

Pengelompokkan Ekspor Kopi Menurut Negara Tujuan Menggunakan Metode *K-Means Clustering* dengan *Silhouette Coefficient*

Suryadi Muzahidi Aziz*, Nur Azizah Komara Rifai

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*Suryadiaziz30@gmail.com, azizah_kr@yahoo.com

Abstract. Indonesia is one of the coffee exporting countries that has reached five continents, namely Asia, Australia, America, and Europe. The purpose of this study is to determine which countries are included in the high coffee export *cluster*, and the low coffee export *cluster*. The method used in this research is the *K-Means Clustering* method with *Silhouette Coefficient*. The advantage of this method is that it can perform analysis of larger samples more efficiently. The data used in this study is secondary data obtained from the Central Statistics Agency (BPS). The data contains data on the amount of coffee exports (net) and the value of FOB (Free On Board) from 2000 to 2020 in Indonesia. The software used in this research is RStudio. The results of this study conclude that, first, the *K-Means Clustering* Method with the *Silhouette Coefficient* can determine the grouping of coffee exports. The data is processed using the silhouette method which in its processing determines the value of k whose results are known to be 2 *clusters*, so the data to be taken is a high export level *cluster*, and a low export level *cluster*. Second, the data centroids for the high export level *cluster* are Japan, the United States and Germany. Third, the data centroids for the low export level *cluster* are Singapore, Malaysia, India, Egypt, Morocco, Algeria, England, Italy, Romania, Georgia, Belgium, the Netherlands, Denmark, and France. Fourth, the Silhouette coefficient value that can be known is 0.73, where $0.7 < SI \leq 1$ is a strong structure *cluster*, so coffee export groupings have very good quality because of their high value.

Keywords: *Clustering, Coffee Export, K-Means, Silhouette Coefficient.*

Abstrak. Indonesia menjadi salah satu negara pengekspor komoditi kopi yang telah menjangkau lima benua, yaitu Asia, Australia, Amerika, dan Eropa. Tujuan penelitian ini adalah menentukan negara yang termasuk ke dalam *cluster* ekspor kopi tinggi, dan *cluster* ekspor kopi rendah. Metode yang digunakan pada penelitian ini adalah metode *K-Means Clustering* dengan *Silhouette Coefficient*. Kelebihan metode ini adalah dapat melakukan analisis sampel dalam ukuran yang lebih besar dengan lebih efisien. Data yang digunakan dalam penelitian ini adalah data Sekunder yang di peroleh dari Badan Pusat Statistik (BPS). Data tersebut berisi data mengenai Jumlah ekspor kopi (netto) dan nilai FOB (Free On Board) dari tahun 2000 sampai 2020 di Indonesia. Software yang digunakan pada penelitian ini adalah RStudio. Hasil penelitian ini menyimpulkan bahwa, pertama, Metode *K-Means Clustering* dengan *Silhouette Coefficient* dapat menentukan pengelompokkan ekspor kopi. Data diolah dengan menggunakan metode silhouette yang dimana dalam pengolahan nya menentukan nilai k yang hasil nya diketahui 2 *cluster*, maka data yang akan di ambil adalah *cluster* tingkat ekspor tinggi, dan *cluster* tingkat ekspor rendah. Kedua, Centroid data untuk *cluster* tingkat ekspor tinggi adalah Negara Jepang, Amerika Serikat dan Jerman. Ketiga, Centroid data untuk *cluster* tingkat ekspor rendah adalah Negara Singapura, Malaysia, India, Mesir, Maroko, Aljazair, Inggris, Italia Rumania, Georgia, Belgia, Belanda, Denmark, dan Perancis. Keempat, Nilai Silhouette coefficient yang dapat diketahui adalah 0.73 yang mana $0.7 < SI \leq 1$ merupakan *cluster* strong structure jadi pengelompokan ekspor kopi mempunyai kualitas yang sangat baik karena nilai nya tinggi

Kata Kunci: *Clustering, Ekspor Kopi, K-Means, Silhouette Coefficient.*

A. Pendahuluan

Indonesia dikenal sebagai negara agraris yang beriklim tropis serta didukung oleh sub sektor pertanian. Salah satu unggulan sub sektor pertanian yang mempunyai peran penting bagi perekonomian Indonesia adalah kopi. Kopi juga merupakan komoditas ekspor Indonesia yang cukup penting sebagai penghasil devisa negara (Badan Pusat Statistik, 2019). Area penanaman kopi di Indonesia menurut Dirjen Perkebunan, Kementerian Pertanian pada Tahun 2017 luas yang diusahakan perkebunan rakyat adalah seluas 1.192.000 hektar, kemudian meningkat pada tahun 2018 seluas 1.210.000 hektar, dan terus meningkat pada tahun 2019 seluas 1.215.000 hektar yang tersebar di beberapa provinsi di Indonesia. Indonesia berada pada urutan ke empat negara yang melakukan ekspor kopi dari total ekspor kopi dunia yaitu sebesar 2,8%, sedangkan luas perkebunan kopi di Indonesia cukup luas (Ginting, 2019). Perlu dilihat perkembangan permintaan ekspor kopi Indonesia dengan melakukan pengelompokan wilayah negara yang menjadi tujuan ekspor kopi (Anjelita, dkk., 2019). Salah satu metode yang dapat digunakan untuk mengelompokkan ekspor kopi adalah metode *clustering* non hirarki yaitu k-means karena perbedaan dalam menentukan *cluster* lebih baik di banding jenis pengelompokan yang lain. Metode *Clustering* ini terbagi menjadi dua pendekatan, yaitu metode hierarki dan non hierarki. Metode hierarki adalah metode pengelompokkan objek atau data berdasarkan kemiripan sifatnya dan *cluster* yang diinginkan belum diketahui banyaknya, sedangkan metode non hierarki adalah pengelompokkan objek dimana banyaknya *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu (Santosa, 2007).

Penggunaan metode non hierarki dalam *clustering* dimulai dengan proses penentuan jumlah *cluster* terlebih dahulu. Metode non hierarki bertujuan mengelompokkan n obyek kedalam kelompok. Kelebihan metode ini adalah dapat melakukan analisis sampel dalam ukuran yang lebih besar dengan lebih efisien. Metode K-Means digunakan sebagai alternatif metode *cluster* untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi dalam proses penelitian dibanding metode hierarki.

Pada penelitian ini menggunakan metode *Silhouette Coefficient*. Metode ini digunakan untuk melihat kualitas dan kekuatan *cluster* yaitu seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Selain itu metode *Silhouette Coefficient* juga digunakan untuk mengukur seberapa dekat relasi antara objek dalam sebuah *cluster* (Prima, 2021).

Beberapa penelitian mengenai *clustering*, Windarto (2017) melakukan penelitian ekspor buah-buahan dilakukan dengan metode *K-Means Clustering* yang menerapkan centroid data untuk menentukan *cluster* ekspor buah buahan tersebut. kemudian pada penelitian Prima (2021) Analisa perbandingan nilai K terbaik untuk *Clustering* K-Means menggunakan pendekatan Elbow dan *Silhouette* pada citra aksara jawa. Oleh karena itu, maka peneliti bermaksud untuk melakukan penelitian dengan judul “Pengelompokkan Ekspor Kopi Menurut Negara Tujuan Menggunakan Metode *K-Means Clustering* dengan *Silhouette Coefficient*”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Menentukan nilai k dengan metode K-Means dan *Silhouette Coefficient* untuk menentukan bagian negara tujuan ekspor?
2. Mengetahui nilai *Silhouette Coefficient* untuk melihat seberapa baik pengelompokan?

B. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data Sekunder yang diperoleh dari data ekspor impor yang dihasilkan oleh Direktorat Jenderal Bea dan Cukai melalui situs <https://www.bps.go.id>. Data tersebut berisi data mengenai Jumlah ekspor kopi (netto) dan nilai FOB (Free On Board) dari tahun 2000 sampai 2020 di Indonesia. Software yang digunakan pada penelitian ini adalah RStudio. Penelitian ini terbagi menjadi delapan tahap, yaitu pengumpulan data, persiapan data, penentuan banyak *cluster*, penentuan centroid, penentuan jarak, pengelompokkan anggota *cluster*, melakukan iterasi, dan evaluasi data.

1. Persiapan data

Pada tahap awal, data jumlah ekspor berat bersih a(netto) dan nilai Free On Board a(FOB) ekspor kopi di setiap negara tujuan ekspor tahun 2000-2020 dijumlahkan kemudian data tersebut digunakan pada tahap *clustering*.

2. Penentuan banyak *cluster*
Banyaknya *cluster* tingkat ekspor tinggi, *cluster* tingkat ekspor sedang, dan *cluster* tingkat ekspor rendah.
3. Penentuan centroid
Penentuan centroid awal dilakukan berdasarkan data Free On Board (FOB). Penentuan centroid pada iterasi berikutnya menggunakan rumus pada persamaan (1)
4. Penentuan jarak
Penentuan jarak dilakukan dengan rumus Euclidean Distance pada persamaan (2)
5. Pengelompokkan anggota *cluster*
Pengelompokkan data ekspor kopi berdasarkan negara tujuan berdasarkan hasil dari euclidean distance
6. Perubahan pada kelompok anggota?
Melihat apakah ada perubahan terhadap anggota *cluster* atau tidak.
7. Melakukan iterasi
Melakukan tahap ke-4 sampai ke- sampai anggota *cluster* pada iterasi terakhir sama dengan anggota *cluster* pada iterasi sebelumnya.
8. Evaluasi data dan hasil *Cluster*
Setelah tidak ada perubahan pada pengelompokkan anggota *cluster* baru, maka proses *clustering* dianggap selesai
9. Pengecekan dengan Silhoutte Coefficient
Melakukan pengecekan seberapa baik nilai *cluster* dengan Silhoutte Coefficient
10. Interpretasi Hasil Silhoutte Coefficient
Melakukan interpretasikan dari hasil nilai Silhoutte Coefficient yang didapat.

C. Hasil Penelitian dan Pembahasan

Persiapan Data

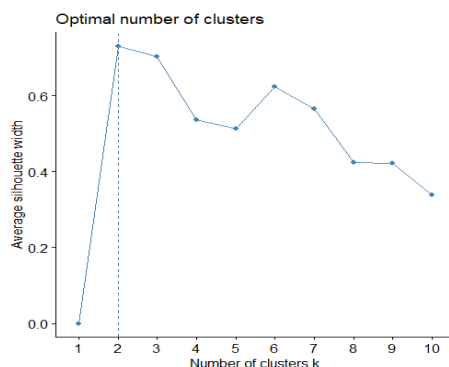
Dalam *clustering*, data yang diperoleh akan dihitung terlebih dahulu berdasarkan jumlah ekspor kopi tahun 2000-2020 berdasarkan negara tujuan. Hasil penjumlahan berdasarkan dua penilaian kriteria bobot bersih dan nilai FOB seperti terlihat pada Lampiran. Data tersebut kemudian diakumulasikan berdasarkan dua kriteria, yaitu berat bersih (*netto*) dan nilai FOB seperti yang ditunjukkan pada tabel 1 dibawah ini:

Tabel 1. Data Akumulasi Ekspor Kopi

Negara Tujuan	Berat Netto (ton)	Nilai FOB (US\$)
Jepang	1.000.081,2	1.883.390,5
Singapura	191.840,3	373.703,3
Malaysia	483.336,1	797.522,8
India	223.635,1	318.339,9
Mesir	309.873,3	538.071,0
Maroko	180.450,9	286.381,2
Aljazair	245.411,6	372.687,9
Amerika Serikat	1.273.017,3	3.874.079,9
Inggris	319.060,8	568.519,4
Jerman	1.019.937,3	1.598.187,2
Italia	594.802,7	968.703,6
Rumania	79.704,7	77.858,6
Georgia	184.596,5	293.309,2
Belgia	280.956,3	549.455,2
Belanda	51.257,3	121.054,9
Denmark	10.870,3	18.136,3
Perancis	103.740,4	169.368,5

Penentuan banyak nilai K menggunakan *Silhouette Coefficient*

Setelah diakumulasikan maka akan didapat nilai seluruh ekspor kopi negara tujuan. Kemudian data akan masuk ke tahap menentukan jumlah nilai k data yang terkumpul akan dimasukkan ke dalam RStudio. Penentuan jumlah nilai k dilakukan dengan menggunakan *Silhouette Coefficient*, dilihat dari garis samar yang menunjukan ke angka 2 pada plot maka dapat diketahui nilai k nya adalah 2 (dua) seperti pada gambar 4.1 dibawah ini:



Gambar 1. Penentuan nilai k dengan *Silhouette Coefficient*

Penentuan Centroid

Penentuan titik pusat *cluster* dilakukan dengan mengambil jumlah terbesar (maksimum) untuk *cluster* tingkat ekspor tinggi (C1), dan jumlah terkecil (minimum) untuk *cluster* tingkat ekspor rendah (C2). Nilai poin dapat dilihat pada Tabel 4.2 dibawah ini:

Tabel 2. Data Awal Centroid

Cluster ekspor tinggi	
Netto	FOB
1.273.017,3	3.874.079,9
Cluster ekspor rendah	
Netto	FOB
10.870,3	18.136,3

Dengan menggunakan centroid dapat mengelompokkan data yang telah didapatkan menjadi 2 *cluster*. Proses *cluster* dengan mengambil jarak terdekat dari setiap data yang diproses.

Penentuan Jarak

Tabel 3. Perhitungan Kesenjangan Pusat *Cluster*

No.	Negara	C1	C2	Shortest Distance
1.	Jepang	2009312,918	2111329,305	2009312,918
2.	Singapura	3663547,468	398971,2175	398971,2175
3.	Malaysia	3176287,139	911409,4846	911409,4846
4.	India	3707356,194	367955,244	367955,244
5.	Mesir	3472261,762	599779,0312	599779,0312
6.	Maroko	3750371,088	317352,968	317352,968
7.	Aljazair	3649071,034	425107,5846	425107,5846
8.	Amerika Serikat	0	4057254,749	0
9.	Inggris	3440459,74	630795,4827	630795,4827
10.	Jerman	2289920,756	1874773,868	1874773,868

11.	Italia	2983485,627	1115596,361	1115596,361
12.	Rumania	3979358,129	91131,37627	91131,37627
13.	Georgia	3742536,392	325424,2115	325424,2115
14.	Belgia	3469483,308	596025,3526	596025,3526
15.	Belanda	3946884,106	110559,2511	110559,2511
16.	Denmark	4057254,749	0	0
17.	Perancis	3884854,569	177471,2196	177471,2196

Pengelompokkan Anggota Cluster

Tabel 4. Pengelompokkan Data Iterasi 1

No.	Negara	C1	C2
1.	Jepang	I	
2.	Singapura		I
3.	Malaysia		I
4.	India		I
5.	Mesir		I
6.	Maroko		I
7.	Aljazair		I
8.	Amerika Serikat	I	
9.	Inggris		I
10.	Jerman		I
11.	Italia		I
12.	Rumania		I
13.	Georgia		I
14.	Belgia		I
15.	Belanda		I
16.	Denmark		I
17.	Perancis		I

Pada tabel 4 proses K-Means akan terus berulang hingga pengelompokkan data sama dengan pengelompokkan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data masuk iterasi terakhir sama dengan data iterasi sebelumnya.

Melakukan Iterasi

Tabel 5. Iterasi Data Centroid 2

Cluster ekspor tinggi	
Netto	FOB
1136549	2878735
Cluster ekspor rendah	
Netto	FOB
285298,2	470086,6

Setelah mendapatkan titik tengah atau centroid seperti pada tabel 5 diatas, proses yang sama dilakukan dengan mencari jarak terdekat. Proses pencarian jarak terdekat, pengelompokkan data pada iterasi 2 dan pengelompokkan data dapat digambarkan dalam tabel 6 berikut:

Tabel 6. Perhitungan Jarak Pusat *Cluster* Iterasi 2

No.	Negara	C1	C2	<i>Shortest Distance</i>
1.	Jepang	1004656	1583775	1004656,459
2.	Singapura	2677249	134254	134253,9648
3.	Malaysia	2181314	382666,3	382666,2502
4.	India	2718278	163796,8	163796,837
5.	Mesir	2482358	72289,78	72289,77948
6.	Maroko	2763046	211519,8	211519,8305
7.	Aljazair	2659774	105249,5	105249,4694
8.	Amerika Serikat	1004656	3544398	1004656,459
9.	Inggris	2450589	104062,1	104062,1284
10.	Jerman	1285847	1346219	1285846,619
11.	Italia	1985374	586866,2	586866,1887
12.	Rumania	2993632	442844,8	442844,7896
13.	Georgia	2755112	203448	203448,0022
14.	Belgia	2481448	79487,28	79487,2764
15.	Belanda	2963555	420236	420235,9923
16.	Denmark	3074114	528743,6	528743,5748
17.	Perancis	2899545	351275,7	351275,7107

Tabel 7. Pengelompokan Data Iterasi 2

No.	Negara	C1	C2
1.	Jepang	I	
2.	Singapura		I
3.	Malaysia		I
4.	India		I
5.	Mesir		I
6.	Maroko		
7.	Aljazair		I
8.	Amerika Serikat	I	
9.	Inggris		I
10.	Jerman	I	
11.	Italia		I
12.	Rumania		I
13.	Georgia		I
14.	Belgia		I
15.	Belanda		I
16.	Denmark		I
17.	Perancis		I

Dari tabel 7 diperoleh hasil pengelompokan iterasi ke-2 dengan hasil yang tidak sama pada iterasi 1. Proses akan berlangsung pada iterasi berikutnya seperti pada tabel 4.8 berikut:

Tabel 8. Iterasi Data Centroid 3

<i>Cluster ekspor tinggi</i>	
Netto	CIF
1097679	2451886
<i>Cluster ekspor rendah</i>	
Netto	CIF

232824	389508
--------	--------

Setelah mendapatkan nilai titik tengah atau centroid, proses yang sama dilakukan dengan mencari jarak terdekat. Proses pencarian jarak terpendek, pengelompokkan data pada iterasi 3 dan pengelompokkan data dapat digambarkan dalam tabel 9 berikut:

Tabel 9. Perhitungan Jarak Pusat *Cluster* Iterasi 3

No.	Negara	C1	C2	<i>Shortest Distance</i>
1.	Jepang	576812,1	1679395	576812,1
2.	Singapura	2267021	43925,55	43925,55
3.	Malaysia	1764748	478782,2	478782,2
4.	India	2305639	71758,85	71758,85
5.	Mesir	2069619	167354,6	167354,6
6.	Maroko	2351748	115663,6	115663,6
7.	Aljazair	2247092	21008,63	21008,63
8.	Amerika Serikat	1432962	3636515	1432962
9.	Inggris	2037968	198700,4	198700,4
10.	Jerman	857231,1	1442377	857231,1
11.	Italia	1566114	683005,2	683005,2
12.	Rumania	2583075	347233,2	347233,2
13.	Georgia	2343752	107610,9	107610,9
14.	Belgia	2070333	167032,4	167032,4
15.	Belanda	2554950	324088,8	324088,8
16.	Denmark	2665387	432643,5	432643,5
17.	Perancis	2489538	255194	255194

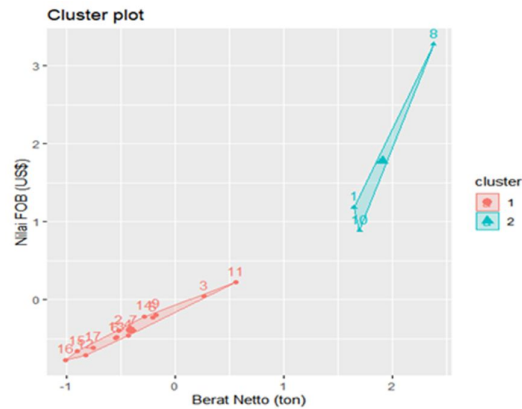
Tabel 10. Pengelompokkan Data Iterasi 3

No.	Negara	C1	C2
1.	Jepang	I	
2.	Singapura		I
3.	Malaysia		I
4.	India		I
5.	Mesir		I
6.	Maroko		I
7.	Aljazair		I
8.	Amerika Serikat	I	
9.	Inggris		I
10.	Jerman	I	
11.	Italia		I
12.	Rumania		I
13.	Georgia		I
14.	Belgia		I
15.	Belanda		I
16.	Denmark		I
17.	Perancis		I

Evaluasi Data dan hasil *cluster*

Pada iterasi 3 pengelompokkan data yang dilakukan pada 2 *cluster* dengan iterasi 3 didapatkan hasil yang sama. Dari 17 data utama ekspor kopi berdasarkan negaratujuan yang dapat diketahui, *cluster* tingkat ekspor tinggi (C1) yaitu negara Jepang Amerika Serikat dan Jerman, dan *cluster* tingkat ekspor rendah (C2) yaitu negara Singapura, Malaysia, India, Mesir, Maroko, Aljazair,

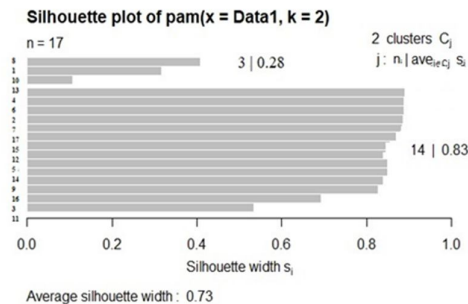
Inggris, Italia Rumania, Georgia, Belgia, Belanda, Denmark, dan Perancis. dibawah merupakan visualisasi negara yang masuk ke ekspor tinggi dan ekspor rendah pada gambar 2.



Gambar 2. Visualisasi Silhoutte Coefficient

Pengecekan dengan *Silhouette Coefficient*

Hasil perhitungan Silhoutte coefficient untuk pengelompokan ekspor kopi dengan bantuan software R, parameter yang digunakan adalah Euclidean. Nilai *Silhouette Coefficient* yang didapat dalam rentang $[-1,+1]$. Plot nilai K-means *Silhouette Coefficient* ditunjukkan pada gambar 3 berikut:



Gambar 3. Plot Nilai *Silhouette Coefficient*

Interpretasi Hasil *Silhouette Coefficient*

Berdasarkan plot nilai Silhoutte Coefficient diatas diketahui bahwa nilai rata-rata Silhoutte coefficient yaitu sebesar 0.73 Dilihat dari hasil nilai Silhoutte coefficient tertinggi yaitu 0.73 maka tergolong dalam strong structure yang mana kualitas *cluster* ekspor kopi sangat baik

D. Kesimpulan

Hasil penelitian dan pembahasan mengenai Pengelompokan Ekspor Kopi Menurut Negara Tujuan Menggunakan Metode *Clustering K-Means* dengan *Silhouette Coefficient* dapat ditarik kesimpulan sebagai berikut:

1. Data akan diolah dengan menggunakan metode *Silhouette Coefficient* yang dimana dalam pengolahan nya menentukan nilai k yang hasil nya diketahui 2 *cluster*, maka data yang akan di ambil adalah *cluster* tingkat ekspor tinggi, dan *cluster* tingkat ekspor rendah. centroid data untuk *cluster* tingkat ekspor tinggi adalah Negara Jepang, Amerika Serikat dan Jerman. Sedangkan centroid data untuk *cluster* tingkat ekspor rendah adalah Negara Singapura, Malaysia, India, Mesir, Maroko, Aljazair, Inggris, Italia Rumania, Georgia, Belgia, Belanda, Denmark, dan Perancis.
2. Nilai Silhoutte coefficient yang dapat diketahui adalah 0.73 yang mana $0.7 < SI \leq 1$

merupakan *cluster* strong scructure jadi pengelompokan ekspor kopi mempunyai kualitas yang sangat baik karena nilai nya tinggi.

Acknowledge

Penulis ucapkan terimakasih kepada orang tua dan keluarga yang sudah mendukung secara penuh untuk menuntaskan pendidikan penulis ditingkat sarjana, terimakasih kepada dosen pembimbing ibu Nur Azizah Komara Rifai, S.Si., M.Si . selaku dosen pembimbing yang sudah meluangkan waktunya untuk menyumbang pikiran, saran dan motivasi. Terakhir, terimakasih kepada institusi khususnya Prodi Statistika Universitas Islam Bandung sebagai wadah penulis untuk menimba ilmu pengetahuan.

Daftar Pustaka

- [1] Badan Pusat Statistik. (2019). Statistik Kopi Indonesia. Statistics Indonesia.
- [2] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A *K-Means Clustering* algorithm. Journal of the royal statistical society. series c (applied statistics), 28(1), 100-108.
- [3] Windarto, A. P. (2017). Penerapan Data mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan *K-Means Clustering* Method. Techno. Com
- [4] Roji, Ihfad Fahrur. (2022). *Penentuan Nilai Premi pada Asuransi Pertanian Berbasis Indeks Iklim di Kabupaten Bogor*, Jurnal Riset Statistika, 2(1), 65-72.