .	70
5.4	Variações das métricas de performance em horas, para os modelos AutoGluon treinados com a base de dados de 22 entradas. . . . .	72
5.5	Variações das métricas de performance em horas, para os modelos AutoGluon treinados com a base de dados de 58 entradas. . . . .	74
5.6	Comparação dos valores reais e previstos para a Constante Dielétrica. . . . .	77
5.7	Comparação dos valores reais e previstos para o Fator de Qualidade. . . . .	78
A.1	Tabela Periódica com os 58 elementos químicos usados na base de dados para treinamento. . . . .	91
A.2	Distribuição dos elementos químicos. . . . .	92
A.3	Média das quantidades. . . . .	93

# Lista de Tabelas

2.1	Exemplos de valores para Constante Dielétrica em diferentes materiais. . . . .	26
4.1	Exemplos de materiais contidos na referência base para este trabalho. . . . .	44
5.1	Diferença entre bases de dados de 58 e 22 colunas considerando a fórmula $\text{Al}_2\text{O}_3$ . . . . .	63
5.2	Valores das métricas de avaliação de performance, para os modelos de Rede Neural treinados com a base de 58 entradas. . . . .	66
5.3	Resultados das métricas de erro entre os métodos estudados para a Constante Dielétrica e a base de dados de 58 entradas. . . . .	75
5.4	Avaliação por quartil dos erros relativos percentuais normalizados para a Constante Dielétrica. . . . .	80
5.5	Avaliação por quartil dos erros relativos percentuais normalizados para o Fator de Qualidade. . . . .	80
B.1	Principais outliers do modelo previsor para a Constante Dielétrica. . . . .	95
B.2	Principais outliers do modelo previsor para o Fator de Qualidade. . . . .	96

# Lista de Algoritmos

4.1 Exemplo de objeto JSON com um item da lista de materiais. . . . .	47
---	----

# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
1.1	Motivação . . . . .	16
1.2	Objetivos . . . . .	17
1.3	Resumo dos resultados . . . . .	18
1.4	Contribuições . . . . .	19
1.5	Organização da dissertação . . . . .	19
<b>2</b>	<b>Fundamentos</b>	<b>21</b>
2.1	Materiais Cerâmicos . . . . .	21
2.2	Propriedades Físicas . . . . .	23
2.2.1	Constante Dielétrica . . . . .	23
2.2.2	Fator de Qualidade . . . . .	26
2.3	Ferramentas para Interface Gráfica . . . . .	28
2.3.1	React . . . . .	29
2.3.2	Canvas.js . . . . .	30
2.3.3	Pacotes JavaScript . . . . .	31
2.4	Ferramentas para Previsão . . . . .	32
2.4.1	Aprendizado de Máquina . . . . .	32
2.4.2	AutoML . . . . .	33
2.4.3	Microserviços . . . . .	34
2.5	Métricas de Erro para Avaliação de Desempenho do Previsor . . . . .	35
2.5.1	MAE . . . . .	35
2.5.2	RMSE . . . . .	37
2.5.3	R2 . . . . .	37
2.5.4	Análise das Métricas . . . . .	38
<b>3</b>	<b>Trabalhos relacionados</b>	<b>40</b>
3.1	Algoritmos Computacionais em Ciência de Materiais . . . . .	41
<b>4</b>	<b>Tratamento e Visualização de Dados</b>	<b>44</b>
4.1	Pré-processamento e estruturação dos dados . . . . .	44
4.2	Esquematização da Interface de Usuário . . . . .	47
4.3	Interface de Usuário . . . . .	50
4.3.1	Análise Gráfica . . . . .	50

4.3.1.1	Tipos de Gráficos . . . . .	52
4.3.2	Informações Frente de Pareto . . . . .	54
4.3.3	Filtros . . . . .	55
4.3.4	Tabela de Materiais Cerâmicos . . . . .	57
4.3.5	Módulo Previsor . . . . .	57
4.3.5.1	Microserviço . . . . .	58
4.3.5.2	Utilização do previsor via interface . . . . .	59
<b>5</b>	<b>Previsão de Propriedades de Materiais Cerâmicos</b>	<b>62</b>
5.1	Estruturas de Bases de Dados para modelos preditivos . . . . .	62
5.2	Redes Neurais . . . . .	64
5.3	AutoGluon . . . . .	66
5.3.1	Utilização . . . . .	67
5.3.2	Validação Cruzada . . . . .	68
5.3.3	Agrupamento de Pilha Multicamadas . . . . .	69
5.3.4	Tipos de modelos base . . . . .	71
5.4	Resultados dos modelos treinados . . . . .	71
5.4.1	Comparação entre valores previstos e reais . . . . .	75
5.4.2	Avaliação do Erro Relativo Percentual Normalizado . . . . .	79
5.4.3	Avaliação de exemplos na base de dados . . . . .	81
<b>6</b>	<b>Conclusões e Trabalhos futuros</b>	<b>82</b>
6.1	Conclusões . . . . .	82
6.2	Trabalhos futuros . . . . .	83
	<b>Referências Bibliográficas</b>	<b>85</b>
	<b>Apêndice A Informações sobre a Base de Dados</b>	<b>90</b>
A.1	Distribuição de Elementos . . . . .	90
A.2	Análise de Composição Média . . . . .	92
	<b>Apêndice B Avaliação de pontos fora da curva (<i>outliers</i>)</b>	<b>94</b>
B.1	<i>Outliers</i> Constante Dielétrica . . . . .	94
B.2	<i>Outliers</i> Fator de Qualidade . . . . .	95

# Capítulo 1

## Introdução

A evolução da ciência se baseou por muitos anos em observações de fenômenos da natureza. Ao longo do tempo, surgiram modelos teóricos e generalizações que deram origem a diversas equações matemáticas para descrição dos fenômenos naturais. Os modelos teóricos evoluíram e se tornaram cada vez mais complexos, dificultando a concepção de novas leis mediante soluções analíticas. Com o avanço de técnicas computacionais e a capacidade de processamento dos computadores, estes modelos teóricos foram levados a um novo patamar, permitindo o desenvolvimento de simuladores capazes de generalizar as análises e auxiliar na descoberta de novas soluções [1]. Mas recentemente, os cientistas passaram a usar os computadores não só para avançar nos estudos teóricos, mas também para tratar, interpretar e descobrir novas informações a partir de dados de experimentos reais. Dentro deste contexto, os modelos de Aprendizado de Máquina têm um papel fundamental, aprofundando a análise dos dados gerados em diversos experimentos e propondo soluções inovadoras [2].

Esse paradigma, juntamente com outras técnicas computacionais, vem sendo aplicados com sucesso na área de Ciência de Materiais. A partir dessa junção, originou-se o campo de estudos conhecido como Informática de Materiais [3]. Essa área de estudos consiste na aplicação dos princípios de Informática para contribuir na seleção, compreensão, desenvolvimento e descoberta de novos materiais.

O processo de criação de novos materiais exige a identificação de elementos químicos que atendam a uma determinada aplicação ou visem otimizar propriedades desejadas. A tradicional forma de criação ocorre por meio da determinação das propriedades relevantes, considerando muitos potenciais candidatos, via experimentos ou cálculos. É realizada uma escolha dos melhores casos para estudos posteriores e otimização.

A utilização de técnicas computacionais está facilitando esse processo. Algoritmos computacionais já podem calcular e realizar previsões de propriedades físicas a partir de um conjunto de entradas pré-estabelecidas. Isso permite que especialistas utilizem ferramentas informatizadas para extrair mais informações, com base em análise de dados e previsão de propriedades, norteadando o processo para descoberta de novos compostos.

## 1.1 Motivação

Nas últimas décadas, diversos avanços científicos e tecnológicos estão sendo incorporados no cotidiano de milhões de pessoas. *Smartphones*, *tablets*, câmeras fotográficas, *smartwatches*, entre outros, já fazem parte da normalidade de muitos. Essas invenções somente se tornaram possíveis através do estudo e descoberta de materiais, como novas ligas metálicas, novos tipos de polímeros ou mesmo cerâmicas. Esse trabalho foca, em especial, no estudo, análise e previsão de propriedades físicas de materiais cerâmicos, visando facilitar e acelerar a criação de novos materiais dessa categoria, com propriedades aperfeiçoadas.

Os cerâmicos são uma classe de materiais associados às ligações químicas mistas (covalentes, iônicos e, às vezes, metálicos) [4]. Basicamente, são os materiais inorgânicos que não são metais ou polímeros. Em geral, são associados a materiais como o cimento, vidros, porcelanas, entre outros. Contudo, existem subclasses desses materiais com aplicações mecânicas, eletroeletrônicas e ópticas [5].

Este trabalho foca nessa classe de cerâmicas com propriedades eletroeletrônicas. Essas cerâmicas avançadas são importantes para a construção de materiais tecnológicos tais como sensores, antenas, capacitores e resistores. Há também aplicações para o desenvolvimento de dispositivos eletrônicos flexíveis [6], que representam novas formas de tecnologias elásticas e deformáveis em formatos curvilíneos como, por exemplo, telas de celulares dobráveis.

No trabalho apresentado nesta dissertação há um foco em duas propriedades físicas dos materiais cerâmicos: Constante Dielétrica e o Fator de Qualidade multiplicado pela Frequência. Para explicar a importância de cada uma, é inicialmente definida a Frequência  $f$ , uma grandeza física que indica o número de oscilações de uma onda, em um determinado tempo, medida em Hertz (Hz). O Fator de Qualidade  $Q$  é um parâmetro adimensional, indicando as perdas de energia no interior de um elemento ressonante [7]. Essas duas propriedades, multiplicadas, dão origem ao  $(Qf)$ , também medido em Hertz. Essa é uma das propriedades alvo de estudo, que para simplificação de nomenclatura será chamada apenas de Fator de Qualidade ou Fator Q. A outra propriedade alvo é a Constante Dielétrica ( $\epsilon_r$ ), medida adimensional de permissividade elétrica, ou seja, a capacidade que um material possui de se polarizar em resposta a um campo elétrico aplicado.

Materiais com alto Fator de Qualidade podem manter a ressonância de uma onda eletromagnética por mais tempo. Materiais com alta Constante Dielétrica são mais permissivos eletricamente, portanto essa propriedade representa o quão propenso é o material para reter cargas elétricas. Avaliando e prevendo ambas as propriedades, é possível determinar a aplicação de novos materiais, como, por exemplo, na utilização de antenas de longo alcance, utilizadas em dispositivos de Internet das Coisas, ou mesmo a criação de

um capacitor que armazena uma quantidade maior de cargas elétricas, ocupando o mesmo espaço físico.

Para atingir o objetivo de apoiar no estudo, análise e criação de novos materiais, é necessário realizar a previsão dos valores dessas duas propriedades. Para isso, foram aplicadas técnicas computacionais de Aprendizado de Máquina, que basicamente são técnicas que permitem um computador aprender um padrão de forma autônoma. Graças ao avanço do poder de processamento dos computadores, aliado à grande disponibilidade de dados, a área de Aprendizado de Máquina foi impulsionada nos últimos anos. Suas aplicações se tornaram populares e estão presentes em diversas áreas de conhecimento. As aplicações são as mais diversas, passando pela detecção de spam em e-mails, detecção de fraudes financeiras, processamento de linguagem, reconhecimento de objetos e biometria.

Praticamente todas as áreas do conhecimento estão tirando proveito das diversas técnicas computacionais para impulsionar suas recentes descobertas. No campo da Ciência de Materiais, diversos estudos e trabalhos têm utilizado das ferramentas de Aprendizado de Máquina para analisar, calcular e prever não só propriedades físicas de materiais, como também a sugestão de criação de novos compostos em si [8].

Especialistas das áreas de química, física e engenharia de materiais podem usufruir do sistema para realizar diversos tipos de trabalhos. É possível utilizar o previsor para direcionar os estudos de criação de novos materiais, sem necessariamente precisar sintetizá-los e medir suas propriedades. É possível extrair informações relevantes que auxiliarão no desenvolvimento e criação de cerâmicas avançadas utilizadas nos mais diversos tipos de aplicações tecnológicas.

## 1.2 Objetivos

Considerando a motivação dos avanços científicos e tecnológicos, que ocorrem a partir da descoberta de novos materiais, este trabalho tem como objetivo geral a criação de um sistema capaz de auxiliar os pesquisadores a analisar, interpretar e criar novos materiais cerâmicos. O sistema foi desenvolvido a partir estudo, análise e processamento de uma base de dados com materiais cerâmicos de baixa perda dielétrica [9]. O foco neste subgrupo ocorre devido às aplicações tecnológicas em eletroeletrônicos, porém toda a metodologia para concepção do sistema foi desenvolvida para abranger diversos tipos de materiais cerâmicos, podendo ser facilmente estendida.

Outro objetivo é facilitar o acesso e a usabilidade, provendo uma experiência Web que não necessita da realização de configurações na máquina local do usuário. A ferramenta tem como premissa ser totalmente interativa, para comparar materiais, realizar

filtros, visualizar gráficos e tabelas, além de trazer informações relevantes para os especialistas de forma prática. Por seguir um conceito de arquitetura de microsserviços, o sistema é de fácil expansão e novos módulos de previsão ou análise podem ser incorporados.

Além disso, o sistema proposto tem como finalidade apoiar na previsão das propriedades físicas do Fator de Qualidade e da Constante Dielétrica de novos materiais. Dessa forma, não há necessidade de sintetizá-los primeiro, no processo empírico de descoberta de materiais, para então aferir as propriedades. O processo tradicional é lento e financeiramente caro. O conceito para a previsão das duas propriedades físicas também foi pensado para facilitar a utilização da ferramenta. A simplicidade em escolher apenas os elementos e suas respectivas quantidades, traz uma experiência de fácil utilização para qualquer pessoa.

### 1.3 Resumo dos resultados

Este trabalho apresenta, como resultado, um sistema Web que permite analisar mais de quatro mil materiais cerâmicos [9], sendo facilmente expansível para incorporar novos itens nessa extensa lista. O sistema conta com 3 gráficos interativos distintos, para auxiliar na análise e visualização dos dados. Além disso, a ferramenta conta com dois tipos de filtros de seleção: um que permite a seleção dos elementos químicos desejados; outro que considera as combinações entre elementos e suas respectivas quantidades, em cada uma das fórmulas químicas utilizadas como base. Cada um dos filtros, conta com duas lógicas de seleção, que concede ainda mais liberdade de escolha para o usuário final.

Todos os materiais presentes no banco de dados do sistema também são apresentados em um formato de tabela aos usuários. Essa funcionalidade facilita a visualização dos dados, pois todas as informações e detalhes de cada material são exibidos em uma única tela. Além disso, a tabela geral facilita na busca por itens específicos, elementos químicos e valores desejados de propriedades físicas.

O sistema foi batizado de CPS (*Ceramics Predictor System*), ou Sistema Previsor de Cerâmicas, pois uma de suas principais funcionalidades está no módulo de previsão de propriedades físicas para novos cerâmicos. O previsor é simples de usar, basta selecionar os elementos químicos desejados e suas respectivas quantidades. Com base nessas informações, são calculadas as duas propriedades físicas de interesse. Cada novo composto criado é exibido em uma tabela para consulta, assim como em um gráfico para ser feita a comparação com os demais cerâmicos.

Os dois modelos previsores utilizados no sistema apresentam boa taxa de acerto. A acurácia é de 85% pelo modelo da Constante Dielétrica. Já para o modelo do Fator

de Qualidade essa taxa corresponde a 75%. O Fator de Qualidade possui valores em escalas maiores e mais variadas. Mesmo com um valor percentual um pouco menor para a segunda propriedade, ainda é possível obter boa previsibilidade, apenas escolhendo os elementos químicos e suas respectivas quantidades.

## 1.4 Contribuições

Com base nos resultados apresentados acima, o trabalho desenvolvido nesta dissertação traz um novo sistema, capaz de contribuir com a comunidade científica. O CPS permite a análise de materiais cerâmicos e suas propriedades físicas, por meio de gráficos interativos, filtros, telas informativas e a base de dados completa, apresentada em tempo real, no formato de tabela. É um sistema totalmente preparado para operar no contexto Web, não necessitando da instalação de programas extras em um computador cliente. Além disso, pode ser disponibilizado para ser acessado de qualquer lugar do mundo e por qualquer usuário interessado em estudar sobre materiais cerâmicos.

Uma das principais contribuições está na criação de dois novos modelos de Aprendizado de Máquina com alta acurácia. Até o momento, na literatura não haviam sido criados modelos para previsão das propriedades de Constante Dielétrica e Fator de Qualidade de materiais cerâmicos. Outra contribuição significativa está na junção de modelos de Aprendizado de Máquina, com Interfaces de Usuário totalmente Web e interativa. É possível utilizar esses modelos no sistema, realizando previsões em tempo real. Tanto especialistas, quanto usuários comuns com pouca experiência, podem criar uma nova fórmula química, obtendo em tempo real o valor das duas propriedades físicas e comparando o novo material com os demais. O formato de utilização do módulo de previsão é muito simples, bastando apenas selecionar os elementos químicos e quantidades desejadas para cada um.

## 1.5 Organização da dissertação

O restante deste trabalho está organizado da seguinte forma.

- **Capítulo 2:** lista os principais conceitos relacionados ao tema do trabalho, como o que são materiais cerâmicos, o que é aprendizado de máquina, o que são as pro-

priedades físicas aqui estudadas e uma visão geral da ferramenta proposta.

- **Capítulo 3:** apresenta os trabalhos cujo contexto consiste no uso de técnicas computacionais em ciência de materiais. Aplicações de aprendizado utilizadas nesse contexto também são descritas, assim como as principais diferenças que distinguem esta dissertação de outras pesquisas.
- **Capítulo 4:** apresenta o tratamento dos dados para utilização neste trabalho. Toda a Interface de Usuário dos sistema desenvolvido e suas funcionalidades também são apresentadas.
- **Capítulo 5:** descreve a modelagem dos modelos previsores para a Constante Dielétrica e o Fator de Qualidade, assim como as técnicas exploradas e os resultados obtidos para os modelos finais adotados.
- **Capítulo 6:** realiza o fechamento desta dissertação, destacando nossas descobertas, contribuições e apontando os rumos de pesquisas futuras.

# Capítulo 2

## Fundamentos

### 2.1 Materiais Cerâmicos

Os cerâmicos são uma classe de materiais não metálicos, não poliméricos, compostos por elementos químicos metálicos e não metálicos, unificados por ligações covalentes, iônicas ou mistas. Essa classe de materiais possui em sua composição elementos metálicos como, por exemplo, magnésio, alumínio e ferro; e não-metálicos, como o oxigênio e silício [10]. Costumam ser resistentes a altas temperaturas, isolantes elétricos, de alta dureza e também frágeis. São comumente representados por vidros, telhas, tijolos, cimento e porcelanas. Por exemplo, a argila pura é um tipo de material cerâmico e possui a fórmula química  $2\text{SiO}_2\text{Al}_2\text{O}_3\text{2H}_2\text{O}$ .

Eles consistem em matrizes de átomos interconectados, não há moléculas discretas. Esta característica distingue a cerâmica de sólidos moleculares, como cristais de iodo (compostos de discretos) e cera de parafina (composta por moléculas de alcano de cadeia longa) [10]. Também exclui o gelo, composto por moléculas discretas de  $\text{H}_2\text{O}$  e frequentemente se comporta como muitas cerâmicas. A maioria das cerâmicas são compostos de metais ou metaloides e não metais. Também, frequentemente são óxidos, nitretos e carbonetos. No entanto, o diamante e a grafite também são classificados como cerâmicas. Essas formas de carbono são inorgânicas no sentido mais básico do termo: não foram preparadas a partir do organismo vivo. Basicamente, cerâmicas são a maioria dos materiais sólidos que não são de metal, plástico, derivados de plantas ou animais.

Estes materiais possuem aplicações em diversas áreas, como, por exemplo, na construção civil, setor automotivo, indústria eletrônica e aeroespacial. Além das aplicações popularmente conhecidas, cerâmicos também são utilizados na construção de resistores, capacitores, próteses biomédicas, combustíveis, antenas, entre outros [11]. Por terem diversas aplicações tecnológicas, o estudo das propriedades físicas e criação de novos materiais é um ramo de grande interesse científico.

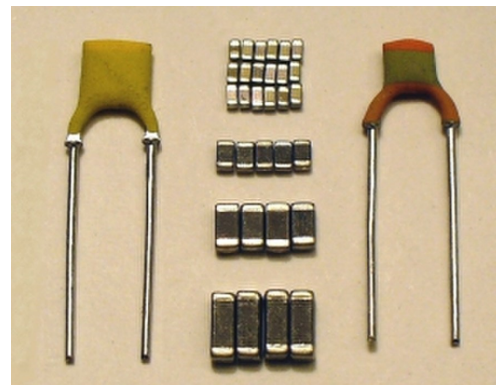
A Figura 2.1 apresenta uma série de exemplos distintos de cerâmicas, demonstrando a importância e aplicabilidade desses materiais no dia a dia. A seguir, temos dois exemplos



Figura 2.1: Diversas aplicações de cerâmicos no cotidiano.



(a) Antena GPS.



(b) Diversos modelos de capacitores.

Figura 2.2: Exemplos de aplicação de cerâmicos em tecnologia.

de cerâmicas avançadas, com aplicações voltadas para o ramo de tecnologia. Na Figura 2.2a apresenta uma antena de navegação GPS. A Figura 2.2b mostra alguns exemplares de capacitores, muito utilizados em circuitos eletrônicos.

## 2.2 Propriedades Físicas

Matéria é tudo aquilo que possui massa e ocupa um lugar no espaço. Cada material possui um conjunto de características específicas, ou seja, suas propriedades [12]. A propriedade física é qualquer propriedade mensurável, cujo valor descreve o estado de um sistema físico. Neste trabalho, nós focaremos no estudo de duas propriedades de materiais cerâmicos que são a Constante Dielétrica e o Fator de Qualidade.

Os materiais utilizados como base para este trabalho são considerados dielétricos de baixa perda, aplicados em telecomunicações sem fio. Os avanços recentes na Internet das Coisas (IoT), Internet Tátil (sistemas sem fio de quinta geração), Internet Industrial, radiodifusão por satélite e sistemas de transporte inteligente (ITS), impulsionaram a demanda por componentes que atendam essas necessidades [9]. As propriedades Constante Dielétrica e Fator de Qualidade, juntas, podem determinar os materiais que são bons candidatos para esse tipo de aplicação tecnológica. Ao maximizarmos ambas propriedades, há uma tendência de encontrarmos melhores antenas, ou seja, componentes menores, com longo alcance e capazes de atender maior faixa de frequências eletromagnéticas. Mais detalhes das duas propriedades serão apresentadas a seguir.

### 2.2.1 Constante Dielétrica

Em eletromagnetismo, a Permissividade é uma propriedade física que descreve como um campo elétrico é afetado ou afeta um meio. A Permissividade Relativa, também conhecida como Constante Dielétrica [13], é a habilidade que um material possui em se polarizar como resposta a um campo elétrico aplicado em seus polos e, consequentemente, cancelar parcialmente o campo no interior do material. Quanto maior a Constante Dielétrica, maior a quantidade de cargas elétricas que podem ser armazenadas.

Esse conceito é amplamente utilizado na produção de capacitores [14], como pode ser exemplificado pelas Figuras 2.3a e 2.3b. Um capacitor com alta permissividade, permite que uma mesma quantidade de carga elétrica seja guardada com um campo elétrico de menor intensidade. O material dielétrico fornece o isolamento entre as placas do capacitor e, consequentemente, eleva a capacitância do mesmo.

A Figura 2.4 apresenta a visão simplificada do circuito de um capacitor. As placas condutivas envolvem o material dielétrico, responsável pelo armazenamento de cargas, assim que uma tensão elétrica é aplicada nas extremidades das placas que o envolvem.

A seguir, as definições importantes para o entendimento dos conceitos necessários



(a) Modelo de capacitor antigo com placas separadas pelo ar como meio dielétrico.



(b) Exemplo de Capacitor Eletrolítico moderno.

Figura 2.3: Evolução da tecnologia de materiais aplicada na construção de capacitores. Atualmente o material cerâmico substitui o ar como camada isolante, aumentando a eficiência.

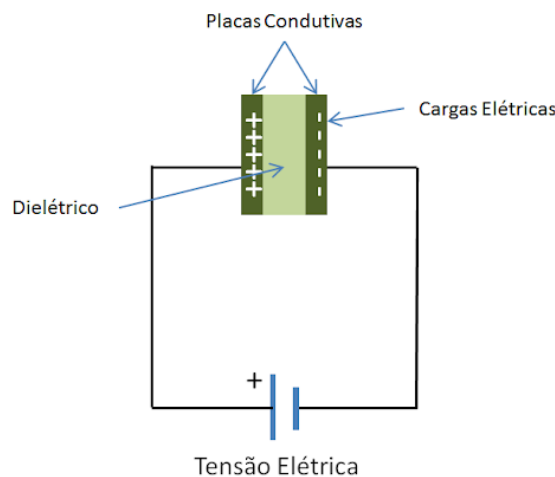


Figura 2.4: Circuito exemplificando a utilização de um capacitor.

serão apresentadas:

- **Densidade de Fluxo Elétrico:** grandeza vetorial que aponta na direção das linhas de fluxo elétrico de um campo. Medida em Coulomb por metro quadrado ( $C/m^2$ ).
- **Intensidade do Campo Elétrico:** é o campo de força provocado pela ação de cargas elétricas. Cargas elétricas colocadas num campo elétrico estão sujeitas à ação de forças elétricas, de atração e repulsão. Medida em Newton por Coulomb ( $N/C$ ).
- **Permissividade:** dada pela razão entre Densidade de Fluxo Elétrico e a Intensidade do Campo Elétrico, simbolizada por  $\epsilon$ . Determina a influência que o meio exerce na criação de um campo elétrico. Medida em Farads por Metro ( $F/m$ ).

- **Permissividade do vácuo:** normalmente simbolizada por  $\epsilon_0$ , é o quociente dos campos  $D/E$  nesse meio. A permissividade do espaço livre é igual a aproximadamente  $8.85 \times 10^{-12}$  Farads por metro (F/m).
- **Constante Dielétrica:** também conhecida como Permissividade Relativa do Material, é definida como a permissividade de um determinado material em relação à permissividade no vácuo. Normalmente é simbolizada por  $\epsilon_r$ . Não possui dimensão e depende de fatores como a temperatura e pureza do material.

Considerando o fato que a permissividade  $\epsilon$  de um meio governa a carga que pode ser armazenada, temos a seguinte Equação 2.1 para determiná-la:

$$\epsilon = \frac{D}{E}, \quad (2.1)$$

em que:

$\epsilon$  = permissividade em Farads por metro;

$D$  = densidade de fluxo elétrico;

$E$  = intensidade do campo elétrico;

Pode ser visto a partir das definições de permissividade que as constantes estão relacionadas de acordo com a seguinte equação 2.2:

$$\epsilon_r = \frac{\epsilon}{\epsilon_0}, \quad (2.2)$$

onde:

$\epsilon_r$  = permissividade relativa;

$\epsilon$  = permissividade da substância em Farads por metro;

$\epsilon_0$  = permissividade do vácuo em Farads por metro.

### Exemplos de áreas de importância:

A Constante Dielétrica e a Permissividade afetam diversos aspectos da tecnologia elétrica e eletrônica. A Permissividade Relativa não é apenas importante quando se trata de capacitores, também existem aplicações em outras áreas da tecnologia:

- **Linhas de transmissão de radiofrequência:** utilizado no interior do cabo coaxial, o dielétrico é colocado entre a capa exterior e o fio condutor. Tem um efeito importante em termos de impedância e influência na velocidade de transmissão.

Tabela 2.1: Exemplos de valores para Constante Dielétrica em diferentes materiais.

Material	Constante Dielétrica
Ar	$\sim 1.0005$
Polietileno	2.25
Óxido de alumínio ( $\text{Al}_2\text{O}_3$ )	9.6
Silício	11.68
Água	55 - 88

- **Propagação de rádio:** mudanças na permissividade relativa da atmosfera podem ter um grande impacto na transmissão de sinais de rádio. Mesmo pequenas mudanças na permissividade relativa podem fazer com que antenas captem sinais de distâncias maiores.

Os casos acima apresentados, entre outros, mostram que a permissividade relativa de um meio pode ter um impacto em muitos aspectos da eletrônica, rádio e outros elementos da tecnologia e da ciência. A Tabela 2.1 exibe valores exemplos de Constante Dielétrica, para alguns materiais cerâmicos e não cerâmicos [15].

### 2.2.2 Fator de Qualidade

Em física e engenharia, o Fator de qualidade, ou Fator Q, é um parâmetro adimensional que descreve o quão amortecido é um oscilador ou ressonador. Representa a energia elétrica armazenada no circuito dividida pela energia dissipada em um determinado período [16]. O amortecimento é uma influência dentro ou sobre um sistema oscilatório com o efeito de reduzir, restringir ou prevenir suas oscilações.

É definido como a razão entre a energia inicial armazenada no ressonador e a energia perdida em um radiano do ciclo de oscilação. O Fator Q é alternativamente definido como a razão entre a frequência central de um ressonador e sua largura de banda quando sujeito a uma força motriz oscilante. Essas duas definições fornecem resultados numericamente semelhantes, mas não idênticos. Fator Q mais alto indica uma taxa mais baixa de perda de energia e as oscilações morrem mais lentamente.

Algumas definições importantes são:

- **Frequência:** é uma grandeza física que indica o número de ocorrências de um evento (ciclos, voltas, oscilações), em um determinado intervalo de tempo medida em Hertz;

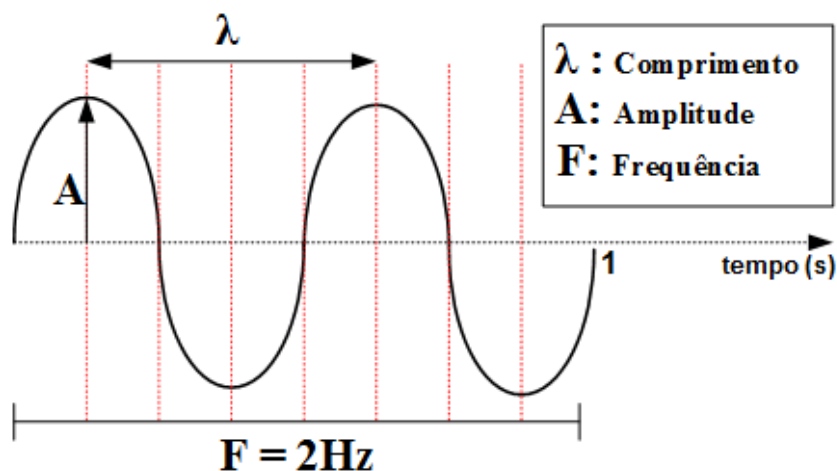


Figura 2.5: Gráfico com uma onda cuja frequência é 2 Hertz, na variação de tempo de 0 a 1 segundo.

- **Comprimento de Onda:** é a distância entre valores repetidos sucessivos num padrão de onda;
- **Amplitude:** corresponde à altura da onda, marcada pela distância entre o ponto de equilíbrio (repouso) até a crista.

Na Figura 2.5<sup>1</sup> é mostrado um exemplo de onda de 2 Hz, com a apresentação das propriedades físicas de Comprimento de Onda, Amplitude e Frequência.

Um exemplo ilustrativo é o de um pêndulo suspenso e oscilando no ar, que possui um Fator Q alto em comparação com um pêndulo imerso em óleo, cujo fator de qualidade baixo demonstra que o pêndulo para de oscilar mais rápido. Ressonadores com Fatores de Qualidade alta têm baixo amortecimento, de modo que tocam ou vibram por mais tempo [17].

A Figura 2.6<sup>2</sup> exemplifica duas ondas, uma com amortecimento rápido e outra com amortecimento lento. Já a imagem mostra a onda de cima com um Fator de Qualidade baixo, o que significa que a oscilação desaparece rapidamente com o passar dos ciclos. A onda de baixo possui Fator de Qualidade alto, representando uma oscilação presente por mais tempo e mais ciclos.

A equação que descreve o Fator de Qualidade é dada por 2.3:

$$Q = \frac{f_r}{\Delta f}, \quad (2.3)$$

em que:

<sup>1</sup>Fonte: [www.fisicanacuca3.blogspot.com/2013/](http://www.fisicanacuca3.blogspot.com/2013/)

<sup>2</sup>Fonte: [www.radiologycafe.com/radiology-trainees/frcr-physics-notes/producing-an-ultrasound-beam](http://www.radiologycafe.com/radiology-trainees/frcr-physics-notes/producing-an-ultrasound-beam)

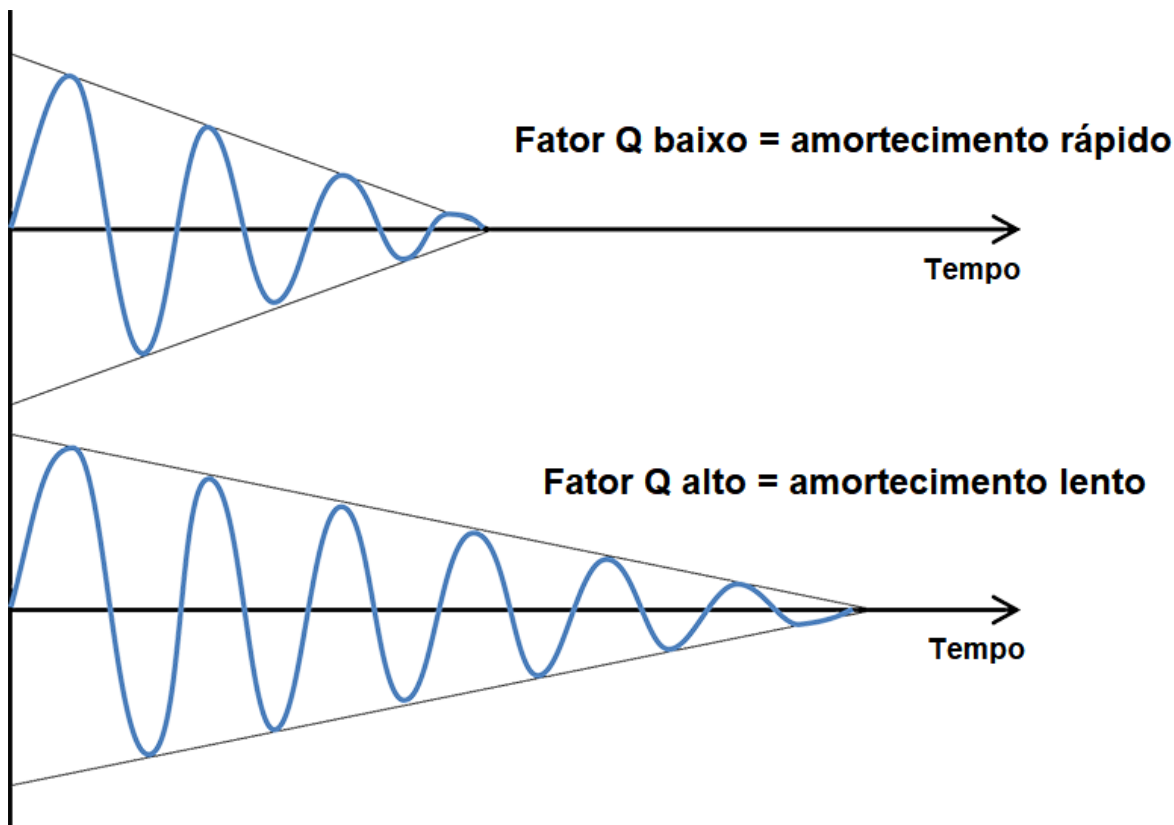


Figura 2.6: Exemplo de ondas com baixo e alto Fator de Qualidade.

$Q$  = Fator de Qualidade;

$f_r$  = Frequência de Ressonância;

$\Delta f$  = Largura da Ressonância. Também referido como Largura a Meia Altura (FWHM), do inglês *Full Width at Half Maximum*.

A Figura 2.7 [18] mostra o conceito da Largura a Meia Altura (FWHM) de uma onda.

## 2.3 Ferramentas para Interface Gráfica

A interface gráfica representa a parte de um sistema que permite a interação do usuário com a ferramenta em si. Neste trabalho apresentamos uma interface construída utilizando a biblioteca React, que possui como base a linguagem de programação JavaScript. A linguagem de programação é um conjunto de regras sintáticas e semânticas que permitem a implementação de um código-fonte de softwares. JavaScript é uma linguagem de alto nível e é a mais presente em navegadores de Internet.

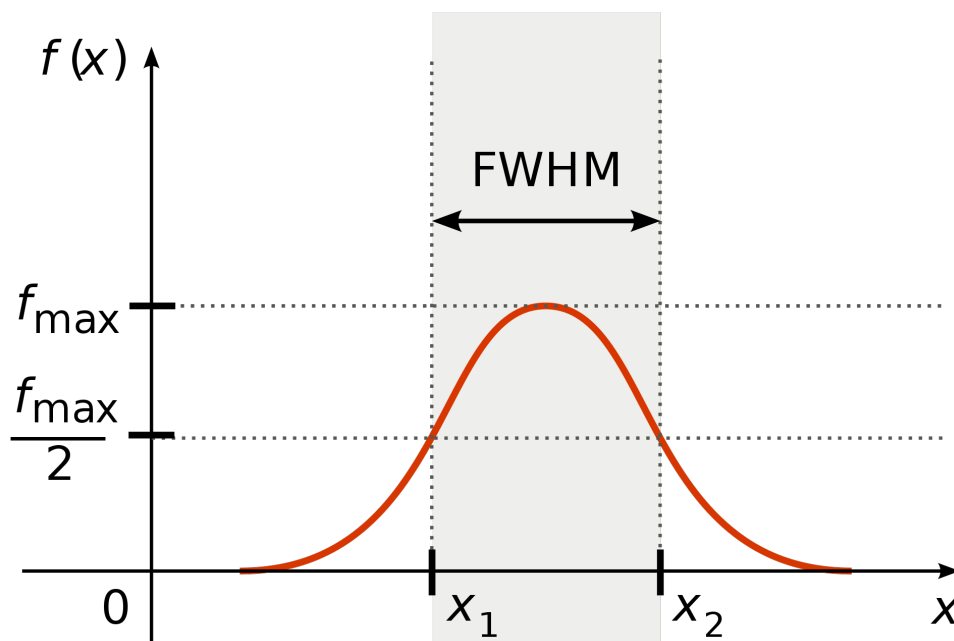


Figura 2.7: A Largura a Meia Altura (FWHM) é o valor que corresponde variação que vai de  $x_1$  a  $x_2$ , marcados pela metade da altura da onda, representado por  $\frac{f_{\max}}{2}$ .

A interface de usuário possui três subdivisões:

1. **Análise Gráfica:** área contendo os gráficos de referência para análise dos materiais. É possível filtrar conforme o conjunto de elementos químicos escolhidos, visualizar a Frente de Pareto dos cerâmicos, obter a Distância Euclidiana de um ponto do gráfico até a Frente de Pareto e também o gráfico com a normalização logarítmica;
2. **Módulo Previsor:** parte da interface que permite escolher um conjunto de elementos químicos para compor a fórmula de um material cerâmico e enviá-la para análise do previsor em si. A resposta obtida é o valor de Constante Dielétrica e Fator de Qualidade exibido em uma tabela e também como um ponto em um gráfico de Frente de Pareto para facilitar a análise do possível novo material;
3. **Tabela Original:** seção que apresenta a tabela com os mais de quatro mil materiais da referência original [5].

### 2.3.1 React

Segundo seus criadores, o React é uma biblioteca JavaScript declarativa, eficiente e flexível para a criação de interfaces de usuário [19]. Uma biblioteca, no contexto de

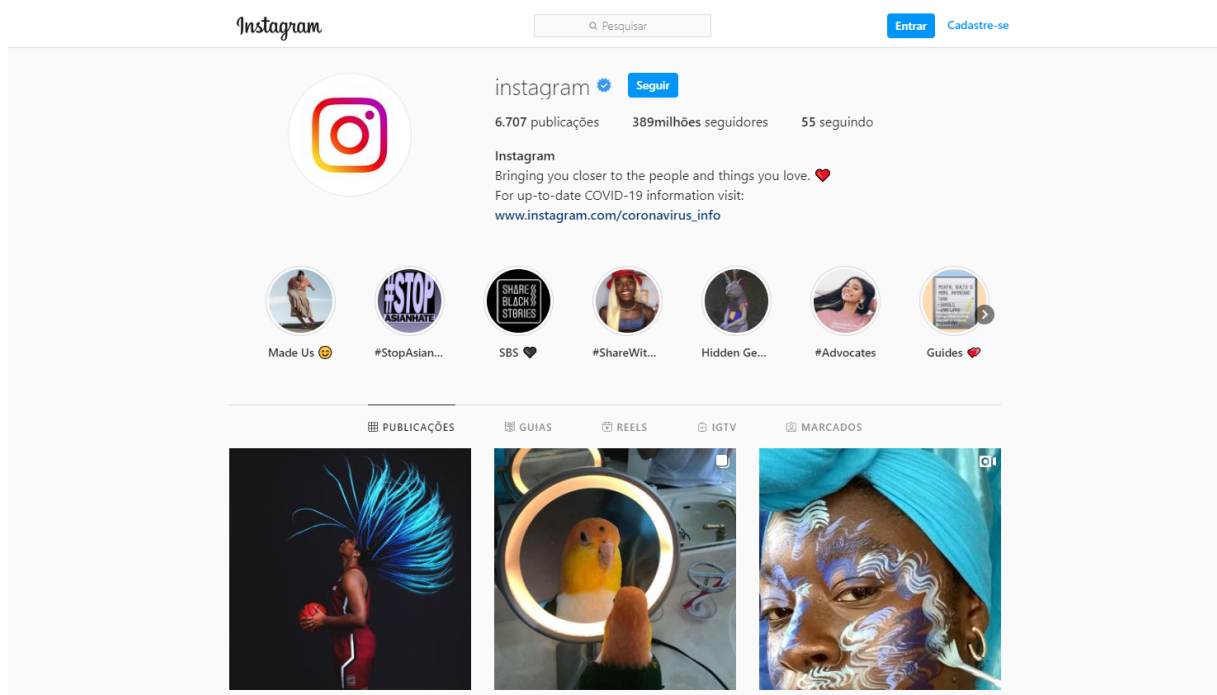


Figura 2.8: A rede social Instagram é uma das plataformas que utilizam a biblioteca React.

computação, é definida como um conjunto de subprogramas, funções e códigos auxiliares pré-escritos, que auxiliam no desenvolvimento de novos programas.

Surgido em 2011, no Facebook, passou a ser utilizada na interface do mural de notícias da rede social e, em 2012, também foi empregada na construção do Instagram, cujo exemplo da página pode ser visto na Figura 2.8. Em 2013 foi lançada oficialmente no formato de código aberto, ou seja, qualquer desenvolvedor que tenha interesse pode acessar o código e contribuir para o projeto. Atualmente é a biblioteca de desenvolvimento de interfaces mais popular do mundo, em pesquisas na ferramenta de busca do Google.

### 2.3.2 Canvas.js

O Canvas.js é uma biblioteca em JavaScript que permite a criação de forma simplificada de gráficos interativos, aplicados em páginas web. É utilizado por organizações como Microsoft, Nasa, Apple e Intel<sup>3</sup>. A licença de utilização é gratuita para uso não comercial, em projetos sem fins lucrativos e estudantis. Na Figura 2.9 é possível ver um exemplo de gráfico criado pela biblioteca em questão.

<sup>3</sup>Fonte: [www.canvasjs.com](http://www.canvasjs.com)

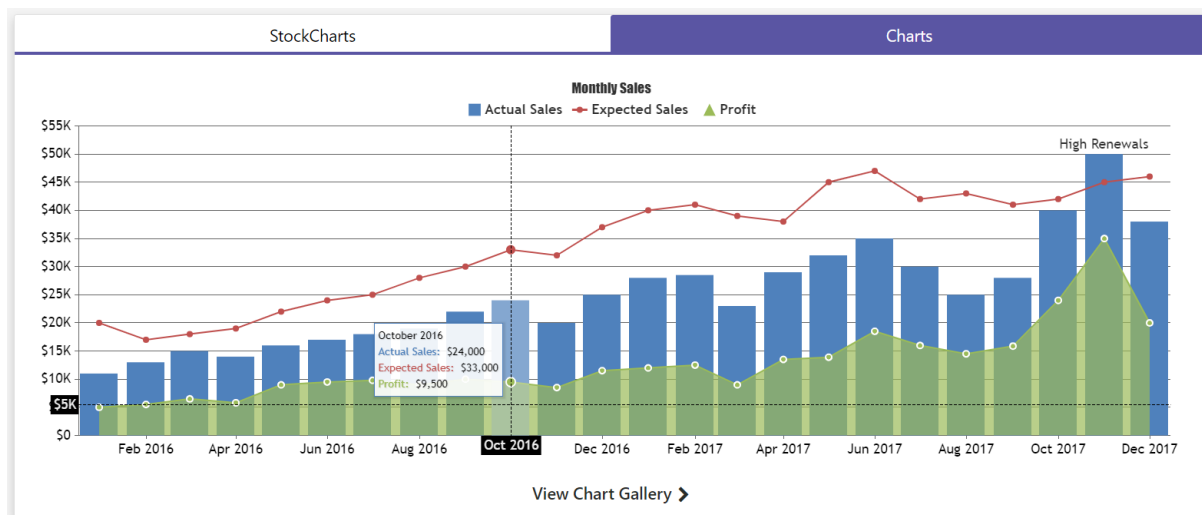


Figura 2.9: Gráfico da biblioteca Canvas.js com dados de exemplo, exibindo valores de vendas, expectativa e lucro ao longo do tempo.

### 2.3.3 Pacotes JavaScript

A interface gráfica deste trabalho foi construída utilizando alguns pacotes de software. O pacote é representado por um conjunto de arquivos que compõem um software já pronto e pode ser instalado em outra aplicação. O projeto foi concebido utilizando o gerenciador de pacotes JavaScript npm (Node Package Manager), que possui uma base de dados online contendo os pacotes [20]. Para utilizar o npm, foi necessário instalar o Node.js, que pode ser definido como um ambiente de execução Javascript do lado do servidor. Isso significa ser possível executar o código JavaScript em uma máquina (servidor), sem a necessidade de um navegador web para funcionar.

O primeiro pacote utilizado foi o Material-UI. Esta é uma biblioteca de componentes React para um desenvolvimento ágil e fácil, permitindo a criação do próprio design de uma página [21]. Ela fornece componentes já estilizados para serem incluídos na criação de uma interface como, por exemplo, botões, campos de texto, caixas de seleção e menus de navegação.

O pacote Lodash também foi utilizado como um auxiliar para a programação do conteúdo da interface. Ele é uma biblioteca JavaScript que fornece funções utilitárias para tarefas comuns de programação, com consistência, modularidade e eficiência.

Para calcular a Frente de Pareto apresentada nos gráficos da interface, foi empregado o pacote *Pareto Frontier*. A Frente de Pareto ou Eficiência de Pareto define um estado de alocação de recursos de um sistema, em que é impossível realocá-los tal que a situação de qualquer participante seja melhorada sem piorar a situação individual de outro participante [22]. Ou seja, representa o conjunto de escolhas de elementos que otimizam um sistema. O pacote calcula o conjunto ideal de pontos em um sistema bidimensional.

O pacote *Euclidean Distance* é usado para calcular a distância euclidiana entre dois vetores n-dimensionais, representando a reta que descreve a menor distância entre dois pontos no espaço [23]. A distância euclidiana entre os pontos  $P(p_1, p_2, \dots, p_n)$  e  $Q(q_1, q_2, \dots, q_n)$  é representada pela Equação 2.4:

$$D(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (2.4)$$

## 2.4 Ferramentas para Previsão

O sistema desenvolvido para este trabalho inclui um módulo de previsão das propriedades Constante Dielétrica e Fator de Qualidade, utilizando como entrada de dados a quantidade de elementos químicos utilizados na composição do material. Ele foi concebido utilizando os conceitos de Aprendizado de Máquina (Machine Learning) e AutoML (Automated Machine Learning - Aprendizado de Máquina Automatizado). A interface gráfica desenvolvida interage diretamente com o previsor, permitindo a escolha dos elementos químicos da composição do material e suas respectivas quantidades, enviadas ao módulo de previsão para retornar os valores das propriedades físicas de interesse.

### 2.4.1 Aprendizado de Máquina

A tecnologia de Aprendizado de Máquina (Machine Learning) é um subcampo da área de Inteligência Artificial, que se refere à capacidade de um computador aprender um padrão de forma autônoma a partir de um algoritmo e um conjunto de dados. Definido em 1959 por Arthur Lee Samuel como "campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados" [24]. Este campo de estudo está ligado à estatística computacional e possui aplicabilidade em realizar previsões a partir do que foi aprendido pela máquina. O Aprendizado de Máquina pode ser subdividido em três categorias [25]:

- **Aprendizado Supervisionado:** os dados são apresentados ao computador informando as entradas e saídas desejadas. O objetivo é aprender um padrão ou regra geral para as informações fornecidas, similar ao trabalho aqui apresentado. Existem basicamente duas subdivisões:

- **Regressão:** consiste na criação de um modelo capaz de prever um valor numérico. Por exemplo, a partir de um conjunto de entradas prever o valor de um imóvel.
- **Classificação:** as entradas da base de dados são divididas em duas ou mais classes pelo modelo desenvolvido. Um exemplo é a classificação de um e-mail, que pode ser “spam” ou “não spam”.
- **Aprendizado Não Supervisionado:** não é apresentado ao computador uma etiqueta para as informações em si, a máquina fica a cargo de encontrar a estrutura sozinha. Mecanismo muito utilizado para descobrir novos padrões em um conjunto de dados. Nesse subconjunto, destacam-se duas técnicas:
  - **Associação:** permite descobrir regras e correlações, identificando conjuntos de itens que ocorrem frequentemente juntos. Utilizado por varejistas em análise de carrinhos de compras.
  - **Agrupamento (Clusterização):** o conjunto todo em análise sofre segmentações em vários grupos, com base nas semelhanças encontradas. Por exemplo, agrupar um conjunto de espécies de seres vivos em um gênero, no conceito de taxonomia da biologia.
- **Aprendizado por reforço:** a aplicação interage com o ambiente dinamicamente, a fim de atingir um objetivo pré-determinado. Enquanto é executado, há uma retroalimentação das informações do meio e ele se corrige automaticamente. Utilizado, por exemplo, na programação de carros autônomos e na inteligência de adversários artificiais em jogos eletrônicos.

## 2.4.2 AutoML

O AutoML (*Automated Machine Learning*), ou Aprendizado de Máquina Automatizado, é o processo que automatiza as tarefas para criar o modelo de aprendizado de máquina, que melhor representa os dados utilizados e atinge o objetivo proposto [26]. O AutoML facilita a criação do modelo representativo final comparando vários algoritmos de Aprendizado de Máquina, assim como avalia o mesmo algoritmo com diferentes parâmetros para obter a melhor configuração. O desenvolvimento de um modelo de Aprendizado de Máquina padrão requer a comparação manual de dezenas de modelos para alcançar o melhor resultado, essa técnica permite alcançar resultados eficazes com economia de tempo, recursos e aproveitando as melhores práticas de ciência de dados.

AutoGluon é um conjunto de ferramentas desenvolvido em Python que permite a fácil utilização do processo de AutoML [27]. Possui foco no aprendizado profundo e em aplicações reais que abrangem imagens, textos ou dados tabulares. Destinado tanto a iniciantes quanto a especialistas em ML, o AutoGluon permite a criação rápida de protótipos de soluções de aprendizado de máquina para os dados escolhidos, com poucas linhas de código de programação. Provê o ajuste automático dos parâmetros, seleção de um modelo e o processamento de dados. O modelo desenvolvido neste trabalho foi construído a partir das ferramentas fornecidas pelo AutoGluon e inserido em um Microserviço para comunicar com a Interface Gráfica.

### 2.4.3 Microserviços

O módulo previsor utiliza o paradigma de arquitetura de microserviços. É uma abordagem de arquitetura de sistemas que visa organizar o software em pequenos serviços independentes que se comunicam usando APIs (*Application Programming Interface* — Interface de Programação de Aplicações) bem definidas [28]. A API é um conjunto de padrões estabelecidos e documentados por uma determinada aplicação, que permitem a utilização de suas funcionalidades por outros sistemas, sem precisar conhecer detalhes da implementação do software. Essas rotinas permitem uma interoperabilidade entre aplicações.

Os microserviços representam módulos autossuficientes, capazes de realizar lógicas computacionais. As arquiteturas de microserviços facilitam a escalabilidade e agilizam o desenvolvimento de aplicativos. O Microserviço desenvolvido neste trabalho carrega em si o modelo de Aprendizado de Máquina, ou seja, o módulo Previsor, que calcula o valor da Constante Dielétrica e Fator de Qualidade a partir dos elementos químicos escolhidos e suas respectivas quantidades. Ele se comunica com a Interface de Usuário via chamadas REST.

REST significa *Representational State Transfer*. Em português, Transferência de Estado Representacional. Trata-se de uma abstração da arquitetura da *World Wide Web*, com um conjunto de princípios que permitem a criação de um projeto com interfaces bem definidas e a comunicação entre sistemas [29]. Ele utiliza o HTTP (*Hypertext Transfer Protocol*), sendo um protocolo de transferência que possibilita que usuários insiram a URL (*Uniform Resource Locator*) do seu site na Web e acessem os conteúdos e dados que nele existem.

O padrão REST apresenta um conjunto de verbos como padrão para serem utilizados no desenvolvimento de aplicações. Os 5 principais verbos ou métodos são:

- **GET:** obter informações;
- **POST:** enviar dados para um sistema;
- **DELETE:** apagar informações;
- **PUT:** utilizado para atualizar completamente um registro;
- **PATCH:** usado para atualizar um registro parcialmente.

O Microserviço deste trabalho foi desenvolvido utilizando a linguagem de programação Python. O framework de desenvolvimento Web conhecido como Flask foi utilizado para possibilitar a implementação dos verbos do padrão REST, que permitem a comunicação com outros sistemas. Além disso, foi utilizado o Docker, o qual é uma plataforma de código aberto para criação de ambientes isolados via contêiner. O Docker permite a disponibilização de uma aplicação independente, como se estivesse funcionando em um computador próprio. A arquitetura final do sistema pode ser vista na Figura 2.10, em que demonstra a união de todas as ferramentas citadas anteriormente.

## 2.5 Métricas de Erro para Avaliação de Desempenho do Previsor

Existem um conjunto de métricas diferentes para medir a acurácia de modelos preditivos. Neste trabalho há um foco em três das principais métricas, se tratando de modelos com variáveis contínuas [30]. São modelos que avaliam a técnica de regressão, cujo objetivo é encontrar o valor numérico que melhor corresponde ao conjunto de dados de entradas fornecidas.

### 2.5.1 MAE

O significado da sigla em inglês é *Mean Absolut Error*. Em tradução livre, é o cálculo do Erro Absoluto Médio entre os valores reais e previstos. A Equação 2.5 representa como o MAE é calculado.

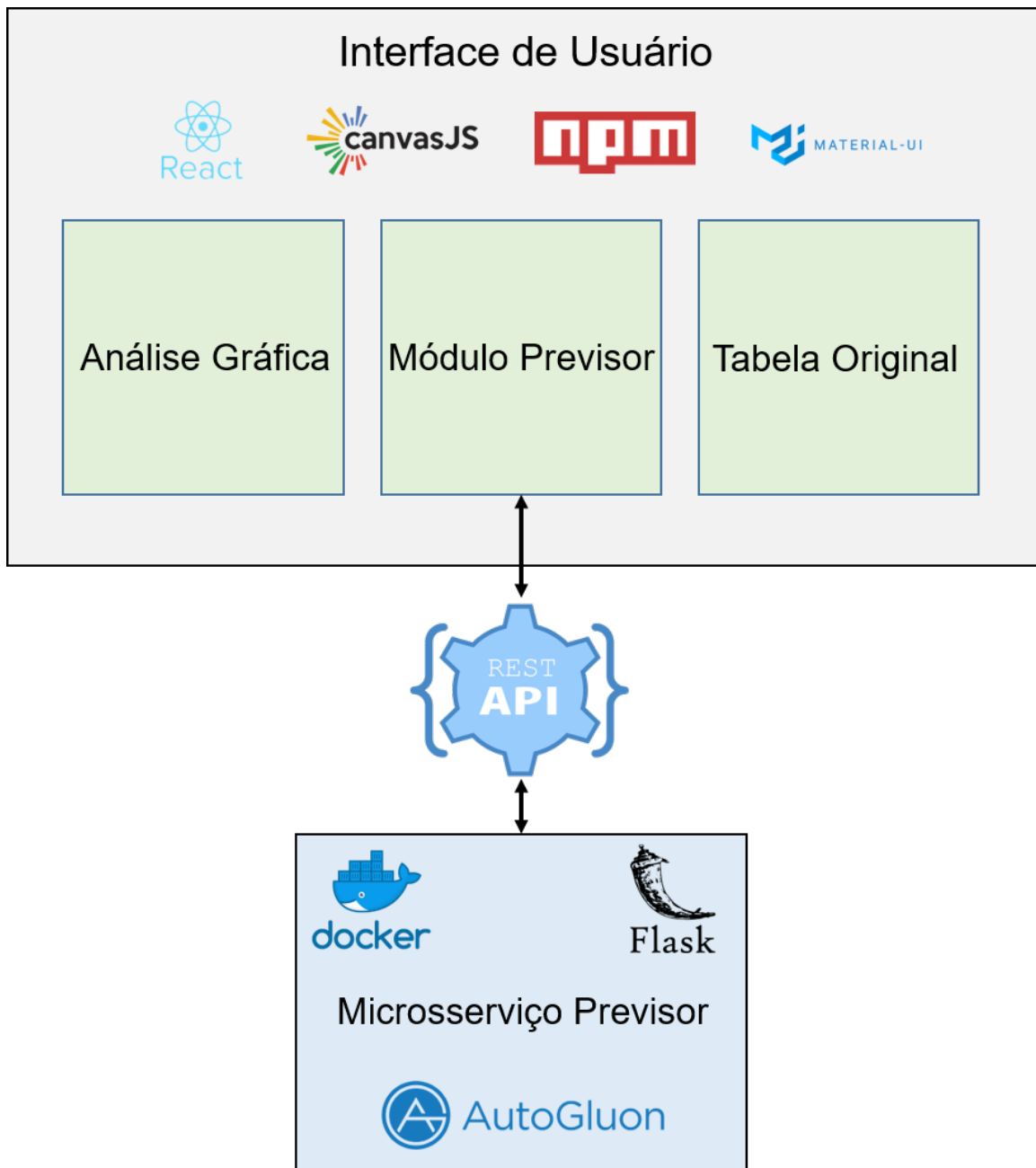


Figura 2.10: O sistema desenvolvido possui a Interface Gráfica de Usuário criada utilizando React, Material-UI, canvas.js e pacotes npm auxiliares. Ela se conecta via requisição API com o Microserviço responsável pela previsão das propriedades físicas, desenvolvido utilizando Docker, Flask e AutoGluon.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2.5)$$

onde:

$y_i$  = valor atual verdadeiro;

$\hat{y}_i$  = valor previsto para  $y_i$ .

A parte interna da Equação 2.5 possui  $y_i - \hat{y}_i$ , representando o cálculo da diferença entre o valor real e o valor previsto, dando origem ao valor do erro em si. Como é uma operação de módulo, não está sendo considerado se o erro é positivo ou negativo, apenas a diferença entre o real e previsto. É realizado o somatório ( $\sum$ ) de todos os erros, considerando cada uma das instâncias analisadas. Por fim, é feita a divisão pelo número total de instâncias  $n$  extraído a média de todos os erros.

### 2.5.2 RMSE

A sigla se refere a *Root Mean squared Error*, é a medida que calcula a Raiz Quadrática Média dos Erros, que ocorrem pela comparação dos valores reais e os valores previstos, dada pela Equação 2.6.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.6)$$

Similar ao MAE, é internamente extraído o erro ao realizar a operação  $y_i - \hat{y}_i$  e então elevado ao quadrado. Também é feito o somatório de todos os erros ao quadrado, cujo resultado é dividido pela quantidade total  $n$  de instâncias. Por fim, é calculada a raiz quadrada da média obtida pela operação interna da equação.

### 2.5.3 R2

Conhecido como *R-squared*, ou R ao quadrado, representa o Coeficiente de Determinação [31]. Determina quão bem os valores se ajustam em comparação com os valores originais. A Equação 2.7 descreve como é realizado o cálculo para R2.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.7)$$

onde:

$\bar{y}$  = valor médio de todos os  $y_i$ .

Na parte superior da Equação 2.7 ocorre a Soma dos Quadrados dos Resíduos representada por  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ . É exatamente o cálculo que ocorre na parte interna da Equação 2.6, que descreve a parte que não é explicada pelo modelo em questão. A parte inferior da equação, dada por  $\sum_{i=1}^n (y_i - \bar{y})^2$ , representa a soma dos quadrados das diferenças entre a média e cada valor observado. O Coeficiente de Determinação pode ser descrito como 1 menos a Soma dos Quadrados dos Resíduos dividido pela Soma dos Quadrados.

#### 2.5.4 Análise das Métricas

Os valores de MAE e RMSE podem ir de 0 a infinito. Eles variam conforme a escala dos valores analisados pelos modelos. Quanto menor o valor de ambos, melhor é o modelo, por apresentar uma menor taxa de erros. Uma característica do RMSE é que os erros (reais — predições) são elevados ao quadrado antes de ter a média calculada. Portanto, pesos diferentes serão atribuídos à soma e, conforme os valores de erros das instâncias aumentam, o índice do RMSE aumenta consideravelmente. Ou seja, se houver um ponto fora da curva no conjunto de dados, seu peso será maior para o cálculo do RMSE e, por conseguinte, prejudicará sua métrica deixando-a maior. Em resumo, MAE utiliza valores absolutos dos erros e é menos sensível à variação que os dados podem apresentar. Já o RMSE tem o benefício em penalizar os erros de maior magnitude.

Em relação à métrica R2, os valores variam de 0 a 1 e podem ser interpretados como porcentagem. Quanto mais próximo de 1, melhor o modelo explica o comportamento dos dados utilizados. Supondo um valor de R2 igual a 0,83, isso implica que 83% da variabilidade das variáveis dependentes no conjunto de dados foram contabilizadas, indicando a precisão do modelo adotado [32]. O R2 compara o ajuste do modelo escolhido com o de uma linha reta horizontal, ou seja, a hipótese nula. Se o modelo escolhido se encaixa pior do que uma linha horizontal, então R2 é negativo. A Figura 2.11 mostra um exemplo no qual os dados são representados pelos pontos azuis no gráfico, a linha vermelha representa a linha horizontal base de comparação e a linha verde o modelo de regressão que prevê esse conjunto de dados. Para esse exemplo, o valor de R2 é mais próximo de 1, por explicar bem o comportamento dos dados listados.

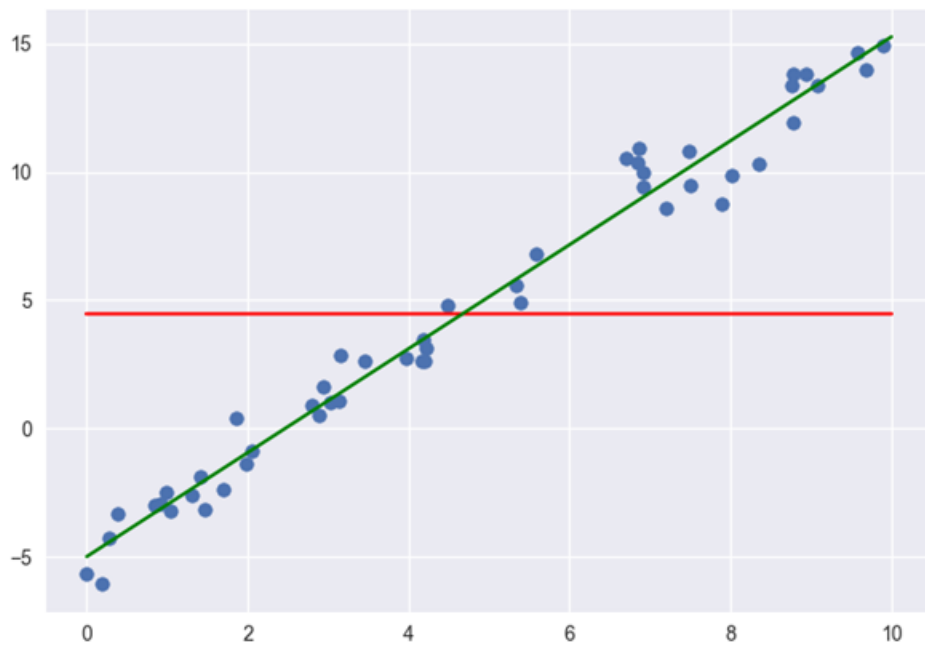


Figura 2.11: Exemplo de modelo com valor da métrica  $R^2$  próximo a 1.

## Capítulo 3

# Trabalhos relacionados

O avanço científico e tecnológico expandiu a quantidade de dados gerados por diversas fontes, o que impactou a área de Ciência de Materiais e seus estudos. A análise e estruturação de quantidades massivas de dados é uma área do conhecimento chamada de *Big Data* (Grandes Dados, em português). Essa área impulsionou a Informática de Materiais, que passou a desenvolver diversas técnicas de predição de propriedades e também descoberta de novos materiais [8]. Se considerarmos apenas os materiais na escala nanométrica, essa área pode ser denominada Nano informática, sendo uma ramificação da Nanotecnologia Computacional [33].

A Informática de Materiais é uma abordagem eficaz que pode acelerar o desenvolvimento de novas estruturas, combinando Ciência de Materiais e técnicas de Inteligência Artificial. Como um exemplo, podemos citar aplicações de sucesso em projetos de materiais térmicos, como materiais de interface térmica para dissipação de calor e materiais termoelétricos para geração de energia [34]. Esses exemplos mostram o progresso das pesquisas associadas aos estudos de previsão de propriedades físicas e descoberta de novos materiais, utilizando técnicas computacionais.

Atualmente, ferramentas modernas de caracterização de materiais geram uma enorme quantidade de dados, que podem ser estruturados e analisados [35]. Além disso, vários modelos físicos sofisticados, fornecem o incentivo para o desenvolvimento e implementação de protocolos baseados em dados, dando suporte aos vários estágios de desenvolvimento de novos materiais. Anteriormente, muitas dessas decisões eram unicamente tomadas com base no conhecimento empírico e nos instintos dos especialistas envolvidos nesses fluxos de trabalho. Este é um dos principais fatores que aumentam drasticamente o custo geral e o tempo para desenvolvimento de novos materiais. Os algoritmos e protocolos baseados em dados podem mitigar o risco inerente em grande medida, não apenas tomando decisões mais objetivas, mas também registrando as falhas e sucessos. Dessa forma, o conhecimento obtido com essas falhas e sucessos pode ser transferidos para trabalhos futuros. A seguir, citaremos alguns exemplos de trabalhos realizados.

## 3.1 Algoritmos Computacionais em Ciência de Materiais

Sistemas computacionais foram utilizados para calcular propriedades físicas de uma subclasse de materiais cerâmicos, conhecidos como Perovskitas de Haletos [36]. Uma das principais aplicações dessa classe de materiais está na construção de células absorventes fotovoltaicas. O algoritmo baseado na Teoria do Funcional da Densidade da Mecânica Quântica foi desenvolvido e aplicado em uma base de dados com cerca de 2000 compostos, sendo posteriormente disponibilizada online para estudos, contendo propriedades estruturais, eletrônicas e de transporte. Trabalhos que envolvem o estudo de óxidos também foram desenvolvidos, em que os autores exploraram 44 óxidos de Perovskita [37]. Algoritmos de química computacional foram criados, baseados em cálculos de primeiros princípios. A partir disso, foram sugeridos cinco materiais para aplicação na absorção de luz em células fotovoltaicas.

Considerando aplicações de Aprendizado de Máquina em Ciência de Matérias, há um conjunto de estudos e trabalhos disponíveis na literatura. Temos, por exemplo, o estudo em que um modelo de Aprendizado de Máquina foi desenvolvido para prever a distribuição local correta de compostos conhecidos como Meio-Heusler [38]. O cálculo foi baseado apenas na composição dos materiais e apresentou uma precisão de 95%, representando uma melhora significativa em comparação com outras heurísticas já implementadas. Para realizar a construção do modelo, foi necessário inspecionar todos os dados individual e meticulosamente, a fim de garantir a qualidade e precisão final. Outro desafio encontrado foi a quantidade de instâncias da base, que pode ser considerada pequena, com um máximo de 103 amostras.

Em um trabalho mais genérico, para lidar com estruturas atômicas e compostos químicos, diferentes metodologias foram desenvolvidas a fim de melhorar o desempenho das tratativas computacionais. Em uma dessas abordagens foi proposta uma "impressão digital", que captura vizinhança atômica única de material [39]. Esta impressão digital pode ser combinada com abordagens de Aprendizado de Máquina, para aprender diversas propriedades de materiais, como energias, forças atômicas, tensões, densidade de carga eletrônica, densidade eletrônica de estados e assim por diante. A impressão digital foi usada para classificar diferentes ambientes atômicos em conjuntos de dados que representam superfícies, defeitos planares e pontuais, fornecendo assim uma métrica confiável para identificar defeitos estruturais. Também foi utilizada para construir um modelo geral de energia potencial de Al (alumínio), cujo desempenho foi validado para um conjunto diversificado de configurações de Al, indo muito além do conjunto de treinamento. A precisão do modelo de energia foi validada para várias propriedades físicas do Al e mostrou estar

em boa concordância.

Outra aplicação consiste na descoberta de materiais, como novas ligas com memória de forma [40]. Esse trabalho utilizou um grande banco de dados gerados por cálculos de alto rendimento. Foram estudados diferentes métodos de Aprendizado de Máquina, como otimização multiobjetivo e de múltiplas finalidades, que permitem propor novos materiais com a otimização simultânea de mais de uma propriedade física como alvo. Esses métodos avançados permitem o uso de menos experimentos e cálculos complexos, combinados com outras estratégias menos dispendiosas para obter uma modelagem comparável às abordagens que utilizam apenas métodos computacionalmente mais caros.

Uma série de estratégias computacionais estão sendo bem-sucedidas e tomando forma na Ciência de Materiais [41], sendo impulsionadas em parte pela *Materials Genome Initiative*<sup>1</sup> e em parte pelos desenvolvimentos algorítmicos e os sucessos de esforços orientados a dados em outros domínios. Estão sendo desenvolvidos modelos de Aprendizado de Máquina que permitem previsões rápidas, baseadas puramente em dados anteriores. Não são utilizadas experimentações diretas por cálculos ou simulações, equações fundamentais são explicitamente resolvidas. Estes métodos estão se tornando úteis para determinar propriedades de materiais que são difíceis de medir ou computar usando métodos tradicionais, devido ao custo, tempo ou esforço envolvido. As previsões são tipicamente interpolativas, envolvendo primeiro a impressão digital de um material numericamente. As impressões digitais, também chamadas de descritores, podem ser de vários tipos e escalas, conforme ditado pelo domínio e pelas necessidades da aplicação. Previsões também podem ser extrapolativas, estendendo-se para novos espaços de materiais, desde que as incertezas das previsões sejam devidamente levadas em consideração.

Outro exemplo interessante, e de certa forma relacionado a este trabalho, é o estudo da temperatura de fusão de materiais cerâmicos por técnicas computacionais. Esta propriedade tem grande influência na avaliação dos limites de temperatura que uma cerâmica suporta durante a execução de um trabalho. Esse subgrupo de materiais são conhecidos como *Ultra-High Temperature Ceramics* (UHTCs), ou seja, Cerâmicas de Temperatura Ultra Alta. A fim de contornar o desafio na medição de pontos de fusão ultra altos, um método para prever a temperatura de fusão desses materiais foi proposto via Aprendizado de Máquina [42]. Um conjunto de dados incluindo mais de dez mil dados de temperatura de fusão foi estabelecido, que cobre 8 elementos e a maioria dos UHTCs não óxidos conhecidos. Uma Rede Neural Artificial com retro propagação foi treinada, obtendo precisão próxima a 90%. Esse tipo de estudo auxilia a criação de novos materiais com temperatura de trabalhos mais altas.

Também com foco no desenvolvimento de novos compostos, uma estrutura de Aprendizado de Máquina de propósito geral foi desenvolvida, para prever as propriedades de uma ampla variedade de materiais [43]. Essa estrutura foi aplicada para descobrir

---

<sup>1</sup>Página: [www.mgi.gov](http://www.mgi.gov)

novos compostos cristalinos com aplicações fotovoltaicas e identificar possíveis novas ligas de vidro metálico. O método funciona usando Aprendizado de Máquina para gerar modelos que preveem as propriedades de um material, em função de uma ampla variedade de atributos projetados, para aproximar os efeitos químicos do mesmo. A técnica consegue criar modelos precisos para propriedades tão diferentes quanto as propriedades eletrônicas de compostos cristalinos e a conformabilidade do vidro de ligas metálicas.

Por fim, uma aplicação interessante consiste em um modelo de Aprendizado de Máquina baseado em biomimética, que abrange materiais cerâmicos, polímeros e metais [44]. A abordagem adotada no trabalho conseguiu criar padrões de microestruturas, que levaram à concepção de materiais mais resistentes. Os resultados foram validados por manufatura aditiva e diversos testes.

Além dos estudos aqui mencionados, há hoje uma grande variedade de trabalhos aplicando técnicas de Inteligência Artificial na descoberta de novos materiais, muitos na escala nanométrica, com propriedades diferenciadas. Cada vez mais, novos trabalhos de Nanotecnologia Computacional, aplicados em materiais, surgem na literatura. Listamos aqui algumas das referências recentes, demonstrando a eficácia dessas técnicas em diferentes contextos, correlacionados ao trabalho desta dissertação.

Este trabalho se difere dos demais por estudar o grupo de materiais cerâmicos dielétricos, visando a previsão e análise das propriedades físicas Constante Dielétrica e Fator de Qualidade. Outro diferencial em relação aos outros trabalhos está na implementação de um sistema inteiramente novo de análise dessas propriedades. O sistema está acoplado ao módulo previsor, diferentemente dos demais.

No capítulo seguinte, será apresentada a metodologia utilizada no tratamento de dados, que originou a parte de Interface de Usuário do sistema proposto. Além disso, serão apresentados os módulos e funcionalidades presentes na interface, que permite a realização de análises com base em gráficos e filtros. Também serão apresentados os exemplos de utilização dos modelos previsores desenvolvidos.

# Capítulo 4

## Tratamento e Visualização de Dados

O trabalho apresentado nesta dissertação pode ser subdividido em duas principais partes: a primeira é relativa ao pré-processamento dos dados e estruturação da interface de usuário para visualização; já a segunda trata da construção dos modelos previsores. A primeira parte é apresentada neste capítulo, em que será discutido como as informações utilizadas neste trabalho foram obtidas e tratadas, permitindo seu posterior uso na interface de usuário. Em seguida, é explicada a estrutura da interface criada utilizando React e pacotes npm. Os resultados e a apresentação da parte visual da interface também são mostrados.

### 4.1 Pré-processamento e estruturação dos dados

Os dados utilizados nesta dissertação foram obtidos a partir do catálogo de mais de quatro mil materiais cerâmicos e suas propriedades [9]. Para cada material são exibidas as propriedades físicas, informações sobre a sua estrutura, propriedades de sintetização do cerâmico e a referência bibliográfica utilizada como fonte desses dados. A Tabela 4.1 contém exemplos dessa base de materiais cerâmicos, mostrando diferentes padrões de fórmulas químicas apresentadas ao longo da referência.

A primeira coluna da Tabela 4.1 apresenta o identificador numérico do material,

Tabela 4.1: Exemplos de materiais contidos na referência base para este trabalho.

No.	Material	TS (°C)	Estrutura Cristalina	$\epsilon_r$	$Qf$ (GHz)	$f_0$	$\tau_f$	Referência
1	AlPO4+5 wt% MgF2	1450	Orthorhombic	3.0	900	5.8	-	1
2	50%Li2CO3-40.24%B2O3-9.76%SiO2	600	glass	3.2	10500		-79	2
6	SiO2 melt method		Fused silica	3.7	122100			5
8	SiO2	1650	Trigonal P3121	3.8	80400		-16	7
13	CaO-B2O3-SiO2 (19.8:30.9:49.3 mol%)	900	Glass	4.1	2000	9.9		8
21	BF33 glass	-	Glass	4.6	860	5		15
73	Al2O3+50 vol% ZBS glass	900	Composite	5.7	17800	12.6		35
444	xMgO-(1-x)B2O3 (x=0.993)	1350/4h	Orthorhombic Pnmm	9.3	773700		-55	93
445	xMgO-(1-x)B2O3 (x=0.995)	1350/4h	Orthorhombic Pnmm	9.3	551700		-56	93
1350	La2.98/3Sr0.01(Mg0.5Sn0.5)O3	1550/4h		20.0	57100		-77	619
3369	0.4CaTiO3-0.6Ca(Mg1/3Nb2/3)O3	1450	Perovskite	54.0	32000	6.7	18	944
3469	Li1+x+yTa1-x-3yTix+4yO3 (x=0.1, y=0.075)	1175/1h	M-Phase	60.5	5014		-5	1399
3585	(Ba4.2Sm9.2) $\alpha$ Ti18-yAlYO54 (y=0.8, $\alpha = 1+y/36$ , x=0.6)	1440	Orthorhombic Tungsten Bronze	72.1	4600		-42	1484
4129	Ba0.6Sr0.4TiO3+0.5 wt% MgCo2(VO4)2		Perovskite	2763.0	300	1		1672

considerando a ordenação pela Constante Dielétrica ( $\epsilon_r$ ), da menor para a maior. A segunda coluna apresenta a fórmula química, ou nome, do material, enquanto a terceira exibe a Temperatura de Sintetização (TS). Na sequência, é apresentada a Estrutura Cristalina do material, Constante Dielétrica e o produto do Fator de Qualidade pelo Frequência,  $Qf$ , citado apenas como Fator Qualidade ao longo deste trabalho, para simplificação. A Frequência é dada por  $f_0$ , enquanto  $\tau_f$  representa o Coeficiente de Variação de Temperatura da Frequência Ressonante (ppm/°C) e, por fim, a referência bibliográfica.

Os dados originais estavam armazenados em um arquivo PDF, sendo então necessário aplicar um software para convertê-lo para CSV. A partir disso, foi possível obter uma base de dados que poderia ser tratada pelos algoritmos desenvolvidos neste trabalho. A partir da criação do CSV, trabalhamos em uma planilha para transformar cada descrição de um material em um único formato padrão. Cada composto apresentado pelo autor, possui uma codificação diferente que representa a concepção do material, ou seja, como ele foi pensado para originar a fórmula. Sendo assim, cada um dos autores utilizou de uma lógica distinta para chegar no composto químico final, como pode ser visto na Tabela 4.1.

Além da ausência de padrões, algumas fórmulas apresentaram erro de digitação e ausência de informações. As instâncias que apresentaram esses problemas foram removidas da base final, por não ser possível usá-las corretamente. A base também apresentou compostos com o nome patenteado do material como, por exemplo, MEMPAX, BF33 e AF32. Para esses casos, foi realizada uma consulta na literatura e nas referências apresentadas para obter a fórmula química original. Foi feita a substituição de todos os termos patenteados pelas fórmulas que as representam, para prosseguir com a estruturação da base de dados.

Os 4129 materiais originais foram inspecionados e, por decisão de projeto, os materiais que possuem mais de um valor para Constante Dielétrica e Fator de Qualidade foram duplicados, dando origem a 4145 instâncias. Devido às particularidades encontradas nos diversos padrões de fórmulas e alguns erros de digitação, não foi possível criar um único algoritmo padrão para calcular todas as fórmulas dos compostos de uma só vez. Foi realizada uma inspeção individual nos 4145 itens, removidas as instâncias com erro de digitação encontradas e substituídos os materiais patenteados pelas fórmulas originais a partir de pesquisas nas referências. Após inspeção, a quantidade final de itens utilizados no sistema de previsão foi de 4084. Os demais itens ainda foram aproveitados para exibição na ferramenta de análise, pois não impactou nos algoritmos utilizados pela interface. Mais informações sobre o conjunto de dados utilizados para o treinamento podem ser encontradas no Apêndice A.

Cada item restante foi então convertido para um padrão único, onde o importante é a quantidade de cada elemento químico que o compõe. Por exemplo, o  $\text{SiO}_2$  fica representado como Si valor 1 e O valor 2. Materiais como  $\text{BaTeO}_3$  e  $\text{Al}_2\text{O}_3$  não necessitam de

ajustes e já estão no formato final. Em alguns casos, há um conjunto de porcentagens ou pesos que descrevem a fórmula. Para chegar ao resultado, esses pesos ou porcentagens são utilizados para multiplicar a quantidade de cada elemento químico. A fórmula  $\text{ZnO}:\text{B}_2\text{O}_3$  (60:40) é um exemplo, no qual a parte à esquerda representa 60% e à direita 40%. O resultado é dado por  $\text{Zn}_{0.6}\text{O}_{2.4}\text{B}_{1.2}$ , em que cada parte foi multiplicada respectivamente por 0,6 e 0,4.

Foi adotado o conceito de dopagem de materiais para realizar o cálculo da fórmula. Vários compostos possuem wt%, vol% e mol%, que indicam respectivamente porcentagem de peso, porcentagem em volume e porcentagem em mol. A conversão direta entre cada uma dessas métricas afeta pouco o valor final da porcentagem. A partir desse conceito foi adotado o seguinte padrão: todos as partes de fórmulas químicas com valores iguais ou inferior a 10% seriam consideradas dopagem e as demais seriam calculados os valores proporcionalmente em todas as partes da fórmula. Para as dopagens abaixo de 10%, foi realizada uma penalização da quantidade total e, ao invés de ser calculado 0,1 vezes o valor apresentado, foi calculado com o multiplicador de 0,001.

#### Exemplo de criação da equação química:

$$\text{Sr}_2\text{Ce}_2\text{Ti}_5\text{O}_{16}(\text{Sr}_{1-3x/2}\text{Ce}_x\text{TiO}_3 \ x = 0.4) + 0.2 \text{ wt}\% \text{ La}_2\text{O}_3, \quad (4.1)$$

$$\text{Sr}_2\text{Ce}_2\text{Ti}_5\text{O}_{16}(\text{Sr}_{1-1.2/2}\text{Ce}_{0.4}\text{TiO}_3) + 0.2 \text{ wt}\% \text{ La}_2\text{O}_3, \quad (4.2)$$

$$\text{Sr}_2\text{Ce}_2\text{Ti}_5\text{O}_{16}(\text{Sr}_{0.4}\text{Ce}_{0.4}\text{TiO}_3) + 0.2 \text{ wt}\% \text{ La}_2\text{O}_3, \quad (4.3)$$

$$\text{Sr}_{2.4}\text{Ce}_{2.4}\text{Ti}_6\text{O}_{19} + 0.2 \text{ wt}\% \text{ La}_2\text{O}_3, \quad (4.4)$$

$$\text{Sr}_{2.4}\text{Ce}_{2.4}\text{Ti}_6\text{O}_{19} + \text{La}_{0.0004}\text{O}_{0.0006}, \quad (4.5)$$

$$\text{Sr}_{2.4}\text{Ce}_{2.4}\text{Ti}_6\text{O}_{19.0006}\text{La}_{0.0004}. \quad (4.6)$$

A Equação 4.1 descreve a representação de um dos materiais presentes na referência original. Nesta fórmula há uma variável  $x$  que precisa ser substituída, informações de porcentagem de peso e frações explícitas não calculadas. O segundo passo representado pela Equação 4.2 mostra o início da transformação da fórmula, para atingir o formato contendo apenas elementos químicos e suas respectivas quantidades. Nessa etapa o valor de  $x$  é substituído em toda a equação, eliminando a variável. A Equação 4.3 representa o cálculo do valor decimal da fração. Na etapa seguinte, representada pela Equação 4.4, todos os elementos do lado esquerdo são unificados, somando as quantidades dos elementos iguais. O próximo passo da Equação 4.5 consiste no cálculo da dopagem da porcentagem de peso (wt%). Todos os valores à esquerda são multiplicados pelo fator 0,0002. Como 0,2% é um valor abaixo de 10%, a penalização da divisão por 1000 é aplicada, reforçando que uma pequena quantidade desses elementos estão presentes na fórmula final. O último passo da Equação 4.6 representa a soma de todos os elementos químicos iguais, em todas as partes da fórmula.

## 4.2 Esquematização da Interface de Usuário

Os dados extraídos da referência base deste trabalho foram transformados para o formato JSON (*JavaScript Object Notation*). Essa estrutura de dados possui um formato leve e de fácil utilização para trocar informações entre sistemas. O JSON com todos os materiais foi utilizado para controle de estado na aplicação React, cujo exemplo pode ser visto no Código 4.1. Esse JSON alimenta os gráficos apresentados pela interface, serve como base para os filtros realizados na aplicação e também para a exibição da tabela completa na plataforma. Várias funcionalidades presentes na interface foram criadas utilizando os módulos de componentes disponibilizados pela biblioteca do Material-UI<sup>1</sup>. Os módulos de componentes prontos facilitaram o desenvolvimento, permitindo a criação de uma interface com maior usabilidade. Já é disponibilizado um padrão para botões, modais, menus, entre outros, facilmente acopláveis ao código da aplicação e personalizáveis.

Toda a interface segue os princípios de componentização e controle de estados do React. Foi criado o componente pai da aplicação que contém o cabeçalho de opções de navegação entre os gráficos, tabela e módulo previsor; além dos componentes filhos (gráficos, tabelas, abas de filtros e resultados da frente de pareto) que sofrem alterações em cascata quando estado pai é modificado. A Figura 4.1 exemplifica o funcionamento da componentização e da renderização de estados do React. Cada nó representa um componente, que pode ter seu próprio estado de controle, além de ter a possibilidade de herdar estados dos pais. As alterações nos estados dos nós pais, podem refletir em alterações nos nós filhos, caso herdem as propriedades modificadas.

Código 4.1: Exemplo de objeto JSON com um item da lista de materiais.

```
1 [ {
2   Id: 16,
3   Material: "Li3AlB206",
4   TemperaturaSintetizacao: 650,
5   EstruturaCristalina: "Triclinic",
6   ConstanteDieletrica: 4.2,
7   FatorQualidade: 12460,
8   FrequenciaMedicao: 16.8,
9   VariacaoTemperaturaFrequenciaRessonante: -290,
10  Referencia: 12,
11  Elementos: "Li,Al,B,O",
12  Compostos: "Li3,Al,B2,O6"
13 }
```

<sup>1</sup>Biblioteca de componentes: [www.material-ui.com](http://www.material-ui.com)

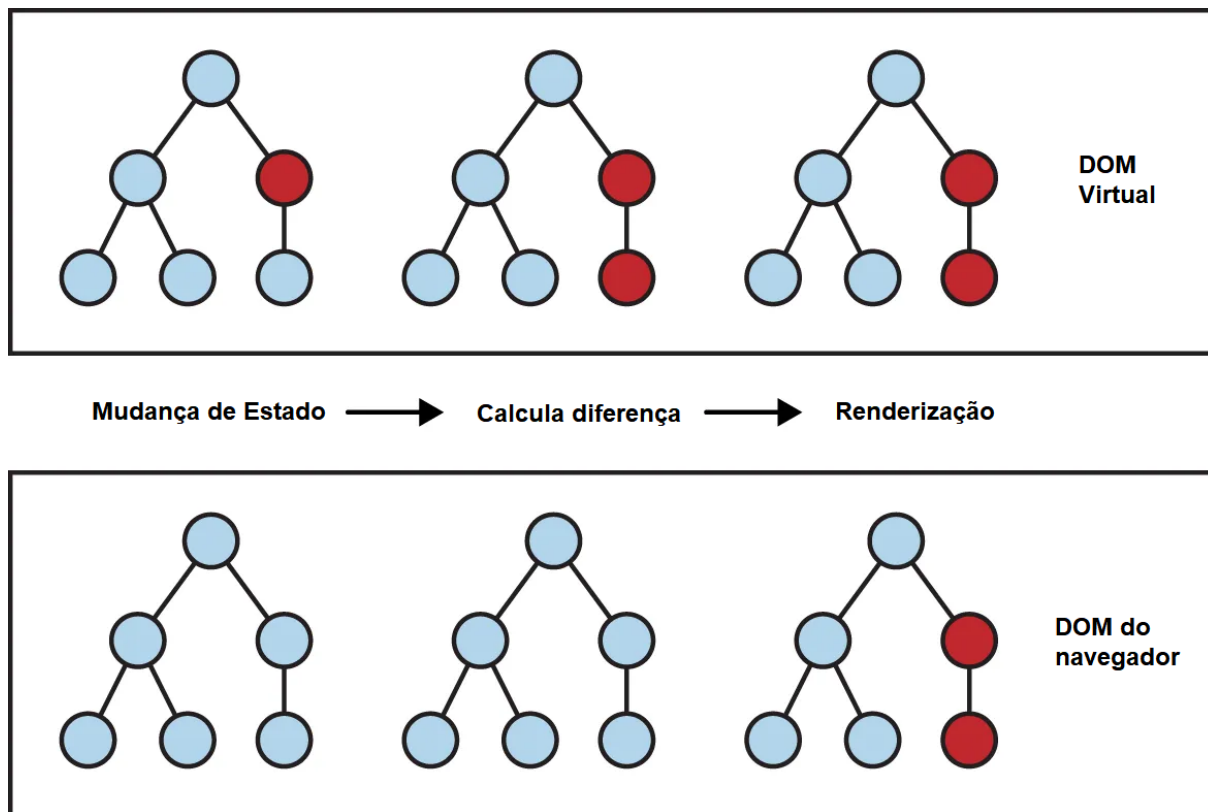


Figura 4.1: Atualização de estados nos componentes do DOM do navegador e do virtual.

A sigla DOM significa *Document Object Model* ou Modelo de Objeto de Documento. O DOM representa a interface de usuário da aplicação. Sempre que há uma mudança no estado, representada nos nós vermelhos na Figura 4.1, o DOM é atualizado para representar essa mudança. O problema é frequentemente manipular o DOM, que afeta o desempenho, tornando-o lento. O DOM é uma estrutura de dados em árvore, qualquer mudança nos nós pais representará alterações em cascata para os nós filhos. Quanto maior o número de ramificações, maior será o número de atualizações. O React utiliza a estrutura do DOM virtual, sendo apenas uma representação do DOM real, presente no navegador. Quando o estado da aplicação é atualizado em algum nó, ou componente, o DOM virtual é atualizado primeiramente.

Quando novos componentes são adicionados à interface, uma árvore do DOM virtual é criada. Cada componente é um nó nesta árvore. Se o estado de qualquer um desses elementos mudar, uma nova árvore DOM virtual será criada. Essa árvore é então comparada com a árvore DOM virtual anterior. A partir disso, o DOM virtual calcula o melhor método possível para fazer essas alterações no DOM real. Isso garante que haja operações mínimas no DOM real. Dessa forma, o custo de desempenho de atualização do DOM real é reduzido.

A interface apresentada neste trabalho possui a estruturação hierárquica de componentes presente na Figura 4.2. Os estados principais da aplicação estão presentes no **App** (Aplicação) e são passados para nós filhos conforme necessidade do componente. Os

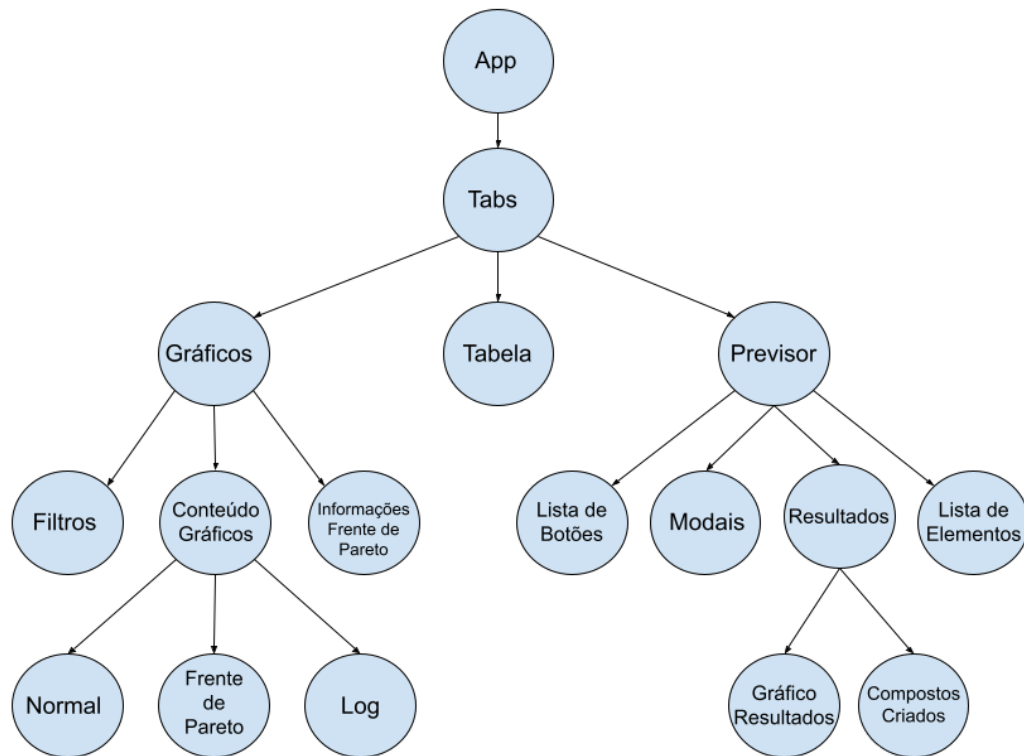


Figura 4.2: Esquematização dos componentes da interface de usuário desenvolvida.

componentes filhos só serão atualizados caso haja alterações nos estados que herdam dos nós pais. As **Tabs** (Abas) possuem a variável de estado que guarda o valor da aba atual, permitindo alternar entre as abas da aplicação. Existem três nós filhos principais:

- **Gráficos:** contém os componentes de **Filtros** por Elementos e por Compostos, menus laterais que permitem escolhas personalizadas para modificar o conteúdo dos gráficos; o **Conteúdo dos Gráficos** em si, com os gráficos **Normal**, **Frente de Pareto** e **Log**; também é pai do componente que possui as **Informações da Frente de Pareto**, menu lateral que pode ser aberto com as informações da Frente de Pareto original ou a gerada através do filtro aplicado;
- **Tabela:** componente responsável por transformar o JSON com os materiais, por meio de um laço de repetição para cada item da lista, em uma tabela no mesmo formato da referência original;
- **Previsor:** possui na parte superior uma **Lista de Botões**: um para abrir o modal de inserção de elementos e suas respectivas quantidades, e também um botão para submeter o material para análise. Os **Modais** estão ocultos na plataforma, são renderizados a partir da mudança de estado para o usuário poder escolher o elemento

e valor que deseja inserir na fórmula; a **Lista de Elementos** guarda todos os valores inseridos pelo usuário antes de submeter para análise, permitindo que a pessoa ainda realize alterações na fórmula química em criação; o componente de **Resultados** é pai das estruturas que contém o **Gráfico de Resultados** no formato Frente de Pareto e da listagem de **Compostos Criados**, exibindo a equação química e os valores previstos.

## 4.3 Interface de Usuário

A interface conta com uma série de ferramentas para facilitar a análise e estudo dos materiais cerâmicos dielétricos. Na tela inicial, são apresentados três gráficos, dois tipos de filtros, duas lógicas distintas por filtro e menus para visualizar os itens que pertencem à Frente de Pareto. Em seguida, é mostrada a tabela contendo todas as informações presentes na referência original, de forma simples e unificada. Por fim, a interface permite que o usuário crie sua própria fórmula química e realize uma simulação para obter as propriedades de Constante Dielétrica e Fator de Qualidade.

### 4.3.1 Análise Gráfica

Como mencionado no Capítulo 2, o pacote Canvas.js foi utilizado para facilitar a criação dos gráficos personalizados. A ferramenta em sua tela inicial de análise apresenta 3 formatos gráficos: o gráfico padrão, contendo no eixo X a Constante Dielétrica e no eixo Y o Fator de Qualidade multiplicada pela Frequência; o segundo gráfico apresenta a Frente de Pareto para observar os materiais dominantes em suas respectivas faixas de valores; por fim o gráfico logaritmo natural, que aplica a normalização na base  $\epsilon$  para os valores do eixo X e Y.

Os gráficos são renderizados considerando o estado atual da aplicação React. É armazenada uma lista com todos os itens inicialmente. Os filtros, que serão apresentados na próxima seção, irão interferir no estado da aplicação, removendo da lista todos os itens que não se enquadram nos critérios escolhidos. Quando a remoção é concluída, o estado da aplicação é alterado, renderizando automaticamente os novos elementos que compõem os gráficos.

A interface de usuário do sistema CPS apresenta como tela inicial a área em que

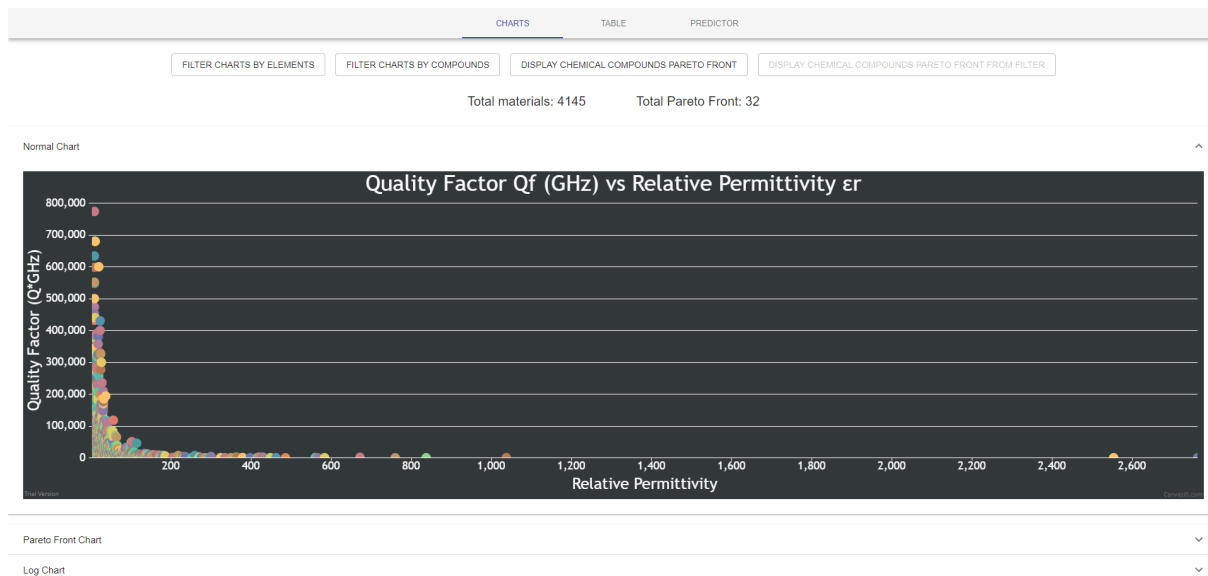


Figura 4.3: Tela inicial da interface de usuário.

serão realizadas as análises dos mais de quatro mil materiais cerâmicos. A Figura 4.3 exemplifica a tela de abertura do sistema, que possui o cabeçalho com a guia de abas para navegar entre os gráficos, a tabela e a seção para realizar previsões de novas fórmulas.

No corpo da aba dos Gráficos (*Charts*) existem quatro botões que permitem a exibição de menus laterais para visualizar informações ou filtrar os itens presentes nos gráficos listados abaixo. Os botões Filtrar Gráficos pelos Elementos (*Filter Charts by Elements*) e Filtrar Gráficos por Compostos (*Filter Charts by Compounds*), ao serem selecionados, abrem um menu flutuante à esquerda da página com as opções para alterar a composição dos gráficos. Os botões Exibir Compostos Químicos Frente de Pareto (*Display Chemical Compounds Pareto Front*) e Exibir Compostos Químicos Pareto na Frente do Filtro (*Display Chemical Compounds Pareto Front from Filter*), abrem um menu lateral flutuante à direita da página, contendo as informações detalhadas dos materiais que fazem parte da Frente de Pareto. O último botão só é habilitado quando um filtro é aplicado, gerando assim uma nova Frente de Pareto para ser analisada.

Logo abaixo dos botões, é exibida a informação com a quantidade Total de Materiais (*Total materiais*) utilizados em todos os gráficos e também o Total Frente de Pareto (*Total Pareto Front*). É importante ressaltar que alguns materiais da referência original foram duplicados por decisão de projeto, por apresentar mais de um valor para a Constante Dielétrica ou para o Fator de Qualidade. Por essa razão, são inicialmente apresentados 4145 materiais. Essa duplicidade não afeta o cálculo para a quantidade de materiais da Frente de Pareto, que exibe a contagem de 32 compostos para toda a base de dados. A página também conta com três gráficos, cada um contido em um painel de expansão próprio, dando liberdade para o usuário minimizar os gráficos quando desejar.

### 4.3.1.1 Tipos de Gráficos

O sistema CPS conta com três tipos de gráficos para a realização de análises. Os círculos coloridos correspondem a um material cerâmico, representados por um ponto entre os eixos X e Y. Os valores do eixo X são dados pela Constante Dielétrica (*Relative Permittivity*), enquanto o eixo Y pela propriedade Fator de Qualidade (*Quality Factor*).

Todos os gráficos possuem internamente a funcionalidade de ampliação de uma área desejada, um modo de navegação pós-ampliação e a opção de restaurar a visualização original completa. Ao mover o cursor sobre os círculos, uma janela é aberta exibindo as informações de identificador, material, Constante Dielétrica ( $\epsilon_r$ ), Fator de Qualidade ( $Qf$ ), lista de elementos e lista de compostos. Essas duas últimas propriedades são consideradas no momento da realização dos filtros, para incluir ou remover o material.

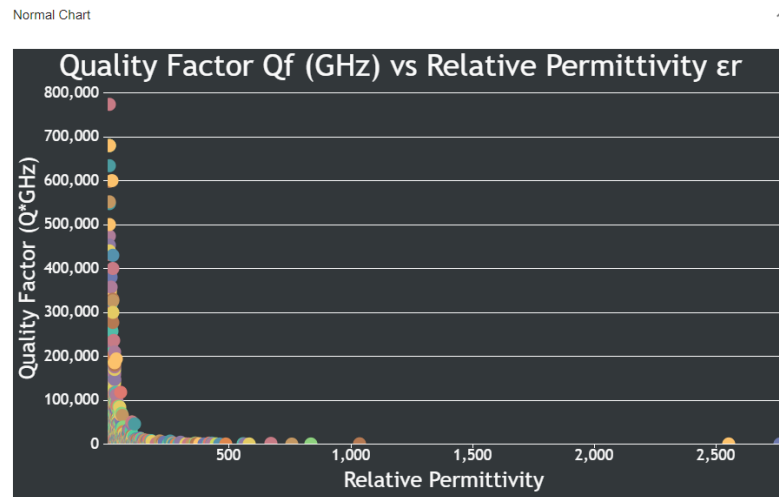
Os três tipos de gráficos, apresentados na Figura 4.4, são:

- **Gráfico Normal:** representa a exibição tradicional de Constante Dielétrica (eixo X) e Fator de Qualidade (eixo Y). A Figura 4.4a exibe o gráfico original, sem a realização de filtros, contendo a listagem dos 4145 itens. Os valores do eixo X variam de 3 até 2763, enquanto no eixo Y vão de 16 até 773700.
- **Gráfico Frente de Pareto:** assim como o Gráfico Normal, exibe os valores de Constante Dielétrica x Fator de Qualidade, porém com a inclusão da Frente de Pareto, criada com o apoio do pacote `npm pareto-frontier`<sup>2</sup>. A Figura 4.4b apresenta o modelo do gráfico, onde os itens presentes na fronteira de Pareto estão listados em vermelho e unidos por uma linha de mesma cor. Eles representam os materiais dominantes de todo o conjunto. Os demais itens estão listados na cor azul. Ao passar o cursor sobre o material, o gráfico também possui a funcionalidade extra de exibir a Distância Euclidiana<sup>3</sup> do ponto até a Frente de Pareto, considerando o ponto da Frente de Pareto mais próximo.
- **Gráfico Log:** neste gráfico, mostrado na Figura 4.4c, os valores dos eixos X e Y são alterados pela função logaritmo natural, gerando uma visualização dos pontos normalizada. Ao passar o cursor por cima do ponto escolhido, os valores originais de  $\epsilon_r$  e  $Qf$  são exibidos, assim como os valores após a aplicação da função logaritmo.

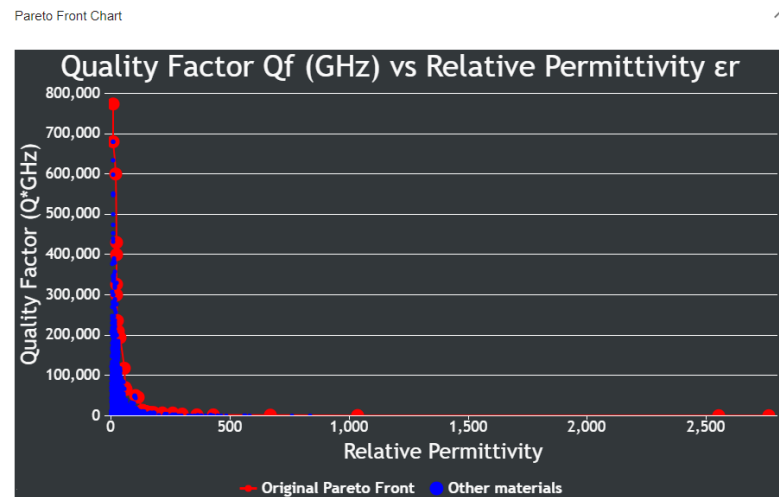
---

<sup>2</sup>Fonte: [www.npmjs.com/package/pareto-frontier](http://www.npmjs.com/package/pareto-frontier)

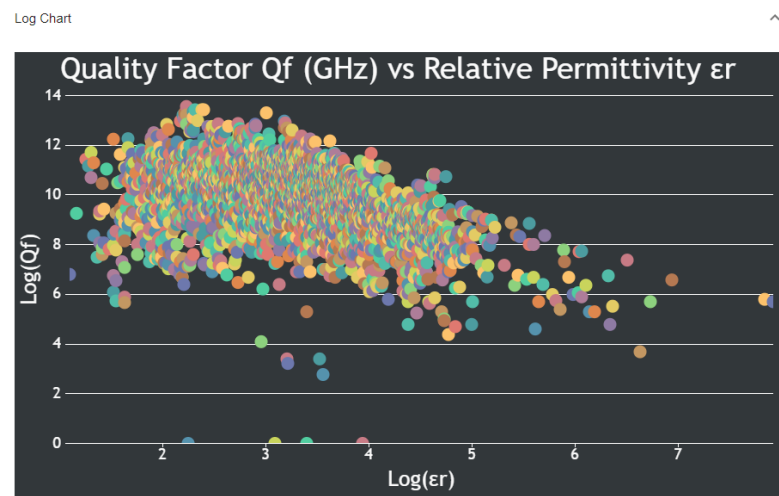
<sup>3</sup>Fonte: [www.npmjs.com/package/euclidean-distance](http://www.npmjs.com/package/euclidean-distance)



(a) Gráfico Normal (*Normal Chart*), presente na referência original.



(b) Gráfico Frente de Pareto (*Pareto Frontier Chart*), com os materiais da Frente de Pareto em vermelho e os demais na cor azul.



(c) Gráfico Log (*Log Chart*), normalizado pelo logaritmo natural.

Figura 4.4: Tipos de gráficos presentes na interface de usuário.

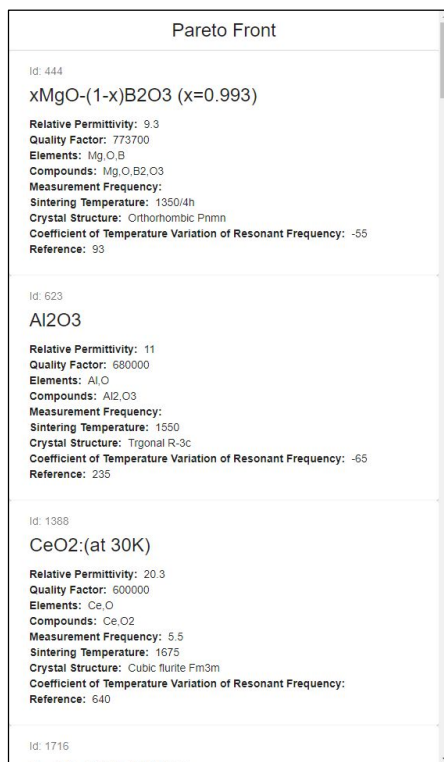
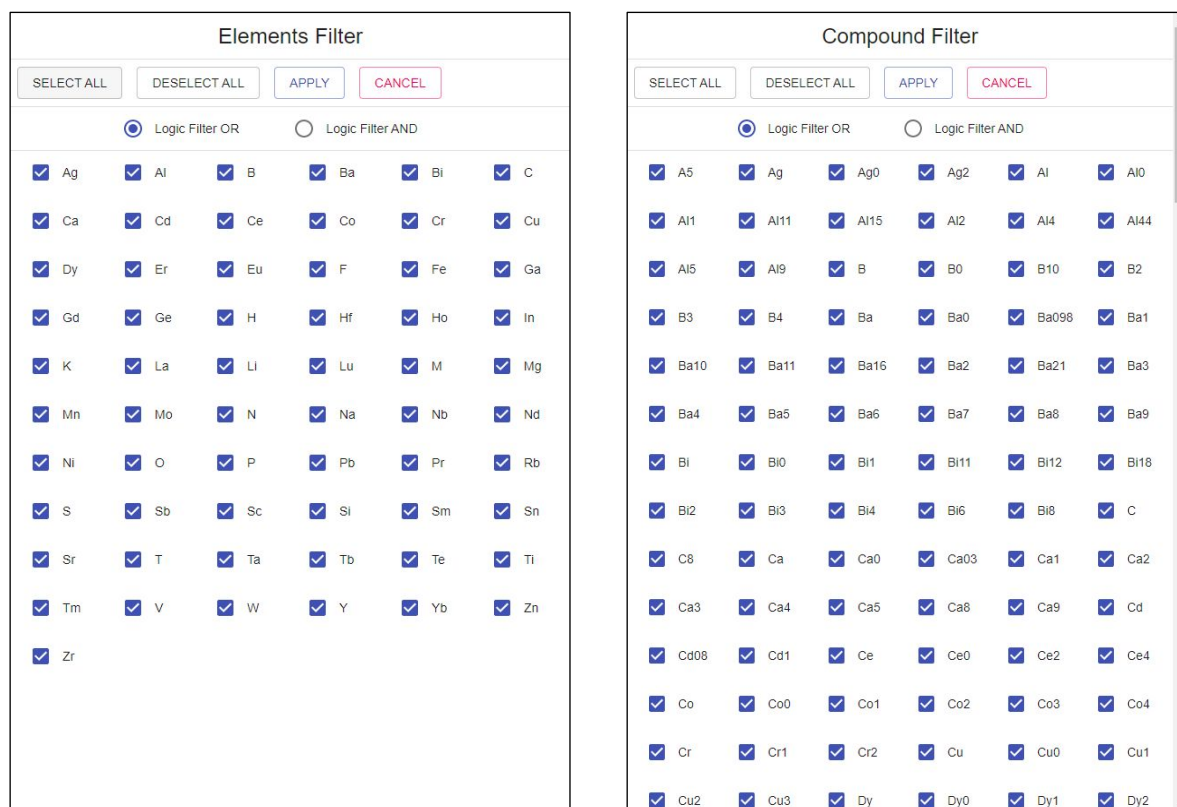


Figura 4.5: Menu contendo os detalhes da lista de materiais da Frente de Pareto.

### 4.3.2 Informações Frente de Pareto

A interface também disponibiliza dois botões para visualizar as informações dos materiais presentes na Frente de Pareto. Ao clicar no botão Exibir Compostos Químicos Frente de Pareto, será aberto um menu lateral na parte direita da interface, cujo conteúdo das informações pode ser visto na Figura 4.5. Os 32 materiais da Frente de Pareto original sempre serão mostrados nessa listagem, que possui uma barra de rolagem para visualizar os demais itens. A listagem está ordenada pelo Id (identificador), que o material aparece na referência de origem, mesma ordem da menor para a maior Constante Dielétrica.

Além do Id, o menu também apresenta a fórmula química em destaque. Todas as informações do material são listadas, sendo as mesmas presentes na tabela de referência. A lista de informações conta com a Constante Dielétrica, Fator de Qualidade, lista de elementos e compostos na fórmula química, Frequência de Medição, Temperatura de Sinterização, Estrutura Cristalina, Coeficiente de Variação de Temperatura da Frequência Ressonante e o número de referência do estudo de origem.



(a) Filtro com os 58 elementos disponíveis.

(b) Filtro de todos os compostos.

Figura 4.6: Tipos de filtros disponibilizados para alterar o conteúdo dos gráficos.

### 4.3.3 Filtros

O sistema conta com dois tipos de filtros para selecionar os materiais desejados na visualização gráfica, exibidos em um menu lateral esquerdo ao clicar nos botões do cabeçalho da página. O primeiro deles é o Filtro por Elementos (*Elements Filter*), presente na Figura 4.6a, que apresenta as opções de seleção para os 58 elementos químicos presentes em todas as fórmulas estudadas. Já o segundo filtro pode ser visto na Figura 4.6b, Filtro por Compostos (*Compound Filter*) que representa uma lista contendo todos os elementos químicos e suas respectivas quantidades. Por exemplo, para o oxigênio temos opções como O0, O, O2, O3, entre outras. Fórmulas que apresentam a quantidade exata, serão listadas via filtro. A opção de elemento com o valor 0, indica um elemento cuja quantidade na fórmula química é inferior a 1, podendo ser, por exemplo, a fração 0,2. Não há distinção dessas frações menores, por decisão de projeto.

No cabeçalho de ambos os filtros há uma lista com quatro botões de ação. Ao Selecionar Todos (*Select All*), automaticamente todos os itens ficarão com um tique azul. Deselecionar Todos (*Deselect All*) fará com que todas as caixas de seleção fiquem em branco. Ao clicar em Aplicar (*Apply*) o filtro será executado alterando o conteúdo de

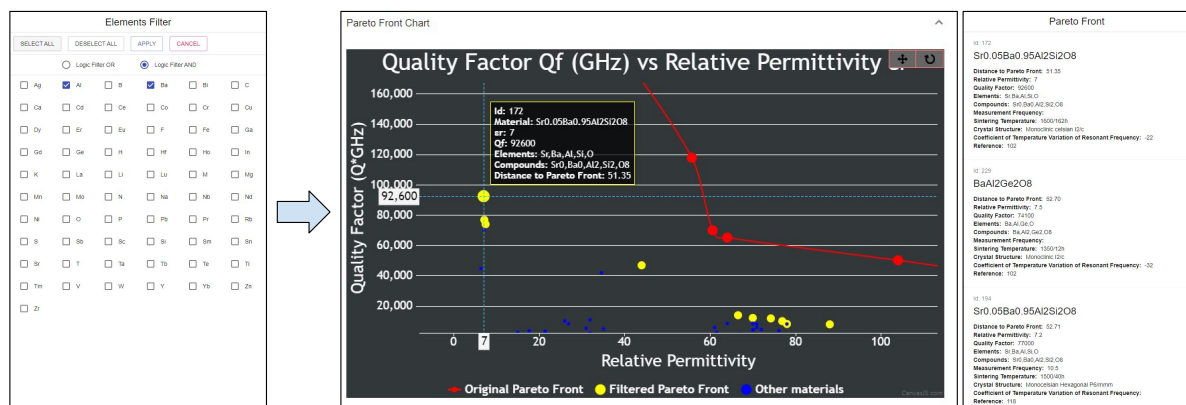


Figura 4.7: Materiais filtrados pelos elementos Al e Ba, via Lógica "E" e utilizando o Filtro por Elementos.

todos os três gráficos. Ao clicar em Cancelar (*Cancel*), o menu será fechado e nenhum filtro será aplicado.

Os dois tipos de filtro também contam com duas lógicas para realizar a operação de seleção de materiais. A primeira lógica é do Filtro Lógico OU (*Logic Filter OR*), que opera da seguinte maneira: caso sejam escolhidos, considerando o Filtro por Elementos, o alumínio (Al) ou bário (Ba), qualquer fórmula química contendo pelo menos um dos dois, ou os dois juntos, será listada. O filtro inicial da plataforma ocorre ao selecionar todos os elementos e aplicar a Lógica OU, listando dessa forma todos os materiais.

A segunda lógica pode ser escolhida selecionando o Filtro Lógico E (*Logic Filter AND*), que funciona conforme o exemplo a seguir: escolhendo novamente alumínio (Al) e bário (Ba) no Filtro de Elementos, serão somente listados os materiais que possuem os dois elementos juntos na fórmula química. O exemplo da aplicação deste último filtro, alterando um dos três gráficos e gerando uma nova Frente de Pareto pode ser visto na Figura 4.7.

Qualquer filtro aplicado na plataforma afetará os três gráficos presentes na seção de análise gráfica e também as informações de Total de Materiais e Total Frente de Pareto. A Figura 4.7 apresenta o funcionamento do filtro, com o exemplo do gráfico Frente de Pareto. Os materiais da Frente de Pareto original continuarão presentes neste gráfico, diferentemente dos gráficos Normal e Log, que apresentarão somente os materiais filtrados. Outra particularidade do gráfico Frente de Pareto é a exibição em amarelo da nova Frente de Pareto filtrada. Ao posicionar o cursor sobre o material amarelo, a Distância Euclidiana até a Frente de Pareto Original também é exibida. A Figura 4.7 também mostra o menu da Frente de Pareto filtrada, exibido ao clicar no botão "Exibir Compostos Químicos Pareto na Frente do Filtro". Essa listagem apresenta os detalhes de todos os itens amarelos do gráfico, incluindo a Distância Euclidiana de cada um.

CHARTS TABLE PREDICTOR										
id	Material	ST (-C)	Crystal structure	ir	Or (GHz)	to	rf	Reference	Elements	Compounds
1	AlPO4+5 wt% MgF2	1450	Orthorhombic	3	900	5.8	-	1	Al, P, O, Mg, F	Al, P, O4, Mg, F2
2	50%Li2CO3-40.24%B2O3-9.76%SiO2	600	glass	3.2	10500		-79	2	Li, C, O, B, Si	Li2, C, O3, B2, O3, Si, O2
3	SiO2 sol-gel	1550	Tetragonal cristobalite	3.5	92400		-15	3	Si, O	Si, O2
4	0.2B2O3-0.8SiO2	1100	Amorphous (fused silica)	3.6	70600			4	B, O, Si	B2, O3, Si, O2
5	SiO2 (solid state method)	1100	fused silica	3.7	44300		-15	5	Si, O	Si, O2
6	SiO2melt method		Fused silica	3.7	122100			5	Si, O	Si, O2
7	10.5CaO-22.2B2O3-67.3SiO2	1035	Glass	3.8	4300	12.9		6	Ca, O, B, Si	Ca, O, B2, O3, Si, O2
8	SiO2	1650	Trigonal P3121	3.8	80400		-16	7	Si, O	Si, O2
9	CaO-B2O3-SiO2 (29.3:9.3:61.4 mol%)	900	Glass	3.9	1800	9.9		8	Ca, O, B, Si	Ca, O, B2, O3, Si, O2
10	27.8CaO-27B2O3-45.2SiO2	875	Glass	4	3200	12.8		8	Ca, O, B, Si	Ca, O, B2, O3, Si, O2
11	KGaGe3O8	890	Monoclinic P21/a	4	10200	13		9	K, Ga, Ge, O	K, Ga, Ge3, O8
12	SrCuSi4O10	1100	Gillespie Tetragonal P4/ncc	4	11500			10	Sr, Cu, Si, O	Sr, Cu, Si4, O10
13	CaO-B2O3-SiO2 (19.8:30.9:49.3 mol%)	900	Glass	4.1	2000	9.9		8	Ca, O, B, Si	Ca, O, B2, O3, Si, O2
14	CaO-B2O3-SiO2 (10.5:22.2:67.3 mol%)	900	Glass	4.1	2600	9.9		8	Ca, O, B, Si	Ca, O, B2, O3, Si, O2
15	Na2MoO4	660	Cubic Fd-3m	4.1	35900		-76	11	Na, Mo, O	Na2, Mo, O4
16	Li3AlB2O6	650	Triclinic	4.2	12460	16.8	-290	12	Li, Al, B, O	Li3, Al, B2, O6
17	27.8CaO-27B2O3-45.2SiO2	850	Glass	4.2	3200	12.8		6	Ca, O, B, Si	Ca, O, B2, O3, Si, O2
18	LiAlSi4O14+1% wt% RfO2	onn	alfa-eucryptite	4.3	62400		-16	13	Li, Al, Si, O	Li, Al, Si, O4

Figura 4.8: Tabela original contendo todos os materiais presentes na interface.

#### 4.3.4 Tabela de Materiais Cerâmicos

A segunda aba da interface do sistema CPS (*Table*) contém a tabela original de referência, para a criação da plataforma e do previsor. A Figura 4.8 apresenta as primeiras linhas da tabela, que contém 4145 instâncias. Essa área de análise facilita a visualização das informações de cada material, possibilitando rolar a página para encontrar qualquer item da listagem e utilizar as ferramentas de busca do próprio navegador para encontrar valores ou palavras-chave rapidamente. A tabela conta com as nove colunas de informações mencionadas anteriormente, além das informações de elementos e compostos das fórmulas, mesmos dados exibidos nos menus laterais da Frente de Pareto.

#### 4.3.5 Módulo Previsor

O módulo previsor possui uma interface simplificada com componentes para abertura de novas janelas para inserção de dados. O estado da aplicação guarda todos os elementos disponíveis para criação de uma nova fórmula química, considerando os 58 elementos presentes na referência base. Ao adicionar um elemento, o estado da aplicação é atualizado na lista de elementos que podem ser enviados para análise. A construção da comunicação entre as plataformas segue os conceitos de REST API apresentados no

capítulo de Fundamentos. Um JSON com os 58 elementos é criado, com os valores escolhidos pelos usuários e deixando como zero os demais elementos. O envio é feito via HTTP Request para o microsserviço com o modelo treinado do AutoGluon, que responde com um JSON contendo a Constante Dielétrica e Fator de Qualidade.

A aplicação guarda esses valores no estado da aplicação, exibindo a nova fórmula criada na lista de compostos e também insere um novo ponto em destaque, representando o composto, no gráfico de Frente de Pareto logo abaixo na mesma tela. A Distância Euclidiana também é apresentada e calculada automaticamente com o pacote npm. Todo o cálculo ocorre a partir da modificação do estado do componente pai, sendo a página do previsor, com relação aos filhos Compostos Criados e Gráfico de Resultados. Ambos aguardam a resposta do modelo previsor, cuja alteração de estado ocorrerá a partir do retorno do microsserviço que calcula o resultado.

#### 4.3.5.1 Microsserviço

O módulo previsor da interface de usuário conecta-se a um microsserviço, que carrega os modelos preditivos gerados pelo AutoGluon e aguarda uma chamada de dados para realizar os cálculos necessários. A conexão entre interface e microsserviço é feita via chamada do tipo POST, disponibilizada pela API da aplicação desenvolvida. O conteúdo da chamada possui uma estrutura de dados JSON com os 58 elementos químicos e suas respectivas quantidades, preenchidas na própria interface. Todos os valores não preenchidos pelo usuário são completados com o valor zero para o algoritmo preditivo poder realizar o cálculo. O AutoGluon oferece o método *predict()* para os modelos treinados, que ao passar o conjunto de dados para as 58 colunas como parâmetro, retorna o valor calculado previsto.

Foi utilizado o pacote Flask<sup>4</sup> em Python, biblioteca que facilita o desenvolvimento da aplicação para comunicar via REST API. O método que recebe o conteúdo do POST está implementado para carregar os 2 modelos preditivos AutoGluon e realizar o cálculo para cada um deles em sequência. Os valores calculados são adicionados em um JSON de retorno e enviados para a aplicação de origem da solicitação.

As ferramentas disponibilizadas pelo Docker<sup>5</sup> são responsáveis por manter o microsserviço como uma aplicação independente, executando em um ambiente de sistema operacional virtual Ubuntu. Uma porta de rede é disponibilizada para que a chamada seja realizada para o Docker, que então redireciona o caminho para a aplicação Flask

<sup>4</sup>Documentação: [www.flask.palletsprojects.com](http://www.flask.palletsprojects.com)

<sup>5</sup>Documentação: [www.docs.docker.com](http://www.docs.docker.com)

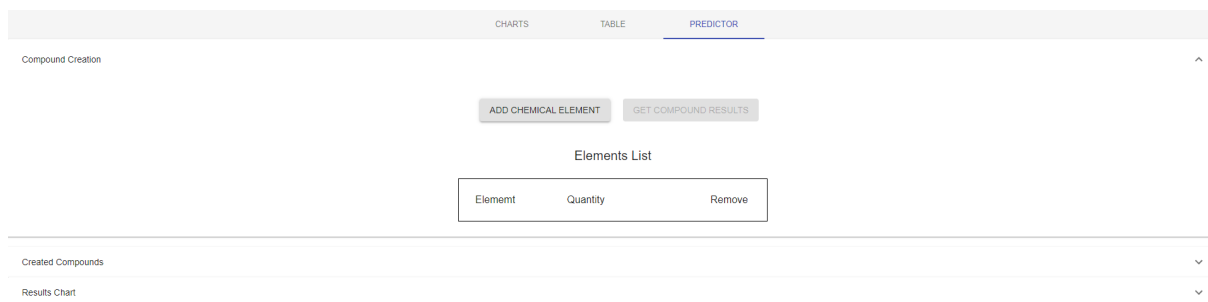


Figura 4.9: Módulo previsor que permite a criação de uma nova fórmula química e submissão para análise.

realizar as operações de cálculo preditivo. As configurações realizadas via Docker instalam os pacotes do AutoGluon, Python, Flask e o sistema operacional Ubuntu de forma automática, fazendo com que a aplicação execute em tempo real.

#### 4.3.5.2 Utilização do previsor via interface

A terceira e última aba do cabeçalho (*Predictor*) leva à parte da interface responsável pela criação de novas fórmulas químicas. A Figura 4.9 mostra a tela apresentada ao usuário da aplicação, ao acessar essa aba. Na área chamada de Criação de Compostos (*Compound Creation*) existem dois botões de ação. O primeiro chamado de Adicionar elemento químico (*Add chemical element*), ao ser selecionado, exibe uma janela para inserção de um dos 58 elementos químicos disponíveis e sua quantidade na fórmula que será criada. O segundo botão Obter resultados do composto (*Get compound results*) envia a fórmula química criada para o microsserviço previsor, a fim de obter os valores para Constante Dielétrica e Fator de Qualidade. Este botão só é habilitado quando pelo menos dois elementos químicos distintos são adicionados. As áreas de Compostos Criados (*Created Compounds*) e Gráfico de Resultados (*Results Chart*) aparecem inicialmente ocultas e para exibi-las basta clicar no painel de expansão de cada uma. Ambas as áreas apresentarão conteúdos relevantes quando algum material for submetido para análise do previsor.

Ao adicionar os elementos, a área de Lista de Elementos (*Element List*) é preenchida, conforme pode ser observado na Figura 4.10. Na imagem temos a criação da fórmula química  $\text{Ba}_4\text{Sm}_{8.33}\text{EuTi}_{18}\text{O}_{54}$ , em que elemento a elemento é adicionado com seu respectivo valor. Caso seja necessário excluir algum elemento, basta clicar no ícone da lixeira presente na mesma linha. É importante ressaltar que ao escolher o valor de algum elemento químico, a mesma regra de dopagem, explicada anteriormente, deve ser aplicada para obter os melhores resultados.

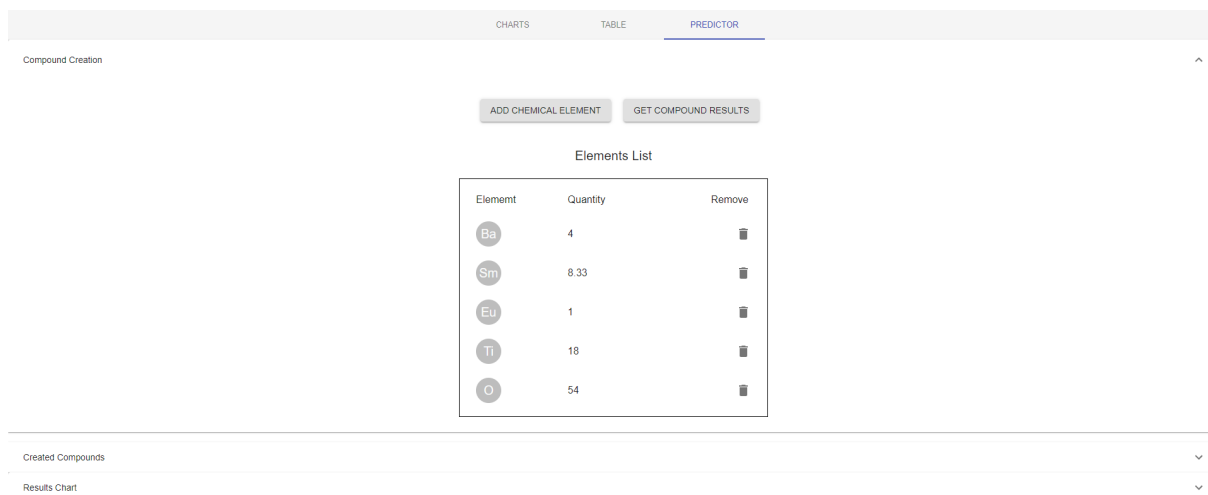


Figura 4.10: Fórmula química inserida na lista de elementos químicos e pronta para ser enviada ao previsor.

Após enviar um elemento para análise e aguardar o cálculo de previsão, o CPS armazena os valores no estado da aplicação. Esses valores então são exibidos a lista de Compostos Criados e também no Gráfico de Resultados. A Figura 4.11 demonstra como os resultados de duas fórmulas químicas novas aparecem na ferramenta. Temos o primeiro composto dado por  $\text{Ba}_4\text{Sm}_{8.33}\text{EuTi}_{18}\text{O}_{54}$ , cujos valores calculados foram 79,85 para a Constante Dielétrica e 9543 para o Fator de Qualidade. Esse é um material real, presente na base de dados com identificador 3673. Os valores reais para este material são 78,7 para Constante Dielétrica e aproximadamente 9560 para o Fator de Qualidade. Como pode ser visto na Figura 4.12, o previsor foi capaz de se aproximar bastante dos valores pretendidos, considerando o valor real como o ponto laranja e o novo valor na cor verde água.

O outro material criado foi  $\text{Al}_2\text{Ti}_2\text{O}_4$ , representando uma instância inteiramente nova para o previsor. Os valores apresentados foram de 18,78 para a Constante Dielétrica e 93048 para o Fator de Qualidade. Os dois materiais permanecem listados na tabela de Compostos Criados, com a fórmula química e os valores previstos ao lado. O Gráfico de Resultados é do tipo Frente de Pareto, para avaliar se os novos materiais se aproximam ou até mesmo podem romper a Fronteira de Pareto. A cor verde água representa os novos materiais neste gráfico, que ao passar o cursor por cima, exibe os detalhes do item e também a Distância Euclidiana calculada.

No capítulo seguinte, serão discutidos os modelos previsores criados neste trabalho. Serão apresentadas as técnicas e ferramentas exploradas para encontrar os modelos mais descritivos para os conjuntos de dados utilizados. Os resultados dos modelos também serão apresentados e discutidos, tanto para a previsão da Constante Dielétrica como para o Fator de Qualidade.

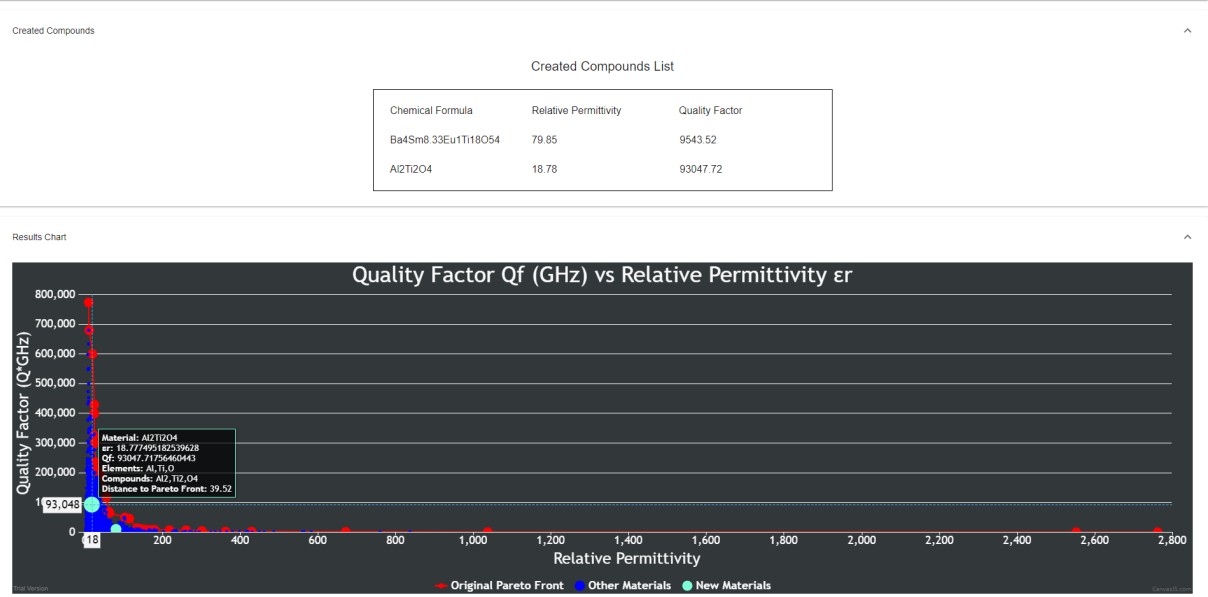


Figura 4.11: Resultados de duas fórmulas químicas enviadas para análise.

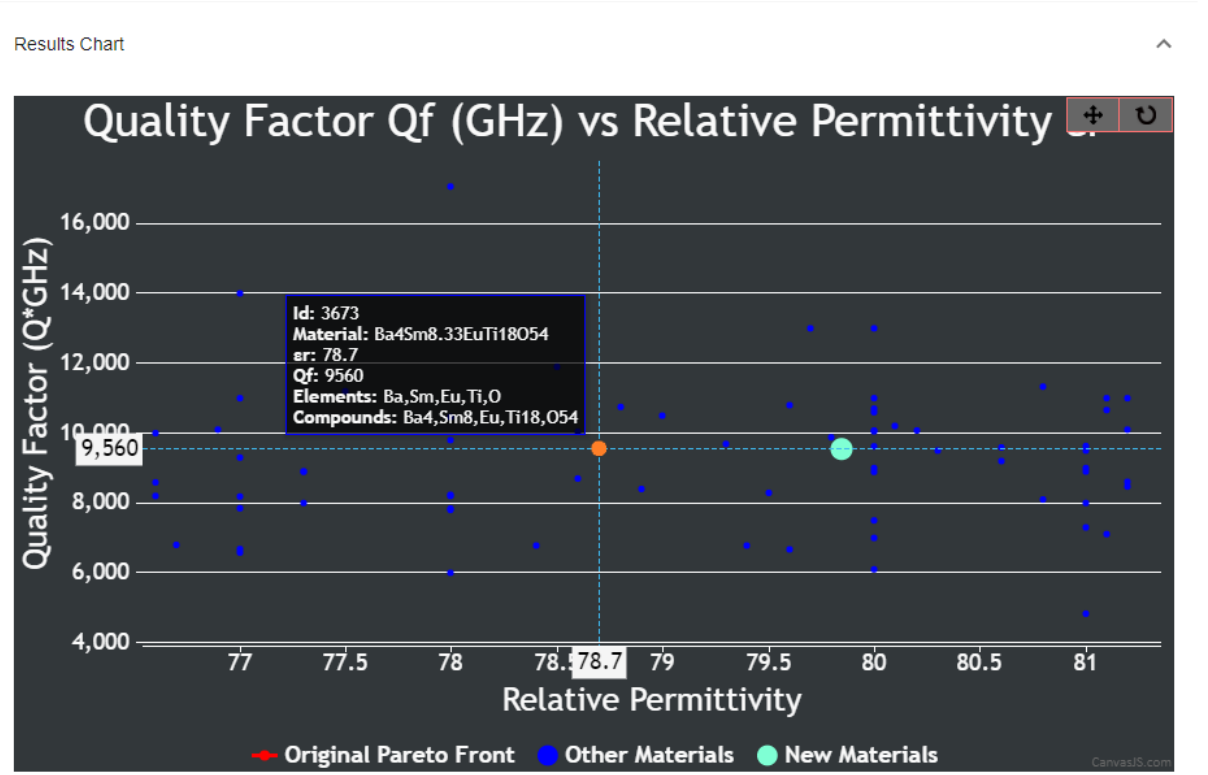


Figura 4.12: Comparação entre o material real e o material previsto pela ferramenta.

## Capítulo 5

# Previsão de Propriedades de Materiais Cerâmicos

Neste capítulo o conjunto de técnicas utilizadas para criação dos modelos previsores, utilizados no sistema CPS, serão apresentados. Primeiramente é mostrado os tipos de bases de dados de modelos preditivos analisados pelas técnicas de Aprendizado de Máquina. Os modelos estudados são apresentados, com ênfase na ferramenta AutoGluon, que deu origem aos modelos preditivos utilizados. Os resultados e análises de erros são discutidos e apresentados neste capítulo.

### 5.1 Estruturas de Bases de Dados para modelos preditivos

A partir do cálculo individual das fórmulas químicas, duas abordagens foram concebidas para servir de base tabular aplicada nos algoritmos e técnicas computacionais preditivas utilizadas neste trabalho. A primeira delas consistiu na separação dos elementos químicos presentes das fórmulas, considerando suas respectivas quantidades. No total, 58 elementos químicos distintos foram encontrados e passaram a representar as colunas dessa primeira tabela. Duas colunas foram adicionadas para representar, respectivamente, a Constante Dielétrica e o Fator de Qualidade, alvos das previsões realizadas. Caso a fórmula química não possua o elemento de uma determinada coluna, esta é preenchida com o valor zero, conforme mostrado na Tabela [5.1a](#).

A segunda abordagem foi concebida visando diminuir a quantidade de entradas para treinamento dos algoritmos de Aprendizado de Máquina, uma vez que possuem alto custo computacional. O modelo criado considerou o tamanho da maior fórmula química existente na base, que possui 11 elementos químicos. A estrutura criada considerou 11 pares de número atômico do elemento químico e sua respectiva quantidade na fórmula,

Tabela 5.1: Diferença entre bases de dados de 58 e 22 colunas considerando a fórmula  $\text{Al}_2\text{O}_3$ .

(a) Modelo de dados para 58 colunas de entrada.

Al	O	Ti	B	Nb	Ag	La	Fe	Co	Hf	...	Constante Dielétrica	Fator de Qualidade
2.0	3.0	0	0	0	0	0	0	0	0	...	10.0	634000

(b) Modelo de dados para 22 colunas de entrada. A letra E representa o Elemento, enquanto V o valor.

E1	V1	E2	V2	E3	V3	...	E11	V11	Constante Dielétrica	Fator de Qualidade
8.0	3.0	13.0	2.0	0	0	...	0	0	10.0	634000

totalizando 22 colunas que representam as entradas de treinamento para as técnicas utilizadas. Mais duas colunas são adicionadas ao final para representar os alvos, ou seja, a Constante Dielétrica e o Fator de Qualidade. Cada linha da tabela foi ordenada, da esquerda para a direita, do menor para o maior número atômico do elemento químico presente na fórmula, cujos valores estão presentes nas colunas ímpares. As colunas pares acompanham o número atômico presente na coluna imediatamente anterior, mantendo a ordenação dos 11 pares de número atômico e quantidade. Nos casos em que as fórmulas químicas não possuem 11 elementos, as demais colunas são preenchidas com o valor zero. A Tabela 5.1b exemplifica a diferença das duas abordagens.

Na Tabela 5.1 temos a primeira base com 58 colunas de entrada de dados, representando todos os elementos químicos presentes na referência. Para a fórmula  $\text{Al}_2\text{O}_3$  apenas duas colunas são preenchidas com o valor diferente de zero presente na fórmula química, considerando o elemento correspondente. As demais colunas possuem o valor 0, representando a não utilização do elemento na fórmula. Já na segunda abordagem, a fórmula  $\text{Al}_2\text{O}_3$  é alocada nas colunas seguindo a ordenação do menor para o maior número atômico dos elementos químicos e cada número atômico é seguido por seu respectivo valor presente na fórmula. Para o oxigênio o número atômico é 8 e o valor 3. Para Al o número atômico é 13 e o valor 2. As demais colunas são preenchidas com o valor 0, até que os 11 pares estejam preenchidos. Tanto no modelo de 58 entradas, como no de 22, as duas últimas colunas são representadas pela Constante Dielétrica e Fator de Qualidade. Essas colunas são consideradas alvos preditivos, portanto não fazem parte do conjunto de dados considerados entradas, ou características (*features*), para o treinamento.

Outras características da base de dados original como a temperatura de síntese, estrutura cristalina, variação da temperatura de frequência ressonante e frequência de medição não foram utilizados nesta versão do trabalho. Em muitos casos, essas colunas não possuem valores definidos, o que poderia comprometer a modelagem em caso de aproximações ou utilização de valores nulos para análise. A base de dados para mo-

delos preditivos final foi estruturada a partir dos elementos químicos e suas respectivas quantidades, além das duas propriedades físicas de interesse.

## 5.2 Redes Neurais

Redes Neurais Artificiais [45] são um dos algoritmos de Aprendizado de Máquina. Seu nome e estrutura são inspirados no cérebro humano, imitando como os neurônios biológicos se comunicam entre si. Esse tipo de modelo é conhecido por gerar bons resultados preditivos, para uma série de aplicações distintas [46].

As Redes Neurais Perceptron são compostas por camadas de nós, contendo uma camada de entrada, uma ou mais camadas ocultas (multicamadas) e uma camada de saída. Cada nó, ou neurônio artificial, se conecta aos outros nós das camadas vizinhas com um peso associado. A Figura 5.1 exemplifica uma arquitetura de rede neural com uma quantidade de entradas "E" igual a "n", duas camadas ocultas com "n" neurônios cada e duas saídas "S1" e "S2". Apesar do exemplo com "n" neurônios em cada camada oculta, esse número de neurônios pode variar.

Elas dependem de dados de treinamento para aprender e melhorar sua precisão ao longo do tempo. O processo de treinamento por iteração consiste em computar os valores ideais dos pesos de cada conexão entre os neurônios da rede. Esses valores fazem parte de uma função que calcula o valor de cada saída da rede. No fim de cada iteração, o valor calculado para a saída é comparado com o valor real, o erro é avaliado e ocorre uma propagação de correções para os valores dos pesos das conexões. A cada repetição os valores são reajustados até que o algoritmo atinja um critério de parada, que pode ser, por exemplo, um valor de erro de alvo, um número de épocas (iterações), entre outros.

Neste trabalho, foi desenvolvido um algoritmo de uma rede neural para realizar o cálculo de regressão e obter os valores de Constante Dielétrica e Fator de Qualidade. Foi utilizado o pacote de código aberto Tensorflow <sup>1</sup> em Python, com a técnica Validação Cruzada *k-fold*, aplicando as bases com 22 e 58 entradas. A Rede Neural serviu como base para análise das outras técnicas de Aprendizado de Máquina utilizadas posteriormente neste trabalho.

A arquitetura da rede neural deste trabalho consiste nos nós de entrada, podendo ser 58 nós ou 22 nós, uma camada interna variável em quantidade de neurônios e uma saída para calcular a Constante Dielétrica ou o Fator de Qualidade. Foi utilizada a técnica de validação cruzada, explicada na Subseção 5.3.2, de 10 *folds* (partições) e 2500 iterações,

---

<sup>1</sup>Página: [www.tensorflow.org](http://www.tensorflow.org)

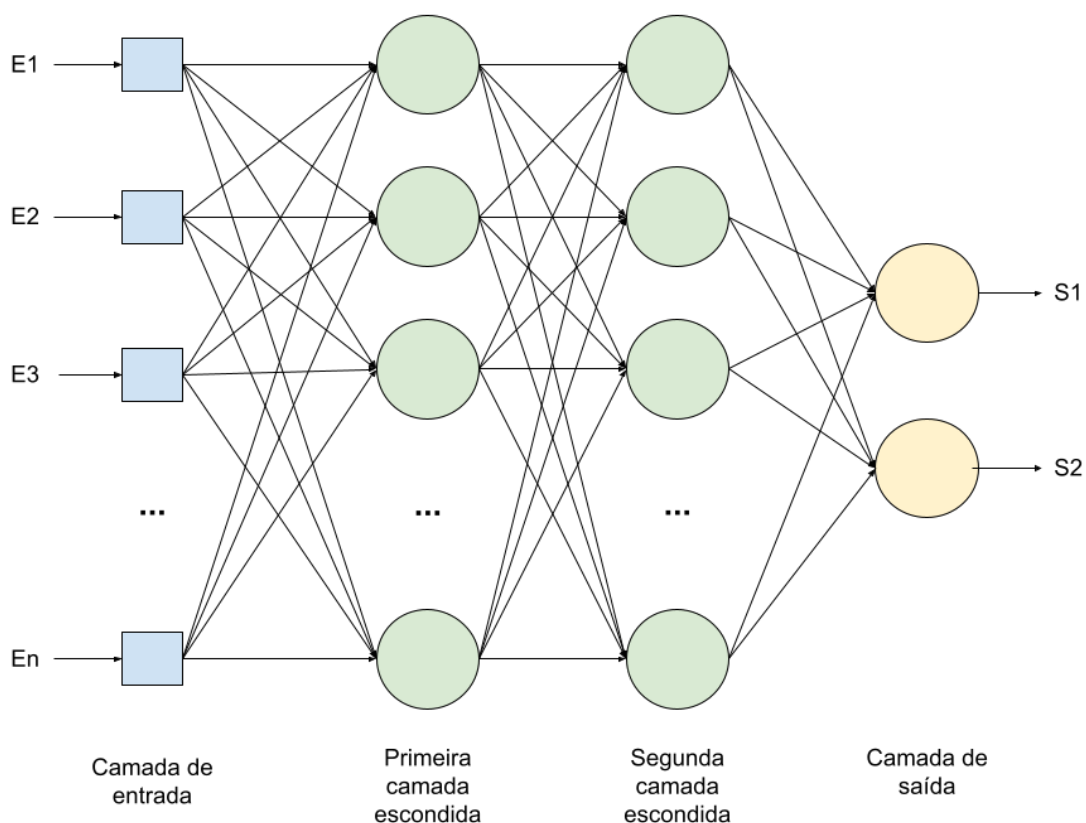


Figura 5.1: Exemplo de arquitetura de rede neural artificial.

ou épocas, por operação. Em cada caso, o número de neurônios da camada interna foi alternado de 1 a 150 para encontrar a melhor arquitetura.

A Rede Neural com uma camada serviu como base de comparação para as demais técnicas utilizadas nos experimentos empíricos, incluindo o modelo final adotado através do AutoGluon. Os seguintes algoritmos também foram explorados de forma empírica, porém sem grande profundidade:

- Rede Neural Artificial com uma camada oculta;
- XGBoost [47];
- XGBoost Otimização Bayesiana [48];
- Auto-keras [49];
- TPOT [50];
- Auto-sklearn [51];
- H2O [52].

Tabela 5.2: Valores das métricas de avaliação de performance, para os modelos de Rede Neural treinados com a base de 58 entradas.

	Constante Dielétrica	Fator de Qualidade
<b>MAE</b>	19,42	26689,38
<b>RMSE</b>	69,30	46696,25
<b>R2</b>	0,15	0,31

Foram realizadas uma série de experimentos variando a quantidade de neurônios da camada oculta de 1 até 150. Para cada propriedade de saída foi treinada uma rede neural própria. Tanto para o Fator de Qualidade, quanto para a Constante Dielétrica, a base organizada em 58 entradas e demonstrou melhores resultados. Os valores de MAE, RMSE e R2 foram significativamente menores aplicando o modelo que considera um elemento químico por coluna. A Tabela 5.2 exibe os melhores resultados encontrados pelos modelos de Rede Neural testados.

É possível observar a partir do valor de R2 que os modelos gerados não foram descritivos, tanto para o Fator de Qualidade com 0,31, quanto para a Constante Dielétrica com 0,15. Para avaliação do MAE e RMSE serão discutidos os resultados da comparação entre diferentes modelos na Tabela 5.3, em que praticamente todos os outros modelos avaliados tiveram valores menores para essas duas métricas e, conseqüentemente, apresentaram melhores resultados.

Também foi realizado um experimento variando os neurônios da camada interna de 1 até 150, com o passo a cada 15 neurônios, para calcular as duas propriedades juntas e considerando os dois tipos de base de dados. Os resultados mostraram valores elevados de MAE e RMSE, indicando que a Rede Neural de uma camada não é uma boa abordagem para prever as duas propriedades juntas. Além disso, os resultados foram piores do que o cálculo por modelo individual, ou seja, um modelo distinto por propriedade alvo desejada.

A Rede Neural não demonstrou bons resultados para servir de modelo preditivo para ambas as propriedades. Apesar disso, os resultados dos experimentos empíricos serviram como base de avaliação para os demais modelos testados, inclusive o modelo final gerado pelo AutoGluon.

## 5.3 AutoGluon

O Aprendizado de Máquina avançou significativamente na última década, se tornando cada vez mais presente em aplicações do cotidiano. Isso levou ao desenvolvimento de novas arquiteturas, modelagem e técnicas interessantes para dimensionar previsores

para grandes conjuntos de dados. No entanto, à medida que as técnicas de última geração crescem em sofisticação, é cada vez mais difícil, mesmo para um especialista, incorporar todas as práticas recomendadas recentes em sua modelagem.

As ferramentas de AutoML oferecem uma alternativa para tratar o problema. Para pessoas iniciantes nas práticas de Aprendizado de Máquina, eles removem muitas das barreiras de implantação de modelos de alto desempenho. Para o especialista, eles oferecem o potencial de implementar as melhores práticas de uma só vez. Essas práticas incluem estratégias para seleção do melhor tipo de modelo, combinação entre modelos (*ensemble*), ajuste de hiperparâmetros, engenharia de características (*features*), pré-processamento de dados, divisão de dados, entre outros. Essa automatização, com a inclusão de melhores práticas, permite a criação de modelos sem a necessidade de intervenção manual frequente, mesmo por especialistas.

O AutoGluon [27] é uma biblioteca Python fácil de usar e altamente precisa para AutoML com dados tabulares. Ele reconhece automaticamente o tipo de dados em cada coluna, para um pré-processamento robusto de dados, incluindo tratamento especial de campos de texto. Ele se adapta a vários modelos, desde árvores otimizadas prontas para uso até modelos de rede neural personalizados. Esses modelos são agrupados através do empilhamento em múltiplas camadas. São treinados de maneira a garantir que os dados brutos possam ser traduzidos em previsões de alta qualidade, considerando uma restrição de tempo. O sobreajuste (*overfitting*) é mitigado ao longo desse processo, dividindo os dados de várias maneiras, com rastreamento de pontos fora da curva (*outliers*).

### 5.3.1 Utilização

Como pode ser visto pelo Algoritmo 1, o AutoGluon é uma biblioteca simples de utilizar para obter um modelo preditivo tabular. Primeiramente é escolhida a coluna da base de dados tabular de alvo para previsão, no exemplo foi selecionada a coluna que representa os valores de Constante Dielétrica para os materiais (linha 2). Em seguida, selecionam-se as colunas que não fazem parte do conjunto de treinamento para previsão, no caso o Fator de Qualidade (linha 3).

O AutoGluon permite a escolha das métricas de erro para calcular a eficácia do modelo (linha 4), uma por vez, e utilizá-la como base de otimização. Neste trabalho foram utilizadas R2, RMSE e MAE. É possível escolher a quantidade de horas que o modelo executará, sendo possível que ele entre em um estado de convergência antes do tempo pré-determinado. Mesmo que esse estado seja atingido, o modelo continua sendo treinado até o fim do tempo escolhido, que no caso do exemplo é 24 horas (linha 5).

---

**Algoritmo 1:** Criação do modelo preditivo para Constante Dielétrica com AutoGluon.

---

**Entrada:** Base de Dados  $D$  com todos os materiais.

```

1 início
2   colunaAlvo = "ConstanteDieletrica"
3   colunaRemover = "FatorQualidade"
4   metrica = "r2" # Métrica de avaliação do erro
5   tempo = 86400 # 86400 segundos representa 24 horas
6   autoStack = verdadeiro # Parâmetro de otimização
7    $D' = \text{RemoveColuna}(D, \text{colunaRemover})$ 
8    $M = \text{CriarModelo}(D', \text{colunaAlvo}, \text{metrica}, \text{tempo}, \text{autoStack})$ 
9    $R = \text{ObterResultados}(M)$ 
10 fim

```

**Saída:** Resultados  $R$  com relatório completo e modelo  $M$  salvo no mesmo diretório de execução.

---

O parâmetro *autoStack* como verdadeiro (linha 6) é importante para obter o melhor modelo possível<sup>2</sup>, mesmo que gaste mais tempo para realizar todo o treinamento. O AutoGluon passa a utilizar a técnica de Validação Cruzada *k-fold* e Agrupamento de Pilhas Multicamadas (*Multi-Layer Stack Ensembling*) automaticamente para aumentar a precisão de do modelo.

A coluna não utilizada é removida (linha 7), transformando a base de dados original  $D$  em  $D'$ . Em seguida, o método para treinar o modelo é chamado (linha 8). Ele executará o método *fit()* da biblioteca, que treina os dados originais e gera o modelo final que será armazenado em  $M$ . A partir de  $M$ , é possível também obter o relatório completo de resultados,  $R$ , dos dados de treinamento e a listagem ordenada, por menor erro, dos melhores modelos encontrados (linha 9).

### 5.3.2 Validação Cruzada

A Validação Cruzada *k-fold* [53] consiste em dividir a base de dados conforme a quantidade  $k$  de partições (*folds*) desejadas. Supondo a divisão de 10 partes, um décimo é reservado para testar o modelo e as nove partes restantes utilizadas no treinamento. O processo é repetido para cada uma dessas frações, fazendo com que cada uma delas seja utilizada como teste uma vez e as restantes utilizadas para treinar o modelo. No final, é computada a média dos resultados para descobrir a eficácia final do modelo treinado. Esse método é conhecido por reduzir a variância dos resultados de previsão e também a

---

<sup>2</sup>Documentação: [www.auto.gluon.ai/tutorials/tabular\\_prediction/tabular-indepth.html](http://www.auto.gluon.ai/tutorials/tabular_prediction/tabular-indepth.html)

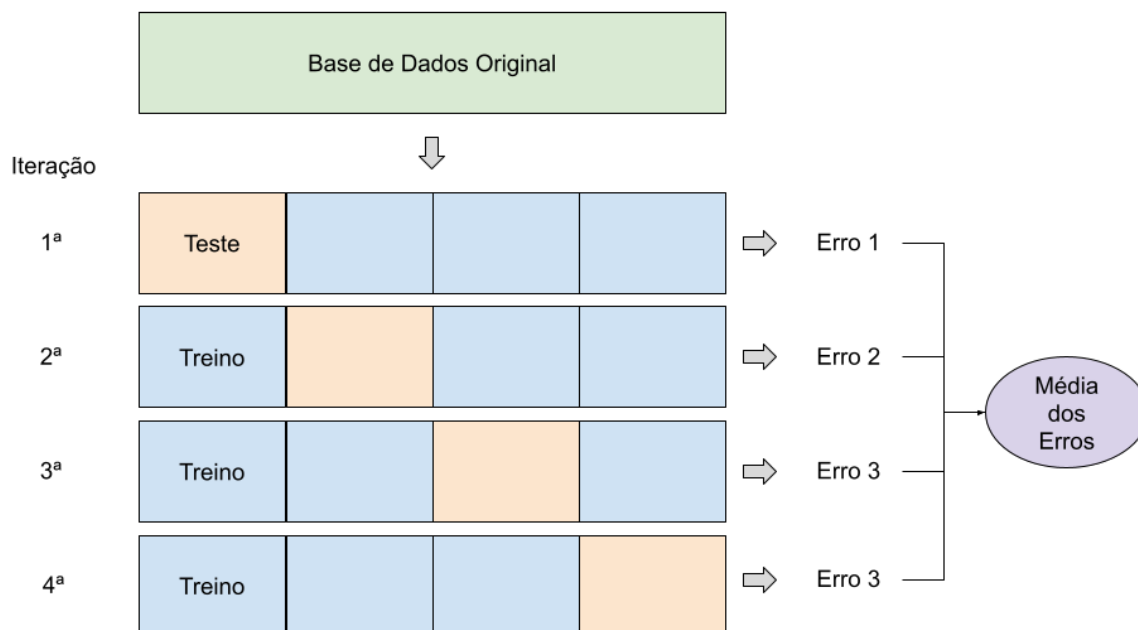


Figura 5.2: Validação cruzada com 4  *folds*.

possibilidade de um modelo sobreajustado (*overfitting*), que representará muito bem o conjunto de dados que serviram de treinamento, mas performará mal em aplicações reais por não ser generalizável.

Na Figura 5.2 há um exemplo do que acontece durante a validação cruzada [54]. O valor de  $k$  no exemplo é 4, representando o número de iterações e subdivisões da base de dados original. Para cada iteração, um modelo é gerado e o erro calculado. A métrica final é dada pela média dos erros de todas as iterações, considerando todos os modelos treinados.

### 5.3.3 Agrupamento de Pilha Multicamadas

O conceito de combinar (*ensemble*) a previsão de modelos individuais, a fim de gerar um único modelo composto, é uma técnica em Aprendizado de Máquina que vem apresentando grande eficácia [55]. O AutoGluon também utiliza essa abordagem de junção de múltiplos modelos, realizando o processo de Agrupamento de Pilha Multicamadas

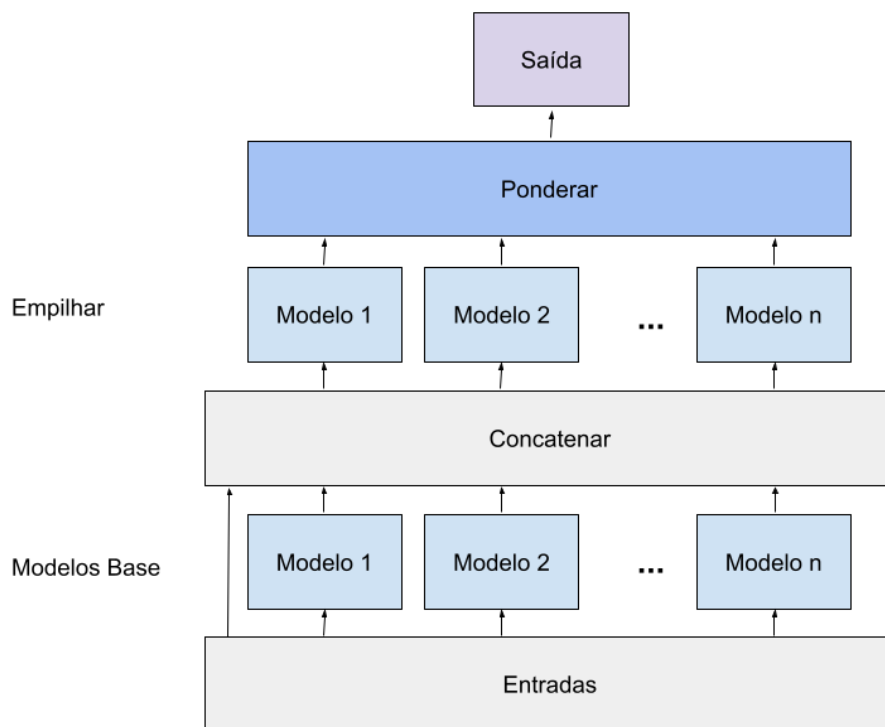


Figura 5.3: Estratégia de empilhamento multicamadas do AutoGluon, exemplificada com duas camadas de empilhamento e  $n$  tipos de modelos de aprendizado básicos.

(*Multi-Layer Stack Ensembling*). Como pode ser visto na Figura 5.3, a primeira camada possui vários modelos individuais de base, que recebem os valores de entradas passadas ao algoritmo no momento do treinamento [27]. Em seguida, é alimentada a próxima camada que consiste em vários modelos de empilhador, pós-concatenação de modelos. Esses empilhadores então atuam como modelos básicos para uma camada adicional.

O AutoGluon reutiliza todos os seus tipos de modelo de camada base, considerando os mesmos valores de hiperparâmetros, como empilhadores. Essa técnica pode ser vista como uma forma alternativa de aprendizado profundo que utiliza o treinamento em camadas, em que as unidades conectadas entre as camadas podem ser modelos de Aprendizado de Máquina arbitrários. Os modelos empilhados tomam como entrada não apenas as previsões dos modelos na camada anterior, mas também os próprios recursos de dados originais, permitindo revisitar os dados originais para melhorar a precisão. A camada de empilhamento final aplica uma técnica conhecida como seleção de conjunto para agregar as previsões dos modelos de empilhador de maneira ponderada, gerando os resultados finais como saída do algoritmo [56]. É importante ressaltar que todo esse processo de treinamento para os modelos utiliza a técnica de Validação Cruzada e cada modelo gerado é salvo em disco, para aumentar a precisão no momento da reutilização

dos modelos em camadas seguintes.

### 5.3.4 Tipos de modelos base

O AutoGluon utiliza um conjunto de modelos sob medida e em uma ordem predefinida. São considerados os seguintes modelos: Redes Neurais, árvores impulsionadas por LightGBM [57], árvores impulsionadas por CatBoost [58], Florestas Aleatórias (*Random Forests*) [59], Árvores Extremamente Aleatórias [60] e kNN (Vizinhos k-Mais Próximos) [61]. A biblioteca usou as implementações do pacote *scikit-learn*<sup>3</sup> desses últimos três modelos. O AutoGluon também permite que os usuários possam facilmente adicionar seus próprios modelos a este conjunto para treinar, ajustar e agrupar de forma automática.

## 5.4 Resultados dos modelos treinados

Neste trabalho foram criados dois modelos previsores, um para a Constante Dielétrica e outro para o Fator de Qualidade, ambos utilizando a ferramenta AutoGluon. Dentre todas as técnicas e ferramentas avaliadas, o AutoGluon apresentou os melhores resultados, considerando as métricas de performance de erros. O nome do modelo, apresentado pela ferramenta, é *weighted\_ensemble\_k0\_l2*, tanto para a Constante Dielétrica, quanto para o Fator de Qualidade. Essa nomenclatura interna é dada, considerando o que ocorre nas etapas de execução do algoritmo, mencionadas anteriormente neste capítulo. Esse modelo final é criado de forma automática, unificando modelos das camadas anteriores, representando uma combinação de modelos de forma ponderada em múltiplos níveis. Apesar de ser o mesmo nome para os dois casos, não representam o mesmo modelo, pois as combinações de modelo base e ponderações são diferentes.

Para chegar ao modelo final, uma série de experimentos foram realizados. Foram consideradas as bases de dados de 22 e 58 entradas para comparação. Em ambos os casos, treinamentos foram realizados por 4 e 6 horas consecutivas, considerando MAE, RMSE e R2 como métricas de avaliação em cada um dos experimentos. O modelo de 58 entradas apresentou melhores resultados e por essa razão mais uma série de experimentos foi realizada considerando o tempo total de 24 horas. Para a base de 22 entradas, 6 experimentos foram realizados e para 58 entradas foram realizados 9 experimentos. O tempo total de

---

<sup>3</sup>Fonte: [www.scikit-learn.org](http://www.scikit-learn.org)

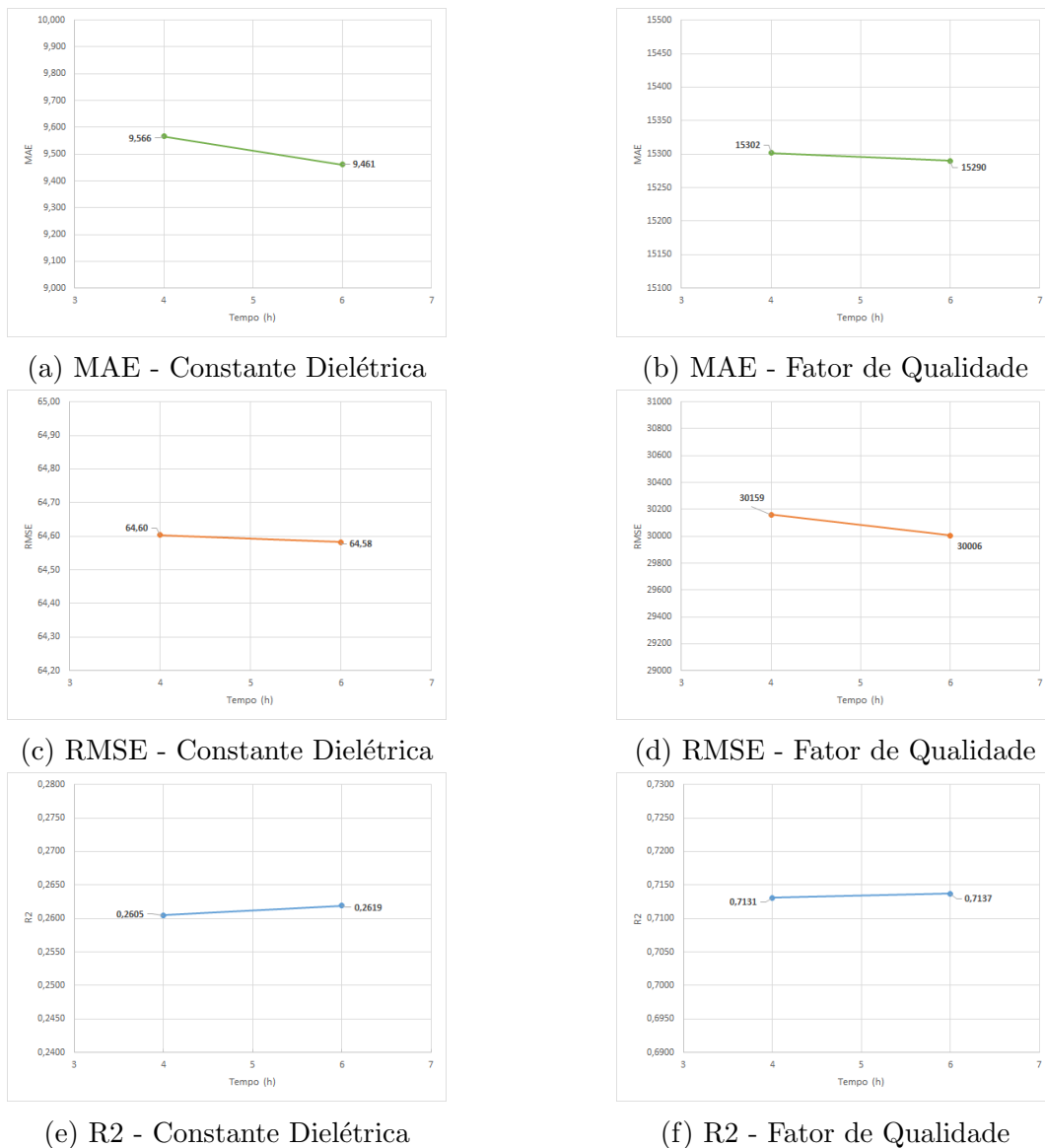


Figura 5.4: Variações das métricas de performance em horas para os modelos AutoGluon, treinados com a base de dados de 22 entradas.

treinamento, considerando todas as avaliações, foi de 132 horas. A Figura 5.4 apresenta a análise dos dados gerados pelos treinamentos com a base de 22 entradas. As Figuras 5.4a, 5.4c e 5.4e, presentes na coluna da esquerda, são referentes à avaliação da Constante Dielétrica. As As Figuras 5.4b, 5.4d e 5.4f, apresentam as avaliações na coluna da direita para o Fator de Qualidade.

Analisando os modelos de treinados para a Constante Dielétrica com 22 entradas, temos a Figura 5.4a que demonstra um pequeno decréscimo no valor do MAE, após duas horas a mais de treinamento. É uma diferença pequena de praticamente 0,10, representando a convergência do modelo. O mesmo ocorre com a pequena melhora apresentada nos valores de RMSE, para a Figura 5.4c, com uma queda de apenas 0,02. Quase não há variação, porém uma pequena melhora neste caso. A Figura 5.4e mostra uma melhora

muito discreta ao longo do tempo, para a métrica R2, de apenas 0,0014. Para todos os casos, os modelos parecem ter convergido, apresentando poucas melhorias ou até mesmo uma pequena piora com mais tempo para treinar.

Considerando os modelos treinados para o Fator de Qualidade e 22 entradas, a Figura 5.4b também demonstra uma melhora do valor do MAE com o passar do tempo, porém muito discreta pela diferença de apenas 12, considerando a escala de valores em 15 mil. Para a métrica RMSE, da Figura 5.4d, ocorre a melhora com a diferença do valor inicial para o final de 153. Já para o valor de R2 na Figura 5.4f há um aumento de apenas 0,0006. Novamente, temos uma convergência nos treinamentos, que após aumentar em 50% o tempo de treinamento, os resultados não apresentaram variações significativas.

A Figura 5.5 exibe os gráficos para os modelos treinados com a base de dados de 58 entradas. Novamente, os gráficos da coluna à esquerda, representados pelas Figuras 5.5a, 5.5c e 5.5e fazem referência à propriedade Constante Dielétrica. Já as Figuras 5.5b, 5.5d e 5.5f, exibem os valores para o Fator de Qualidade.

Analisando a coluna à esquerda contendo os valores de Constante Dielétrica, a Figura 5.5a apresenta queda de 0,03 para o MAE, após mais 20 horas de treinamento com relação ao modelo inicial. O comportamento é similar para a melhora do R2 na Figura 5.5e, com um aumento de apenas 0,0008. O modelo avaliado com a métrica RMSE na Figura 5.5c, foi o que apresentou o maior ganho, diminuindo 7,71. Este valor corresponde uma redução de mais de 20%. Porém, essa melhora com relação ao modelo treinado de 6h é muito baixa, mostrando que o ponto de convergência foi atingido com muito menos tempo de treinamento.

Na coluna à direita, com as métricas correspondentes para o Fator de Qualidade, a Figura 5.5b, ao final temos uma melhora de 13, representando uma melhora de cerca de apenas 1%. Esse comportamento se repete para o RMSE na Figura 5.5d, com uma melhora aproximada de apenas 3%. O mesmo ocorre para o R2 na Figura 5.5f, que apresenta um ganho baixo de 0,0003.

É possível observar que mesmo com mais tempo para o treinamento, os modelos já apresentam grau de convergência quando treinados por 4h. Comparando as duas bases, temos os melhores resultados sempre nos modelos treinados com 58 entradas, independente da métrica avaliada. Nos casos para a Constante Dielétrica a diferença é grande ao observar as métricas R2, com valores de 0,26 para 22 entradas e 0,86 para 58 entradas. O mesmo corre para o RMSE, com diferença discrepante entre 64 e 28. No caso do Fator de Qualidade, o RMSE apresenta uma diferença de cerca de 10% entre as duas abordagens. Apesar disso, as diferenças não são tão discrepantes como ocorre com a Constante Dielétrica, apesar do modelo de 58 entradas sobressair em todos os casos.

Embora a convergência tenha ocorrido com menos horas de treinamento, os modelos finais adotados foram os que apresentaram os melhores resultados, ou seja, pertencem ao grupo de treinamento de 24 horas para a base de 58 entradas. Por decisão de projeto,

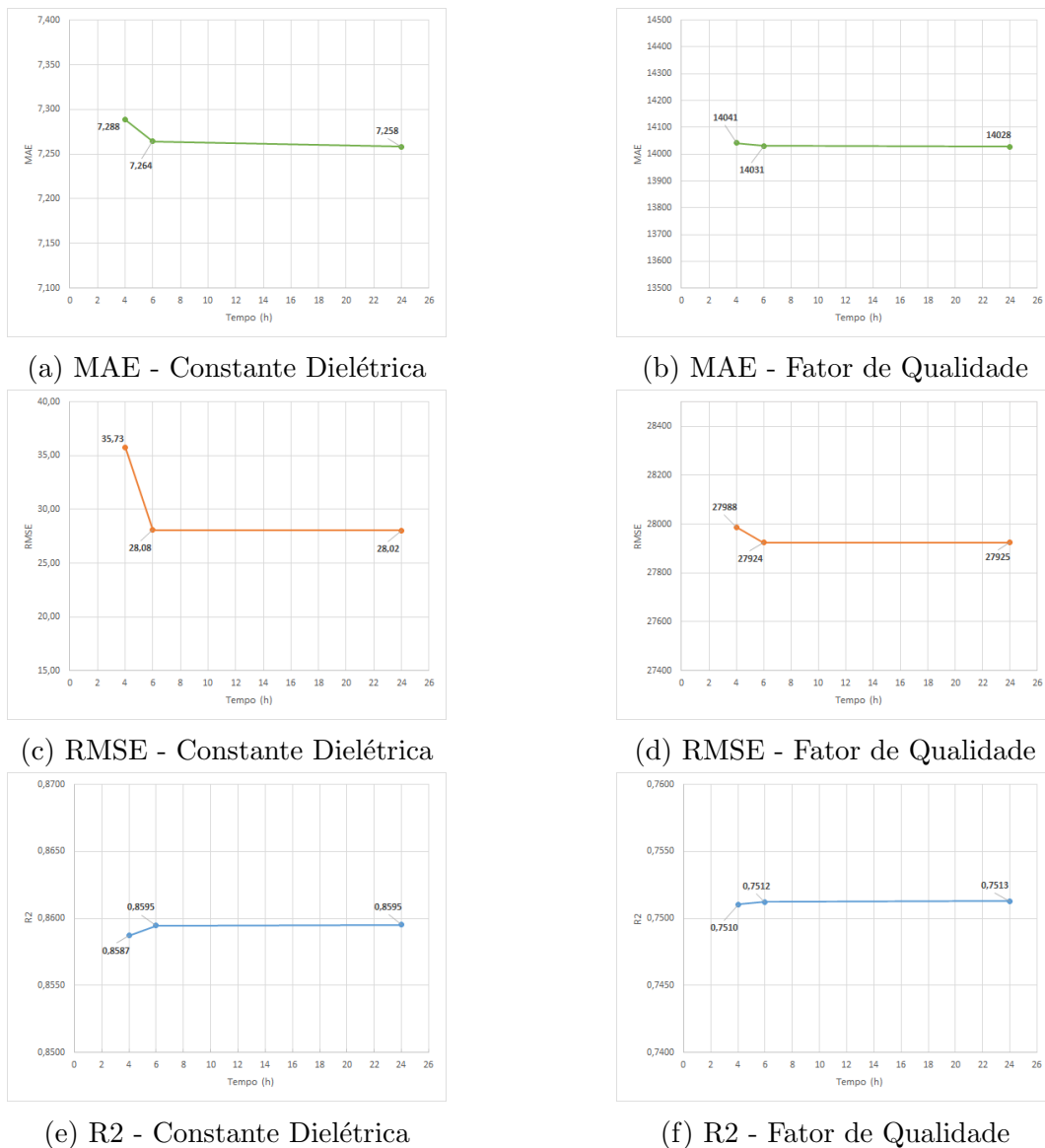


Figura 5.5: Variações das métricas de performance em horas para os modelos AutoGluon, treinados com a base de dados de 58 entradas.

os modelos finais escolhidos foram os treinados considerando a métrica de avaliação R2, que indica que o modelo explica a maioria da variabilidade dos dados de resposta em torno de sua média, conforme explicado no Capítulo 2. No caso do Fator de Qualidade, 75% dos dados se ajustam ao modelo de regressão, enquanto para a Constante Dielétrica esse valor é de 86%.

Outros modelos também foram avaliados durante a experimentação empírica. Apesar dos resultados apresentados na Tabela 5.3, os valores gerados foram a partir de poucos treinamentos. Não foi aprofundado em possibilidades, iterações, configurações e tempo de treinamento, como ocorreu com as Redes Neurais e o AutoGluon. Apesar disso, alguns valores de performance interessantes foram extraídos, permitindo a comparação com os modelos estudados com maior profundidade.

Tabela 5.3: Resultados das métricas de erro entre os métodos estudados para a Constante Dielétrica e a base de dados de 58 entradas.

Método / Métrica	MAE	RMSE	R2
<b>Redes Neurais</b>	19,42	69,30	0,15
<b>XGBoost</b>	13,72	53,86	0,20
<b>XGBoost Otimização Bayesiana</b>	11,31	85,74	0,40
<b>Auto-keras</b>	29,70	-	-
<b>TPOT</b>	-	-	0,55
<b>Auto-sklearn</b>	9,90	28,20	0,78
<b>H2O</b>	11,80	38,20	0,42
<b>AutoGluon</b>	<b>7,30</b>	<b>28,08</b>	<b>0,86</b>

A partir da Tabela 5.3, vemos que o AutoGluon foi o melhor modelo em todas as três métricas de desempenho. Os modelos gerados pelo TPOT e Auto-keras foram explorados apenas com uma das métricas, por essa razão alguns valores aparecem em branco. Também é possível observar que mesmo com os vários treinamentos de Rede Neural Artificial com uma camada, quase todas as outras abordagens apresentaram menor MAE e RMSE, além de valores de R2 mais altos.

#### 5.4.1 Comparação entre valores previstos e reais

Nesta seção serão apresentadas as comparações de todos os valores previstos e realizados, para os dois modelos previsores criados. O primeiro modelo avaliado é relativo à propriedade Constante Dielétrica, que apresentou um ótimo resultado de precisão. O resultado da métrica R2 para esse modelo foi de 0,86, em outras palavras, 86% da variância dos dados pode ser explicada neste contexto. Para fins de comparação, na Figura 2.11, explicada anteriormente, há um modelo de regressão linear que representa um R2 mais próximo do valor 1, que descreve bem o conjunto de dados apresentados.

A Figura 5.6a apresenta o conjunto de valores reais, em azul, e previstos, em laranja, para a Constante Dielétrica considerando o modelo final adotado. É possível observar que há um acompanhamento muito próximo dos valores em uma escala geral, demonstrando que o modelo consegue ser muito descritivo para o conjunto de dados adotados para treinamento. Ao observar a divisão por quartis, temos que o Primeiro Quartil mostrado na Figura 5.6b possui uma dispersão de pontos que acompanha as previsões e possui 2 grandes *outliers*, que estão explicados no Apêndice B. No geral, há um erro

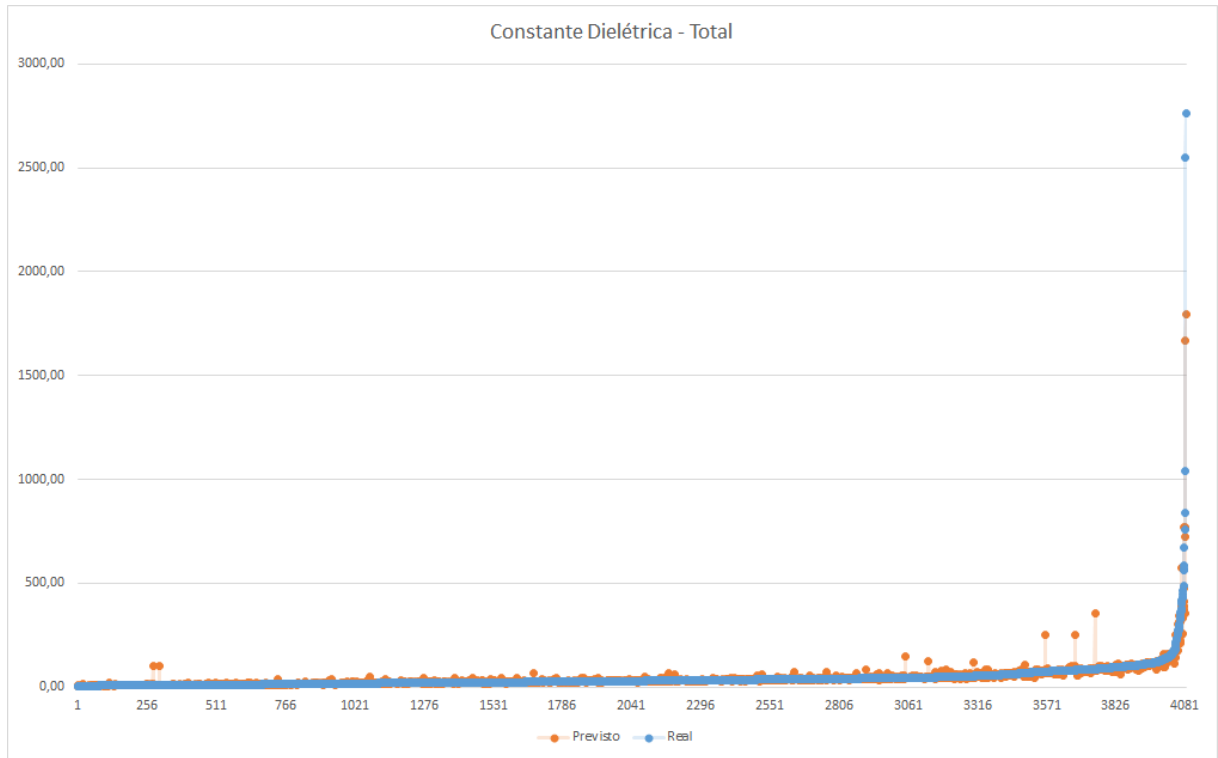
maior, considerando um exemplo no qual o valor real é 3 e o previsto foi 9, o erro foi 3 vezes o esperado. Mas mesmo assim, o previsor continua apontando que o material possui baixa Constante Dielétrica, considerando que a média de toda a base de valores reais é de 80 e a maioria está classificada abaixo de 20.

O Segundo Quartil, na Figura 5.6c, também segue a linha de tendência, em que a maioria dos materiais em laranja acompanha a linha azul, com exceção de alguns *outliers*. O Terceiro Quartil, na Figura 5.6d, tem uma aproximação entre real e previsto ainda maior que o quartil anterior. Já o Quarto Quartil, na Figura 5.6e possui a linha de tendência que melhor acompanha o conjunto de materiais. É interessante notar que os materiais previstos com maiores valores de Constante Dielétrica formam um pequeno grupo abaixo da linha azul, ao final da curva, porém muito próximos do real. Eles apresentam a tendência de curva crescente, mostrando que o modelo conseguiu prever em quais situações podemos encontrar Constantes Dielétricas maiores.

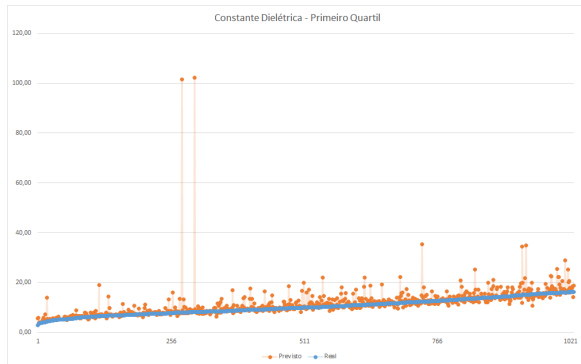
Considerando o modelo criado para o Fator de Qualidade, temos o valor de R2 que corresponde à 0,75, ou seja, 75% da variância dos dados sendo explicada. Apesar de ser mais baixo do que o modelo da Constante Dielétrica, ainda há uma boa precisão e é possível ver que o modelo é descritivo para o conjunto de dados, conforme mostrado na Figura 5.7a. Os pontos em laranja, valores previstos, seguem a tendência dos pontos azuis, que formam uma linha representando os valores reais. O previsor acompanha os valores mais altos e também determina que certos materiais terão o Fator de Qualidade mais baixos.

Analisando o Primeiro Quartil, na Figura 5.7b, é possível observar uma dispersão dos pontos previstos, mas que seguem a tendência crescente. A maioria dos valores se encontra na faixa abaixo de 20 mil, sendo considerados valores baixos de Fator de Qualidade em relação ao contexto geral de valores, cuja média é de 80 mil e máximo é superior a 700 mil. Devido à grande diferença de escala na base, existem valores reais como, por exemplo, 16 que podem ser previstos como 1 mil. Mesmo sendo classificado como um valor baixo para o contexto geral, o erro nesse caso seria de 62,5 vezes o valor original.

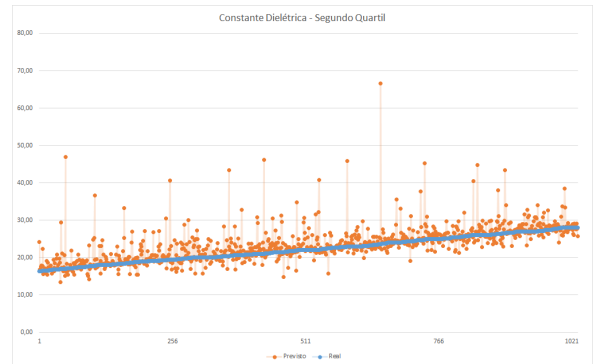
O Segundo Quartil, presente na Figura 5.7c, também possui materiais que são *outliers* e tendência da dispersão de pontos laranjas é próxima ao Primeiro Quartil. A grande maioria das previsões está alocada abaixo da linha dos 50 mil e mesmo com o erro associado, o previsor não tem a tendência de apontar valores muito altos para esse quadrante. O Terceiro Quartil, exibido na Figura 5.7d, também possui alguns pontos fora da curva, mas a tendência de dispersão dos pontos laranja diminui significativamente. Nesses casos, o previsor está apresentando uma taxa de acerto mais alta. O último quartil, na Figura 5.7e, é o que apresenta melhor tendência de previsão. Para valores mais altos o modelo consegue ser mais preciso e também acompanhar a linha de tendência crescente.



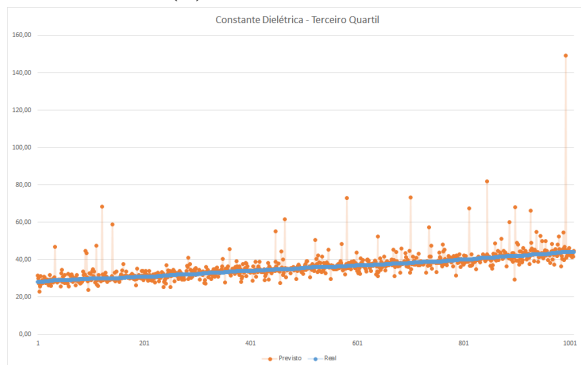
(a) Todos os valores.



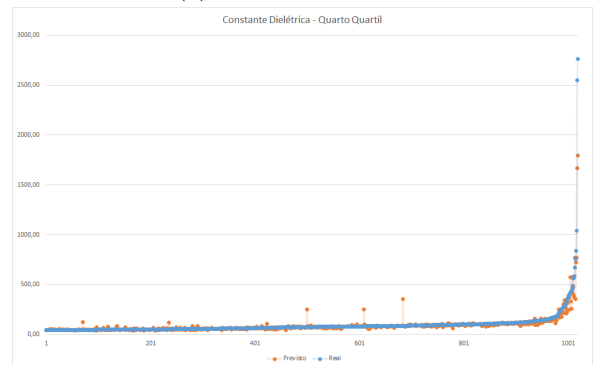
(b) Primeiro quartil.



(c) Segundo quartil.

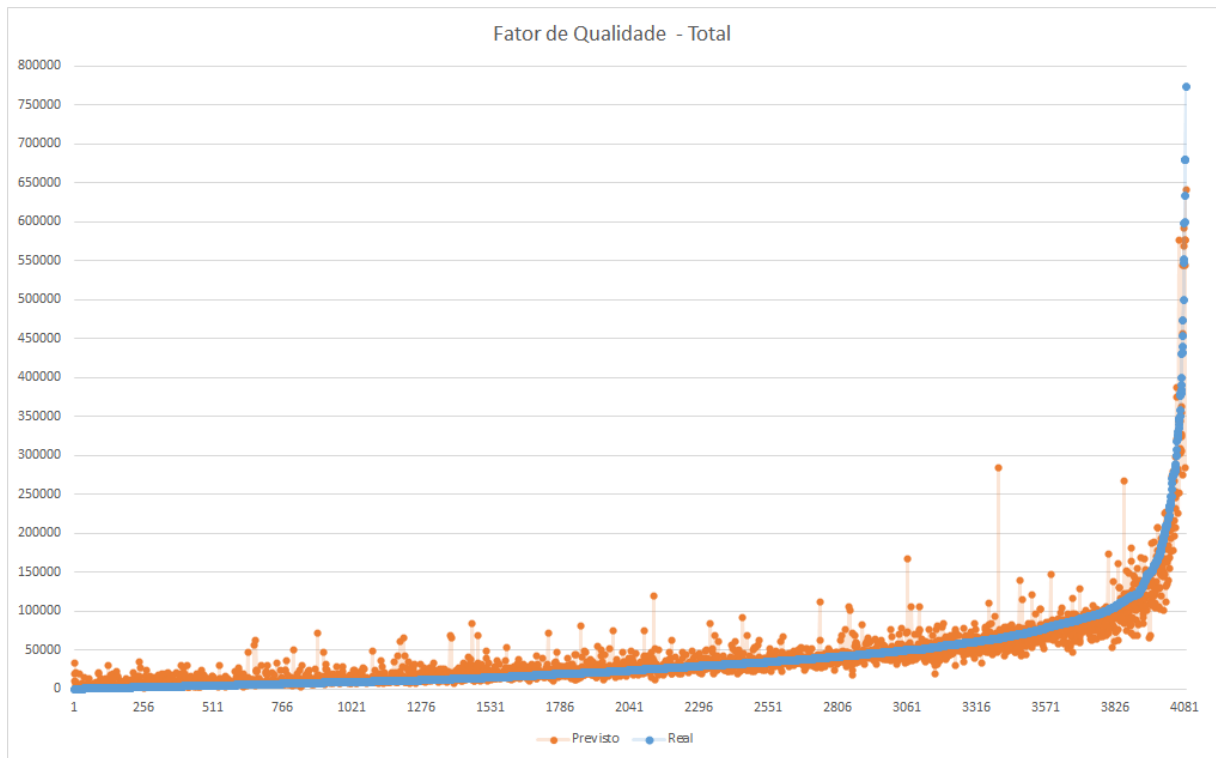


(d) Terceiro quartil.

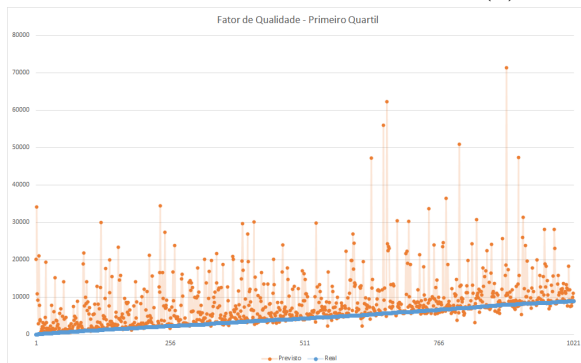


(e) Quarto quartil.

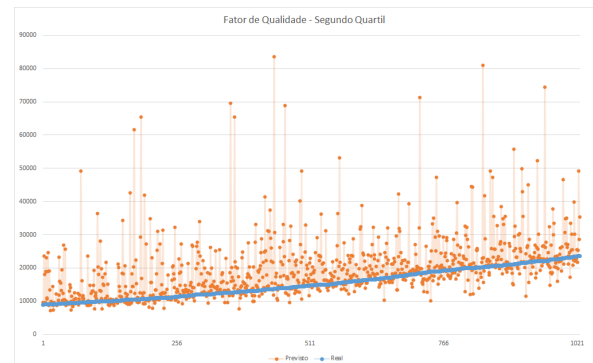
Figura 5.6: Comparação dos valores reais e previstos para a Constante Dielétrica.



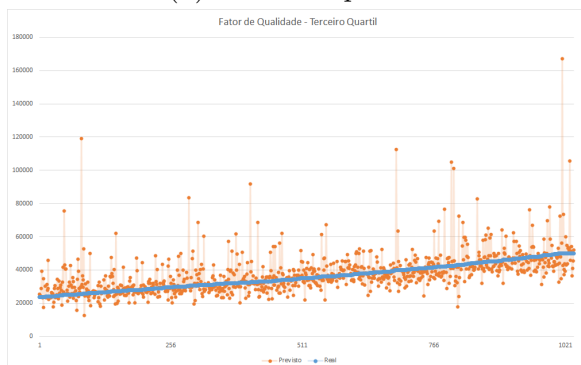
(a) Todos os valores.



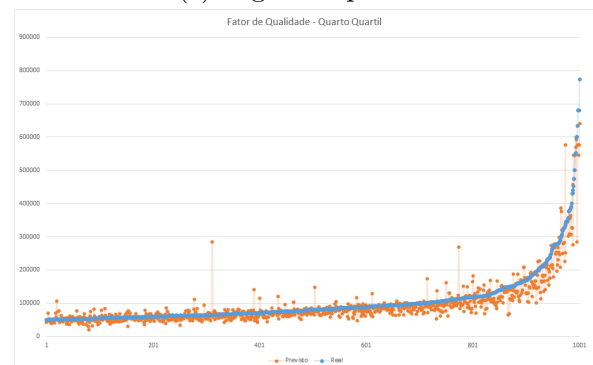
(b) Primeiro quartil.



(c) Segundo quartil.



(d) Terceiro quartil.



(e) Quarto quartil.

Figura 5.7: Comparação dos valores reais e previstos para o Fator de Qualidade.

### 5.4.2 Avaliação do Erro Relativo Percentual Normalizado

Para avaliar a consistência dos dois modelos finais escolhidos, mais uma abordagem foi adotada para descobrir os valores de erros e outras métricas estatísticas. Primeiramente o conjunto de valores reais e previstos foram normalizados, seguindo a seguinte abordagem explicada pela Equação 5.1:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (5.1)$$

onde:

$x_{norm}$  = valor normalizado;

$x$  = valor atual a ser normalizado;

$x_{min}$  = menor valor de todo o conjunto de dados;

$x_{max}$  = maior valor de todo o conjunto de dados.

Após a normalização, foi calculado o Erro Relativo Percentual, explicado pela Equação 5.2:

$$e_p = \frac{v_e - v_r}{v_r} * 100, \quad (5.2)$$

em que:

$e_p$  = erro relativo percentual;

$v_e$  = valor estimado ou previsto;

$v_r$  = valor real.

Em resumo, todos os valores da base de dados, tanto os previstos como reais, passaram pela normalização que os transformaram em valores que vão de 0 a 1 ( $x_{norm}$ ). Em seguida, para cada par de valores real ( $v_r$ ) e previsto ( $v_e$ ), foi calculado o Erro Relativo Percentual correspondente.

A Tabela 5.4, apresenta a análise dos Erros Relativos Percentuais encontrados para a Constante Dielétrica e separados por quartil, observando as métricas em relação ao erro: maior, menor, média, mediana, desvio padrão e a quantidade de materiais. É possível observar que os menores erros encontrados em todos os casos é muito próximo a zero. O Primeiro Quartil, que corresponde ao intervalo das Constantes Dielétricas reais de 3 a 16,3, apresentou maiores valores médios e de desvio padrão para esse intervalo. Ainda assim, como é possível constatar na distribuição de valores apresentada na Figura 5.6b, o predictor determina que existe uma probabilidade grande de o material sendo analisado

Tabela 5.4: Avaliação por quartil dos erros relativos percentuais normalizados para a Constante Dielétrica.

Métricas	Primeiro Quartil (25% percentil)	Segundo Quartil (50% percentil)	Terceiro Quartil (75% percentil)	Quarto Quartil (100%)
<b>Intervalo</b>	3 a 16,3	16,3 a 28	28 a 44	44 a 2763
<b>Menor</b>	0,0019	0,0007	0,0025	0,0061
<b>Maior</b>	9400239	214	257	325
<b>Média</b>	9185	9,72	6,55	7,81
<b>Mediana</b>	7,47	4,34	3,29	3,46
<b>Desvio Padrão</b>	293327	17,37	13,75	18,11
<b>Quantidade</b>	1026	1032	1007	1019

Tabela 5.5: Avaliação por quartil dos erros relativos percentuais normalizados para o Fator de Qualidade.

Métricas	Primeiro Quartil (25% percentil)	Segundo Quartil (50% percentil)	Terceiro Quartil (75% percentil)	Quarto Quartil (100%)
<b>Intervalo</b>	16 a 8946,5	8946,5 a 23600	23600 a 50000	50000 a 773700
<b>Menor</b>	0,0005	0,004	0,02	0,01
<b>Maior</b>	259034772237	518	367	338
<b>Média</b>	253707680	34,23	15,81	14,63
<b>Mediana</b>	45,39	13,78	8,93	10,96
<b>Desvio Padrão</b>	8102749435	55,81	24,37	16,62
<b>Quantidade</b>	1021	1025	1037	1001

possuir de fato uma Constante Dielétrica pequena, abaixo de 20, em comparação com todos os materiais.

Os valores médios para o Segundo, Terceiro e Quarto quartil são inferiores a 10%, apresentando ótima assertividade nas previsões realizadas. Os desvios padrões para esses quartis variam de 13,75 a 18,11 %, demonstrando um bom intervalo de confiabilidade. É possível observar que o modelo predictor criado apresenta os melhores resultados para Constantes Dielétricas mais elevadas, principalmente as que estão no Terceiro Quartil, com média de erro de apenas 6,55%.

Considerando o modelo predictor para o Fator de Qualidade, cujos resultados estão listados na Tabela 5.5, temos um comportamento bastante similar. Para valores muito pequenos, há um erro associado maior. O intervalo de valores para o Fator de Qualidade é grande, indo de 16 a 773700. Assim como ocorre para o modelo da Constante Dielétrica, o predictor associa esses materiais com valores de Fator de Qualidade mais baixos, em que a maioria está abaixo do valor de 35 mil, como pode ser visto na Figura 5.7b.

Para o Segundo Quartil, há uma média de erros de 33%, com desvio padrão de 55,81%, mostrando que existe uma melhora nas previsões realizadas para esse intervalo. O predictor continua indicando que os materiais não possuem valores elevados e a grande maioria encontra-se abaixo de 50 mil. Os Terceiro e Quarto Quartis apresentam as maiores taxa de assertividade, com o destaque para o último quartil com um erro médio de apenas 14,63%. Para as duas propriedades analisadas, essas últimas faixas de valores são as mais importantes, pois quanto melhores os predictores forem para valores altos, menor será a chance de erros para encontrar novos materiais que podem se aproximar ou ultrapassar a Frente de Pareto.

### 5.4.3 Avaliação de exemplos na base de dados

Nota-se que para o Fator de Qualidade, existe um erro relativo percentual maior, assim como valores maiores do desvio padrão e mediana. Essa propriedade física demonstrou-se mais sensível, tanto a pequenas alterações na composição química quanto às variações de outras características encontradas na base de dados original. A Tabela 4.1, apresentada no Capítulo 4, contém dois ótimos exemplos dessa variação.

O primeiro exemplo consiste na comparação entre os materiais com Id 6 e 8. Para o modelo previsor, ambos possuem a mesma composição química ( $\text{SiO}_2$ ). A primeira diferença entre os dois pode ser vista na descrição do material Id 6, que apresenta o comentário "*melt method*", indicando o método de criação do composto. Em outros casos da base isso também ocorre, assim como para indicar a temperatura ambiente no momento da medição dos valores. Essas informações foram desconsideradas para criar o modelo de fórmula química final. Outras diferenças ocorrem na TS, Estrutura Cristalina e  $\tau_f$ . O Fator de Qualidade do material Id 6 foi de 122100, enquanto para o Id 8 foi 80400. Essa variação corresponde a mais de 50% do valor de um para o outro.

Outro caso interessante ocorre nos materiais Id 444 e 445, na mesma Tabela 4.1. Os materiais possuem as mesmas propriedades, exceto  $\tau_f$  que é 55 para o Id 444 e 56 para o Id 445. A diferença na fórmula química entre os dois materiais é muito pequena, sendo representada pela diferença entre o valor de  $x$  da fórmula, que no primeiro caso é 0,993 e no segundo 0,995. Apesar de possuírem quase a mesma composição química, o valor do Fator de Qualidade para o Id 444 é de 773700, enquanto para o Id 445 é 551700. Há uma diferença de 222000 entre os dois, enquanto a Constante Dielétrica permaneceu a mesma. Esse tipo de situação impacta na construção do modelo preditivo. Mesmo que fossem adotadas todas as características da tabela de referência original, essa pequena variação em alguns casos eleva drasticamente o valor do Fator de Qualidade. Na mesma base de dados, há casos em que a composição química é muito similar e os Fatores de Qualidade são muito próximos.

No próximo capítulo, serão discutidos a Conclusão e os Trabalhos Futuros. Serão avaliados os resultados apresentados pela Interface de Usuário e dos modelos previsores criados, conforme o objetivo e motivação propostos. Sugestões de continuidades desta dissertação também serão apresentadas, tanto para trabalhos relacionados, como para melhorias e evolução do sistema criado.

# Capítulo 6

## Conclusões e Trabalhos futuros

### 6.1 Conclusões

Devido ao avanço do poder de processamento dos computadores e dos crescentes evoluções na área de Aprendizado de Máquina, novas soluções e sistemas têm emergido nos mais diversos campos da ciência. Neste trabalho, apresentamos um sistema de Informática de Materiais. O CPS é um Sistema Web que fornece ferramentas de análise e previsão de propriedades físicas, para materiais cerâmicos dielétricos.

O sistema conta com funcionalidades que facilitam o estudo e a avaliação de materiais cerâmicos. Possui uma área para análise dos materiais, contendo três modelos principais de gráficos. É possível realizar filtros por elementos e compostos, além de visualizar a Frente de Pareto, que exibe o conjunto dominante de materiais. A aplicação de filtros faz com que a ferramenta calcule uma nova Frente de Pareto, exibindo também o cálculo da Distância Euclidiana, que representa o quão próximo o material está do conjunto ótimo original. É possível acessar a tabela completa por dentro da própria interface, visualizando todos os detalhes das propriedades dos cerâmicos.

Além disso, o CPS conta com o módulo de previsão de propriedades físicas. Nesta seção, é possível criar uma nova fórmula química e submetê-la para análise dos modelos preditivos. Existem duas áreas para visualizar a informação dos modelos de previsão: a primeira delas uma tabela com a lista de fórmulas químicas, seus respectivos valores de Constante Dielétrica e Fator de Qualidade; a segunda conta com um gráfico com Frente de Pareto, exibindo os detalhes do item ao passar o cursor sobre o mesmo, assim como o cálculo da Distância Euclidiana até a Fronteira de Pareto.

Dois modelos preditivos distintos foram criados utilizando o AutoGluon, um para a Constante Dielétrica e outro para o Fator de Qualidade. Considerando as métricas de desempenho avaliadas, o modelo final da Constante Dielétrica conta com MAE de 7,3 e RMSE de 28,08. Esse modelo demonstrou ótimos resultados preditivos, sendo bastante assertivo, mesmo com as variações de outras propriedades do material. Já o modelo do Fator de Qualidade apresentou MAE 14028 e RMSE de 27925. Apesar dessa propriedade

ser mais sensível às variações de outras características do material, demonstrou-se um ótimo modelo para previsão, mesmo tendo um desempenho menor em comparação com a Constante Dielétrica. Através das métricas de performance R2, o previsor para Constante Dielétrica conta com 86% da variância explicada pelo modelo, enquanto o previsor do Fator de Qualidade possui o valor de 75%. Ambos possuem uma boa taxa de acerto, gerando previsões próximas ao esperado. A ferramenta possibilita estimar as propriedades de novos materiais cerâmicos, sem ser necessário sintetizar e posteriormente aferir, cujo custo financeiro e de tempo é muito mais alto do que a previsão computacional.

## 6.2 Trabalhos futuros

Apesar dos resultados promissores encontrados nos modelos preditivos e também nas ferramentas de análise já desenvolvidas, observaram-se algumas possibilidades de melhoria. O sistema CPS está preparado para receber novas abas em um sua interface, permitindo a criação de outras ferramentas analíticas e funcionalidades. Além disso, a arquitetura de microsserviços favorece a criação de novos módulos independentes, em qualquer linguagem de programação. Podem ser facilmente acopláveis ao sistema final, através da conexão a serviços de dados REST. A seguir, são listados alguns trabalhos que podem ser desenvolvidos na sequência:

- Incorporar mais características ao modelo de treinamento: a simplificação do modelo preditivo, para considerar apenas os elementos químicos e suas respectivas quantidades, é boa para o usuário final criar seus próprios compostos, sem conhecer muitos detalhes. Uma possibilidade é treinar um modelo com outras características, tais como temperatura de sintetização e estrutura cristalina, melhorando o modelo preditivo. Algumas considerações importantes:
  - Caso o usuário deseje informar essas características extras, a ferramenta pode utilizar o modelo mais completo e caso não seja informado, o treinamento pode ser feito com o modelo original. O novo modelo poderia ser utilizado de opcionalmente para melhorar a precisão;
  - O Fator de Qualidade se mostrou menos resiliente a variações em comparação à Constante Dielétrica, devido a duplicidades de fórmulas. Para um mesmo composto, valores muito distintos de Fator de Qualidade foram encontrados. Uma possibilidade é distinguir essas fórmulas mediante outras características, tratando o problema;

- Explorar outros modelos de AutoML mais a fundo: neste trabalho alguns modelos foram estudados, porém, existem vários outros métodos, abordagens e configurações a serem exploradas. Mesmo alguns dos modelos citados brevemente neste trabalho podem ser treinados com outras configurações e por mais tempo. Existe a possibilidade de encontrar previsores ainda mais eficazes;
- Incorporar filas de mensagens: é possível deixar a arquitetura do sistema mais robusta, aplicando o conceito de filas para comunicação assíncrona entre serviços. As mensagens são armazenadas na fila até serem processadas e excluídas. Cada mensagem é processada uma única vez, por um único consumidor. As filas de mensagens podem ser usadas para dissociar processamento pesado;
- Modelo de previsão inversa: uma possibilidade interessante é fazer o caminho inverso deste trabalho, encontrando as fórmulas químicas sugeridas a partir do par de valores de Constante Dielétrica e Fator de Qualidade. Seria possível desenvolver um microserviço apenas para esse novo previsor e desenvolver a interface em uma nova aba da ferramenta;
- Análise das outras características da referência original: é possível desenvolver uma nova área na interface para analisar, com gráficos, tabelas e filtros, os materiais a partir de temperaturas de síntese, estruturas cristalinas, frequência de medição, entre outros.
- Incorporar novos materiais cerâmicos: o sistema foi desenvolvido de forma que novos materiais, de outros referenciais teóricos e estudos, possam ser incorporados a ele. Os materiais podem ser listados tanto nas visualizações gráficas, como também utilizados para aprimorar os modelos preditivos.

# Referências Bibliográficas

- [1] Anthony JG Hey, Stewart Tansley, Kristin Michele Tolle, et al. *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA, 2009.
- [2] Omar Paranaíba Vilela Neto. Intelligent computational nanotechnology: the role of computational intelligence in the development of nanoscience and nanotechnology. *Journal of Computational and Theoretical Nanoscience*, 11(4):928–944, 2014.
- [3] Krishna Rajan. Materials informatics. *Materials Today*, 8(10):38–45, 2005.
- [4] William Callister. *Ciência E Engenharia de Materiais: Uma Introdução*. Grupo Gen-LTC, 2000.
- [5] Mailadil T Sebastian, Rick Uvic, and Heli Jantunen. *Microwave materials and applications*. John Wiley & Sons, 2017.
- [6] MT Sebastian and LK Namitha. Rubber–ceramic composites. *Microwave Materials and Applications, 2 Volume Set*, 1:537, 2017.
- [7] James H Harlow. *Electric power transformer engineering*. CRC press, 2012.
- [8] Ankit Agrawal and Alok Choudhary. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. *Apl Materials*, 4(5):053208, 2016.
- [9] MT Sebastian, R Uvic, and H Jantunen. Low-loss dielectric ceramic materials and their properties. *International Materials Reviews*, 60(7):392–412, 2015.
- [10] C Barry Carter, M Grant Norton, et al. *Ceramic materials: science and engineering*, volume 716. Springer, 2007.
- [11] Shigeyuki Somiya. *Handbook of advanced ceramics: materials, applications, processing, and properties*. Academic press, 2013.
- [12] Simon Saunders and Harvey R Brown. *The philosophy of vacuum*. 1991.
- [13] LG Hector and HL Schultz. The dielectric constant of air at radiofrequencies. *Physics*, 7(4):133–136, 1936.
- [14] John Bird. *Electrical and electronic principles and technology*. Taylor & Francis, 2017.

- 
- [15] Hugh D Young and Roger A Freedman. Física iii: eletromagnetismo. *São Paulo: Person Education do Brasil*, 2009.
- [16] Ian Hickman. *Analog Electronics: Analog Circuitry Explained*. Newnes, 2013.
- [17] Jurgen Garcke, Chris K Dyer, Patrick T Moseley, Zempachi Ogumi, David AJ Rand, and Bruno Scrosati. *Encyclopedia of electrochemical power sources*. Newnes, 2013.
- [18] Peter Hausler, Carina Roth, Thomas Vitzthumecker, and Rudolf Bierl. Miniaturized surface plasmon resonance based sensor. *Optics, Photonics and Laser Technology 2018*, 223:169, 2020.
- [19] Cory Gackenheimer. *Introduction to React*. Apress, 2015.
- [20] Erik Wittern, Philippe Suter, and Shriram Rajagopalan. A look at the dynamics of the javascript package ecosystem. In *Proceedings of the 13th International Conference on Mining Software Repositories*, pages 351–361, 2016.
- [21] Adam Boduch. *React Material-UI Cookbook: Build captivating user experiences using React and Material-UI*. Packt Publishing Ltd, 2019.
- [22] Alexander V Lotov and Kaisa Miettinen. Visualizing the pareto frontier. In *Multi-objective optimization*, pages 213–243. Springer, 2008.
- [23] John Tabak. *Geometry: the language of space and form*. Infobase Publishing, 2014.
- [24] John McCarthy and Edward A Feigenbaum. In memoriam: Arthur samuel: Pioneer in machine learning. *AI Magazine*, 11(3):10–10, 1990.
- [25] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. Practical machine learning tools and techniques. *Morgan Kaufmann*, page 578, 2005.
- [26] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.
- [27] Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*, 2020.
- [28] Nicola Dragoni, Saverio Giallorenzo, Alberto Lluch Lafuente, Manuel Mazzara, Fabrizio Montesi, Ruslan Mustafin, and Larisa Safina. Microservices: yesterday, today, and tomorrow. *Present and ulterior software engineering*, pages 195–216, 2017.
- [29] Mark Masse. *REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces*. "O'Reilly Media, Inc.", 2011.

- [30] Cort J Willmott. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63(11):1309–1313, 1982.
- [31] James P Barrett. The coefficient of determination—some limitations. *The American Statistician*, 28(1):19–20, 1974.
- [32] Olivier Renaud and Maria-Pia Victoria-Feser. A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference*, 140(7):1852–1862, 2010.
- [33] Omar Paranaiba VILELA NETO and Marco Aurélio Cavalcanti PACHECO. Nanotecnologia computacional inteligente: Concebendo a engenharia em nanotecnologia. *PUCRio: Interciência*, 2012.
- [34] Xiao Wan, Wentao Feng, Yunpeng Wang, Haidong Wang, Xing Zhang, Chengcheng Deng, and Nuo Yang. Materials discovery and properties prediction in thermal transport via materials informatics: a mini review. *Nano letters*, 19(6):3387–3395, 2019.
- [35] Surya R Kalidindi and Marc De Graef. Materials data science: current status and future outlook. *Annual Review of Materials Research*, 45:171–193, 2015.
- [36] Yao Cai, Wei Xie, Yin Ting Teng, Padinhare Cholakkal Harikesh, Biplab Ghosh, Patrick Huck, Kristin A Persson, Nripan Mathews, Subodh G Mhaisalkar, Matthew Sherburne, et al. High-throughput computational study of halide double perovskite inorganic compounds. *Chemistry of Materials*, 31(15):5392–5401, 2019.
- [37] Donghoon Kim, Hyeon Han, June Ho Lee, Jin Woo Choi, Jeffrey C Grossman, Hyun Myung Jang, and Donghun Kim. Electron- hole separation in ferroelectric oxides for efficient photovoltaic responses. *Proceedings of the National Academy of Sciences*, 115(26):6566–6571, 2018.
- [38] Alexander S Gzyl, Anton O Oliynyk, Lawrence A Adutwum, and Arthur Mar. Solving the coloring problem in half-Heusler structures: Machine-learning predictions and experimental validation. *Inorganic chemistry*, 58(14):9280–9289, 2019.
- [39] Rohit Batra, Huan Doan Tran, Chiho Kim, James Chapman, Lihua Chen, Anand Chandrasekaran, and Rampi Ramprasad. General atomic neighborhood fingerprint for machine learning-based methods. *The Journal of Physical Chemistry C*, 123(25):15859–15866, 2019.
- [40] JE Gubernatis and T Lookman. Machine learning in materials design and discovery: Examples from the present and suggestions for the future. *Physical Review Materials*, 2(12):120301, 2018.

- [41] Rampi Ramprasad, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials*, 3(1):1–13, 2017.
- [42] Nan Qu, Yong Liu, Mingqing Liao, Zhonghong Lai, Fei Zhou, Puchang Cui, Tianyi Han, Danni Yang, and Jingchuan Zhu. Ultra-high temperature ceramics melting temperature prediction via machine learning. *Ceramics International*, 45(15):18551–18555, 2019.
- [43] Logan Ward, Ankit Agrawal, Alok Choudhary, and Christopher Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2(1):1–7, 2016.
- [44] Grace X Gu, Chun-Teh Chen, Deon J Richmond, and Markus J Buehler. Bioinspired hierarchical composite design using machine learning: simulation, additive manufacturing, and experiment. *Materials Horizons*, 5(5):939–945, 2018.
- [45] Sun-Chong Wang. Artificial neural network. In *Interdisciplinary computing in java programming*, pages 81–100. Springer, 2003.
- [46] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11):e00938, 2018.
- [47] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 2015.
- [48] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.
- [49] Haifeng Jin, Qingquan Song, and Xia Hu. Auto-keras: An efficient neural architecture search system. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1946–1956, 2019.
- [50] Randal S Olson and Jason H Moore. Tpot: A tree-based pipeline optimization tool for automating machine learning. In *Workshop on automatic machine learning*, pages 66–74. PMLR, 2016.
- [51] Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Tobias Springenberg, Manuel Blum, and Frank Hutter. Auto-sklearn: efficient and robust automated machine learning. In *Automated Machine Learning*, pages 113–134. Springer, Cham, 2019.
- [52] Erin LeDell and Sebastien Poirier. H2o automl: Scalable automatic machine learning. In *Proceedings of the AutoML Workshop at ICML*, volume 2020, 2020.

- [53] Bambang Parmanto, Paul W Munro, and Howard R Doyle. Reducing variance of committee prediction with resampling techniques. *Connection Science*, 8(3-4):405–426, 1996.
- [54] Mariia Fedotenkova. *Extraction of multivariate components in brain signals obtained during general anesthesia*. PhD thesis, Université de Lorraine, 2016.
- [55] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.
- [56] Rich Caruana, Alexandru Niculescu-Mizil, Geoff Crew, and Alex Ksikes. Ensemble selection from libraries of models. In *Proceedings of the twenty-first international conference on Machine learning*, page 18, 2004.
- [57] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30:3146–3154, 2017.
- [58] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *arXiv preprint arXiv:1706.09516*, 2017.
- [59] Adele Cutler, D Richard Cutler, and John R Stevens. Random forests. In *Ensemble machine learning*, pages 157–175. Springer, 2012.
- [60] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.
- [61] Yunsheng Song, Jiye Liang, Jing Lu, and Xingwang Zhao. An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing*, 251:26–34, 2017.

# Apêndice A

## Informações sobre a Base de Dados

Esta seção tem o objetivo de clarificar informações relevantes a respeito da base de dados utilizada para treinamento dos modelos preditivos. Ao todo foram utilizados 4084 compostos químicos, representados por 58 elementos e suas respectivas quantidades. Ao todo, são 118 elementos, evidenciando a utilização de quase 50% do total. A Figura A.1 contém uma Tabela Periódica que mostra em azul todos os elementos presentes nas fórmulas químicas utilizadas neste trabalho. Em cinza, estão representados os demais elementos que não foram contemplados, conforme a referência original adotada.

### A.1 Distribuição de Elementos

A primeira análise, para exemplificar a distribuição das informações na base de dados, diz respeito à quantidade de elementos que aparecem mais vezes em diferentes compostos. O objetivo é identificar quais os átomos mais comuns presentes na composição dos materiais. O oxigênio (O) está em primeiro lugar como o mais utilizado. Todas as fórmulas químicas possuem o elemento. Em segundo lugar observamos o titânio (Ti), presente em mais da metade das composições com o total de 2133 ocorrências. Acima de 1 mil ocorrências temos bário (Ba), nióbio (Nb) e cálcio (Ca). Esses são os 5 átomos mais utilizados em todas as fórmulas químicas, como pode ser observado pela Figura A.2a. Já os menos utilizados, temos os elementos que aparecem uma única vez como o enxofre (S) e rubídio (Rb). Apenas duas ocorrências para lutécio (Lu) e escândio (Sc). Com apenas 6 composições químicas distintas temos o túlio (Tm), compondo o time dos 5 átomos que menos aparecem nos registros.

Para enriquecer os exemplos e análise, são apresentados os elementos que aparecem nas composições químicas com maiores valores de Constante Dielétrica e Fatores de Qualidade. Foram selecionadas as 1% fórmulas químicas de maiores valores, de acordo com cada propriedade. Ou seja, são os 40 primeiros compostos ao ordenar dos maiores para os menores valores. A Figura A.2b mostra os elementos encontrados ao ordenar seguindo

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	<b>1</b> <b>H</b> Hidrogênio 1,008														
2	<b>3</b> <b>Li</b> Lítio 6,940	<b>4</b> <b>Be</b> Berílio 9,012											<b>5</b> <b>B</b> Boro 10,810	<b>6</b> <b>C</b> Carbono 12,011	<b>7</b> <b>N</b> Nitrogênio 14,007
3	<b>11</b> <b>Na</b> Sódio 22,990	<b>12</b> <b>Mg</b> Magnésio 24,305											<b>13</b> <b>Al</b> Alumínio 26,982	<b>14</b> <b>Si</b> Silício 28,085	<b>15</b> <b>P</b> Fósforo 30,974
4	<b>19</b> <b>K</b> Potássio 39,098	<b>20</b> <b>Ca</b> Cálcio 40,078	<b>21</b> <b>Sc</b> Escândio 44,956	<b>22</b> <b>Ti</b> Titânio 47,867	<b>23</b> <b>V</b> Vanádio 50,942	<b>24</b> <b>Cr</b> Cromio 51,996	<b>25</b> <b>Mn</b> Manganês 54,938	<b>26</b> <b>Fe</b> Ferro 55,845	<b>27</b> <b>Co</b> Cobalto 58,933	<b>28</b> <b>Ni</b> Níquel 58,693	<b>29</b> <b>Cu</b> Cobre 63,546	<b>30</b> <b>Zn</b> Zinco 65,380	<b>31</b> <b>Ga</b> Gálio 69,723	<b>32</b> <b>Ge</b> Germânio 72,630	<b>33</b> <b>As</b> Arsênio 74,922
5	<b>37</b> <b>Rb</b> Rubídio 85,468	<b>38</b> <b>Sr</b> Estrôncio 87,620	<b>39</b> <b>Y</b> Ítrio 88,906	<b>40</b> <b>Zr</b> Zircônio 91,224	<b>41</b> <b>Nb</b> Nióbio 92,906	<b>42</b> <b>Mo</b> Molibdênio 95,950	<b>43</b> <b>Tc</b> Técncio 97,000	<b>44</b> <b>Ru</b> Rutênio 101,070	<b>45</b> <b>Rh</b> Ródio 102,910	<b>46</b> <b>Pd</b> Paládio 106,420	<b>47</b> <b>Ag</b> Prata 107,870	<b>48</b> <b>Cd</b> Cádmio 112,410	<b>49</b> <b>In</b> Índio 114,820	<b>50</b> <b>Sn</b> Estanho 118,710	<b>51</b> <b>Sb</b> Antimônio 121,760
6	<b>55</b> <b>Cs</b> Césio 132,910	<b>56</b> <b>Ba</b> Bário 137,330	<b>57 - 71</b>	<b>72</b> <b>Hf</b> Háfio 178,490	<b>73</b> <b>Ta</b> Tântalo 180,950	<b>74</b> <b>W</b> Tungstênio 183,840	<b>75</b> <b>Re</b> Rênio 186,210	<b>76</b> <b>Os</b> Ósmio 190,230	<b>77</b> <b>Ir</b> Írídio 192,220	<b>78</b> <b>Pt</b> Platina 195,080	<b>79</b> <b>Au</b> Ouro 196,970	<b>80</b> <b>Hg</b> Mercúrio 200,590	<b>81</b> <b>Tl</b> Tálio 204,380	<b>82</b> <b>Pb</b> Chumbo 207,200	<b>83</b> <b>Bi</b> Bismuto 208,980
7	<b>102</b> <b>Fr</b> Frâncio 223,000	<b>102</b> <b>Ra</b> Rádio 226,000	<b>89 - 103</b>	<b>104</b> <b>Rf</b> Rutherfordio 267,000	<b>105</b> <b>Db</b> Dubnio 268,000	<b>106</b> <b>Sg</b> Seabórgio 269,000	<b>107</b> <b>Bh</b> Bohrio 270,000	<b>108</b> <b>Hs</b> Hássio 269,000	<b>109</b> <b>Mt</b> Meitnério 277,000	<b>110</b> <b>Ds</b> Darmashtio 281,000	<b>111</b> <b>Rg</b> Roentgenio 282,000	<b>112</b> <b>Cn</b> Copernício 285,000	<b>113</b> <b>Nh</b> Nihônio 286,000	<b>114</b> <b>Fl</b> Flavínio 290,000	<b>115</b> <b>Mc</b> Moscóvio 290,000
			<b>57 - 71</b>	<b>57</b> <b>La</b> Lantânio 138,910	<b>58</b> <b>Ce</b> Cério 140,120	<b>59</b> <b>Pr</b> Praseodímio 140,910	<b>60</b> <b>Nd</b> Neodímio 144,240	<b>61</b> <b>Pm</b> Promécio 145,000	<b>62</b> <b>Sm</b> Samário 150,360	<b>63</b> <b>Eu</b> Európio 151,960	<b>64</b> <b>Gd</b> Gadolínio 157,250	<b>65</b> <b>Tb</b> Térbio 158,930	<b>66</b> <b>Dy</b> Disprósio 162,500	<b>67</b> <b>Ho</b> Hólmio 164,930	<b>68</b> <b>Er</b> Érbio 167,260
			<b>89 - 103</b>	<b>89</b> <b>Ac</b> Actínio 227,000	<b>90</b> <b>Th</b> Tório 232,040	<b>91</b> <b>Pa</b> Protactínio 231,040	<b>92</b> <b>U</b> Urânio 238,030	<b>93</b> <b>Np</b> Neptúlio 237,000	<b>94</b> <b>Pu</b> Plutônio 244,000	<b>95</b> <b>Am</b> Améριο 243,000	<b>96</b> <b>Cm</b> Cúrio 247,000	<b>97</b> <b>Bk</b> Berquélio 247,000	<b>98</b> <b>Cf</b> Califórnio 251,000	<b>99</b> <b>Es</b> Einstênio 252,000	<b>100</b> <b>Fm</b> Férmio 257,000

Nº Atômico  
**Símbolo**  
 Nome  
 Massa

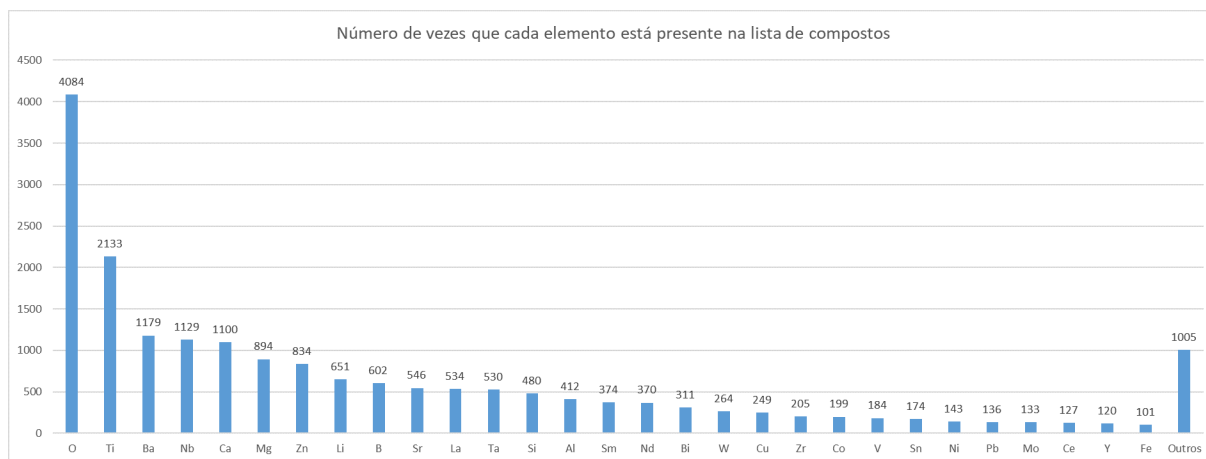
Elementos utilizados

Elementos não utilizados

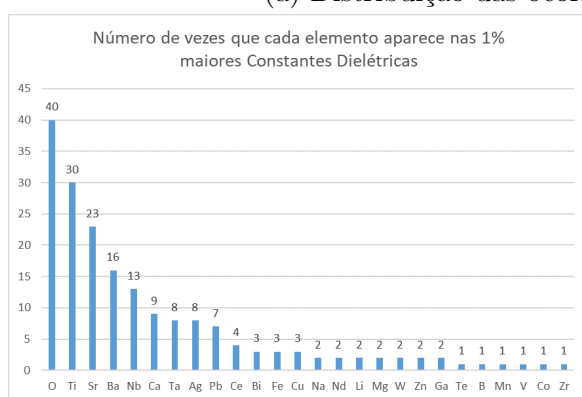
Figura A.1: Tabela Periódica com os 58 elementos químicos usados na base de dados para treinamento.

a maior Constante Dielétrica. Seguindo o padrão geral, temos o oxigênio, titânio, bário e nióbio nos 5 primeiros. Porém, em terceiro lugar encontramos o elemento estrôncio (Sr), sendo apenas o décimo terceiro colocado ao analisar todas as fórmulas químicas. O cálcio passa a ocupar a sexta posição, enquanto tântalo (Ta), prata (Ag) e chumbo (Pb) possuem destaque com mais de 5 ocorrências.

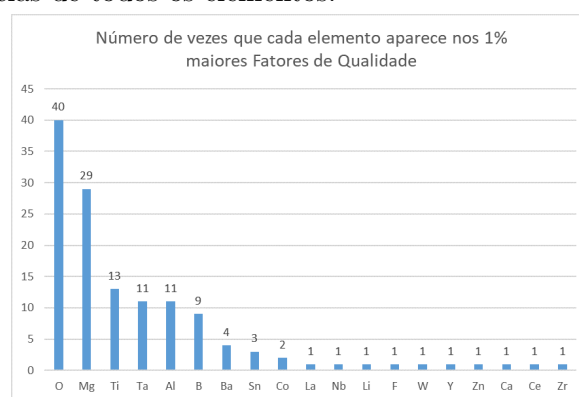
Ao analisarmos o Fator de Qualidade, observamos um comportamento um pouco diferente. A Figura A.2c mostra em segundo lugar o magnésio (Mg), aparecendo em mais da metade das composições que estão no topo. Titânio também aparece nessa lista em terceiro lugar, assim como está presente nas listas anteriores. O tântalo aparece em destaque no quarto lugar e o alumínio (Al) aparece pela primeira vez entre os 5 mais utilizados, na quinta colocação.



(a) Distribuição das ocorrências de todos os elementos.



(b) Distribuição dos elementos das 1% maiores Constantes Dielétricas.



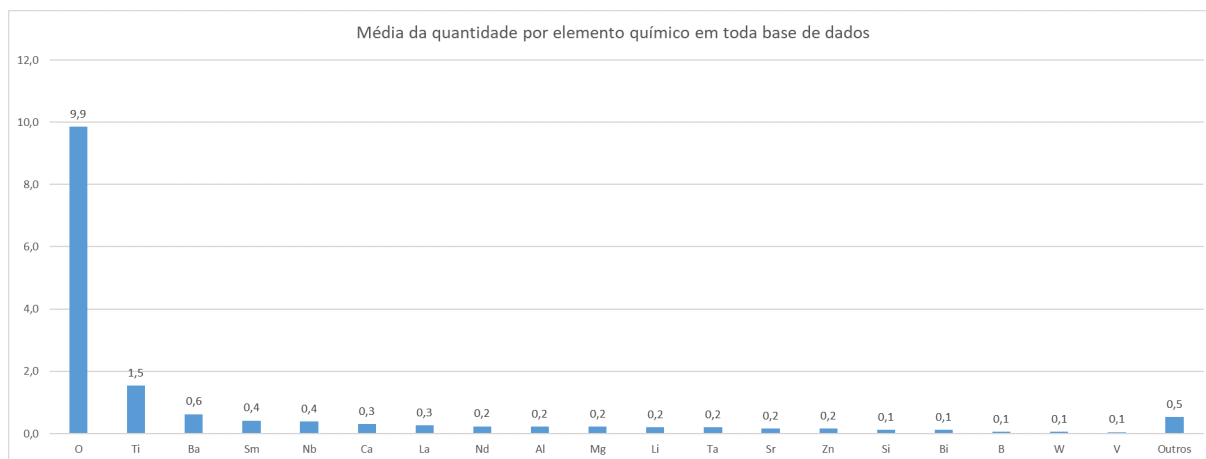
(c) Distribuição dos elementos dos 1% maiores Fatores de Qualidade.

Figura A.2: Distribuição dos elementos químicos.

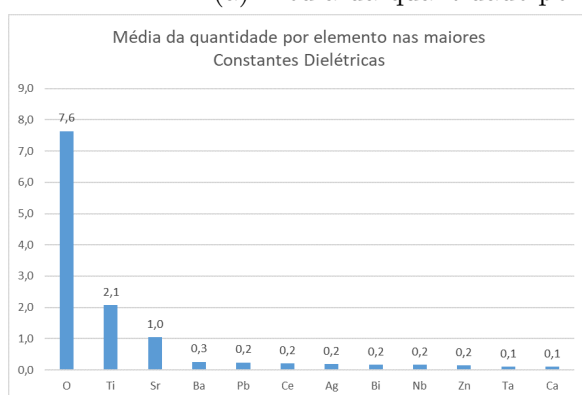
## A.2 Análise de Composição Média

A próxima análise tem o objetivo de elucidar as quantidades mais comuns presentes nos compostos químicos. Para além dos elementos mais comuns, é necessário verificar suas respectivas quantidades, fator importante para a base de dados adotada. A Figura A.3a exibe que o oxigênio é o elemento que aparece em maior quantidade, com um valor médio de 9,9 nas fórmulas químicas. Em segundo lugar temos o titânio com cerca de 1,5 em quantidade. Em terceiro o bário, logo em seguida o samário (Sm) que aparece pela primeira vez no topo, mostrando que apesar de não aparecer com tanta frequência em compostos, aparece em abundância quando aplicado. E completando os 5 primeiros temos o nióbio. Os que menos aparecerem em quantidade seguem quase o mesmo padrão do que foi apresentado pelo número de repetições. Em ordem de menor para maior quantidade temos: enxofre, rubídio, escândio e lutécio. Completando os 5 em menor quantidade, representando dopagens nos materiais, temos o carbono (C).

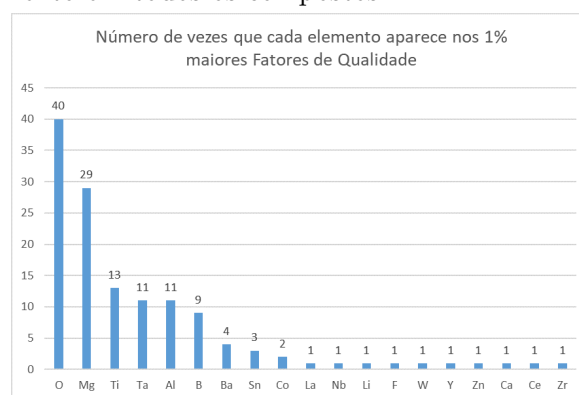
Analisando a Figura A.3b, podemos observar a utilização de cerca de 7 a 8 átomos



(a) Média da quantidade por elemento em todos os compostos.



(b) Média da quantidade por elemento das 1% maiores Constantes Dielétricas.



(c) Média da quantidade por elemento dos 1% maiores Fatores de Qualidade.

Figura A.3: Média das quantidades.

de oxigênio para obter os maiores valores de Constante Dielétrica. Assim como a aplicação de 2 átomos de titânio e 1 átomo de estrôncio. Dessa forma, como na análise pelo número de ocorrências, temos uma quantidade um pouco acima 1 aplicada nas fórmulas líderes em Constante Dielétrica. Fechando os 5 primeiros colocados temos o bário e o chumbo. A média encontrada para a Constante Dielétrica, nos 40 primeiros, é de aproximadamente 503. Existem outros elementos químicos utilizados nesses compostos, mas foram omitidos no gráfico para visualização e por apresentarem quantidades muito próximas a zero.

Já na Figura A.3c, observamos que o oxigênio é aplicado em menores quantidades para atingir maiores valores de Fator de Qualidade, tendo uma média de aproximadamente 3,7. Em segundo lugar, existe a aplicação de boa quantidade de magnésio nesses compostos. Seguido pelo alumínio, tântalo e titânio. Assim como o oxigênio aparece em maior quantidade para os compostos de maior Constante Dielétrica, o titânio segue o mesmo padrão. A média de Fator de Qualidade obtida para os 40 maiores foi 421960. Demais compostos com quantidades próximas a zero também foram ocultados para simplificação.

# Apêndice B

## Avaliação de pontos fora da curva (*outliers*)

No Capítulo 5 são apresentadas as comparações entre valores reais e valores previstos, divididas por quartis, para os dois modelos criados. O objetivo é apresentar algumas análises a partir da base de dados original, observando alguns padrões, materiais duplicados, elementos químicos específicos e subseções de fórmulas químicas de materiais.

A Tabela B.1 apresenta um conjunto de *outliers* para os valores previstos de Constante Dielétrica e a Tabela B.2 para o Fator de Qualidade. Foram escolhidas as três previsões de maior erro para cada quartil, totalizando 12 materiais de cada previsor. Todos os identificadores citados podem ser obtidos através da referência original.

### B.1 *Outliers* Constante Dielétrica

1. O elemento Ag (prata) é muito utilizado em materiais com alta Constante Dielétrica na faixa de 300 a 500. Exemplos de identificadores de materiais na referência original: 4112, 4113, 4114, 4119 e 4120;
2. Mesmo caso acima, elemento Ag presente na composição do material;
3. Outros materiais utilizando Ga em faixas mais elevadas como  $\text{La}_5\text{GaTi}_3\text{O}_{15}$ , com Constante Dielétrica no valor 35;
4. Possível utilização de  $\text{Al}_2\text{O}_3$ , presente em vários materiais de diferentes faixas de valores e também muitos materiais diferentes na composição;
5. Muitos materiais utilizando  $\text{SrTiO}_3$  com valores acima de 40;
6. Possivelmente influenciado por materiais com Nd (neodímio), assim como a combinação desses elementos estão presentes na casa de valores entre 60 a 70;

Tabela B.1: Principais outliers do modelo previsor para a Constante Dielétrica.

Identificador	Posição	Fator de Qualidade	Constante Dielétrica Real	Constante Dielétrica Prevista	Quartil	Material
1	280	33000	7,9	101,58	1	NaAgMoO4
2	304	17000	8,1	102,14	1	Ag2MoO4
3	740	234700	12,3	35,47	1	Sm3Ga5O12+TiO2
4	1081	800	17	47,00	2	30 vol% Al2O3+BaO-ZnO-SrOCaO-Nd2O3-TiO2-B2O3-SiO2 glass
5	1463	112000	21	46,22	2	0.6Mg4Nb2O9-0.4SrTiO3
6	1689	86300	23,6	66,67	2	Sr1-xNd1-xAl1-xTi04 (x=0.6)
7	2661	10800	36,6	72,95	3	0.7SrTiO3-0.3LaAlO3+10 wt% ZnO-B2O3
8	2928	10400	41	81,96	3	Bi12TiO20
9	3075	3600	44	149,30	3	Sr2Zn4Ti15O36
10	3598	2500	74	250,19	4	Ca0.8Sr0.2TiO3-Li0.5Sm0.5TiO3+x wt% [YB2O3-(1-y)CuO] (x=7.5, y=0.5)
11	3711	3000	80,4	250,19	4	Ca0.8Sr0.2TiO3-Li0.5Sm0.5TiO3+x wt% [YB2O3-(1-y)CuO] (x=7.5, y=0.5)
12	3785	3050	86	355,49	4	BaTi0.5Ga0.25Nb0.25O3

7. Possivelmente pela dopagem com ZnO–B<sub>2</sub>O<sub>3</sub>, que está presente em outros materiais cuja faixa de valores varia entre 5,9 e 70;
8. Inconclusivo. Possivelmente fórmulas com quantidades altas de oxigênio.
9. Existem materiais na faixa dos 150 que possuem estruturas semelhantes, utilizando elementos como Sr, Zn, Ti e abundância de oxigênio. Identificadores: 4088, 4047, 4068;
10. Materiais com estrutura semelhante e valores acima de 200, cujos identificadores são: 4092, 4098;
11. Mesmo caso acima, valores acima de 200 para compostos similares;
12. Aparece 2 vezes, uma com valor 86 e outra com 760 na posição 4125.

## B.2 *Outliers* Fator de Qualidade

1. Mesmo material aparece 5 vezes na base. Uma delas apresenta o valor de Fator de Qualidade de 148100;
2. Possui material igual (Mg<sub>2</sub>Al<sub>4</sub>Si<sub>5</sub>O<sub>18</sub> *Indialite recrystallized*) com valor 207800 e identificador 24;
3. Possivelmente materiais com quantidades próximas como Zn<sub>2</sub>SiO<sub>4</sub> sol-gel, cujo valor é 198400 influenciam no aumento do valor da previsão;
4. Inconclusivo. Talvez algum subconjunto de elementos esteja sendo utilizado em fórmulas com valores maiores;
5. Possivelmente influenciado pelo Al<sub>2</sub>O<sub>3</sub> que tem valores altos de Fator de Qualidade;

Tabela B.2: Principais outliers do modelo predictor para o Fator de Qualidade.

Identificador	Posição referência original	Constante Dielétrica	Fator de Qualidade Real	Fator de Qualidade Previsto	Quartil	Material
1	29	4,7	10600	65430	2	K0.9Ba0.1Ca1.1Ge2.9O8
2	100	6,2	40000	112434	3	Mg2Al4Si5O18 cordierite
3	124	6,5	5700	62310	1	(Zn1-xCox)2SiO4 (x=0.05)
4	360	8,5	50000	167160	3	Mg0.8Co0.2Al2O4
5	451	9,4	13700	83668	2	Al2O3+20 wt% CaSiO3
6	523	10	335000	576804	4	Al2O3
7	666	11,5	25520	119209	3	Mg4Nb0.5Ta.1.5O9
8	698	11,8	8000	71303	1	0.9Al2O3-0.1TiO2+0.3 wt% Nb2O5
9	1388	20,3	600000	284839	4	CeO2;(at 30K)
10	1536	22	5600	56000	1	Sr(Mg1/3Ta2/3)O3
11	3067	43,9	20200	80924	2	Ca4MgNb2TiO12-xCaTiO3 (x=0.3)
12	3395	23	65000	284839	4	CeO2

6. Aparece novamente na base, com identificador 516, com valor de 634000;
7. Inconclusivo. Talvez algum subconjunto de elementos esteja sendo utilizado em fórmulas com valores maiores;
8. O material  $0.9\text{Al}_2\text{O}_3-0.1\text{TiO}_2$  aparece quatro vezes na base sem dopagem. Possui valores como 148000 e 117000. Nesse caso a dopagem teve grande influência, mesmo com valores tão pequenos acrescentados no material.
9. Possui temperatura ambiente de medição. Mesma formula química aparece outras vezes na base, com valores menores.
10. Possivelmente materiais com quantidades próximas. Material com identificador 1716 possui valor de 430000;
11.  $\text{CaTiO}_3$  aparece em inúmeras dopagens, com valores muitas vezes entre 30 mil e 50 mil. Mas há casos como material identificador 955 (20% de  $\text{CaTiO}_3$ ), cujo valor é 390500. Existem 203 ocorrências de  $\text{CaTiO}_3$ ;
12. Existe esse mesmo material na base, porém com valor de 700 mil por estar resfriado a uma temperatura muito baixa. Mesmo material do item 9.