



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
COLEGIADO DO CURSO DE ENGENHARIA DE COMPUTAÇÃO

Gabriel dos Santos Sereno

**Classificadores especializados com seleção de
atributos SHAP: aplicação no processo
Tennessee Eastman e Reator-tanque Agitado
Contínuo (CSTR)**

Vitória, ES

2023

Gabriel dos Santos Sereno

**Classificadores especializados com seleção de atributos
SHAP: aplicação no processo Tennessee Eastman e
Reator-tanque Agitado Contínuo (CSTR)**

Dissertação apresentada ao Curso de Inteligência Artificial do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Inteligência Artificial.

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Colegiado do Curso de Engenharia de Computação

Orientador: Prof. Dr. Thomas Walter Rauber

Coorientador: Prof. Dr. Celso José Munaro

Vitória, ES

2023

Gabriel dos Santos Sereno

Classificadores especializados com seleção de atributos SHAP: aplicação no processo Tennessee Eastman e Reator-tanque Agitado Contínuo (CSTR)/ Gabriel dos Santos Sereno. – Vitória, ES, 2023-

34 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Thomas Walter Rauber

Monografia (PG) – Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Colegiado do Curso de Engenharia de Computação, 2023.

1. Diagnóstico de falhas. 2. Seleção de Características. 3. Seleção SHAP. 4. Processo Tennessee Eastman. 5. Classificadores especializados. I. Rauber, Thomas Walter. II. Universidade Federal do Espírito Santo. IV. Classificadores especializados com seleção de atributos SHAP: aplicação no processo Tennessee Eastman e Reator-tanque Agitado Contínuo (CSTR)

CDU 02:141:005.7

Gabriel dos Santos Sereno

Classificadores especializados com seleção de atributos SHAP: aplicação no processo Tennessee Eastman e Reator-tanque Agitado Contínuo (CSTR)

Dissertação apresentada ao Curso de Inteligência Artificial do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Inteligência Artificial.

Trabalho aprovado. Vitória, ES, 25 de setembro de 2023:

Prof. Dr. Thomas Walter Rauber
Orientador

Professor
Convidado 1

Professor
Convidado 2

Vitória, ES
2023

Agradecimentos

Primeiramente agradecer a Deus que me guiou e fez ser quem eu sou e sou muito grato a tudo o que aconteceu em minha vida. Agradeço também principalmente aos meus pais e ao meu irmão que sempre me apoiaram em tudo e me aconselharam em todas as minhas decisões. A minha esposa Maria que me apoiou e me deu forças para enfrentar todos os tipos de barreiras. Ao o meu professor Dr. Thomas Walter Rauber que me ensinou a ser um futuro mestre e pelo direcionamento para desenvolver esse estudo. Ao Dr. Celso José Munaro que também fez parte desse estudo, o meu muito obrigado. Por fim, agradeço a UFES pela oportunidade e também por toda estrutura para o apoio ao estudante.

“Quem não vive para servir não serve para viver.”
(Autor Desconhecido)

Resumo

Nesse estudo, classificadores especializados são treinados utilizando o método de seleção de característica SHAP no objetivo de identificar falhas nos processos industriais. Nesse contexto, os classificadores recebem dados de entradas a partir da decomposição do problema multi-classe utilizando a técnica One-vs-All. Essa metodologia foi empregada em dois benchmarks muito utilizados no meio academico: Processo Tennessee Eastman e o Reator-tanque Agitado Contínuo. O método SHAP obteve um desempenho de aproximadamente 89% na métrica F1-score no Processo Tennessee Eastman, selecionando em torno de 4 características em média por falha. Já no Reator-tanque Agitado Contínuo, o algoritmo obteve 90% na métrica F1-score e selecionou cerca de 3 características em média por falha. Apesar do estudo mostrar que o método Recursive Feature Elimination (RFE) obteve resultados parecidos em comparação ao SHAP em ambos os benchmarks. Entretanto, o método RFE tende a selecionar mais características para ter o mesmo desempenho. Por fim, o estudo sugere que o SHAP pode reduzir a dimensionalidade da base de dados com um bom desempenho na métrica F1-Score, podendo ser utilizado na área industrial e acadêmica.

Palavras-chaves: Diagnóstico de Falhas. Seleção de Características. SHAP. Processo Tennessee Eastman. Reator-tanque Agitado Contínuo. Classificadores Especializados.

Lista de ilustrações

Figura 1 – Fases do classificador	20
Figura 2 – A árvore de decisão e sua estrutura.	21
Figura 3 – O Random Forest e sua estrutura	21
Figura 4 – Dividindo os dados de treinamento multi-classe das flores para sub- bases de treinamento binário para cada flor	25

Lista de tabelas

Tabela 1 – Relação entre a falha e as características selecionadas pelo SHAP no CSTR (exceto a característica 13)	33
Tabela 2 – Relação entre a falha e as características selecionadas pelo SHAP no processo Tennessee Eastman	34

Sumário

1	INTRODUÇÃO	10
1.1	Objetivos	11
1.2	Estrutura da dissertação	12
2	REFERENCIAL TEÓRICO	13
2.1	Seleção de Características	13
2.1.1	Shapley additive explanations (SHAP)	13
2.1.2	Recursive Feature Elimination (RFE)	15
2.1.3	Selection from Model (SFM)	16
2.1.4	Mutual Information (MI)	17
2.1.5	ANOVA F-Value	18
2.2	Classificador	18
2.3	Decompondo conjuntos de dados multiclasse	22
2.3.1	One-vs-All (OVA)	22
2.3.2	One-vs-One (OVO)	23
3	METODOLOGIA DA CONSTRUÇÃO DO MODELO	24
3.1	Dados para os classificadores	24
3.2	Decomposição dos dados multi-classe	24
3.3	Seleção das características	25
3.4	Treinamento dos classificadores	25
3.5	Seleção do melhor classificador	26
4	CONSIDERAÇÕES FINAIS	27
	REFERÊNCIAS	28
	APÊNDICES	32

1 Introdução

Com o constante avanço tecnológico na área industrial, progressivamente o número de máquinas e processos para produzir um produto aumenta substancialmente (BRYNJOLFSSON; MCAFEE, 2011). Em decorrência disso, o custo de manutenção também cresce, tornando menos lucrativo todo o processo industrial.

Dessa forma, a predição, detecção e identificação de falhas torna-se de suma importância, já que com menos investimento pode-se prevenir gastos exorbitantes com consertos de máquinas caras que tem um papel importante no processo industrial (GARCÍA; PEDREGAL; ROBERTS, 2010).

A área de detecção e diagnóstico de falhas tem como objetivo em analisar componentes industriais em busca de falhas e determinar o instante que o ocorreu. Além disso, com a detecção dessa falha, o diagnóstico pode ser aplicado para investigar a causa do problema e também a sua magnitude (REPPA; TZES, 2011).

Nesse contexto, algoritmos que detectam previamente a ocorrência de uma falha requerem dados uteis para desempenhar sua função corretamente. Entretanto, uma máquina ou um processo industrial em sua complexidade. Portanto, a seleção dos melhores dados contribui para o aumento da taxa de detecção dessas falhas (KARABULUT; ÖZEL; IBRIKCI, 2012).

Desse modo, o presente estudo propõe a utilização do algoritmo Shapley Value Framework (SHAP), na qual seleciona as melhores características presentes em uma base de dados de acordo com a sua fórmula que calcula os valores Shapley (LUNDBERG; LEE, 2017). Esses valores são produzidos baseados no modelo, ou seja, o SHAP é um algoritmo post-hoc que utiliza os dados de treinamento gerados pelo modelo para calcular a importância das características com os valores Shapley. Esse algoritmo é focado principalmente na interpretação de modelos caixa-preta, em outras palavras, modelos onde não há conhecimento profundo sobre o seu funcionamento, baseando apenas nos dados de entrada e saída para sua interpretação.

Inicialmente, o SHAP foi criado para a Teoria de Jogos, com o objetivo de calcular o efeito da cooperatividade entre jogadores. Nisso, o estudo trás esse conceito para o meio industrial, para verificar sua efetividade em máquinas de processos químicos como o Tennessee Eastman e o Reator-tanque Agitado Contínuo em conjunto com o classificador Random Forest utilizando a técnica de decomposição, separando a base de dados em classes binárias para aumentar a eficácia da identificação das classes.

Além disso, como o SHAP gera valores de importância para cada característica

contida na base de dados baseado no classificador Random Forest, para cada máquina foi criado mapas de calor, bem como tabelas para determinar a relação entre a falha e as características. Com isso, é possível determinar onde está o problema mesmo se a falha não for conhecida, como acontece no processo Tennessee Eastman.

1.1 Objetivos

Este estudo tem como objetivo reduzir a dimensionalidade desses dados, na qual problemas multi-classe são decompostos em subproblemas de classe binária, bem como encontrar o conjunto de dados com o melhor desempenho nesse modelo. Com isso, este estudo utiliza o Shapley Value Framework (SHAP) como principal método de seleção de características em conjunto com o modelo Random Forest, além dos dados serem decompostos seguindo a técnica One-vs-All. Para fomentar o poder do algoritmo SHAP nesse modelo, foi feito um comparativo entre o SHAP e métodos de seleção de características que são relevantes no meio acadêmico e também industrial. São esses métodos: Recursive Feature Elimination (RFE), Selection from Model (SFM), Mutual Information (MI), e o ANOVA F-value.

Os algoritmos de seleção de características foram aplicados em duas bases de dados que são utilizadas como benchmark: Processo Tennessee Eastman e o Reator-tanque Agitado Contínuo. Ambos simulam máquinas em um processo químico, onde é possível alterar suas entradas no intuito de gerar falhas nos componentes.

A ideia do estudo é utilizar os resultados obtidos de ambas as bases para cada método de seleção de característica e analisar a eficiência do SHAP diante a metodologia proposta, tanto em seu desempenho como em sua redução de características utilizadas para detectar uma falha.

O estudo tem como objetivo secundário analisar os valores produzidos pelo SHAP e relacionar as características que tiveram maior importância para o algoritmo em comparação com a falha. Assim é possível verificar se a interpretabilidade é eficiente com o SHAP. Caso essa afirmação seja positiva, o SHAP pode apresentar grande potencial para área industrial, já que será possível verificar qual a intensidade e qual componente está afetando para produzir essa falha, portanto, aumentando a eficácia da manutenção dos equipamentos.

Com esse resultado podemos afirmar que o SHAP pode ser efetivo, principalmente, no meio industrial, já que os estudos iniciaram recentemente, aumentando a partir do ano de 2021. Outros autores fizeram a aplicação do SHAP em outros tipos de máquinas industriais, também com o objetivo de identificar quais componentes desses equipamentos apresentam algum defeito a partir dos resultados do SHAP. Por exemplo, ([BRITO et al., 2022](#)) utilizou o SHAP para identificar problemas em máquinas de alta rotação. Já

(UTAMA et al., 2023) utilizou o SHAP para explicar problemas em máquinas fotovoltaica.

1.2 Estrutura da dissertação

Este trabalho está dividido em mais quatro capítulos: No capítulo 2 é realizado uma revisão bibliográfica dos algoritmos de seleção de características, bem como o modelo Random Forest e as técnicas de decomposição. Já no capítulo 3 descreve a ideia e a construção do modelo, explicando o passo-a-passo de cada processo até o resultado. No capítulo ?? é apresentado a metodologia proposta aplicada a dois estudos de caso e seus resultados. E por fim, o capítulo 4 apresenta as considerações finais e também trabalhos futuros.

2 Referencial Teórico

Neste capítulo são apresentados os métodos e conceitos necessários para o desenvolvimento do estudo. A seção 2.1 detalha todos os métodos de seleção de características utilizados no comparativo para a demonstração dos resultados. A seção 2.2 explica o classificador utilizado no modelo proposto, bem como seu funcionamento interno. Por fim, a seção 2.3 faz um comparativo entre duas técnicas populares na literatura de decomposição de problemas multi-classe.

2.1 Seleção de Características

A seleção de características é basicamente formada por um critério de seleção e uma estratégia de busca. Um algoritmo simples seleciona um subconjunto das variáveis que descrevem o processo, preservando apenas as características mais importantes para o modelo, a fim de remover as características supérfluas e maximizar as pontuações de desempenho. O critério de desempenho pode ser a precisão do diagnóstico de falhas. No entanto, em um desequilíbrio extremo entre a disponibilidade de exemplos para o comportamento normal e de falha, métricas alternativas, como F_1 , podem ser mais apropriadas. Considere, por exemplo, um conjunto de dados em que uma em cada 100 amostras é uma falha. Nesse caso, uma precisão de 99% seria equivalente a um resultado de uma classificação aleatória.

2.1.1 Shapley additive explanations (SHAP)

O SHAP é um método baseado na teoria dos jogos para descrever o desempenho de um modelo de aprendizado de máquina. O principal objetivo dessa técnica é produzir um modelo interpretável e, em seguida, usar as características mais importantes detectadas que contribuíram principalmente para os resultados da classificação, especificamente no diagnóstico de falhas.

No artigo (LUNDBERG; LEE, 2017), o modelo interpretável é definido como f e o modelo de explicação como g . O objetivo é explicar o modelo interpretável com uma única entrada x . O modelo de explicação usa entradas simplificadas x' para mapear as entradas originais por meio de uma função de mapeamento:

$$x = h_x(x'), \quad (2.1)$$

onde h_x remapeia uma entrada simplificada para a entrada original, em outras palavras, converte um vetor binário de entradas interpretáveis para o espaço de entrada original.

Portanto, a função de mapeamento tenta garantir que

$$g(z') \approx f(h_x(z'))$$

sempre que,

$$z' \approx x',$$

onde z' é um subconjunto das entradas não nulas x' .

Além disso, o algoritmo se concentra em três princípios: precisão local, ausência de informações e consistência. A precisão local é a propriedade de aproximar o modelo original f para uma entrada específica x . Isso requer que o modelo de explicação g corresponda à saída $f(x)$ com a entrada simplificada x' , garantindo que essa saída seja a soma das atribuições de características, onde ocorre a multiplicação da entrada simplificada x' com o efeito ϕ , que ocorre quando a função de mapeamento 2.1 é usada.

Além disso, o modelo de explicação, sendo uma função linear de variáveis binárias que garante a precisão local, pode ser definido da seguinte forma:

$$f(x) = g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i, \quad (2.2)$$

onde $z' \in 0, 1^M$ é um vetor binário, com M sendo o número de características da entrada simplificada, e ϕ é o efeito para cada característica. Esse efeito é o valor de Shapley, sendo calculado a partir da equação 2.4, determinando se a característica é relevante ou não. Em seguida, pode ser comprovado que o modelo de explicação $g(x')$ corresponde ao modelo original $f(x)$ quando $x = h_x(x')$.

A propriedade de ausência de informações impede que a entrada original não tenha nenhum impacto após uma entrada simplificada com características ausentes ter sido submetida ao modelo. Essa propriedade é alcançada quando

$$x'_i = 0 \Rightarrow \phi_i = 0, \quad (2.3)$$

significando que uma entrada zero específica de uma característica x_i não causa nenhum efeito ϕ_i dessa característica.

Por fim, a consistência é alcançada quando um modelo altera f' de modo que uma entrada simplificada x' não mude ou diminua sua importância, independentemente de outras entradas. Se

$$f_x(z') = f(h_x(z')).$$

Então

$$z' = 0, \quad f'_x(z') - f'_x(z' \setminus i) \geq f_x(z') - f_x(z' \setminus i).$$

Isso implica para todas as entradas $z' \in 0, 1^M$, que

$$\phi_i(f', x) \geq \phi_i(f, x).$$

A única condição que suporta essa ideia é se:

$$\phi_i = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z' \cup \{i\}) - f_x(z')], \quad (2.4)$$

onde $|z'|$ representa o número de entradas não nulas em z' , e $z' \subseteq x'$ representa todos os vetores z' com um subconjunto de entradas não nulas x' . Portanto, ϕ_i da equação 2.4 são os valores de Shapley que medem a contribuição marginal de cada característica para o resultado previsto. Essa fórmula é calculada considerando todas as combinações possíveis de características e as diferenças no modelo. Além disso, o SHAP tem várias versões, como Kernel SHAP, Deep SHAP e Tree SHAP, que foram usadas neste artigo em conjunto com o modelo Random Forest.

2.1.2 Recursive Feature Elimination (RFE)

O RFE (Recursive Feature Elimination) é um método de seleção de característica baseado no sistema de ranqueamento de acordo com a importância de cada característica. Essa importância representa a quantificação da contribuição de cada característica para resolução do problema, em outras palavras, o classificador com as características mais importantes, tem grandes chances de determinar corretamente a classe do dado fornecido. O algoritmo inicialmente analisa todas as características e, em seguida, define em ordem decrescente o ranque de cada uma. Em seguida, identifica a característica mais fraca com base em métricas de avaliação do modelo (como F_1 -score) e a remove (GRANITTO et al., 2006). Como esse processo é iterativo, o modelo é ajustado novamente até atingir o número desejado de características a serem selecionadas, definido por hiperparâmetro. O procedimento de remoção iterativa é necessário porque, após a eliminação da característica, a importância de cada característica muda substancialmente, exigindo um recálculo (GUYON et al., 2002).

Algorithm 1 Recursive Feature Elimination (RFE)

Require: Base de dados \mathcal{D} com características \mathbf{X} , classificador \mathcal{C} , número de características para selecionar k

Ensure: Características selecionadas $\mathbf{X}_{\text{selecionado}}$

- 1: Inicializar $\mathbf{X}_{\text{selecionado}} \leftarrow \mathbf{X}$
 - 2: **while** Número de características em $\mathbf{X}_{\text{selecionado}}$ é maior que k **do**
 - 3: Treinar \mathcal{C} com $\mathbf{X}_{\text{selecionado}}$
 - 4: Computar as importâncias da características utilizando \mathcal{C}
 - 5: Identificar a característica com a menor importância e remove-la de $\mathbf{X}_{\text{selecionado}}$
 - 6: **end while**
 - 7: **return** $\mathbf{X}_{\text{selecionado}}$
-

As importâncias de cada característica podem ser calculadas de diversas formas, como por coeficientes, pesos e permutação. Os coeficientes são principalmente utilizados

em modelos lineares, como no estudo de (FONTI; BELITSER, 2017), já que esses coeficientes recebem uma característica para quantificar a mudança na variável alvo. Com isso, características que afetam fortemente essa variável tende a serem classificadas como mais importantes. Já na utilização de pesos, são unicamente utilizados em modelos de árvores, onde maiores pesos determinam características que dividem melhor os dados para classificar nos nós das árvores, como utilizado no estudo (ZHOU et al., 2021). E por fim a permutação, introduzido inicialmente por (BREIMAN; CUTLER, 2001), pode ser utilizado em diversos modelos, na qual consiste em modificar a ordem de apenas uma característica, deixando as demais estáticas. Caso o algoritmo identifique que o desempenho diminuiu drasticamente a partir de métricas como a acurácia, o RFE determina que essa característica é de suma importância para o modelo.

Por fim, o RFE como método de seleção de características tem como benefício a simplicidade e também a rapidez na seleção. Ideal para problemas que não requerem uma seleção de características tão complexas e que exigem performances melhores em troca do uso computacional e tempo. Além disso, o método é de simples interpretação, onde o utilizador consegue entender porque tal característica não influencia no atual problema proposto (CHEN; JEONG, 2007). Outro benefício do RFE pode ser utilizado por vários classificadores e dependendo do classificador escolhido, o resultado das características selecionadas pode se tornar diferente.

Portanto, devido ao bom desempenho nos estudos feitos por diversos autores como: (SHAO et al., 2017) que utilizou o método para selecionar melhores características na predição de valores a serem pagos em contas de energia, já (RICHHARIYA et al., 2020) que utilizou para selecionar melhores características no diagnóstico de Alzheimer e (ZHOU et al., 2021) que utilizou para selecionar melhores características no mapeamento de suscetibilidade, o RFE foi usado na comparação com o SHAP proposto neste estudo.

2.1.3 Selection from Model (SFM)

O Selection from Model é um algoritmo básico de seleção de características baseado em modelo criado pelos autores. Para isso, para cada característica foi atribuído uma importância de acordo com o modelo. Por exemplo, a importância de cada característica utilizando o modelo Random Forest são representados pelo ganho de informação, mais especificamente pelo ganho de impureza do índice de Gini de cada característica (BREIMAN et al., 1984). Dessa forma, com a atribuição dessa importância para cada característica, é feito o ranqueamento e seleção das N primeiras características para compor o novo conjunto de dados.

O principal benefício desse algoritmo é a sua simplicidade, por apenas utilizar os N primeiras características mais importantes, não requer quase nenhum poder computacional. Entretanto, a eficiência do método é ser comprometida por não haver nenhum

parâmetro no momento da seleção, apenas o número máximo de características informado pelo usuário.

Foi introduzido esse algoritmo como base no comparativo, permitindo a comparação da força entre um algoritmo que emprega cálculos e métodos na seleção.

2.1.4 Mutual Information (MI)

O Mutual Information foi descrito por (VERGARA, 2014) como um poderoso método de seleção de características que possui duas propriedades principais. Em primeiro lugar, qualquer tipo de relação entre variáveis aleatórias pode ser medida. Em segundo lugar, o Mutual Information preserva a ordem dos elementos originais, mesmo após a aplicação de uma transformação, como rotações ou translações de dados.

O Mutual Information estima a informação mútua entre duas características aleatórias, onde valores mais altos indicam uma dependência maior e um valor aproximadamente zero indica uma dependência menor. O método é baseado na estimativa de entropia a partir da distância dos k-vizinhos mais próximos, que foi originalmente proposto por (KRASKOV; STÖGBAUER; GRASSBERGER, 2011). Este método foi escolhido para ser comparado devido à quantificação da importância entre duas características aleatórias, o que pode resultar em resultados significativamente diferentes.

Os benefícios na utilização desse método de seleção de variáveis são focados na principal característica do Mutual Information, a análise da correlação entre os dados. O método consegue capturar relacionamentos não-lineares, tornando possível a identificação de dependências complexas que não são detectadas por métodos lineares. Além disso, é mais suscetível que o método selecione características é melhor em conjunto com outras características do que a característica individualmente. Por fim, o Mutual Information pode ser aplicado na grande maioria dos domínios, como classificação, regressão e clustering sem grandes modificações no método e nem no dado de entrada (BAO-GANG; YONG, 2008).

Entretanto, o método por ser poderoso ao estimar os relacionamentos através de cálculos matemáticos pode ser computacionalmente custoso com base de dados grandes ou com alta dimensionalidade. Esse problema é explorado neste estudo científico (SHAMS; BARNES, 2007), onde é utilizado NVIDIA CUDA para aprimorar o desempenho do algoritmo. Ademais, por focar extremamente na correlação dos dados, a seleção de características pode ser comprometida, pois o algoritmo não consegue definir se essa combinação é relevante ou não para o problema proposto. Por fim, o Mutual Information é limitado a verificar a correlação entre duas variáveis, não sendo possível identificar um conjunto maior que contem mais interações (BAO-GANG; YONG, 2008).

Portanto, além de entender os pontos fortes e fracos do método, foi verificado

estudos reais que utilizaram Mutual Information como um método de seleção de características. Como por exemplo, o estudo de (AMIRI et al., 2011) que utiliza o método integrando um modelo com o objetivo de detectar intrusões em sistemas. Além desse estudo, (VERRON; TIPLICA; KOBİ, 2008) utiliza o Mutual Information justamente no mesmo princípio deste estudo, a detecção e identificação de falhas no processo químico Tennessee Eastman. E por fim, a utilização do método em diagnósticos assistidos por computador (TOURASSI et al., 2001).

2.1.5 ANOVA F-Value

De acordo com (DOAN; JEONG; JI, 2020), o *ANOVA F-Value* é uma análise estatística que seleciona características com base na dependência de cada característica usando as estatísticas F. Essa dependência pode ser determinada calculando a variação em termos de categorias em relação às variações individuais. Uma abordagem sólida para remover características que não possuem importância é calcular a significância estatística, onde $p\text{-value} < 0.05$ ou $p\text{-value} < 0.01$, conforme relatado por (CAMPBELL et al., 2018).

Os benefícios de utilizar esse método são intrinsecamente relacionados ao cálculo do $p\text{-value}$ entre características, na qual determina se a característica será retirada ou não dependendo da sua significância estatística. Com isso, o ANOVA F-Value possui boa interpretabilidade, pois a partir das características em que o método determinou como significante, é possível concluir a razão da seleção para um determinado caso.

Embora o ANOVA F-Value tenha importantes qualidades, o método requer que os dados estejam normalmente distribuídos e que a variação dentro dos grupos sejam homogêneas e categóricas. Ao contrário do método Mutual Information, o ANOVA F-Value não considera potenciais relações entre as características, podendo remover importantes características que em conjunto, poderiam demonstrar melhor a classe correta.

O *ANOVA F-Value* foi escolhido para compor o comparativo pela sua relevância entre os estudos científicos como (DING et al., 2014) que utilizou o método na identificação de proteínas de bacteriófagos. Além desse estudo, (ELSSIED; IBRAHIM; OSMAN, 2014) utilizou o *ANOVA F-Value* como método de seleção de características para auxiliar na classificação de spam de e-mails. Por fim, (SHEIKHAN; BEJANI; GHARAVIAN, 2013) que utilizou o método para selecionar as melhores características para detectar a emoção através da voz. Portanto, o *ANOVA F-Value* é largamente utilizado no meio acadêmico e científico, além de ser reconhecido como um método eficaz para seleção de características.

2.2 Classificador

Em aprendizado de máquinas, um classificador é um algoritmo que recebe dados de entrada e o determina como uma dentre várias classes baseado no relacionamento dos da-

dos de treinamento, na qual foram separados antes da execução do algoritmo. O principal objetivo de um classificador é identificar a classe de um dado nunca utilizado anteriormente e o classificando através da generalização vinda do padrão aprendido durante a fase de treinamento (PEREIRA; MITCHELL; BOTVINICK, 2009). Nesse sentido, o classificador possui quatro fases (Figura 1) até a predição da classe (MOHAMMED; KHAN; BASHIER, 2016):

- **Treinamento:** Durante essa fase, é introduzido ao classificador uma base de dados que contém rótulos. Como por exemplo, um classificador que tem como objetivo identificar números escritos a mão, esses dados provavelmente serão o valor do pixel de uma imagem junto a um rótulo, que nesse caso será um número.
- **Aprendizado de Padrões:** Com a informação recebida dos dados de treinamento, o classificador utiliza diversas técnicas e fórmulas matemáticas para identificar padrões através do relacionamento entre os dados. Por exemplo, um classificador pode identificar os padrões do número 1, que pode ser escrito de diferentes formas por diversas pessoas, mas é reconhecível porque há um padrão na escrita desse número. Ao aprender com esses padrões durante a fase de treinamento, o classificador se torna capaz de generalizar e classificar corretamente novos dados nunca antes vistos.
- **Teste:** Depois do classificador treinado, é introduzido ao classificador uma base de dados de testes que não contém rótulos. Esses dados introduzidos nunca foram vistos pelo classificador durante o treinamento. Neste momento, é calculado principalmente a acurácia e outras métricas importantes na intenção de verificar quão correto o classificador avaliou os novos dados.
- **Predição:** Após todas essas etapas e ajustes no classificador, o classificador estará pronto para o funcionamento, identificando novos dados e rotulando-os de acordo com o padrão analisado.

Então, com um classificador é possível utilizar para diversas finalidades, como o estudo de (PAL, 2005) que utiliza classificadores para identificar em uma determinada área da zona rural, quais são os tipos de plantação através de imagens. Além disso, o estudo (ZHANG et al., 2018) utiliza classificadores que são modelos de Deep Learning para identificar a quantidade de objetos presente no ambiente em imagens, como árvores, construções, rios. E também para a finalidade de detectar e identificar falhas como nos trabalhos de (WANG et al., 2016), (NEUPANE; SEOK, 2020) e (SAUFI et al., 2019). Isso é apenas possível pois o classificador identifica padrões e a relação entre os dados.

Há diversos tipos de classificadores e neles incluem (AYODELE, 2010):

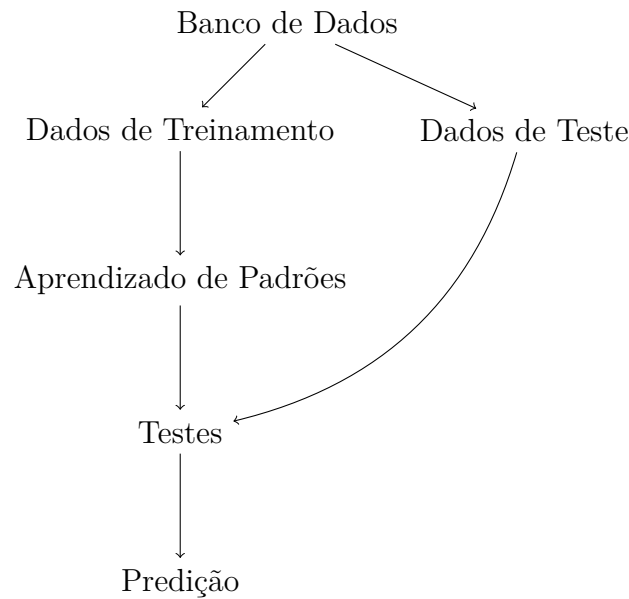


Figura 1 – Fases do classificador

- Regressão Linear: Utiliza a matemática da regressão linear para classificação binária através de uma reta trassada entre os dados e separando-os.
- Arvore de Decisão: Estruturas em formato de árvore que utiliza sequencia binária de decisões para classificar tanto problemas multi-classe tanto binário.
- Random Forest: Um conjunto de Arvores de Decisão, na qual utiliza o resultado da maioria das arvores como a classificação final.
- Naive Bayes: Baseado no teorema de Bayes e comumente utilizado para classificação de textos.
- Neural Networks: São modelos de Deep Learning em que lidam com padrões de dados complexos e em grande quantidade.

A escolha de um classificador depende de vários fatores, como a natureza do problema, a quantidade de dados disponível, o nível de interpretabilidade, o tempo de treinamento e demais fatores. Cada classificador possui pontos fortes e fracos, então é essencial o experimento e a escolha apropriada para a tarefa a ser resolvida (RAY, 2019).

Nesse trabalho, é necessário o uso de métodos de seleção de características que se baseiam no classificador. Por isso, o classificador Random Forest foi o mais apropriado, já que a partir dos dados coletados, o classificador se ajusta para escolher as melhores características para diagnosticar a classe do dado, além de ter a vantagem do modelo ser de fácil interpretação. Para o entendimento desse comportamento, a próxima subseção irá entrar em detalhes como o algoritmo de Random Forest opera.

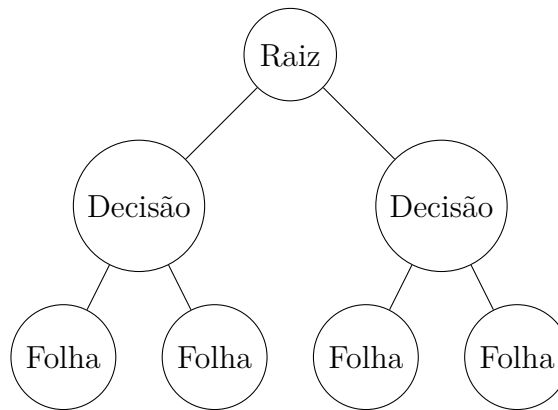


Figura 2 – A árvore de decisão e sua estrutura.

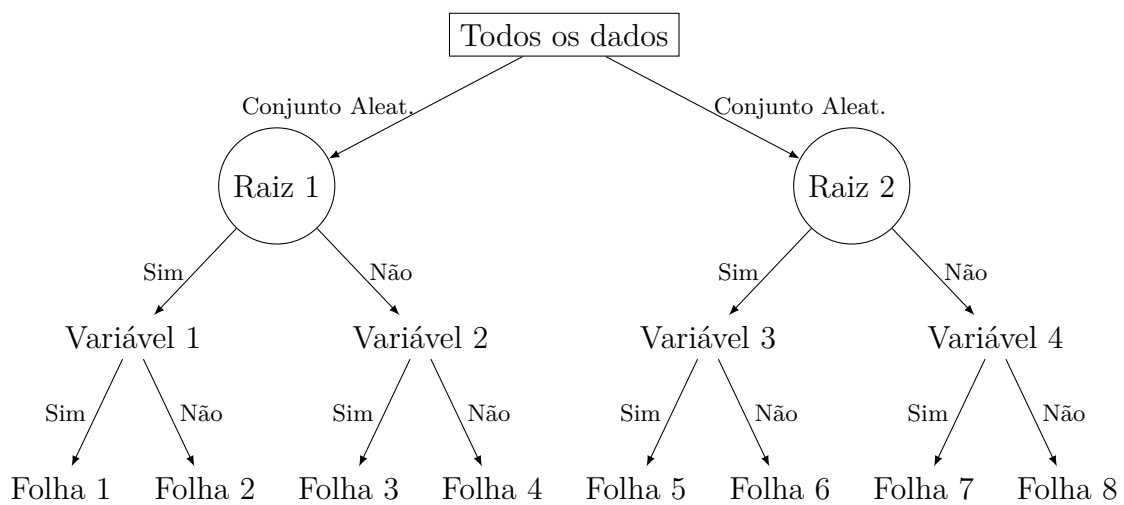


Figura 3 – O Random Forest e sua estrutura

O Random Forest é um classificador formado por um conjunto de árvores de decisão. Para o entendimento profundo do classificador Random Forest, precisamos entender primeiro as árvores de decisão. Uma árvore de decisão é um grafo acíclico direcionado em que cada nó é dividido por seus sucessores, podendo uma delas ser uma nó folha ou nó de divisão (MYLES et al., 2004), como mostra na Figura 2.

O nó de uma árvore de decisão, é o ponto onde contem um teste condicional, na qual determina os seus sucessores. Dessa forma, as árvores de decisão são fáceis de interpretar e visualizar. Além de poder classificar tanto dados categóricos e numéricos. Entretanto, as árvores de decisão podem sofrer com overfitting, especialmente quando é criado diversos nós, criando mais complexidade na árvore, pois os nós contem testes condicionais muito específicos, perdendo o atributo de generalização e portanto a taxa de acurácia. Uma forma de prevenir esse problema é efetuar a técnica de poda, onde propositalmente é removido esses nós para a árvore ser mais genérica.

Como mostra na Figura 3, o Random Forest é um conjunto de árvores de decisão. Cada árvore é treinado com uma porção aleatória dos dados e a predição dessas árvores são unidas para a decisão final.

O algoritmo Random Forest é altamente eficiente na identificação das complexidades presentes nos dados, permitindo que as árvores de decisão possuam uma maior capacidade de generalização. Ao contrário de uma única árvore de decisão que abrange todas as características, o Random Forest subdivide essas características entre várias árvores, permitindo que cada uma delas explique melhor as relações presentes nos dados.

Uma das principais vantagens do Random Forest é a melhora significativa na classificação dos dados. Isso é alcançado por meio da combinação das decisões de todas as árvores em um processo de votação, onde a classe mais frequente é selecionada como o resultado final. Essa abordagem coletiva garante uma maior robustez e precisão no processo de classificação, tornando o algoritmo uma poderosa ferramenta para diversas tarefas de aprendizado de máquina ([BREIMAN; CUTLER, 2001](#)).

2.3 Decompondo conjuntos de dados multiclasse

Em Aprendizado de Máquina, diversos cenários são extremamente complexos e que requer um ótimo classificador para identificar diversas classes de um banco de dados. Para isso, existe a técnica de decomposição, na qual particiona o problema em pequenas partes.

Em classificação multi-classe, a estratégia de decomposição é uma valiosa ferramenta para habilitar o uso de classificadores binários em situações onde anteriormente não eram viáveis devido à complexidade do problema. Com essa abordagem, torna-se possível testar múltiplos classificadores binários e verificar se algum deles contribui de forma mais eficaz para a identificação das diferentes classes ([ELKANO et al., 2016](#)).

Para lidar com essa limitação, existem dois tipos de métodos de decomposição que transformam problemas multi-classe em subproblemas binários, são eles One-vs-All (OVA) e One-vs-One (OVO). Para a escolha, é necessário entender os fatores em que uma técnica tem maiores benefícios que a outra, pois dependendo do número de classes, características da base de dados e recurso computacional, uma de ambas as técnicas pode sobressair e ter melhores resultados ([LI et al., 2020](#)).

2.3.1 One-vs-All (OVA)

Em OVA, para N classes, N classificadores binários são treinados, sendo que cada um tem como objetivo distintivo identificar uma classe específica em relação a todas as outras agrupadas. Por exemplo, dado um classificador binário especializado na classe A é treinado para identificar apenas dados da classe A como positiva e todas as demais negativas ([ZHANG et al., 2019](#)). Portanto, durante a predição dos dados, cada classificador gera a probabilidade exclusiva de sua classe, e, dessa forma, é verificado qual classe corresponde a uma determinada entrada com base no maior score obtido entre os classificadores.

2.3.2 One-vs-One (OVO)

Em OVO, para N classes, $N(N-1)/2$ classificadores binários são treinados, sendo que cada um tem como objetivo distintivo identificar uma classe específica em relação a outra. Por exemplo, durante o treinamento, um classificador binário especializado na classe A recebe como entrada a classe A e a classe B para serem treinados e identificar apenas a classe A como positiva e a classe B como negativa (PAWARA et al., 2020). Portanto, durante a predição dos dados, cada classificador faz sua própria decisão, e a classe com o maior voto dentre os classificadores é escolhido para ser a predição final.

3 Metodologia da construção do modelo

A metodologia apresentada aplica informações provenientes de máquinas de processamento químico, com objetivo de identificar quaisquer falhas no decorrer do processo.

Inicialmente, os dados são coletados da maquina, como o volume dos produtos químicos utilizados, temperatura, abertura de válvula e varias outras características. Com o controle de cada característica e modificando-as intencionalmente, é possível gerar falhas e assim, agrupar os dados gerados e rotula-los de acordo com a falha apresentada. Então, para cada falha rotulada nesse conjunto de dados, um classificador é especializado em identificar a falha selecionada. Esse conjunto de dados é introduzido em todos os classificadores, embora, a classe positiva é correspondente falha, permanecendo as demais classes como negativa.

Com o objetivo de diagnosticar as falhas, a metodologia proposta foca em características que mais se manifesta durante a ocorrência da falha e o identifica de acordo com o padrão dos dados. Portanto, foi utilizado selecionadores de características para realizar um comparativo, a respeito de qual deles detectam melhor essa variação quando está ocorrendo a falha.

A aplicação dessa proposta está descrito no Capítulo ??, onde foram aplicados em duas máquinas de processo químico para comparar tais selecionadores de características. Nas próximas subseções, é descrito como foi preparado os dados, selecionadores, classificadores e também todos outros detalhes relevantes para a construção completa do modelo.

3.1 Dados para os classificadores

Os dados para os classificadores proposto nessa metodologia é separado em duas bases entre treinamento e testes. Entretanto, não foram divididos igualmente, pois a base de treinamento requer mais dados para possibilite que os classificadores aprenderem o padrão dos dados e assim classifica-los. Portanto, os dados de teste servem para metrificar a eficiência de um determinado classificador com dados novos e nunca introduzidos no classificador.

3.2 Decomposição dos dados multi-classe

Após a divisão dos dados em bases de treinamento e de testes, ambos precisam ser decompostos, de acordo com a metodologia proposta. Para isso, deve-se modificar a

rotulação dos dados multi-classe para binário. Para isso, cada caso analisado neste presente estudo científico, foi criado, de acordo com o número de classes, bases iguais, mas com rótulos diferentes. Para melhor entendimento, caso o estudo fosse baseado nos dados das flores Iris, onde temos três tipos de flores, seriam feitos três novas bases para essas três classes, entretanto, a rotulação é diferente para cada base, como ilustra a figura 4. Com isso, para a primeira flor, os dados que correspondem a segunda e a terceira flor, seriam rotulados como dados negativos e para a primeira, positivos e assim por diante. Portanto, em cada estudo foram divididos em sub-bases binárias, tanto a base de treinamento, tanto a de teste para que cada classificador seja especializado em identificar apenas um tipo de falha.

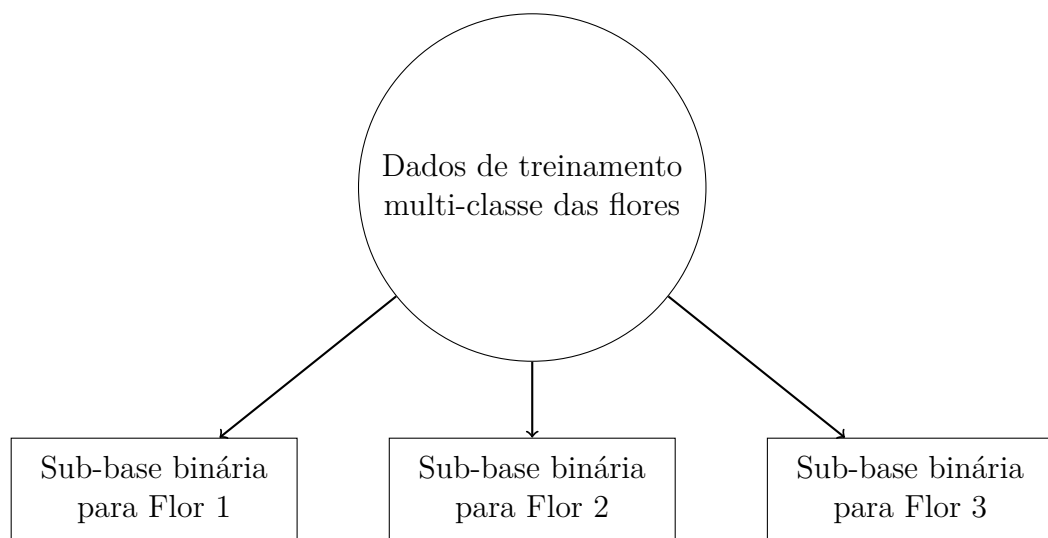


Figura 4 – Dividindo os dados de treinamento multi-classe das flores para sub-bases de treinamento binário para cada flor

3.3 Seleção das características

Com a base de treinamento pronta e sub-divida, é feito a seleção das características para cada classificador de cada falha. Portanto, as características são ranqueadas pelo nível de manifestação obtida através dos dados quando uma falha ocorre. Neste estudo científico, foi utilizado cinco métodos de seleção e estes receberam os mesmos dados sem modificações.

3.4 Treinamento dos classificadores

Depois da retirada de características que não contribuem para identificação, os classificadores são treinados com os dados de sua respectiva falha. Nesse caso, o classificador torna-se especialista da classe positiva que corresponde a falha selecionada, conseguindo identificar essa falha dentre as demais. Nesse processo, foi criado diversos classificadores

para a mesma falha, mas com a quantidade de características diferentes, seguindo o ranqueamento feito na seleção. Com isso, esse conjunto de classificadores terão diferentes taxas de acerto, mas apenas o melhor é selecionado.

3.5 Seleção do melhor classificador

Após o treinamento dos classificadores, é metrificado a acurácia com a base de testes. Nisso, são selecionados os melhores classificadores baseados em sua taxa de acerto para cada falha. Utilizando novamente como exemplo a base das flores, suponha que para a primeira flor, foi criado diversos classificadores binários especializados em identificar esse tipo de flor, mas com diferentes quantidades de características que foram utilizadas em seu treinamento. Entretanto, apenas um classificador foi escolhido para ser utilizado em identificar novas flores. Portanto, esse classificador é selecionado para compor o conjunto de classificadores que serão responsáveis em identificar as três flores.

4 Considerações Finais

Neste trabalho foi apresentado um estudo comparativo na área de diagnóstico de falhas em processos industriais, utilizando métodos de seleção de característica em conjunto com a técnica de decomposição One-vs-All e o classificador Random Forest.

A metodologia foi testada em benchmarks utilizados em diversos trabalhos científicos, como o Processo Tennessee Eastman e o Reator-tanque Agitado Contínuo com objetivo de evidenciar a eficácia do método de seleção de característica SHAP sobre os demais, bem como demonstrar que o método pode ser aplicado não somente na área de jogos, mas também na área industrial e acadêmica. Ao decorrer da apresentação do trabalho, evidencia-se que o SHAP obteve um dos melhores desempenhos, selecionando menos característica. Além disso, o SHAP é um algoritmo poderoso e fácil de ser aplicado em classificadores suportados, sem que os dados originais sejam alterados.

As limitações na identificação apresentadas em algumas falhas ocorreram devido aos algoritmos de seleção escolheram características similares, tanto no Processo Tennessee Eastman e o Reator-tanque Agitado Contínuo. Isso não ocorre devido ao método de seleção, mas pelo classificador. Esse fato evidencia-se na apresentação dos resultados, onde todos os métodos obtiveram pontuações aproximados.

Portanto, o SHAP demonstra que é capaz de operar além na área de jogos, mas também na área industrial e acadêmica, onde pode ser mais uma opção simples para aumentar a precisão do classificador, bem como explicar as suas decisões através das características que se manifestaram durante a ocorrência da falha. Além disso, o uso de dados se torna mais eficiente, reduzindo a complexidade e tamanho da base dados.

Para o aprimoramento da metodologia, recomenda-se como trabalho futuro utilizar outros tipos de modelo, como técnicas de Deep Learning para prevenir que o classificador tenham dificuldades na identificação onde a base de dados contem classes com características extremamente similares. Além disso, recomenda-se a utilização de outros métodos de seleção de característica para comparação, já que os utilizados neste estudo são classificadores clássicos da literatura, podendo abranger classificadores mais novos ou criados por outros autores. Para abranger mais o estudo comparativo, pode ser empregado a técnica One-vs-One com objetivo de verificar se houve melhora no desempenho com a técnica mais complexa. Por fim, a metodologia pode ser aplicada a outros benchmarks ou até mesmo em um problema real para testar a eficácia do algoritmo de uma forma mais ampla.

Referências

- AMIRI, F. et al. Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications*, Elsevier, v. 34, n. 4, p. 1184–1199, 2011. Citado na página 18.
- AYODELE, T. O. Types of machine learning algorithms. *New advances in machine learning*, IntechOpen Rijeka, v. 3, p. 19–48, 2010. Citado na página 19.
- BAO-GANG, H.; YONG, W. Evaluation criteria based on mutual information for classifications including rejected class. *Acta Automatica Sinica*, Elsevier, v. 34, n. 11, p. 1396–1403, 2008. Citado na página 17.
- BREIMAN, L.; CUTLER, R. Random forests machine learning [j]. *journal of clinical microbiology*, v. 2, p. 199–228, 2001. Citado 2 vezes nas páginas 16 e 22.
- BREIMAN, L. et al. *Classification and Regression Trees*. [S.l.]: Taylor & Francis, 1984. (The Wadsworth and Brooks-Cole statistics-probability series). ISBN 9780412048418. Citado na página 16.
- BRITO, L. C. et al. An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. *Mechanical Systems and Signal Processing*, Elsevier, v. 163, p. 108105, 2022. Citado na página 11.
- BRYNJOLFSSON, E.; MCAFEE, A. *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. [S.l.]: Brynjolfsson and McAfee, 2011. Citado na página 10.
- CAMPBELL, Z. et al. Differentially private anova testing. In: IEEE. *2018 1st International Conference on Data Intelligence and Security (ICDIS)*. [S.l.], 2018. p. 281–285. Citado na página 18.
- CHEN, X.-w.; JEONG, J. C. Enhanced recursive feature elimination. In: IEEE. *Sixth international conference on machine learning and applications (ICMLA 2007)*. [S.l.], 2007. p. 429–435. Citado na página 16.
- DING, H. et al. Identification of bacteriophage virion proteins by the anova feature selection and analysis. *Molecular BioSystems*, Royal Society of Chemistry, v. 10, n. 8, p. 2229–2235, 2014. Citado na página 18.
- DOAN, D. M.; JEONG, D. H.; JI, S.-Y. Designing a feature selection technique for analyzing mixed data. In: IEEE. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. [S.l.], 2020. p. 0046–0052. Citado na página 18.
- ELKANO, M. et al. Fuzzy rule-based classification systems for multi-class problems using binary decomposition strategies: on the influence of n-dimensional overlap functions in the fuzzy reasoning method. *Information Sciences*, Elsevier, v. 332, p. 94–114, 2016. Citado na página 22.

- ELSSIED, N. O. F.; IBRAHIM, O.; OSMAN, A. H. A novel feature selection based on one-way anova f-test for e-mail spam classification. *Research Journal of Applied Sciences, Engineering and Technology*, Maxwell Scientific Publications, v. 7, n. 3, p. 625–638, 2014. Citado na página 18.
- FONTI, V.; BELITSER, E. Feature selection using lasso. *VU Amsterdam research paper in business analytics*, Business Analytics Master Amsterdam, The Netherlands, v. 30, p. 1–25, 2017. Citado na página 16.
- GARCÍA, F. P.; PEDREGAL, D. J.; ROBERTS, C. Time series methods applied to failure prediction and detection. *Reliability Engineering & System Safety*, Elsevier, v. 95, n. 6, p. 698–703, 2010. Citado na página 10.
- GRANITTO, P. M. et al. Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products. *Chemometrics and intelligent laboratory systems*, Elsevier, v. 83, n. 2, p. 83–90, 2006. Citado na página 15.
- GUYON, I. et al. Mach learn 46: 389. doi: 10.1023. A: 1012487302797, 2002. Citado na página 15.
- KARABULUT, E. M.; ÖZEL, S. A.; IBRIKCI, T. A comparative study on the effect of feature selection on classification accuracy. *Procedia Technology*, Elsevier, v. 1, p. 323–327, 2012. Citado na página 10.
- KRASKOV, A.; STÖGBAUER, H.; GRASSBERGER, P. Erratum: estimating mutual information [phys. rev. e 69, 066138 (2004)]. *Physical Review E*, APS, v. 83, n. 1, p. 019903, 2011. Citado na página 17.
- LI, Q. et al. Multiclass imbalanced learning with one-versus-one decomposition and spectral clustering. *Expert Systems with Applications*, Elsevier, v. 147, p. 113152, 2020. Citado na página 22.
- LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, v. 30, 2017. Citado 2 vezes nas páginas 10 e 13.
- MOHAMMED, M.; KHAN, M. B.; BASHIER, E. B. M. *Machine learning: algorithms and applications*. [S.l.]: Crc Press, 2016. Citado na página 19.
- MYLES, A. J. et al. An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, Wiley Online Library, v. 18, n. 6, p. 275–285, 2004. Citado na página 21.
- NEUPANE, D.; SEOK, J. Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *IEEE Access*, IEEE, v. 8, p. 93155–93178, 2020. Citado na página 19.
- PAL, M. Random forest classifier for remote sensing classification. *International journal of remote sensing*, Taylor & Francis, v. 26, n. 1, p. 217–222, 2005. Citado na página 19.
- PAWARA, P. et al. One-vs-one classification for deep neural networks. *Pattern Recognition*, Elsevier, v. 108, p. 107528, 2020. Citado na página 23.

- PEREIRA, F.; MITCHELL, T.; BOTVINICK, M. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, Elsevier, v. 45, n. 1, p. S199–S209, 2009. Citado na página 19.
- RAY, S. A quick review of machine learning algorithms. In: IEEE. *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. [S.l.], 2019. p. 35–39. Citado na página 20.
- REPPA, V.; TZES, A. Fault detection and diagnosis based on parameter set estimation. *IET control theory & applications*, IET, v. 5, n. 1, p. 69–83, 2011. Citado na página 10.
- RICHHARIYA, B. et al. Diagnosis of alzheimer’s disease using universum support vector machine based recursive feature elimination (usvm-rfe). *Biomedical Signal Processing and Control*, Elsevier, v. 59, p. 101903, 2020. Citado na página 16.
- SAUFI, S. R. et al. Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *Ieee Access*, IEEE, v. 7, p. 122644–122662, 2019. Citado na página 19.
- SHAMS, R.; BARNES, N. Speeding up mutual information computation using nvidia cuda hardware. In: IEEE. *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*. [S.l.], 2007. p. 555–560. Citado na página 17.
- SHAO, Z. et al. A new electricity price prediction strategy using mutual information-based svm-rfe classification. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 70, p. 330–341, 2017. Citado na página 16.
- SHEIKHAN, M.; BEJANI, M.; GHARAVIAN, D. Modular neural-svm scheme for speech emotion recognition using anova feature selection method. *Neural Computing and Applications*, Springer, v. 23, p. 215–227, 2013. Citado na página 18.
- TOURASSI, G. D. et al. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical physics*, Wiley Online Library, v. 28, n. 12, p. 2394–2402, 2001. Citado na página 18.
- UTAMA, C. et al. Explainable artificial intelligence for photovoltaic fault detection: A comparison of instruments. *Solar Energy*, Elsevier, v. 249, p. 139–151, 2023. Citado na página 12.
- VERGARA, E. Vergara jr, estévez pa. *A review of feature selection methods based on mutual information*, *Neural Computing and Applications*, v. 24, n. 1, p. 175–186, 2014. Citado na página 17.
- VERRON, S.; TIPLICA, T.; KOBI, A. Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control*, Elsevier, v. 18, n. 5, p. 479–490, 2008. Citado na página 18.
- WANG, L. et al. Wind turbine gearbox failure identification with deep neural networks. *IEEE Transactions on Industrial Informatics*, IEEE, v. 13, n. 3, p. 1360–1368, 2016. Citado na página 19.
- ZHANG, C. et al. Multi-imbalance: An open-source software for multi-class imbalance

learning. *Knowledge-Based Systems*, Elsevier, v. 174, p. 137–143, 2019. Citado na página [22](#).

ZHANG, C. et al. A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 140, p. 133–144, 2018. Citado na página [19](#).

ZHOU, X. et al. Landslide susceptibility mapping using hybrid random forest with geodetector and rfe for factor optimization. *Geoscience Frontiers*, Elsevier, v. 12, n. 5, p. 101211, 2021. Citado na página [16](#).

Apêndices

Tabela 1 – Relação entre a falha e as características selecionadas pelo SHAP no CSTR (exceto a característica 13)

Falha	Descrição da Falha	Características selecionadas
2	Obstrução na saída do tanque	Contro. de Nível de Saída Tempe. da E. do Líquido de Refri.
3	Obstrução no revestimento	Ponto de Ajuste do Líquido de Refri. Pressão da Entrada do Líquido de Refri.
4	Vazamento do revestimento para o ambiente	Ponto de Ajuste do Líquido de Refri. Vazão do Líquido de Refrigeração
5	Vazamento do revestimento para o tanque	Vazão do Líquido de Refrigeração
6	Vazamento da bomba	Contro. de Nível de Saída Vazão do Produto
7	Perda de pressão na bomba	Contro. de Nível de Saída
8	Incrustação na superfície de troca de calor do revestimento	Temperatura do Reator Ponto de Ajuste do Líquido de Refri.
10	Energia de ativação da reação primária	Concentração do Produto A Ponto de Ajuste do Líquido de Refri.
12	Vazão de alimentação anormal	Ponto de Ajuste do Líquido de Refri. Contro. de Nível de Saída
13	Temperatura de alimentação anormal	Temperatura da entrada Ponto de Ajuste do Líquido de Refri.
14	Concentração de alimentação anormal	Ponto de Ajuste do Líquido de Refri. Vazão do Líquido de Refrigeração
15	Temperatura anormal da água de resfriamento	Ponto de Ajuste do Líquido de Refri.
16	Pressão anormal da água de resfriamento	Pressão da Entrada do Líquido de Refri. Concentração de entrada
19	Ponto de ajuste anormal do controlador de nível	Nível do reator Concentração do Produto A

Tabela 2 – Relação entre a falha e as características selecionadas pelo SHAP no processo Tennessee Eastman

Falha	Descrição da Falha	Características selecionadas
1	Razão de alimentação A/C	Alimentação do reagente A (fluxo 1) Vazão de reagente A (fluxo 1)
2	Composição de B	Taxa de purga (fluxo 9) Válvula de purga (fluxo 9)
3	Temperatura de alimentação de D	Válvula de vapor do removedor Componente E
4	Temperatura de entrada do refrigerante do reator	Vazão de refrigerante do reator Temperatura do removedor
5	Temperatura de entrada do refrigerante do condensador	Vazão de refrigerante do condensador Fluxo do removedor (fluxo 11)
6	Perda de alimentação de A	Vazão de reagente A (fluxo 1)
7	Redução na disponibilidade de C	Vazão dos reagentes A e C (fluxo 4) Válvula de vapor do removedor
8	Composição de alimentação de A, B e C	Válvula de vapor do removedor Componente E
9	Temperatura de alimentação de D	Válvula de vapor do removedor Componente E
10	Temperatura de alimentação de C	Vazão de gás do removedor Válvula de vapor do removedor
11	Temperatura de entrada do refrigerante do reator	Vazão de refrigerante do reator Temperatura do reator
12	Temperatura de entrada do refrigerante do condensador	Válvula de vapor do removedor Componente E
13	Cinética das reações	Vazão de gás do removedor Válvula de vapor do removedor
14	Válvula do Reator da Água de Refrigeração	Temperatura do reator Temperatura de saída do refrigerante do reator
15	Válvula do Condensador da Água de Refrigeração	Temperatura do removedor Vazão de gás do removedor
16	Desconhecido	Válvula de vapor do removedor Vazão de gás do removedor
17	Desconhecido	Temperatura de saída do refrigerante do reator Temperatura do reator
18	Desconhecido	Temperatura de saída do refrigerante do separador Válvula de vapor do removedor
19	Desconhecido	Vazão de reciclagem (fluxo 8) Válvula do compressor
20	Desconhecido	Válvula do compressor Pressão do separador
21	Válvula (Stream 4)	Vazão dos reagentes A e C (fluxo 4)