

# PRINCIPAL COMPONENT DAN K-MEANS CLUSTER ANALYSIS UNTUK DATA SPEKTRUM BLACK TEA GRADES GUNA PENILAIAN KUALITAS ALTERNATIF

Aditya Toraismaya<sup>1</sup>, Leopoldus Ricky Sasongko<sup>2\*</sup>, Ferdy Samuel Rondonuwu<sup>3</sup>

<sup>1,2,3</sup>Program Studi Matematika, Universitas Kristen Satya Wacana, Salatiga, Indonesia

Email : <sup>1</sup>662016014@student.uksw.edu, <sup>2</sup>leopoldus.sasongko@uksw.edu, <sup>3</sup>ferdy.rondonuwu@uksw.edu

\*Penulis Korespondensi

**Abstract.** The aim of this article is to apply the method to measure or evaluate the sampling quality of black tea in determining its category or class based on the spectrum of the sampled data. The number of the spectrum of the variable of data reduced by the *Principal Components Analysis* (PCA) method becomes a new variable that will be classified later by using *K-Means Clustering* method. This research use 120 sample of tea from Fanning II (F-II), Pekoe Fanning (PFANN), and Broken Orange Pekoe Fannings (BOPF) with 90 sample used for training and 30 sample used for validation. The method and the analysis used in this research gave effective and efficient performance in measuring/evaluating the quality of the black tea sample to determine its class as it showed that the accuracy of *K-Means Clustering* results is larger than 50%.

**Kata Kunci:** Spectrum data, *K-Means Clustering*, PCA, black tea.

**Abstrak.** Tujuan dari makalah ini adalah mengaplikasikan metode untuk mengukur atau menilai kualitas sampel teh hitam untuk menentukan kategori atau kelas mereka berdasarkan sampel data spektrum. Jumlah variabel data spektrum dikurangi dengan metode *Principal Components Analysis* (PCA) menjadi variabel baru yang kemudian akan dikelompokkan dengan metode *K-Means Clustering*. Penelitian ini menggunakan 120 sampel teh dari Fanning II (F-II), Pekoe Fanning (PFANN), dan Broken Orange Pekoe Fannings (BOPF) dengan 90 sampel sebagai data pelatihan dan 30 sampel sebagai data validasi. Metode dan analisis yang dilakukan dalam penelitian ini cukup efektif dan efisien untuk mengukur/menilai kualitas sampel teh hitam untuk menentukan kelasnya karena memiliki tingkat keakuratan hasil *K-Means Clustering* diatas 50%.

**Kata Kunci:** Data Spektrum, *K-Means Clustering*, PCA, Teh Hitam.

## I. PENDAHULUAN

Indonesia adalah negara yang menempati urutan kedua di ASEAN dan urutan kedelapan di dunia sebagai produsen teh. Hal itu membuat Indonesia menjadi eksportir teh urutan kedua di ASEAN dan urutan ketujuh di dunia. Volume ekspor teh Indonesia pada tahun 2015 adalah sebesar 61.915 ton dengan nilai jual mencapai US \$126.051 [1]. Penilaian kualitas sampel teh perlu dilakukan guna mengetahui teh itu masuk ke dalam kategori jenis atau *grade* tertentu agar kemudian teh itu layak untuk diekspor. Efektivitas dan efisiensi suatu metode penilaian kualitas tersebut perlu untuk menjadi perhatian.

Penilaian kualitas teh hitam dilakukan oleh suatu badan khusus dan meliputi uji-uji seperti uji kimia, fisika, visual (organoleptik), serta mikrobiologi [2]. Kualitas teh hitam ditentukan oleh beberapa kriteria seperti ukuran partikel, warna, liquor taste, dan tampilan saat infusikan [3]. Alur prosedur yang panjang dalam penilaian kualitas suatu teh membutuhkan biaya dan waktu yang tidak sedikit. Sementara itu, ada uji penilaian yang bersifat subyektif karena mengandalkan panca indera manusia. Dapat dijumpai pula kesalahan penentuan kategori *grade* teh hitam jika ada banyak sampel teh yang diuji. Pemalsuan teh juga dapat dilakukan oleh orang-orang yang tidak bertanggungjawab.

Pengukuran *Near Infrared Spectroscopy* (NIR) dalam kombinasi dengan kemometri, telah banyak digunakan untuk penilaian kualitas pengobatan dan untuk mendeteksi obat palsu selama bertahun-tahun [4]. Metode yang melibatkan *Near Infrared Spectroscopy* (NIRS) dapat menjadi metode alternatif uji penilaian kualitas teh hitam guna memberikan solusi terhadap permasalahan atau kelemahan yang ada pada uji-uji lainnya karena metode yang melibatkan NIRS adalah relatif cepat, ramah lingkungan, serta memiliki penilaian yang obyektif [5,6]. NIRS memanfaatkan informasi hasil *spectroscopy* suatu bahan organik guna dapat dilakukan analisis kualitasnya [7]. Sampel teh hitam yang dilarutkan dalam air diletakan di dalam spektrometer guna ditembak oleh gelombang inframerah dengan panjang 10000 sampai dengan  $4000\text{ cm}^{-1}$  yang mana hasil pengukuran dari spektrometer itu kemudian disebut sebagai data spektrum. Ukuran matriks data spektrum yang besar akan direduksi dengan *Principal Components Analysis* (PCA) sehingga didapati peubah-peubah baru dengan banyak dimensi yang lebih sedikit. Kemudian analisis dapat dilanjutkan dengan menggunakan *K-Means Cluster Analysis* guna penentuan kelompok-kelompok jenis atau *grade* teh hitam yang didasarkan pada peubah-peubah baru hasil reduksi PCA. Penelitian ini menggunakan 120 sampel teh dari *Fanning II* (F-II), *Pekoe Fanning* (PFANN), dan *Broken Orange Pekoe Fannings* (BOPF) dengan 90 sampel sebagai data pelatihan dan 30 sampel sebagai data validasi.

## II. METODE

*Principal Component Analysis* (PCA) adalah salah satu metode yang dapat digunakan untuk mereduksi suatu data berdimensi tinggi atau data dengan banyak variabel. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi informasi pada data tersebut secara signifikan [8-11]. Metode PCA memiliki keuntungan untuk menghilangkan korelasi tanpa kehilangan variabel aslinya [12]. Nilai variabel-variabel pada data yang mungkin saling berkorelasi dikenai transformasi linier dan direduksi sehingga menjadi variabel-variabel baru dengan dimensi yang sedikit dan tidak saling berkorelasi. Transformasi dilakukan sedemikian sehingga proyeksi nilai variabel-variabel pada data di *Principal Axes* memiliki variansi yang maksimum pada tiap variabel. Variabel-variabel baru itu kemudian disebut *Principal Components* (PCs) atau *PC Scores*.

### Algoritma untuk Perolehan Principal Components (PCs) [13, 14, 15]

1. Inialisasi matriks data  $X$  (*centered*) yang berukuran  $n \times p$  dengan  $n$  adalah banyak sampel,  $p$  adalah banyak variabel; dan  $k \leq p$  dengan  $k$  adalah banyaknya *Principal Axes*.
2. Tentukan semua nilai eigen  $\alpha$  terurut dari matriks  $X^T X$  yang berukuran  $X^T X$  sehingga dimiliki paling banyak  $p$  nilai eigen dengan cara:
  - 2.a. Hitung semua solusi  $\alpha$  dari persamaan  $\det(X^T X - \alpha I) = 0$ ,
  - 2.b. Urutkan semua solusi  $\alpha$  secara descending,  $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \dots \geq \alpha_p$ .

3. Tentukan matriks eigenvector  $V$  dengan cara:
  - 3.a. Hitung  $v_j^*$ , vektor berukuran  $p \times 1$ , yang merupakan solusi dari persamaan  $X^T X v = \alpha_j v$  untuk  $j = 1, 2, 3, \dots, p$ ,
  - 3.b. Hitung panjang atau norm  $v_j^*$  yaitu  $v_j^*$  untuk  $j = 1, 2, 3, \dots, p$ ,
  - 3.c. Hitung vektor unit  $\hat{v}_j = v_j^* / \|v_j^*\|$  untuk  $j = 1, 2, 3, \dots, p$ ,
  - 3.d. Peroleh  $V = [\hat{v}_1 \ \hat{v}_2 \ \hat{v}_3 \ \dots \ \hat{v}_p]$ .
4. Peroleh PCs untuk k kolom pertama dengan cara:
  - 4.a. Tentukan  $V_k$  yaitu k kolom pertama dari matriks  $V$ ,
  - 4.b. Hitung  $XV_k$ , PCs untuk k kolom pertama.

*Clustering* adalah proses di mana titik-titik data serupa dikelompokkan dalam grup yang sama sedangkan titik data yang berbeda dikelompokkan pada grup yang berbeda berdasarkan beberapa gagasan tentang 'kesamaan' [16]. Terdapat dua jenis teknik pengelompokan yaitu hierarki dan partisi pengelompokan [17]. *K-Means* adalah salah satu analisis pengelompokan yaitu hierarki yang paling populer dan algoritma pengelompokan yang banyak digunakan karena reliabel dan simpel [18].

#### Metode *K-Means Clustering* [19, 20]

Misal diketahui data sebanyak  $n$  titik yaitu  $Y_1, Y_2, Y_3, \dots, Y_n$  yang mana setiap titik itu memiliki sebanyak  $k$  nilai peubah yaitu  $Y_i = (y_{i1}, y_{i2}, y_{i3}, \dots, y_{ik})$ , untuk  $i = 1, 2, 3, \dots, n$ . Kemudian  $n$  titik itu akan dikelompokkan ke dalam  $q$  kelompok yaitu  $C_1, C_2, C_3, \dots, C_q$  dengan titik-titik pusat awalnya berurutan adalah  $\mu_1, \mu_2, \mu_3, \dots, \mu_q$  dengan  $\mu_l = (\mu_{l1}, \mu_{l2}, \mu_{l3}, \dots, \mu_{lk})$ , untuk  $l = 1, 2, 3, \dots, q$ . Pengelompokan melalui metode *K-Means Clustering* untuk data itu dapat dilakukan dengan meminimumkan fungsi tujuan

$$D = \sum_{l=1}^q \sum_{Y_i \in C_l} d(Y_i, \mu_l) \quad (1)$$

dengan  $d(Y_i, \mu_l)$  menyatakan jarak setiap titik terhadap titik pusat kelompoknya. Pengukuran jarak dijabarkan sebagai berikut:

$$d(Y_i, \mu_l) = \sqrt{\sum_{h=1}^k (Y_{ih} - \mu_{lh})^2} \quad (2)$$

Algoritma *K-Means Clustering* dapat dijabarkan melalui pseudocode sebagai berikut:

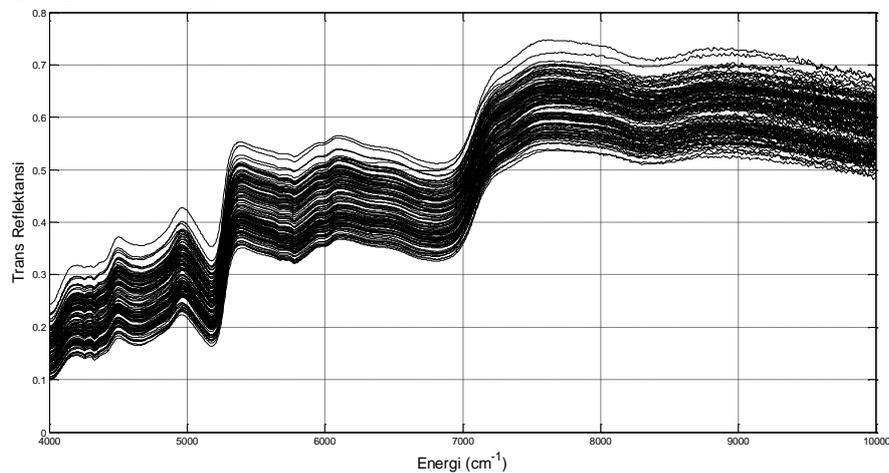
1. Inisialisasi  $Y_1, Y_2, Y_3, \dots, Y_n$ ;  $q$ ; dan  $\mu_1, \mu_2, \mu_3, \dots, \mu_q$ .
2. Ulangi langkah-langkah berikut ini hingga D pada (1) konvergen, yaitu:
  - 2.a. Tentukan tiap titik  $Y_i$  ke kelompok dengan jarak  $Y_i$  terhadap  $\mu_l$  adalah terdekat, sehingga dapat diperoleh  $C_1, C_2, C_3, \dots, C_q$ .
  - 2.b. Hitung D seperti pada (1).
  - 2.c. Tentukan titik pusat baru tiap kelompok yaitu

$$\mu_l = \frac{1}{|C_l|} \sum_{Y_i \in C_l} Y_i \quad (3)$$

dengan  $|C_l|$  menyatakan banyak titik di kelompok  $C_l$ .

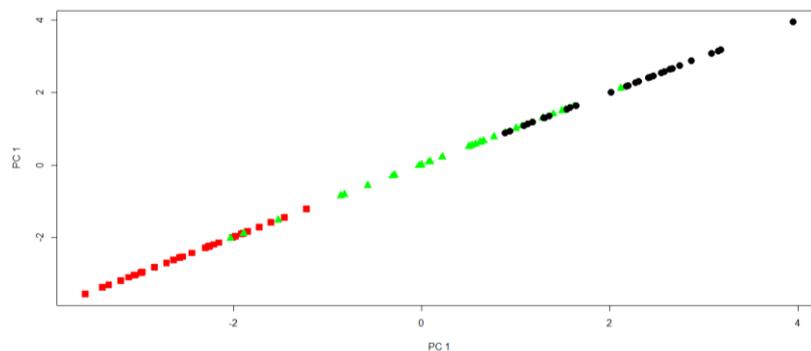
### III. PEMBAHASAN

Data spektrum dari 120 sampel teh yang terdiri dari jenis Fanning II (F-II), Pekoe Fanning (PFANN), dan Broken Orange Pekoe Fannings (BOPF) memiliki 1501 variabel pada setiap sampelnya yang dapat dilihat pada Gambar 1.

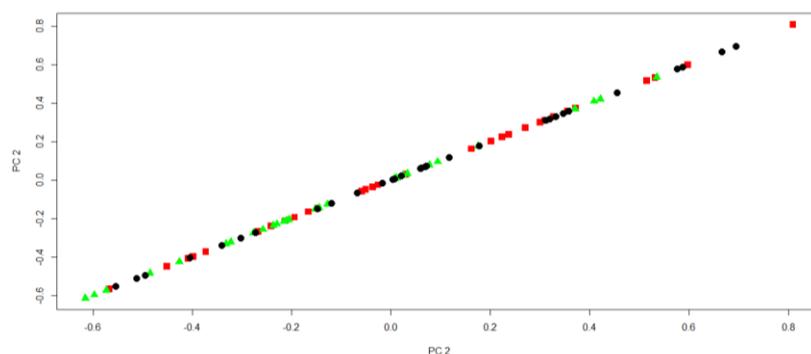


Gambar 1. Visualisasi data spektrum.

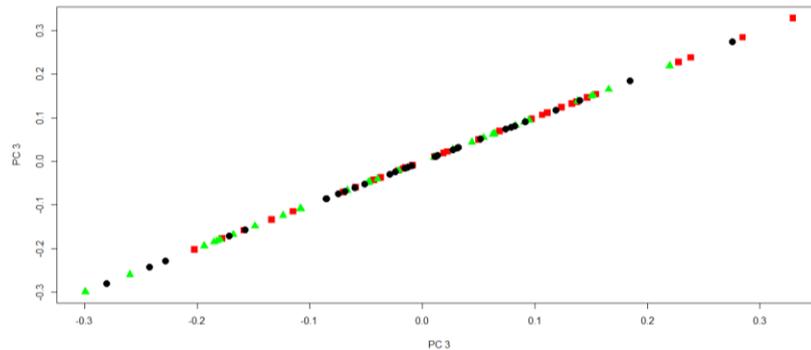
Ukuran matriks data spektrum yang besar akan direduksi dengan algoritma untuk perolehan *Principal Components* (PCs) sehingga didapati peubah-peubah baru yaitu PC ke-1, PC ke-2, dan PC ke-3. Persentase variansi dari nilai-nilai PC ke-1 adalah sebesar 96.61%, PC ke-2 sebesar 2.7%, dan PC ke-3 sebesar 0,42% yang memiliki arti bahwa presentasi tersebut dapat mewakili dari data keseluruhan.



Gambar 2.1. Scatterplot univariat PC ke-1.

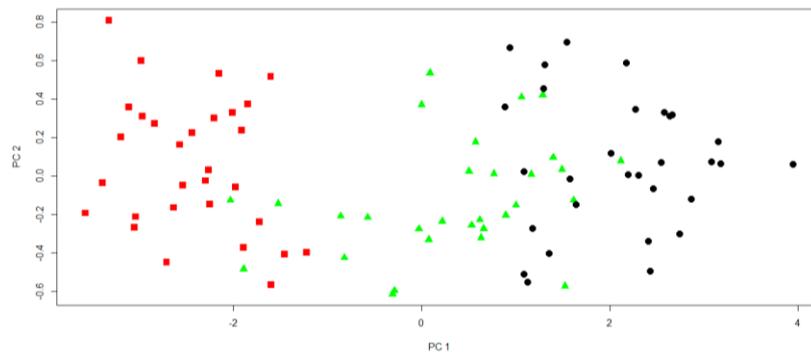


Gambar 2.2. Scatterplot univariat PC ke-2

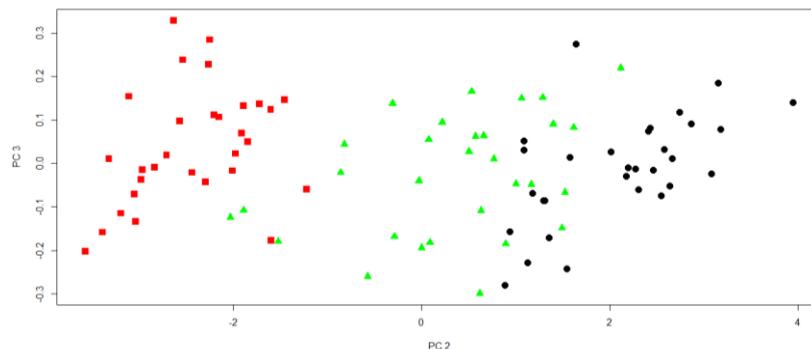


Gambar 2.3. Scatterplot univariat PC ke-3.

Gambar 2.1. menunjukkan bahwa nilai-nilai PC ke-1 dari sampel-sampel *grade* F-II nampak heterogen terhadap PFANN (merah) begitu juga halnya F-II (hijau) terhadap BOPF (hitam). Nilai-nilai PC ke-1 dari sampel-sampel *grade* PFANN juga nampak heterogen terhadap BOPF meskipun terdapat sampel PFANN terlihat terkumpul bersama BOPF. Sebaran nilai-nilai PC ke-2 dan PC ke-3 diberikan pada Gambar 2.2. dan Gambar 2.3. Nilai-nilai PC ke-2 dan PC ke-3 nampak homogen untuk semua sampel dari semua *grade* namun PC ke-2 penyebaran nilai-nilai lebih melebar dibandingkan PC ke-3 yang penyebaran nilai-nilainya lebih terpusat.



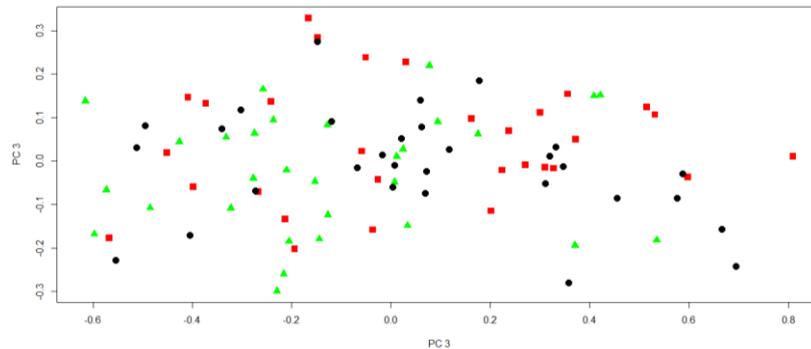
Gambar 3.1. Scatterplot bivariat PC ke-1 dan PC ke-2.



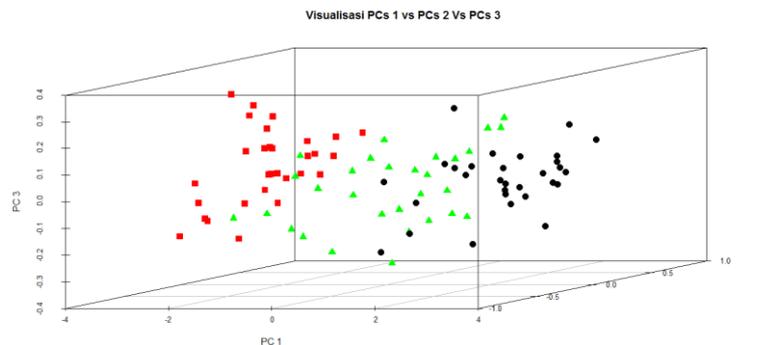
Gambar 3.2. Scatterplot bivariat PC ke-1 dan PC ke-3.

Gambar 3.1. dan 3.2. yang berurutan adalah data bivariat PC ke-1 dan ke-2 dengan persen-tase variansi sebesar  $96.61\% + 2.7\% = 99.31\%$  serta data bivariat PC ke-1 dan ke-3 yang mana memiliki persentase variansi sebesar  $96.61\% + 0.42\% = 97.03\%$  memperlihatkan sebaran nilai-nilai PCs dari sampel *grade* F-II, PFANN, dan BOPF yang heterogen. Berbeda hal pada Gambar 3.3. yang merupakan data bivariat PC ke-2 dan ke-3 dengan persentase variansi sebesar  $2.7\% + 0.42\% = 3.12\%$ , homogenitas muncul pada sebaran nilai-nilai PCs dari semua *grade*. Ter-kait dengan jumlah persentase variansi PCs, ini dimungkinkan bahwa

semakin besar jumlah persentase variansi PCs maka semakin pula dijumpai heterogenitas pada PCs terkait untuk *grade-grade* teh hitam.



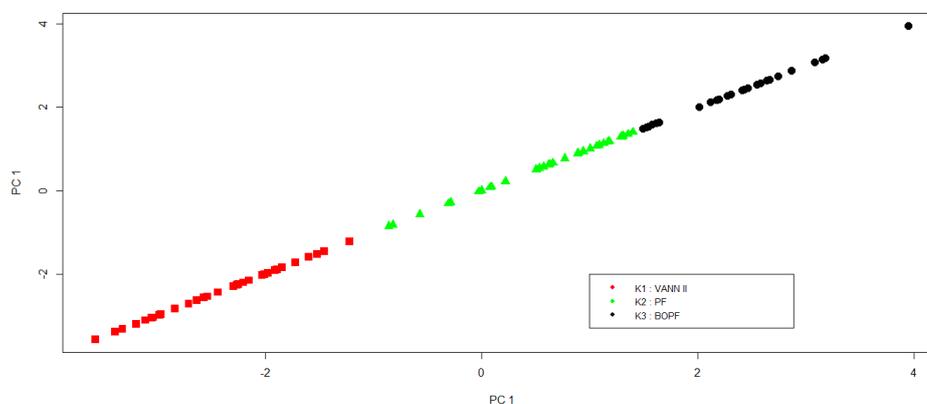
Gambar 3.3. Scatterplot bivariat PC ke-2 dan PC ke-3.



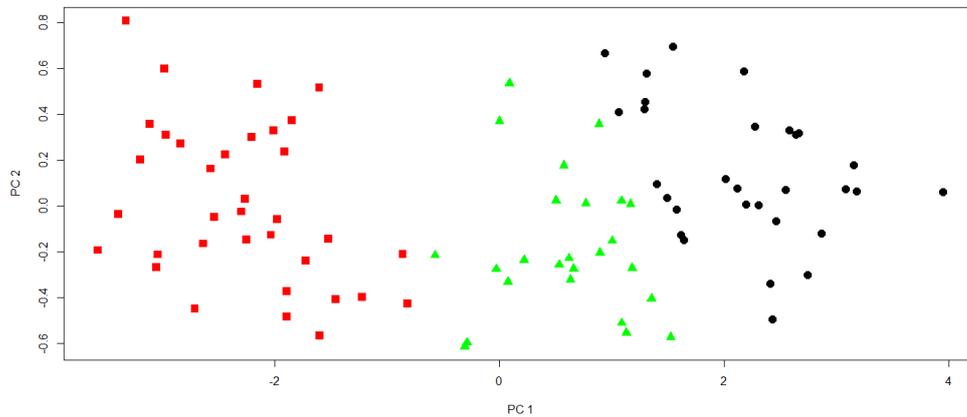
Gambar 4. Scatterplot Trivariat PC ke-1, ke-2, dan ke-3.

Gambar 4 menunjukkan data trivariat PC ke-1, ke-2, dan ke-3 dengan jumlah persentase variansi  $96.61\% + 2.7\% + 0.42\% = 99.73\%$  memperlihatkan sebaran nilai PCs dari *grade* F-II, PFANN, dan BOPF yang heterogen.

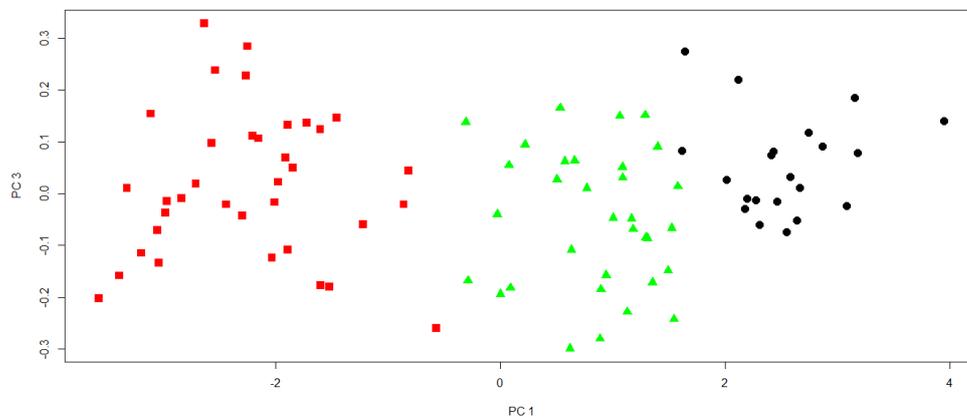
Kemudian analisis dilanjutkan dengan menggunakan K-Means Cluster Analysis berdasarkan univariat: PC ke-1, bivariat: PC ke-1 dan PC ke-2 serta PC ke-1 dan PC ke-3, dan trivariat karena terlihat heterogen. Hasil Pengelompokan dapat dilihat pada Gambar 5, Gambar 6, dan Gambar 7.



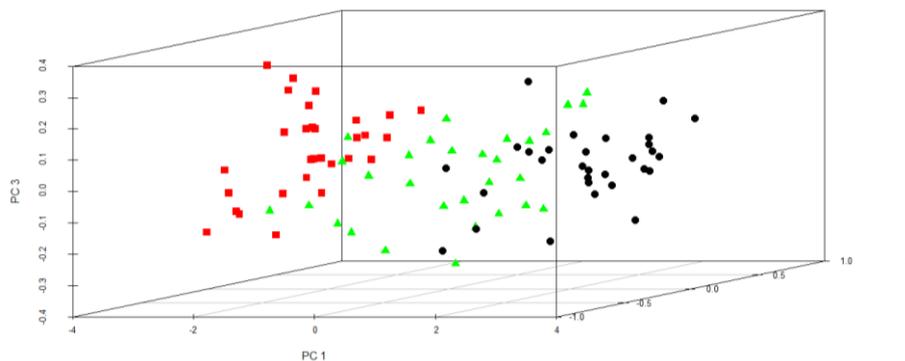
Gambar 5. Scatterplot Hasil Pengelompokan PC ke-1.



Gambar 6. Scatterplot Hasil Pengelompokan PC ke-1 dan PC ke-2



Gambar 7 Scatterplot Hasil Pengelompokan PC ke-1 dan PC ke-3.



Gambar 8. Scatterplot Hasil Pengelompokan PC ke-1, PC ke-2 dan PC ke-3.

Untuk mengetahui tingkat keakuratan hasil pengelompokan *K-Means Clustering* maka akan dibandingkan dengan sampel teh yang diketahui. Presentase keakuratan dapat dilihat pada Tabel 1. Sedangkan untuk mengetahui tingkat keakuratan hasil pengelompokan *K-Means Clustering* dengan data validasi dengan cara menghitung jarak terpendek antara tiap pusat kluster dengan data validasi yang ditunjukkan pada Tabel 2.

Didapati metode alternatif untuk pengukuran/penilaian *grade-grade* teh hitam melalui *Principal Component Analysis (PCA)* dan *K-Means Clustering*. Data yang direduksi menjadi 3 Principal Components (PCs) yang diperoleh melalui algoritma perolehan PCs kemudian dianalisis secara univariat, bivariat, serta trivariat.

Tabel 1. Persentase Keakuratan Hasil *K-Means Clustering*.

Kluster	Grade	Keakuratan (%)			
		PC ke-1	PC ke-1 dan PC ke-2	PC ke-1 dan PC ke-3	PC ke-1, ke-2, dan ke-3
1	F-II	100%	100%	100%	100%
2	PFANN	76,67%	63,33%	73,33%	60%
3	BOPF	70%	80%	63,33%	80%

 Tabel 2. Persentase Keakuratan Hasil *K-Means Clustering* pada Data Validasi.

Kluster	Grade	Keakuratan (%)			
		PC ke-1	PC ke-1 dan PC ke-2	PC ke-1 dan PC ke-3	PC ke-1, ke-2, dan ke-3
1	F-II	100%	90%	100%	100%
2	PFANN	90%	80%	90%	90%
3	BOPF	70%	60%	70%	70%

Sementara itu, nilai persentase variansi PCs juga mempengaruhi heterogenitas nilai-nilai PCs. Semakin besar persentase variansi suatu PC (univariat), maka nilai-nilai PC *grade-grade* teh hitam semakin heterogen. Untuk kasus bivariat, semakin besar jumlah presentase variansi PCs, maka nilai-nilai PCs juga semakin heterogen. Untuk kasus trivariat, karena yang dibahas adalah 3 PCs pertama yang mana jumlah persentase variansi 3 PCs itu juga yang terbesar, maka sifat heterogen diperoleh. Heterogenitas itu dijumpai untuk nilai-nilai PCs dari *grade* F-II, PFANN, dan BOPF. Validasi juga dilakukan pada sampel data yang khusus digunakan untuk validasi metode dan analisis yang diperoleh. Titik pusat kelompok yang merupakan hasil *K-Means Clustering* pada data training digunakan sebagai acuan nilai-nilai PCs dari data validasi untuk ditentukan guna masuk ke dalam kelompok-kelompok berdasarkan jarak minimumnya terhadap tiap titik pusat itu. Didapati kembali bahwa nilai-nilai PCs dari data validasi *grade* F-II terkelompok dg baik dari *grade-grade* yang lain. Sementara itu, dijumpai nilai-nilai PCs dari data validasi *grade* PFANN dan BOPF masih *misplaced* atau ada nilai-nilai PCS yang homogen di interval atau daerah nilai-nilai tertentu.

#### IV. KESIMPULAN

Didapati metode alternatif untuk pengukuran/penilaian *grade-grade* teh hitam melalui *Principal Component Analysis* (PCA) dan *K-Means Clustering*. Data yang direduksi menjadi 3 Principal Components (PCs) yang diperoleh melalui algoritma perolehan PCs dengan persentase variansi dari nilai-nilai PC ke-1 adalah sebesar 96.61%, PC ke-2 sebesar 2.7%, dan PC ke-3 sebesar 0,42% kemudian dianalisis secara univariat, bivariat, serta trivariat. Analisis secara univariat dapat diketahui nilai-nilai PC ke-1 dari *grade* F-II, PFANN, dan BOPF nampak heterogen sedangkan PC 2 dan PC 3 terlihat homogen. Secara bivariat, nilai-nilai PC ke-1 dan ke-2 serta PC ke-1 dan ke-3 memperlihatkan sebaran nilai-nilai PCs dari *grade* F-II, PFANN, dan BOPF yang heterogen. Secara trivariat, nilai nilai PC ke-1, ke-2, dan ke-3 memperlihatkan sebaran nilai-nilai PCs dari *grade* F-II, PFANN, dan BOPF yang heterogen. Analisis dilanjutkan dengan menggunakan *K-Means Cluster Analysis* berdasarkan univariat: PC ke-1, bivariat: PC ke-1 dan PC ke-2 serta PC ke-1 dan PC ke-3, dan trivariat karena terlihat heterogen. Persentase keakuratan hasil *K-Means Clustering* yang didapat untuk univariat: PC ke-1, bivariat: PC ke-1 dan PC ke-2 serta PC ke-1 dan PC ke-3, dan trivariat secara berturut-turut untuk kluster 1 (F-II) sebesar 100% secara univariat, bivariat, dan trivariat; kluster II

(PFANN) sebesar 76,67%, 63,33%, 73,33%, dan 60%; serta kluster 3 (BOPF) sebesar 70%, 80%, 63,33%, dan 80%. Sedangkan persentase keakuratan hasil *K-Means Clustering* pada data validasi secara berturut-turut untuk kluster 1 sebesar 100%, 90%, 100%, dan 100%; kluster 2 sebesar 90%, 80%, 90%, dan 90%; serta kluster 3 sebesar 70%, 60%, 70%, dan 70%. Metode alternatif untuk pengukuran/penilaian *grade-grade* teh hitam melalui *Principal Component Analysis* (PCA) dan *K-Means Clustering* yang dilakukan dalam makalah ini cukup efektif dan efisien karena memiliki tingkat keakuratan diatas 50%.

## REFERENSI

- [1] Pusdatin, "Outlook Teh Komoditas Pertanian Subsektor Perkebunan," 27 Desember 2016. [Online]. Available: <http://epublikasi.setjen.pertanian.go.id/arsip-outlook/75-outlook-perkebunan/417-outlook-teh-2016>. [Diakses: 3 April 2020].
- [2] D. A. Wibowo, "Prosedur Pengujian dan Sertifikasi Mutu Teh untuk Ekspor pada Balai Pengujian dan Sertifikasi Mutu Barang (BPSMB) di Surakarta," 25 July 2013. [Online]. Available: <https://eprints.uns.ac.id/9557/>. [Diakses: 12 November 2019].
- [3] R. O. Anindya, J. Muninggar dan F. S. Rondonuwu, "Indonesian Black Tea Classification Using Fourier-Transform Near-Infrared Spectroscopy and a Principal Component Analysis," in *International Conference on Mathematics, Science, and Education*, Malang, 2018.
- [4] O. Ye Rodionova, A.V. Titova, K.S. Balyklova, A.L. Pomerantsev, "Detection Of Counterfeit And Substandard Tablets Using Non-Invasive NIR And Chemometrics - A Conceptual Framework For A Big Screening System," *Talanta*, vol. 205, 2019, 120150.
- [5] S. Samadia, S. Wajizaha, dan A. A. Munawar, "Rapid and Simultaneous Determination of Feed Nutritive Values by Means of Near Infrared Spectroscopy," *Tropical Animal Science Journal*, vol. 41, no. 2, pp. 121-127, 2018.
- [6] Y. Ozaki, C. Huck, dan K. Bec, "Near-IR Spectroscopy and Its Applications," *Molecular and Laser Spectroscopy*, Elsevier, 2017, pp. 11-38.
- [7] M. E. N. M. Ruah, N. F. Rasaruddin, S. S. Fong, M. Z. Jaafar, "Data Pre-processing Methods of FT-NIR Spectral Data for the Classification Cooking Oil," *American Institute of Physics Conference Series*, Vol. 1635, no. 1, pp. 890-897, 2014.
- [8] F. Ismawan, "Hasil Ekstraksi Algoritma Principal Component Analysis (PCA) untuk Pengenalan Wajah dengan Bahasa Pemrograman Java Eclipse IDE," *Jurnal Sisfotek Global*, vol. 5, no. 1, 2015.
- [9] S. P. Mishra, et al, "Multivariate Statistical Data Analysis Principal Component Analysis (PCA)," *International Journal of Livestock Research*, vol. 7, no. 5, 2017.
- [10] M. Z. Nasution, "Penerapan Principal Component Analysis (PCA) Dalam Penentuan Faktor Dominan Yang Mempengaruhi Prestasi Belajar Siswa," *Jurnal Teknologi Informasi*, vol. 3, no.1, 2019.
- [11] J. Ma, Y. Yuan, "Dimension Reduction Of Image Deep Feature Using PCA," *Journal of Visual Communication and Image Representation*, vol. 63, August 2019, 102578.
- [12] F. M. Kusuma, dan A. Wibowo, "Principal Component Analysis (PCA) Untuk Mengatasi Multikolinieritas Terhadap Faktor Angka Kejadian Pneumonia Balita Di Jawa Timur Tahun 2014," *Jurnal Biometrika dan Kependudukan*, vol. 6, no. 2, 2017.
- [13] Husaini, Huzaeni, dan Fahmi, "Algoritma Principal Componen Analysis Dalam Pemrosesan Sinyal Elektrokardiogram," *Jurnal Infomedia*, vol. 3, no. 1, 2018.
- [14] D. R. Hidalgo, B. B. Corte's, dan E. C. Bravo, "Dimensionality Reduction Of

- Hyperspectral Images Of Vegetation And Crops Based On Self-Organized Maps," *Information Processing in Agriculture*, in press. Available: <https://www.sciencedirect.com/science/article/pii/S2-21431732030189X>.
- [15] S. Ayesha, M. K. Hanif, dan R. Talib, "Overview And Comparative Study Of Dimensionality Reduction Techniques For High Dimensional Data," *Information Fusion*, vol. 59, pp. 44-58, 2020.
- [16] A. Ahmad, dan S. S. Khan, "InitKmix-A Novel Initial Partition Generation Algorithm For Clustering Mixed Data Using K-Means-Based Clustering," Juli 2020. [Online]. Available: <https://doi.org/10.13140/RG.2.2.21979.62244>. [Diakses: 9 November 2020].
- [17] A. Kaur, S. K. Pal, dan A. P. Singh, "Hybridization Of Chaos And Flower Pollination Algorithm Over K-Means For Data Clustering," *Applied Soft Computing Journal*, 2019, 105523
- [18] S. Tayfura, N. Alvera, S. Abdib, S. Saatçı, A. Ghiamid, "Characterization of concrete matrix/steel fiber de-bonding in an SFRC beam: Principal component analysis and k-mean algorithm for Clustering AE data," *Engineering Fracture Mechanics*, vol. 194, pp. 73-85, 2018.
- [19] O. R. Kristyaningrum, A. Setiawan, dan L. R. Sasongko, "Analisis Pengelompokan K-Means untuk Data Bivariat Laju Kunjungan dan Rasio Rujukan," *Jurnal Riset dan Aplikasi Matematika*, vol. 2, no. 1, 2018.
- [20] J. O. Ong, "Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, 2013.