



BLACK TEA GRADE CLASSIFICATION USING PROBABILISTIC NEURAL NETWORK (PNN)

Evi Indah Sari¹, Nurul Huda Prasetya², Riri Syafitri Lubis³

Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received October 11, 2020
Revised November 12, 2020
Accepted December 23, 2020

Keywords:

Probabilistic Neural Network,
K-Means Clustering,
Black Tea Grade

ABSTRACT

North Sumatera tea is known as black tea, one which is produced by PT Perkebunan Nusantara IV Unit Bah Butong which produce 16 types of black tea. This research aims to classify black tea grades using Probabilistic Neural Network method and determine the accuracy value of black tea classification using Probabilistic Neural Network method. Data used are data of characteristics of 16 black tea types with the attributes id, types of tea, colour density, partcel weight, partcel size and partcel shape. To get the best accuracy result, training data and testing data are divided using K-Means Clustering. Futhermore, classify the testing data using Probabilistik the result obtained a grade 1 classification class totaling 7 records with a brownish black appearance and granutes partcel, grade 2 totalling 7 record with a brownish appearance and grade 3 totaling 2 records with reddish appearance and not determine its shape and obtain an accuracy value of 80,00%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Evi Indah Sari,
Department of Mathematics,
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: eviindahsari05@gmail.com

1. INTRODUCTION

Tea is one of the plantation commodities that has an important role in Indonesia's economic activities. Tea is also one of Indonesia's export commodities which is quite important as a foreign exchange earner in addition to oil and gas. As a beverage ingredient, tea has more value compared to other beverages, considering that tea is rich in minerals and vitamins needed by the body. Various health benefits of tea have also been recognized by nutrition experts. In addition to the increasingly open export opportunities, the domestic tea market is still quite large even though it has not been explored to its full potential. Domestic market opportunities are increasingly open, if followed by improving the quality of tea (BPS Indonesian Tea Statistic, 2017).

Black tea in North Sumatra is currently produced by PT Perkebunan Nusantara IV which consists of 3 Business Units, namely the Sidamanik, Bah Butong, and Tobasari Business Units which have been combined into 1 (one) business unit, namely the Tea Business Unit. The black tea processing system at PT Perkebunan Nusantara IV currently uses the Orthodox system. Fresh tea shoots are processed into black tea with the combined Orthodox-Rotor Vane system. The processing capacity of the Bah Butong factory is 1530 Kg of dry tea per hour and the capacity of the shoots of Wet Tea is ± 100 tons per day. Through a sorting process, PT Perkebunan Nusantara IV Unit Bah Butong produces 16 types of black tea which are

then classified into 3 grades, namely grade 1, grade 2, grade 3 (Overview of PT Perkebunan Nusantara IV Unit Bah Butong, 2019).

Currently, the black tea production process, especially in the sorting and grading process, still relies on the foreman's ability as a manual quality elevator. By relying on humans as the only quality elevator, the accuracy rate for recognizing product defects is relatively low, it is very rare to find a single inspector who successfully detects defects exceeding 50-60% (Ulhaq, 2015). At PT Perkebunan Nusantara IV unit Bah Butong, in the grading process it is not known the value of the classification accuracy of each grade. Therefore, it is necessary to build a system to identify the quality grade of black tea based on practical standards, namely by using the Probabilistic Neural Network method.

2. RESEARCH METHODE

2.1 K-Means Clustering

K-Means Clustering is a non-hierarchical data grouping method that partitions existing data into two or more groups so that data with the same characteristics are put into the same group and data with different characteristics are grouped into other groups. The purpose of doing data grouping is to minimize the objective function that has been set in a group and maximize the variation between groups. The steps of K-Means Clustering method can be described below (Dilla, 2020).

1. Determine the value of k as the number of clusters to be formed.
2. Initialize k as initial centroid randomly.
3. Calculate the distance of each data based on equation (1).

$$DL(x_2 - x_1) = || x_2 - x_1 || \quad (1)$$

To calculate the distance using the Euclidean Distance equation can be seen in equation (2).

$$DL(x_2 - x_1) = || x_2 - x_1 ||_2 \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad (2)$$

4. Group each data based on the distance between the data and its centroid.
5. Determine the position of the new centroid.
6. Return to step 3 if the position of the new centroid is not the same as the old centroid.

2.2 Probabilistic Neural Network

Probabilistic Neural Network was first developed by Donald Specht. Probabilistic Neural Network is an artificial neural network method that uses supervised training. Probabilistic Neural Network is an Artificial Neural Network which is generally used for classification and pattern recognition problems. The Probabilistic Neural Network classification method is formed based on statistical principles to overcome problems in the backpropagation method prone to the wrong minimum value (Alex, 2018).

The PNN structure consists of four layers, namely the input layer, pattern layer, summation layer, and decision layer.

1. Input Layer

An input containing test data that will be used for PNN calculations.

2. Pattern Layer

Performing the calculation of the distance between the test data and the training data can be seen in equation (3).

$$y(x) = e^{-\frac{(x-X)^2}{\sigma^2}} \quad (3)$$

with : x = test data atribut value
 X = average value of training data
 σ^2 = standard deviation of training data

3. Summation Layer

Calculate the average per class so that it will be obtained how many possibilities of an input into a class can be seen in equation (4).

$$g(x) = \sum y(x) \quad (4)$$

4. Output Layer

For the output of the PNN method is the largest value from the other classes from the calculation results of the summation layer can be seen in equation (5).

$$\text{Output } (z) = \max(gx). \quad (5)$$

3. RESULT AND ANALYSIS

Data characteristics of black tea PT. Perkebunan Nusantara IV 2020, which contains variables of color density, particle weight, particle size and particle shape.

Table 1. Characteristic of Black Tea Data

Id	Types of Black Tea	Color Density	Particle Weight	Particle Size	Particle Shape
1	BOPI	Fairly Black	350	12	Grainy
2	BOP	Brownish	350	14	Grainy
3	BOPF	Brownish	335	18	Grainy
4	BP	Brownish	250	22	Grainy
5	BT	Brownish	420	30	Grainy
6	PF	Brownish	295	24	Grainy
7	DUST	Brownish	225	30	Grainy
8	BP II	Brownish	260	24	Leafy
9	BT II	Fairly Black	350	18	Leafy
10	PF II	Brownish	290	22	Leafy
11	DUST II	Brownish	245	60	Leafy
12	DUST III	Reddish	240	40	Leafy
13	DUST IV	Brownish	240	60	Leafy
14	FANN	Brownish	295	30	Leafy
15	BM	Reddish	430	18	Rough
16	FLUFF	Reddish	370	30	Rough

3.1 Grouping Training Data and Data Testing Using K-Means Clustering

Table 2. Cluster distance and Position

Id	C ₁	C ₂	C ₃	Min distance	Cluster
1	0,493	0,959	1,457	0,493	1
2	0,387	0,753	1,187	0,387	1
3	0,380	0,660	1,210	0,380	1
4	0,617	0,480	1,424	0,480	2
5	0,556	0,896	1,147	0,556	1
6	0,463	0,521	1,300	0,463	1
7	0,751	0,427	1,521	0,427	2
8	0,509	0,287	1,097	0,287	2
9	0,373	0,827	1,184	0,373	1
10	0,381	0,364	0,987	0,364	2
11	1,025	0,522	1,442	0,522	2
12	0,976	0,469	1,149	0,469	2
13	1,037	0,525	1,458	0,525	2
14	0,425	0,268	0,998	0,268	2
15	1,021	1,203	0,000	0,000	3
16	0,975	0,953	0,385	0,385	3

From Table 2 above, information is obtained that with the K-Means Clustering calculation, it is known that cluster 1 has 6 records, cluster 2 has 8 records, and cluster 3 has 2 records. Then the distribution of training data and testing data is carried out with a predetermined percentage of 70% training data and 30% of testing data.

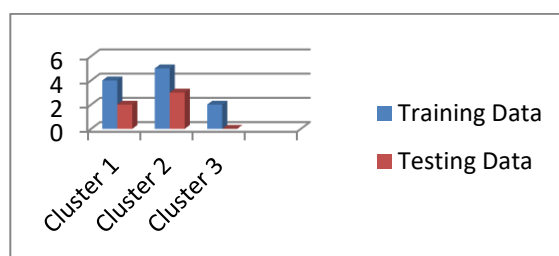


Figure 1. Grouping of training data and testing data

in this study there are 16 data, in accordance with the comparison of training data and testing data that have been determined then,

1. Training data = lots of data x percentage of training data
 $= 16 \times 70\%$
 $= 11$ Training Data
2. Testing data = lots of data x percentage of training data
 $= 16 \times 30\%$
 $= 5$ Testing Data

3.2 Classification using Probabilistic Neural Network

Test results with the Probabilistic Neural Network method can be seen in Table 3.

Table 3. PNN Test Result

Id	Color Density	Particle Weight	Particle Size	Particle Shape	Statement
1	0,000	0,610	0,000	0,000	Valid
2	0,500	0,610	0,042	0,000	Valid
7	0,500	0,000	0,375	0,000	Valid
8	0,500	0,171	0,250	0,500	Valid
13	0,500	0,073	1,000	0,500	Tidak Valid

In Table 5, it can be seen that the results obtained are as follows:

1. The maximum value of type 1 tea is at grade 1, so it is classified into grade 1, this is in line with the test results using K-Means clustering so that type 1 tea is given a valid statement.
2. The maximum value of type 2 tea is at grade 1, so it is classified into grade 1, this is in line with the results of the test using K-Means clustering so that type 2 tea is given a valid statement.
3. Type 7 tea has a maximum value of grade 2 so it is classified into grade 2, this is in line with the results of the test using K-Means clustering so that type 7 tea is given a valid statement.
4. Type 8 tea has a maximum value of grade 2, so it is classified into grade 2, this is in line with the test results using K-Means clustering so that type 8 tea is given a valid statement.
5. The maximum value of type 13 tea is at grade 1, so it is classified into grade 1, this is not in line with the test results using K-Means clustering where type 13 tea should be at grade 2 so that type 13 tea is given the Invalid statement.

As for through this test, the accuracy value can be calculated as follows,

$$\% \text{ Test identification success} : \frac{4}{5} \times 100 \% = 80.00 \%$$

$$\% \text{ Test identification failure} : \frac{1}{5} \times 100 \% = 20.00 \%$$

4. CONCLUSIOON

Based on the results of the analysis of the black tea classification of PT Perkebunan Nusantara IV Bah Butong Unit that has been carried out, it can be concluded that:

1. By using the Probabilistic Neural Network method obtained from training data and test results of data testing, 3 classification classes are obtained, namely:
 - Grade 1 consists of 7 records with brownish black appearance and granular particles.
 - Grade 2 consists of 7 records with the appearance of a brownish color and rather large particles.
 - Grade 3 consists of 2 records with the appearance of a reddish color and erratic particle shape.
2. In the Probabilistic Neural Network Method, an accuracy of 80.00% was obtained from the test results of 5 test data and 11 training data with a failure percentage of 20.00%.

REFERENCES

- [1] Avicienna, "Klasifikasi Teh Hitam Menggunakan Tekstur Tamura dan Klasifikasi *Naive Bayes*." *jurnal Bogor Agricultural University*, 2015.
- [2] Badan Pusat Statistik Republik Indonesia, Statistik Teh 2017, Sumatera Utara: Badan Pusat Statistik Provinsi Sumatera Utara, 2017.
- [3] Kurniati, Dilla. "Penerapan *Probabilistic Neural Network* dan *Bacpropagation Neural Network* Untuk Klasifikasi Produksi Garam Di Indonesia." Skripsi Uin Suska, 2020.
- [4] Mitsa, Theophano, "Themporal Data Mining. United State of America : Taylor and Francis Group
- [5] Nurbaiti. 2017. Identifikasi Bibit Pada Tanaman Laan Gambut Berdasarkan Bentuk Daun Menggunakan *Probabilistic Neural Network (PNN)* Berbasis *Webside*." *Jurnal Universitas Tanjungpura*, 2019.
- [6] Nugroho, S, "Pengantar Statistika Matematika." Edisi Pertama. UNIB Press, 2008.
- [7] Rahmani, "Clustering of Image Data Using *K-Means* and *Fuzzy K-Means*. *International journal of Advanced Computer Science and Application*." 2014.
- [8] Selayang Pandang PT Perkebunan Nusantara IV Unit Bah Butong, 2019.
- [9] Tindaon., R. Fandy, "Identifikasi Sistem Produksi Teh Hitam Di PT. Perkebunan Nusantara IV Kebun Bah Butong" Skripsi Universitas Sumatera Utara, 2009.
- [10] Wijaya, Alex, "Klasifikasi Spesies Ikan Menggunakan Probabilistic Neural Network." Skripsi Universitas Sumatera Utara, 2018.
- [11] Yudi, Sari, "Penerapan Algoritma Neural Network Untuk Klasifikasi Kardiotokografi." *Jurnal BSI University*, 2018.