Advancing Medjool Date Research using the Date Sorting Machine

by

Matthew Adam
Ryan Antes
Connor Gaudette
Eli Hoffberg
Advancing Medjool Date Research using the Date Sorting Machine

An Interactive Qualifying Project submitted to the Faculty of WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the degree of Bachelor of Science

by
Mathew Adam
Ryan Antes
Connor Gaudette
Eli Hoffberg

Date:
27/2/2023

Report Submitted to:

Booky Katz
Southern Arava R&D

Professors Isa Bar-On and Erin Solovey and Graduate Student Co-advisor Tess Meier
Worcester Polytechnic Institute

This report represents the work of one or more WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on the web without editorial or peer review.
Abstract

Southern Arava Research and Development studies agricultural practices on Medjool dates to support local farmers and protect the industry from water shortages and plant diseases. The current data collection process consumes much of the researcher's time that could be used elsewhere. This paper discusses the retrofit of a machine designed to automate the data collection process. The post-retrofit machine was able to classify dates based on weight and maturity. It can also export data on individual dates to an Excel sheet. The workings of the machine that were learned or added during this project were documented in a manual to allow researchers to use the machine for years to come.
Acknowledgements

The Date Sorting Team would like to thank all the people we worked with at Southern Arava R&D. Dr. Michal Adler-Agmon, Yaara Danino, and our sponsor, Booky Katz, who all proved integral to the success of our project by understanding our work and donating their time to push our project to the finish line. Yaara and Michal furthered our understanding of date fruit characteristics. They guided us in understanding what they needed to improve Medjool date research, which was critical in the development of our algorithms. Booky Katz helped with all aspects of the project, from driving us to the bus station to getting us the equipment we needed to be successful.

We would also like to thank Tamir and Yuval. Tamir taught us how to use the date sorting machine and was a great resource when we ran into issues with the machine. Yuval answered any questions we had relating to dates or the Southern Arava Region, which helped us understand the motivations behind our project.

Southern Arava R&D felt like a community; everyone we met was kind and willing to help us. We wish to show great appreciation to the people of Southern Arava R&D.

We would like to thank our advisors, Professor Isa Bar-On, Professor Erin Solovey, and Tess Meier, from Worcester Polytechnic Institute, for their time teaching and mentoring us throughout the project. From bringing us to an agricultural exhibit to laughing in our weekly meetings, their belief in us was felt by all. For that, we cannot thank them enough.

We would like to thank the Arava Institute for Environmental Studies for hosting us and providing programming to help us build connections in the region.

Finally, we would like to thank Worcester Polytechnic Institute for allowing us to complete such an amazing project in a place as historically significant as Israel.
Authorship Page

Matthew Adam, Ryan Antes, Connor Guadette, and Eli Hoffberg contributed to the research and writing of the following IQP report. See below for a further breakdown of how the IQP report was written.

Matthew Adam was responsible for writing the Introduction section, the Integration and Instructional Media section, and Appendix B. He was also the primary editor of the Additions to the DSM section. Mr. Adam also wrote the section of the manual on integrating the PLC with the desktop computer.

Ryan Antes was responsible for writing the first draft of the background section containing information about Southern Arava R&D. He also wrote the first draft of the Additions to the DSM and Discussion sections, and over half of the DSM Manual which is included in Appendix A. Mr. Antes formatted the entire paper and made sure to include the required portions of the WPI IQP report.

Connor Gaudette was responsible for drafting the Abstract and Conclusion. He also added to the Background section containing the Fruit Sorting with Computer Vision section. Mr. Gaudette also wrote parts of the code involved in Appendix B.

Eli Hoffberg was responsible for writing the Date Fruit Growth portion of the Background section. He also wrote the Understand section. Mr. Hoffberg wrote the section of the manual containing the SQL instructions and parts of the code in Appendix B. He edited the MOP section of the Background, the Integration and Instructional Media section, and the Discussion section.

Matt Adam, Ryan Antes, Connor Gaudette, and Eli Hoffberg worked together to edit the IQP report for grammatical and spelling errors, correct format, and paper flow.
Table of Contents

Abstract ii
Acknowledgements iv
Authorship Page v
Table of Figures vii
Chapter 1: Introduction 8
Chapter 2: Background 10
MOP - Southern Arava Research and Development 10
Date Fruit Growth 11
Harvest Cycle 11
Date Development 11
Relevance of Date Research 13
Fruit Sorting with Computer Vision 13
Machine Learning and Artificial Intelligence Training 14
K-Nearest Neighbor 14
Structured Query Language 14
Chapter 3: Understand 15
Formative Assessment 15
DSM Before Modification 15
Chapter 4: Additions to the DSM 18
Collecting and Storing information from the DSM 18
Image Isolation 19
Automated Method to Distinguish Maturity 21
K-Nearest-Neighbor 21
Training Data 22
Chapter 5: Integration and Instructional Media 23
Integration of the Modified DSM into the Research Process 23
Instructional Media for Future Use of the DSM 23
Chapter 6: Discussion 25
Benefits of the DSM 25
Recommendations 25
Limitations 27
Chapter 7: Conclusion 28
REFERENCES 29
APPENDIX A: Instructional Media 31
Item 1: Information Flow Diagram 31
Item 2: User Manual 32
Appendix B: Project Code 62
Item 1: Code Summary 62
Item 2: Github Repository 62
# Table of Figures

Figure 1: Medjool Date Maturity ......................................................................................... 12  
Figure 2: Medjool Date Maturity Stages ............................................................................. 12  
Figure 3: The DSM with Labeled Components .................................................................. 16  
Figure 4: DSM Flow Diagram ............................................................................................. 17  
Figure 5: Example of Excel Sheet Produced by the Updated DSM ................................. 19  
Figure 6: Before and After Image Subtraction .................................................................... 20  
Figure 7: Methods used for Isolation .................................................................................. 21
Chapter 1: Introduction

The Southern Arava region is one of the least populated areas of Israel, yet it is home to one of the most flourishing agricultural industries in the country. Medjool dates grown in Israel consist of 50% of the world market for Medjool dates (Shahbandeh, M. 2022, November 21). Since the introduction of Medjool dates to the region in 1985, the industry has flourished because of the work of the people living in the Southern Arava region (E. Solowey, personal communication, February 2023). Researching date fruit growth and farming maintains and protects this industry and the livelihood of the people in the Southern Arava region.

Date agriculture requires many limited resources, which raises concerns with people who rely on the industry. Some of the concerns of this industry are plant disease and water efficiency. Each date tree was bred to be genetically identical, making the industry more susceptible to plant diseases. Scarcity in the desert also leads to increased water and fertilizer costs. Researchers throughout the Southern Arava region work to address these concerns.

Southern Arava Research and Development (MOP) is conducting this research to focus on improving agricultural practices using an automated Date Sorting Machine (DSM). However, the DSM fell into disuse because it did not have all the functions required by the researchers. The DSM only weighed dates in groups and did not classify them based on quality and maturity. Studies about how growing techniques affect the quality and maturity of dates require a lot of data about individual dates to draw statistically relevant results that were not captured by the DSM. Researchers at MOP turned to manually weighing and sorting dates to get the information needed for their studies because of the inadequacies of the DSM. To help the researchers collect their data, the DSM must now store and collect information on the weight of individual dates. It would also aid the researchers if the DSM could recognize the stage of maturity of a date. Documentation on the workings of the DSM would allow for it to be more easily updated to match any changes needed by the researchers.
The updated DSM can help researchers learn how changes in date-growing methods improve and protect the date industry. Farmers can use those methods to improve and protect their livelihood. This continues to allow people to be secure in their future and live comfortably in the desert.

**Goal Statement:**
The goal of this project was to automate the collection of Medjool date maturity and weight to facilitate research on how growing and treatment techniques affect the economic viability of the date harvest.

To achieve our goal, we aimed to meet the following objectives:

- Understand the needs of date research at Southern Arava R&D
- Automatically record collected weight and maturity data on individual dates in an Excel sheet.
- Develop an automated method to distinguish maturity of dates.
- Create instructional media for future use of the DSM.
- Integrate the modified DSM into the research process.
Chapter 2: Background

MOP - Southern Arava Research and Development

The Southern Arava Region of Israel has recently become an agricultural powerhouse for the country. Developments made in drip irrigation, desalination, and pest control have allowed for a variety of plants to grow in the hyper-arid desert. Israel has invested in desert agriculture research, and by funding Southern Arava R&D (MOP), advancements have been made in Medjool date fruit treatments, using plants to control pests and water-saving practices.

Researchers at MOP provide comprehensive agricultural support to farmers by experimenting with different methods to improve nutrients in the soil, limit water usage, and maximize plant growth. The Southern Arava Research and Development farm, established in 1964 as the Arava Experimental Station, is in Kibbutz Yotvata, along the Jordanian border. At first, small experiments were conducted that aimed to find crops that were well suited for the arid desert land. The station was not funded well and remained small until 1985. In 1985, Southern Arava Research and Development was founded by “the Board of Governors of the Jewish Agency, in the framework of the Negev and Arava R&D. The R&D extends over the entire area of the Hevel Eilot Regional Council and is operated by Ardom Management and Holdings” (R&D, 2019). Local and national Israeli government entities contribute to the funding for MOP, while the Jewish National Fund is the largest donor to MOP.

MOP focuses on increasing the profit of farmers in the Southern Arava desert. Reducing freshwater consumption is a main objective of MOP because the arid climate of the Negev desert makes water expensive, especially for farmers. One-way researchers measure the effectiveness of an agricultural experiment is by the quality of the fruit produced. Fruit that is of higher quality will sell for more money, increasing the profits of the farmers. A big part of the research at MOP focuses on Medjool date growth and treatment aiming to save farmers money and improve fruit quality.

Date research at MOP focuses on learning how changing water and fertilizer quantities affect date growth and post-harvest quality. Researchers at MOP also study how post-harvest
treatments, such as rehydration and blister treatments can improve the quality of a harvest. As part of this research, information collected from the experiment’s dates includes their weight, stage of growth, and percentage of damaged skin, which is performed by researchers who individually look at the dates and categorize them based on color and texture.

**Date Fruit Growth**

**Harvest Cycle**

Dates undergo an intensive growth process; caring for Medjool trees is very labor driven. In the winter months the date palms must be pruned for thorns to protect workers harvesting the dates in spring and summer. During early summer the date bunches must be trimmed to prevent the trees from producing too many dates and not providing enough nutrients to each fruit. The first harvest comes late in the summer when the higher-date bunches are ready to be picked. Early fall marks the second harvest when the lower date bunches can be harvested. Some trees may even require a third harvest in the middle of fall. Pollinating date trees is a specialized process, as depending on the size and age of the tree, a certain amount of pollen is given. This reduces the number of dates on a tree, thereby increasing the size of the dates. By researching the exact amount of pollen needed to maximize the growth potential of a date palm, fruit can be picked at all three harvests with little drop-off in production.

**Date Development**

The “perfect” date follows a growth stage that is evident by color and other physical characteristics. First, the dates develop in small bunches and are a greenish color. After a few weeks, the date enters the Kimiri stage, keeping a green color. When the date enters the Khalal stage after about 18 weeks, it starts to yellow. Between the Khalal and Rutub growth stages, the date is considered “smooth”. The date begins to digest the complex sugars stored in the fruit and changes color from yellow to brown in the Rutub stage at about 22 weeks. Between the Rutub and Tamer growth stages, the date is considered “moist”. After the dates digest the complex sugars, they lose moisture and become wrinkly. This is known as the Tamer stage, also called “juicy”, in which the fruit is ready to be consumed and is deemed ready for sale. The entire ripening process takes around 26 weeks, as displayed in Figure 1. It is possible that Khalal
dates can dry up before digesting their complex sugars, and this is demonstrated in the flow chart in Figure 2 as “skipped a step”.

**Figure 1: Medjool Date Maturity**

Figure 1 shows the maturity stages dates go through in a normal growing season. Starting around April as tiny buds, they grow to a light brown color by harvest season, which is August to September (Alhejjaj, Hashem & Ayad, Jamal, 2018).

**Figure 2: Medjool Date Maturity Stages**

Figure 2 gives a clear visual representation of Medjool date maturity stages. It is important to note that in the yellow and half-yellow stages, the dates could skip a stage and would then need treatment.
Relevance of Date Research

Understanding how to grow more profitable dates is vital for the success of the Arava Valley economy. Measuring weight and maturity are two concrete ways to define the more profitable dates. Larger dates sell for more and immature dates need extra treatment before they are ready for sale (Fancy, n.d.). Researchers focus on these categories to improve the profit for the farmers in the region.

Fruit Sorting with Computer Vision

As a group, we visited a date packaging facility called Ardom that is doing something similar to our project on a larger scale. Ardom also has the goal to improve date farmers' profit in the Southern Arava region. Ardom packages most of the dates grown in the Southern Arava region and uses computer vision to sort the dates it packages. Computer vision is a tool used in many automation systems to identify the object of interest within an image and then perform analysis. Computer vision breaks down the image into pixels that store color information. These pixels are often stored in a Red-Green-Blue (RGB), Hue-Saturation-Value (HSV), or grayscale format. RGB stores the color of a pixel in the levels of red, green, and blue that compose it. HSV stores the color of the pixel by using the base color on the color wheel called hue, the amount of that color compared to white, called saturation, and the amount of black in the color, called value. Grayscale images do not contain any color and instead contain only shades of black and white. These forms of image storage are useful because computer logic can be easily programmed to analyze particular features of an image in some forms of storage than others. For example, edges are easier to detect in greyscale than in HSV or RGB.

To isolate the relevant data from an image, subtraction, color masks, and contouring techniques are common. Image subtraction compares an “empty” image, which contains the typical background of the camera, with a new image, where the object of interest is in the frame. Pixels, where the images differ, are kept as regions of interest, and pixels where the image remains the same are changed to black as uninteresting regions. Color masks mark the areas of interest by looking for pixels that are in a specified range called a colorspace. If the pixel is in the colorspace, then it is assumed that the pixel is part of an object that is interesting. Contouring techniques find the edges of objects in the image by locating large changes in pixel colors and
define a complex shape from these edges. The pixels within the shape are retained as interesting and the pixels outside of the shape are changed to black as uninteresting.

**Machine Learning and Artificial Intelligence Training**

Machine learning and Artificial Intelligence (AI) algorithms are useful prediction tools that use previous knowledge to form their predictions. This previous knowledge is called training data. Machine learning and AI algorithms are trained using large collections of data with known outcomes. More training data often makes for a more accurate algorithm. Machine learning algorithms are common practice when classifying objects in an image.

**K-Nearest Neighbor**

K-Nearest Neighbor (KNN) is a machine learning technique that can be used to predict the classification of a data point. KNN abstracts the data field as a coordinate system, with each feature of the data representing a different dimension of the coordinate system. In the case of an HSV image, KNN would abstract the data field as a three-dimensional coordinate system with the hue on one axis, the saturation on a different axis, and the value on the third axis. Training data for a KNN algorithm is a list of data points and their true classifications. KNN then predicts the classification of new data based on a predefined number of nearest neighbors, called “k”. The mode of the k nearest neighbors is the predicted classification of the new data point. KNN, when used for classifying with a low number of fields, can often use less training data than more complicated techniques.

**Structured Query Language**

Structured Query Language (SQL) is used in writing and communicating with secure databases for information storage. To access an SQL database, the user needs to enter a username and password that are both defined when the database is created, or by a user with permission to create new users. A database consists of tables of information, in which the columns are defined by a name and the type of data that column can contain, and the rows are defined by a unique, or primary, key or set of keys that are not found in any other row in the same column of that table. SQL is typically used in relational databases, which allow tables to interact with each other through sharing columns that consist of unique keys of other tables.
Chapter 3: Understand

Formative Assessment

Informal interviews were held with the researchers at MOP to discuss their research process and what they need from the DSM. We learned that the researchers wanted the DSM to be able to sort dates into six different classes of maturity: “juicy”, “moist”, “smooth”, “dry”, “yellow/halfyellow”, and “skipped a step”. They needed the DSM to collect and report data on individual dates. For the DSM to be viable it had to meet the goal accuracy of at least 92% accurate. A few important things we learned during the first few weeks at MOP were that the dates are frozen after harvest until they are needed for analysis, and the technician gave us a walk-through of the DSM’s function before modification.

If we wanted the DSM to be able to distinguish between the different stages of maturity, we needed to think of physical characteristics, like color, that can be used to help categorize different classes of maturity. This meant the DSM would need to sort between dates with overlapping color spaces. This similarity in color spaces made distinguishing between maturity classes a non-trivial task.

DSM Before Modification

The DSM is a tool that was developed to help researchers quickly weigh and sort dates into five weight categories. Brought to Southern Arava R&D in 2017, the DSM was left with no documentation and MOP employees knew very little about it. The DSM is a technologically advanced machine, employing sensors, pneumatic valves, a load cell, and a rotating arm to collect data on Medjool date characteristics. Two conveyor belts move the dates upwards towards a mechanical vibration plate. If too many dates are on the vibration plate, the excess will fall into runoff bins. Past the vibration plate, the dates fall into a box that weighs them. After they are weighed, a trapdoor in the sensor area opens and a swivel mechanism guides the dates into the correct chute determined by weight. Figure 3 visualizes the important components of the DSM.
Figure 3: The DSM with Labeled Components

The labeled components, as shown above, are the most important parts of the DSM. The components all work together to weigh and sort dates.

The DSM synchronizes all its operations using the B&R Programmable Logic Controller (PLC) X20CP1382, which communicates with different elements of the DSM and can process the weight measurements and classify dates. The first communication point is between motion-activated sensors located at different spots on the DSM. Once the motion is detected by the sensors, a command is relayed to the PLC to turn on two electric motors, which are connected to two conveyor belts. Each electric motor is connected to an Invertek Opti-Drive, which regulates the motion of each conveyor belt. After the conveyor belt, each date falls onto a vibration plate. The vibration plate ensures that the dates are brought to the sensor area individually. A motion sensor above the sensor area recognizes when a date falls onto the scale. At the sensor area, the load cell (See Appendix A: User Manual) recognizes the weight change and transmits each date’s weight. A B&R T30 Power Panel, located on the side of the DSM, visualizes the weight of the date (See Appendix A: User Manual for more information on communication). Figure 4 shows the process that the DSM goes through when weighing and sorting each date.
Figure 4: DSM Flow Diagram

Figure 4 uses color to describe how the DSM works. Blue indicates the conveyor motion of the DSM. Orange is where the date is weighed. Green is where the dates are sorted into separate bins.

The DSM before modification stored the total weight of the dates in each weight class, but not the weight of each date. The researchers required the individual weight to test if their experiments had any merit. Thus, the total weight in each class was useless. The DSM could not do more complicated analyses, such as measuring quality level, maturity level, color, texture, or humidity. This resulted in researchers resorting to manual sorting, rather than relying on the DSM.

Manual Analysis

In the absence of an automatic solution, the researchers needed to weigh each date manually. The researchers used a digital scale and software called “Shekel” to record the individual weight of each date in an Excel sheet. They visually assessed the date to determine the date’s stage of maturation, using the color to distinguish between stages. Squeezing the date to feel its compliance is also a technique the researchers used to differentiate between stages of maturity. The presence of blisters is marked on the Excel sheet during this assessment, noting the approximate percent of blisters covering the date. Humidity is analyzed by crushing the date into a paste and using a different piece of equipment. This manual process to collect data from a single harvest of one tree takes the researchers about three days. With over 250 trees on the farm, the researchers spent nearly 6,000 hours per harvest collecting the data they needed for their research. This is time the researchers could spend doing more thought-intensive research.
Chapter 4: Additions to the DSM

Collecting and Storing information from the DSM

Given what we understood from the researcher’s needs, we made modifications to collect the weight from the PLC and transfer it to an Excel Sheet, made a program that can store tree identification and barcode in a database, and attached a camera to classify dates automatically. The data collected from the scale is relayed to the B&R PLC. The PLC communicates with the computer through "PVI" software (B&R Industrial Automation, n.d.). The PLC does not store the software logic in a readable human format, it only stores the variables and their value at any given time in machine language. The PVI software can read machine language and transmit the value of any given variable, but not the logic that controls the DSM and changes variables. The variables that store the weight and sensor information of the PLC were identified for use in our Python program. Using the PVI bridge, the python software can find the weight value and the status of the cell, indicating if there is a date in it or not. Python is then able to take pictures at the correct time and format this data in the way that the researchers requested.

The tree identification number and the day the dates were harvested are inputted by the user and merged with the rest of the data in SQL software. A script in Python uploads the data from the PLC into an SQL database. A second Python script sends SQL statements to the database to return the data in an Excel sheet format. The data was made accessible in Excel so that researchers could analyze the data according to their experiment design. A SQL database archives the data permanently which allows researchers to look at historical data and make better conclusions on new experiments. Researchers can easily run Python scripts to generate an Excel sheet from the same data being stored in the SQL database; the researchers are familiar with an Excel Sheet format so both options are available. This ensures that there is always a record of the original data even if the Excel sheets are changed. The dates are analyzed by batch analysis but are measured individually thus we needed to incorporate a SQL database to include each date in a group.

The HP Computer which was given to us at the beginning of the project is located on the right side of the DSM. We requested it to write our Python scripts and to give the DSM more technological capability. The computer can store data in an Excel sheet that is accessible to the
researchers. The Excel sheet, as shown in Figure 5, includes the maturity categories: juicy, moist, smooth, dry, yellow/half-yellow, and skipped a step. The weight of each date in grams was uploaded from the DSM to the database automatically. Figure 5 is in the same layout that the researchers were using before the DSM was updated and output a similar Excel sheet.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>date</td>
<td>barcode</td>
<td>weight</td>
<td>ready/juicy</td>
<td>moist</td>
<td>yellow</td>
<td>firm</td>
<td>blisters</td>
</tr>
<tr>
<td>488</td>
<td>13/09/21</td>
<td>101758</td>
<td>21.5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>489</td>
<td>13/09/21</td>
<td>101758</td>
<td>21.8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>490</td>
<td>13/09/21</td>
<td>101758</td>
<td>16.6</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>491</td>
<td>13/09/21</td>
<td>101758</td>
<td>27.5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>492</td>
<td>13/09/21</td>
<td>101758</td>
<td>23.4</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>493</td>
<td>13/09/21</td>
<td>101758</td>
<td>15.7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>494</td>
<td>13/09/21</td>
<td>101758</td>
<td>23.3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>495</td>
<td>13/09/21</td>
<td>101758</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>496</td>
<td>13/09/21</td>
<td>101758</td>
<td>24.7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>497</td>
<td>13/09/21</td>
<td>101758</td>
<td>18.8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>498</td>
<td>13/09/21</td>
<td>101758</td>
<td>25.8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>499</td>
<td>13/09/21</td>
<td>101758</td>
<td>17.6</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>13/09/21</td>
<td>101758</td>
<td>14.7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>13/09/21</td>
<td>101758</td>
<td>29.8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>502</td>
<td>13/09/21</td>
<td>101758</td>
<td>18.3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>503</td>
<td>13/09/21</td>
<td>101758</td>
<td>18.5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>13/09/21</td>
<td>101758</td>
<td>12.7</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>505</td>
<td>13/09/21</td>
<td>101758</td>
<td>10.5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>506</td>
<td>13/09/21</td>
<td>101758</td>
<td>8.3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>507</td>
<td>13/09/21</td>
<td>101758</td>
<td>19.2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5: Example of Excel Sheet Produced by the Updated DSM**

The Excel sheet in Figure 5 is the final output of the updated DSM. It was adapted from the researcher's Excel sheet which gives the sorted categories.

The DSM can save the weight into the SQL database in real-time while the DSM is running. Another python script classifies each entry in the database after all entries for the batch of dates are collected.

**Image Isolation**

We added an Alien Out of Space Technology camera and mount it to the DSM to get visual information about each date. We built a camera mount that is attached by a bolt to the sensor area, and the camera sits directly above the scale. A lamp is attached to the mount to ensure the entire sensor area is properly lit. For the machine learning algorithm to work effectively, the date in each image needs to be isolated from the background. This makes it easier for the machine learning algorithm to correctly identify the type of each date because it is only looking
at the pixels that differ from other images of dates when classifying each date. The camera is connected to the computer through USB. The uniformly lit images are sent to the computer to be analyzed.

The computer and DSM communicate via ethernet cable and use a Python script (see Appendix B) and an open-source library called OpenCV (OpenCV, n.d.) to recognize the maturity of each date. The camera image contained unnecessary data, such as shadows, so we needed to isolate the date using OpenCV subtraction and contouring functions. Image subtraction left the date completely visible in the image, as shown in Figure 6. From this mostly isolated date, we were able to isolate the date further using a detection method that finds the edges of the date so that only the date is shown. As depicted in Figure 7, these techniques were used together to isolate all the interesting pixels for classification.

![Figure 6: Before and After Image Subtraction](image)

Shown are two pictures, the left picture is the date in the sensor area before image subtraction, and the right is after complete image subtraction.
This diagram shows the methods for image isolation used in the order in which the image is processed.

**Automated Method to Distinguish Maturity**

**K-Nearest-Neighbor**

Once the date is isolated, we can extract features from it and use these features to determine how mature the date is. KNN determines the maturity category of each date. This technique uses a set of training images with known output categories for quality and maturity, and known input features of Hue, Saturation, and Value; the known values are stored in the SQL database and can be accessed easily. For each unclassified date image, we measured each input variable and compared it to a subset of the training data. If the features placed the date closer to more dates previously classified in one maturity category, we determined that the image was also in that category. These categories were assumed to be mutually exclusive, meaning the KNN algorithm could predict that only the most common label of the neighbors was correct.

The Alien Out of Space Technology camera obtains consistent images of Medjool date fruits because it has consistent lighting, never moves position, and takes a photo every time a date is in the sensor area. The image analysis is accurate enough for the researcher's purposes with the KNN form of classification and the combined forms of isolation. The classification enumeration for the KNN algorithm is as follows: 0 = dry, 1 = juicy or ready, 2 = moist, 3 = firm or smooth, 4 = yellow, 5 = half firm, and 6 = skipped a stage. These classifications were chosen because the researchers already classify based on these characteristics.
Training Data

The KNN algorithm was trained on a minimum of thirty images of dates in each maturity classification category. These images were taken using the DSM main_logger function, we then changed the file location and removed the data from the collected data table to avoid having unneeded data stored on the computer. A script was written to take the average hue, saturation, and value of each picture in the file. This outputted an array which was entered into a different function that adds the average features to the TrainingData table in SQL.

To keep the KNN algorithm accurate the TrainingData table in SQL must be maintained. If the algorithm produces a wrong prediction, it is recommended that this data point also be added to the training data database. When a new camera is installed on the DSM, the KNN algorithm needs to be retrained. To retrain the algorithm, all images from the database need to be removed and a new set of images need to be uploaded to SQL. It is recommended that every year, at least thirty new images be added to the training database to make sure that classification remains accurate.

Evaluation

After testing the updated DSM using a sample of 150 dates, the images were isolated enough to be classified by the KNN algorithm. The DSM was able to collect all the required data from these dates at a rate of one-hundred dates per ten minutes. Classification worked well with all of the tested date samples. We developed a data set for training KNN containing 180 images. The data set was broken up into classification data. When we left, the juicy and moist dataset had 50 images each, and the smooth and dry datasets had 40 images each.
Chapter 5: Integration and Instructional Media

Integration of the Modified DSM into the Research Process

We contacted the researchers at MOP regularly during the process of learning and updating the DSM. We had 3 meetings as follows: First Meeting: learn about the research they are conducting on dates and what they require from the updated DSM to use it. Meeting 2: Sorted dates into different categories required by researchers to train our KNN classification method correctly. Meeting 3: gave a progress report on the work we had completed so far and asked if anything we were doing was incorrect or if something else needed to be completed. These 3 meetings confirmed with the researchers that our project was viable and could be integrated into their research process. Researchers can now more quickly produce information on how successful their experiments were.

Instructional Media for Future Use of the DSM

We prepared instructional media for future users of the DSM. A manual contains detailed descriptions of the software, hardware, and modifications made to existing systems within the DSM and is usable to people with many different backgrounds (See Appendix A for Manual). Documentation for both old and new parts of the DSM were created and distributed to MOP. This included the technical manuals from all the separate parts of the DSM and an instructional video showing how anyone can use the DSM from a user perspective. We recorded a video in Hebrew so that any Hebrew-speaking user will be able to use the DSM correctly.

The user manual can be used by workers at MOP to use the DSM with minimal experience. The user manual contains five main sections which explain all the important aspects of using the DSM. The first section is System Overview. All of the components that the DSM uses are labeled, and graphics are used to point out specific components. The second section is Basic Use, which includes step-by-step instructions on how to use the DSM to weigh and classify dates. Section three is Troubleshooting, which addresses every issue that we encountered with the DSM and gives detailed instructions about how to fix these issues. Section four is Technical Specifications of the DSM. This section goes in-depth into what we learned about the specifications of each technical component of the DSM. Section five, Advanced Use and
Modification, explains how to train our KNN algorithm so that the highest accuracy rates can be preserved. The main contents of the user manual prevent a situation where the DSM falls out of use because there is no one to teach the workers how to use it.
Chapter 6: Discussion

Benefits of the DSM

The integration with DSM data collection allows researchers to get data more efficiently for their experiments. The addition of a camera with image classification methods produces actionable data for researchers to analyze. Creating this system streamlined the data collection and analysis process, allowing MOP researchers to use time effectively.

Our project was successful in meeting our objectives and the DSM was improved so it can be used by researchers. The overall weighing and classification time was reduced from 24 hours to 10 hours per tree per harvest. By saving 14 hours per harvest, the DSM completed its task of improving time management.

Currently, several areas of the DSM could use replacement, the first being the PLC. PLCs used as an automation solution are old, outdated, and are rarely used in the industry anymore. This leads to information about how to program and fix PLCs being lost and software upgrades being non-existent. The current camera is not the best for classification either, as it is a web camera and does not have the best capabilities for taking photographs of dates. The web camera has low frame and pixel rates compared to industry-standard automation cameras. The current mounting fixture could be upgraded. Currently, the mount cannot be adjusted, so it is hard to change the position of the camera. Recommendations and replacements play a large part in any modification effort, so we would like to point out the old components and give future recommendations.

Recommendations

We recommend that a component located in the PLC cabinet labeled Ewon Cozy 131 be removed from the DSM and replaced with an ethernet splitter.

The Ewon Cozy 131 (Ewon) is a piece of technology that does not match the needs of the DSM and is currently serving little purpose. At first, the Ewon seemed to be storing data from the PLC, but upon further analysis, we realized it was being used only as a convoluted ethernet splitter. Instead, one could add an Ethernet splitter that would allow for the computer and Power
Panel to be connected to the PLC simultaneously so that cords would not have to be constantly switched.

Also, the absence of the Ewon would use less power and free up cabinet space if another component needed to be added in the future.

**We recommend that future modifications of the DSM include real-time classification of dates.**

During our time at MOP, we learned that the source code of the B&R PLC is not stored in a readable format. This prevented us from integrating real-time classification. To classify in realtime, the PLC’s source code must be found. Real-time classification would save even more time than current upgrades and would further improve the analysis speed of the DSM because it would be classified instantly. At Ardom, each date is classified immediately, which allows for more training images and higher rates of classification accuracy.

**We recommend implementing a newer model of PLC that can store the logic in a readable format.**

We discovered through our research that the PLC currently used in the DSM is 23 years old. As a result, the PLC is outdated and does not cooperate well with new technology. Without the computer that wrote the original program, there is no way to change how the DSM functions. A newer PLC would store logic that can be changed, making updates to the DSM simpler and allowing for a wider variety of experiments.

**We recommend developing software that allows the DSM to open two-way communication with the computer.**

The DSM does not consider input from the computer when making logical decisions. This means that the DSM cannot sort dates based on the information gathered from the camera. We recommend rewriting the PLC program so that the source code is known and can be modified as needed to fit the requirements of MOP.

**We recommend purchasing PVI software and investing in POWERLINK cameras.**
We recommended purchasing the PVI software from B&R to ensure that the connection never has to be refreshed. Also, look into POWERLINK cameras because the PLC already has a POWERLINK port which supports easy integration with supported devices.

**We recommend automating how humidity and blistering are determined.**

Talking to researchers, we learned that they would like humidity and blistering detection in dates to be automated as well. Looking into ways that humidity and blistering can be determined using Python and computer vision is an option. For these reasons, we recommend talking to researchers to find the best ways to implement these features.

**Limitations**

This project was limited by the status of the DSM before the team arrived. The DSM was not regularly used for the six years before our arrival. This coincided with the company that had initially built the DSM dissolving, leaving no points of contact to instruct us on the technical details of the DSM. There was no documentation of the DSM to learn from. Additionally, the team had a limited number of classified dates at their disposal, which limited the number of original pictures taken of each date. However, the accuracy of the KNN algorithm was able to be preserved by taking multiple pictures of each date and compiling a substantial amount of training data.
Chapter 7: Conclusion

The old DSM did not fit the needs of the researchers at MOP. It was difficult to use and unable to gather the information the researchers needed, which resulted in the researchers using slower, manual data collection processes. The post-retrofit DSM was made possible by studying the old DSM’s antiquated processes along with input from the researchers who use it. Understanding the researchers’ goals for their studies was helpful when designing the system so it could be implemented in a way that could integrate it into their research workflow. New processes and instructional media were created so that this DSM could be integrated into the research process.

The new systems implemented with the DSM included a camera, computer vision to classify dates, automatic export of data to excel, and a database for easy access to information. These additions made it possible to automatically record collected data on individual dates in an Excel sheet and develop an automated method to distinguish maturity. The DSM is now easy to use, fast, and can gather most of the information the researchers need.
REFERENCES


APPENDIX A: Instructional Media

Item 1: Information Flow Diagram

Show above is an information flow diagram of how the DSM components work in conjunction to weigh and sort dates. Starting at the left of the page we can see where the date entered the DSM. The conveyor system is the first stage of the DSM, using two motion sensors, two NGB induction motors, and two Invertek Opti-Drive E^3s the conveyor belt system can bring dates to the vibration plate. The conveyor system is connected to the PLC as shown in the cabinet stage via the Invertek Opti-Drive. The PLC controls the Opti-Drive to regulate the speed at which the motors spin. The date then enters the sensor area stage. After the vibration plate drops the individual date onto the trapdoor the sensor activates the load cell and records the weight of the date. When the date is on the trapdoor a picture is taken. The camera is not directly connected to the cabinet stage however the load cell is. The last stage the date moves into is the sorting system. The PLC uses logic to determine what weight category the date should be sorted into. Depending on the user's class size, every date will be dropped into a bin. The swivel mechanism
is connected to the cabinet as the PLC tells what location it should be after the weight is recorded. It is important to note that the computer and camera operate separately from the PLC.

The program developed to classify the dates was not able to be implemented directly into the PLC, so it is a separate process.

**Item 2: User Manual**

Date Sorting Machine Manual

Written February 2023

By WPI Students:

Matt Adam

Ryan Antes

Connor Gaduette

Eli Hoffberg
Table of Contents

System Overview 4
  Parts Used 5

Basic Startup 6
  1. Turning on The Machine 6
  2. Calibration 7
  3. Connecting to the Computer 7
  4. Running the Machine 8

Troubleshooting 10
  Connecting Power Panel To PLC 10
  Connecting Computer to PLC 11
  Connecting the HP computer to the Internet via Ethernet 12

PLC to Computer Interfacing 13

How to Change SQL Database 14
  Username and Password for the MySQL Database 14
  Instruction for " MySQL Workbench" 14
  Startup 14
  Checking For a Server Connection 15
  Creating/Checking For a Schema 15
  Creating the necessary tables 16
  User-Table Interaction in MySQL 19
  Table Viewing 19
  Table Editing 21
  Table Deletion 21

MySQL Data 21

Technical Specifications of the DSM 22
  Date Sorting Machine Specifications 22
    Labeled Components of the Date Sorting Machine 23
  B&R Programmable Logic Controller (PLC) 26
  B&R T30 Power Panel 26
  Invertek Opti-Drive E^3 27
  Load Cell 28
  Camera 28
  Computer 28

Advanced Use and Modification 30
  Using Camera to Train Data 30

GitHub Information 31
  Code 31

Manuals 32

Contact Us 33
**System Overview**

The date sorting machine (DSM) uses 4 main components to weigh and sort dates.

**Component 1: Conveyor System**
- Transport system to move dates into position to be weighed individually.

**Component 2: Sensor Area**
- The sensor area photographs, weighs, and determines which weight category the date falls into. The weight categories are set on the T-30 power panel (See “Troubleshooting” section).

**Component 3: Sorting System**
- Taking the weight from the sensor area, the sorting system guides the date into the correct chute, which sends the data into buckets with dates within the same weight class.

**Component 4: Cabinet and Computer**
- The cabinet and computer are the brains of the DSM. The cabinet coordinates all the actions of the first 3 components and the computer produces the final output of the DSM.
Parts Used

Details of these parts can be found in the manuals in the “manuals” section at the end.

In **component 1**, a variety of parts are used to bring the dates to the sensor area. These include:

- Two Invertek Opti-Drive E^3s
- Three GRTB18S motion sensors
- Two NGB induction motors
- A wide conveyor belt
- A narrow conveyor belt

In **component 2** a few more complicated parts are used:

- Tedea Huntleigh 142 load cell
- Frisko air pressure valve
- Air Compressor
- One GRTB18S motion sensor
- Alien “Out of Space Technology” Camera and mount

In **component 3** one main part is used:

- The swivel system on a stepper motor rotates depending on the weight of each date

**Component 4** has the technology of the machine and is located in the cabinet below the machine and a computer to the right of the machine:

- B&R PLC X20CP1382 with 3 buses
- Everything is connected to the PLC besides the camera
- HP computer
- The cabinet contains the Invertek Opti-Drives
Basic Startup

1. Turning on The Machine

1. Go behind the building and plug the yellow air compressor into the blue extension cord. Plug the air hose into the female end of the air compressor. You should hear the compressor turn on at this point.
2. Make sure the lever at the bottom of the air tank is horizontal so that air is not leaking out. The compressor should look like the picture below.

3. Come back to the building and turn on the machine by making sure the switch is on (vertical) and then push the green button.
4. Ensure the switch on the right-hand side is angled 45 degrees from the upward-facing position. It should not be moved when the machine is turned off.
5. Inspect the red emergency stop lever and make sure it is pulled all the way out.
6. You will know the machine is on when you hear the conveyor belts moving.

2. Calibration

Calibration of the DSM is very important because the data will be skewed if calibration is incorrect. To complete calibration follow the steps below.

**Important: Make sure the air hose is not connected to DSM during calibration**
1. Switch the ethernet cable from the computer to Power Panel.
2. Go to the Power Panel on the left side of the machine.
3. On the bottom of the screen press the “cal” tile.
4. The screen will say “last weight” and “actual weight”. This is old data already collected; Assume it is accurate.
5. Hit next, and you will see “remove all weight from scale” then make sure the scale is free of any weight.
6. Hit next again then you may insert the 200g block into the scale found directly below the power panel, the screen will prompt you for the weight in grams. This is auto-filled with 200. Only change this number if using a different weight.
7. Hit next again, then remove the weight from the DSM.
8. Hit the “main” tile and resume normal weighing functions.

3. Connecting to the Computer

1. Confirm the ethernet cable is plugged into the machine and computer.
2. Confirm the ethernet cable is plugged into the machine. Please note that you will have to unplug the Power Panel from the machine to do this.
3. Open the PVI monitor app located on the desktop. Press File>Start Manager. If you intend to operate the machine for over 2 hours, ensure you remember to repeat this process in 2 hours or less. You can see the elapsed time in the PVI Monitor main window.
4. Running the Machine

1. Place 5 bins under each chute.
2. Run the “Main_Machine.py” script located on the desktop. Follow the instructions to complete the initial setup.
   a. The program will ensure that the camera is connected, the PLC is connected, and the database is running. See “Troubleshooting” if the program informs you that one of these systems is not configured properly.
   b. The program will ask you for the date the batch was picked as well as the batch ID number.
   c. Press ENTER when prompted to start recording dates.
3. Put the dates into the bottom of the large conveyor.
4. Once all dates have gone through the machine, press CTRL+C to begin the next phase of the program. It is now safe to turn off the machine and air compressor.
5. The classification script will begin running and print “You are now moving to the classification section of the program.” From here on the process will not require input.
   a. Before running the classification section make sure that Trainingdb in SQL has enough training data (at least 30 images per classification).
6. When the classification is done the program will print “The excel sheet is stored at” followed by the file path of the excel sheet into the terminal.
7. Navigate to the file path specified to find the final excel sheet produced by the system.
Troubleshooting

Connecting Power Panel To PLC

If the connection is ever lost between the PLC and the Power Panel on the side of the machine follow these steps to solve the issue, you will know if the connection is lost if the power panel says unable to connect to the socket, trying to reconnect.

1. Open the cabinet and ensure the ethernet cable with a blue mark at the end of it is plugged into the left ethernet port on the PLC.
2. Press the hand button in the bottom right corner of the Power Panel.
3. Once on the settings page navigate to the tile labeled “Network”.
4. In “Network” specify the hostname as HMI, and leave the DHCP and Activate DNS boxes unchecked.
5. On the IP address line enter 10.10.10.11, this is the IP address for the Power Panel.
6. On the subnet mask line enter 255.255.255.0.
7. On the default gateway line enter 10.10.10.1.
8. Navigate to the “VNC” tile on the left side of the screen.
9. On the server line, enter 10.10.10.10. This is the IP address for the PLC.
10. Navigate to the save and exit tile at the bottom of the screen. Make sure to hit “Save changes and Exit”.

Left Picture: Correct settings for the “Network” Tile. Right picture: Correct “VNC” tile settings.
Connecting Computer to PLC

Follow these steps to connect the PLC to the computer on the machine's right side. This needs to be performed every time the computer is switched from the internet to the machine via an ethernet port. These steps must be executed to run the program that classifies and stores date weights and qualities.

1. Plug the ethernet cable labeled “Machine” into the ethernet port at the back of the computer. Depending on which operation is occurring then the ethernet cable may already be plugged in, if the internet cable is plugged in follow these steps.
2. Open the cabinet and plug the other end of the ethernet cable into the left ethernet port on the PLC.
3. Turn the machine on by turning the knob on the cabinet to the right so it is vertical.
4. Push the green button and make sure you hear a click.
5. Move to the computer
6. Open the application labeled Automation Studio.
7. On the startup page, select the most recent project.
8. Once in the project navigate to the tile labeled Online. It is at the top of the screen.
9. Once in Online navigate to settings then click the button with a magnifying glass next to it.
10. On the right side of the screen, the PLC will appear.
11. The IP address and subnet mask will appear in red.
12. On the computer navigate to the control panel.
13. Open the network and sharing center and click on ethernet 4 settings.
14. Click on Properties and scroll down until Internet Protocol Version 4 (IPV 4) is visible.
15. Click on IPV 4 and hit the button Use the following IP address.
16. Change the IP address to 10.10.10.16 and the subnet mask to 255.255.255.0.
17. Click ok on the first two windows that appear and close on the third.
18. Return to Automation Studios online settings and hit the green connect button in the top left program.
19. You are now connected to the PLC. To make sure you are connected to the machine look for RUN in the bottom right corner.
20. Video instruction is saved on the computer. Navigate to downloads and look for instructional videos- Computer to PLC.

Connecting the HP computer to the Internet via Ethernet
This needs to be done anytime you want to access the internet after running the DSM because the computer does not have wifi capabilities and only has one ethernet port. Follow the steps below to complete.

1. Make sure the ethernet cable labeled internet is plugged into the computer, and the cable is located behind the HP computer.
2. On the computer navigate to control panel -> network and internet -> network and sharing center -> ethernet 4.
3. Double-click in ethernet 4 settings and click properties in the bottom left.
4. Then navigate to Internet protocol version 4 (IPV4) and double-click. Highlighted below is what Internet Protocol Version 4 looks like.
5. In IPV4 make sure the settings shown in the right picture are selected. Click obtain an IP Address Automatically and Obtain DNS server address automatically.
6. Once done, hit ok on the first two pop-up boxes and close on the third.
7. Wait a few seconds and you are now connected to the internet.

![Image of Internet Protocol Version 4 (TCP/IPv4) Properties]

**PLC to Computer Interfacing**

The PLC to PVI interface is used to watch the local variable that houses weight. PVI is used to transfer the weight from the PLC to the computer which is key for getting the individual weight of each date.
PVI is B&R’s way to connect PLCs to common devices like computers. PVI is a software download available through B&R and is specifically made to communicate easily with PLCs.

B&R built a low-level bridge between the PLC and higher-level things like windows and python. Automation Studio 4 as well as the libraries we use in our python code use PVI to communicate (read/write) with the variables stored in the PLC. Unfortunately, the trial license of PVI only allows for communication with the PLC for 2 hours before it has to be reset.

PVI comes installed with Automation Studio 4 and consists of a program called PVI manager and PVI monitor. Through the PVI monitor, you can reset the instance if the 2 hours end before you are done using the PVI software.

**How to Change SQL Database**

**Username and Password for the MySQL Database**

SQL username: root  
Password is stored in a .txt file named “MySQL_Password” on the desktop of the computer.

**Instruction for "MySQL Workbench"**

**Startup**

1. On startup, left-click the box under MySQL connections that says both “root” and “localhost:3306”.

2. A popup should appear prompting the user to type the MySQL Database password (specified in “GitHub Information”).
3. Type in the password and click “OK”.

Checking For a Server Connection

1. In the left bar under “Navigator”, click “Server Status”. A new tab should pop up.
2. If the “Server Status” is not “Running”, click “Startup/Shutdown” in the left bar under “Navigator”.
3. In the window that pops up, click “Start Server” to start the server.

Creating/Checking For a Schema

1. Open the "MySQL Workbench" application.
2. Left-click "Schemas" in the left bar, under "Navigator", next to "Administration", and above "Information".
3. If there is not an object named "datesdb", right-click the empty space under the other objects and left-click "Create Schema...".

4. In the middle of the screen, there should be a box labeled "Name:".
5. Replace the text in the box with "datesdb" and left-click the box that says "Apply". There should now be a pop-up that says "Apply SQL Script to Database".
6. Left-click the box that says "Apply", then left-click the box that says "Finish".
7. Double-click the object in the left bar under "SCHEMAS" named "datesdb".
8. Right-click "datesdb" and left-click "Set as Default Schema".
   You now have a schema, which can be thought of as a list of tables.
Creating the necessary tables

1. Create a table in which to store your information on dates if you do not already have one.
2. Left-click "File" at the top of the window and left-click "New Query Tab" or press "Ctrl" and "t" on the keyboard at the same time. This will create a new tab for creating scripts.
3. Run the following SQL script by copying and pasting it into the new tab to create a table named "Dates":

```sql
CREATE TABLE Dates (  
    imageAddress VARCHAR(255) NOT NULL, #image location  
    harvestDay DATE NOT NULL, #day of tree harvest, format YYYY-MM-DD  
    measureDay DATE NOT NULL, #day of date classification, format YYYY-MM-DD  
    barCode INT(6) UNSIGNED NOT NULL, #tree identifier  
    weight DOUBLE(6,3) NOT NULL,  
    readyOrJuicy BOOLEAN,  
    moist BOOLEAN,  
    yellow BOOLEAN,  
    halfFirm BOOLEAN,  
    firm BOOLEAN, #(AKA smooth)  
    /*  
    blisterPercent DOUBLE,  
    */  
    /*  
    blisterAmount VARCHAR(4),  
    */  
    blistered BOOLEAN,  
    skippedAStage BOOLEAN,  
    dry BOOLEAN,  
    /*  
    CONSTRAINT blisterPercent_CHECK CHECK (blisterPercent>=0 AND blisterPercent<=100),  
    */  
    /*  
    CONSTRAINT blisterAmount_CHECK CHECK (blisterAmount='none' OR blisterAmount='low' OR blisterAmount='mid' OR blisterAmount='high'),  
    */  
    CONSTRAINT Dates_PK PRIMARY KEY (imageAddress)
);```
*It should be noted that this code contains items that are commented out which may be implemented in further iterations of this project.

4. To run this script and create the table, click the yellow button near the top-left of the tab.

![Query 1](image)

The purpose of the next table is to store information for the KNN algorithm for classifying images.

5. Repeat the steps for the previous table, but paste the following script into the tab instead to create a table named "TrainingData":

```sql
CREATE TABLE TrainingData (  
    imageAddress VARCHAR(255) NOT NULL,  
    averageHue FLOAT(11,8) NOT NULL,  
    averageSaturation FLOAT(11,8) NOT NULL,  
    averageValue FLOAT(11,8) NOT NULL,  
    skippedStage BOOLEAN,  
    dry BOOLEAN,  
    CONSTRAINT blistersPercent CHECK CHECK (blisterPercent >= 0 AND blisterPercent <= 100)  
    CONSTRAINT blistersAmount CHECK CHECK (blisterAmount = 'none' OR blistersAmount >= 0),  
    CONSTRAINT Dates_PK PRIMARY KEY (imageAddress)
  )
```

# classifications are as follows:

# 0 = dry
# 1 = readyOrJuicy
# 2 = moist
# 3 = firm (AKA smooth)
# 4 = yellow
# 5 = halfFirm
# 6 = skippedAStage
classification INT NOT NULL,
    CONSTRAINT classification_CHECK CHECK (classification>=0 AND classification <=7),
CONSTRAINT averageHue_CHECK CHECK (averageHue>=0 AND averageHue<=180),
CONSTRAINT averageSaturation_CHECK CHECK (averageSaturation>=0 AND averageSaturation<=255),
CONSTRAINT averageValue_CHECK CHECK (averageValue>=0 AND averageValue<=255),
CONSTRAINT Dates_PK PRIMARY KEY (imageAddress)
);

*, Unlike the previous table, the comments in this table creation script are to help the user understand the meaning of the fields in the table.

6. To run this script and create the table, click the yellow button near the top-left of the tab.

User-Table Interaction in MySQL

If needed, tables or the data in them may need to be changed. Right-click a table's name in the column on the left labeled "SCHEMAS" to interact with it.

*Tables cannot be changed while the python script is running. If needed, press “Ctrl + c” to stop the python script.

Table Viewing

1. Right-click a table's name in the column on the left labeled "SCHEMAS" to interact with it.
2. Left-click "Select Rows" to open up a tab with a menu for changing table information.
3. Rows in the table can be filtered by typing specific values found in the table into the box next to the words “Filter Rows:”.
4. Individual rows can be changed as needed in the table by double-clicking on a value within that row or by clicking any of the 3 buttons to the right of the word “Edit:”.
   a. The leftmost “Edit:” button will allow the user to edit the current row selected.
   b. The middle “Edit:” button will allow the user to add a new row to the table.
   c. The rightmost “Edit:” button will allow the user to delete selected rows from the table.

5. Another row editing method is to click “Form Editor” under “Result Grid” to edit the currently selected row in the table. The other “Edit:” buttons for adding and removing rows in the table are also visible in the “Form Editor”.

6. When done with changes, click "Apply" to confirm changes or click "Revert" to cancel changes.
7. Clicking the blue “Refresh Button” will update the current view of the table to express recently made changes.
8. From this view, the table can also be exported to a file by clicking the button immediately left of “Export/Import” or imported to a file by clicking the second leftmost button from “Export/Import”.

9. Right-click a table’s name in the column on the left labeled "SCHEMAS" to interact with it.
10. Left-click "Alter Table..." to open up a tab with a menu for changing table information.
11. In the bottom left of the tab, click "Columns" to view and edit column information in the table.
12. Click "Indexes" next to "Columns" to view and edit information on table entries.
13. When done with changes, click "Apply" to confirm changes or click "Revert" to cancel changes.

Table Editing
Table Deletion

1. Right-click a table's name in the column on the left labeled "SCHEMAS" to interact with it.
2. Left-click "Drop Table" to delete a table and all the information stored in that table.
3. To confirm the deletion, click "Drop Now".

MySQL Data

Data from MySQL is stored at the following file address in .dbi files: "C:\ProgramData\MySQL\MySQL Server 8.0\Data"
Technical Specifications of the DSM

Explanation of parts used in each component are represented below with images.

Date Sorting Machine Specifications

The date-sorting machine uses many components for sorting and classifying dates. The various technical parts used in the machine are, a B&R Programmable Logic Controller (PLC) model number XCP1382, two Invertek Opti-Drive E3s, EWON Cosy 131, A B and R T30 power panel, and two NGB induction motors.

Labeled Components of the Date Sorting Machine

Besides the components mentioned previously, the date sorting machine also uses sensors, conveyor belts, a load cell, swivel mechanisms, and a vibration plate to weigh effectively and sort dates. Shown below the components are pictured and labeled. The first stage components are pictured below.

Stage 1 labeled components:
Stage 2 components are below. The vibration plate, load cell, motion-activated sensor, camera, and trap door are mentioned. It is essential to know that these two components need each other to work effectively. The data will be skewed if more than one date is dropped on the load cell. The vibration plate minimizes this risk by forcing the dates into a single file line.
Motion activated sensor: When date passes by weight is recorded.

Load Cell: used to record weight of date

Area where date is weighed

Vibration Plate: Forces date to separate into a single file line

Runoff for excess dates

Trap Door
Stage 2 camera and mount.

Stage 3 swivel mechanism. Run by stepper motor rotates depending on weight

B&R Programmable Logic Controller (PLC)

Stage 4 components are located in this section. The B&R PLC acts as the brain of the date-sorting machine. The PLC is comprised of three buses, one controls the swivel mechanism of the DSM. A second is connected to the scale and the third is connected
to the Invertek Opti-Drive. The PLC is labeled with all three buses and what they are connected to.

B&R T30 Power Panel

The T-30 power panel works in conjunction with the PLC. Connected via an ethernet cable, the Power Panel displays what the PLC is doing. On the panel, a human-machine interface (HMI) application is loaded. HMI applications are usually customized to the specific needs of the project. Shown above the power panel with the installed HMI application is shown. Currently, the program displays the date's weight and stores the weight for each category in a log. The power panel also has many features that can be changed by the user. On the main screen, the size of the bins can be changed by clicking on the green boxes and inputting your desired weight. 5 different size settings can be changed depending on your experiment. The dates will be dropped into these bins after they are weighed.
Invertek Opti-Drive E^3

The DSM uses two Invertek Opti-Drive E^3s to control the speed of two NGB induction motors. The NGB induction motors are electric and power the two conveyor belts that bring the dates to the scale. The Opti-Drives are controlled by the X20 SM 1436 PLC bus so everything can work in conjunction.
Load Cell

The load cell is used to weigh the date. The current model being used is a Tedea-Huntleigh 142. The load cell is connected to the PLC via the X20 AI 1744 bus. It is possible to remove the wiring from this bus and connect it to an Arduino or Raspberry Pie. An increase in weight triggers an increase in voltage.

Camera

We used an Alien Out of Space Technology web camera mounted above the scale to image the dates. The camera is plugged into a USB port on the back of the computer using an extension cord to reach the mount position. The main_logger function will find the camera plugged into the computer regardless of the port. The main_logger function calls an openCV function to capture images from the camera and saves them locally to analyze the data. The images are processed and edited using OpenCV functions to isolate the date. The isolated images are analyzed in the KNN functions. This camera was chosen because it was USB connected and it was easy to mount. The camera was mounted using wood planks and 13mm bolts to secure the mount to the machine.

Computer

The computer that is located on the right side of the machine is used to run the code that is classifying the dates, connects to the PLC, and outputs the weights of the dates. If you want to connect to the internet make sure that the blue-ended ethernet cable is plugged into the computer and not the white-ended cable.
Advanced Use and Modification

Using Camera to Train Data

The camera currently on the machine is used to take pictures as the dates are being called in logger_main. These pictures are sent to our isolation and KNN algorithms to determine what class the dates fall into.

To train the KNN database with a new camera, use MySQL software to delete all of the existing training data in the table named TrainingData and delete all of the images in the TrainingData folder. Take at least thirty new photos of each category and enter these into the TrainingData folder. These photos can be obtained by running the dates through the machine and copying or moving images from the folder they were stored into the folder named “TrainingImages”. This folder is in the same folder as all of the date images called “DateImages”. Use the file_AverageHSV function in the knn_functions file to get an array of the average HSV values for each image in the folder. Remember that the “file path” parameter must be a string of the file path where all “\” are replaced with “\\” and end with “\”. Enter these HSV values into training in the order returned by the function. Enter the classification enumeration in the column as you are entering the HSV values. You can also use the trainKNN function in MyApp to add a large array of HSV values that are all the same classification.

You can test the new KNN algorithm by running a sample of dates through the machine as normal, checking to see if the dates are properly classified. If the algorithm is not accurate enough, add more images to the folder and run the fileAverageHSV function again adding only the new rows into the SQL database. It is recommended that the images added are the images that were improperly classified.
GitHub Information

Put GitHub login info here and any other needed info to make it easier for the next group
GitHub username: DateMachine
Password is stored in a .txt file named “AccountInformation” on the desktop of the computer.

Code

Using python via Visual Studio Code we developed algorithms aimed at isolating and classifying dates. The first step we used was isolation which removed the entire background and left just the date. The code is attached below. Using K nearest neighbor and mask algorithms we began to classify the dates.

[Image: Date before Isolation and Completely Isolated Date]
Attached are the manuals that we found for all the components of the Date Sorting Machine. Each title color coordinates with the manual that it's named.

T-30 Power Panel Manual

Power Panel T30

B&R X20CP1382 Manual

Make sure to only look at the CP1382 section of the manual!


Invertek Opti-Drive E^3 Manual


Tede-Huntleigh Load Cell Specs

Contact Us

If this manual does not answer the questions you may have, do not hesitate to reach out to any of us.

For the computer to PLC, Power Panel to PLC connection, and any other hardware help reach out to:
ryanantes488@gmail.com

For isolation, python code, GitHub, and extracting source code from PLC help reach out to:
matthew.adam20@gmail.com

For K nearest neighbor, computer vision and camera help reach out to:
connorxc20@gmail.com

For database and coding help reach out to:
ebhoffberg@gmail.com
Appendix B: Project Code

Item 1: Code Summary

The program is designed to log weight and images in real-time with the DSM and to classify each date after the DSM has gathered the information about them. The program logs all this information in an SQL database locally on the DSM computer. It is also designed to automatically create an excel sheet with the gathered information in the format requested by the researchers.

It requires a connection to the PLC through the PVI bridge, a connection to the MySQL database, and a connection to a camera. The program connects to this equipment automatically or informs the user if the equipment is not running correctly.

Find detailed documentation about how to use the program within the DSM manual.

Item 2: Github Repository

ConnorBMERBE (n.d.). ConnorBMERBE/DateIQP: Project Repository for the "Advancing Medjool Date Research using the Date Sorting Machine" IQP at WPI. GitHub. Retrieved February 24, 2023, from https://github.com/connorBMERBE/DateIQP