

Sistem Informasi Pengklasifikasian Hasil Produksi Teh Menggunakan Metode K-Means

Umi Hayati¹, Arsyia Mutia Karuniani²
STMIK "AMIK BANDUNG"

umi.haya41@gmail.com¹, arsyia.mutia@gmail.com²

Intisari - Sistem informasi untuk pengklasifikasian dibutuhkan oleh sebuah perusahaan yang memproduksi suatu barang. Tujuan penggunaan sistem informasi pengklasifikasian adalah untuk memudahkan didapaknya hasil produksi secara tepat. Permasalahan yang dihadapi secara spesifik yaitu ketika bahan baku datang, kemudian ditimbang, bahan baku tersebut tidak dapat dihitung secara tepat berapa jumlah bubuk teh yang dihasilkan sesuai dengan standar kualitas PTPN VIII, sehingga hanya dapat dikira-kira saja. Untuk itu dirancang sebuah sistem informasi, di mana sistem ini dapat menentukan jumlah bubuk teh yang dihasilkan oleh PTPN VIII. Sistem informasi yang dimaksud adalah sistem informasi yang menggunakan metode *K-Means*. Cara kerja *K-Means* adalah memperhitungkan sejumlah data yang didapat dari penelitian. Data tersebut diolah dengan cara mengklusterkan data awal menjadi 2 klaster. Percobaan dilakukan dengan mengolah 90 data. Hasil pengolahan dihasilkan 2 klaster, yaitu klaster 1 dengan kriteria berat daun teh basah rata-rata dikisaran 307.867 ton s.d. 658.761 ton, setelah diolah menggunakan metode *K-Means* menghasilkan bubuk teh seberat rata-rata 277.080 ton s.d. 592.885 ton dengan menggunakan suhu 100°C s.d. 250°C menghasilkan kepekatan warna yang rendah. Klaster 2 dengan kriteria berat daun teh basah rata-rata 180.876 ton s.d. 246.246 ton, setelah diolah menggunakan metode *K-Means* menghasilkan bubuk teh seberat rata-rata 162.788 ton s.d. 221.621 ton dengan suhu 250°C s.d. 350°C dan menghasilkan teh yang lebih pekat. Dengan demikian dapat disimpulkan bahwa penggunaan sistem informasi untuk pengklasifikasian hasil produksi teh dengan menggunakan metode *K-Means* sangat baik.

Kata Kunci : *K-Means, klaster, Kuantitas, berat teh basah, berat bubuk teh, suhu, warna.*

Abstrack -Information systems for classification are needed by a company that produces an item. The purpose of using a classification information system is to make it easier to obtain accurate production results. The specific problem faced is that when the raw materials arrive, then weigh them, the raw materials cannot be calculated exactly how much tea powder is produced in accordance with the quality standards of PTPN VIII, so it can only be estimated. For this reason, an information system was designed, in which this system could determine the amount of tea powder produced by PTPN VIII. The information system in question is an information system that uses the K-Means method. The way K-Means works is to take into account a number of data obtained from research. The data is processed by clustering the initial data into 2 clusters. The experiment was carried out by processing 90 data. The results of the processing resulted in 2 clusters, namely cluster 1 with the criteria for the average weight of wet tea leaves in the range of 307,867 tons up to 658,761 tons, after being processed using the K-Means method, it produces tea powder weighing an average of 277,080 tons to up to 277,080 tons. 592,885 tonnes using a temperature of 100°C s.d. 250°C results in low color density. Cluster 2 with criteria for the average weight of wet tea leaves is 180,876 tons to s.d. 246,246 tons, after being processed using the K-Means method, it produces tea powder weighing an average of 162,788 tons up to. 221,621 tons with a temperature of 250°C to. 350°C and produces a more concentrated tea. Thus, it can be concluded that the use of information systems for the classification of tea production using the K-Means method is very good.

Keywords: K-Means, cluster, quantity, weight of wet tea, weight of tea powder, temperature, color.

I. PENDAHULUAN

Keberhasilan perusahaan dalam menghasilkan produk tidak hanya terlihat dari segi kualitas SDM, sarana dan prasarana yang tersedia, tetapi juga dari teknologi yang digunakan, karena saat ini teknologipun berperan penting dalam keberhasilan sebuah perusahaan. PT Perkebunan Nusantara VIII, disingkat PTPN VIII, adalah Badan Usaha Milik Negara yang bergerak di bidang perkebunan teh, karet, kina, kakao, dan kelapa. sawit. Kantor pusat perusahaan berada di Bandung dengan wilayah operasi di Jawa Barat, sedangkan untuk perkebunan tehnya terletak di desa Ciater Jalan Cagak, Kabupaten Subang, Jawa Barat dengan ketinggian ± 1.100 di atas permukaan laut. Tanaman perkebunan merupakan salah satu komoditas yang bisa diandalkan sebagai sentra bisnis yang menggiurkan. Terlebih produk cukup ramai permintaannya, baik di pasar dalam negeri maupun luar negeri. Selain itu harga jual yang tinggi membuat perkebunan menjadi penyumbang devisa negara yang cukup besar. Saat ini sudah ada banyak komoditas jenis perkebunan yang cukup potensial, salah satunya adalah teh, [1].

PTPN VIII saat ini masih belum menerapkan penggunaan teknologi untuk mengolah dan menyimpan data yang sifatnya permanen dan penting. Sehingga permasalahan yang ditemui adalah ketika bahan baku datang, kemudian ditimbang dan dijumlahkan, bahan baku tersebut tidak dapat dihitung secara tepat berapa jumlah bubuk teh yang dihasilkan sesuai dengan standar kualitas yg ditetapkan, sehingga hanya dapat dilakukan secara perkiraan saja. Untuk itu diperlukan pengelolaan yang tepat, hal ini dapat dilakukan salah satunya melalui perancangan sistem informasi khususnya pengelolaan pengklasifikasian hasil teh dengan memanfaatkan metoda *K-Means*.

II. LANDASAN TEORI

A. Sistem

Sistem adalah suatu kumpulan objek atau unsur-unsur atau bagian-bagian yang memiliki arti berbeda-beda yang saling memiliki hubungan, saling bekerja sama dan saling memengaruhi satu sama lain serta memiliki keterikatan pada rencana yang sama dalam suatu tujuan tertentu pada lingkungan yang kompleks [2].

Manfaat sistem yaitu untuk menyatukan /mengintegrasikan semua unsur yang ada dalam suatu ruang lingkup, di mana komponen-komponen tersebut tidak dapat berdiri sendiri. Komponen atau subsistem harus saling berintegrasi dan saling berhubungan untuk membentuk satu kesatuan sehingga sasaran dan tujuan sistem tersebut dapat tercapai [2].

B. Informasi

Informasi adalah sekumpulan fakta-fakta yang telah diolah menjadi bentuk data, sehingga dapat menjadi lebih berguna dan dapat digunakan oleh siapa saja yang membutuhkan data-data tersebut sebagai pengetahuan ataupun dapat digunakan dalam pengambilan keputusan. Informasi bisa dikatakan sebagai pengetahuan

yang didapatkan dari belajar, pengalaman atau instruksi. Dalam ilmu komputer, informasi adalah data yang disimpan, diproses atau ditransmisikan. Para ahli meneliti konsep informasi tersebut sebagai pengetahuan yang didapatkan dari pembelajaran, pengalaman maupun instruksi [3].

C. Sistem Informasi

Sistem Informasi adalah suatu sistem yang terdiri dari kumpulan komponen sistem, yaitu software, hardware dan brainware yang memproses informasi menjadi sebuah output yang berguna untuk mencapai suatu tujuan tertentu dalam suatu organisasi [2].

D. K-Means

K-Means merupakan salah satu algoritma *clustering*. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma, komputer mengelompokkan sendiri data-data yang menjadi masukan tanpa mengetahui terlebih dulu target kelasnya. Algoritma ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut [4].

Algoritma untuk melakukan *K-Means Clustering* adalah:

1. Pilih k buah titik *centroid* secara acak
2. Kelompokkan data sehingga terbentuk k buah *cluster* dengan titik *centroid* dari setiap *cluster* merupakan titik *centroid* yang telah dipilih sebelumnya
3. Perbaharui nilai titik *centroid*
4. Ulangi langkah 2 dan 3 sampai nilai dari titik *centroid* tidak lagi berubah [5].

Proses pengelompokkan data ke dalam suatu *cluster* dapat dilakukan dengan cara menghitung jarak terdekat dari suatu data ke sebuah titik *centroid*. Pembaharuan suatu titik *centroid* dapat dilakukan dengan rumus berikut :

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad \dots\dots (I)$$

Di mana:

μ_k = titik *centroid* dari cluster ke-K

N_k = banyaknya data pada cluster ke-K

x_q = data ke-q pada cluster ke-K

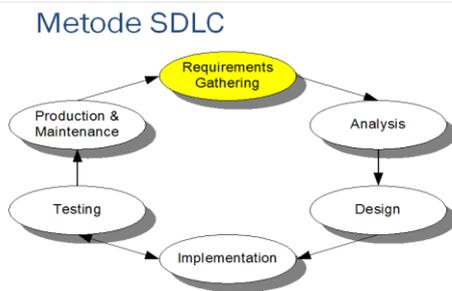
III. METODOLOGI PENELITIAN

Untuk mendukung penelitian, digunakan algoritma *K-Means*. Tujuan yang diharapkan adalah kemudahan dalam menentukan berapa jumlah hasil daun teh basah menjadi bubuk teh jadi, dengan batasan:

- a. Input: Jumlah data diambil selama 3 bulan dengan memiliki data sebanyak 90 data;
- b. Memiliki 4 variabel, berupa berat daun teh basah, berat bubuk teh jadi, suhu, dan kepekatan warna.

- c. Proses: Pemograman menggunakan bahasa *PHP* (*Hypertext Preprocessor*);
- d. Output: Hasil jumlah bubuk teh dengan kepekatan rendah dan bubuk teh jadi dengan kepekatan tinggi.

Dalam mengembangkan sistem yang akan dirancang, digunakan metode *SDLC*. Metode *SDLC* (System Development Life Cycle) atau Siklus Hidup Pengembangan Sistem adalah proses pembuatan dan perubahan sistem serta model dan metodologi yang digunakan untuk mengembangkan sistem tersebut. Berikut tahapan metode *SDLC*: [6]



Gambar 1 Metode SDLC

- a. Perencanaan Sistem, lebih menekankan pada aspek studi kelayakan pengembangan sistem (feasibility study). Penentuan prioritas teknologi dan pemilihan aplikasi.
- b. Analisis Sistem, tahap di mana dilakukan beberapa aktivitas dari melakukan studi literatur sampai mendefinisikan kebutuhan sistem.
- c. Perancangan Sistem, tahap ini, features dan operasi-operasi pada sistem dideskripsikan secara detail, dari Menganalisa sampai merancang user interface.
- d. Implementasi Sistem, mengimplementasikan rancangan dari tahap-tahap sebelumnya dan melakukan uji coba.
- e. Pemeliharaan Sistem, dilakukan oleh admin yang ditunjuk untuk menjaga sistem tetap mampu beroperasi secara benar melalui kemampuan sistem dalam mengadaptasikan diri sesuai dengan kebutuhan.

Untuk memperkuat analisis, dilakukan pendekatan dalam usaha untuk memecahkan permasalahan yang ada dengan menggunakan beberapa metode Observasi, Wawancara dan Studi Literatur.

II. HASIL PENELITIAN

A. Proses Data Mining

Tahapan yang dilakukan untuk melakukan *K-means* dalam proses Data Mining: [7]

1. *Pembersihan Data* (untuk membuang data yang tidak konsisten dan noise). Dalam tahap ini dilakukan seleksi data, membuang data-data yang memiliki nilai yang sama. Hal ini dilakukan untuk mengambil sample dari data yang diperoleh,
2. *Integrasi data* (penggabungan data dari beberapa sumber). Dalam tahap ini dilakukan penggabungan data. Data yang digabungkan adalah data hasil kebun (daun teh basah), data hasil pabrik (bubuk teh jadi), suhu, dan juga

kepekatan warna untuk dapat melakukan sampling data. Adapun data yang sudah digabungkan adalah:

Tabel 1. Data Sampling

TANGGAL	SELISIH			
	KEBUN (TON)	PABRIK (TON)	SUHU (Celcius)	WARNA (KEPEKATAN)
01/09/17	202.833	182.550	2	9
02/09/17	201.681	181.513	4	6
03/09/17	307.867	277.080	8	3
04/09/17	220.285	198.257	2	3
05/09/17	108.967	98.070	4	6
06/09/17	309.867	278.880	8	9
07/09/17	658.761	592.885	2	9
08/09/17	216.810	195.129	4	6
09/09/17	285.635	257.072	8	3
10/09/17	324.546	292.091	2	3
11/09/17	180.876	162.788	4	6
12/09/17	180.011	162.010	8	9
13/09/17	180.876	162.788	2	9
14/09/17	202.833	182.550	4	6

3. *Transformasi data* (data diubah menjadi bentuk yang sesuai untuk di *mining*). Pada tahap ini dilakukan tahap awal untuk klastering. Merubah data awal menjadi data yang sesuai untuk data mining maka dibuat *centroid* awal dan beberapa perhitungan yang sesuai dengan rumus data mining. Ketika menentukan *centroid* awal dilakukan secara acak. Karena memiliki 4 variabel dan mempunyai 3 proses disetiap variabel maka ketika menentukan *centroid* pun harus memiliki 12 proses. proses 1 (untuk hasil daun teh basah) : CK1, CK2, CK3 | proses 2 (untuk hasil bubuk teh): CU1, CU2, CU3 | proses 3 (untuk penentuan suhu) | proses 4 (untuk perhitungan warna): CW1, CW2, CW3) yang berbeda karena setiap 1 proses memiliki 3 *centroid* yang berbeda. Ketika sudah menentukan *centroid* maka harus ada perhitungannya kembali. Dengan rumus:
 - a. Membuat perhitungan dan dinamai dengan JC1, JC2 dan JC3 (ini untuk menghitung hasil kebun yang dihitung dengan *centroid*)
 - b. Membuat perhitungan dan dinamai dengan JU1, JU2, dan JU3 (ini untuk menghitung hasil pabrik yang dihitung dengan *centroid*)
 - c. Membuat perhitungan dan dinamai dengan JS1, JS2, dan JS3 (ini untuk menghitung suhu dengan *centroid*)
 - d. Membuat perhitungan dan dinamai dengan JW1, JW2, dan JW3 (ini untuk menghitung kepekatan warna)
 - e. Jika tahap di atas sudah dilakukan maka selanjutnya hitung untuk penentuan cluster dengan table C1, C2, C3, dan C4
 - f. Jika sudah ada hasilnya, lalu tentukan jumlah minimalnya dengan rumus MIN (C1:C4)
 - g. Tahap terakhir menemukan jumlah data pada klaster, dengan rumus Jika (jumlah MIN = jumlah MIN, 1,0)

Dari tahapan tersebut maka hasil yang didapat adalah:

Tabel 2. Data awal yang sudah ada klasternya

TANGGAL	SELISIH				KLASTER		
	KEBUN	PABRIK	SUHU	WARNA	K1	K2	K3
01/09/17	202.833	182.550	2	9	0	1	0
02/09/17	201.681	181.513	4	6	0	1	0
03/09/17	307.867	277.080	8	3	0	0	1
04/09/17	220.285	198.257	2	3	0	1	0
05/09/17	108.967	98.070	4	6	1	0	0
06/09/17	309.867	278.880	8	9	0	0	1
07/09/17	658.761	592.885	2	9	0	0	1
08/09/17	216.810	195.129	4	6	0	1	0
09/09/17	285.635	257.072	8	3	0	0	1
10/09/17	324.546	292.091	2	3	0	0	1
11/09/17	180.876	162.788	4	6	1	0	0
12/09/17	180.011	162.010	8	9	1	0	0
13/09/17	180.876	162.788	2	9	1	0	0
14/09/17	202.833	182.550	4	6	0	1	0
				JUMLAH	4	5	5

KETERANGAN :	
Suhu 2 => 100°C	Warna 3 => Kepekatan Rendah
Suhu 4 => 200°C	Warna 6 => Kepekatan Sedang
Suhu 5 => 300°C	Warna 9 => Kepekatan Tinggi

4. Aplikasi teknik Data Mining dan proses ekstraksi pola dari data yang ada,

Pada tahap ini dilakukan tahap iterasi 1, dengan menentukan titik *centroidnya* dimana sudah dapat dihitung berdasarkan klaster awal yang sudah ada. Cara menghitung untuk menentukan centroid baru melihat dari jumlah data sesuai pada klasternya masing-masing. Ada pun rumusnya sebagai berikut :

Centroid hasil kebun (CK1) = $(108967 + 180876 + 180011 + 180876) / 4$

Centroid hasil kebun (CK1) = $(202833 + 201681 + 220285 + 216810 + 202833) / 5$

Centroid hasil kebun (CK1) = $(307867 + 309867 + 658761 + 285635 + 324546) / 5$

Centroid hasil pabrik (CK1) = $(98070 + 162788 + 162010 + 162788) / 4$

Centroid hasil pabrik (CK1) = $(182550 + 181513 + 198257 + 195129 + 162788) / 5$

Centroid hasil pabrik (CK1) = $(277080 + 278880 + 592885 + 257072 + 292091) / 5$

Centroid hasil suhu (CK1) = $(4+4+8+2)/4$

Centroid hasil suhu (CK1) = $(2+4+2+4+4)/5$

Centroid hasil suhu (CK1) = $(8+8+2+8+2)/5$

Centroid hasil warna (CK1) = $(6+6+9+9)/4$

Centroid hasil warna (CK1) = $(9+6+3+6+6)/5$

Centroid hasil warna (CK1) = $(3+9+9+3+3)/5$

Jika telah diketahui semua jumlah centroidnya maka hitung tahapan sesuai rumus yang sudah ada ditahap awal. Ulangi tahap ini sehingga menemukan posisi klaster yang tidak berubah.

5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan). Pada tahap ini menemukan titik posisi klaster yang tidak berubah. Posisi klaster yang tidak berubah terjadi pada iterasi ke 7.

Perubahan posisi klaster pada iterasi 1 dan 7 sebagai berikut:

Perubahan posisi pada Iterasi Kesatu

Tabel 3. Posisi Klaster pada Iterasi Satu

TANGGAL	SELISIH				KLUSTER		
	KEBUN	PABRIK	SUHU	WARNA	K1	K2	K3
01/09/17	202833	182550	2	9	0	1	0
02/09/17	201681	181513	4	6	0	1	0
03/09/17	307867	277080	8	3	0	0	1
04/09/17	220285	198257	2	3	0	1	0
05/09/17	108.967	98070	4	6	1	0	0
06/09/17	309.867	278880	8	9	0	0	1
07/09/17	658.761	592885	2	9	0	0	1
08/09/17	216.810	195129	4	6	0	1	0
09/09/17	285.635	257072	8	3	0	1	0
10/09/17	324.546	292091	2	3	0	0	1
11/09/17	180.876	162788	4	6	1	0	0
12/09/17	180.011	162010	8	9	1	0	0
13/09/17	180.876	162788	2	9	1	0	0
14/09/17	202.833	182550	4	6	0	1	0
				JUMLAH	4	6	4

Perubahan posisi pada Iterasi Ketujuh:

Tabel 4. Posisi Klaster pada Iterasi ke Tujuh

TANGGAL	SELISIH				KLASTER		
	KEBUN (TON)	PABRIK	SUHU	Warna	K1	K2	K3
01/09/17	202833	182550	2	9	0	1	0
02/09/17	201681	181513	4	6	0	1	0
03/09/17	307867	277080	8	3	1	0	0
04/09/17	220285	198257	2	3	0	1	0
05/09/17	108967	98070	4	6	0	1	0
06/09/17	309867	278880	8	9	1	0	0
07/09/17	658761	592885	2	9	1	0	0
08/09/17	216810	195129	4	6	0	1	0
09/09/17	285635	257072	8	3	1	0	0
10/09/17	324546	292091	2	3	1	0	0
11/09/17	180876	162788	4	6	0	1	0
12/09/17	180011	162010	8	9	0	1	0
13/09/17	180876	162788	2	9	0	1	0
14/09/17	202833	182550	4	6	0	1	0
				JUMLAH	5	9	0

6. Presentasi pengetahuan (dengan teknik visualisasi), melakukan percobaan pada sampling 14 data maka terdapat 2 klaster dengan jumlah data pada klaster 1 terdapat 5 data dan klaster 2 terdapat 9 data dengan perbandingan sebagai berikut:

Tabel 5. Hasil Iterasi pada Data Sampling Klaster 1

Hasil Kebun	Hasil Pabrik	Suhu	Warna
307867	277080	8	3
309867	278880	8	9
658761	592885	2	9
285635	257072	8	3
324546	292091	2	3

Tabel 6. Hasil Iterasi pada Data Sampling Kluster 2

Hasil Kebun	Hasil Pabrik	Suhu	Warna
202833	182550	2	9
201681	181513	4	6
220285	198257	2	3
108967	98070	4	6
216810	195129	4	6
180876	162788	4	6
180011	162010	8	9
180876	162788	2	9
202833	182550	4	6

Setelah melakukan implementasi terhadap 90 data maka dapat terlihat perbedaannya sebagai berikut :

Tabel 7. Hasil pada Kluster 1

KLASTER 1			
Berat Kebun	Berat Pabrik	Suhu	Warna
307.867 s.d. 658.761	277.080 s.d. 592.884	100°C s.d. 150°C	Rendah

Tabel 8. Hasil pada Kluster 2

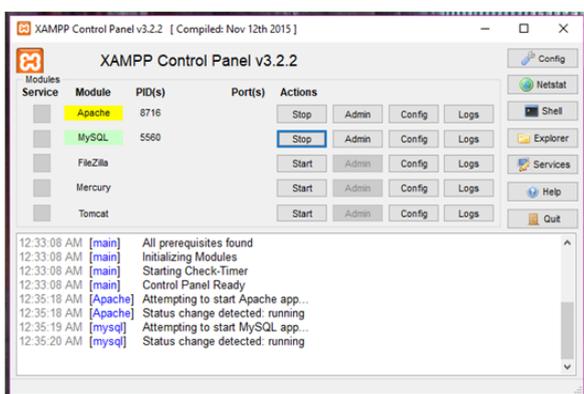
KLASTER 1			
Berat Kebun	Berat Pabrik	Suhu	Warna
108.967 s.d. 277.080	98.070 s.d. 249.372	200°C s.d. 250°C	Pekat

B. VALIDASI HASIL

Hasil penelitian yang sudah dilakukan selanjutnya diimplementasikan kepada sistem. Berikut langkah- langkah dan hasil validasi atau hasil program yang dibuat:

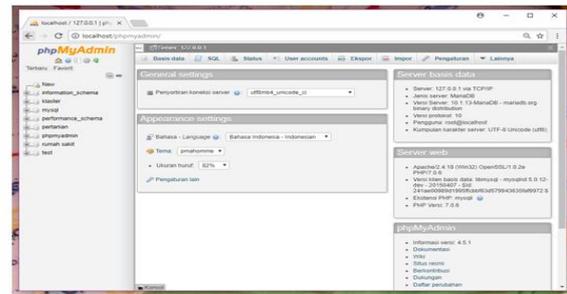
1. Tahap Pertama.

- a. Pada tahap ini dilakukan penginstalan *xampp* agar membantu dalam tahap pembuatan aplikasi. *Merunning apache dan mysql*. Contoh tampilannya dapat dilihat setelah di *Runing*:



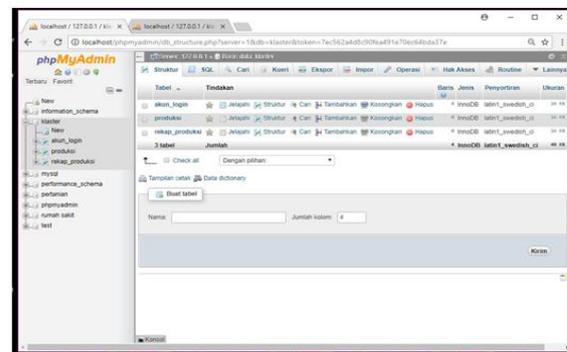
Gambar 3. Tampilan XAMPP

- b. Buka *database* atau *phpmyadmin* dengan membuka mozilla atau google chrome dengan *address* : localhost/phpmyadmin/



Gambar 2. Tampilan PHPMyAdmin

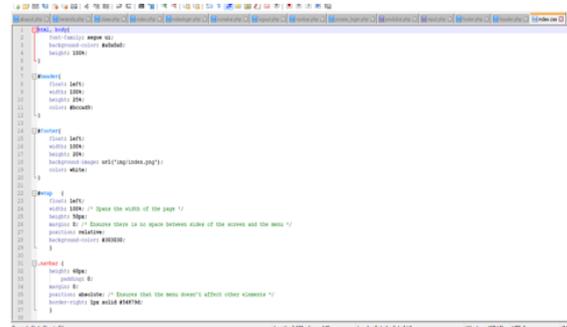
- c. Buat *database* dengan nama “kluster”. Di dalam *database* tersebut dibuatlah 3 tabel atau *structure* dengan nama:
 - *akun_login*. Berfungsi untuk mensinkronkan antara *database* dan juga *syntax* yang akan dibuat nantinya dan akan memudahkan tampilan pada aplikasi sesuai yang sudah dirancang. Di dalam *structure* *akun_login* terdapat 3 data (*id, username, password*)
 - *produksi*
 - *rekap_produksi*.



Gambar 4. Tampilan Database yang sudah dibuat

2. Tahap Kedua.

- a. Membuat tampilan atau biasa disebut *css*. Untuk *css* sendiri bisa *download*. Untuk file ini bisa disimpan dengan nama : *index.css*.



Gambar 5. Rancangan Tampilan Aplikasi

- b. Membuat file menu login agar terkoneksi ke dalam *database* untuk login. *File* ini disimpan dengan nama : *proses_login.php*.

V. KESIMPULAN

Berdasarkan hasil penelitian serta pembahasan dan implementasi aplikasi yang dibuat dapat ditarik kesimpulannya:

1. Metode *K-Means* dapat digunakan dalam penentuan kluster/hasil dari kualitas teh.
2. Jumlah data pada kluster 1 yaitu 49 data dengan kriteria berat kebun rata-rata 307.867 s.d. 658.761 dan berat pabrik rata-rata 277.080 s.d. 592.884 dengan suhu 100°C s.d. 150°C menunjukkan kepekatan warna yang rendah.
3. Jumlah data pada kluster 1 yaitu 49 data dengan kriteria berat kebun rata-rata di 108.967 s.d. 277.080 dan berat pabrik rata-rata 98.070 s.d. 249.372 dengan suhu 200°C s.d. 250°C menghasilkan kepekatan warna yang cukup pekat.

III. REFERENSI

- [1] Penebas Swadata Tim, (2008), *Agribisnis Tanaman Perkebunan*. Wisma Hijau: Depok.
- [2] James A. O Brien , George M. Marakas (2017), *Pengantar Sistem Informasi*, Salemba Empat.
- [3] Agus Mulyanto (2016), *Konsep Sistem Informasi & Aplikasi Pustaka Pelajar*, Yogyakarta.
- [4] Suyanto, (2018), *Data Mining untuk klasifikasi dan klusterisasi data*, penerbit Informatika, Bandung.
- [5] Johan Oscar Ong, (2013), “ Implementasi Algoritma K Means Clustering Untuk Menentukan Strategi Marketing President University, *Jurnal Ilmiah Teknik Industri*, Vol.12, No.1, Juni 2013, ISSN 1412-6869
- [6] Andri Kristanto. (2008). *Perancangan Sistem Informasi dan Aplikasinya*, penerbit Gava Media, Yogyakarta.
- [7] Retno Tri Wulandari, (2017). *Data Mining, Teori dan Aplikasi Rapidminer*. Gava Media, Yogyakarta.