A picture containing calendar

Description automatically generated

Raisins Quality Detection based on image data

Final Project Report

**PROGRAMMING IN PYTHON [B]**

Submitted by:

|  |  |
| --- | --- |
| **NAME** | **ID** |
| Rahman, Md Ashikur | 18-38519-2 |
| Hasan, Jahidul | 18-38489-2 |

Submitted to:

**DR. AKINUL ISLAM JONY**

**Associate Professor**

**Faculty of Science and Technology**

**American International University-Bangladesh (AIUB)**

# [Dataset : https://www.muratkoklu.com/datasets/Raisin\_Dataset.zip](https://www.muratkoklu.com/datasets/Raisin_Dataset.zip) PROJECT OBJECTIVES

This project's primary goal is to provide students a solid understanding of machine learning and the process of creating machine learning models. One of this project's goals may be described by its principal focus, which is supervised machine learning. Additionally, many Python libraries are utilized here, which is likewise done in order to have a firm understanding of them. This project's primary focus is picture detection, which is a classification issue. The knowledge of these kinds of difficulties will grow as a result of the project's completion. The primary goal of this study is a greater understanding of machine learning and various algorithms.

# DATA PREPROCESSING

Preprocessing data is one of the most important tasks involved in building a machine learning model. The reason being that utilizing outdated data makes it incredibly difficult to create an appropriate machine learning model. There are several ways to clean up data, including looking for null values, figuring out which columns in a data collection contain the most empty values, and replacing null data with the mean, median, and mode value. With no null values and every row filled with genuine data, the dataset used in our research was completely clean.

# DATA SPLIT

Data splitting is essential since without it, the model cannot be produced. The data must be divided into two parts: a training dataset and a testing dataset. In total, 900 different data instances were used in this project. The number 630 is used to train the model. The remaining 270 data points are used to confirm the accuracy of the model. Since training and testing are necessary for the model to be precise, spitting mostly depends on the user.

# MODEL CREATION

The machine learning model has to be created after the first data preparation, which included the formation of a feature matrix and the separation of the target variable. We employed a total of two different categorization models. the DT and the SVM. The model was tested using them. The SVM provided the best accuracy, 83.7 percent, after the total testing. The latter of which is neither the best nor the worse.

# DISCUSSION AND CONCLUSION

This model, which forecasts the quality of raisins, was developed using a dataset of raisin details. This model's accuracy is 83.7 percent. It isn't the worst but isn't the finest either. The best decisions, such as choosing the most significant column to test the data, were made in order to attain this precision. Although the accuracy is below the recommended level of 95%, it's still feasible that the dataset is the problem. To improve forecasting, it may, for instance, include more columns with a greater connection with other variables. To correctly train the model, it could additionally contain additional rows of data. This paradigm has expanded our understanding of machine learning, particularly supervised machine learning.