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

The main purpose of doing this project is to get a proper idea about the machine learning and the model creation in machine learning. This project mainly focuses on the supervised machine learning so one of its purposes can be defined by it. Moreover, different types of Python libraries are used here which also serve a purpose to get a good grip of them. This project mainly focuses on image detection which can be defined as a classification problem. With completing this project, it will increase the knowledge about these type problems. This project’s main outcome is to increase the knowledge about the machine learning and different algorithms.

# DATA PREPROCESSING

One of the most critical jobs in the creation of a machine learning model is data preprocessing. Because it's nearly impossible to build an accurate machine learning model using stale data. There are numerous data cleaning strategies, such as checking for null values, determining which columns of a data set have the emptiest values, and replacing null data with the mean, median, and mode value. The dataset in our project was fully clean, with all of the rows filled with actual data and no null values.

# DATA SPLIT

Data splitting is critical since the model cannot be generated without it. The data must be split into two sections: one for training the dataset and the other for testing it. There were 900 data instances in total in this project. The model is trained with the number 630. The remaining 270 data points are utilized to validate the model's correctness. Spitting is largely dependent on the user, as both training and testing are required for the model to be accurate.

# MODEL CREATION

After the initial data preprocessing also with the creation of feature matrix and separating the target variable it is time for creating the machine learning model. Total 2 types of classification models were used. The SVM and DT. Those were used to test the model. After the overall testing, the SVM gave the highest accuracy which is 83.7%. Which is not the best but not the worst as well.

# DISCUSSION AND CONCLUSION

This model is created based on a raisin details dataset which predicts the quality of raisins. The accuracy of this model is 83.7 percent. Which isn't the best, but it's also not the worst. This precision was achieved by making all of the best options possible, such as selecting the most important column to test the data. Still, because the accuracy isn't around 95 percent, which is considered ideal, it's possible that the dataset is the issue. It may, for example, contain additional columns with a higher correlation with other factors in order to forecast better. It could also have extra rows of data to properly train the model. This model has increased the knowledge about machine learning specifically the supervised machine learning.