

## MACHINE LEARNING NA PREDIÇÃO DE ATRIBUTOS DE QUALIDADE EM ÁGUA DE COCO COM USO DA ESPECTROSCOPIA Vis-NIR

LURENIELLE FERREIRA MORAES DA SILVA<sup>1</sup>, RAFAEL FERREIRA MOARIS  
DA SILVA<sup>2</sup>, KARLA DOS SANTOS MELO DE SOUSA<sup>3</sup>, DANIEL DOS SANTOS  
COSTA<sup>4</sup>, CÍCERO HENRIQUE DE SÁ<sup>5</sup>

<sup>1</sup> Eng. Agrônoma, Mestranda em Engenharia Agrícola, Depto. de engenharia Agrícola e Ambiental, UNIVASF, Juazeiro - BA, laurenielle276@gmail.com

<sup>2</sup> Eng. Agrícola e Ambiental, Depto. de engenharia Agrícola e Ambiental, UNIVASF, Juazeiro – BA.

<sup>3</sup> Eng. Agrícola, Profa. Assist. Doutora, Depto. de engenharia Agrônômica, UNIVASF, Petrolina - PE.

<sup>4</sup> Eng. Agrícola e Ambiental, Prof. Assist. Doutor, Depto. de engenharia Agrícola e Ambiental, UNIVASF, Juazeiro – BA.

<sup>5</sup> Graduando em Engenharia Agrônômica, Depto. de engenharia Agrônômica, UNIVASF, Petrolina - PE.

Apresentado no  
LII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2023  
18 a 21 de outubro de 2023 – Ribeirão Preto - SP, Brasil

**RESUMO:** Este trabalho tem por objetivo o desenvolvimento de modelos para a predição de atributos físico-químicos e classificação do estágio de maturação em água de coco com uso da espectroscopia Vis-NIR associada a técnicas de machine learning. Foram utilizadas 200 unidades de cocos (*Cocus nucifera* L.), da variedade anã, sendo 100 unidades no estágio de maturação verde e 100 unidades no estágio de maturação maduro. Os atributos analisados foram o pH, acidez e sólidos solúveis através da metodologia tradicional, além das amostras submetidas a análise espectral. Antes de serem gerados os modelos, foi necessário a aplicação de algumas técnicas de pré-processamento. Foi possível observar a influência do estágio de maturação nos parâmetros físico-químicos da água de coco. O modelo gerado com o pré-tratamento MSC e apresentar resultados iguais para todos os parâmetros analisados. A taxa de verdadeiro positivo para os dois modelos ficou próximas da unidade e mostra a relação das amostras que foram classificados corretamente. A espectroscopia Vis-NIR associada com o algoritmo de modelagem SMO e pré-tratamento MSC, proporcionaram a discriminação dos estádios de maturação e a predição dos atributos de qualidade do coco.

**PALAVRAS-CHAVE:** *Cocus nucifera* L., algoritmos, espectro

## MACHINE LEARNING IN THE PREDICTION OF QUALITY ATTRIBUTES IN COCONUT WATER USING VIS-NIR SPECTROSCOPY

**ABSTRACT:** This work aims to develop models for the prediction of physicochemical attributes and classification of maturation stage in coconut water using Vis-NIR spectroscopy associated with machine learning techniques. A total of 200 units of coconuts (*Cocus nucifera* L.), of the dwarf variety, were used, 100 units in the green ripeness stage and 100 units in the mature ripeness stage. The attributes analyzed were pH, acidity, and soluble solids through the traditional methodology, besides the samples submitted to spectral analysis. Before the models were generated, it was necessary to apply some pre-processing techniques. It was possible to observe the influence of the maturation stage on the physicochemical parameters of coconut water. The model generated with the pre-treatment MSC and present equal results for all analyzed parameters. The true positive rate for both models was close to unity and shows the ratio of samples that were correctly classified. Vis-NIR spectroscopy associated

with the SMO modeling algorithm and MSC pre-treatment provided the discrimination of maturation stages and the prediction of coconut quality attributes.

**KEYWORDS:** *Cocos nucifera* L., algorithms, spectrum

**INTRODUÇÃO:** Em 2019, dados da FAO (Food and Agriculture Organization of the United Nations - Organização das Nações Unidas para a Alimentação e a Agricultura) mostraram que o Brasil ocupava a quinta posição mundial na produção de coco com um total de 3,9% da produção total. A Indonésia, Filipinas, Índia e Sri Lanka detém as primeiras posições na produção mundial. Segundo Brainer e Ximenes (2020), no Brasil, a maior região produtora é o Nordeste, com 81,3% da área cultivada e 71,2% da produção nacional. A produtividade dos coqueiros está relacionada ao fim que o produto será destinado, a variedade cultivada, a tecnologia a ser utilizada, os insumos que serão usados, os fatores edafoclimáticos, dentre outros. O coco cultivado no Brasil se destina principalmente para coco seco in natura, coco ralado, leite de coco e água de coco. Estima-se que 10% da produção de coco é destinada ao processamento de água de coco para envase e posteriormente distribuição (SUCUPIRA, 2016). Com o aumento de consumo de água de coco e importação dessa matéria-prima, o Ministério da Agricultura, Pecuária e Abastecimento determinou a instrução normativa nº 27 de 22 de julho de 2009 para estabelecer os padrões de identidade e qualidade para a água de coco e respectivos parâmetros analíticos. Em 2020, essa instrução foi revogada e atualizada para a instrução nº 9 de 30 de janeiro de 2020 (MAPA, 2020). Diante disso, os produtores de água de coco verde in natura passaram a padronizar seus produtos de acordo com os parâmetros dispostos na legislação. Desta forma, as empresas devem ter processos adequados para padronização e métodos analíticos e mão de obra qualificada para garantir a qualidade de seus produtos. Uma das problemáticas desses métodos de análises são os custos gerados pelo mantimento desses equipamentos e materiais de análise, como vidrarias e reagentes. Esses métodos, por serem dependentes de um analista, acabam demandando tempo para serem realizadas, o que pode atrasar a tomada de decisão caso um desses parâmetros fiquem fora do padrão. Nesse sentido, surgiu a necessidade de desenvolvimento de métodos analíticos rápidos, preciso e de baixo custo. Assim, a espectroscopia Vis-NIR vem sendo cada vez mais utilizada em escala laboratorial, industrial e na agricultura, pois traz vantagens como análises em tempo até menores que 60 segundos, não necessitando de pessoas especializadas, e tem uma ampla aplicação (BARRETO, 2020). Portanto, pesquisas que envolvam desenvolvimento de métodos que facilitem a obtenção dos atributos físico-químicos e classifiquem o estágio de maturação da água de coco, elevarão o crescimento das agroindústrias. Deste modo, este trabalho teve por objetivo principal o desenvolvimento de modelos para a predição de atributos físico-químicos e classificação do estágio de maturação em água de coco com uso da espectroscopia Vis-NIR associada a técnicas de machine learning.

**MATERIAL E MÉTODOS:** As amostras de coco foram adquiridas em duas propriedades no município de Petrolina-PE (09° 23' 55" de latitude sul e 40° 30' 03" de longitude oeste). O clima é classificado como semiárido quente (do tipo BSh na classificação climática de Köppen-Geiger), com regime de chuvas de primavera-verão. Foram utilizadas 200 unidades de cocos (*Cocos nucifera* L.) da variedade anã, sendo 100 unidades no estágio verde (cocos de 5 e 6 meses) e 100 unidades no estágio maduro (cocos de 8 a 10 meses). As análises foram realizadas em dois dias, no primeiro dia foram analisados os cocos verdes, e posteriormente os cocos maduros. Para a retirada da água de coco dos frutos, foi utilizado um facão. Após abertos, a água de coco era filtrada em uma peneira de granulometria de 0,02mm para retirada do bagaço gerado pelo método de abertura, e a água era colocada em recipientes identificados.

Por fim, seguiam para análises espectrais e físico-químicas de referência. A determinação do espectro de refletância foi realizada pelo espectrômetro FieldSpec 3 com faixa de utilização de comprimento de onda de 350 a 2500 nm, resolução de 3 a 10 nm, tempo de varredura de 100 ms e precisão de  $\pm 1$  nm, com acessório Hi-Brite MugLight que possui lâmpada de quartzo-tungstênio-halogênio de 4 W e um recipiente para dispor a amostra com fundo de vidro de diâmetro de 12mm, no qual a refletância das amostras de água de coco eram adicionadas para leitura. O equipamento foi ligado com 30 min de antecedência à aquisição dos espectros para aumentar a homogeneidade dos dados. O sensor óptico era posicionado de forma a ser colocado o volume de amostra a ser lido. Foram removidos das análises os comprimentos de onda 1351 a 2500 devido ao excessivo ruído aleatório. As análises dos parâmetros físico-químicos das amostras de água de coco e de suas diluições foram realizadas em triplicata, seguindo a metodologia padrão da AOAC (Association of Official Analytical Chemists). Foram analisados os parâmetros de acidez titulável, pH e sólidos solúveis. Para a análise de acidez titulável mediu-se 10ml de água de coco com o auxílio de uma pipeta graduada, transferiu-se para um Erlenmeyer e adicionado 2 gotas de fenolftaleína 1% como indicador. Em seguida, titulou-se com hidróxido de sódio (NaOH) 0,1N até a coloração levemente rósea. A acidez foi determinada em g/100g de ácido cítrico da amostra. Para a determinação de sólidos solúveis foi realizada leitura direta em um refratômetro analógico de campo ABBE® com variação de 0 a 30°Brix e precisão de  $\pm 0,2\%$ . Com o auxílio de uma pipeta de Pasteur descartável foram adicionadas 2 a 3 gotas de amostras no prisma do equipamento e realizado a leitura. A análise de pH foi realizada com o uso de um pHmetro digital Quimis®. A leitura foi feita colocando o eletrodo equipamento diretamente na amostra de água de coco. Antes de serem gerados os modelos, foi necessário a aplicação de algumas técnicas de pré-processamento. Para realizar o pré-processamento dos espectros foi utilizado o software Unscrambler X 10.4 (CAMO ASA, Oslo, Noruega) com o objetivo de remoção dos ruídos e melhoramento dos sinais espectrais. Para as correções devido à instabilidade do instrumento, ao espalhamento da radiação durante a realização das medidas e as propriedades físicas das amostras, foram utilizadas as seguintes técnicas: suavização com filtro Savitzky-Golay (SGolay); variação normal padrão (Standard Normal Variate, SNV); e correção multiplicativa de sinal (Multiplicative Signal Correction, MSC). Foi gerada uma base de dados para cada um dos tratamentos de pré-processamento para seguir para o programa com os algoritmos de aprendizagem de máquina junto da base de dados sem nenhum tratamento. Todas as amostras foram utilizadas no processo de calibração e validação cruzada (FERREIRA, 2015). As quatro bases com os dados espectrais sem tratamento e com tratamento foram submetidas aos algoritmos de classificação e regressão no programa Weka (Waikato, Nova Zelândia). Os melhores modelos gerados para cada atributo e tipo de tratamento foram selecionados, utilizando diferentes parâmetros estatísticos para classificação e regressão.

**RESULTADOS E DISCUSSÃO:** Na Tabela 1 estão apresentados os valores médios dos parâmetros de acidez total, pH e sólidos solúveis das amostras de água de coco verde e madura. Pode-se observar a influência do estágio de maturação nos parâmetros físico-químicos da água de coco.

TABELA 1. Resultado das análises físico-químicas da água de coco anã verde e maduro.

Água de coco verde			Água de coco maduro		
AT (g/100g)	pH	SS	AT (g/100g)	pH	SS

<b>Média</b>	0,0945	4,58	5,11	0,0715	5,34	4,38
<b>Máximo</b>	0,1314	4,91	6,20	0,2050	5,85	7,00
<b>Mínimo</b>	0,0691	4,41	4,30	0,0267	4,05	3,00
<b>Desvio Padrão</b>	0,0097	0,1180	0,3805	0,0327	0,3428	0,9328
<b>CV</b>	0,1027	0,0257	0,0743	0,4566	0,0642	0,2129

AT – Acidez titulável. CV – Coeficiente de variação. pH – Potencial Hidrogeniônico. SS – Sólidos Solúveis.

Segundo AROUCHA et al. (2014), os ácidos presentes na água de coco é um atributo que confere sabor característico e é importante para avaliação sensorial e aceitação do consumidor, que varia de acordo com o estágio de maturação e como é conservado o fruto. Os principais ácidos presentes na água de coco são o málico, cítrico e tartárico. Nas amostras analisadas a acidez do coco verde apresentou valores médios de 0,09 g/100g e com menor variação entre as amostras conforme o coeficiente de variação de 0,1027. Ademais, para coco maduro a acidez ficou com uma média de 0,07 g/100g e com maior variação entre as amostras, de acordo como o coeficiente de variação de 0,4566. O pH informa o potencial de íons de hidrogênio (H<sup>+</sup>) presente na água de coco e verifica-se a diminuição desse potencial com a maturação do fruto (Tabela 1). O aumento do pH nas amostras de coco maduro, pH médio de 5,34, reflete a mudança do endosperma líquido (água de coco) para o sólido (albume) em que componentes da água vão se concentrando próximo ao endocarpo e formando a massa branca, que foi observado no coco com maior maturação. O mesmo fenômeno afeta a quantidade de sólidos solúveis presentes na água de coco (SUCUPIRA, 2016). Nas amostras a média de sólidos solúveis na água de coco verde foi de 5,11 °Brix e na água de coco maduro de 4,38 °Brix. De acordo com o descrito na legislação as amostras de água de coco verde, ficaram com os resultados dentro do padrão estabelecido. Já nas amostras de coco maduro uma das amostras de sólidos solúveis ficou abaixo no parâmetro da normativa. Na Tabela 2 estão apresentados os melhores modelos gerados para a classificação dos estádios de maturação do coco dos dados sem tratamento e com os dados tratados. Observa-se que não houve muita diferença entre os modelos para a classificação da água de coco.

TABELA 2. Principais modelos gerados para a classificação dos estádios de maturação na água de coco

	<b>Sem tratamento</b>	<b>MSC</b>
<b>PARÂMETROS</b>	<b>SMO</b>	<b>SMO</b>
<b>Taxa de verdadeiros positivos</b>	0,985	0,990
<b>Taxa de falsos positivos</b>	0,015	0,010

<b>Instâncias classificadas corretamente</b>	197	198
<b>Instâncias classificadas incorretamente</b>	3	2
<b>Estatística Kappa</b>	0.97	0.98
<b>Raiz do erro quadrático médio</b>	0.12	0.10
<b>Erro absoluto relativo</b>	3,00%	2,00%
<b>Erro quadrado relativo de raiz</b>	14.33%	14.33%

MSC – Multiplicative Signal correction (Correção de sinal multiplicativa). PRC – Precision Recall curve (Curva de Rechamada de Precisão). ROC – Receiver Operating Characteristic (Características operacionais do receptor). SMO - Sequential Minimal Optimization (Otimização mínima sequencial). SNV – Standard Normal Variate (Variável Normal Padrão).

A taxa de verdadeiro positivo para o modelo ficou próximo da unidade e mostra a relação das amostras que foram classificadas corretamente, sendo confirmada pelo parâmetro de instâncias classificadas corretamente com 198 de um total de 200 amostras. Tanto o erro médio absoluto, que trata do erro na predição, quanto o erro absoluto relativo, que trata do erro na previsão, mostram a eficiência da utilização da espectroscopia para classificação da maturação do coco. Lima et al. (2017) mostraram em seu trabalho<sup>27</sup> sobre classificação de eucalipto que modelos com valores estatísticos de Kappa acima de 0,8 são considerados bons. Outro ponto importante a se observar é que os dados mesmos sem tratamentos, para retirada dos ruídos, já apresentaram resultados elevados.

**CONCLUSÕES:** A espectroscopia Vis-NIR associada com o algoritmo de modelagem SMO (Regressão de Otimização Mínima Sequencial) e pré-tratamento MSC (Correção Multiplicativa de Sinal), proporcionaram a discriminação dos estádios de maturação e a predição dos atributos acidez titulável, pH e sólidos solúveis em água de coco.

**AGRADECIMENTOS:** Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a UNIVASF pela concessão da bolsa de iniciação científica à primeira autora.

#### **REFERÊNCIAS:**

AROUCHA, Edna Maria Mendes et al. Análise físico-química e sensorial de água-de-coco em função de estágio de maturação das cultivares de coco anão verde e vermelho. **Agropecuária científica no semiárido**, v. 10, n. 1, p. 33-38, 2014.

BARRETO, Naiane Santos. Aplicação da técnica de espectroscopia do visível e infravermelho próximo (VIS/NIR) no controle físico-químico da qualidade de vinhos produzidos no submédio do Vale do São Francisco. 2020.

FAOSTAT. Food and Agriculture Organization of the United Nations. World Production. 2019. Disponível em: <http://faostat.fao.org/> acessado em novembro – 2021.

**Instrução Normativa**, nº 9, de 30 de janeiro de 2020. Estabelecer os padrões de identidade e qualidade para a água de coco. 2020. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO.

LIMA, Elizeu de Souza, *et al.* Classification of the initial development of Eucalyptus using data mining techniques. **CERNE**. Lavras, v. 23, n. 2, Abr/jun 2017.

SUCUPIRA, Natália Rocha. Efeito do processo UHT (Ultra High Temperature) nas características químicas, enzimáticas e sensoriais de água de coco (*Cocos nucifera* L.). 2016.

BRAINER, M. S. C. P.; XIMENES, L. F. Produção de coco: Soerguimento das áreas tradicionais do Nordeste. **Caderno Setorial ETENE**. Ano 5, nº 127. Agosto de 2020.