



**Implementasi K-Means & Google Data Studio Untuk Clustering
Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat**

TUGAS AKHIR

Zulfikar Awaluddin Habibie
41518010016

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Zulfikar Awaluddin Habibie
41518010016

UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010016

Nama : Zulfikar Awaluddin Habibie

Judul Tugas Akhir : Implementasi K-Means & Google Data Studio Untuk
Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota
di Jawa Barat

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 04 Juli 2022



Zulfikar Awaluddin Habibie

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Zulfikar Awaluddin Habibie
NIM : 41518010016
Judul Tugas Akhir : Implementasi K-Means & Google Data Studio
Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat.

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi ✓	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat	
	ISSN	2476-9843	
	Link Jurnal	https://journal.universitatumigora.ac.id/index.php/matrik/authorDashboard/submission/2245	
	Link File Jurnal Jika Sudah di Publish		

- Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
- Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 04 Juli 2022



Zulfikar Awaluddin Habibie

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Zulfikar Awaluddin Habibie
NIM : 41518010016
Judul Tugas Akhir : Implementasi K-Means & Google Data Studio
Untuk Clustering Produktivitas Pertanian Pada 10
Kabupaten/Kota di Jawa Barat.

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non Eksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 04 Juli 2022



Zulfikar Awaluddin Habibie

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010016
Nama : Zulfikar Awaluddin Habibie
Judul Tugas Akhir : Implementasi K-Means & Google Data Studio
Untuk Clustering Produktivitas Pertanian Pada 10
Kabupaten/Kota di Jawa Barat.

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 3 Agustus 2022



(Umniy Salamah, S.T., MMSI)

UNIVERSITAS
MERCU BUANA

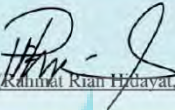
LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010016
Nama : Zulfikar Awaluddin Habibie
Judul Tugas Akhir : Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat.

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28-Juli-2022

TANDA TANGAN DIGITAL DOSEN PENGUJI


(Rahmat Rian Hidayat, ST, MMSI)

UNIVERSITAS
MERCU BUANA

Universitas Mercu Buana

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010016
Nama : Zulfikar Awaluddin Habibie
Judul Tugas Akhir : Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat.

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28-Juli-2022

TANDA TANGAN DIGITAL DOSEN PENGUJI



(Muhammad Rifqi, S.Kom, M.Kom)

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

NIM : 41518010016
Nama : Zulfikar Awaluddin Habibie
Judul Tugas Akhir : Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat.

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28-Juli-2022

Menyetujui,



(Wawan Gusawan, S.Kom, MT)
Dosen Pembimbing

Mengetahui,



(Wawan Gusawan, S.Kom, MT)
Koord. Tugas-Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Kai Prodi Teknik Informatika

MERCU BUANA

KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat” dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Orang tua. Ayah dan Ibu, yang selalu mendukung dan percaya bahwa saya bisa menyelesaikan kuliah dengan baik untuk proses meraih gelar Sarjana Komputer
2. Bapak Wawan Gunawan S.Kom., M.T selaku Dosen Pembimbing Tugas Akhir saya yang telah memberikan masukan saat bimbingan dan meluangkan waktunya untuk membimbing saya dalam penyusunan tugas akhir ini hingga selesai.
3. Bapak Sabar Rudiarto, M.Kom selaku Dosen Pembimbing Akademik
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat kepada saya selama perkuliahan.
5. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
6. Sahabat serta teman dekat, yang selalu memberikan dukungan kepada saya dan percaya bahwa saya bisa melewati dan menyelesaikan Tugas Akhir ini.
7. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, penulis sadar bahwa hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 04 Juli 2022
Zulfikar Awaluddin Habibie



DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI.....	v
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	14
BAB 1. LITERATUR REVIEW	15
BAB 2. ANALISIS DAN PERANCANGAN.....	20
BAB 3. SOURCE CODE	23
BAB 4. DATASET.....	46
BAB 5. TAHAPAN EKSPERIMEN	49
BAB 6. HASIL SEMUA EKSPERIMEN.....	62
DAFTAR PUSTAKA	66
LAMPIRAN DOKUMEN HAKI.....	68
LAMPIRAN KORESPONDENSI	70

NASKAH JURNAL

Implementasi K-Means dan Google Data Studio Untuk Clustering Potensi Pertanian Jawa Barat

Zulfikar Awaluddin H¹, Wawan Gunawan²

¹Universitas Mercu Buana, Indonesia

Article Info	ABSTRAK
<p>Genesis Artikel: Diterima, Tanggal Bulan Tahun Direvisi, Tanggal Bulan Tahun Disetujui, Tanggal Bulan Tahun</p>	<p>Pertanian adalah suatu hal yang sangat dekat dengan Indonesia karena Indonesia adalah negara tropis yang memiliki potensi alam yang sangat baik, berdasarkan data dari Badan Pusat Statistik Jawa Barat dijelaskan bahwa Jawa Barat adalah salah satu provinsi yang memiliki potensi pertanian yang tinggi, Berdasarkan hal tersebut maka peneliti melakukan eksperimen tentang daerah potensial berdasarkan produktivitas dan pembagian kluster wilayah di Jawa Barat yang memiliki potensi pertanian yang tinggi pada tiap komoditi dengan menggunakan Algoritma <i>K-Means</i> untuk mengklusterisasi wilayah Kabupaten/Kota di Jawa Barat, serta dilakukan visualisasi dengan Google Data Studio untuk mempermudah melihat sebaran wilayah hasil klusterisasi. Data yang diperoleh adalah hasil akurasi terbaik sebesar 91% dan hasil terendah 63% dari total data sebanyak 1535 baris dengan kolom yang menjadi pusat penilaian. Pusat penelitian ini adalah hasil pertanian dan luas lahan pertanian pada komoditi mencakup bawang daun, melon, semangka, bayam, kangkung, buncis, blewah, lobak, paprika dan lainnya.</p>
<p>Kata Kunci: <i>Potensi Pertanian</i> <i>K-Means</i> <i>Produktivitas</i></p>	
<p>Keywords: <i>Agricultural Potential</i> <i>K-Means</i> <i>Productivity</i></p>	
<p>ABSTRACT</p> <p><i>Agriculture is something that is very close with Indonesia because Indonesia is a tropical country which has excellent natural potential. Based on West Java Central Bureau of Statistic, West Java is one of provinces which has high potential in Agriculture. Authors decided to do experiment about potential area and eluster division at West Java which have high Agriculture potential on each commodity using the K-Means Algorithm to cluster the area which have high agriculture potential and the visualization was carried out with Google Data Studio to better see the distribution of clustered regions. Data result showing off the best accuracy results were 91% and the lowest results were 63% of the total data as 1535 rows with the column being the center of the assessment being agricultural products. The data of this research are agricultural yield and land area on commodities which include green onions, melons, watermelons, spinach, kale, beans, cantaloupe, radishes, peppers etc.</i></p> <p style="text-align: right;"><i>This is an open access article under the CC BY-SA license.</i></p>	
<p>Penulis Korespondensi: Zulfikar Awaluddin Habibie, Program Studi Teknik Informatika, Universitas Mercu Buana, Indonesia Email: 41518010016@student.mercubuana.ac.id</p>	

1. PENDAHULUAN

Pertanian adalah sebuah aktivitas untuk menghasilkan suatu hasil bumi. Kegiatan pertanian ataupun bercocok tanam di Indonesia menjadi suatu hal yang lazim ditemui mengingat Indonesia adalah salah satu negara agraris [1] dan hampir tiap provinsi di Indonesia memiliki jumlah produksi pertanian yang besar, salah satunya adalah Jawa Barat.

Dengan potensi pertanian pada provinsi Jawa Barat yang tinggi sesuai dengan data (BPS Provinsi Jawa Barat, 2019), untuk itu klusterisasi wilayah di Jawa Barat yang memiliki produksi pertanian tinggi pada tiap komoditi hasil pertanian, akan menjadi sangat penting, guna membantu pemerintah provinsi Jawa Barat untuk membuat keputusan ekspor dan ketahanan pangan.

Oleh karena itu penulis akan melakukan klusterisasi wilayah Jawa Barat berdasarkan potensi pertanian dan komoditi hasil pertanian pada tiap kabupaten/kota di Jawa Barat, dengan menggunakan algoritma *K-Means* [2] kemudian hasil klusterisasi akan divisualisasikan dengan menggunakan Google Data Studio, yang akan memberikan output pemetaan wilayah potensial hasil pertanian di Jawa Barat pada tiap kabupaten/kota berdasarkan komoditi hasil pertanian dan tahun.

Variabel yang akan digunakan pada penelitian ini untuk proses klusterisasi dengan *K-MEANS* adalah hasil pertanian dan produktivitas pertanian tiap komoditi di kabupaten/kota pada Jawa Barat.

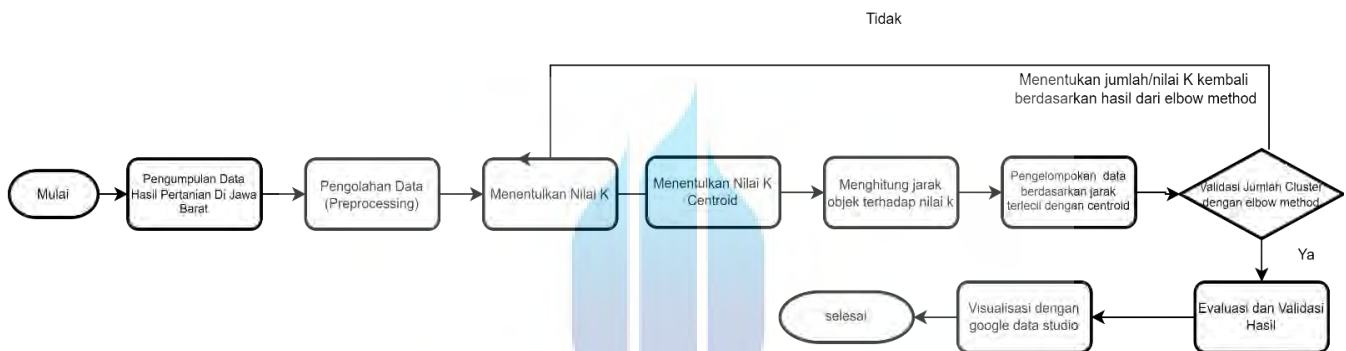
Beberapa penelitian telah memberikan solusi dalam proses klusterisasi wilayah hasil pertanian, diantaranya adalah penelitian yang dilakukan oleh Yani Sugiyani dengan menerapkan metode data mining algoritma *K-Means* dalam studi kasus pengelompokan wilayah berdasarkan potensi hasil pertanian di kota Cilegon, dengan label pembandingan adalah luas panen dan produksi pertanian kemudian membagi wilayah berdasarkan kecamatan [3], selain penelitian yang telah dilakukan oleh Yani Sugiyani, terdapat pula penelitian yang dilakukan oleh Putrama Alkhairi dan Agus Perdana Windarto dengan menerapkan metode data mining algoritma *K-Means* dalam studi kasus klusterisasi wilayah potensial pertanian di Sumatra Utara dengan hasil pertanian karet, pada penelitian ini label yang dipergunakan untuk pembandingan adalah Luas dan Produktivitas pertanian pada tiap kabupaten di provinsi Sumatra Utara, hasil yang diperoleh ialah wilayah Sumatra Utara dibagi menjadi 3 kluster berdasarkan produktivitas pertanian [2]. Serta penelitian yang dilakukan oleh Rachman & Handayani pada penelitian ini memperoleh, nilai *Accuracy* 81.81%, *Precision* 63.63%, dan *Recall* 100% dengan hasil AUC 0.800 studi kasus yang diangkat pada penelitian ini adalah memprediksi tingkat kelancaran pembayaran sewa teras UMKM [4].

Pada penelitian ini penulis mencoba memvisualisasikan hasil klusterisasi dengan menggunakan Google Data Studio [5] yang akan memberikan output pemetaan wilayah potensial hasil pertanian di Jawa Barat pada tiap kabupaten/kota berdasarkan komoditi hasil pertanian dan tahun, serta komoditi hasil pertanian yang akan diklusterisasi lebih dari satu komoditi, yaitu bawang daun, melon, semangka, bayam, kangkung, buncis, bawang merah, blewah, cabai besar, bawang putih, cabai rawit, kacang merah, kacang panjang, kembang kol, ketimun,

kubis, labu siam, lobak, paprika. Diharapkan banyaknya data dari hasil pertanian dapat menghasilkan data yang lebih kompleks serta lengkap.

2. METODE PENELITIAN

Metode Penelitian yang dilakukan oleh penulis kali ini adalah penelitian kuantitatif yang mana penelitian ini akan berdasarkan pada jumlah data dan nilai yang ada pada variabel data yang diteliti secara terstruktur. Sumber penelitian ini akan melalui portal open data Jawa Barat (<https://opendata.jabarprov.go.id/>) dan data yang diambil adalah data mengenai hasil pertanian dan luas wilayah pertanian per-komoditi seperti bawang daun, melon, semangka, bayam, kangkung, buncis, blewah, bawang merah, bawang putih, cabai besar, cabai rawit, kacang merah, kacang panjang, kembang kol, ketimun, kubis, labu siam, lobak dan paprika pada tiap tahun serta pada tiap kabupaten/kota di Jawa Barat.



Gambar 1. Tahapan Penelitian

Seperti yang terlihat pada gambar 1, tahapan yang akan dilakukan oleh penulis adalah sebagai berikut:

A. Pengumpulan Data

Tahapan ini dilakukan dengan mengambil data yang telah tersedia di portal open data Jawa Barat dengan cakupan data yaitu Kabupaten/Kota di Jawa Barat. Parameter pada penelitian ini adalah Luas Lahan Pertanian, Hasil Pertanian, Produktivitas Pertanian, Komoditi hasil pertanian dan tahun. Kemudian sampel yang akan dipergunakan datanya pada penelitian ini adalah bawang daun, melon, semangka, bayam, kangkung, buncis, blewah, bawang merah, bawang putih, cabai besar, cabai rawit, kacang merah, kacang panjang, kembang kol, ketimun, kubis, labu siam, lobak, paprika. Data yang terkumpul adalah sebanyak 3941 baris/data. Dengan cakupan wilayah seluruh kabupaten atau kota terdiri dari 10 Kabupaten/Kota pada tiap interval tahun nya dari tahun 2013-2020, seperti yang terlihat pada tabel 1.

Tabel 1. List Kabupaten/Kota

nama_kabupaten_kota
KABUPATEN BOGOR
KABUPATEN SUKABUMI
KABUPATEN CIANJUR
KABUPATEN BANDUNG
KABUPATEN GARUT
KABUPATEN TASIKMALAYA
KABUPATEN CIAMIS
KABUPATEN KUNINGAN
KABUPATEN MAJALENGKA
KABUPATEN SUMEDANG

Sumber : (<https://opendata.jabarprov.go.id/>)

B. Pengolahan Data (*Preprocessing*)

Setelah data yang dibutuhkan telah terkumpul maka tahapan selanjutnya yang akan dilakukan adalah *Preprocessing* data [6]. proses ini dilakukan dengan menghapus data yang bernilai kosong atau null, ini dilakukan agar menghindari kesalahan dalam perhitungan pada proses selanjutnya[7], selain itu akan dilakukan proses penyesuaian dari tipe dari dataset yang dipergunakan[8], penyesuaian tipe data ini adalah mengganti tipe data float pada kolom hasil pertanian dan luas lahan menjadi integer.

C. Implementasi Algoritma *K-Means*

Pada saat implementasi dengan Algoritma *K-Means*, dipergunakan jumlah data jadi sebanyak 1535 baris, yang mana data ini akan menjadi keseluruhan untuk data yang akan siap diolah dengan algoritma K-Means sedangkan kolom yang akan diperhitungkan dalam proses pengolahan data dengan algoritma K-Means adalah hasil pertanian dan produktivitas pertanian, selain itu Proses dari Algoritma *K-Means* adalah dengan penentuan nilai K terlebih dahulu [9], kemudian dilanjutkan dengan penentuan dari nilai centroid[10], setelahnya akan dilakukan proses perhitungan tiap-tiap nilai objek ke masing-masing centroid dari tiap cluster, setelahnya akan dikelompokan data tersebut berdasarkan jarak minimum dengan centroid, lalu dilakukan perulangan untuk menentukan centroid baru dan akan diulang hingga posisi centroid baru sama. Seperti yang terlihat pada persamaan 1.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

Dimana :

C_{ij} = Pusat Cluster

C_{kj} = Data

D. Pengujian Jumlah Cluster

Tahapan selanjutnya adalah pengujian jumlah cluster, ditahapan ini akan menguji dan menentukan jumlah cluster terbaik dengan menggunakan elbow method. Elbow method akan dapat menghasilkan dan menentukan jumlah cluster terbaik berdasarkan jumlah *SSE* (Sum of Square Error) terkecil [11], sehingga akan diperoleh jumlah cluster terbaik.

E. Evaluasi dan Validasi Model

Proses evaluasi dan validasi model adalah tahapan untuk melihat seberapa jauh ketepatan model yang telah dibuat, proses evaluasi dan validasi model ini akan dilihat dengan Confusion Matrix, hasilnya akan tercermin dari nilai *Accuracy*, *Precision* dan *Recall* [12] seperti yang terlihat pada gambar 3, dengan 3 parameter nilai tersebut maka akan diketahui apakah model yang telah dibuat hasilnya baik atau tidak.

F. Visualisasi Hasil *Clustering* Dengan Google Data Studio.

Setelah Mendapatkan data hasil proses algoritma K-Means dengan akurasi yang baik, maka selanjutnya data dengan label cluster yang terbentuk akan divisualisasikan dengan Google Data Studio untuk membuat tampilan dashboard yang akan mempermudah dalam melihat hasil clusterisasi dan melihat kecenderungan data untuk bisa mengambil keputusan.

3. HASIL DAN ANALISIS

A. Tahap Pengumpulan Data

Penelitian ini mengambil data pada portal (<https://opendata.jabarprov.go.id/>) yang disesuaikan dengan kebutuhan. Parameter yang digunakan adalah Luas Lahan Pertanian, sampel Hasil Pertanian mencakup 19 hasil pertanian, produktivitas pertanian dan komoditi hasil pertanian dan tahun. Proses ini dilakukan dengan mengunduh data hasil generate pada portal open data jawa barat kemudian dilakukan proses penggabungan pada tiap hasil komoditi menggunakan python panda data frame. Data yang terkumpul adalah sebanyak 3941 baris/data. Representasi data penelitian terdapat pada tabel 2 berikut:

Tabel 2. Representasi Data Penelitian

nama_provinsi	nama_kabupaten_kota	hasil_panen	satuan	tahun	komoditas	luas_panen	satuan_luas	produktivitas
JAWA BARAT	KABUPATEN BOGOR	3245.0	TON	2013	BAWANG DAUN	374.0	HEKTAR	8.676.471
JAWA BARAT	KABUPATEN TASIKMALAYA	895.0	TON	2013	LABU SIAM	60.0	HEKTAR	14.916.637
JAWA BARAT	KABUPATEN GARUT	13324.0	TON	2013	LABU SIAM	368.0	HEKTAR	36.206.522
JAWA BARAT	KABUPATEN CIANJUR	3282.0	TON	2013	LABU SIAM	146.0	HEKTAR	22.479.452
JAWA BARAT	KABUPATEN BOGOR	5756.0	TON	2013	BUNCIS	526.0	HEKTAR	10.942.963
...
JAWA BARAT	KABUPATEN SUBANG	2438.0	TON	2013	SEMANGKA	129.0	HEKTAR	18.899.225
JAWA BARAT	KABUPATEN KARAWANG	1.0	TON	2013	SEMANGKA	4.0	HEKTAR	1.250.000
JAWA BARAT	KABUPATEN BEKASI	67.0	TON	2013	SEMANGKA	6.0	HEKTAR	11.163.637
JAWA BARAT	KABUPATEN PURWAKARTA	147.0	TON	2013	SEMANGKA	9.0	HEKTAR	16.333.333
JAWA BARAT	KABUPATEN PURWAKARTA	5.0	TON	2013	LOBAK	1.0	HEKTAR	5.000.000

B. Pengolahan Data (*Preprocessing*)

Dilakukan dengan membuang kolom yang tidak digunakan [13] kemudian mengubah tipe data produktivitas dari float menjadi integer seperti yang terlihat pada tabel 2. Hal ini dilakukan untuk mempermudah proses clustering dan melihat kecenderungan angka. karakteristik dari tipe data float adalah terdapat angka koma maka ini akan mempersulit proses perhitungan dan visualisasi, maka dari itu diubah menjadi tipe data integer.

Tabel 3. Data Missing Value

nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	hasil_panen	satuan	tahun	komoditas	luas_panen	satuan_luas	produktivitas
JAWA BARAT	3274	KOTA CIREBON	0.0	TON	2020	CABAI RAWIT	0.0	HEKTAR	NaN
JAWA BARAT	3279	KOTA BANJAR	0.0	TON	2020	PAPRIKA	0.0	HEKTAR	NaN

Tabel 4. Jumlah Data Setelah *Preprocessing*

no	nama provinsi	nama kabupaten kota	hasil panen	satuan	tahun	komoditas	luas panen	satuan luas	produktivitas
1	JAWA BARAT	KABUPATEN BOGOR	3245.0	TON	2013	BAWANG DAUN	374.0	HEKTAR	8.676.471
2	JAWA BARAT	KABUPATEN TASIKMALAYA	895.0	TON	2013	LABU SIAM	60.0	HEKTAR	14.916.637
3	JAWA BARAT	KABUPATEN GARUT	13324.0	TON	2013	LABU SIAM	368.0	HEKTAR	36.206.522
4	JAWA BARAT	KABUPATEN CIANJUR	3282.0	TON	2013	LABU SIAM	146.0	HEKTAR	22.479.452
5	JAWA BARAT	KABUPATEN BOGOR	5756.0	TON	2013	BUNCIS	526.0	HEKTAR	10.942.963
...
2351	JAWA BARAT	KABUPATEN SUBANG	2438.0	TON	2013	SEMANGKA	129.0	HEKTAR	18.899.225
2352	JAWA BARAT	KABUPATEN KARAWANG	1.0	TON	2013	SEMANGKA	4.0	HEKTAR	1.250.000
2353	JAWA BARAT	KABUPATEN BEKASI	67.0	TON	2013	SEMANGKA	6.0	HEKTAR	11.163.637
2354	JAWA BARAT	KABUPATEN PURWAKARTA	147.0	TON	2013	SEMANGKA	9.0	HEKTAR	16.333.333
1535	JAWA BARAT	KABUPATEN PURWAKARTA	5.0	TON	2013	LOBAK	1.0	HEKTAR	5.000.000

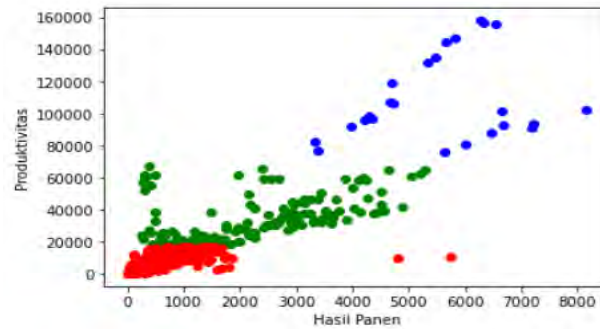
Seperti yang terlihat pada gambar 2, telah dilakukan *preprocessing* data dimana Jumlah baris data sebelum preprocessing adalah 3941 rows, setelah preprocessing menjadi 1535 rows, yang mana data 1535 baris ini adalah data yang akan siap masuk pada proses perhitungan dengan algoritma K-Means. Data yang hilangkan pada proses ini adalah data yang memiliki nilai null atau missing value.

C. Tahap Implementasi Algoritma *K-Means*

Dilakukan penentuan nilai K terlebih dahulu, kemudian dilanjutkan dengan penentuan dari nilai centroid [14], setelahnya akan dilakukan proses perhitungan tiap-tiap nilai objek ke masing-masing centroid dari tiap cluster, lalu akan dilakukan pengelompokan data berdasarkan jarak minimum dengan centroid serta dilakukan perulangan untuk menentukan centroid baru dan akan diulang hingga posisi centroid baru sama. Titik centroid yang akan menjadi acuan terdapat pada gambar

Cluster	luas_panen	hasil_panen
0	263.0	3325.0
1	5546.0	109512.0
2	2291.0	36599.0

Gambar 2. Titik *Centroid*.



Gambar 3. Visualisasi Data Cluster

Seperti yang terlihat pada gambar 3, adalah hasil visualisasi dari data cluster sesuai dengan titik centroid pada gambar 2, dapat dijelaskan bahwa titik berwarna merah adalah praduga data dengan hasil cluster kelompok 1, titik berwarna hijau adalah praduga data dengan hasil cluster kelompok 2 dan titik berwarna biru adalah praduga data dengan hasil cluster kelompok 3. Yang mana kelompok 1 adalah kelompok data yang memiliki produktivitas rendah, sedangkan kelompok 2 adalah kelompok data yang memiliki produktivitas sedang dan kelompok 3 adalah kelompok data yang memiliki produktivitas tinggi.

Setelah ditetapkan nilai centroid data maka akan dilakukan perhitungan dengan persamaan 1 sebagai berikut:

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

Dimana :

C_{ij} = Pusat Cluster

C_{kj} = Data

Dimana :

C_{ij} = Pusat Cluster

C_{kj} = Data

Hasil dari perhitungan tersebut adalah :

$$D11 = \sqrt{((374 - 263,1)^2 + (3245 - 3325,3))^2} \\ = 136,919$$

$$D12 = \sqrt{((374 - 5546,4)^2 + (3245 - 109511,9))^2} \\ = 3361,2$$

$$D13 = \sqrt{((374 - 2290,9)^2 + (3245 - 36599,4))^2} \\ = 33.409,437$$

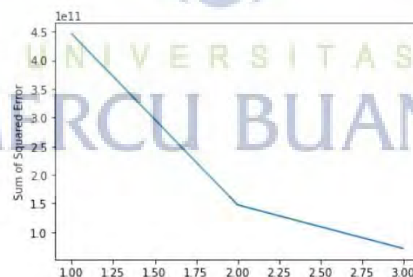
Berdasarkan perhitungan diatas maka data pada baris pertama ini, akan masuk kedalam cluster pertama, karena memiliki nilai terkecil dengan jarak kluster pertama, perhitungan akan dilanjutkan hingga data ke 1535, kemudian data hasil perhitungan ini akan dikelompokkan berdasarkan hasil cluster yang diperoleh. Seperti pada tabel 5 berikut:

Table 4. Hasil Pengelompokan Data

nama provinsi	kode kabupaten kota	nama kabupaten kota	hasil panen	satuan	tahun	komoditas	luas panen	satuan luas	produktivitas	cluster
JAWA BARAT	3205	KABUPATEN GARUT	155701	TON	2013	KUBIS	6546	HEKTAR	23	1
JAWA BARAT	3207	KABUPATEN CIAMIS	5937	TON	2013	CABAI BESAR	577	HEKTAR	10	0
JAWA BARAT	3205	KABUPATEN GARUT	93142	TON	2013	CABAI BESAR	7107	HEKTAR	12	1
...
JAWA BARAT	3208	KABUPATEN KUNINGAN	3383	TON	2013	CABAI RAWIT	193	HEKTAR	17	0
JAWA BARAT	3204	KABUPATEN BANDUNG	17524	TON	2013	CABAI BESAR	719	HEKTAR	24	0
JAWA BARAT	3271	KABUPATEN BOGOR	391	TON	2013	CABAI RAWIT	12	HEKTAR	32	0

D. Pengujian Jumlah Cluster

Pengujian jumlah cluster, ditahapan ini akan menguji dan menentukan jumlah cluster terbaik dengan menggunakan *elbow method*. *Elbow method* akan dapat menghasilkan dan menentukan jumlah cluster terbaik berdasarkan jumlah *SSE* (Sum of Square Error) terkecil [15]. Pada penelitian ini hasil *SSE* jatuh pada nilai $k=3$ sesuai dengan kaidah dari *Elbow Method*.



Gambar 4. Visualisasi perhitungan *SSE* terhadap nilai k

Seperti yang terlihat pada gambar 4 dapat dilihat bahwa nilai k terbaik adalah 3 dengan penurunan nilai *sse* yang stabil, dikatakan stabil pada angka tiga karena penilaian pada *SSE* akan diambil dari patahan siku signifikan [16], ini terlihat pada angka $k=3$ Visualisasi perhitungan *SSE* terhadap nilai k menunjukkan penilaiannya yang stabil.

E. Evaluasi dan Validasi Model

Proses evaluasi ini dilakukan dengan menggunakan Confusion Matrix, tahapan ini adalah proses untuk melihat seberapa baik hasil penelitian yang dihasilkan, hasil

ini berupa seberapa baik Algoritma K-Means dalam menentukan kluster, yang mana representasi hasilnya akan tercermin dari nilai *accuracy*, Berdasarkan hasil pada tabel 5, hasil dari kinerja model terbaik jatuh pada percobaan pertama dengan nilai K-Range = 3, memperoleh **akurasi sebesar 91**.

No	Percobaan	Hasil	K-Range
1	Percobaan 1	91%	4
2	Percobaan 2	63%	10

Tabel 5. Data Percobaan

F. Visualisasi Hasil *Clustering* Dengan Google Data Studio.

Seperti yang terlihat pada tabel 5, hasil evaluasi dari klasifikasi *K-means* dinyatakan baik dengan nilai akurasi 91%. Data dengan hasil akurasi 91% ini akan masuk ke tahap terakhir yaitu visualisasi data *Clustering* dengan *Google Data Studio* untuk mempermudah melihat data hasil clustering. Hasilnya akan berbentuk peta persebaran wilayah potensial hasil pertanian, Dari hasil visualisasi ini dapat pada gambar 5, sebaran daerah pada seluruh kabupaten/kota di Jawa Barat pada filter nama_kabupaten_kota, dan tiap komoditas pada filter komoditas serta tahun pada dropdown filter.



Gambar 5. Hasil Visualisasi data

4. KESIMPULAN

Kesimpulan dari percobaan algoritma K-Means implementasi Elbow Methode ditentukan bahwa proses implementasi K-Means akan membuat klaster data menjadi 3 dengan Sum of Squared Error terkecil. Hasil dari kinerja model terbaik adalah memperoleh akurasi sebesar 91%. Untuk percobaan dengan hasil akurasi terendah adalah 63%, dengan mengganti range untuk perhitungan Sum of Squared Error terkecil menjadi 10, penurunan akurasi ini terjadi karena jumlah data yang dipergunakan tidak cocok untuk menggunakan range sebesar 10, karena diawal penulis mencoba memperdiksi jumlah cluster adalah 3 dan baris data latih yang dipergunakan juga hanya sebesar 1535 rows, sehingga akan memperoleh hasil yang kurang baik. Serta salah satu hasil visualisasi dengan Google Data Studio dapat menunjukkan Filter Dashboard untuk tahun 2013 dan Kabupaten Bandung dan Garut adalah daerah produktivitas hasil pertanian untuk komoditi Kubis Terbaik.

REFERENSI

- [1] Aryawati Ni Putu Riska and M. K. S. Budhi, “Pengaruh Produksi, Luas Lahan, dan Pendidikan Terhadap Pendapatan Petani dan Alih Fungsi Lahan Provinsi Bali,” *E-JURNAL EKONOMI PEMBANGUNAN UNIVERSITAS UDAYANA*, VOL. 7, No.9 September 2018, 2018.
- [2] P. Alkhairi and A. P. Windarto, “Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara,” *Seminar Nasional Teknologi Komputer & Sains*, pp. 762–767, 2019.
- [3] Y. Sugiyani, “Pengelompokan Wilayah Berdasarkan Potensi Hasil Pertanian Menggunakan Algoritma K-means Di Kota Cilegon,” *Jurnal ProTekInfo*, vol. 3, no. September, pp. 60–67, 2016.
- [4] R. Rachman and R. N. Handayani, “Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM,” *Jurnal Informatika*, vol. 8, no. 2, pp. 111–122, 2010, doi: 10.31294/ji.v8i2.10494.
- [5] D. Apriani, M. Aan, and W. E. Saputra, “Data Visualization Using Google Data Studio,” *International Journal of Cyber and IT Service Management*, vol. 2, no. 1, pp. 11–19, 2022, doi: 10.34306/ijcitsm.v2i1.68.
- [6] S. Handoko, F. Fauziah, and E. T. E. Handayani, “Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering,” *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–91, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [7] S. Rudiarto, “Implementasi Algoritma K-Means Clustering Pada Aplikasi Pencari Pelanggan Potensial Pada Restoran Xyz,” *Jurnal Ilmu Teknik dan Komputer*, vol. 2, no. 1, pp. 39–62, 2018.
- [8] W. M. Baihaqi, M. Pinilih, and M. Rohmah, “Kombinasi K-Means Dan Support Vector Machine (Svm) Untuk K-Means and Support Vector Machine (Svm) Combination To Predict Sara Elements on Tweet,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 501–510, 2020, doi: 10.25126/jtiik.202071026.
- [9] D. Triyansyah and D. Fitriana, “Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing,” *Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [10] S. S. Nagari and L. Inayati, “Implementation of Clustering Using K-Means Method To Determine Nutritional Status,” *Jurnal Biometrika dan Kependudukan*, vol. 9, no. 1, p. 62, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.
- [11] N. Putu, E. Merliana, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means,” pp. 978–979.
- [12] I. W. Gamadarenda and I. Waspada, “Implementasi Data Mining untuk Deteksi Penyakit Ginjal Kronis (PGK) menggunakan K-Nearest Neighbor (KNN) dengan Backward Elimination,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, p. 417, 2020, doi: 10.25126/jtiik.2020710896.
- [13] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

- [14] W. Mega, “CLUSTERING MENGGUNAKAN METODE K-MEANS UNTUK MENENTUKAN STATUS GIZI BALITA,” 2015.
- [15] Y. Darmi, A. Setiawan, J. Bali, K. Kampung Bali, K. Teluk Segara, and K. Bengkulu, “PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK,” 2016.
- [16] D. Rusjyanthi, “Identifikasi Biometrika Telapak Tangan Menggunakan Metode Pola Busur Terlukalisasi, Block Standar Deviasi, Dan K-Means Clustering,” *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 4, no. 2, pp. 265–276, 2013.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten/Kota di Jawa Barat.”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.



UNIVERSITAS
MERCU BUANA