XXXIII Congresso de Iniciação Científica da Unicamp

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE RESINAS PLÁSTICAS ATRAVÉS DE SINAIS DE ÁUDIO

Palavras-Chave: RESINAS PLÁSTICAS, APRENDIZADO DE MÁQUINA, PROCESSAMENTO DE SINAIS

Autores(as):
ANNA CAROLINA RUTSATZ BARTZ, FEQ, UNICAMP
Drª. ANA CLÁUDIA OLIVEIRA E SOUZA, FEQ, UNICAMP
Prof. Dr. FLÁVIO VASCONCELOS DA SILVA, FEQ, UNICAMP

INTRODUÇÃO

"O plástico é a maior descoberta do milênio" (Pilapitiya e Ratnayake, 2024). Esta frase é, sem dúvidas, inegável, dada a versatilidade e os avanços tecnológicos que os plásticos permitiram à sociedade. Plásticos rígidos, maleáveis, termofixos, entre outros, permitem o avanço da pesquisa científica e da produção de bens que são hoje essenciais para a nossa sobrevivência, como carros, aviões, celulares, casas e utensílios domésticos. Entretanto, o consumo destes polímeros aumentou cerca de 180 vezes de 1960 a 2019, de 2 milhões de toneladas a 368 milhões de toneladas (Pilapitiya e Ratnayake, 2024), resultando em uma considerável quantidade de resíduos gerados a cada ano que contaminam não somente o ambiente terrestre, mas também o marítimo e atmosférico.

Diante deste contexto, existem diversas alternativas ao descarte imediato de resíduos plásticos, entre elas, a reciclagem. No Brasil, cooperativas e associações de catadores possuem um papel importante não somente no contexto da reciclagem, mas também no quesito socioeconômico do país, permitindo que classes, geralmente marginalizadas e excluídas da sociedade, tenham acesso a oportunidades de empregos. Segundo o Anuário da Reciclagem de 2024, as mais de 3 mil organizações de catadores foram responsáveis pela recuperação de cerca de 1,6 milhões de resíduos.

Apesar da reciclagem ser uma alternativa viável e relevante para o cenário brasileiro, ela ainda está longe de se tornar uma solução para o problema da poluição por plásticos, pois apenas 9% de todo o plástico gerado no mundo é reciclado (ONU, 2019). Isso ocorre pois nem todo plástico pode ser reciclado e existe uma dificuldade em relação a separação correta de cada tipo de plástico (Assis e Santos, 2020), problemática esta que dá origem a este trabalho.

Em vista da dificuldade de separação dos plásticos, o presente trabalho propõe facilitar esta tarefa através de sinais de áudio provenientes de sete tipos de plástico, sendo eles: (0) polietileno tereftalato (PET), (1) polietileno de alta densidade (HDPE), (2) policloreto de vinila (PVC), (3) polietileno de baixa densidade (LDPE), (4) polipropileno (PP), (5) poliestireno (PS) e (6) Outros. Para tal, foi proposta a avaliação de quatro diferentes tipos de modelos de aprendizado de máquina: k-vizinhos mais próximos (*k*-NN – *k-nearest neighbors*), as máquinas de vetores de suporte (SVM – *support vector machines*), a regressão logística (LR – *logistic regression*) e as florestas aleatórias (RF – *random forests*).

METODOLOGIA

Para o desenvolvimento dos modelos, o banco de dados obtido por Tessarini e Fileti (2022) foi utilizado. Nele há 1400 dados de áudio, sendo 700 deles plásticos cortados, colocados em uma sacola de pano e amassados com uma morsa, e

os outros 700 plásticos na sua forma natural e amassados com a mão. Em ambos casos, há 100 amostras de cada tipo de plástico. Os modelos foram treinados com os dois tipos de dados juntos.

Para a abertura dos arquivos, os quais estavam em formato *.wav,* utilizou-se a biblioteca *librosa* do Python no ambiente de desenvolvimento Google Colab, sendo ambos, a linguagem de programação e o ambiente utilizados durante todo o desenvolvimento do trabalho. Uma amostra aleatória do banco de dados, processada com tal biblioteca, é apresentada na Figura 1. Assim, esta imagem configura o sinal bruto obtido do arquivo de áudio.

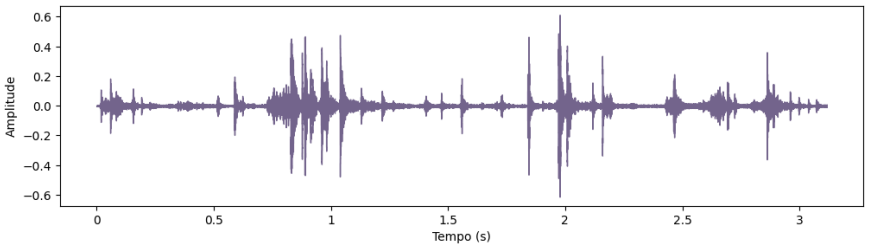

Figura 1 – Sinal bruto de uma amostra aleatória do banco de dados.

Apenas o sinal de áudio não é suficiente para o treinamento dos modelos, foi preciso tratar este sinal em busca de características e informações relevantes para a entrada do modelo. Diversos trabalhos no ramo de classificação através de sinais de áudio utilizam os coeficientes de Mel (MFCCs - *Mel-Frequency Cepstral Coefficients*) para tal tarefa, como identificação de ruídos em Grama e Rusu (2017) ou sistemas de detecção de intrusos de Ghiurcau *et al.* (2011).

Os coeficientes de Mel são obtidos a partir de uma transformada de Fourier, seguida da extração da magnitude e, logo, do logaritmo de um sinal de áudio. Em seguida, o espectro é transformado para a escala de Mel e outra transformada é realizada, a transformada discreta de cosseno (DCT) (Mitrović, 2010). O produto desta DCT é uma série de vetores logaritmizados que representam os MFCCs. Os primeiros coeficientes gerados, 12 a 24 segundo Furui (2010) ou 8 a 13 segundo Mitrović (2010), são aqueles que contém a maior informação do sinal, mas o número ótimo para cada situação pode variar (Mitrović, 2010). Sendo assim, foram extraídos 30 MFCCs e, em cada algoritmo testado, avaliou-se a quantidade ideal de MFCCs.

Após a extração dos MFCCs, foi necessário reduzir a dimensão deste conjunto de dados que descreve o áudio, pois os modelos investigados aceitam apenas entradas no formato de uma matriz "dados *x* características" e o que se tem, logo após a extração dos MFCCs, é uma matriz no formato "dados *x* características *x* ~300 valores", pois cada dado possui características que tem o formato de um vetor com cerca de 300 valores. Assim, a redução de cada MFCC pela sua média aparece como uma opção bem consolidada para classificação através de sinais de áudio (Küçübay e Sert, 2015; Ghiurcau *et al.*, 2012). Desta forma, para cada modelo em estudo, foram extraídos os MFCCs de cada sinal de áudio e, em seguida, esses foram reduzidos a suas médias, configurando a entrada dos modelos.

Os quatro modelos utilizados no estudo se adaptam ao cenário multiclasse em questão. O modelo *k*-NN, um dos mais simples algoritmos de aprendizado de máquina (Müller e Guido, 2017), consiste apenas em guardar os dados de treino e, para realizar uma classificação, encontrar os *k* pontos mais próximos, identificando suas classes e predizendo a classe do ponto a ser classificado a partir da distância medida, conforme a Equação 1.

$$d = \left(\sum_{i=1}^{n} \left(\left| x_i - y_i \right| \right)^p \right)^{1/p} \quad (1)$$

O modelo SVM, por outro lado, busca o melhor hiperplano que separa os dados de entrada. Como, neste caso, a separação entre classes não é linear, utiliza-se extensões das características, isto é, a combinação de características através de multiplicações entre elas, por exemplo (Müller e Guido, 2017). A função SVC do *scikit-learn* exige o ajuste dos hiperparâmetros *kernel* e *C*, que consiste no parâmetro de regularização. Existem diversos tipos de *kernel* e, a depender do

tipo, é necessário configurar distintos hiperparâmetros. No caso do *kernel* polinomial, por exemplo, ajusta-se o *degree*, já no caso dos *kernels* de função de base radial, polinomial ou sigmóide, o *gamma* é ajustado (Pedregosa *et al.*, 2011). Em contrapartida, o modelo RL, originalmente se refere a uma classificação binária, mas é generalizado pela chamada *Sofmax Regression* (SR) (Géron, 2017), continuando, entretanto, com a mesma ideia da classificação binária. Neste modelo, é necessário ajustar a regularização *C* do modelo bem como o tipo de *solver* que será utilizado. A depender do *solver*, pode-se utilizar um ou mais tipos de penalizações, L1, L2 ou um conjunto das duas (Pedregosa *et al.*, 2011) Por fim, o modelo RF é um modelo do tipo *ensemble* que junta as classificações de várias árvores de decisão para decidir a classificação final. Este modelo, diferentemente dos demais, consegue indicar quais características do áudio são mais importantes, isto é, quais MFCCs tem maior relevância na construção das árvores, visto que é provável que as características mais importantes se encontram perto da raiz da árvore, enquanto as menos importantes devem se encontrar mais longe, perto das folhas, por exemplo (Géron, 2017). Assim, foram analisados o número máximo de árvores, o número mínimo de amostras para divisão e o número mínimo de amostras por folha. A profundidade máxima da árvore não foi estipulada previamente, a fim de permitir que o algoritmo encontre uma profundidade ótima.

Em todos os modelos, o melhor conjunto de hiperparâmetros e número de MFCCs foi definido quando a acurácia no conjunto de validação do *GridSearch* com 10 folds era mais alta, a fim de evitar o *overfitting* durante a construção dos modelos. Assim, variou-se os hiperparâmetros de forma a encontrar o modelo que ajustasse melhor os dados. Entretanto, antes de realizar o *GridSearch* realizou-se uma análise dos hiperparâmetros e como eles influenciavam o conjunto de treino e validação. Variou-se, por exemplo, o tipo de *solver* na RL ou a quantidade de árvores na RF para avaliar o efeito gerado.

Por fim, comparou-se os modelos através da acurácia obtida por estes nos dados de teste.

RESULTADOS E DISCUSSÃO

Com respeito ao modelo *k*-NN, o melhor modelo obtido foi aquele com a média de 18 MFCCs como entrada e hiperparâmetros $k = 5$, $p = 3{,}2$ e *weights = uniform,* definidos no *GridSearch*, que obteve 89,33% de acurácia média nos 10 *folds* de validação utilizados. Este algoritmo de classificação obteve 93,53% de acurácia com os dados de treino e 92,86% com os dados de teste, sendo a matriz de confusão do teste disposta na Figura 2 (a).

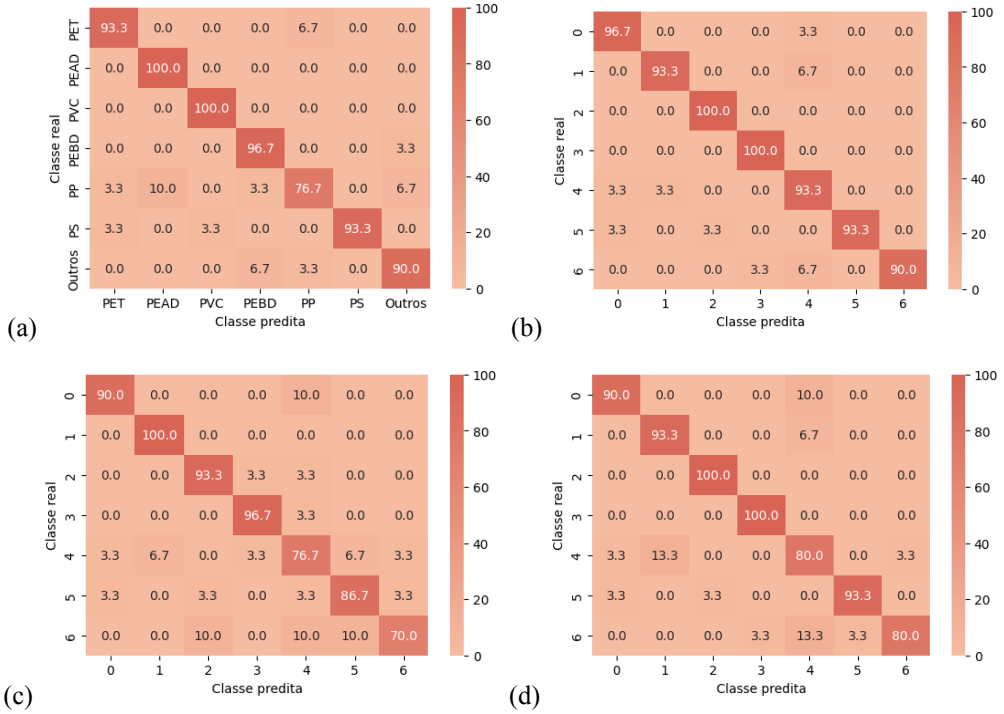

Figura 2 – Matriz de confusão dos testes dos modelos (a) k-NN, (b) SVM, (c) RL e (d) RF.

Por outro lado, o modelo SVM, o qual obteve a maior acurácia no teste, teve, como entrada, a média de 13 MFCCs, e hiperparâmetros como *kernel* = *RBF*, *C* = 2 e *gamma* = 11,7. No *GridSearch,* os dados de validação dos *folds* apresentaram uma acurácia média de 92,34%. Este modelo obteve 98,99% de acurácia no treino e 95,23% no teste. A matriz de confusão referente ao teste do modelo SVM pode ser visualizada na Figura 2 (b) acima.

Em contrapartida, o modelo de RL trouxe o resultado menos favorável em termos de acurácia no teste, com 86,97% de acurácia no treino e 87,62% no teste, com hiperparâmetros definidos em *solver* = *saga*, *penalty* = L1 e C = 1, obtendo uma acurácia média nos *folds* do *GridSearch* em 76,77%. A matriz de confusão dos dados de teste do modelo se encontra na Figura 2 (c). É interessante notar, para além do resultado da acurácia, que este modelo, diferentemente dos demais, utilizou um maior número de MFCCs para gerar o melhor resultado: 27 coeficientes. Isto pode indicar que, eventualmente, os MFCCs não seriam a melhor técnica de extração de características para este modelo. A análise da influência do número de coeficientes na acurácia dos *folds* do *GridSearch* para o modelo RL pode ser visualizada abaixo, na Figura 3.

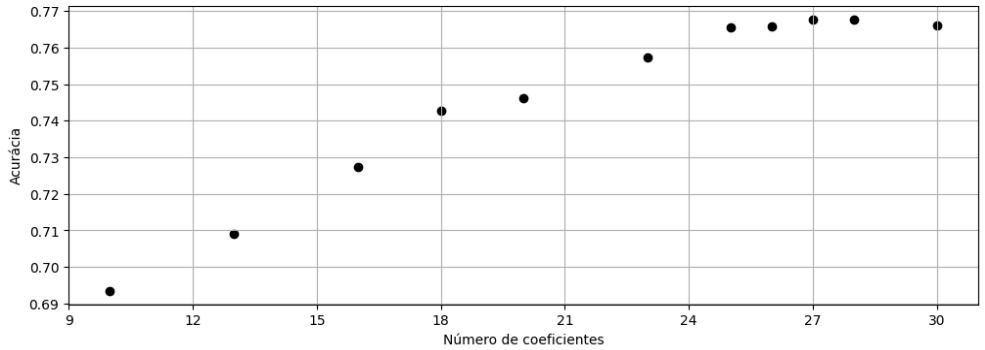

Figura 3 – Análise da influência do número de MFCCs no resultado de acurácia no GridSearch do modelo RL..

Por fim, o modelo RF obteve resultados medianos, porém permitiu uma análise da importância de cada um dos MFCCs na classificação. Na Figura 4 abaixo, pode-se observar a importância de cada coeficiente no modelo gerado, identificando que o terceiro coeficiente é o que tem maior relevância e que os três primeiros guardam grande parte da informação necessária para a classificação pela floresta. Ademais, a melhor RF obteve, com 16 MFCCs, 99,66% de acurácia no treino e 90,95% de acurácia no teste, com número de árvores = 80, número mínimo de amostras para divisão = 2 e número mínimo de amostras por folha = 2. Este conjunto de hiperparâmetros expressou uma acurácia de 88,99% nos dados de validação dos *folds* do *GridSearch* e a matriz de confusão dos dados de teste é apresentada na Figura 2 (d).

Comparativamente, percebe-se que o modelo que melhor classifica os dados é o SVM, seguido do *k*-NN, da RF e, então, da RL. As métricas de desempenho no treino, na validação dos *folds* do *GridSearch* e no teste de cada modelo analisado estão resumidas na Tabela 1 abaixo.

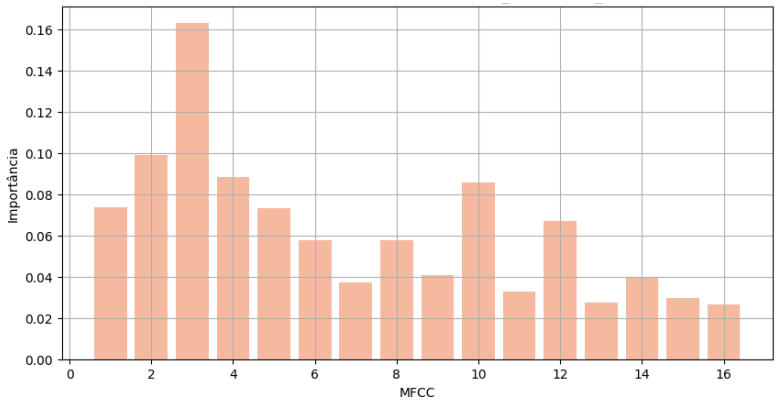

Figura 4 – Análise da importância de cada MFCC na construção do modelo RF..

Modelo	Acurácia no treino	Acurácia nos *folds* de validação do *GridSearch*	Acurácia no teste
k-NN	93,53%	89,33%	92,86%
SVM	98,99%	92,34%	95,23%
RL	86,97%	76,77%	87,62%
RF	99,66%	88,99%	90,95%

Tabela 1 – Tabela comparativa do desempenho dos quatro modelos testados.

CONCLUSÕES

O objetivo deste trabalho era classificar com a maior acurácia possível através dos modelos estudados os sete tipos de plásticos presentes no banco de dados para que, desta forma, em futuros trabalhos um aplicativo de celular ou *software* possa ser criado e passe a auxiliar a tomada de decisões nas cooperativas de reciclagem do Brasil. De fato, o objetivo foi atingido, especialmente, mas não somente, através do modelo SVM, o qual classificou corretamente 95,23% dos dados de teste a ele expostos, configurando uma excelente opção para o desenvolvimento das aplicações futuras mencionadas.

BIBLIOGRAFIA

ANUÁRIO DA RECICLAGEM 2024. Disponível em: https://anuariodareciclagem.eco.br/. Acesso em: 21 jul. 2025.

ASSIS, Marcos Wilson Vicente de; SANTOS, Taides Tavares dos. Propriedades químicas, problemas ambientais e reciclagem de plástico: uma revisão de literatura. **Jornal Interdisciplinar de Biociências.** Santarém, p. 31-37. 2020.

FURUI, Sadaoki. Chapter 7 - Speaker Recognition in Smart Environments. In: AGHAJAN, Hamid; DELGADO, Ramón López-Cózar; AUGUSTO, Juan Carlos. **Human-Centric Interfaces for Ambient Intelligence.** Academic Press, 2010. p. 163-184.

GÉRON, Aurélien. ***Hands‑On Machine Learning with Scikit‑Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.*** 2. ed. (ilustrada). Sebastopol: O'Reilly Media, 2019. 819 p. ISBN 978‑1‑492‑03264‑9.

GHIURCAU, Marius Vasile et al. **Audio based solutions for detecting intruders in wild areas.** Signal Processing, [S.L.], v. 92, n. 3, p. 829-840, mar. 2012. Elsevier BV. http://dx.doi.org/10.1016/j.sigpro.2011.10.001.

GRAMA, Lacrimioara; RUSU, Corneliu. **Choosing an accurate number of mel frequency cepstral coefficients for audio classification purpose.** IEEE, Ljubljana, p. 225-230, set. 2017. IEEE. http://dx.doi.org/10.1109/ispa.2017.8073600.

KÜÇÜKBAY, Selver Ezgi; SERT, Mustafa. **Audio-based event detection in office live environments using optimized MFCC-SVM approach.** IEEE, Anaheim, p. 475-480, fev. 2015. IEEE. http://dx.doi.org/10.1109/icosc.2015.7050855.

MITROVIć, Dalibor *et al*. Chapter 3 - Features for Content-Based Audio Retrieval. In: ZELKOWITZ, Marvin V. **Advances in Computers**. Viena: Elsevier, 2010. p. 71-150.

MÜLLER, Andreas C.; GUIDO, Sarah. **Introduction to Machine Learning with Python: a guide for data scientists**. 1. ed. Sebastopol: O'Reilly Media, 2016. 376 p.

PEDREGOSA, F.. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**. p. 2825-2830, 2011.

PILAPITIYA, P.G.C. Nayanathara Thathsarani; RATNAYAKE, Amila Sandaruwan. The world of plastic waste: a review. **Cleaner Materials**, [S.L.], v. 11, mar. 2024. Elsevier BV. http://dx.doi.org/10.1016/j.clema.2024.100220.

TESSARINI, Letícia; FILETI, Ana Maria Frattini. **Banco de dados de áudio: classificação de resinas plásticas para reciclagem usando processamento de sinais acústico e redes neurais artificiais.** Repositório de Dados de Pesquisa Unicamp, 2022. Disponível em: https://redu.unicamp.br/dataset.xhtml?persistentId=doi:10.25824/redu/UGMPJR.

UNIDAS, Nações. **ONU Meio Ambiente aponta lacunas na reciclagem global de plástico**. 2019. Disponível em: https://brasil.un.org/pt-br/82048-onu-meio-ambiente-aponta-lacunas-na-reciclagem-global-de-pl%C3%A1stico. Acesso em: 21 jul. 2025.