



Classificação de Sucos, Chás e Chocolates por Análise de Padrões RGB

Classification of Juices, Teas, and Chocolates by RGB Pattern Analysis

Patrick Guilherme Roza¹, Vanderlei Aparecido de Lima²

RESUMO

Neste trabalho classificou-se amostras de quatro sabores de suco, três sabores de chás, duas marcas de chocolates com e sem presença de cacau, por meio de seus padrões RGB (Red, Green, Blue), utilizando softwares de fácil acesso e manuseio. Foram preparadas 104 amostras de suco, 312 amostras de chás, 160 amostras de chocolate da marca A e 240 amostras da marca B. A preparação das amostras se fez por meio do registro fotográfico e então recortadas por meio do programa GIMP 2.10.30. Após essa etapa, seus padrões RGB foram extraídos pelo software ChemoStat e as amostras foram classificadas em função de seus atributos menos correlacionados no software Weka 3.8.6, utilizando o algoritmo de aprendizagem de máquina *Random Forest* (RF). Os resultados da classificação apresentaram acertos significativos, conseguindo diferenciar, mesmo com alguma taxa de erro, amostras de cores muito próximas. A acurácia global foi acima de 80%, dependendo do tipo de matriz analisada.

PALAVRAS-CHAVE: Aprendizado de máquina, classificação, quimiometria, RGB, padrão

ABSTRACT

In this study, samples of four flavors of juice, three flavors of tea, two brands of chocolate with and without cocoa, were classified using their RGB (Red, Green, Blue) patterns. The classification was carried out with user-friendly software. We prepared a total of 104 juice samples, 312 tea samples, 160 chocolate samples from brand A, and 240 samples from brand B. The samples were initially captured through photography and then processed using the GIMP 2.10.30 program. Following this step, their RGB patterns were extracted using the ChemoStat software, and the samples were categorized based on their least correlated attributes using the Weka 3.8.6 software. The classification was performed with the Random Forest (RF) machine learning algorithm. The results of the classification exhibited significant success, even though there was a slight margin of error in distinguishing very similar color samples. The overall accuracy exceeded 80%, and this accuracy rate varied depending on the type of matrix being analyzed.

INTRODUÇÃO

Tendo seu começo entre 1950 e 1960, a quimiometria se mostrou um método eficiente trazendo velocidade de análise, mostrou-se viável, pois por meio de algoritmos duas grandes tarefas podem ser realizadas, (1) regressão em modelos preditivos ou (2) classificação de matrizes químicas. Na quimiometria, dados de diversos tipos de matrizes são analisados computacionalmente. Essa área da química está presente ainda em outras diversas áreas tais como indústrias de alimentos e de materiais, áreas ambiental, na medicina e na área agrícola. (BRUNI, 2023).

Por meio da quimiometria é possível modelar dados químicos de natureza multivariados utilizando-se algoritmos, que correlacionam variáveis entre si, gerando modelos de regressão ou

¹ Bolsista do(a) PIBIC. Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil. E-mail: Patrickbarbosa742@gmail.com. ID Lattes: <http://lattes.cnpq.br/8579529381581768>

² Orientador de IC. Universidade Tecnológica Federal do Paraná, Pato Branco, Paraná, Brasil. E-mail: valima@utfpr.edu.br. ID Lattes: <http://lattes.cnpq.br/9090461949264421>



modelos de classificação de matrizes. Existem vários programas computacionais que são utilizados para modelar dados de matrizes químicas. (WOLD, 1995.)

Um desses software que é comumente utilizado na quimiometria e para tratamento de dados é o software ChemoStat, que pode analisar tanto dados espectrais como dados de imagem. Dados espectrais podem ser por meio de infravermelho médio ou próximo (NIR) e dados de imagens de ensaios que desenvolvem coloração. Nesse último imagens de ensaios (coloridos) são decompostos em tons de cinza, de seus respectivos canais RGB (Red/ Green/ Blue). Este software é gratuito, leve, fácil de instalação e de utilização fácil. (HELPER, BOCK, MATDER, FURTADO, Da COSTA, FERRÃO, 1995.)

O ChemoStat utiliza métodos como análise de agrupamento hierárquico (HCA), análise de componentes principais (PCA), como outros métodos de análise de dados. Esse software aplica funções para transformação dos dados e detecção de amostras anômalas (outliers), entre outros. (DAMACENO, 2016.)

Outro programa de importância no tratamento de dados multivariados é o WEKA (Waikato Environment for Knowledge Analysis), um software que foi desenvolvido por meio da linguagem java. O Weka possui uma gama de algoritmos com diversas técnicas de mineração de dados como métodos de classificação, como Random Forest que constrói diversas árvores de decisão, selecionando amostras aleatoriamente e não sua totalidade, e também métodos de predição numérica, agrupamento e de associação sendo um software de ampla utilização no meio acadêmico. Praticamente dados de quaisquer matrizes podem ser analisados nesse software até mesmo dados de adulteração de matrizes. (DAMACENO, 2016.)

Crimes de adulteração de alimentos e produtos comerciais são comuns no Brasil. Essa área de pesquisa ainda está em desenvolvimento e ainda há muitas lacunas para serem preenchidas nas análises de adulterações de alimentos. Essa falta de metodologias que consigam identificar padrões de adulteração em alimentos motiva pesquisadores para se determinar tais padrões de adulteração. No Brasil podemos considerar que temos problemas persistente de crimes de adulteração de todos os tipos de produtos, produtos alimentícios e consumíveis como combustível e cosméticos. (VELOSO, FERREIRA, FERREIRA, 2002.)

Outro ponto chave para a investigação em matrizes de alimentos é a falta de padronização de suas matrizes. A coloração de produtos alimentícios pode mostrar falta de padronização em sua escala industrial. Produtos padronizados podem levar o consumidor a consumir mais de certo produto e produtos com baixa padronização podem levar o consumidor a consumir em menor escala tal matriz alimentícia.

Diante do exposto, o objetivo desta pesquisa foi a classificação de três tipo de alimentos: chás, sucos em pós (para serem preparados com água) e chocolates com diferentes porcentagens de cacau.

MATERIAIS E MÉTODOS

Para aquisição das imagens dos diferentes sabores de suco foram adquiridos quatro sabores de suco em pó no mercado local, abacaxi, laranja, manga e maracujá, diluído cada um deles em 1 L de água. Usou-se duas bandejas de plástico com 13 pequenos recipientes, após preencher esses recipientes com determinado sabor de suco, foi capturado uma imagem por meio de um smartphone ASUS Zenfone Max Shot, em duplicada, considerando cada recipiente da

bandeja preenchida com suco constituindo uma amostra, e esse processo foi realizado para todos os sabores totalizando 2 imagens para cada sabor.

Para aquisição das imagens dos chás, um processo bem semelhante foi utilizado, sendo adquiridos três sabores de chás, sendo eles maracujá, erva-doce e capim cidreira. Foram feitos quatro xícaras de 240mL de chá, em seguida foi preenchido as bandejas totalizando 26 amostras, e então, com um fundo branco, capturado imagem por meio de um smartphone ASUS Zenfone Max Shot, em duplicata para cada xícara, e esse processo foi realizado com os 3 sabores de chás, totalizando 8 imagens para cada sabor.

Na aquisição das imagens dos chocolates foram comprados no mercado local uma barra de chocolate da marca A ao leite e uma barra de chocolate A 60% cacau, além de uma barra do chocolate da marca B ao leite e uma barra do chocolate B 40%. Os chocolates foram cortados com uma faca de cozinha em pequenas barras relativamente iguais, obtendo 10 barras de cada chocolate da marca A e 15 de cada chocolate da marca B.

Em um fundo branco, então foi realizado a captura das imagens, após cada captura, o celular foi girado 45 graus, até completar os 360 graus, totalizando 8 imagens para cada chocolate.

Utilizando o programa GIMP 2.10.30, as imagens foram recortadas em uma dimensão de 40x40 *pixels*, cada recipiente preenchido na bandeja, assim como cada quadrado dos chocolates, foi recortado, gerando assim 26 amostras de cada sabor de suco, 104 amostras de cada sabor de chá, 240 amostras do chocolate da marca B e 160 amostras do chocolate da marca A.

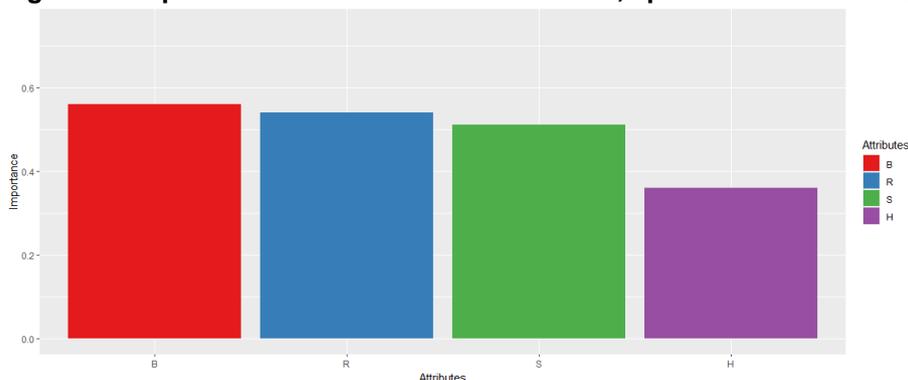
Por meio do programa ChemoStat, foi extraído e analisado os padrões RGB das amostras, que foram em seguida avaliados no RStudio, para seleção dos melhores atributos menos correlacionados, por meio do VIF (Variance Inflation Factor).

O software Weka foi usado para classificação das matrizes, utilizando-se seus padrões de cores (RGB). Vários algoritmos foram testados para as classificações das matrizes, mas apenas os resultados do algoritmo *Random Forest* (RF) foram mostrados nesse artigo

RESULTADOS E DISCUÇÕES

Os atributos menos correlacionados e, portanto, os mais importantes para as imagens da matriz chocolate foram B, R, S e H (Figura 1) e, assim baseado nesses atributos a classificação realizada no software Weka com RF, apresentou 92,5% de acurácia global (Quadro 1).

Figura 1 - Importância dos atributos com VIF < 5,0 para os sabores de chocolate



Fonte: Dados gerados no RStudio. (2023)



Quadro 1 - Matriz de confusão para classificação das amostras de chocolates, em suas classes prevista e esperadas

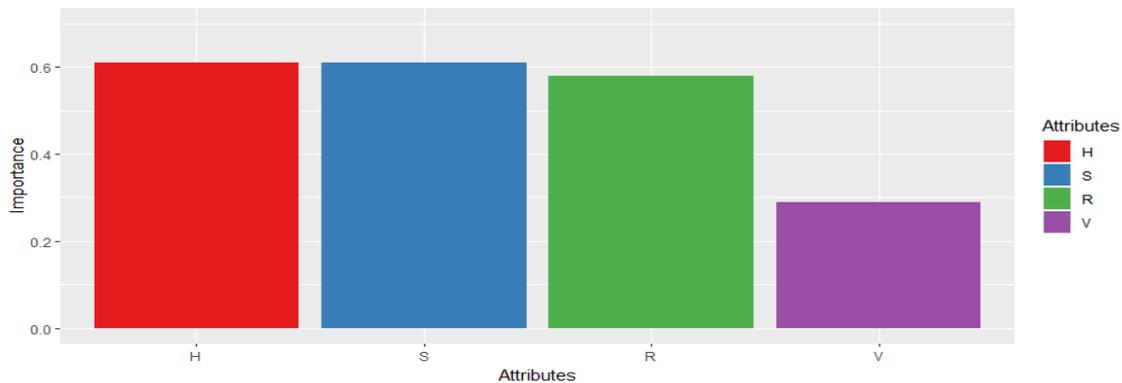
Chocolates		Classe esperada			
		A60	AL	B40	BL
Classe prevista	A60	20	0	5	0
	AL	0	19	0	0
	B40	3	0	40	0
	BL	0	1	0	32

Fonte: Software Weka, análise por Random Forest. (2023)

O total de 20 amostras de chocolate A60 foram corretamente classificadas (80% de acerto). As outras cinco amostras foram classificadas como chocolate B40 (chocolate meio amargo com 40% de cacau). Por outro lado, três amostras B40 (93,02% de acerto) foram classificadas como A60, também uma amostra BL foi classificada como AL, e no grupo do AL obtive-se acerto de 100%, na classificação. O chocolate AL foi classificado com 100% de acerto. Pode-se observar que as classificações errôneas foram sempre do mesmo grupo de amostras; chocolate ao leite e chocolate meio amargo com maiores teores de cacau.

Para os sucos preparados, a porcentagem de acerto global foi de 87,5%. As amostras de abacaxi apresentaram 100% de acerto na classificação, o que já era esperado pois esse sabor apresenta uma coloração bem diferente dos demais sucos.

Figura 2 - Importância dos atributos com VIF < 5,0 para os sabores de sucos



Fonte: Dados gerados no RStudio. (2023)

Quadro 2 - Matriz de confusão para a classificação das amostras de sucos

Sucos		Classe esperada			
		Abacaxi	Laranja	Manga	Maracujá
Classe prevista	Abacaxi	19	0	0	0
	Laranja	0	15	1	3
	Manga	0	2	15	0
	Maracujá	0	3	0	14

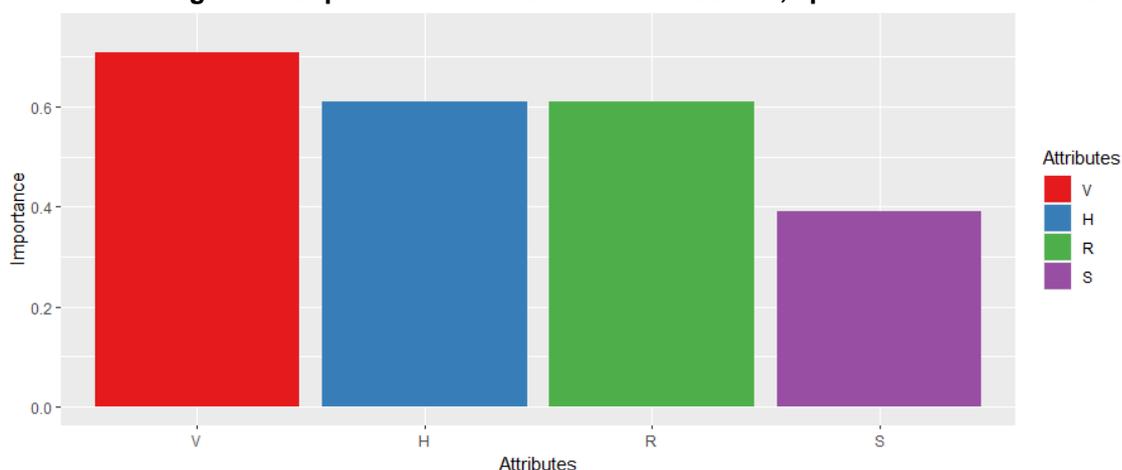
Fonte: Software Weka, análise por Random Forest. (2023)

Quatro amostras de suco de laranja foram classificadas como suco de manga (1 amostra) e como suco de maracujá (3 amostras). A classificação correta, para o suco de laranja foi de

78,94%. A classificação correta para o suco de manga foi de 88,23% e porcentagem de acerto para o suco de maracujá foi de 82,35%. Além dos sucos e chocolates, as classificações para os chás também foram avaliadas.

A taxa de acerto global para os chás foi de 97,7%. Observou-se que apenas duas amostras de capim-cidreira classificadas erroneamente como chá de erva doce. Por outro lado, três amostras de chá de erva-doce, foram erroneamente classificadas como capim-cidreira. O chá de camomila apresentou 100% de acerto

Figura 3 - Importância dos atributos com VIF < 5,0 para os sabores de chás



Fonte: Dados gerados no RStudio. (2023)

Quadro 3 - Matriz de confusão para a classificação das amostras de chás, em suas classes prevista e esperada

Chás		Classe esperada		
		Capim-cidreira	Erva-doce	Camomila
Classe prevista	Capim-cidreira	61	2	0
	Erva-doce	3	70	0
	Camomila	0	0	82

Fonte: Software Weka, análise por Random Forest. (2023)

CONCLUSÕES

Constatou-se que amostras com cores mais próximas apresentaram uma taxa de erro maior das demais em todos os tipos de amostras analisadas. Isso ocorreu com amostras de chocolates ao leite, e o chocolate meio amargo, que apresentou cacau, em sua composição. O suco de abacaxi assim, como o chá de camomila apresentaram 100% de correta classificação, devido ao seu padrão de coloração.

Os diferentes números de amostras em cada grupo deixam perceptível que a classificação se torna mais assertiva com alto número de amostra. N amostral alto implica em melhor aprendizagem dos valores dos atributos pelo algoritmo utilizado na classificação.

Utilizando-se apenas imagens capturadas por smartphones e algoritmos de aprendizado de máquina, como Random Forest, pode-se prever as classes de matrizes alimentícias. Essa metodologia pode ser utilizada no setor de alimentos para controle de qualidade de matrizes alimentícias.



Agradecimentos

Meus agradecimentos vão para o PIBIC e a UTFPR pelo importante apoio dado a nós alunos, ao meu orientador Doutor Vanderlei Aparecido de Lima pela oportunidade de aprendizagem, orientação e confiança, e também aos meus pais por tornar possível meus estudos na UTFPR.

Conflito de interesse

Os autores do artigo declaram não haver conflito de interesse..

REFERÊNCIAS

BRUNI T. A. Quimiometria Aula 02 Histórico **E Definições**. E-Disciplinas, São Paulo, 2023.

Disponível em:

<https://edisciplinas.usp.br/pluginfile.php/7640696/mod_resource/content/1/Aula02Quimio_2023.pdf>. Acesso em: 09/08/2023.

DAMACENO M. Introdução a Mineração de Dados Utilizando o Weka. **Instituto Federal de Educacao, Ciencia e Tecnologia do Rio Grande do Norte/Campus**, 03/09/2016. Disponível em

<Macau,<http://connepi.ifal.edu.br/ocs/anais/conteudo/anais/files/conferences/1/schedConfs/1/papers/258/public/258-4653-1-PB.pdf>>. Acesso em: 09/08/2023.

HELPER A G.. BOCK F. MATDER L. FURTADO C. J. Da COSTA B. A. FERRÃO F. M.

Chemometrics; what do we mean with it, and what do we want from it? **Química Nova**. Volume 30, Issue 1, Pages 109-115, November 1995. Disponível em:

<<https://www.sciencedirect.com/science/article/pii/0169743995000429?via%3Dihub>>. Acessado em:09/08/2023

VELOSO A. C. A.; TEIXEIRA N.; FERREIRA M. P. L. V. O. I.; FERREIRA A. m.; Detecção de adulterações em produtos alimentares contendo leite e/ou proteínas lácteas. **Química Nova**.

Porto, Portugal, 25/07/2002; 609–15. Disponível em: <<https://doi.org/10.1590/S0100-40422002000400016>>. Acessado em: 09/08/2023.

WOLD. S. Chemometrics; what do we mean with it, and what do we want from it?. **Química Nova**.

Recebido 04/01/1995, Aceitado: 10/05/1995, disponível online 21/02/2000. Disponível em:

<[https://doi.org/10.1016/0169-7439\(95\)00042-9](https://doi.org/10.1016/0169-7439(95)00042-9)>. Acesso em: 09/08/2023