

VI SEMANA UNIVERSITÁRIA DA URCA XXIV SEMANA DE INICIAÇÃO CIENTÍFICA DA URCA

13 a 17 de Dezembro de 2021

Tema: “Centenário de Paulo Freire: contribuição da divulgação científica e tecnológica em defesa da vida, da cidadania e da educação”

ESTUDOS E TÉCNICAS COMPUTACIONAIS PARA AVALIAR A EFICIÊNCIA DO SOFTWARE PALEOIMAGE

Pedro Henrique Miranda Sales¹, Kátia Pires Nascimento do Sacramento²

Resumo: O software Paleoimage foi criado em busca de fazer a identificação biométrica de fósseis, para tal, ele processa uma determinada imagem de um fóssil e retira as características dessa imagem, desta forma o programa cria para ele uma identidade, isso porque ao extrair as características de uma imagem, o software é capaz de diferenciá-la de qualquer outra. A fim de implementar novos modelos e avaliar a eficiência do software, o presente trabalho tem como objetivo explorar novas técnicas computacionais. Para isso, foi feito levantamento bibliográfico sobre temas como inteligência artificial, machine learning, entre outros. Já que não se pôde ir a campo por consequência da pandemia do Covid-19, o meio usado para obter os resultados foi um conjunto de dados da ferramenta de aprendizagem de máquina, Weka, com as características de três tipos de plantas íris; estas características estão relacionadas à largura e ao comprimento da pétala e da sépala das flores de cada uma das plantas. Ao aplicar o conjunto de dados à ferramenta Weka, usando o algoritmo J48, foi possível obter a árvore de decisão e a matriz de confusão geradas pelo teste.

Palavras-chave: Inteligência artificial. Machine learning. Weka.

1. Introdução

A Inteligência Artificial (IA) busca simular a inteligência humana através de algoritmos e teve início após a segunda guerra mundial, especificamente em 1956. De acordo com Cossetti (2018) “Podemos definir a inteligência artificial, no grosso modo, como a capacidade das máquinas de pensarem como seres humanos: aprender, perceber e decidir quais caminhos seguir, de forma racional, diante de determinadas situações”. Existem diversas tecnologias por trás da IA, tais como data mining, machine learning, data science, deep learning, entre outras. Dentre essas diversas tecnologias, foi dado ênfase ao machine learning.

O subcampo da inteligência artificial, machine learning ou aprendizado de máquina, consiste na capacidade de o computador aprender de acordo com dados e através disso, realizar atividades como classificação, regressão, associação, agrupamento, entre outras. Cada uma dessas tarefas possui um objetivo específico, em que a classificação tenta prever o atributo classe, que é o atributo de interesse, onde este atributo é sempre o último da relação, ou seja, o último atributo do conjunto de dados; na classe se encontram os resultados históricos e através deles a máquina prever novos resultados, nesta tarefa o valor previsto é nominal. Assim como na classificação, a regressão tenta prever o atributo classe, porém, o valor previsto é numérico, ou seja, os valores da classe são numéricos. A associação é uma tarefa que não possui o atributo

1 Universidade Regional do Cariri, email: pedro.miranda@urca.br

2 Universidade Regional do Cariri, email: katia.pires@urca.br

VI SEMANA UNIVERSITÁRIA DA URCA XXIV SEMANA DE INICIAÇÃO CIENTÍFICA DA URCA

13 a 17 de Dezembro de 2021

Tema: "Centenário de Paulo Freire: contribuição da divulgação científica e tecnológica em defesa da vida, da cidadania e da educação"

classe, pois seu objetivo é apenas encontrar relações entre os atributos. O agrupamento ou clustering, busca encontrar padrões e verificar as relações entre os dados de um conjunto de dados, a fim de criar grupos entre os que tiverem mais semelhanças. A estes grupos é dado o nome de clusters ou agrupamentos.

Destas tarefas de aprendizagem, esta pesquisa trabalhou com a classificação, em busca de classificar as flores íris, que formam o atributo classe no conjunto de dados utilizado. Este banco de dados é o Iris e é disponibilizado no *Machine Learning Repository* pelo *Center for Machine Learning and Intelligent Systems*. A análise foi realizada por meio da ferramenta de aprendizagem Weka, especificamente pelo algoritmo de construção de árvore de decisão, J48.

2. Objetivo

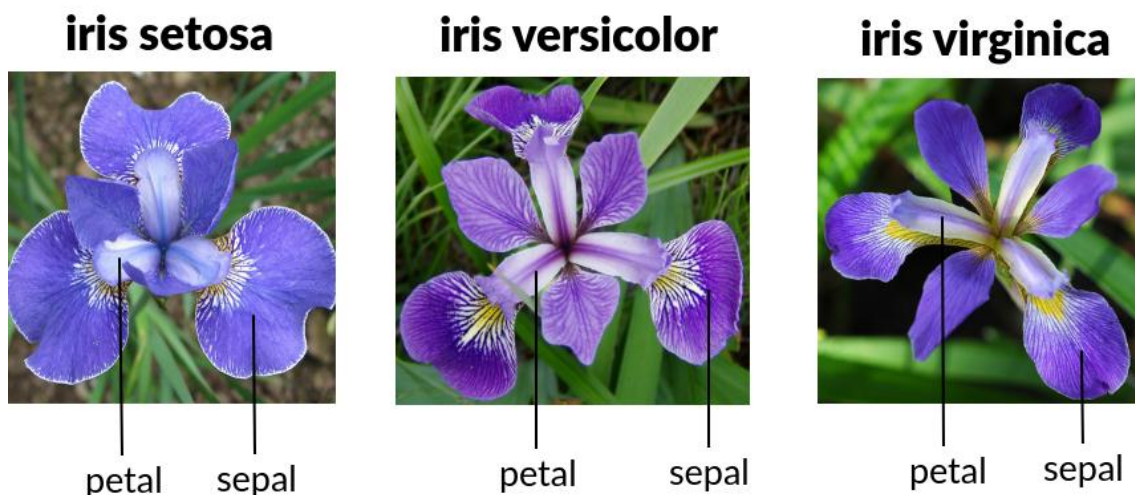
O objetivo geral deste estudo é avaliar novas técnicas computacionais, a fim de implementar novos modelos e medir a eficiência do software Paleoimage. Em relação aos objetivos específicos, têm-se: Usar a classificação através da ferramenta Weka para comparar as características das flores das plantas Iris Setosa, Iris Versicolor e Iris Virginica, a fim de identificar cada uma e classificá-las; por fim, analisar o comportamento do software Weka com o algoritmo J48.

3. Metodologia

O método utilizado foi de caráter exploratório, com uma abordagem teórica quantitativa, envolvendo revisão bibliográfica em artigos.

Esta pesquisa teve como entrada o conjunto de dados Iris, criado pelo estatístico e biólogo britânico Ronald Fisher, em 1936. O mesmo possui 150 instâncias, divididas em três classes de 50 instâncias cada, onde cada classe se refere a um tipo de planta de íris. A figura 1 mostra os três tipos de íris, indicando as partes da pétala e da sépala, as quais serão levadas em conta como as principais características de cada uma.

Figura 1 - Imagem dos três tipos de íris



VI SEMANA UNIVERSITÁRIA DA URCA XXIV SEMANA DE INICIAÇÃO CIENTÍFICA DA URCA

13 a 17 de Dezembro de 2021

Tema: "Centenário de Paulo Freire: contribuição da divulgação científica e tecnológica em defesa da vida, da cidadania e da educação"

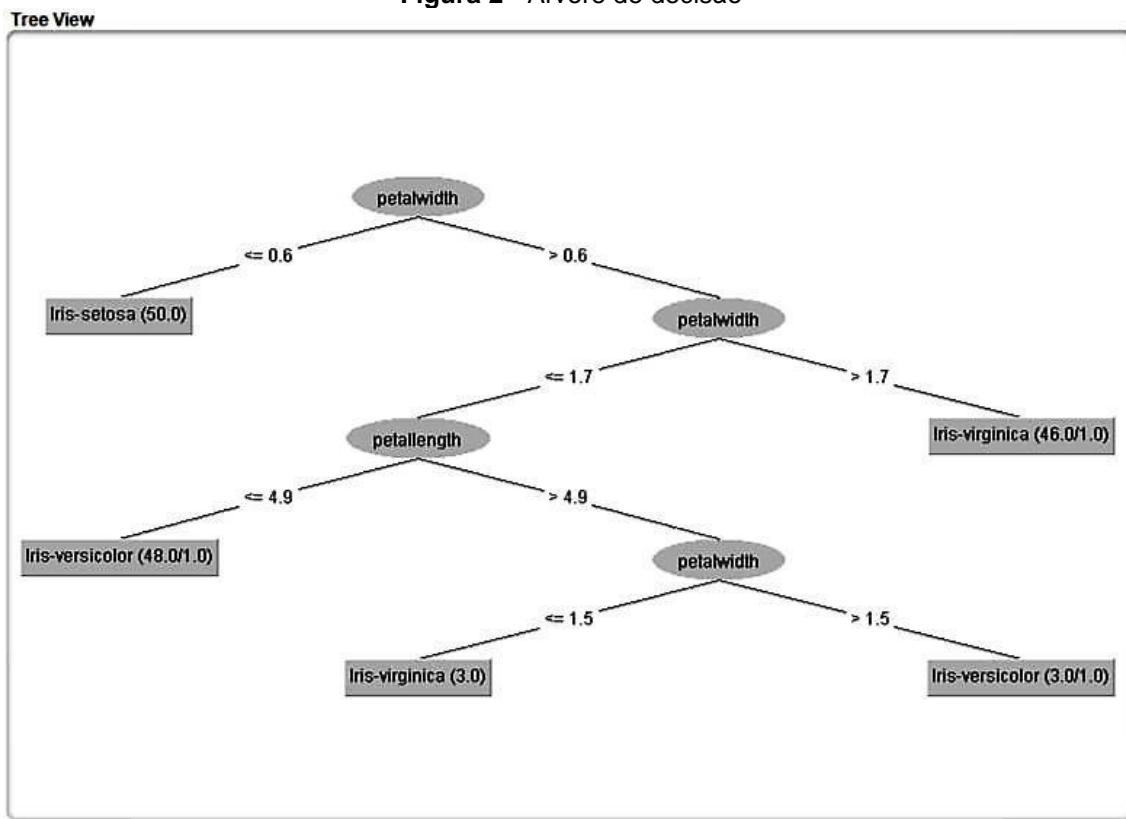
Fonte: Lucas Fonseca Müller, 2020

Além do atributo classe, formado pelos tipos de íris, tem-se no conjunto de dados, as medidas das principais características de cada flor, ou seja, da pétala e da sépala, formando os outros atributos.

4. Resultados

A partir do conjunto de dados foram realizados os testes usando o algoritmo J48 da ferramenta Weka. Com a aplicação desses dados ao algoritmo, obteve-se um índice de 96% de classificações corretas e 4% de classificações errôneas, desta forma, classificou corretamente 144 das 150 instâncias aplicadas ao teste. A figura 2 mostra a árvore de decisão gerada. Foi usado petalength para o comprimento da pétala e petalwidth para a largura da pétala.

Figura 2 - Árvore de decisão



Fonte: Próprios autores

Valores estatísticos resultantes do teste, disponibilizados pelo Weka:

- Erro médio absoluto: 0,035
- Erro quadrático: 0,1586
- Erro absoluto relativo: 7,8705%
- Raiz relativa ao erro quadrático: 33,6353%

VI SEMANA UNIVERSITÁRIA DA URCA XXIV SEMANA DE INICIAÇÃO CIENTÍFICA DA URCA

13 a 17 de Dezembro de 2021

Tema: "Centenário de Paulo Freire: contribuição da divulgação científica e tecnológica em defesa da vida, da cidadania e da educação"

- Número total de instâncias: 150

Em busca de permitir uma melhor visualização do desempenho do algoritmo, o Weka mostra, através de uma tabela, uma matriz de confusão. A tabela 1 apresenta a matriz de confusão produzida pelo teste.

Tabela 1. Matriz de confusão

A	B	C	Classificado como
49	1	0	A = Íris Setosa
0	47	3	B = Íris Versicolor
0	2	48	C = Íris Virginica

A matriz de confusão está relatando como o modelo J48 se saiu bem em relação aos erros e acertos. Usou A para Íris Setosa, B para Íris Versicolor e C para Íris Virginica.

Análise detalhada da matriz de confusão:

- Na primeira linha, o modelo classificou 49 instâncias corretamente, em que, das 50 instâncias que eram A ele classificou uma como B, errando apenas essa.
- Na segunda linha, das 50 instâncias que eram B, o modelo classificou 47 corretamente e errou 3 delas, classificando-as como C.
- Na terceira linha, o modelo classificou 48 instâncias corretamente, em que, das 50 instâncias que eram C ele errou apenas duas, classificando-as como B.

5. Conclusão

Esta pesquisa possibilitou conhecer novas técnicas computacionais capazes de avaliar a eficiência do software Paleoimage. A ferramenta de aprendizagem de máquina Weka, demonstrou através do algoritmo J48 ser bastante precisa, gerando uma árvore de decisão em que 96% das classificações foram corretas. Tendo em vista a precisão dessa ferramenta, fica claro que ela pode ser usada para avaliar a eficiência com que o software Paleoimage trabalha.

6. Referências

COSSETTI, Melissa Cruz. O que é inteligência artificial? **Tecnoblog**, 2018. Disponível em: <<https://tecnoblog.net/263808/o-que-e-inteligencia-artificial/>>. Acesso em: 14 de set. de 2021.

MÜLLER, Lucas Fonseca. Conceitos introdutórios de Machine Learning: análise do dataset Iris e a busca por um modelo ótimo de predição utilizando python. **LinkedIn**, 2020. Disponível em: <https://www.linkedin.com/pulse/parte-2-conceitos-introdu%C3%B3rios-de-machine-learning-e-fonseca-m%C3%BCI-ler?trk=public_profile_article_view>. Acesso em: 23 de set. de 2021.

**VI SEMANA UNIVERSITÁRIA DA URCA
XXIV SEMANA DE INICIAÇÃO CIENTÍ-
FICA DA URCA**

13 a 17 de Dezembro de 2021

Tema: "Centenário de Paulo Freire: contribuição da divulgação científica e tecnológica em defesa da vida, da cidadania e da educação"

UCI Machine Learning Repository. **Center for Machine Learning and Intelligent Systems**, 2017. Disponível em: <<http://archive.ics.uci.edu/ml/datasets/Iris>>. Acesso em: 22 de set. de 2021.