



Quality Sorting of Green Coffee Beans from Wet Processing by Using The Principle of Machine Learning

Thanapat Thongnop, Tanayut Perpaman, Panchanit Kansiri, Warathep Nuchda, Supachai Peungsungwan*

Department of Electrical Technology Education, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi, Thailand

* Correspondence: E-mail: supachai.pue@kmutt.ac.th

ABSTRACT

Coffee beans are processed in a variety of ways such as beverages, foods, sweets, etc. Technology to sort coffee beans is still automated. We, therefore, proposed the concept of machine learning to be applied in coffee bean sorting. We consider the moisture content, size, colour, and characteristics of the coffee bean using image processing. Finally, we sorted the green coffee beans by the Thai coffee grading standards. There are three grades: A, X, and Y. Our machine has an accuracy of 85%. To improve the quality of our machines, the datasets used to train machines must be increased.

© 2021 Universitas Pendidikan Indonesia

ARTICLE INFO

Article History:

Submitted/Received 01 May 2021

First revised 02 Jun 2021

Accepted 04 Aug 2021

First available online 06 Aug 2021

Publication date 01 Sep 2021

Keyword:

Coffee beans dataset,

coffee beans,

Coffee dataset,

Green coffee beans,

Machine Learning,

Quality sorting.

1. INTRODUCTION

Coffee grading is known to be a very challenging task. The acquisition of quality coffee involves a variety of factors, including consumer trust. Machine learning technology is applied to a wide variety of data-related tasks. Most of today's problems are related to data and predictions. Designing a process that allows a machine to handle large amounts of data and have a learning process is extremely important (Faruk & Cahyono, 2018; Riza et al., 2021; Mediayani et al., 2019). Applying machine learning to coffee grading is an extremely challenging and interesting task. The process of acquiring features is important as the attributes extracted from the big data must be attributes that can promote the machine learning process in a positive direction (Arboleda et al., 2020). Assessing coffee quality is another point of great interest. Generally, coffee products all over the world are sold in two ways: domestic sales and exports abroad. Assessing coffee quality with machine learning has been of great interest to researchers (Feria-Morales, 2002).

This research presented the application of machine learning for grading coffee. This research studied the characteristics that can be used to solve coffee grading problems. Although there is a small amount of dataset that is suitable for small to large coffee businesses very well, the proposed coffee grading process consists of 6 sub-processes.

2. METHODS

The quality sorting of green coffee beans are following steps:

- (i) Research the data of green coffee beans. Data of green coffee beans were obtained from the wet process and the data of Thai coffee beans grading standards. Thai coffee beans grading standards have sorted the green coffee beans into 3 grades: A, X and Y.
- (ii) Get moisture data of green coffee beans. Moisture data of green coffee beans was to sort out the good green coffee beans. Good green coffee beans must have moisture less than 12%. The moisture data were obtained using a moisture sensor.
- (iii) Get 15 datasets of green coffee beans using image processing. 3 data are necessary to classify the quality of coffee beans: size, colour, and characteristics.
- (iv) Grading the green coffee beans. In this step, the received data was plotted using Python. On the graph, the X-axis is colour, the Y-axis is characteristics, and the Z-axis is the size of green coffee beans.
- (v) Get 20 datasets of green coffee beans, including size, colour, and characteristics using image processing to test the machine.
- (vi) Grading the green coffee beans from 20 datasets using Python for the principle of machine learning that we analyze in **Figure 4** to calculate the error and accuracy.
- (vii) We summarized all the gathered data.

3. RESULTS AND DISCUSSION

From the train of 15 datasets by plot graph, the grouping of coffee substances in each grade was obtained (**Figure 2**). The cluster represented by green is A-grade coffee compounds with a size greater than 5.6 mm and dark green. Intact beans humidity is not more than 12%. The group represented by blue is X-grade coffee compounds with a size greater than 5.6 mm and dark green. Intact beans humidity is not more than 12%. The group represented by black is Y-grade coffee compounds with a size greater than 5.6 millimetres and dark green. Intact beans humidity is not more than 12%. In **Figure 2**, X-axis is denoted by grain appearance, Y-axis is denoted by grain size, and Z-axis is denoted by the colour of green coffee beans. We found errors from this curve since there are red points in every cluster. Red points are datasets of

bad green coffee beans (moisture more than 12%). We can not sort out the bad green coffee beans from every cluster. Thus, moisture data cannot be used for grading green coffee beans. We took the moisture data to sort the bad green coffee beans and graded the remaining good coffee beans (**Figure 3**).

Figure 3 shows the clustering of data without moisture data. We can divide the points shown on the graph into 3 clusters by referring to the data from Thai coffee bean grading standards. After that, we will use the datasets in Table 2 to test the grading accuracy shown in **Figure 4**. From **Figure 4**, there are 3 orange points outside the cluster. We calculated an error of 15% and an accuracy of 85%. Our experimental limitation is due to less data information, reducing the accuracy of coffee bean grading work. It is necessary to add more information that will be used to train the machine to achieve greater accuracy and precision.

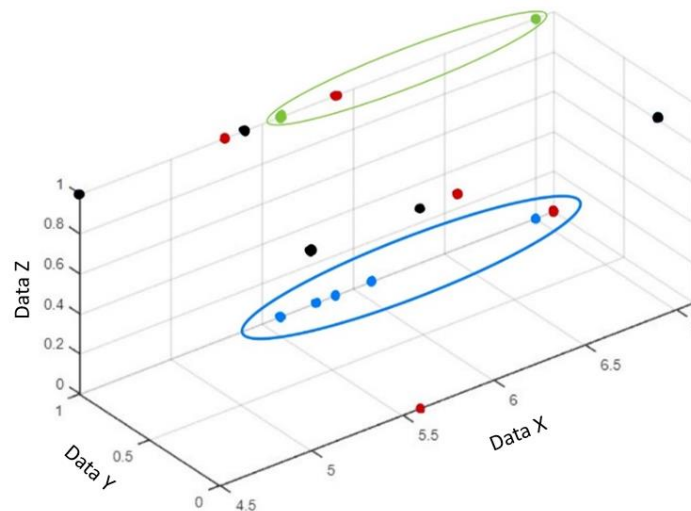


Figure 2. The clustering of data. Green, blue, black, and red colours are clustering of A, X, Y, and bad green coffee beans (moisture more than 12%), respectively.

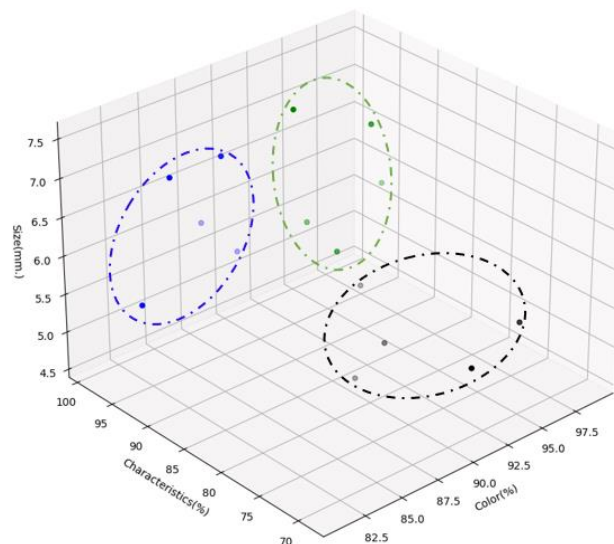


Figure 3. Show the clustering of data from table 1 (without moisture). Green, blue, and black colours are clustering of A, X, and Y, respectively.

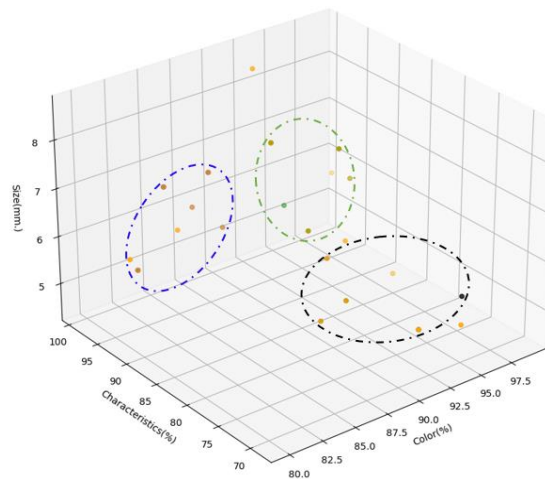


Figure 4. Show the clustering of data. Green, blue, black, and orange colours are clustering of A, X, Y, and 20 datasets for testing, respectively.

4. CONCLUSION

We have researched the information necessary for grading coffee beans to apply machine learning principles to bean grading based on size, colour, characteristics, and moisture content. These parameters are important for grading, classified as A, X and Y based on the Thai standard coffee beans.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

7. REFERENCES

- Arboleda, E. R., Fajardo, A. C., and Medina, R. P. (2020). Green coffee beans feature extractor using image processing. *Telkomnika*, 18(4), 2027.
- Faruk, A., and Cahyono, E. S. (2018). Prediction and classification of low birth weight data using machine learning techniques. *Indonesian Journal of Science and Technology*, 3(1), 18-28.
- Feria-Morales, A. M. (2002). Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. *Food Quality and Preference*, 13(6), 355-367.
- Mediayani, M., Wibisono, Y., Riza, L. S., and Pérez, A. R. (2019). Determining trending topics in twitter with a data-streaming method in R. *Indonesian Journal of Science and Technology*, 4(1), 148-157.
- Riza, L. S., Rosdiyana, R. A., Pérez, A. R., and Wahyudin, A. The K-Means Algorithm for Generating Sets of Items in Educational Assessment. *Indonesian Journal of Science and Technology*, 6(1), 93-100.