CSED 226 Introduction to Data Analysis

2023 Final Exam

Kaggle Competition: Binary Classification

1. Kaggle URL

- URL: https://www.kaggle.com/t/b69bf0d502c740ada9c5989bf061aa0b

2. Overview

- The goal is to improve your **binary classifier**'s performance as much as possible.
- You can improve your model's performance through data preprocessing, and hyperparameter tuning, etc.
- You can submit your model's results to Kaggle and view your scores in real-time on the leaderboards.

3. Rule

- **Note that you cannot use K-Nearest-Neighbors (KNN) based models.**

4. Evaluation

- Submissions are evaluated on the **Accuracy** between the ground truth labels of the data and your predicted labels.

5. Dataset: Dried fruit

- You are given data with 7 features to classify species of fruits.

Column	Meaning of features	Type
A	Surface Area	Numerical
B	Longer Axis Length	Numerical
C	Shorter Axis Length	Numerical
D	Sharpness*	Numerical / Value range: 0~1
E	Surrounding Area*	Numerical
F	Shape Factor*	Numerical / Value range: 0~1
G	Circumference	Numerical

Table 1: Features

→ Sharpness*: When it is close to 1, the shape would be a line, and when close to 0, the shape would be a circle.

→ Surrounding Area*: The area of the smallest convex bounding surface surrounding the fruit.

→ Shape Factor*: It provides the proportion of the area created by the fruit relative to the total pixels within the bounding box.