

UNIVERSIDADE ESTADUAL PAULISTA

“JÚLIO DE MESQUITA FILHO”

Instituto de Geociências e Ciências Exatas - IGCE

Curso de Bacharelado em Ciências da Computação

GABRIEL LUIZ

PROPOSTA TRABALHO FINAL

Professora: Dra. Adriane Beatriz de Souza Serapião

Rio Claro - SP

2020

1 Proposta Trabalho Final

Como proposta de trabalho final para a disciplina de aprendizado de máquina é proposto desenvolver um código utilizando da linguagem python e alguns frameworks como pandas, para realizar o processamento de imagens de flores com o intuito de classificá-las de acordo com seu tipo i.e. processar uma imagem de uma flor e dizer se é um dente de leão ou uma margarida, etc.

1.1 Dataset

Para realizar o trabalho será utilizado um dataset encontrado na base de dados Kaggle (<https://www.kaggle.com/>), onde o dataset se encontra em (<https://www.kaggle.com/alxmamaev/flower-recognition>) este dataset possui cinco classes de tipos de flores, que estão separados em pastas, que são:

- Margaridas
- Dente de leão
- Rosas
- Girassol
- Tulipa.

1.2 Algoritmos

Para este trabalho é proposto ser utilizado dois tipos de algoritmos para realizar o treinamento e classificação das imagens, sendo eles, Convolutional Neural Networks (CNN) e Random Forest e comparar os resultados obtidos para se dizer qual é o melhor algoritmo para a tarefa em questão.

1.3 Convolutional Neural Networks (CNN)

Para se realizar o processamento das imagens com o algoritmo de CNN será utilizado a biblioteca Keras, que utiliza como core o TensorFlow, ela possibilita criar e treinar modelos de aprendizado profundo, sendo fácil de usar e fácil de estender.

1.4 Random Forest

Para se realizar o processamento das imagens com o algoritmo de Random Forest será utilizado a biblioteca Skylearn que possui uma classe implementada com o algoritmo em questão.

1.5 Pré Processamento

Para o pré-processamento de ambos os algoritmos será utilizado uma biblioteca do python cv2, que permite ler as imagens, redimensionar e descrever as imagens de uma maneira que possa ser utilizada em ambos algoritmos. E também para processar os arquivos nas pastas as bibliotecas glob e os

1.6 Validação

Para validar qual dos dois modelos é o melhor para o problema apresentado serão utilizadas as métricas de precisão, recall, f1-score e suporte. Onde estas metricas serão executadas para ambos os algoritmos e comparado os resultados.