CLASSIFYING FRUITS (APPLE VS. ORANGE) BASED ON SIZE AND COLOR

A PROJECT REPORT

Submitted by

SAKTHI VISHAL C (2303811724321095)

in partial fulfillment of requirements for the award of the course

AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **"CLASSIFYING FRUITS (APPLE VS. ORANGE) BASED ON SIZE AND COLOR"** is the bonafide work of **SAKTHI VISHAL C (2303811724321095)** who carried out the project work during the academic year 2024 - 2025 under my supervision.

SIGNATURE

Dr.T. AVUDAIAPPAN, M.E.,Ph.D.,

HEAD OF THE DEPARTMENT

PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

SIGNATURE

Ms.S.Murugavalli., M.E.,(Ph.D).,

SUPERVISOR

ASSISTANT PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

Submitted for the viva-voce examination held on **02.06.2025**.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION

I declare that the project report on **"CLASSIFYING FRUITS (APPLE VS. ORANGE) BASED ON SIZE AND COLOR"** is the result of original work done by me and best of my knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGI1252 - FUNDAMENTALS OF DATA SCIENCE USING R.**

Signature

SAKTHI VISHAL C

Place: Samayapuram

Date:02.06.2025

ACKNOWLEDGEMENT

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing us with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E.,Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I express our deep expression and sincere gratitude to our project supervisor **Ms.S.Murugavalli., M.E.,(Ph.D).,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated us to carry out this project.

I render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

INSTITUTE

Vision:

- To serve the society by offering top-notch technical education on par with global standards.

Mission:

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

DEPARTMENT

Vision:

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

Mission

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

PROGRAM EDUCATIONAL OBJECTIVES (PEO)

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

PROGRAM SPECIFIC OUTCOMES (PSOs)

- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

PROGRAM OUTCOMES (POs)

Engineering students will be able to:

1. **Engineering knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals, and an engineering specialization to develop solutions to complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development.
3. **Design/development of solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required.
4. **Conduct investigations of complex problems:** Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions.
5. **Engineering Tool Usage:** Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems.
6. **The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment.

7. **Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws.

8. **Individual and Collaborative Team work:** Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

9. **Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences.

10. **Project management and finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

11. **Life-long learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change.

ABSTRACT

This project presents an interactive Shiny application in R designed to classify fruits, specifically Apples and Oranges, based on their size and color. The classification is achieved using a machine learning model, Random Forest, trained on a carefully structured dataset. The app enables users to input fruit parameters manually or upload a dataset and receive instant predictions on the fruit type. By automating the fruit classification process, this project demonstrates an efficient and scalable approach that can be extended to industrial applications in agriculture, food processing, and retail. The system combines machine learning with a user-friendly web interface, illustrating the practical application of data science in real-world scenarios. It showcases how fundamental features like size and color can be harnessed effectively to make accurate predictions.

ABSTRACT WITH POs AND PSOs MAPPING

CO 5 : BUILD DATA SCIENCE USING R PROGRAMMING FOR SOLVING REAL-TIME PROBLEMS.

ABSTRACT	POs MAPPED	PSOs MAPPED
This project presents an interactive Shiny application in R designed to classify fruits, specifically Apples and Oranges, based on their size and color. The classification is achieved using a machine learning model, Random Forest, trained on a carefully structured dataset. The app enables users to input fruit parameters manually or upload a dataset and receive instant predictions on the fruit type. By automating the fruit classification process, this project demonstrates an efficient and scalable approach that can be extended to industrial applications in agriculture, food processing, and retail. The system combines machine learning with a user-friendly web interface, illustrating the practical application of data science in real-world scenarios. It showcases how fundamental features like size and color can be harnessed effectively to make accurate predictions.	**PO1 -3** **PO2 -3** **PO3 -3** **PO4 -3** **PO5 -3** **PO6 -2**	**PSO1 -3** **PSO2 -3**

Note: 1- Low, 2-Medium, 3- High

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

11.1 Objective

The primary objective of this project is to design and develop an intelligent fruit classification system using the R programming language and the Shiny web application framework. The system aims to accurately distinguish between apples and oranges based on two key features: the size (in centimeters) and the color (red or orange) of the fruit.

Key goals of this project include:

- **Automating fruit classification**: To reduce human error and time-consuming manual sorting processes in agricultural and retail environments.
- **Enhancing operational efficiency**: By leveraging machine learning models that can make accurate predictions in real time.
- **User-friendly interface**: Providing a simple, accessible Shiny application that allows users to input data manually or upload a CSV file for batch predictions.
- **Educational value**: Demonstrating the practical application of data science concepts such as feature extraction, model training, prediction, and evaluation.
- **Scalability**: Laying the groundwork for future extensions that can handle more fruit types and integrate with advanced input sources like image data or sensor readings.

Ultimately, the objective is to showcase how even a basic machine learning model, when combined with a structured dataset and an intuitive interface, can solve real-world classification problems effectively and efficiently.

11.2 Overview

Manual classification of fruits can be inconsistent and time-consuming. The ever-growing need for automation in agricultural and food supply chains necessitates smart, efficient, and reliable systems. Leveraging machine learning with visual interfaces, this project addresses these issues by providing a system that not only classifies fruits accurately but also operates interactively for user convenience. The Random Forest model is well-suited for this task due to its robustness and ability to handle categorical variables like color. R Shiny provides the ideal platform for developing an interactive dashboard that is both functional and educational.

11.3 R Programming Concepts Used

- **Data Preprocessing**: Cleaning and formatting raw data to ensure consistency.
- **Feature Engineering**: Utilizing size (numerical) and color (categorical) as predictive features.
- **Machine Learning**: Implementation of Random Forest algorithm for binary classification.
- **Model Evaluation**: Testing model accuracy using confusion matrices and performance metrics.
- **Shiny Framework**: Building a reactive web interface for seamless interaction.
- **UI Design**: Structuring input panels, outputs, and file handling mechanisms in Shiny.

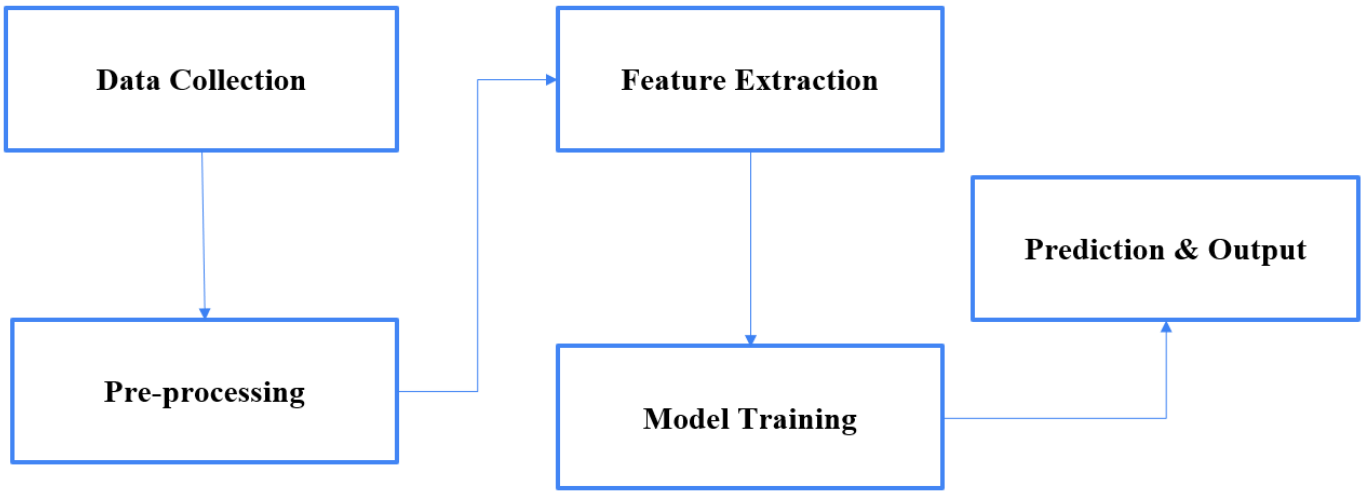
CHAPTER 2
PROJECT METHODOLOGY

2.1 Proposed Work

The proposed system follows a structured pipeline of steps that include:

1. **Data Collection**: Creation of a labeled dataset including features like size and color for both apples and oranges.
2. **Data Preprocessing**: Encoding categorical variables, normalizing numerical data, and removing inconsistencies.
3. **Model Training**: Using a Random Forest classifier trained on the dataset to identify decision patterns.
4. **Application Development**: Creating a dynamic web application using R's Shiny package.
5. **User Interaction**: Allowing users to manually input data or upload a CSV file for batch prediction.
6. **Model Prediction**: Performing classification using the trained model.
7. **Evaluation and Feedback**: Displaying prediction results and model accuracy.

2.2 Block Diagram



CHAPTER 3
MODULE DESCRIPTION

3.1 Data Acquisition Module

This module accepts user-provided inputs for fruit size and color or reads uploaded CSV files containing fruit attributes. It ensures that the input format is consistent with model requirements and prepares the data for analysis.

3.2 Preprocessing Module

This module transforms raw input into clean, structured data. It converts color labels into factors, checks data types, and ensures numerical values fall within expected ranges. For uploaded datasets, missing values are imputed or discarded based on severity.

3.3 Feature Extraction Module

This module isolates the input features relevant to the model. For this application, two features are used:

- **Size**: A continuous numerical variable representing the diameter of the fruit in centimeters.
- **Color**: A categorical variable representing the visible color (Red or Orange).

3.4 Classification Module

This module applies the pre-trained Random Forest model to classify the fruit based on input features. The Random Forest classifier aggregates the decisions from multiple decision trees and outputs the most likely fruit class (Apple or Orange).

3.5 Output Module

This module displays the prediction res-ult on the UI and calculates the model's classification accuracy. If a CSV is uploaded, it processes the entire dataset and outputs predictions for all rows.

CHAPTER 4
CONCLUSION & FUTURE SCOPE

Conclusion

This project successfully demonstrates how simple features like size and color can be used with machine learning to classify fruits. By developing a Shiny web application, the system is made interactive and easy to use, even for non-technical users. The use of Random Forest ensures high accuracy and robustness. The application highlights the integration of machine learning and web technologies using R and provides a scalable base for more complex classification systems.

Future Scope

In the future, this project can be extended in several ways:

- **Expansion of Fruit Categories**: Include more fruit types and varieties.
- **Incorporate Additional Features**: Such as weight, texture, shape, and ripeness.
- **Image Processing**: Use computer vision and CNNs to classify fruits based on photos.
- **Cloud Deployment**: Host the Shiny app on servers like shinyapps.io for global accessibility.
- **Real-Time Classification**: Integrate with camera sensors or industrial scanners.
- **Mobile Application**: Create an Android/iOS version of the app for field use.

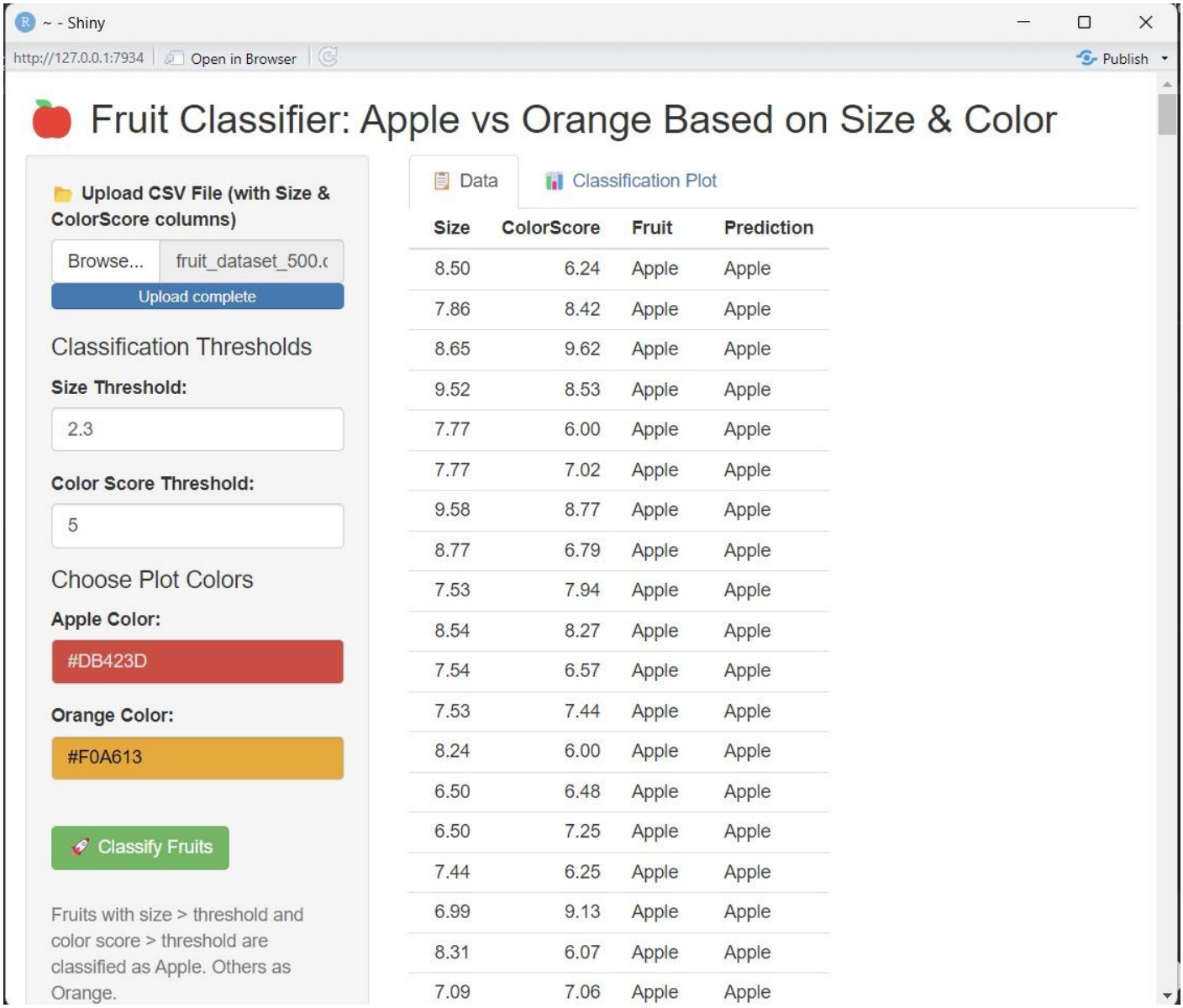
CHAPTER 5
Appendix A – Source Code

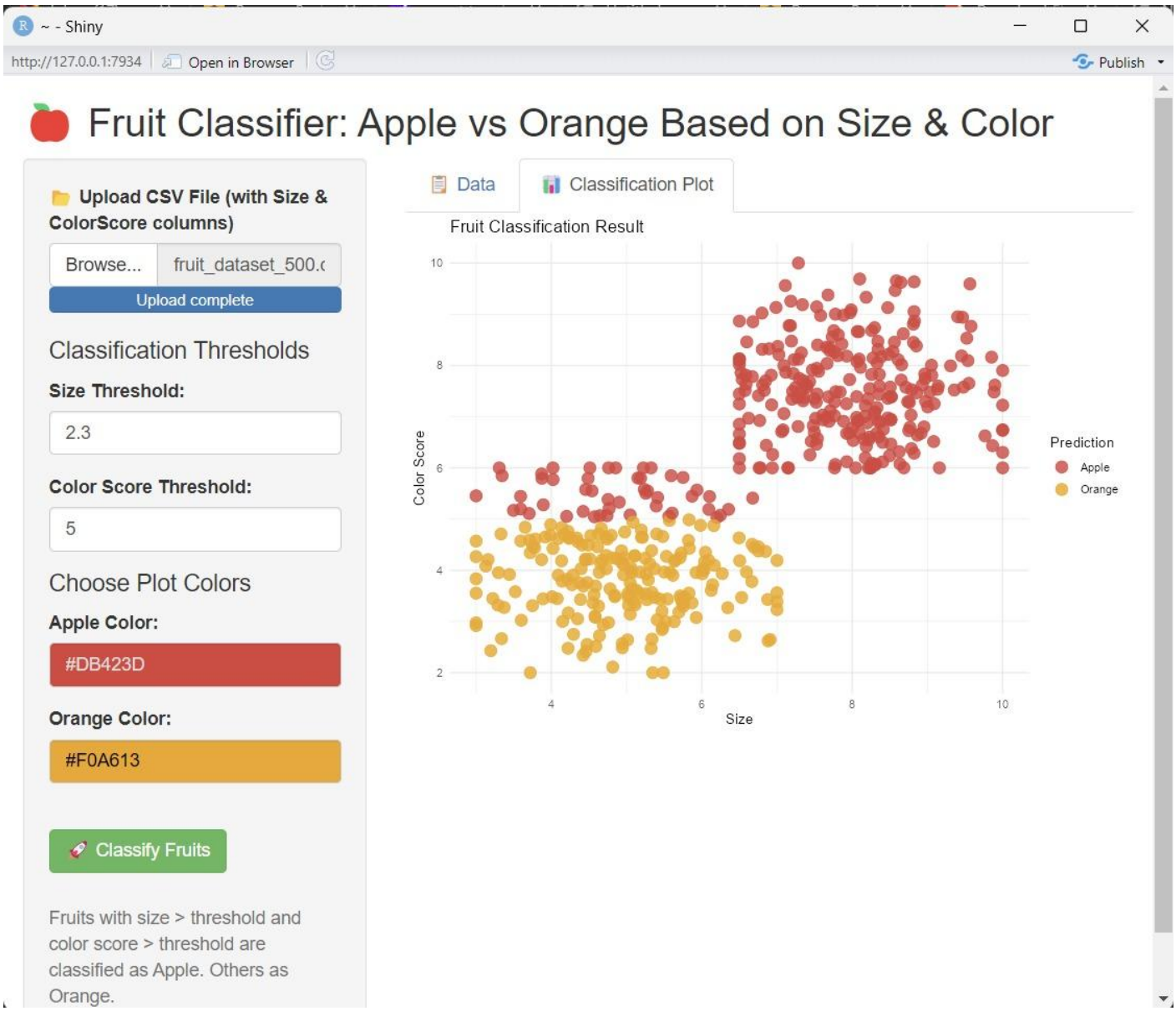
```
# Required Libraries
library(shiny)
library(ggplot2)
library(dplyr)
library(colourpicker)  # For colourInput widget

# Define UI
ui <- fluidPage(
  titlePanel("🍎 Fruit Classifier: Apple vs Orange Based on Size & Color"),

  sidebarLayout(
    sidebarPanel(
      fileInput("file", "📂 Upload CSV File (with Size & ColorScore columns)", accept = ".csv"),
      h4("Classification Thresholds"),
      numericInput("sizeThreshold", "Size Threshold:", value = 7.0, min = 1, max = 15, step = 0.1),
      numericInput("colorThreshold", "Color Score Threshold:", value = 5.0, min = 1, max = 10, step = 0.1),
      h4("Choose Plot Colors"),
      colourInput("appleColor", "Apple Color:", value = "red"),
      colourInput("orangeColor", "Orange Color:", value = "orange"),
      br(),
      actionButton("run", "🚀 Classify Fruits", class = "btn btn-success"),
      br(), br(),
      helpText("Fruits with size > threshold and color score > threshold are classified as Apple. Others as Orange.")
    ),

    mainPanel(
      tabsetPanel(
```

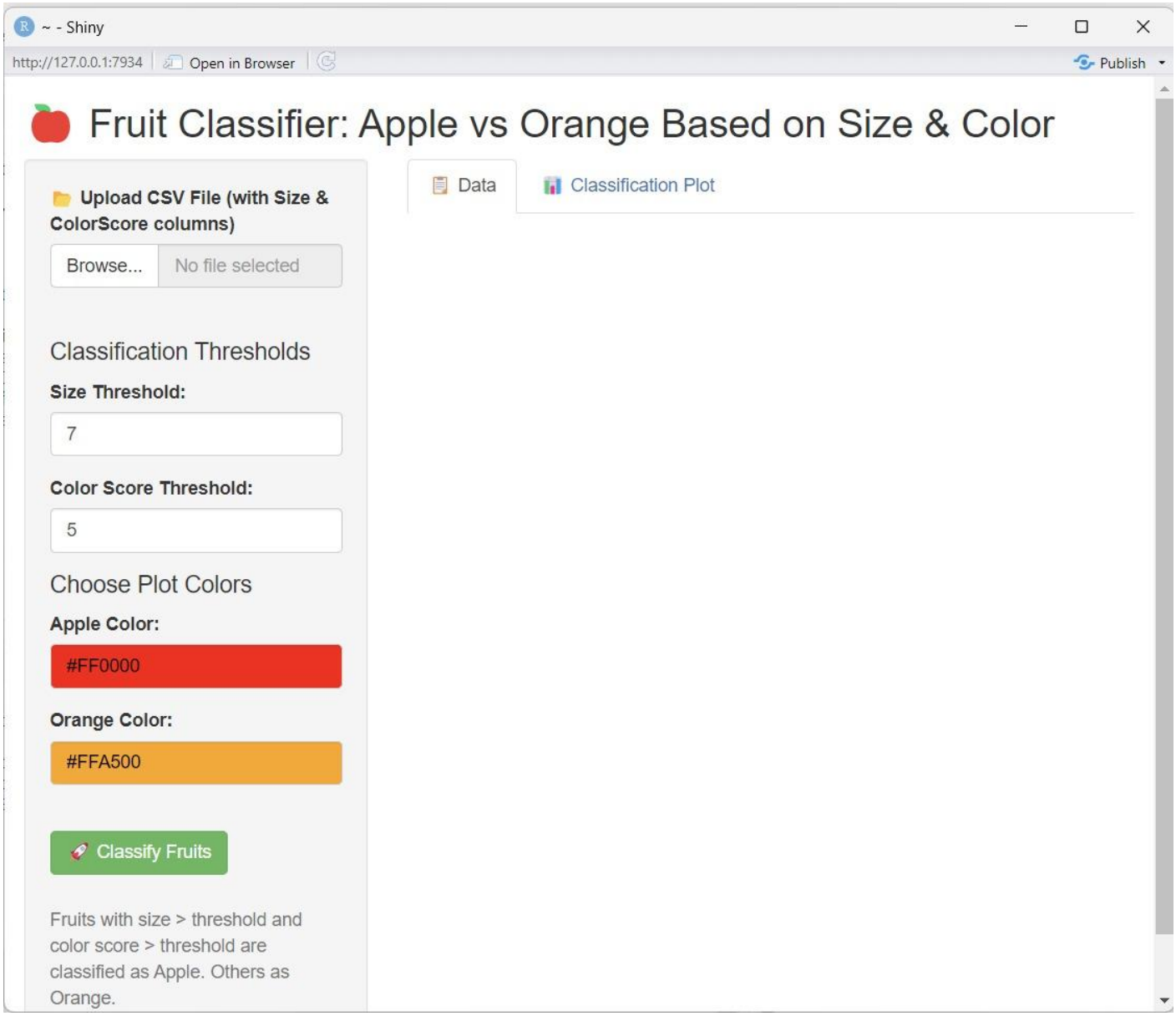
```
        tabPanel("📋 Data", tableOutput("dataTable")),
        tabPanel("📊 Classification Plot", plotOutput("fruitPlot"))
      )
    )
  )
)

# Define Server
server <- function(input, output) {

  userData <- eventReactive(input$run, {
    req(input$file)
    df <- read.csv(input$file$datapath)
    req("Size" %in% names(df), "ColorScore" %in% names(df))

    # Classification logic
    df$Prediction <- ifelse(df$Size > input$sizeThreshold & df$ColorScore > 
input$colorThreshold,
                            "Apple", "Orange")
    df
  })

  output$dataTable <- renderTable({
    userData()
  })

  output$fruitPlot <- renderPlot({
    df <- userData()
    ggplot(df, aes(x = Size, y = ColorScore, color = Prediction)) +
      geom_point(size = 4, alpha = 0.8) +
      scale_color_manual(values = c("Apple" = input$appleColor, "Orange" = 
input$orangeColor)) +
      labs(title = "Fruit Classification Result",
           x = "Size", y = "Color Score") +
      theme_minimal()
```

```
  })
}

# Run the App
shinyApp(ui = ui, server = server)
```

Appendix B – Screenshots




~ - Shiny
http://127.0.0.1:7934
Open in Browser
Publish
Fruit Classifier: Apple vs Orange Based on Size & Color
Upload CSV File (with Size & ColorScore columns)
Browse...
fruit_dataset_500.c
Upload complete
Classification Thresholds
Size Threshold:
2.3
Color Score Threshold:
5
Choose Plot Colors
Apple Color:
#DB423D
Orange Color:
#F0A613
Classify Fruits
Fruits with size > threshold and color score > threshold are classified as Apple. Others as Orange.
Data
Classification Plot
Fruit Classification Result
Color Score
Size
Prediction
Apple
Orange


~ - Shiny
http://127.0.0.1:7934
Open in Browser
Publish
Fruit Classifier: Apple vs Orange Based on Size & Color
Upload CSV File (with Size & ColorScore columns)
Browse...
No file selected
Classification Thresholds
Size Threshold:
7
Color Score Threshold:
5
Choose Plot Colors
Apple Color:
#FF0000
Orange Color:
#FFA500
Classify Fruits
Fruits with size > threshold and color score > threshold are classified as Apple. Others as Orange.
Data
Classification Plot

REFERENCES

1. Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5-32.
2. Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2021). *Shiny: Web Application Framework for R*. RStudio.
3. James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning*.
4. Kaggle Datasets: https://www.kaggle.com
5. R Documentation: https://www.rdocumentation.org/