

IMPLEMENTAÇÃO DE INTELIGÊNCIA ARTIFICIAL UTILIZANDO MAQUINAS DE SUPORTE DE VETORES (SVM) NA CLASSIFICAÇÃO DE VIDROS

Marcelo Pontes Rodrigues¹; Marcelo de Oliveira Rego²

¹ Programa de Pós-Graduação em Tecnologias Aplicadas a Animais de Interesse Regional–UFPI;

² Programa de Doutorado em Ciência da Computação Associação UFMA/UFPI/DCCMAPI/CCET

¹marcelo.rodrigues@ufpi.edu.br; ²oliveira@ufdpar.edu.br;

Resumo

A pesquisa sugeriu uma abordagem para classificação de tipos de vidro utilizando o conjunto de dados "Glass Identification" disponível na UCI (University of California, Irvine) (German 1987). O objetivo é identificar a composição de vidro com base em variáveis químicas e físicas, diferenciando entre tipos comuns e especiais. Para isso, o dataset foi pré-processado e normalizado, e foram projetadas Maquinas de Suporte de Vetores (SVM) utilizando as funções de Kernel, onde foram treinadas para distinguir entre as classes. Para isso, foram abordadas neste trabalho quatro hipóteses todas considerando dados normalizados pelo método MinMaxScaler: (i) Hipótese 1, dados com as 7 classes e todos os atributos; (ii) Hipótese 2.1, reduzido a 3 classes e todos os atributos e Hipótese 2.2, problema reduzido a 3 classes e retirando a coluna "Ca"; (iii) Hipótese 3, divisão em Flutuado e Não Flutuado excluindo a classe "Outros" sem retirar atributos. Com isso, culminou em um modelo eficaz com a Hipótese 3, com Acurácia 88,64%. Os achados sugerem que SVMs são ferramentas promissoras para a análise composicional de materiais.

Palavras-chave: vidro, classificação, machine, learning, aprendizado, máquina, vetores e suporte.

INTRODUÇÃO

O objetivo desta pesquisa é desenvolver um modelo de classificação de vidros com base em sua composição química. Utilizou-se o conjunto de dados "Glass Identification" do repositório UCI Machine Learning (German 1987), que contém informações detalhadas sobre a composição de diferentes tipos de vidros, como os usados em janelas e em ambientes especiais. Originalmente criado para fornecer uma base confiável e padronizada para investigações forenses e estudos em engenharia de materiais, este conjunto de dados permite a classificação automatizada de amostras, o que pode auxiliar na identificação rápida e precisa de materiais, economizando tempo e recursos. Neste estudo, exploramos máquinas de suporte de vetores (SVM) para a classificação dos tipos de vidro, com base nas propriedades químicas e físicas do material. O uso de SVMs é especialmente adequado, pois essas máquinas são capazes de detectar padrões complexos e não lineares em dados multidimensionais (Cortes e Vapnik 1995), uma característica essencial para analisar a diversidade nas composições químicas deste dataset.

METODOLOGIA

Inicialmente foi realizado uma avaliação dos dados e a correlação das features para auxiliar nas opções a serem escolhidas no modelo de SVMs e definir as hipóteses, conforme Figura 1.

O estudo foi conduzido de forma abrangente, revisitando cada etapa diversas vezes para avaliar, ajustar e validar as hipóteses propostas. Em cada iteração, o tamanho do conjunto de testes foi mantido constante em 30% para efeitos de comparação, e análises detalhadas foram realizadas para decidir sobre a aceitação ou rejeição das hipóteses. Foram abordadas neste trabalho quatro hipóteses:

- (i) Hipótese 1, dados com as 7 classes e todos os atributos;
- (ii) Hipótese 2 foi dividida em Hipótese 2.1, reduzindo a 3 classes e todos os atributos; e Hipótese 2.2, problema reduzido a 3 classes e retirando a coluna "Ca"; e
- (iii) Hipótese 3, divisão em Flutuado e Não Flutuado excluindo a classe "Outros" sem retirada de atributos.

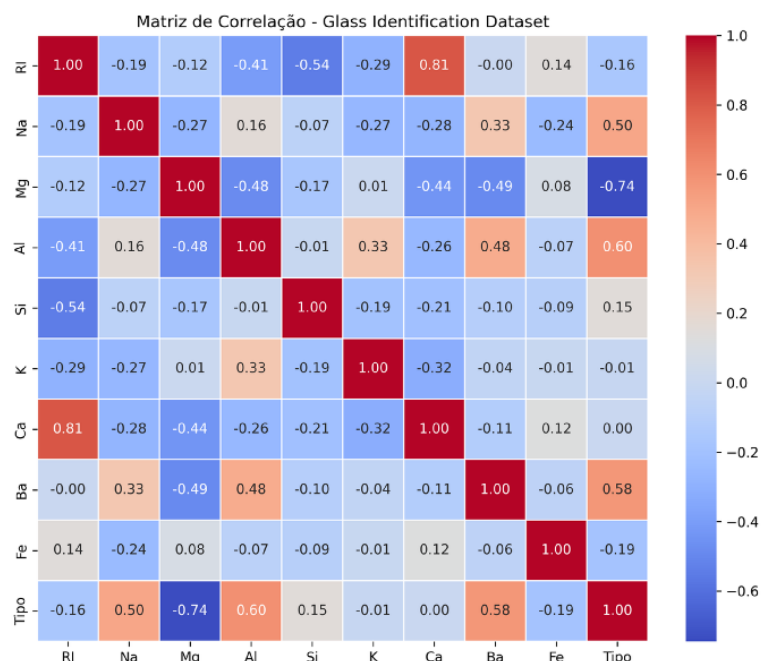


Figura 1: Matriz de Correlação

O estudo foi implementado na linguagem python e está disponível no repositório do GitHub (Pontes e Rego 2024)

RESULTADOS E DISCUSSÕES

Os resultados obtidos após a aplicação das diferentes hipóteses e ajustes de parâmetros demonstraram variações significativas nas métricas de desempenho. Percebemos que com 30% dos dados sendo destinados para o teste obtivemos as melhores métricas, portanto fizemos todos os testes subsequentes já com essa proporção, também fizemos uso de `class_weight=Balanced`, que ajusta o peso atribuído a cada classe durante o treinamento para lidar com dados desbalanceados, onde uma classe possui significativamente mais exemplos que outra. Isso ajuda o modelo a tratar todas as classes de maneira mais justa, evitando que ele favoreça a classe majoritária.

Na Tabela 1 abaixo expomos as hipóteses, hiperparâmetros usados bem como a acurácia de cada modelo.

Hipótese	Kernel	C	Gamma	Degree	Acurácia (%)
1	Poly	100	Scale	3	78,46
	RBF	100	Scale	-	78,46
2.1	Poly	100	1	5	76,92
	RBF	100	3	-	76,92
2.2	Poly	100	4	4	73,85
	RBF	100	3	-	73,85
3	Poly	40	Scale	3	88,64
	RBF	50	3	-	84,09

Tabela 1: Resultados das hipóteses testadas com diferentes configurações de kernel e acurácias.

CONCLUSÃO

O estudo confirmou a eficácia de abordagens específicas para otimizar a precisão e robustez do modelo de classificação. A Hipótese 3, divisão em Flutuado e Não Flutuado excluindo a classe "Outros" foi que apresentou melhor acurácia 88,64%. Esses resultados indicam que o SVM obteve melhores resultados sem simplificação do conjunto de variáveis como demonstrado nas hipóteses 2.1 e 2.2.

Referências

CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, Springer, v. 20, n. 3, p. 273–297, 1995.

GERMAN, B. *Glass Identification*. 1987. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5WW2P>.

PONTES, M.; REGO, M. *Repositório de Inteligência Artificial utilizando Árvore de Decisão no Diagnóstico Câncer de Mama*. 2024. Acessado em: 21 de novembro de 2024. Disponível em: <https://github.com/infopontes/inteligencia_artificial/>.