

MODELO DE MACHINE LEARNING PARA CLASSIFICAÇÃO DA FLOR DA PITAYA (*Hylocereus undatus*)

ALAN TOMÉ DE OLIVEIRA¹; CÉSAR AUGUSTO MATOS LADEIRA¹; ELOIZA MARTINS PRIMO CAPELOCI²;

¹ Discente em Big Data no Agronegócio na FATEC Pompéia “Shunji Nishimura”, Pompeia-SP, alan.oliveira33@fatec.sp.gov.br; cesar.ladeira@fatec.sp.gov.br

² Docente em Big Data no Agronegócio na FATEC Pompéia “Shunji Nishimura”, Pompeia-SP, eloiza.capeloci@fatec.sp.gov.br

RESUMO

A utilização de tecnologias capazes de auxiliar o produtor na otimização dos processos, têm sido cada vez mais implementadas na Agricultura. Na Fruticultura, algumas espécies frutícolas que só eram encontradas nos períodos de safra, com ajuda da tecnologia já podem ser produzidas em qualquer época do ano, mesmo na entressafra e em condições desfavoráveis. Este trabalho tem como objetivo a criação de um modelo de Machine Learning (ML) para auxiliar na polinização da flor aberta da pitaya (*Hylocereus undatus*), modelo esse que pode ser utilizado junto a uma Inteligência artificial, com o objetivo de melhorar e otimizar a polinização artificial que por consequência aumenta a produção. A pitaya é uma fruta originalmente encontrada nas regiões Norte, Sul e Central da América e abre sua flor em período predominantemente noturno e o produtor precisa fazer a polinização de forma manual nesse período para que a produção tenha maior qualidade e quantidade, pois os insetos que fazem polinização noturna não conseguem fazê-la de forma tão eficiente quanto o processo manual. O modelo de Machine Learning utilizado na presente pesquisa foi criado no sistema Orange Data Mining, para mostrar a viabilidade e possibilidade da implementação e criação de uma técnica para melhorar a eficiência, qualidade, comodidade e dessa forma aumentar a produtividade do fruto, atendendo as necessidades dos produtores.

Palavras chaves: Aprendizado de Máquina; Pitaya; Polinização artificial; Orange Data Mining.

INTRODUÇÃO

Na sociedade atual onde a transformação digital e tecnológica se faz cada dia mais presente na agricultura mundialmente, a preocupação dos agricultores está no aumento da produtividade e na eficiência operacional das máquinas nas mais diversas áreas do ciclo produtivo, plantio, aplicação de agrotóxicos e na polinização em culturas específicas como a pitaya.

Na fruticultura já são utilizadas tecnologias como sistemas robóticos que trabalham em conjunto com algoritmos de inteligência artificial, resultando em novos métodos para classificar e reconhecer imagens, utilizando imagens multi e

hiperespectrais, sendo possível analisar o conteúdo da água, quantidade de plantas e níveis de nutrientes. Entre as vantagens do uso dessas tecnologias está a detecção de doenças e/ou pragas, pela redução da biomassa, coloração das folhas, através da matriz de coocorrência em imagens e para análise de texturas, o uso de técnicas de inteligência artificial como redes neurais e convolucionais (CNN) artificiais (GASPAR, 2020).

Dessa forma, faz-se necessário cada vez mais tecnologias capazes de otimizar todo o ciclo da cultura e auxiliar na tomada de decisão dos produtores. Para tal, já vem sendo utilizado o uso de algoritmos, Inteligência Artificial (IA), *Big Data* e Machine Learning (FIELDVIEW, 2022).

O aprendizado de máquina é uma área fundamental da IA, onde técnicas computacionais de aprendizado são aplicadas na construção de sistemas, que possam adquirir conhecimento de forma automática e sejam capazes de tomar decisões baseadas em experiências anteriores (MONARD et al. 2003). Em algumas culturas já tem sido utilizado IA em conjunto com técnicas de *Machine Learning* para melhoria e otimização, como por exemplo, controles e identificação de pragas.

Tendo em vista o aprendizado supervisionado, para tal projeto o uso do aprendizado por transferência é essencial para a construção do modelo, sendo este um modelo usado quando se tem um número limitado de exemplos rotulados (BOMBARDELLI, 2021).

Esta pesquisa demonstra a utilização de um modelo de *Machine Learning*, com aprendizado por transferência para criar um modelo que poderá ser usado com uma inteligência artificial para facilitar o manejo da polinização da pitaya (*Selenicereus undatus*), fazendo a identificação da flor aberta para a realização da polinização artificial.

Alguns estudos mostram que a pitaya quando submetida a polinização artificial há uma melhora na qualidade do fruto, sendo necessário que a flor da mesma esteja aberta, algo que só ocorre uma vez por fruto em horários predominantemente noturnos das 20h às 8h da manhã. Fazendo assim, com que seja necessário visitas noturnas as plantações o que muitas vezes ocasiona no incômodo do produtor que realizaria a polinização artificial, já que não se sabe quando a flor da pitaya estará aberta durante seu tempo produtivo que vai de dezembro a maio (MASCARENHAS, 2018).

PITAYA E A POLINIZAÇÃO ARTIFICIAL

A pitaya (*Hylocereus undatus*) ou fruto do dragão (Figura 1) é uma planta pertencente à família das Cactáceas originalmente proveniente da América do Norte, do Sul e Central. As pitayas são comumente caracterizadas como plantas perenes com hábitos de trepadeiras. Entre os diversos tipos do fruto da pitaya, as principais comercializadas são especificamente duas: a de casca vermelha - *Hylocereus undatus* e a de casca amarela - *Selenicereus megalanthus*. A vermelha possui outras 25 espécies, entre elas muitas parecidas (DONADIO, 2009).

Figura 1. Exemplo de pitayas



Fonte: COSTA (2021).

As flores (Figura 2) são laterais com cerca de 20 a 35 cm de comprimento, possuem hábitos de abertura noturnos, elas são brancas, completas e perfumadas, podem ser polinizadas por insetos, porém com a polinização artificial esse processo se torna muito mais eficiente, possui mais de 800 estames em uma única flor, arranjos em duas fileiras, ao redor do pistilo formado por 14 a 28 estiletes de cor creme, os pólenes são predominantemente de cor amarela e as sépalas de cor verde-clara (DONADIO, 2009).

Figura 2. Flor da pitaya aberta.



Fonte: Elaborado pelos autores (2022).

Para que ocorra a polinização cruzada ou autopolinização é necessário que a flor se abra, o que ocorre à noite, e devido ao horário de abertura da flor o procedimento de polinização é dificultado pela ausência de polinizadores naturais, como as abelhas (DONADIO, 2009).

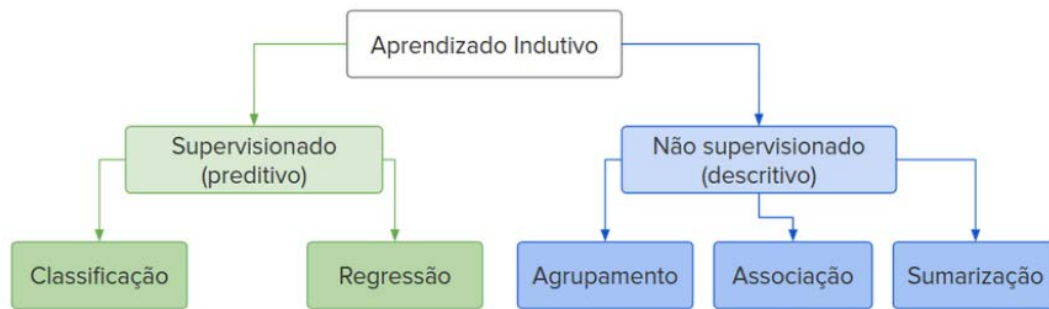
A intenção da polinização artificial na pitaya é garantir maior pegamento, melhor qualidade das frutas e, conseqüentemente, resultar em uma maior produtividade. Isso porque alguns fatores dificultam sua polinização, como a distância entre estigma (parte feminina da flor) e estames (parte masculina), influência climática. Quando ocorrem chuvas, a polinização é inviabilizada, além de não ocorrer a posterior fecundação; outro detalhe é que o estigma fica muito acima dos estames então, há necessidade de transportar esse pólen e a polinização artificial assegura esse procedimento o melhor pegamento das frutas, e melhoria na qualidade, tanto na aparência quanto no tamanho. (MASCARENHAS, 2018).

TIPOS DE APRENDIZADO

O aprendizado de máquina é uma área da IA, que objetiva a criação de algoritmos e técnicas computacionais que possibilitem a tomada de decisão pelas máquinas, sendo dividida em duas categorias de aprendizado a supervisionada e a não supervisionada, quando os exemplos são rotulados, o mais indicado dos modelos a serem utilizados é o supervisionado, pois assim os dados identificam padrões e induzem a um resultado de acordo com a base de dados, por outro lado

quando os exemplos não são rotulados, usa-se o aprendizado não supervisionado que identifica padrões nos dados buscando por similaridades entre os mesmos (SANCHES, 2003).

Figura 3. Tipos de Aprendizado.



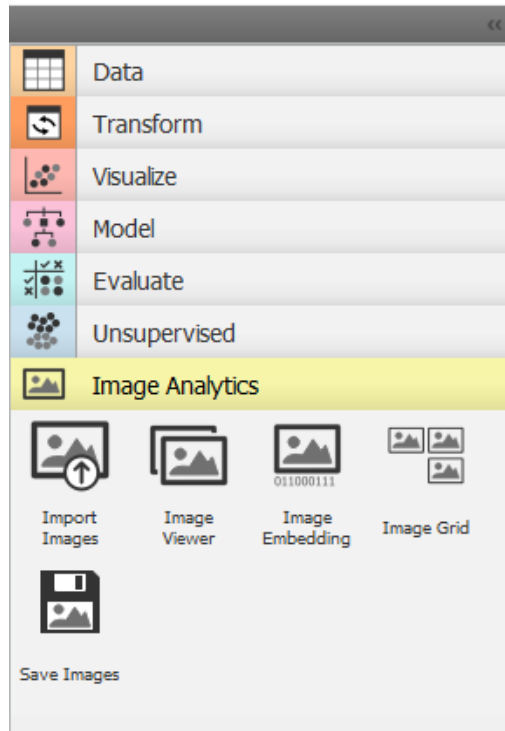
Fonte: BARRETO (2018).

MATERIAL E MÉTODOS

O projeto foi elaborado para ser um modelo de aprendizado de máquina (Machine Learning) que possa ser implementado numa Inteligência Artificial. Para o início do projeto foi feita a instalação do Orange Data Mining para facilitar o desenvolvimento do modelo de Machine Learning. O Orange Data Mining é uma ferramenta de código aberto, que por sua vez ajuda na criação do modelo, sem a necessidade de utilizar código; possui uma interface de fácil utilização, além de ter várias funções, algoritmos de Machine Learning e inúmeros métodos matemáticos que facilitam a implementação de projetos voltados para dados, modelos entre outras aplicações (BATISTA, 2019).

Além disso a ferramenta possibilita aprender e tratar dados de forma intuitiva e dinâmica, a ferramenta disponibiliza ícones (e widgets) para *Data*, *Model*, *Evaluate* sendo estes dados, visualização, modelagem e não supervisionado, respectivamente entre outras, sendo possível utilizar as funcionalidades, ou seja, os widgets apenas arrastando para o centro da tela onde deseja usá-lo. No desenvolvimento do projeto essa ferramenta foi utilizada em todo o processo de aplicação e treinamento do modelo.

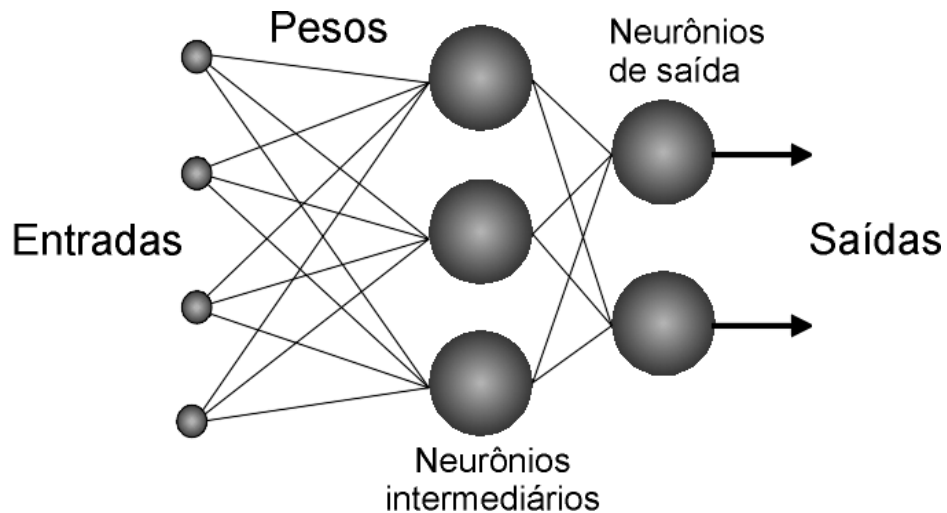
Figura 4. Exemplos de *widgets* Orange.



Fonte: Orange Datamining (2022).

Foram utilizados algoritmos para o desenvolvimento do modelo. O algoritmo Neural Network foi utilizado inicialmente como teste para classificar as imagens das flores da pitaya. O Neural Network é um método de Machine Learning que ensina computadores a processar dados de uma forma semelhante ao cérebro humano. É um tipo de processo de Machine Learning que usa neurônios interconectados em uma estrutura em camadas, semelhante ao cérebro, criando uma rede neural. Sendo um sistema adaptável que os computadores usam para aprender com os erros e se aprimorar continuamente como um humano. As redes neurais artificiais tentam solucionar problemas diversos, como classificar se algo está aberto ou fechado com grande precisão (AWS, 2022).

Figura 5. Algoritmo *Neural Network*.



Fonte: Brain & Mind (1998).

Também foi utilizado para testes o algoritmo Logistic Regression que é um modelo estatístico usado para determinar a probabilidade de um evento acontecer. Ele mostra a relação entre os recursos e, em seguida, calcula a probabilidade de um determinado resultado, possibilitando obter previsões precisas (TIBCO, 2022).

O algoritmo Random Forest é descrito como supervisionado resolvendo problemas de regressão e classificação. Este algoritmo faz o agrupamento das previsões de várias árvores de decisão, treinadas cada uma individualmente, com o objetivo de criar o melhor modelo preditivo para resolver um mesmo problema, diminuindo a variância. Desta forma esse modelo funciona com exatidão em grandes conjuntos de dados, usando as decisões finais de cada "nó" para chegar a uma conclusão própria, encontrando assim uma média entre os valores. Utiliza do coeficiente *Geni* para identificar qual ramificação é mais provável de acontecer, desta forma o algoritmo analisa o resultado de várias árvores de decisão e traz um novo resultado como resposta com base nessas árvores (LEITE; MORAES; LOPES, 2020). Foi também utilizado para os testes de classificação.

O algoritmo Tree (árvore de decisão) foi também testado. É do tipo supervisionado, utilizado tanto para classificação quanto para regressão, ou seja, pode ser usado para prever variáveis categorias como por exemplo sim ou não ou valores numéricos como lucros por exemplo. Ele cria vários pontos de decisão sendo eles os "nós" da árvore, onde em cada um o resultado será seguir por um caminho(decisão), os caminhos neste caso são os ramos, sendo este mais eficiente

com um menor número de dados, com uma árvore de decisão simples imagem(SACRAMENTO, 2021).

O iLoveIMG é um editor de imagem e foi usado para padronizar o tamanho das imagens, pois nas tentativas de classificar as imagens, os algoritmos usam diversas formas. Uma maneira de realizar isso é pelo tamanho da imagem onde caso exista um padrão de tamanho entre as imagens pode acabar viciando o modelo, no nosso caso se as imagens de flor aberta por coincidência sempre tiverem uma maior resolução que as imagens de flor fechada ao colocar uma imagem de flor fechada com uma grande resolução pode ocasionar o erro do modelo o fazendo acusar ser uma flor aberta, fazendo assim que essa padronização seja de suma importância.

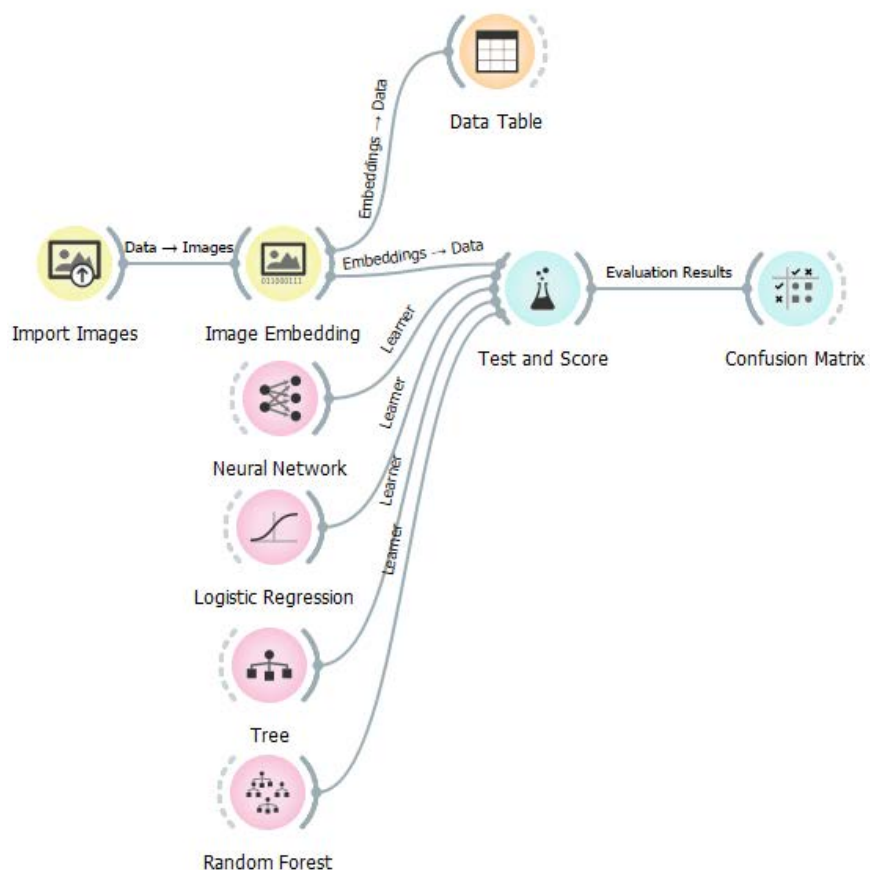
No decorrer da criação deste projeto, os algoritmos de Machine Learning mostrados foram usados para fazer um teste de precisão com as imagens da flor da pitaya para que fosse possível avaliar a eficiência e precisão dos mesmos e escolher o que melhor se ajusta ao modelo, na questão de classificar a flor da pitaya aberta. Dessa forma, foi de grande importância conhecer os conceitos de aprendizado dos algoritmos de Machine Learning.

Foi feito uma comparação entre os algoritmos para que fosse possível, identificar aquele com o melhor desempenho e ser possível implementá-lo, no modelo, e assim, fosse possível analisar como os algoritmos se comportaram, para prever qual seria o estado da flor da pitaya, aberta ou fechada e assim concluir qual o algoritmo que seria o melhor para ser utilizado em nosso modelo de Machine Learning.

RESULTADOS E DISCUSSÃO

O modelo foi criado de acordo com o esquema demonstrado na figura 6, visando descobrir qual algoritmo teria o melhor desempenho, na identificação da flor aberta ou fechada da pitaya.

Figura 6. Modelo esquema de Machine Learning Orange.



Fonte: Elaborado pelos autores (2022).

Neste modelo uma pasta é importada tendo dentro duas outras pastas contendo as imagens, em nosso caso uma com o nome Aberta e outra com o nome Fechada, tendo as imagens dentro dessas pastas um tamanho padrão de 200px para que o tamanho das imagens não influencie nas decisões dos algoritmos, onde a quantidade de pastas seriam as categorias. É possível analisar na figura 7 que existem dez flores abertas e dez flores fechadas respectivamente, que passam pelo Image Embedding (incorporação de imagem), onde essas imagens são transformadas em dados, sendo possível serem visualizados no Datatable (tabela de dados) ilustrado na figura 7.

Figura 7. Data Table para visualização da Base de Dados.

hidden origin	category	image name	image	size	width	height
1	AbertaP	10A	AbertaP\10A.jpg	9931	200	200
2	AbertaP	1A	AbertaP\1A.jpg	20550	200	200
3	AbertaP	2A	AbertaP\2A.jpg	8106	200	200
4	AbertaP	3A	AbertaP\3A.jpg	23265	200	200
5	AbertaP	4A	AbertaP\4A.jpg	13993	200	200
6	AbertaP	5A	AbertaP\5A.jpg	8434	200	200
7	AbertaP	6A	AbertaP\6A.jpg	10794	200	200
8	AbertaP	7A	AbertaP\7A.jpg	11392	200	200
9	AbertaP	8A	AbertaP\8A.jpg	7465	200	200
10	AbertaP	9A	AbertaP\9A.jpg	8589	200	200
11	FechadaP	10F	FechadaP\10F.j...	9617	200	200
12	FechadaP	1F	FechadaP\1F.jpg	6671	200	200

Fonte: Elaborado pelos autores (2022).

Logo após a transformação das imagens na figura 7, os dados passam pelo Test and Score (teste e pontuação), onde o treinamento é feito com os algoritmos Neural NetWork, Logic Regression, Tree e Random Forest, para que o modelo possa identificar se a flor está aberta ou fechada e ser possível analisar qual dos algoritmos teve a maior precisão na identificação da flor, que neste teste foram os algoritmos Neural NetWork e Logic Regression com 0,857 para ambos como mostrado na figura 8.

Figura 8. Resultado do Modelo.

Test and Score - Orange

Number of folds: 3

Stratified:

Cross validation by feature:

Random sampling:

Repeat train/test: 10

Training set size: 66 %

Stratified:

Leave one out:

Test on train data:

Test on test data:

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.990	0.900	0.899	0.917	0.900
Logistic Regression	0.950	0.900	0.899	0.917	0.900
Random Forest	0.875	0.750	0.749	0.753	0.750
Tree	0.650	0.650	0.649	0.652	0.650

Compare models by: Area under ROC curve

Negligible diff.: 0.1

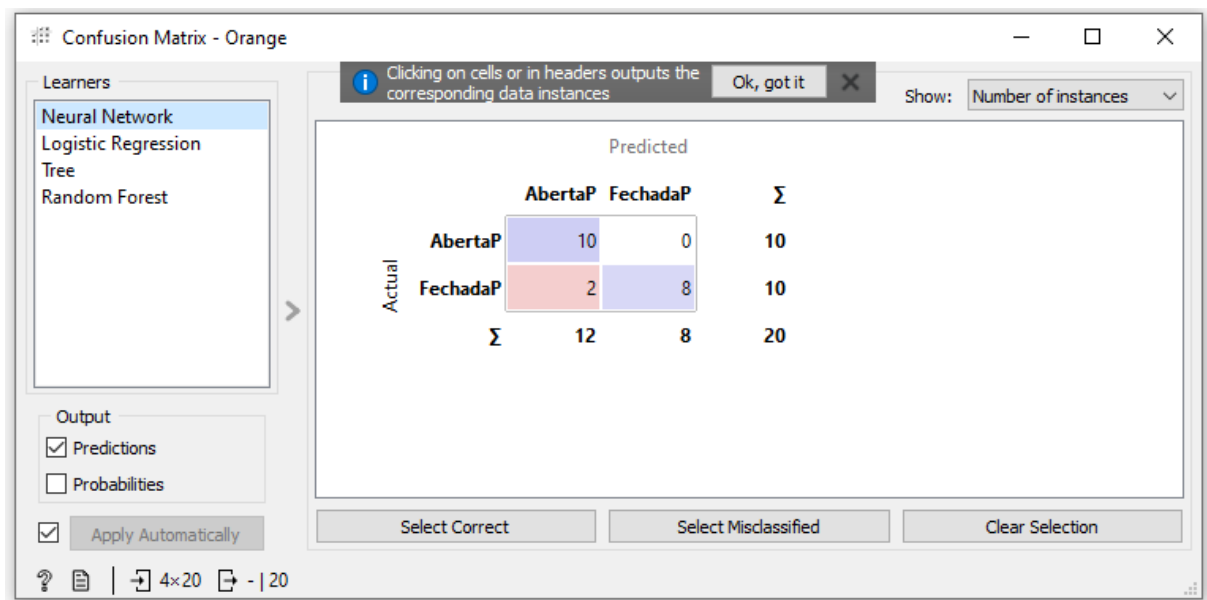
	Tree	Random Forest	Neural Network	Logistic Regressi...
Tree		0.068	0.013	0.013
Random Forest	0.932		0.187	0.187
Neural Network	0.987	0.813		0.500
Logistic Regression	0.987	0.813	0.500	

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

Fonte: Elaborado pelos autores (2022).

De acordo com o processo que identificou a precisão dos algoritmos apresentados na figura 8, foi gerado uma Confusion Matrix (matriz de confusão), para que essas informações fossem mais fáceis de visualizar. Sendo mostrado apenas o Neural Network que foi o algoritmo que teve o melhor desempenho/precisão, onde de dez flores abertas conseguiu identificar todas e de dez flores fechadas identificou oito, como mostrado na figura 9.

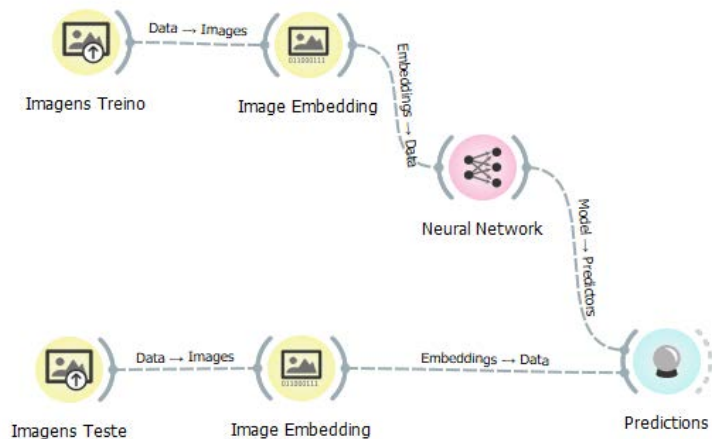
Figura 9. Matriz de Confusão.



Fonte: Elaborado pelos autores (2022).

Após ver que o modelo teve um bom desempenho foi testado, através dos componentes import images, que realizou a importação da imagem para teste, o image embedding que faz a transformação dos dados e o Neural network, que foi o algoritmo que teve a melhor precisão entre os usados e por fim temos a implementação do Prediction que faz a predição dos dados, ou seja demonstra com base no algoritmo, se a imagem da flor está aberta ou fechada, sendo possível ver o processo de treinamento de forma simples, como mostra a figura 10.

Figura 10. Teste do modelo.



Fonte: Elaborado pelos autores (2022).

Como visto anteriormente o que teve melhor resultado nessa análise foi o Neural Network, por isso apenas esse algoritmo foi usado nessa etapa, para o treino foram usadas as imagens da base anterior, e para o Teste novas imagens para que assim o modelo tente fazer sua classificação com base nas informações que possui e não por ter uma imagem exatamente igual àquela que já foi usada para o treinamento, como pode ser observado na figura 11.

Figura 11. Visualização do Predictions

Predictions - Orange						
Show probabilities for (None) Restore Original Order						
Neural Network	image name	image	size	width	height	n0
1 Fechada	teste1F	teste1F.jfif	7930	248	203	0.447501
2 Aberta	teste2A	teste2A.jfif	97754	1201	1600	0.089298
3 Fechada	teste3F	teste3F.jfif	8574	300	168	0.110632
4 Aberta	teste4A	teste4A.jfif	9883	259	194	0.0973898
5 Aberta	teste5A	teste5A.jfif	4466	275	183	0.0328861

Fonte: Elaborado pelos autores (2022).

De acordo com o teste é notório a eficiência do modelo nesse caso, onde o resultado dos testes foi perfeito, com o modelo sendo capaz de classificar as cinco imagens de forma correta. Mostrando assim ser um modelo eficiente para a identificação da flor da pitaya, estando ela aberta ou fechada.

CONCLUSÕES

Diante do trabalho realizado, a utilização do modelo de Machine Learning para a classificação da flor da pitaya se torna totalmente viável e possível de ser realizado, tendo em vista que o processo faz com que o tempo para a verificação da flor seja preciso. O modelo melhora o processo de polinização artificial, que é uma das etapas mais importantes do ciclo da cultura, sendo necessário ter uma base de dados relativamente grande com imagens da flor aberta e fechada.

Dessa forma é viável a utilização de tecnologias como Machine Learning e Inteligência Artificial, no cultivo de pitaya, pois assim, é possível aumentar a produção e qualidade do fruto, com maior manejo do processo de polinização.

Como aprimoramento do modelo será implementado câmeras na área de cultivo de pitaya e a criação de um aplicativo para que assim que as flores de pitaya se abram, o produtor seja notificado para a realização da polinização artificial, não sendo necessário monitorar as flores presencialmente durante todo seu ciclo produtivo.

REFERÊNCIAS

AWS. **O que é uma rede neural? Guia de IA e ML - AWS.** 2022. Disponível em: <<https://aws.amazon.com/pt/what-is/neural-network/>>. Acesso em: 22. set. 2022.

BARRETO, Cephias Alves da Silveira. **Uso de Técnicas de Aprendizagem de Máquina para Identificação de Perfis de Uso de Automóveis baseados em Dados Automotivos.** 2018. Disponível em: https://www.researchgate.net/publication/335988002_Uso_de_Tecnicas_de_Aprendizado_de_Maquina_para_Identificacao_de_Perfis_de_Uso_de_Automoveis_Baseado_em_Dados_Automotivos. Acesso em: 12. out. 2022.

BATISTA, B. **Machine Learning sem código.** 2019. Disponível em: <<https://medium.com/ensina-ai/machine-learning-sem-c%C3%B3digo-636d1a8f9081#:~:text=Orange%20Data%20Mining%20%C3%A9%20uma>>. Acesso em: 03. out. 2022.

BOMBARDELLI, J. C. **Como o aprendizado por transferência pode tornar o Machine Learning mais eficiente** - Blog da Gama Academy. 2021. Disponível em: <<https://blog.gama.academy/aprendizado-por-transferencia-machine-learning/>>. Acesso em: 03. out. 2022.

COSTA, Márcia. **Colheita de pitaya deve crescer 30% em SC**. 2021. Disponível em: <<https://economiasc.com/2021/01/15/colheita-de-pitaya-tem-previsao-de-crescimento-de-30-em-sc/>>. Acesso em: 15. out. 2022.

DONADIO, L. C. **Pitaya**. Revista Brasileira de Fruticultura, v. 31, n. 3, set. 2009. Disponível em: <<https://www.scielo.br/j/rbf/a/cq7JmmnPgWXb369PYvmMKYn/?lang=pt>>. Acesso em: 02. out. 2022.

FIELDVIEW. **Como usar inteligência artificial na agricultura. 2022**. Disponível em: <<https://blog.climatefieldview.com.br/inteligencia-artificial-agricultura#:~:text=O%20que%20%C3%A9%20intelig%C3%Aancia%20artificial%20na%20agricultura%3F>>. Acesso em: 10. out. 2022.

GASPAR, D, P. **Fruticultura 4.0 : Novas tecnologias na fruticultura**. 2020. Disponível em: <<http://vozdocampo.pt/2020/06/08/fruticultura-4-0-novas-tecnologias-na-fruticultura/>> Acesso em: 15. out. 2022.

IBM. **O que são Redes Neurais?** . 2020. Disponível em: <<https://www.ibm.com/br-pt/cloud/learn/neural-networks>>. Acesso em: 22. set. 2022.

ILOVEPDF. **ILoveIMG | ferramentas online para edição de imagens**. 2022. Disponível em: <<https://www.iloveimg.com/pt#:~:text=O%20iLoveIMG%20%C3%A9%20a%20solu%C3%A7%C3%A3o>>. Acesso em: 23 out. 2022.

LEITE, D. R. A; MORAES, R. M. DE; LOPES, L. W. **Método de Aprendizado de máquina para Classificação da intensidade do desvio vocal utilizando Random Forest**. Journal of Health Informatics, v. 12, 2020. Disponível em: <<https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/814/456>>. Acesso em 22. set. 2022.

MASCARENHAS, Karina. **Polinização artificial na pitaia garante melhor qualidade das frutas na colheita e rentabilidade ao produtor** | Arquivo de notícias anteriores a Jan/2018 – DCOM UFLA. Disponível em: <<https://www.ufla.br/dcom/2018/02/23/polinizacao-artificial-na-pitaia-garante-melhor-qualidade-das-frutas-na-colheita-e-rentabilidade-ao-produtor/>>. Acesso em: 01. out. 2022.

SACRAMENTO, G. **Árvore de decisão: entenda esse algoritmo de Machine Learning**. 2021. Disponível em: <<https://blog.somostera.com/data-science/arvores-de-decisao>>. Acesso em 22. set. 2022.

SANCHES, M. **Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados.** [s.l.: s.n.]. 2003. Disponível em: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-12102003-140536/publico/Dissertacao_MKS.pdf>. Acesso em: 10. out. 2022.

TAFNER, A, M. **ARTIFICIAL NEURAL NETWORKS** 1998. Disponível em: <https://cerebromente.org.br/n05/tecnologia/rna_i.htm>. Acesso em: 28 nov. 2022.

TIBCO. **O que é regressão logística?**. 2022 Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-logistic-regression>>. Acesso em: 22. set. 2022.