

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA DE CIÊNCIAS EXATAS E DA COMPUTAÇÃO
GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO



**DETERMINAÇÃO DE CONCENTRAÇÃO DE AMIDO EM *CURCUMA LONGA*
UTILIZANDO *MACHINE LEARNING***

AIRTON CHAGAS DE SOUSA

GOIÂNIA
2021

AIRTON CHAGAS DE SOUSA

**DETERMINAÇÃO DE CONCENTRAÇÃO DE AMIDO EM *CURCUMA LONGA*
UTILIZANDO *MACHINE LEARNING***

Trabalho de Conclusão de Curso apresentado à Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador(a):

Prof. Dr. Clarimar José Coelho

Banca examinadora:

Prof. Dr. Arlindo Rodrigues Galvão Filho

Prof. Dr. Tiago da Silva Curtinhas

GOIÂNIA
2021

AIRTON CHAGAS DE SOUSA

**DETERMINAÇÃO DE CONCENTRAÇÃO DE AMIDO EM *CURCUMA LONGA*
UTILIZANDO *MACHINE LEARNING***

Trabalho de Conclusão de Curso aprovado em sua forma final pela Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Ciência da Computação, em ____/____/_____.

Prof.^a Ma. Ludmilla Reis Pinheiro dos Santos
Coordenador(a) de Trabalho de Conclusão de Curso

Banca examinadora:

Orientador: Prof. Dr. Clarimar José Coelho

Coorientador: Prof. Dr. Arlindo Rodrigues Galvão Filho

Prof. Dr. Tiago da Silva Curtinhas

GOIÂNIA
2021

Dedico este trabalho à minha família, meus amigos, meu orientador e coorientador por terem me dado todo o apoio necessário para que eu chegasse até aqui.

AGRADECIMENTOS

Ao Professor Clarimar, pela orientação, pelos conhecimentos e pelas sugestões que tornaram este trabalho possível.

Ao Professor Arlindo Galvão, pelas aulas ministradas no decorrer do curso, por toda instrução, apoio e confiança repassada.

Ao Isaac Yves, coorientador que, juntamente ao Laboratório de Análise Farmacêutica e Ambiental (LAFAM), disponibilizou os materiais utilizados neste estudo e pelo apoio a este trabalho.

Ao Laboratório de Computação Científica (LCC) por disponibilizar os equipamentos utilizados neste trabalho.

Aos meus colegas Rafael Justino e Walacy Xavier por terem trilhado essa jornada acadêmica comigo, pelo apoio, incentivo e por todos os momentos vivenciados durante o curso.

A todos os meus familiares pelo apoio imprescindível.

A todos que direta ou indiretamente colaboraram para realização deste trabalho.

RESUMO

As indústrias alimentícias e farmacêuticas são estritamente regulamentadas de modo a garantir a qualidade dos produtos fabricados e industrializados. A metodologia habitual aplicada nestes processos envolve procedimentos químicos que podem ser descartáveis e onerosos. Assim, é de grande importância o uso de métodos computacionais para contribuir e otimizar o processo de qualidade. O açafrão-da-terra (*Curcuma longa*) é uma planta herbácea de raiz tuberosa. Em diversos lugares do mundo, sua raiz seca e moída é utilizada principalmente como condimento, corante e fármaco. Há diversos estudos que apontam seus benefícios à saúde, desde atividades antioxidantes e anti-inflamatórias até o auxílio no combate de doenças como câncer, artrite, Alzheimer e doenças cardiovasculares. O açafrão está entre os 7 alimentos mais adulterados nos últimos 30 anos, tendo o amido como seu principal contaminante. Portanto, este trabalho explora o uso de aprendizagem de máquina para a determinação da concentração de amido em imagens digitais de amostras de açafrão-da-terra. Foram testados os seguintes modelos: regressão linear múltipla, árvore de decisão e máquina de vetores de suporte. O modelo de máquina de vetores de suporte obteve os melhores resultados, com raiz quadrática média de 6,52, erro médio absoluto de 5,78 e coeficiente de determinação de 0,93. Conclui-se que máquina de vetor de suporte é uma boa ferramenta para a determinação da concentração de amido para o açafrão-da-terra.

Palavras-Chave: *Açafrão-da-terra, Amido, Curcuma longa, Aprendizagem de Máquina, Regressão Linear Múltipla, Árvore de Decisão, Máquina de Vetores de Suporte.*

ABSTRACT

The food and pharmaceutical industries are strictly regulated in order to guarantee the quality of the manufactured and industrialized products. The usual methodology applied in these processes involves chemical procedures that can be wasteful and costly. Thus, the use of computational methods to contribute and optimize the quality process is of great importance. Turmeric (*Curcuma longa*) is a tuberous root herbaceous plant. In many places around the world, its dried and ground root is used mainly as a condiment, dye, and drug. There are several studies that point out its health benefits, from antioxidant and anti-inflammatory activities to help combat diseases such as cancer, arthritis, Alzheimer's, and cardiovascular diseases. Saffron is among the 7 most adulterated foods in the last 30 years, with starch as its main contaminant. Therefore, this work explores the use of machine learning for the determination of starch concentration in digital images of turmeric samples. The following models were tested: multiple linear regression, decision tree, and support vector machine. The support vector machine model obtained the best results, with a root mean square of 6.52, a mean absolute error of 5.78, and a coefficient of determination of 0.93. It is concluded that support vector machine is a good tool for the determination of starch concentration for turmeric.

Keywords: *Turmeric, Starch, Curcuma longa, Machine Learning, Multiple Linear Regression, Decision Tree, Support Vector Machine.*

LISTA DE ILUSTRAÇÕES

Figura 1 - Diagrama com as etapas do trabalho	12
Figura 2 - Demonstração bidimensional da imagem digital	14
Figura 3 - Demonstração do espaço de cores do RGB	15
Figura 4 - Demonstração do sistema de cores HSV	15
Figura 5 - Demonstração das etapas de ML	16
Figura 6 - Demonstração de uma árvore de decisão	18
Figura 7 - Demonstração do SVM	19
Figura 8 - Transformação de dados no kernel.....	20
Figura 9 – Raiz e pó do açafão-da-terra após processo de secagem e moagem	20
Figura 10 - Amostras dos tipos de açafão-da-terra (de A a J).....	21
Figura 11 - Amostras de açafão-da-terra do tipo A (puro e adição de amido).....	22
Figura 12 - Ambiente de manipulação com amostra de açafão-da-terra	22
Figura 13 - Ambiente de aquisição das imagens digitais	23
Figura 14 - Imagens das concentrações de açafão-da-terra e amido puro.....	23
Figura 15 - Imagem da amostra A puro cortada com tamanho de 150x150 pixels ...	24
Figura 16 - Demonstração do recorte das imagens de 150x150 pixels.....	24
Figura 17 - Demonstração do espelhamento nas imagens de 50x50 pixels	24
Figura 18 - Demonstração da técnica de rotação nas imagens de 150x150 pixels...	25
Figura 19 - Gráfico da PCA com os scores da PC1 x PC2 x PC3.....	26
Figura 20 - RLM: gráfico predito pelo esperado	31
Figura 21 - Árvore de decisão: gráfico predito pelo esperado	34
Figura 22 - SVM: gráfico predito pelo esperado	37

LISTA DE TABELAS

Tabela 1 - RLM: predito pelo esperado e acurácia.....	29
Tabela 2 - Arvore de decisão: predito pelo esperado e acurácia	33
Tabela 3 - SVM: predito pelo esperado e acurácia	36
Tabela 4 - Acurácia dos modelos	38

LISTA DE SIGLAS

ANVISA	Agência Nacional de Vigilância Sanitária
ML	<i>Machine Learning</i>
RGB	<i>Red, Green e Blue</i>
HSV	<i>Hue, Saturation e Value</i>
ROI	<i>Region Of Interest</i>
AI	<i>Artificial Intelligence</i>
SLR	<i>Simple Linear Regression</i>
MLR	<i>Multiple Linear Regression</i>
SVM	<i>Support Vector Machine</i>
LAFAM	Laboratório de Análise Farmacêutica e Ambiental
LCC	Laboratório de Computação Científica
PCA	<i>Principal Components Analysis</i>
RMSE	<i>Root Mean Square Error</i>
MAE	<i>Mean Absolute Error</i>

SUMÁRIO

1	INTRODUÇÃO	11
1.1	ESCOPO DO TRABALHO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	PROCESSAMENTO DIGITAL DE IMAGENS	14
2.1.1	<i>Representação de Imagem Digital</i>	14
2.1.2	<i>Sistemas de Cores RGB e HSV</i>	15
2.1.3	<i>Região de Interesse</i>	16
2.2	<i>MACHINE LEARNING</i>	16
2.2.1	<i>Regressão Linear Múltipla</i>	17
2.2.2	<i>Árvore de Decisão</i>	18
2.2.3	<i>Support Vector Machine</i>	19
3	MATERIAIS E MÉTODOS	20
3.1	ESTUDO DE CASO	20
3.2	AMOSTRAS	21
3.3	AMOSTRAGEM	22
3.4	EXTRAÇÃO DA REGIÃO DE INTERESSE	23
3.5	AUMENTO DE DADOS	24
3.6	ANÁLISE DE COMPONENTES PRINCIPAIS	25
3.7	TREINAMENTO	26
3.8	MÉTRICAS DE AVALIAÇÃO DOS MODELOS	27
4	RESULTADOS	28
4.1	REGRESSÃO LINEAR MÚLTIPLA	28
4.2	ÁRVORE DE DECISÃO	32
4.3	<i>SUPPORT VECTOR MACHINE</i>	35
4.4	COMPARATIVO ENTRE OS MODELOS	38
5	CONCLUSÃO	39
	REFERÊNCIAS	40

1 INTRODUÇÃO

Após a revolução industrial ocorrida entre os séculos XVIII e XIX, as indústrias alimentícias e farmacêuticas aumentaram sua produção em massa, visando números de vendas e lucros cada vez mais altos (ALMEIDA, 2005). Este cenário tornou-se propício o consumo de produtos com um baixo grau de fidelidade (CARLOS, 2012). Como forma de tratativa, o controle de qualidade foi criado no século XX nos Estados Unidos para adotar medidas que parametrizam os requisitos mínimos de qualidade dos produtos (SELEME; STANDLER, 2012). A ANVISA (Agência Nacional de Vigilância Sanitária) foi criada no Brasil em 1999, com o objetivo de ser um órgão responsável por criar normas e regulamentos para fiscalização em todo o país (CONTROLARE, 2020).

A análise química vem sendo utilizada de forma clássica no controle de qualidade pois entrega alto nível de segurança por conta da precisão e eficácia em análises qualitativas e quantitativas (CROWL; LOUVAR, 2001). Essa forma de análise pode se tornar um processo moroso e de um custo elevado pois depende de laboratórios especializados, profissionais qualificados e materiais que muitas das vezes, devem ser descartáveis (PENATTI et al., 2008).

Com o avanço da tecnologia e estudos computacionais, surgiram métodos alternativos que facilitam e aprimoram os processos de controle de qualidade (LUCAS, 2020). Existem estudos que propõem o uso de modelos computacionais como por exemplo a utilização de imagens digitais para identificar impurezas e misturas no pó do café, proposto por Eduardo et al. (2002) e a determinação da qualidade da água usando técnicas de aprendizagem de máquina (*Machine learning*, ML) proposto por Jefferson e Mia (2019).

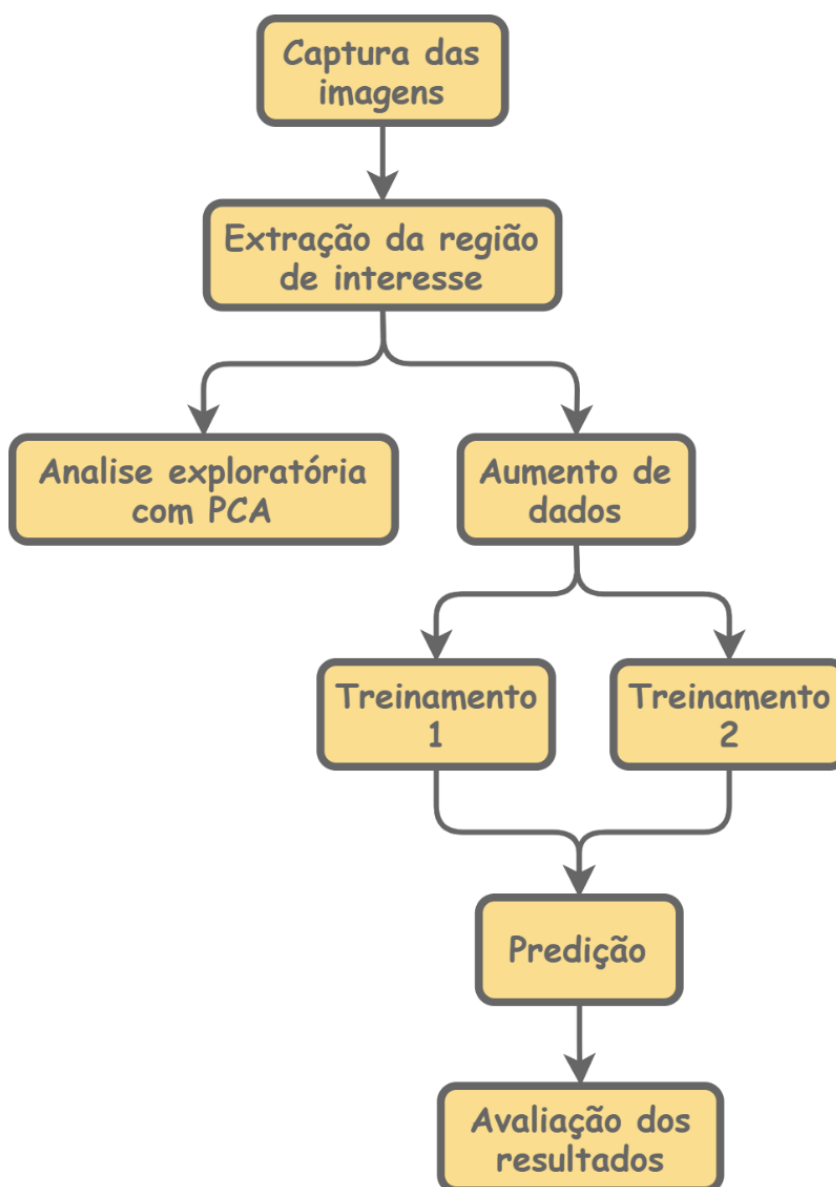
Uma pesquisa publicada no *Jornal of Food Sciences* (JEFFREY et al., 2012) revela que o açafrão-da-terra (*Curcuma longa*) está listado como um dos 7 alimentos mais adulterados nas últimas 3 décadas. Existem diversos estudos que apontam seus benefícios à saúde, que vão desde atividades antioxidantes e anti-inflamatórias, até o auxílio no combate de doenças como câncer, artrite, Alzheimer e cardiovasculares (SINGLETERY, 2010). A pesquisa também apresenta que os principais contaminantes do açafrão-da-terra são o amido, corante e a gelatina. Uma forma comum de adulteração do açafrão é a adição do amido ao seu pó. Isso acontece devido ao amido ser uma substância de baixo custo e já estar presente na composição natural do açafrão (PARVATHY et al., 2015).

Diante disso, este trabalho tem como objetivo o estudo de calibração a partir do uso de imagens digitais em conjunto com técnicas de ML para determinar a concentração de amido em amostras de açúcar-da-terra.

1.1 Escopo do trabalho

Na Figura 1 é apresentado o diagrama de blocos com as etapas que foram seguidas neste trabalho.

Figura 1 - Diagrama com as etapas do trabalho



Fonte: Autoria Própria (2021)

O processo foi iniciado com a captura das imagens, onde foram obtidas 459 imagens. Em seguida foi realizado a extração da região de interesse dessas imagens, no qual cada imagem foi cortada na região central do batoque em um tamanho de 150x150 pixels. As imagens cortadas foram utilizadas em uma etapa de análise exploratória utilizando PCA e em outra etapa de aumento de dados, onde foi utilizado as técnicas de corte, espelhamento e rotação que totalizaram em 6930 imagens. O treinamento foi dividido em duas etapas, na primeira somente a marca A foi utilizada para treinar e testar os modelos. Na segunda todas as amostras da marca A foram utilizadas para treinamento e o teste foi realizado com as demais marcas (B até a J). Ao final foi utilizado as métricas RMSE, MAE e R2 para avaliar a predição dos modelos.

2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica foi dividida em duas seções: Processamento Digital de Imagens e Aprendizagem de Máquina.

2.1 Processamento Digital De Imagens

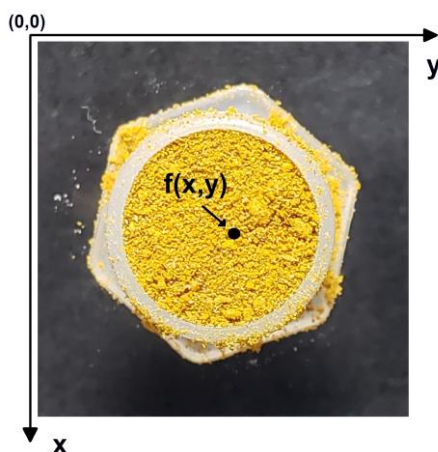
Processamento digital de imagens é um conjunto de técnicas para capturar, representar e transformar imagens por meio de um computador (PEDRINI; SCHWARTZ, 2008).

O processo de aquisição de imagens digitais consiste na captura da imagem utilizando equipamentos que transformam sinais analógicos em dados digitais. Esses equipamentos podem ser câmeras, tomógrafos médicos, scanners, satélites, dentre outros (SCHWARTZ, 2008).

2.1.1 Representação de Imagem Digital

A representação de uma imagem na forma digital é feita através de matrizes, que mapeiam objetos e cores do mundo real para o digital. As matrizes são compostas por elementos pictóricos, também conhecidos como *pixels* (contração de *Picture Elements*) (GONZALEZ; WOODS, 2006). De acordo com Smith (1995), o *pixel* representa um ponto de amostragem nos eixos $f(x,y)$, contendo valores que representam as cores naquela coordenada, conforme ilustra a Figura 2.

Figura 2 - Demonstração bidimensional da imagem digital

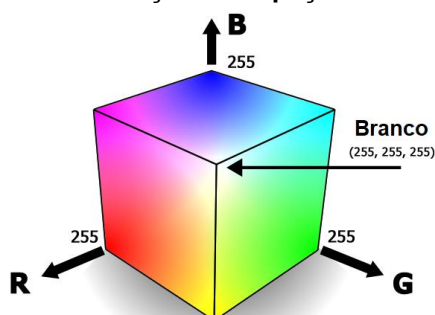


Fonte: Autoria própria (2020)

2.1.2 Sistemas de Cores RGB e HSV

Uma forma de representar imagens digitais coloridas é a utilização do sistema aditivo. Ele é composto por três canais para cada pixel: RGB (*Red Green Blue*), ou em português, Vermelho, Verde e Azul. A combinação dessas três cores pode gerar todas as outras cores (PEDRINI, 2008), como demonstra a Figura 3.

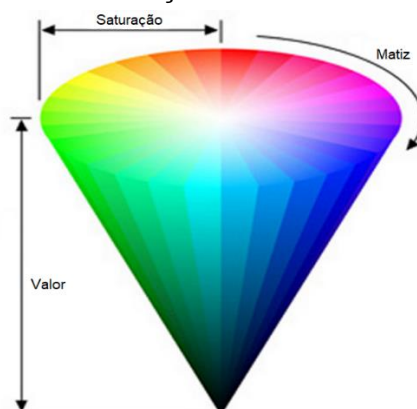
Figura 3 - Demonstração do espaço de cores do RGB



Fonte: Medialooks (2020)

Por haver a necessidade de se especificar numericamente as propriedades das cores e utilizar características de uma cor de forma mais intuitiva, Alvy Ray Smith criou em 1974 (SMITH, 1978) o sistema de cores HSV (*Hue, Saturation and Value*), representado na Figura 4. Esse sistema representa as cores baseadas em matiz, saturação e valor ou brilho. A matiz define a tonalidade dominante, diferenciando o azul do verde, por exemplo. A saturação determina a pureza da cor, como por exemplo, a distinção entre rosa e vermelho. Por fim, o brilho que representa a luminância da cor, representando quão perto do preto a cor está (GONZALEZ; WOODS 2006).

Figura 4 - Demonstração do sistema de cores HSV



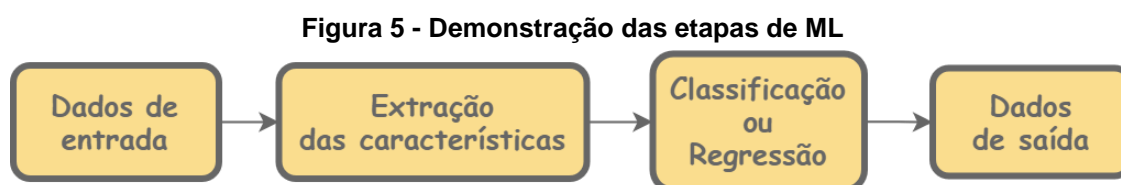
Fonte: Researchgate (2020)

2.1.3 Região de Interesse

A Região de Interesse, do inglês *Region Of Interest* (ROI), é a região da imagem com maior relevância para o experimento ou estudo. Ela pode ser definida de forma manual ou de forma automática, que por sua vez, utiliza características da própria imagem. Um dos seus principais usos é na remoção do fundo da imagem (RUSS, 1995).

2.2 Machine Learning

ML é uma subárea da inteligência artificial (*Artificial Intelligence*, AI) que utiliza técnicas capazes de aprender informações empregando o princípio de indução, no qual é possível obter conclusões genéricas a partir de um conjunto particular de dados (BRINK; RICHARDS, 2014). Na figura 5 está representado o fluxo de atividades de ML.



Fonte: Adaptado de StudyToNight.

Disponível em: <https://www.studytonight.com/post/machine-learning-versus-deep-learning>

A indução é o recurso mais utilizado pelo cérebro humano para a criação de novos conhecimentos. Para algum conceito ser aprendido pela indução, são geradas várias hipóteses de conhecimento a partir dos exemplos analisados que podem ou não ser verdadeiras (MONARD; BARANAUSKAS, 2003).

A quantidade de dados necessária para aplicar a técnica de indução por meio de algoritmos de ML depende diretamente do problema e do modelo utilizado. Dessa forma, na maioria dos casos, quanto maior o volume de dados amostrais maior é a capacidade dos modelos de identificar padrões e gerar melhores resultados (LORENA et al., 2011).

Os tipos mais comuns de ML são: o supervisionado e o não-supervisionado. No supervisionado, o resultado esperado é utilizado pelo modelo a fim de minimizar a diferença entre ele e o resultado obtido (MONARD; BARANAUSKAS, 2003). Já no aprendizado não-supervisionado não são utilizados resultados esperados, assim, os resultados são obtidos a partir dos agrupamentos dos dados e suas similaridades (SHALEV-SHWARTZ e BEN-DAVID, 2014).

Além dos problemas de classificação, onde existe uma quantidade finita de possíveis respostas (quando a amostra é rotulada), também existem os problemas que não possuem respostas finitas (a amostra recebe um valor numérico ou probabilístico). Esses são denominados problemas de regressão (CHAPELLE et al., 2002).

2.2.1 *Regressão Linear Múltipla*

A análise de regressão linear é um modelo estatístico que consiste no estudo entre uma variável dependente y estar relacionada a uma ou mais variáveis independentes (x_1, x_2, \dots, x_n) (ALTMAN; KRZYWINSKI 2015). Um modelo de regressão linear determina uma reta de regressão com a finalidade de minimizar os desvios dos dados reais (BROWN, 2009). Modelos que utilizam somente uma variável independente, são denominados regressão linear simples (*Simple Linear Regression*, SLR), já aqueles que utilizam mais de uma variável independente são denominados regressão linear múltipla (*Multiple Linear Regression*, MLR). A equação da regressão linear múltipla é definida por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i \quad i = 1, 2, \dots, n \quad (1)$$

dado que, y_i seja o valor observado para a variável dependente; β_0 a constante de regressão; β_k os coeficientes de regressão; x_{ik} a variável independente e e_i os valores residuais e possíveis erros.

Para encontrar o vetor de coeficientes de regressão, é utilizada a fórmula dada pela seguinte expressão:

$$\beta = (X^T X)^{-1} X^T Y \quad (2)$$

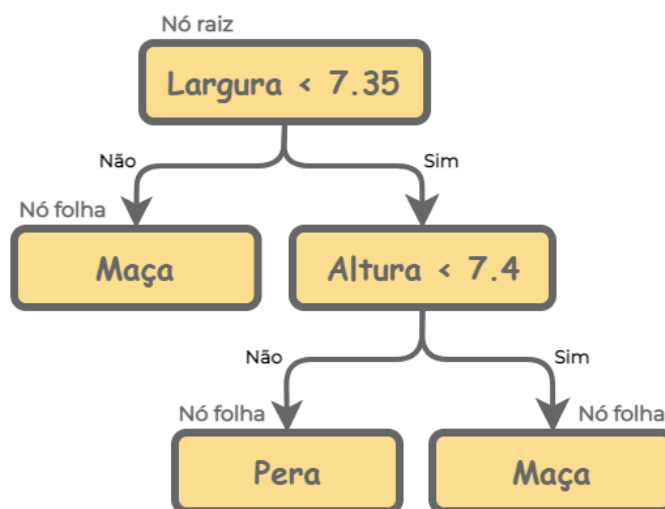
dado que X seja a matriz de dados; X^T seja a matriz de dados transposta e Y o vetor de respostas.

2.2.2 Árvore de Decisão

Árvore de decisão é um modelo de aprendizado de máquina amplamente adotado por ser considerado robusto e escalável (COPPIN, 2010). Em sua construção é utilizado um conjunto de treinamento formado por entradas e saídas, onde as entradas são os dados a serem analisados durante o treinamento e as saídas são as classes ou valores esperados (WITTEN; FRANK, 2000).

Conforme apresentado na figura 6, o modelo é composto por vários pontos de decisão chamados “nós”, em cada um deles há duas possibilidades de saída onde apenas uma delas é escolhida. Esses nós são interligados por caminhos chamados de “ramos”. O primeiro nó é o que contém o conjunto completo dos dados, nomeado de nó raiz. A partir do nó raiz, os dados são subdivididos dentro de outros sub-nós. Já os últimos nós, onde não é mais possível a subdivisão dos dados, são chamados de folha (RODRIGUES, 2015).

Figura 6 - Demonstração de uma árvore de decisão



Fonte: Adaptado de towardsDataScience.

Disponível em: <https://towardsdatascience.com/decision-trees-60707f06e836>

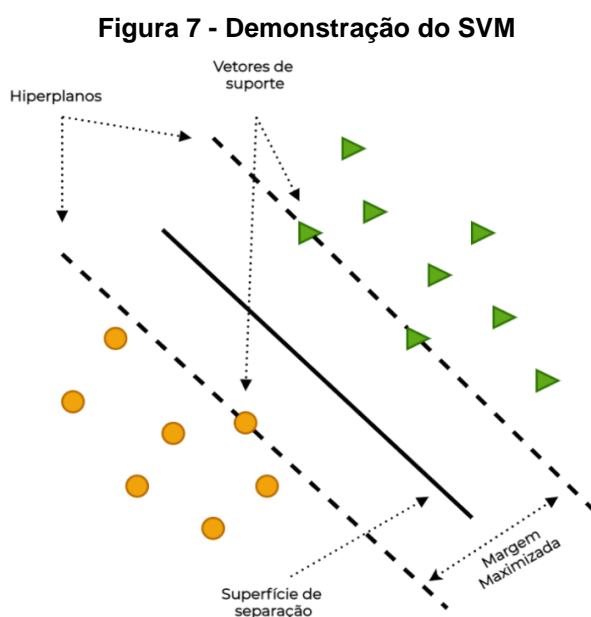
Uma árvore de decisão utiliza a estratégia de “dividir para conquistar”, onde um problema complexo é decomposto em subproblemas mais simples. Esse processo é feito de forma recursiva para cada subproblema, possibilitando a análise de um grande conjunto de dados (WITTEN; FRANK, 2000).

2.2.3 Support Vector Machine

Máquina de Vetores de Suporte (*Support Vector Machine*, SVM), é um modelo de ML utilizado para o reconhecimento de padrões definidos sobre o espaço vetorial. O modelo consiste em separar a superfície entre hiperplanos a partir dos dados de duas ou mais classes (WATTERS, SHEPERD, 2002).

Os hiperplanos que possuem fronteiras de maior margem são definidos como a solução ótima. Essas fronteiras são maximizadas a partir de um subconjunto de pontos que são denominados de vetores de suporte (MANNING; RAGHAVAN; SCHÜTZE, 2009).

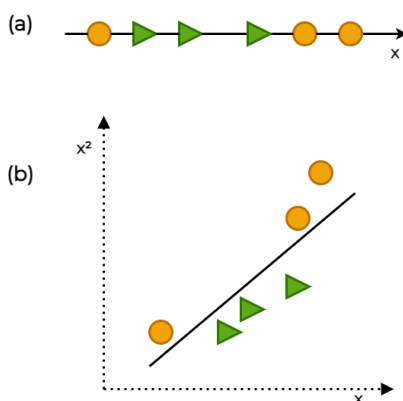
A Figura 7 representa os vetores de suporte de duas classes. A partir desses vetores, a máquina SVM define os hiperplanos de cada classe com a maior margem possível.



Fonte: Adaptado de scikit-learn. Disponível em:
<https://scikit-learn.org/stable/modules/svm.html>

Existem casos em que não é possível obter a separação linear entre os hiperplanos, conforme é apresentado na Figura 8. A solução para este problema é a utilização de funções *kernel*. Estas funções têm a finalidade de transformar os vetores de características em um espaço com maior dimensão e assim, neste novo espaço, é traçado um hiperplano ótimo delimitando as classes (MANNING; RAGHAVAN; SCHÜTZE, 2009).

Figura 8 - Transformação de dados no kernel



Fonte: Adaptado de scikit-learn. Disponível em: <https://scikit-learn.org/stable/modules/svm.html>

3 MATERIAIS E MÉTODOS

3.1 Estudo De Caso

O açafrão-da-terra, também conhecido como *Cúrcuma longa*, é uma planta herbácea da família do gengibre, com sua origem na Ásia (Índia e Indonésia). A sua raiz seca e moída (Figura 9) é utilizada como condimento ou corante amarelo na culinária e no preparo de medicamentos. E sobre sua composição, um estudo de 1996 apontou que 35,3% dela é amido presente de forma natural (CECÍLIO & VILLAS BOAS, 1996).

Figura 9 – Raiz e pó do açafrão-da-terra após processo de secagem e moagem



Fonte: Inteligência Nutritiva. Disponível em: <https://inteligencianutritiva.com/beneficios-da-curcuma/>

Para fins medicinais o açafrão-da-terra se destaca pela ação de antienvhecimento, antioxidante e anti-inflamatória, também auxilia no tratamento de

diversos tipos de câncer. Segundo uma pesquisa da Universidade da Califórnia (LEIGH, 2018), nos Estados Unidos, o açafão-da-terra é capaz de reduzir o risco da doença de Alzheimer.

3.2 Amostras

Foram utilizadas amostras de açafão-da-terra e de amido para o desenvolvimento deste trabalho. Elas foram fornecidas pelo Laboratório de Análise Farmacêutica e Ambiental da Universidade Federal de Goiás (LAFAM). As amostras são compostas por 10 marcas de açafão-da-terra e o amido puro. Para as amostras de açafão, as marcas foram nomeadas com letras de A à J conforme a Figura 10.

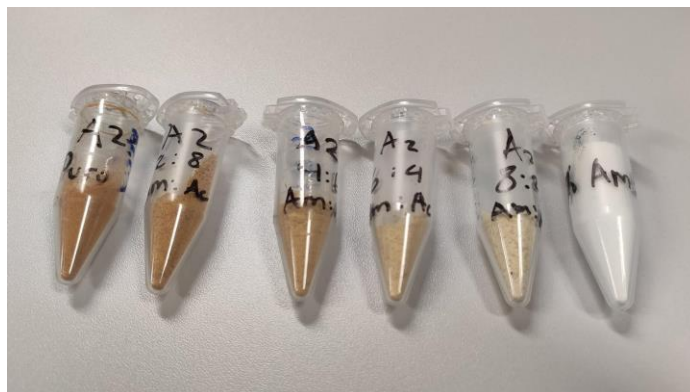
Figura 10 - Amostras dos tipos de açafão-da-terra (de A a J)



Fonte: Autoria Própria (2020)

Para cada tipo de açafão-da-terra, foram disponibilizados 5 níveis de concentração de amido: o açafão-da-terra puro (considerando o amido de forma natural) e outros 4 níveis de adição de amido por açafão (2:8, 4:6, 6:4 e 8:2) como mostra a Figura 11.

Figura 11 - Amostras de açafão-da-terra do tipo A (puro e adição de amido).

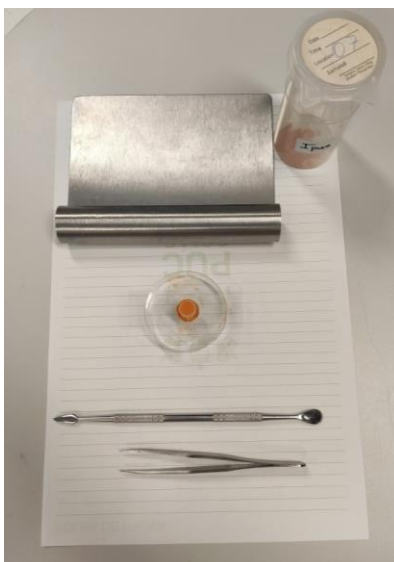


Fonte: Autoria Própria (2020)

3.3 Amostragem

O ambiente de manipulação das amostras (Figura 12) foi preparado com uma folha de papel A4, placas de Petri, lâminas de vidro, batoque de tamanho M, pinças, borrifador com álcool 70% e borrifador com água.

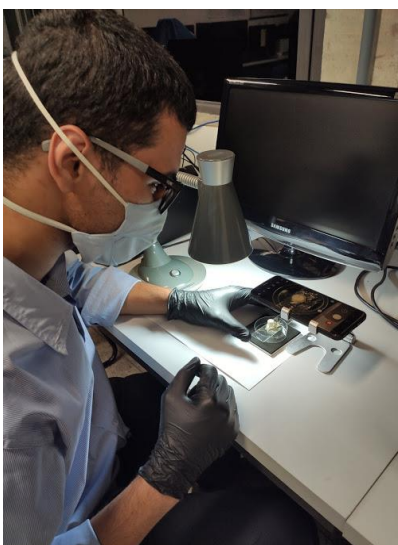
Figura 12 - Ambiente de manipulação com amostra de açafão-da-terra



Fonte: Autoria própria (2020)

O ambiente de captura (Figura 13) foi preparado com uma folha A4, suporte de celular, luminária, superfície preta utilizada para servir como fundo da imagem e um celular Samsung Galaxy S9 Plus para aquisição das imagens.

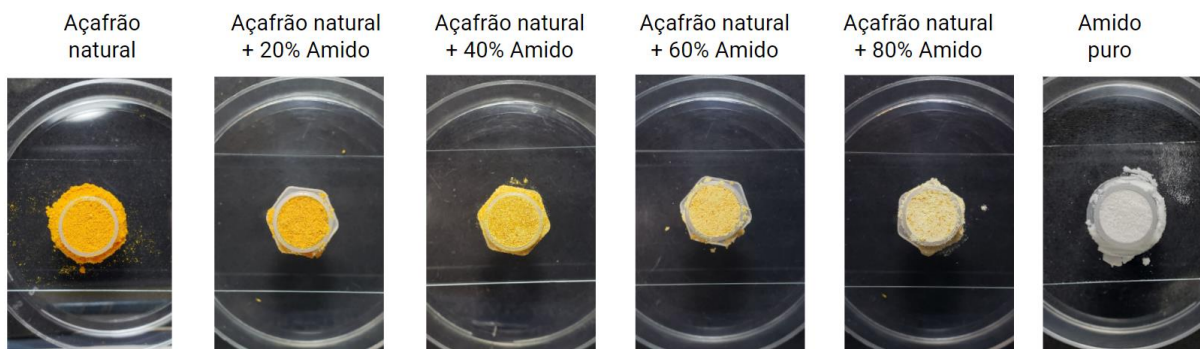
Figura 13 - Ambiente de aquisição das imagens digitais



Fonte: Autoria própria (2020)

Para cada amostra foram preenchidos três batoques, onde cada um teve três aquisições de imagens em rotações diferentes. Como resultado, gerou-se 9 fotos para cada amostra, totalizando 459 imagens. A Figura 14 apresenta uma foto de cada amostra da marca A.

Figura 14 - Imagens das concentrações de açafrão-da-terra e amido puro



Fonte: Autoria própria (2020)

3.4 Extração Da Região De Interesse

Com a finalidade de extrair as informações de maior importância das fotos, elas foram cortadas na região central do batoque, com um tamanho de 150x150 *pixels*, resultando em imagens semelhantes à Figura 15.

Figura 15 - Imagem da amostra A puro cortada com tamanho de 150x150 pixels

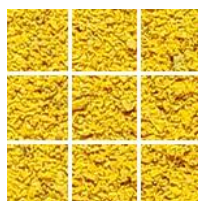


Fonte: Autoria própria (2020)

3.5 Aumento De Dados

Como técnica de aumento de dados, foram utilizados para todas as amostras, cortes das imagens de tamanho 150x150 *pixels* em 9 imagens de tamanho 50x50 *pixels*, totalizando 81 imagens por amostra como apresenta a Figura 16.

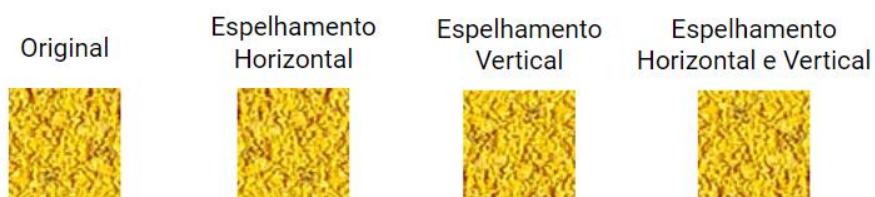
Figura 16 - Demonstração do recorte das imagens de 150x150 pixels



Fonte: Autoria própria (2020)

Após isso, foi utilizada a técnica de espelhamento vertical, horizontal e vertical com horizontal nas imagens de 50x50 *pixels*, resultando em 324 imagens por amostra conforme a Figura 17.

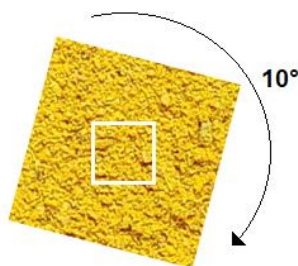
Figura 17 - Demonstração do espelhamento nas imagens de 50x50 pixels



Fonte: Autoria própria (2020)

As imagens de tamanho 150x150 *pixels* também foram utilizadas em uma técnica de rotação, representada pela figura 18, as quais foram rotacionadas 35 vezes de 10 em 10 graus retirado a região central de tamanho 50x50 *pixels* das imagens resultantes.

Figura 18 - Demonstração da técnica de rotação nas imagens de 150x150 pixels



Fonte: Autoria própria (2020)

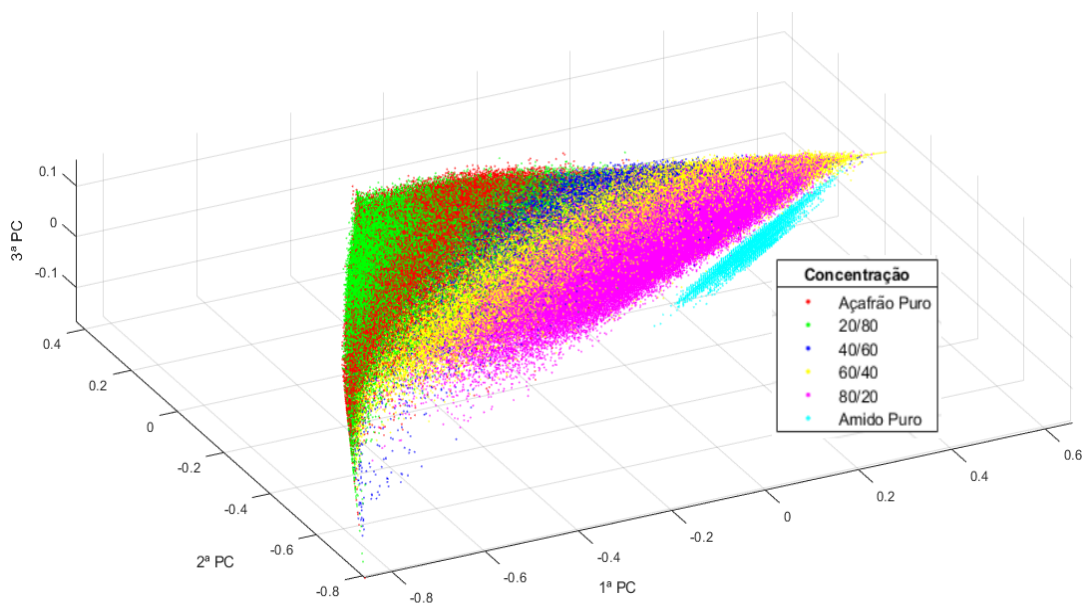
Somando as imagens geradas pelas três técnicas (corte, espelhamento e rotação) a base de treinamento resultou em 639 imagens por amostra.

3.6 Análise De Componentes Principais

A análise de componentes principais (*Principal Component Analysis*, PCA) é uma técnica estatística utilizada para reduzir a dimensionalidade de um conjunto de dados que possui muitas variáveis relacionadas. Essa redução ocorre a partir de um novo conjunto de variáveis não correlacionadas chamadas de Componentes Principais (PCs) (WARNE et al. 2004). As PCs são ortogonais entre si, assim a primeira PC irá obter as informações na direção onde aponta a maior variância, a segunda PC irá obter as informações em um sentido ortogonal a primeira PC e assim sucessivamente (MINGOTI 2005).

Foi realizada uma análise exploratória utilizando a PCA nas amostras de açafreão-da-terra. Na Figura 19 é apresentada os scores da PC1 x PC2 x PC3.

Figura 19 - Gráfico da PCA com os scores da PC1 x PC2 x PC3



Fonte: Autoria própria (2020)

3.7 Treinamento

Para empregar técnicas de aprendizado de máquina, foram utilizados os modelos de Regressão Linear Múltipla, Árvore de Decisão e SVM pelo *software* Matlab versão 2020a.

Os parâmetros utilizados nos modelos foram: valores dos canais RGB e HSV como variáveis independentes no modelo de Regressão Linear Múltipla, 30 nós no modelo de Arvore de Decisão e, para o modelo de SVM, foi utilizada a função *kernel* linear. Todos esses parâmetros foram definidos empiricamente.

O processo de treinamento foi dividido em duas etapas: treino e teste com a marca A e treino com a marca A e teste com as demais. A primeira etapa foi realizada com 60% das imagens para o treinamento, 20% para validação e 20% para teste. Na segunda etapa 80% das imagens foram para treino, 20% para validação e o teste foi realizado com as demais marcas, de B a J. Nas duas etapas, a separação de treino, validação e teste foram realizadas aleatoriamente. O valor real da concentração das amostras foi obtido a partir de uma pesquisa realizada por Macêdo et al. (2020).

3.8 Métricas De Avaliação Dos Modelos

A capacidade de predição dos modelos propostos foi avaliada por três métricas: O erro de raiz quadrático médio (*Root Mean Square Error*, RMSE), onde N é a quantidade de valores totais, \hat{y}_i são os valores preditos e y_i os valores observados. Quanto mais próximo de zero o RMSE for, melhor é o modelo. A Equação 3 apresenta sua fórmula.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

O erro médio absoluto, (*Mean Absolute Error*, MAE), apresentada na Equação 4, onde N é a quantidade de valores totais, \hat{y}_i são os valores preditos e y_i os valores observados. Quanto mais próximo de zero for o resultado, melhor será o modelo.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (4)$$

E o Coeficiente de Determinação (R^2), onde N é a quantidade de valores totais, \hat{y}_i são os valores preditos e y_i os valores observados. No R^2 quanto mais próximo de um estiver o resultado, melhor será o modelo. A Equação 5 apresenta a fórmula do R^2 .

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (5)$$

4 RESULTADOS

Para avaliação dos resultados obtidos neste trabalho, a acurácia de cada modelo foi analisada utilizando as métricas RMSE, MAE e R^2 com os valores esperados fornecido pelo trabalho de Macêdo et al. (2020) e seus respectivos valores preditos. São apresentados a seguir, os resultados de cada modelo utilizado seguido da comparação entre todos eles.

4.1 Regressão Linear Múltipla

Conforme apresentado na Tabela 1, a marca que obteve melhor resultado foi a C com 0,17 de RMSE, 0,13 de MAE e 0,54 de R^2 . O pior resultado foi da marca F com 0,25 de RMSE, 0,21 de MAE e 0,01 de R^2 .

Tabela 1 - RLM: predito pelo esperado e acurácia

Marca	Esperado					Predito					Acurácia		
	X-P \pm DP ^a	X-2 \pm DP ^b	X-4 \pm DP ^b	X-6 \pm DP ^b	X-8 \pm DP ^b	X-P \pm DP ^c	X-2 \pm DP ^d	X-4 \pm DP ^d	X-6 \pm DP ^d	X-8 \pm DP ^d	RMSE	MAE	R ²
A	29,82 \pm 0,56	43,86 \pm 0,14	66,32 \pm 0,12	86,53 \pm 0,08	97,31 \pm 0,07	41,04 \pm 7,5	44,54 \pm 1,1	64,96 \pm 11,6	78,49 \pm 11,3	94,81 \pm 8,6	0,12	0,09	0,74
B	66,05 \pm 0,36	72,84 \pm 0,12	83,7 \pm 0,14	93,48 \pm 0,09	98,7 \pm 0,05	30,93 \pm 9,6	41,14 \pm 9,5	53,52 \pm 9,5	66,45 \pm 8,7	86,24 \pm 5,7	0,3	0,27	0,26
C	33,99 \pm 2,41	47,19 \pm 0,27	68,31 \pm 0,13	87,33 \pm 0,07	97,47 \pm 0,11	35,64 \pm 6,5	42,6 \pm 14,2	53,86 \pm 17,3	71,82 \pm 16,3	90,16 \pm 14,3	0,17	0,13	0,54
D	36,9 \pm 0,27	49,52 \pm 0,11	69,71 \pm 0,1	87,88 \pm 0,07	97,58 \pm 0,02	29,64 \pm 5,2	35,69 \pm 11,8	48,19 \pm 15,8	55,21 \pm 16,1	78,84 \pm 13,5	0,24	0,2	0,28
E	22,14 \pm 0,44	37,71 \pm 0,14	62,63 \pm 0,05	85,05 \pm 0,1	97,01 \pm 0,03	31,11 \pm 5,8	43,38 \pm 13,9	49,51 \pm 15,2	65,04 \pm 13,6	85,14 \pm 11,1	0,18	0,14	0,42
F	27,56 \pm 0,08	42,05 \pm 0,1	65,23 \pm 0,07	86,09 \pm 0,02	97,22 \pm 0,04	31,57 \pm 3,3	32,38 \pm 7,3	40,87 \pm 9,9	51,29 \pm 12,9	68,65 \pm 11,4	0,25	0,21	0,01
G	39,3 \pm 0,61	51,44 \pm 0,12	70,87 \pm 0,1	88,35 \pm 0,03	97,7 \pm 0,05	21,24 \pm 6,4	37,24 \pm 12,3	50,32 \pm 12,1	62,79 \pm 11,3	85,38 \pm 9,3	0,21	0,19	0,5
H	24,21 \pm 1,25	39,37 \pm 0,21	63,62 \pm 0,11	85,45 \pm 0,04	97,09 \pm 0,03	34,76 \pm 6,8	38,92 \pm 12,0	50,44 \pm 13,0	65,83 \pm 15,3	85,31 \pm 13,2	0,18	0,14	0,43
I	53,98 \pm 1,67	63,19 \pm 0,23	77,91 \pm 0,15	91,16 \pm 0,07	98,23 \pm 0,09	22,74 \pm 7,3	34,92 \pm 11,2	48,95 \pm 11,3	60,4 \pm 13,4	84,32 \pm 9,7	0,29	0,27	0,32
J	70,32 \pm 1,11	76,26 \pm 0,22	85,75 \pm 0,18	94,3 \pm 0,08	98,86 \pm 0,08	44,09 \pm 13,1	60,69 \pm 12,3	72,96 \pm 10,4	85,71 \pm 9,5	95,04 \pm 8,4	0,19	0,14	0,44

Fonte: Autoria própria (2020)

^a Amido por 100 gramas de pó de açafração sem mistura \pm desvio padrão.

^b Mistura de X-2 até X-8 correspondente à 2:8 até 8:2 (amido:pó de açafração), respectivamente.

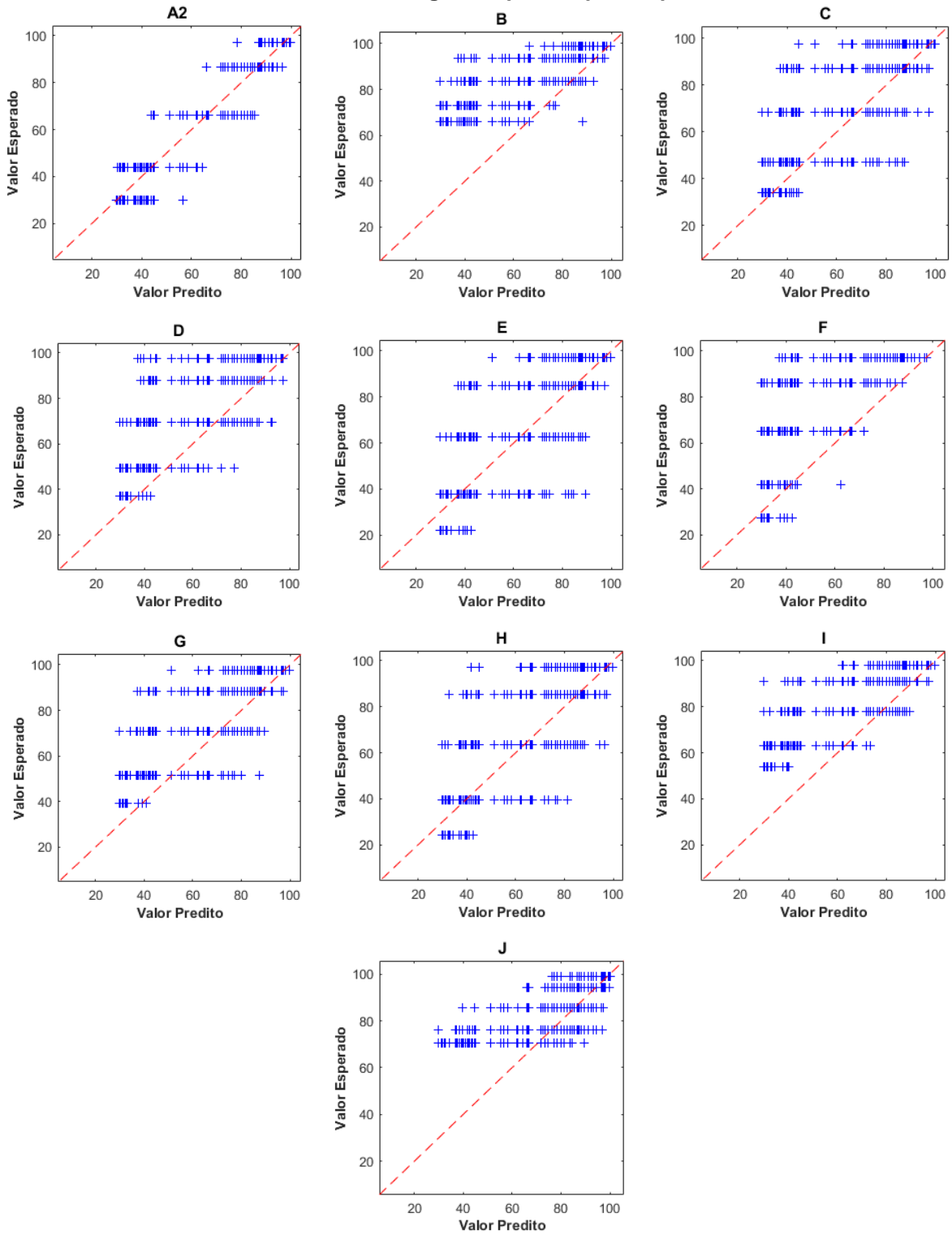
^c Média do resultado da predição da concentração de amido para amostra de açafração sem mistura \pm desvio padrão.

^d Média do resultado da predição da concentração de amido para amostra de açafração com mistura de X-2 até X-8 correspondente à 2:8 até 8:2 amido:pó de açafração, respectivamente \pm desvio padrão.

Nos gráficos da Figura 20, o eixo x representa os valores preditos e o eixo y os valores esperados. A reta diagonal em vermelho representa o valor ideal (quando o valor predito é exatamente o esperado) e cada ponto azul representa o valor predito para cada imagem.

Dentre os modelos testados, RLM foi um dos que apresentaram maiores desvios padrão, com o valor máximo de 16,3. É possível visualizar tal comportamento na Figura 18, onde os valores preditos de mesma concentração estão consideravelmente dispersos no eixo x.

Figura 20 - RLM: gráfico predito pelo esperado



Fonte: Autoria própria (2020)

4.2 Árvore De Decisão

A Tabela 2 mostra que o melhor resultado foi da marca C, onde obteve 0,14 de RMSE, 0,10 de MAE e 0,67 de R^2 e o pior foi da marca F com 0,22 de RMSE, 0,17 de MAE e 0,18 de R^2 .

Tabela 2 - Arvore de decisão: predito pelo esperado e acurácia

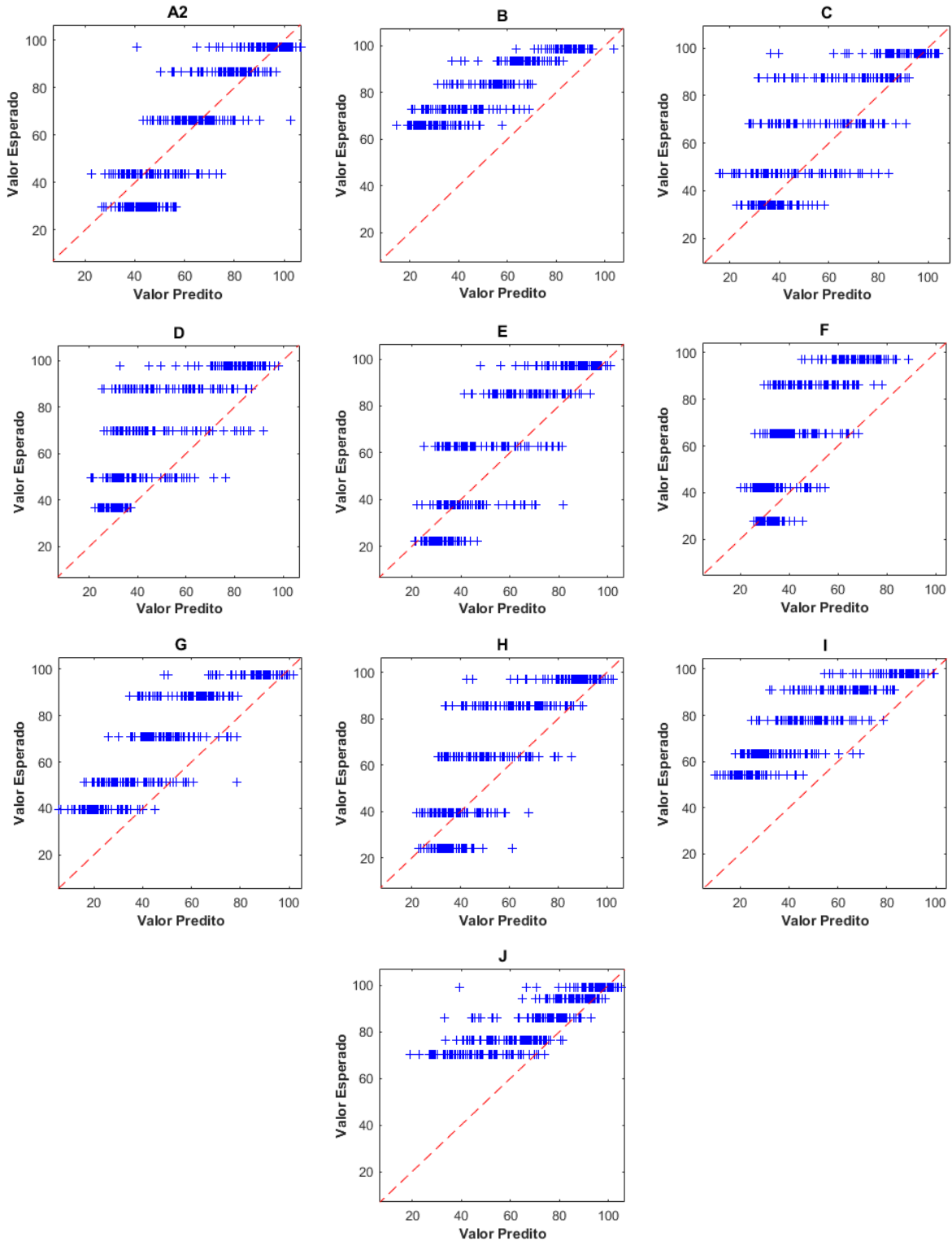
Marca	Esperado					Predito					Acurácia		
	X-P ± DP ^a	X-2 ± DP ^b	X-4 ± DP ^b	X-6 ± DP ^b	X-8 ± DP ^b	X-P ± DP ^c	X-2 ± DP ^d	X-4 ± DP ^d	X-6 ± DP ^d	X-8 ± DP ^d	RMSE	MAE	R ²
A	29,82 ± 0,56	43,86 ± 0,14	66,32 ± 0,12	86,53 ± 0,08	97,31 ± 0,07	42,53 ± 6,2	69,02 ± 8,4	83,66 ± 6,2	95,23 ± 4,8	42,53 ± 6,7	0,08	0,06	0,89
B	66,05 ± 0,36	72,84 ± 0,12	83,7 ± 0,14	93,48 ± 0,09	98,7 ± 0,05	49,6 ± 11,4	61,87 ± 12,3	76,38 ± 12,4	93,54 ± 5,1	49,6 ± 11,3	0,24	0,2	0,34
C	33,99 ± 2,41	47,19 ± 0,27	68,31 ± 0,13	87,33 ± 0,07	97,47 ± 0,11	49,63 ± 12,4	58,96 ± 14,1	77,78 ± 13,3	87,07 ± 10,8	49,63 ± 12,7	0,14	0,1	0,67
D	36,9 ± 0,27	49,52 ± 0,11	69,71 ± 0,1	87,88 ± 0,07	97,58 ± 0,02	40,58 ± 8,1	55,87 ± 13,7	65,74 ± 15,2	83,91 ± 12,2	40,58 ± 8,6	0,18	0,14	0,5
E	22,14 ± 0,44	37,71 ± 0,14	62,63 ± 0,05	85,05 ± 0,1	97,01 ± 0,03	51,71 ± 12,6	57,69 ± 13,2	72,62 ± 12,1	89,13 ± 10,5	51,71 ± 12,4	0,15	0,11	0,58
F	27,56 ± 0,08	42,05 ± 0,1	65,23 ± 0,07	86,09 ± 0,02	97,22 ± 0,04	33,07 ± 5,2	43,08 ± 10,7	55,91 ± 13,6	74,95 ± 12,3	33,07 ± 5,4	0,22	0,17	0,18
G	39,3 ± 0,61	51,44 ± 0,12	70,87 ± 0,1	88,35 ± 0,03	97,7 ± 0,05	49,34 ± 12,2	63,21 ± 11,4	76,06 ± 11,5	88,75 ± 8,7	49,34 ± 12	0,13	0,1	0,61
H	24,21 ± 1,25	39,37 ± 0,21	63,62 ± 0,11	85,45 ± 0,04	97,09 ± 0,03	41,45 ± 9,7	56,93 ± 14,6	72,22 ± 13,8	89,12 ± 9,2	41,45 ± 9,8	0,13	0,1	0,63
I	53,98 ± 1,67	63,19 ± 0,23	77,91 ± 0,15	91,16 ± 0,07	98,23 ± 0,09	44,07 ± 11,6	60,07 ± 13,2	69,42 ± 13,1	89,98 ± 10,4	44,07 ± 11,2	0,22	0,19	0,46
J	70,32 ± 1,11	76,26 ± 0,22	85,75 ± 0,18	94,3 ± 0,08	98,86 ± 0,08	70,36 ± 12	82,97 ± 10,8	91,5 ± 8,3	96,34 ± 4,3	70,36 ± 12,4	0,13	0,09	0,51

Fonte: Autoria própria (2020)

^a Amido por 100 gramas de pó de açafração sem mistura ± desvio padrão.^b Mistura de X-2 até X-8 correspondente à 2:8 até 8:2 (amido:pó de açafração), respectivamente.^c Média do resultado da predição da concentração de amido para amostra de açafração sem mistura ± desvio padrão.^d Média do resultado da predição da concentração de amido para amostra de açafração com mistura de X-2 até X-8 correspondente à 2:8 até 8:2 amido:pó de açafração, respectivamente ± desvio padrão.

Na Figura 21, em que apresenta os valores preditos pelos esperados de cada teste para cada marca, é possível identificar que a marca C expressa maior desvio padrão conforme valores já apresentados na Tabela 2.

Figura 21 - Árvore de decisão: gráfico predito pelo esperado



Fonte: Autoria própria (2020)

4.3 *Support Vector Machine*

A Tabela 3 mostra que a marca C obteve o melhor resultado, com 0,10 de RMSE, 0,08 de MAE e 0,72 de R^2 e a marca F com o pior resultado 0,20 de RMSE, 0,17 de MAE e 0,19 de R^2 .

Tabela 3 - SVM: predito pelo esperado e acurácia

Marca	Esperado					Predito					Acurácia		
	X-P \pm DP ^a	X-2 \pm DP ^b	X-4 \pm DP ^b	X-6 \pm DP ^b	X-8 \pm DP ^b	X-P \pm DP ^c	X-2 \pm DP ^d	X-4 \pm DP ^d	X-6 \pm DP ^d	X-8 \pm DP ^d	RMSE	MAE	R ²
A	29,82 \pm 0,56	43,86 \pm 0,14	66,32 \pm 0,12	86,53 \pm 0,08	97,31 \pm 0,07	33,57 \pm 3,2	41,72 \pm 4,8	68,85 \pm 2,4	83,44 \pm 3,5	92,25 \pm 4,1	0,06	0,05	0,93
B	66,05 \pm 0,36	72,84 \pm 0,12	83,7 \pm 0,14	93,48 \pm 0,09	98,7 \pm 0,05	36,34 \pm 8,4	45,72 \pm 3,6	58,3 \pm 3,1	71,61 \pm 3,4	91,29 \pm 3,8	0,24	0,22	0,31
C	33,99 \pm 2,41	47,19 \pm 0,27	68,31 \pm 0,13	87,33 \pm 0,07	97,47 \pm 0,11	36,2 \pm 2,4	46,65 \pm 5,2	54,88 \pm 6,5	76,21 \pm 5,2	88,25 \pm 4,3	0,1	0,08	0,72
D	36,9 \pm 0,27	49,52 \pm 0,11	69,71 \pm 0,1	87,88 \pm 0,07	97,58 \pm 0,02	35,16 \pm 1,9	39,45 \pm 2,3	52,33 \pm 2,8	62,05 \pm 3,5	81 \pm 4,3	0,17	0,14	0,52
E	22,14 \pm 0,44	37,71 \pm 0,14	62,63 \pm 0,05	85,05 \pm 0,1	97,01 \pm 0,03	32,51 \pm 2,4	47,44 \pm 4,8	54,35 \pm 4,4	70,58 \pm 4,4	84,68 \pm 5,5	0,12	0,11	0,6
F	27,56 \pm 0,08	42,05 \pm 0,1	65,23 \pm 0,07	86,09 \pm 0,02	97,22 \pm 0,04	38,53 \pm 1,6	38,28 \pm 1,7	46,28 \pm 2,8	57,87 \pm 4,9	72,6 \pm 4,8	0,2	0,17	0,19
G	39,3 \pm 0,61	51,44 \pm 0,12	70,87 \pm 0,1	88,35 \pm 0,03	97,7 \pm 0,05	19,06 \pm 7,5	41,59 \pm 3,1	56,95 \pm 4,2	67,72 \pm 4,2	86,49 \pm 4,1	0,17	0,15	0,65
H	24,21 \pm 1,25	39,37 \pm 0,21	63,62 \pm 0,11	85,45 \pm 0,04	97,09 \pm 0,03	34,57 \pm 3,6	39,87 \pm 5,9	51,17 \pm 4,5	69,03 \pm 5,3	85,15 \pm 5,3	0,12	0,11	0,62
I	53,98 \pm 1,67	63,19 \pm 0,23	77,91 \pm 0,15	91,16 \pm 0,07	98,23 \pm 0,09	22,45 \pm 6,4	37,49 \pm 4,3	53,95 \pm 3,2	65,74 \pm 5,8	83,11 \pm 4,5	0,25	0,24	0,43
J	70,32 \pm 1,11	76,26 \pm 0,22	85,75 \pm 0,18	94,3 \pm 0,08	98,86 \pm 0,08	50,17 \pm 5,8	65,09 \pm 5,8	79,7 \pm 4,6	84,94 \pm 3,7	91,5 \pm 4,2	0,13	0,11	0,55

Fonte: Autoria própria (2020)

^a Amido por 100 gramas de pó de açafrão sem mistura \pm desvio padrão.

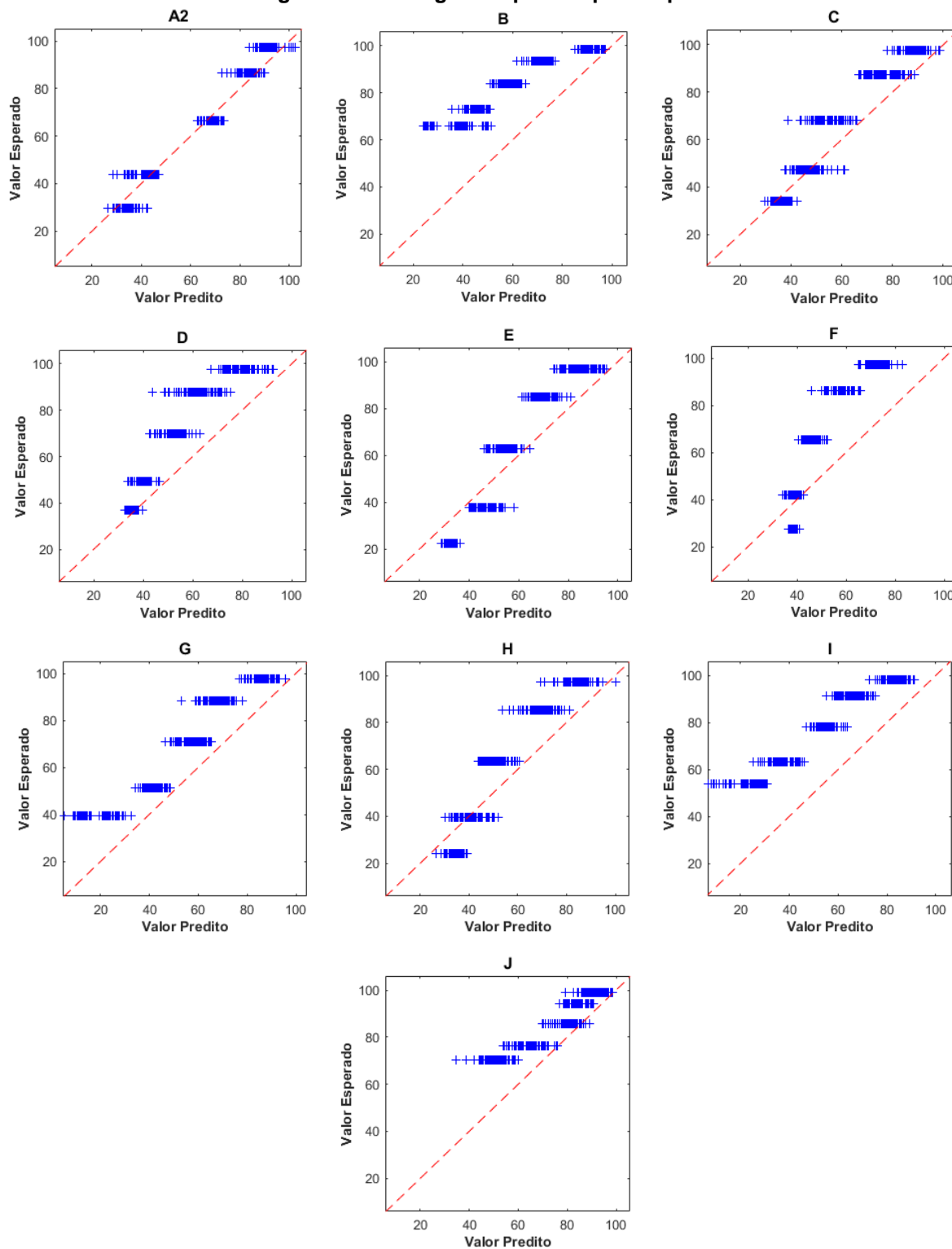
^b Mistura de X-2 até X-8 correspondente à 2:8 até 8:2 (amido:pó de açafrão), respectivamente.

^c Média do resultado da predição da concentração de amido para amostra de açafrão sem mistura \pm desvio padrão.

^d Média do resultado da predição da concentração de amido para amostra de açafrão com mistura de X-2 até X-8 correspondente à 2:8 até 8:2 amido:pó de açafrão, respectivamente \pm desvio padrão.

O modelo SVM foi o que apresentou os melhores desvios padrão, que visualmente é apresentado na figura 22.

Figura 22 - SVM: gráfico predito pelo esperado



Fonte: Autoria própria (2020)

4.4 Comparativo Entre os Modelos

O resultado apresentado na Tabela 4 mostra a comparação entre as métricas obtidas de cada a marca e modelo, e ao final da tabela é apresentado a média destas métricas.

Tabela 4 - Acurácia dos modelos

Marca	RLM			Arvore de Decisão			SVM		
	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²
A	0,12	0,09	0,74	0,08	0,06	0,89	0,06	0,05	0,93
B	0,3	0,27	0,26	0,24	0,2	0,34	0,24	0,22	0,31
C	0,17	0,13	0,54	0,14	0,1	0,67	0,1	0,08	0,72
D	0,24	0,2	0,28	0,18	0,14	0,5	0,17	0,14	0,52
E	0,18	0,14	0,42	0,15	0,11	0,58	0,12	0,11	0,6
F	0,25	0,21	0,01	0,22	0,17	0,18	0,2	0,17	0,19
G	0,21	0,19	0,5	0,13	0,1	0,61	0,17	0,15	0,65
H	0,18	0,14	0,43	0,13	0,1	0,63	0,12	0,11	0,62
I	0,29	0,27	0,32	0,22	0,19	0,46	0,25	0,24	0,43
J	0,19	0,14	0,44	0,13	0,09	0,51	0,13	0,11	0,55
*	0,21	0,18	0,39	0,16	0,13	0,54	0,16	0,14	0,55

Fonte: Autoria própria (2020)

* Média das métricas de todas as marcas e modelos.

5 CONCLUSÃO

O presente trabalho fez uso de métodos computacionais, especificadamente técnicas de processamento digital de imagens e *machine learning*, para determinar a concentração de amido em amostras de açafirão-da-terra. Comparando as métricas aplicadas aos resultados de cada modelo (Regressão Linear Múltipla, Árvore de Decisão e SVM), o SVM obteve o melhor resultado geral, apresentando média entre todas as marcas de 0,16 de RMSE, 0,14 de MAE e 0,55 de R2. O açafirão-da-terra da marca C apresentou melhores resultados e o da marca F apresentou os piores resultados para todos os modelos abordados.

Como sugestão para trabalhos futuros, recomenda-se o uso de outros modelos de aprendizagem de máquina ou então o uso de aprendizado profundo. Outra sugestão é utilizar uma quantidade maior de amostras de dados, assim como a aplicação de melhorias nas técnicas de aumento de dados. Por fim, sugere-se também o treinamento com diferentes marcas de açafirão-da-terra.

REFERÊNCIAS

ALMEIDA, P. R. O Brasil e a nanotecnologia: rumo à quarta revolução industrial. 2005. **Espaço Acadêmico**, Maringá, a. VI, n. 52, set. 2005.

ALTMAN, NAOMI; KRZYWINSKI Martin. **Points of Significance: Simple linear regression. Nature Methods**, v.12, n.11, p.999-1000, 2015

BASU, C.; WATTERS; SHEPERD, M. **Support Vector Machines for Text Categorization**. IEEE Hawaii International Conference on System Sciences, Hawaii, v. 36, jan. 2002.

BRINK, H.; RICHARDS, J. **Real World Machine Learning**. [S.l.]: Manning Publications C.O, 2014.

BROWN, S. H. **Multiple Linear Regression Analysis: A Matrix Approach with MATLAB. Alabama Journal of Mathematics**, p. 1 – 3, 2009.

CARLOS E. **A Era do Consumo**, O tempo. Disponível em: <<https://www.otempo.com.br/opiniaio/meu-dinheiro/a-era-do-consumo-1.202141>> Acesso em: 12 novembro 2020

CONTROLARE. **Vigilância Sanitária é um entre os mais de 10 órgãos reguladores. 2020**. Disponível em: <https://www.controlare.com.br/blog/seguranca-alimentar/vigilancia-sanitaria-e-um-entre-os-mais-de-10-orgaos-reguladores-de-restaurantes>. Acesso em: 02/04/2021.

COPPIN, Ben. **Inteligência artificial**. Rio de Janeiro, RJ: LTC, 2010. 610 p.

CROWL, D.A., LOUVAR, J.F.. **Chemical Process Safety. Fundamentals with Applications**, 2001, Prentice Hall, USA.

EDUARDO D. A.; EDSON E. S.; SILVIA A. R.; TÂNIA B. S.; HILDA R. R. Identificação de Impurezas e Misturas em pó de Café por meio de Comportamento Espectral e Análise de Imagens Digitais; **Pesquisa Agropecuária Brasileira**, 2002

JEFFERSON L.; MIA V. V.; **Internati Pattern Extraction of Water Quality Prediction Using Machine Learning Algorithms of Water Reservoir; Jonal Journal of Mechanical Engineering and Robotics Research** Vol. 8, No. 06, novembro, 2020

JEFFREY C. M.; JOHN S.; MARKUS L *Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010*, *Journal of Food Science*, Vol. 77, Nr. 4, 2012

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma Introdução às *Support Vector Machines*. **Revista de Informática Teórica e Aplicada**, vol.14, no2, pp 43-67, 2007. Disponível em: <http://seer.ufrgs.br/index.php/rita/article/viewPDFInterstitial/rita_v14_n2_p43-67/3543>. Acesso em: 16 outubro. 2020.

LUCAS B. M., Qualidade em tempos de indústria 4.0, **Revista Ferramental**, 2020. Disponível em: <https://www.revistaferramental.com.br/?cod=artigo/qualidade-em-tempos-de-industria-4-0> acesso em: 14 novembro 2020

MACÊDO, Isaac Yves Lopes de; MACHADO, Fabio Bahls; RAMOS, Gabrielle Santos; COSTA, André Gabriel do Carmo; BATISTA, Rayssa Dias; FILHO, Arlindo Rodrigues Galvão; ASQUIERI, Eduardo Ramirez; DE SOUZA, Aparecido Ribeiro; DE OLIVEIRA, Anselmo Elcana; GIL, Eric de Souza. *Starch adulteration in turmeric samples through multivariate analysis with infrared spectroscopy*. **Food Chem**, [S. l.], v. 340, p. 1-8, 25 ago. 2020. DOI <https://doi.org/10.1016/j.foodchem.2020.127899>.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. *Introduction to Information Retrieval*. Online Edition. Cambridge, Inglaterra: Cambridge University Press, 2009. Disponível em: . Acesso em 15 março de 2021.

MINGOTI, SUELI APARECIDA. **Análise de dados através de métodos de estatística multivariada**, uma abordagem aplicada. Editora UFMG 2005.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre Aprendizado de Máquina. In: MONARD, Maria Carolina; BARANAUSKAS, José Augusto. **Sistemas Inteligentes Fundamentos e Aplicações**. Barueri, Sp: Manole Ltda, 2003. p. 39-56. Disponível em: . Acesso em: 14 janeiro 2021.

O. CHAPELLE; V. VAPNIK; O. BOUSQUET; S. MUKHERJEE (2002). *Choosing multiple parameters for support vector machines*. **Machine Learning**. p46. 2002.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, algoritmos e aplicações**. São Paulo: THOMSON Learning, 2008.

PENATTI, F. E.; GUIMARÃES, S. L.; SILVA, P. M. DA. Gerenciamento de resíduos químicos em laboratórios de análises e pesquisa: O desenvolvimento do sistema em laboratórios da área química. **Workshop In-ternacional em Indicadores de Sustentabilidade – WIPIS II**. v. 2, p.106–119, 2008. São Carlos, Brasil: Universidade de São Paulo.

PLATT, JOHN. *Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods*. **Advances in Large Margin Classifiers**. 1999.

RODRIGUES, M. A. S. **Árvores de classificação**. 2005. Monografia - Departamento de Matemática da Universidade dos Açores, Ponta Delgada. 2005.

RUSS, J. C. **The Image Processing Handbook**. 2. ed. [S.l.]: CRC Press, 1995.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning: From Theory to Algorithms**. Cambridge: Cambridge University Press, 2014.

SMITH, ALVY RAY, *A Sprite Theory of Image Computing*, **Tech Memo 5**, Microsoft, Jul 1995.

SMITH, ALVY RAY, *Color Gamut Transform Pairs*, **New York Institute of Technology**, *Conference Proceedings*, agosto 1978.

LEIGH HOPPER, **Universidade Da Califórnia**. Sistema de Bibliotecas. *Curcumin improves memory and mood, new study says*, janeiro 2018. Disponível em: <<https://www.universityofcalifornia.edu/news/curcumin-improves-memory-and-mood-new-ucla-study-says>>. Acesso em: 16 outubro. 2020.

WARNE, K., G PRASAD, N. SIDDIQUE & L. MAGUIRE. **Development of a hybrid pca measurement system for monitoring product quality in the coating industry**. IEEE Internation Conference on Systems, Man and Cybernetics, 2004.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations**. 2nd. ed. San Francisco, CA: Morgan Kaufmann, 2000. ISBN 0-12-088407-0. Citado 13 vezes nas páginas 20, 23, 36, 39, 40, 41, 44, 45, 50, 53, 58, 73 e 77.



**PUC
GOIÁS**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
GABINETE DO REITOR

Av. Universitária, 1069 ● Setor Universitário
Caixa Postal 86 ● CEP 74605-010
Goiânia ● Goiás ● Brasil
Fone: (62) 3946.1000
www.pucgoias.edu.br ● reitoria@pucgoias.edu.br

RESOLUÇÃO n° 038/2020 – CEPE

ANEXO I

APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O(A) estudante Airton Chagas de Sousa
do Curso de Ciência da Computação, matrícula 20161002800621,
telefone: 62994138541 e-mail chagasairton0208@gmail.com, na qualidade de titular dos
direitos autorais, em consonância com a Lei n° 9.610/98 (Lei dos Direitos do autor),
autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o
Trabalho de Conclusão de Curso intitulado
Determinação de concentração de amido em carcuma longa utilizando machine learning
_____, gratuitamente, sem ressarcimento dos direitos autorais, por 5
(cinco) anos, conforme permissões do documento, em meio eletrônico, na rede mundial
de computadores, no formato especificado (Texto (PDF); Imagem (GIF ou JPEG); Som
(WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros, específicos da
área; para fins de leitura e/ou impressão pela internet, a título de divulgação da
produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 09 de junho de 2021.

Assinatura do(s) autor(es):  _____

Nome completo do autor: Clarimar José Coelho

Airton Chagas de Sousa

Assinatura do professor-orientador:  _____

Nome completo do professor-orientador: Clarimar José Coelho