FRUIT PUNCH

CPSC 481 - Artificial Intelligence

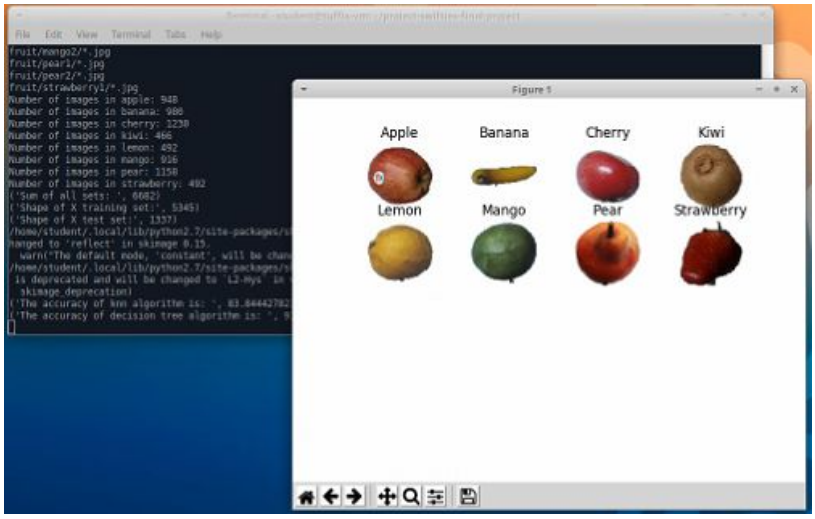
Christian Jimenez (SCRUM Master)
Ryan Talactac (Technical Manager)
Matthew Zuniga (Project Manager)
Pooja Gajjar (Product Manager)

Design Document Outline

I. Project Overview
II. Objectives / Outcomes
III. Project Design Pipeline
IV. Requirements
V. Use Cases
VI. Milestones

I. Project Overview

Fruit Punch is a machine learning program that trains on a dataset of fruit images. The program identifies a fruit based on an image using machine learning algorithms. Data visualization based on the accuracy of the machine learning model.



II. Objectives & Outcome

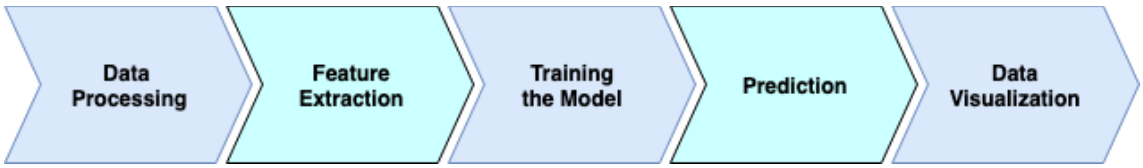
A. Objectives

1. To apply knowledge learned from material in CPSC 481
2. Identify an AI algorithm to solve a complex problem
3. Design and Implement data structures that will represent important features of the environment
4. Implement an AI algorithm that utilizes a concept discussed in class

B. Expected Outcomes

1. To learn more about a supervised machine learning algorithm
2. Gain experience with Python and its AI libraries
3. For the supervised algorithm to accurately predict a Fruit by images

III. Project Design Pipeline



A. Data Processing

1. We used data from Kaggle*:* https://www.kaggle.com/moltean/fruits#fruits-360_dataset.zip
2. We created a function which reads images as the input from the specific folders and appends them to a list.

B. Feature Extraction

1. In order to determine what kind of fruit it is, we need to retrieve the features of the image. In this case, the approach we took was to label each fruit in order to categorize it into its respective category.

C. Training Model

1. We load the test data using sklearn's **`test_train_split`**. Then we must split the data into two sets: **Test** and **Train Set** using **`test_train_split`**. We split the data in 80% training and 20% testing. Lastly, we randomize the data in order to ensure that the model is non-deterministic.

D. Prediction

1. We used "sklearn" knn algorithm and decision tree algorithms.

E. Data Visualization

1. We will visualize the data with matplotlib and test against the accuracy of the model in identifying these images.

IV. Requirements

A. <u>Performance Requirements:</u> The user must install the following requirements in order for the program to function.

1. Python 3.7+ (If you're on Windows, ensure that you add Python to your PATH variables when installing.)
2. matplotlib **`pip install --user matplotlib`**
3. sk-image **`pip install --user scikit-image`**
4. scipy **`pip install --user scipy`**

5. sklearn `pip install --user scikit-learn`
6. opencv `pip install --user opencv-python`

V. Use Cases

A. Program scans the image and categorizes it into the appropriate category.

VI. Milestones

A. **Milestone 1:**
   1. Establish roles & responsibilities for each member
   2. Create a method to load data set from images
   3. Add pandas integration to show data visualization
   4. Create method to show image from train/test set
   5. Add imports.py to main.py
   6. Create Google Colab if possible to run script line-by-line for class demo

B. Milestone 2:
   1. Refactor code
   2. Add KNN Algorithms

C. Milestone 3:
   1. Create Software Design Document
   2. Ensure that program is fully functional

VII. Supplementary Documentation

A. Tools Used
   1. Google Colab, Github

B. Technical Implementations
   1. **Languages:** *Python 3.7+*
   2. **Libraries:** *pandas, numpy, matplotlib, sklearn, opencv2*

C. References
   1. The following dataset was used to test our application: https://www.kaggle.com/moltean/fruits#fruits-360_dataset.zip