BAB 4
HASIL PENELITIAN

Proses penelitian dilakukan dengan menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*). CRISP-DM merupakan metodologi standar dalam data mining yang terdiri dari enam tahapan utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.

Setiap tahapan tersebut telah dilaksanakan secara sistematis untuk menjawab tujuan penelitian serta menghasilkan informasi yang dapat digunakan sebagai dasar pengambilan keputusan. Dalam bab ini, peneliti akan memaparkan hasil dari masing-masing tahapan CRISP-DM, mulai dari pemahaman konteks bisnis hingga penyajian model akhir yang dihasilkan dari proses analisis data. Hasil yang ditampilkan mencakup pemahaman awal terhadap permasalahan dan kebutuhan *stakeholder*, karakteristik dan kondisi data yang digunakan, proses pembersihan serta transformasi data, penerapan algoritma atau teknik analisis, evaluasi terhadap performa model, hingga interpretasi serta rekomendasi berdasarkan temuan yang diperoleh.

4.1 BUSINESS UNDERSTANDING

Tahap *business understanding* merupakan langkah awal dalam proses CRISP-DM yang berfokus pada pemahaman tujuan bisnis secara menyeluruh serta identifikasi permasalahan yang ingin diselesaikan melalui analisis data. Pada tahap ini, peneliti menggali kebutuhan dan harapan stakeholder, merumuskan tujuan analisis yang selaras dengan kebutuhan bisnis, serta menetapkan indikator keberhasilan dari proyek analisis yang dilakukan.

4.1.1 *Business Problem*

Pada tahap ini penulis mengidentifikasi permasalahan bisnis yang ada pada Lampiran 1 dari feedback pemangku kepentingan yaitu bapak Cahya Daru Saputro selaku Kabid Sub Program Dinas Tata Ruang Daerah Istimewa Yogyakarta. Pada

gambar 4.1 merupakan diskusi Bersama Bapak Cahya Daru Saputro selaku Kabid Sub Program Dinas Tata Ruang Daerah Istimewa Yogyakarta.

Gambar 4. 1 Diskusi Bersama kabid Sub Program Dinas Tata Ruang

Dari hasil diskusi tersebut terdapat permasahalan bisnis yang disimpulkan penulis yaitu:

Bagaimana menciptakan pengelompokan yang jelas terhadap UMKM di Yogyakarta, sehingga membantu pemangku kebijakan dalam mengambil keputusan dan juga wisatawan dapat mudah menemukan informasi tentang UMKM yang sesuai dengan minat mereka?.

4.1.2 *Business Objective*

Dalam konteks pengembangan potensi UMKM di Yogyakarta, khususnya yang bergerak di sektor kuliner, kerajinan tangan, dan *fashion*, diperlukan pendekatan strategis yang dapat mengelompokkan dan menyajikan informasi secara tepat

sasaran. Melihat permasalahan bisnis yang ada penulis mengidentifikasi tujuan bisnis yang ingin dicapai yaitu: Mengelompokan UMKM di Yogyakarta berdasarkan kategori seperti sektor usaha (kuliner, seni, kerajinan), dan klasifikasi usaha. Serta menyajikan pengelompokkan kategori UMKM menggunakan visualisasi sederhana

4.1.3 *Business Impact*

Pengelompokan UMKM berdasarkan kategori diharapkan tidak hanya memberikan manfaat dalam hal penyusunan data, tetapi juga berdampak langsung terhadap pengembangan sektor ekonomi kreatif dan pariwisata di Yogyakarta. Adapun dampak bisnis yang diproyeksikan dari implementasi hasil penelitian ini adalah sebagai berikut:

1. Meningkatkan daya tarik UMKM di Yogyakarta sebagai destinasi wisata berbasis ekonomi kreatif, yang pada gilirannya dapat mendorong peningkatan kunjungan wisatawan domestik dan internasional.
2. Dapat menjadi dasar bagi pemerintah daerah untuk menentukan prioritas pengembangan wilayah berbasis potensi UMKM dan minat wisatawan.

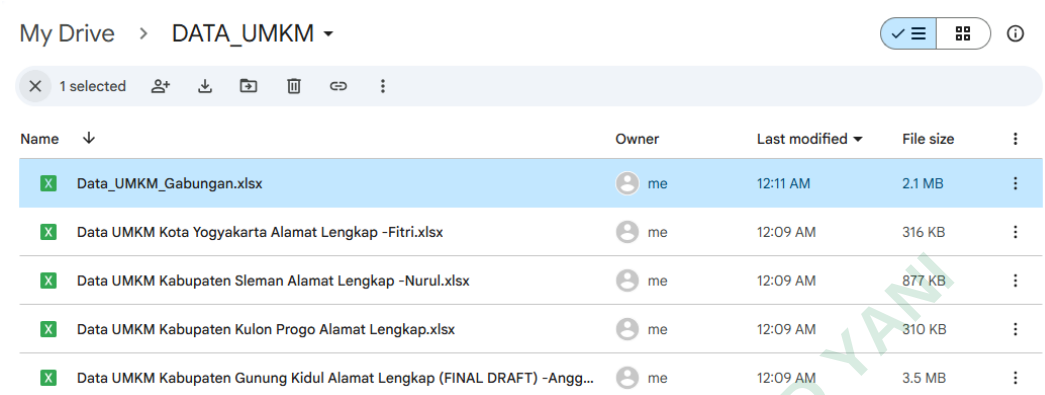
4.2 DATA UNDERSTANDING

Pada tahap *Data Understanding*, penulis melakukan eksplorasi awal terhadap data UMKM di Yogyakarta untuk memahami struktur, karakteristik, dan kualitas data yang tersedia. Proses ini mencakup pengumpulan data, eksplorasi data awal serta analisis statistik dasar dan juga distribusi kategori usaha seperti kuliner, seni, dan kerajinan, serta pemeriksaan kelengkapan informasi seperti lokasi desa atau kecamatan.

4.2.1 Pengumpulan Data

Pada tahap pengumpulan data, penulis mengumpulkan data dari sumber sekunder yang diperoleh melalui Dinas Pertanahan dan Tata Ruang Daerah Istimewa Yogyakarta yang menyediakan informasi seputar UMKM di Provinsi Daerah Istimewa Yogyakarta dimana terdapat 5 tabel data UMKM, yaitu UMKM

kota Yogyakarta, Kab. Sleman, Kab. Bantul, Kab. Kulon Progo dan juga Kab. Gunungkidul. Pada gambar 4.2 Merupakan tabel data UMKM yang ada.



Gambar 4. 2 Tabel data UMKM

Dari gambar 4.2 terlihat data UMKM belum cukup terstruktur sehingga diperlukan penggabungan antara tabel-tabel yang ada menjadi 1 tabel, untuk melakukan penggabungan tabel penulis menggunakan kode sebagai berikut:

```
import pandas as pd
import os

# Path folder di Google Drive
folder_path = '/content/drive/MyDrive/DATA_UMKM/'
# Gabungkan semua file menjadi satu DataFrame
combined_df =
pd.concat([pd.read_excel(os.path.join(folder_path, f)) for f
in file_list], ignore_index=True)
```

Setelah data UMKM berhasil digabungkan maka dilihat bahwa data tersebut berjumlah sebanyak 25.065 yang meliputi nama lengkap, jenis kelamin, disabilitas, nama usaha, kegiatan usaha, nama ekraf, nama sektor pergub, alamat usaha, nama kabupaten, kecamatan serta desa. Contoh data tersebut bisa di lihat pada Tabel 4.1

Tabel 4. 1 Contoh Data UMKM Provinsi Daerah Istimewa Yogyakarta

Nama Lengkap	Jenis Kelamin	Nama Usaha	Klasifikasi Usaha	Nama Ekraf	Alamat Usaha
Keni Astuti	Perempuan	Nasi Box/Cup Dan Snack	Mikro	Kuliner	Bausasran Dn 3/ 829 Yogyakarta

Nama Lengkap	Jenis Kelamin	Nama Usaha	Klasifikasi Usaha	Nama Ekraf	Alamat Usaha
Binti Istikomah	Perempuan	Isse Craft	Mikro	Kerajinan	Kp. Bausasran Dn.Iii/912, Rt.046 Rw.012
Dimas Septianto	Laki-Laki	Lullabic	Mikro	Fashion	Bausasran Dn.Iii/718, Rt044/011, Bausasran, Danurejan, Diy
Nasfayanti	Perempuan	Aisyahcraft	MIKRO	Kerajinan	Bausasran DN III/816
Wiraswati Yuliani	Perempuan	Pt.Lawe Adi Warna Etnika	Kecil	Kerajinan	Jl. Prof. Dr. Ki Amry Yahya No 88
Tiwik Hartyaningsih	Perempuan	Keripik Kentang	Mikro	Kuliner	Bausasran, Dn 3/594, Rt 41 Rw 11
Michael Sigit Wicaksono A.K	Laki-Laki	Keripik Gettook	Mikro	Kuliner	Jl. Dr Sutomo No. 11a, Bausasran, Danurejan, Yogyakarta
Priaji Iman Prakoso	Laki-Laki	I.P Project	Mikro	Kerajinan	Bausasran Dn/Iii, Rt 41/11
Sutini	Perempuan	Kuliner	Mikro	Kuliner	Bausasran Dn 3/643 Rt 33 Rw 09 Yg
Elisabeth Sulistyowati	Perempuan	Batik Sumoatmodjo	Mikro	Fashion	Bausasran Dn 3/773 Rt.038/Rw.010

4.2.2 Ekplorasi Data Awal

Penulis melakukan pemeriksaan menyeluruh terhadap struktur dan isi data UMKM yang telah dikumpulkan pada tahap Ekplorasi data awal ini. Langkah pertama yang dilakukan adalah memastikan bahwa data telah terhubung dengan *Google Collab* dan *Google Drive*. Untuk memastikan keterhubungan tersebut, dilakukan pemanggilan lima data awal dan 5 data terakhir dari dataset dengan kode sebagai berikut.

```
umkm.head(5)
umkm.tail(5)
```

Hasil dari pemanggilan lima data awal dan terakhir ditampilkan pada gambar 4.3 dan gambar 4.4

	nama lengkap	jenis kelamin	disabilitas	nama usaha	kegiatan usaha	klasifikasi usaha	nama ekraft
0	KENI ASTUTI	Perempuan	Tidak Ada	Nasi Box/Cup dan Snack	Penyediaan makanan dan minuman	MIKRO	Kuliner
1	Binti Istikomah	Perempuan	Tidak Ada	Isse Craft	Membuat Souvenir dari kain batik	MIKRO	Kerajinan
2	Dimas Septianto	Laki-Laki	Tidak Ada	Lullabic	Pembuatan Baju Muslim	MIKRO	Fashion
3	Nasfayanti	Perempuan	Tidak Ada	aisyahcraft	Alat makan kayu	MIKRO	Kerajinan
4	Wiraswati Yuliani	Perempuan	Tidak Ada	PT.Lawe Adi Warna Etnika	Pengembangan Produk Berbahan Lurik	KECIL	Kerajinan

Gambar 4. 3 Hasil lima data Teratas

	nama lengkap	jenis kelamin	disabilitas	nama usaha	kegiatan usaha	klasifikasi usaha	nama ekraft
25060	ROCHIMAH IMAWATI	Perempuan	Tidak Ada	YUMNA SNACK	Usaha kuliner	MIKRO	Kuliner
25061	Suratinah	Perempuan	Tidak Ada	Warung Nasi Goreng	Usaha kuliner	MIKRO	Kuliner
25062	Jatiyem	Perempuan	Tidak Ada	Usaha kuliner	Usaha kuliner	MIKRO	Kuliner
25063	ENY MARWANTI	Perempuan	Tidak Ada	Usaha kuliner	Usaha kuliner	KECIL	Kuliner
25064	Murtini	Perempuan	Tidak Ada	Usaha kuliner	Usaha kuliner	MIKRO	Kuliner

Gambar 4. 4 Hasil lima data Terakhir

Dari hasil gambar 4.3 dan 4.4 menampilkan 5 data teratas dan terbawah yang menandakan bahwa data telah terhubung antara *Google Collab* dan juga *Google Drive* dan siap di lakukan untuk proses selanjutnya.

Kemudian pada tahap eksplorasi data awal juga kita bisa melihat nama kolom yang ada dengan kode sebagai berikut.

```
kolom_df = pd.DataFrame(umkm.columns, columns=["Nama Kolom"])
kolom_df
```

pada Tabel 4.2 merupakan nama beserta penjelasan tiap kolom.

Tabel 4. 2 Nama kolom Pada Data UMKM

No	Nama Kolom	Penjelasan
0	Nama Lengkap	Nama lengkap pemilik UMKM atau pelaku usaha.
1	Jenis Kelamin	Jenis kelamin pemilik usaha (Laki-laki atau Perempuan).
2	Disabilitas	Status apakah pemilik usaha merupakan penyandang disabilitas atau tidak.
3	Nama Usaha	Nama dari usaha/brand yang dimiliki.
4	Kegiatan Usaha	Aktivitas atau jenis bisnis yang dijalankan (contoh: kuliner, fashion).
5	Klasifikasi Usaha	Kategori ukuran usaha, misalnya Mikro, Kecil, Menengah.
6	Nama Ekraf	Nama sub-sektor ekonomi kreatif yang sesuai, seperti kerajinan, desain.
7	Nama Sektor Pergub	Nama sektor ekonomi berdasarkan peraturan gubernur (spesifik daerah).
8	Alamat Usaha	Alamat lengkap lokasi usaha dijalankan.
9	Nama Kabupaten	Nama kabupaten tempat usaha berada.
10	Nama Kecamatan	Nama kecamatan tempat usaha berada.
11	Nama Desa	Nama desa atau kelurahan tempat usaha berada.

Pada tahap eksplorasi data awal, juga dilakukan analisis untuk memahami struktur dan karakteristik dari dataset yang digunakan. Salah satu langkah awal yang dilakukan adalah dengan melihat informasi dasar dari setiap kolom yang terdapat pada data. Informasi ini mencakup nama kolom, tipe data, serta jumlah data yang tersedia (*non-null*) pada masing-masing kolom. Pada Gambar 4.5,

ditampilkan hasil informasi dari seluruh kolom dalam data UMKM yang berasal dari Provinsi Daerah Istimewa Yogyakarta. Informasi ini memberikan gambaran umum mengenai struktur dataset, seperti apakah kolom bertipe data numerik, teks (*object*), atau kategori lainnya.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25065 entries, 0 to 25064
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   nama lengkap        25065 non-null  object
 1   jenis kelamin       25064 non-null  object
 2   disabilitas         24914 non-null  object
 3   nama usaha          25044 non-null  object
 4   kegiatan usaha      25049 non-null  object
 5   klasifikasi usaha   25065 non-null  object
 6   nama ekraft         25065 non-null  object
 7   nama sektor pergub  25065 non-null  object
 8   alamat usaha        25065 non-null  object
 9   nama kabupaten      25065 non-null  object
 10  nama kecamatan      25065 non-null  object
 11  nama desa           25065 non-null  object
```

Gambar 4. 5 Informasi Dataset UMKM

4.2.3 Analisis Statistik Dasar

Analisis statistik dasar adalah tahap awal dalam eksplorasi data yang bertujuan untuk memahami karakteristik umum dari dataset yang akan dianalisis. Melalui analisis ini, kita dapat memperoleh informasi penting seperti jumlah data (*count*), nilai minimum dan maksimum, nilai rata-rata (*mean*), standar deviasi, serta distribusi data melalui kuartil (Q1, median/Q2, Q3). Selain itu, untuk data kategorikal, statistik dasar mencakup jumlah kategori unik, kategori yang paling sering muncul (*modus*), dan frekuensinya.

Tahap ini sangat penting karena membantu mengidentifikasi pola awal dalam data, mendeteksi potensi anomali atau data kosong (*missing values*), serta memberikan gambaran mengenai penyebaran dan kecenderungan data. Dengan memahami struktur dan sifat dasar data, analis dapat menentukan metode yang tepat untuk pengolahan lebih lanjut, seperti teknik pembersihan data (*data cleaning*),

transformasi data, hingga pemodelan prediktif, untuk melakukan tahapan tersebut kitab isa menggunakan kode sebagai berikut.

```
umkm.describe()
```

Pada Gambar 4.6 merupakan hasil statistic dasar dari dataset UMKM provinsi Daerah Istimewa Yogyakarta

	nama lengkap	jenis kelamin	disabilitas	nama usaha	kegiatan usaha	klasifikasi usaha	nama ekraft	nama sektor pergub	alamat usaha	nama kabupaten	nama kecamatan	nama desa
count	25065	25064	24914	25044	25049	25065	25065	25065	25065	25065	25065	25065
unique	18432	2	5	19902	6104	3	4	16	13218	5	77	415
top	SRI LESTARI	Perempuan	Tidak Ada	Usaha kuliner	Usaha kuliner	MIKRO	Kuliner	Industri Pengolahan	PARANGTRITIS	KABUPATEN BANTUL	PANDAK	PARANGTRITIS
freq	43	15258	24867	771	4864	21865	15130	8746	234	10579	887	435

Gambar 4. 6 Hasil Statistik Dasar data UMKM

Dari tabel di atas diketahui :

1. Nama Lengkap: Ada 25.065 data nama, dengan 18.432 nama unik. Nama terbanyak adalah Sri Lestari sebanyak 43 kali.
2. Jenis Kelamin: Ada 25.064 data, terdiri dari 3 kategori. Mayoritas adalah Perempuan dengan 15.258 data.
3. Nama Usaha: Tercatat 25.044 data nama usaha, dengan 19.902 nama unik. Nama usaha yang paling sering muncul adalah Usaha kuliner sebanyak 771 kali.
4. Klasifikasi Usaha: Terdapat 25.065 data klasifikasi usaha, dengan 3 kategori. Klasifikasi Mikro mendominasi dengan 21.865 data.
5. Nama Ekraf: Ada 25.065 data nama ekonomi kreatif, terbagi dalam 4 kategori. Kuliner menjadi kategori terbanyak dengan 15.130 data.
6. Alamat Usaha: Terdapat 25.065 data alamat usaha, dengan 13.218 alamat unik. Alamat terbanyak adalah Parangtritis sebanyak 234 kali.
7. Nama desa : terdapat 25.065 data Nama desa, dengan 415 nama unik, nama desa terbanyak adalah Parangtritis sebanyak 435 kali.

Tidak adanya Q1, Q2, atau Q3 pada data tersebut disebabkan oleh sifat data yang bersifat kategorikal, seperti nama usaha, jenis kelamin, dan nama desa. Kuartil

(Q1, Q2, Q3) digunakan untuk data numerik yang dapat diurutkan, seperti harga, usia, atau jumlah, dan membagi data tersebut menjadi empat bagian yang hampir sama besar. Dalam perhitungan kuartil, data harus terlebih dahulu diurutkan secara numerik, kemudian dibagi menjadi empat bagian, yang masing-masing mencakup 25% dari total data. Namun, variabel-variabel kategorikal seperti nama usaha atau jenis kelamin tidak memiliki urutan numerik yang bisa diurutkan, sehingga tidak bisa diterapkan perhitungan kuartil pada data tersebut. Sebagai contoh, untuk data seperti "jenis kelamin" atau "nama desa", kita hanya bisa menghitung frekuensi kemunculan tiap kategori, bukan menghitung posisi kuartil. Oleh karena itu, Q1, Q2, dan Q3 hanya berlaku untuk data numerik yang terurut dan tidak dapat diterapkan pada data kategorikal.

4.3 DATA PRE-PARATIONS

Pada tahap ini, dilakukan proses *data preparation* yang bertujuan untuk memastikan data yang digunakan dalam penelitian sudah bersih, konsisten, dan siap untuk dianalisis. Langkah-langkah yang dilakukan meliputi pemeriksaan data hilang (*missing values*), pembersihan data dari duplikasi, serta standarisasi format data agar seragam. Selain itu, dilakukan juga pengolahan awal seperti *encoding* untuk data kategorikal dan normalisasi data numerik jika diperlukan. Proses ini penting untuk meminimalisir potensi kesalahan analisis dan meningkatkan akurasi hasil yang akan diperoleh pada tahap berikutnya.

Setelah data dibersihkan, dilakukan pemilihan fitur (*feature selection*) untuk menentukan variabel-variabel yang relevan dengan tujuan penelitian. Fitur yang kurang berpengaruh atau mengandung banyak *noise* dihapus agar model yang dibangun menjadi lebih optimal. Dalam beberapa kasus, dilakukan juga teknik rekayasa fitur (*feature engineering*) untuk menciptakan variabel baru yang dapat meningkatkan kinerja analisis. Seluruh proses data preparation ini mendukung tercapainya hasil penelitian yang valid, terukur, dan dapat diandalkan.

4.3.1 *Cleaning Data*

Proses *cleaning data* atau pembersihan data dilakukan untuk memastikan bahwa data yang digunakan dalam analisis berada dalam kondisi yang bersih,

konsisten, dan siap untuk diproses lebih lanjut. Pembersihan data merupakan langkah krusial dalam *data preprocessing* karena data mentah sering kali mengandung nilai kosong (*missing values*), data duplikat, atau kesalahan entri lainnya yang dapat memengaruhi akurasi hasil analisis. Beberapa langkah yang dilakukan dalam proses *cleaning data* antara lain:

1. Handling Missing Value

Data di periksa untuk mengetahui apakah terdapat nilai kosong pada kolom tertentu menggunakan fungsi

```
missing_data = umkm.isnull().sum()
missing_table = pd.DataFrame({
    'Kolom': missing_data.index,
    'Jumlah Nilai Hilang': missing_data.values
})

missing_table
```

Hasilnya ditampilkan dalam bentuk tabel agar memudahkan identifikasi kolom mana saja yang memiliki jumlah nilai kosong. Kolom atau baris dengan nilai kosong kemudian dianalisis lebih lanjut apakah perlu di hapus atau di isi dengan nilai tertentu, tergantung konteks dan relevansi data tersebut. Pada Gambar 4.7 merupakan hasil dari pemeriksaan data kosong dari data UMKM Provinsi Daerah Istimewa Yogyakarta.

	Kolom	Jumlah Nilai Hilang
0	nama lengkap	0
1	jenis kelamin	1
2	disabilitas	151
3	nama usaha	21
4	kegiatan usaha	16
5	klasifikasi usaha	0
6	nama ekraft	0
7	nama sektor pergub	0
8	alamat usaha	0
9	nama kabupaten	0
10	nama kecamatan	0
11	nama desa	0

Gambar 4. 7 Hasil *Missing Value*

Dari gambar di atas kita bisa lihat terdapat *missing value* (data kosong) pada kolom jenis kelamin sebanyak 1 data, kolom disabilitas sebanyak 151 data, kolom nama usaha sebanyak 21 data dan juga kolom kegiatan usaha sebanyak 16 data. *Missing values* tersebut akan dihapus semuanya dikarenakan data bersifat kategorikal yang artinya tidak bisa di isi oleh nilai rata-rata ataupun nilai tengah, untuk penghapusan *missing value* bisa menggunakan kode sebagai berikut.

```
# Load data dan ubah nilai aneh jadi NaN
umkm = pd.read_excel('/content/drive/MyDrive/dataset1/DATA
UMKM FULL.xlsx', na_values=[' ', 'NA', 'null', '-'])

# Hapus baris yang mengandung NaN
umkm = umkm.dropna()

umkm.isnull().sum()
```

setelah dilakukannya penghapusan *missing value* dan juga pengecekan setelah penghapusan maka hasil tersebut dapat dilihat pada gambar 4.8

Gambar 4. 8 Hasil setelah penghapusan Missing Value

Dari gambar di atas bisa di lihat bahwa sudah tidak ada data yang kosong di semua kolom yang ada pada data UMKM, data yang semulanya berjumlah 25.065 menjadi 24.803 setelah penghapusan *missing value* (data yang kosong) dan sudah bisa lanjut ke tahap pengecekan duplikat data.

2. Handling Duplikat Value

Penulis melakukan penanganan nilai duplikat untuk memastikan kualitas dan keakuratan data yang akan digunakan dalam analisis. Duplikat dapat menyebabkan bias atau distorsi hasil, terutama jika data yang sama dihitung lebih dari sekali. Oleh karena itu, langkah yang diambil meliputi pemeriksaan data menggunakan fungsi `umkm.duplicated().sum() #Chek duplikate value` untuk mengidentifikasi baris-baris yang memiliki informasi identik pada seluruh kolom. Setelah itu, baris duplikat yang terdeteksi dihapus menggunakan fungsi `drop_duplicates()`. Proses ini penting agar analisis yang dilakukan selanjutnya, seperti clustering atau visualisasi data, dapat mencerminkan

kondisi UMKM secara lebih valid dan representatif. Pada gambar 4.9 merupakan hasil pengecekan *duplikat value*.

	nama lengkap	jenis kelamin	disabilitas	nama usaha	kegiatan usaha	klasifikasi usaha	nama ekraft
1936	MUJIYO	Laki-laki	Tidak Ada	MUJIYO	Makanan siap saji	MIKRO	Kuliner
2080	MUSILAH	Perempuan	Tidak Ada	MUSILAH	Makanan	MIKRO	Kuliner
2513	DWI RIANDINI ANGGITYA BUDI	Laki-laki	Tidak Ada	WARUNG AYAM GORENG	Ayam goreng	MIKRO	Kuliner
4769	MUJILAH	Perempuan	Tidak Ada	WARUNG MAKAN	Makanan & minuman	KECIL	Kuliner
5305	SUROSO	Laki-Laki	Tidak Ada	SUROSO	Anyaman bambu	KECIL	Kerajinan
9142	Jumiyem	Perempuan	Tidak Ada	Sembung batik	IKM Batik	MIKRO	Fashion
9159	Murtini	Perempuan	Tidak Ada	Sembung batik	IKM Batik	MIKRO	Fashion
10743	Marnoto	Laki-laki	Tidak Ada	Aneka anyaman Marnoto	Kerajinan anyaman	MIKRO	Kerajinan
10751	Sugirah	Perempuan	Tidak Ada	Aneka makanan Sugirah	Penyediaan makanan dan minuman	MIKRO	Kuliner
11352	WASITO	Laki-laki	Tidak Ada	Pengrajin Bambu WASITO	Kerajinan bambu/caping/tambir	MIKRO	Kerajinan
24101	YITNO UTOMO	Laki-laki	Tidak Ada	MEMBUAT KRENENG	Usaha kerajinan tangan	MIKRO	Kerajinan
24170	SAMIYONO	Laki-laki	Tidak Ada	MEMBUAT KRENENG	Usaha kerajinan tangan	MIKRO	Kerajinan
24199	DARMI	Perempuan	Tidak Ada	MEMBUAT KRENENG	Usaha kerajinan tangan	MIKRO	Kerajinan

Gambar 4. 9 Hasil data duplikat

Dikarenakan terdapat data duplikat yang berjumlah 13 data maka data duplikat tersebut akan kita hapus menggunakan kode

```
# Menghapus duplikat
umkm = umkm.drop_duplicates()
```

setelah melakukan penghapusan data duplikat data yang semulanya berjumlah 24.803 menjadi 24.790, kemudian penulis Kembali melakukan pengecekan terhadap data UMKM untuk melihat apakah masih terdapat data duplikat, pada gambar 4.10 merupakan hasil pengecekan duplikat.

```
np.int64(0)
```

Gambar 4. 10 Hasil Pengecekan Duplikat

Dari gambar tersebut terlihat jelas bahwa duplikat value pada data UMKM sudah tidak ada dan siap dilanjutkan ke tahap selanjutnya.

4.3.2 *Data Transformation*

Data transformation merupakan tahap penting dalam proses persiapan data yang bertujuan untuk mengubah data mentah menjadi format yang sesuai dan optimal untuk analisis lebih lanjut maupun pemodelan *machine learning*. Transformasi data dilakukan untuk meningkatkan kualitas data, menyederhanakan struktur, serta memperbaiki skala nilai sehingga model dapat memahami pola dalam data secara lebih efektif.

1. Encoder variable kategorikal

Penulis melakukan proses *encoding* terhadap variabel kategorikal guna mengubah data non-numerik menjadi format numerik yang dapat diproses oleh algoritma *machine learning*. Variabel kategorikal seperti jenis kelamin, nama ekraf, kategori usaha tidak dapat digunakan secara langsung dalam analisis karena bersifat string atau label. Oleh karena itu, diperlukan transformasi agar nilai-nilai tersebut dapat direpresentasikan secara numerik tanpa menghilangkan informasi yang terkandung di dalamnya.

Metode *encoding* yang digunakan meliputi *Label Encoding* untuk variabel dengan jumlah kategori terbatas dan tidak memiliki hirarki khusus, serta *One-Hot Encoding* untuk variabel dengan banyak kategori dan bersifat nominal. Langkah ini menjadi bagian penting dalam tahapan pra-pemrosesan data sebelum dilakukan analisis lebih lanjut. Untuk melakukan *Label encoding* kitab isa menggunakan code seperti berikut untuk kolom jenis kelamin

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
umkm["jenis_kelamin_encoded"] = le.fit_transform(umkm["jenis
kelamin"])
```

Pada gambar 4.11 merupakan hasil *label encoding* dimana terdapat penambahan kolom dengan nama” jenis kelamin encoded”.

	nama lengkap	jenis kelamin	kegiatan usaha	klasifikasi usaha	nama desa	jenis_kelamin_encoded
0	KENI ASTUTI	Perempuan	Penyediaan makanan dan minuman	MIKRO	BAUSASRAN	1
1	Binti Istikomah	Perempuan	Membuat Souvenir dari kain batik	MIKRO	BAUSASRAN	1
2	Dimas Septianto	Laki-Laki	Pembuatan Baju Muslim	MIKRO	BAUSASRAN	0
3	Nasfayanti	Perempuan	Alat makan kayu	MIKRO	BAUSASRAN	1
4	Wiraswati Yuliani	Perempuan	Pengembangan Produk Berbahan Lurik	KECIL	BAUSASRAN	1
...	...	...	...	...	...	...
25060	ROCHIMAH IMAWATI	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1
25061	Suratinah	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1
25062	Jatiyem	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1
25063	ENY MARWANTI	Perempuan	Usaha kuliner	KECIL	TRIMURTI	1
25064	Murtini	Perempuan	Usaha kuliner	MIKRO	TRIRENGGO	1

24790 rows × 6 columns

Gambar 4. 11 Hasil Label *encoding* kolom jenis kelamin

Kemudian setelah melakukan *label encoding* terhadap kolom jenis kelamin selanjutnya kita menggunakan *label encoding* terhadap kolom klasifikasi usaha, pada gambar 4.12 merupakan hasil *label encoding* terhadap kolom klasifikasi usaha dimana terdapat kolom baru Bernama"klasifikasi_encoded".

	nama lengkap	jenis kelamin	kegiatan usaha	klasifikasi usaha	nama desa	jenis_kelamin_encoded	klasifikasi_encoded
0	KENI ASTUTI	Perempuan	Penyediaan makanan dan minuman	MIKRO	BAUSASRAN	1	0
1	Binti Istikomah	Perempuan	Membuat Souvenir dari kain batik	MIKRO	BAUSASRAN	1	0
2	Dimas Septianto	Laki-Laki	Pembuatan Baju Muslim	MIKRO	BAUSASRAN	0	0
3	Nasfayanti	Perempuan	Alat makan kayu	MIKRO	BAUSASRAN	1	0
4	Wiraswati Yuliani	Perempuan	Pengembangan Produk Berbahan Lurik	KECIL	BAUSASRAN	1	1
...	...	...	...	...	...	...	...
25060	ROCHIMAH IMAWATI	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1	0
25061	Suratinah	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1	0
25062	Jatiyem	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1	0
25063	ENY MARWANTI	Perempuan	Usaha kuliner	KECIL	TRIMURTI	1	1
25064	Murtini	Perempuan	Usaha kuliner	MIKRO	TRIRENGGO	1	0

24790 rows × 7 columns

Gambar 4. 12 Hasil Label *Encoding* kolom Klasifikasi usaha

Selanjutnya penulis melakukan proses *Label encoding* pada kolom nama ekraf dengan kode sebagai berikut

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
umkm["ekraft_encoded"] = le.fit_transform(umkm["nama
ekraft"])
```

hasil dari *Label encoding* tersebut bisa dilihat pada gambar 4.13

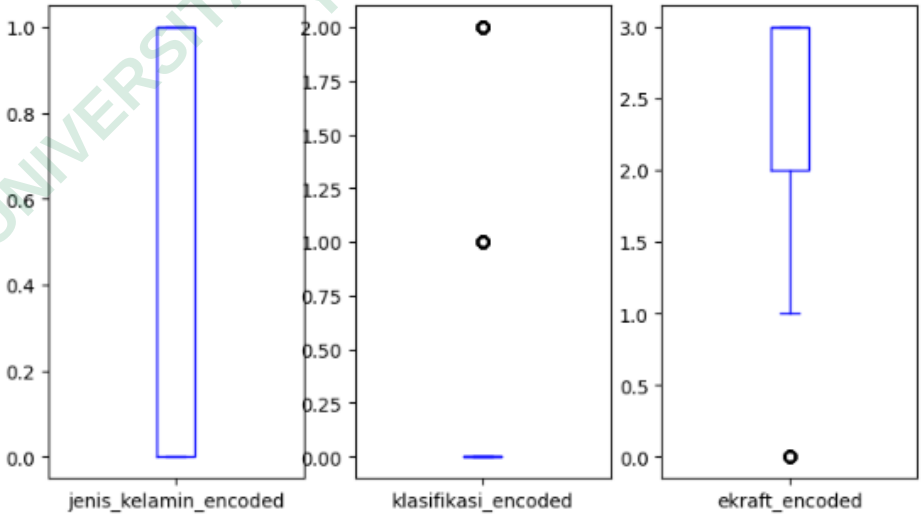
	nama lengkap	jenis kelamin	kegiatan usaha	klasifikasi usaha	nama desa	jenis_kelamin_encoded	klasifikasi_encoded	ekraft_encoded
0	KENI ASTUTI	Perempuan	Penyediaan makanan dan minuman	MIKRO	BAUSASRAN	1	0	3
1	Binti Istikomah	Perempuan	Membuat Souvenir dari kain batik	MIKRO	BAUSASRAN	1	0	2
2	Dimas Septianto	Laki-Laki	Pembuatan Baju Muslim	MIKRO	BAUSASRAN	0	0	1
3	Nasfayanti	Perempuan	Alat makan kayu	MIKRO	BAUSASRAN	1	0	2
4	Wiraswati Yuliani	Perempuan	Pengembangan Produk Berbahan Lurik	KECIL	BAUSASRAN	1	1	2
...	...	...	...	...	...	...	...	...
25060	ROCHIMAH IMAWATI	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1	0	3
25061	Suratinah	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1	0	3
25062	Jatiyem	Perempuan	Usaha kuliner	MIKRO	TRIMURTI	1	0	3
25063	ENY MARWANTI	Perempuan	Usaha kuliner	KECIL	TRIMURTI	1	1	3
25064	Murtini	Perempuan	Usaha kuliner	MIKRO	TRIRENGGO	1	0	3

24790 rows × 8 columns

Gambar 4. 13 Hasil label *encoding* kolom nama ekraf

2. *Hendling Outlier*

Setelah dilakukan proses *encoding* terhadap variabel kategorikal, tahapan selanjutnya adalah melakukan pengecekan *outlier*. Pada tahap ini, penulis melakukan pengecekan *outlier* untuk mengidentifikasi adanya nilai ekstrem atau data yang menyimpang secara signifikan dari distribusi mayoritas. Meskipun *outlier* umumnya lebih relevan diterapkan pada data numerik, dalam penelitian ini pengecekan juga dilakukan terhadap variabel kategorikal yang telah diubah ke bentuk numerik, seperti variabel jenis kelamin**,** klasifikasi usaha**,** dan nama ekraf, guna memastikan tidak terdapat nilai yang tidak wajar hasil dari proses *encoding*. Pada gambar 4.14 merupakan visualisasi pengecekan *outlier* pada variable.



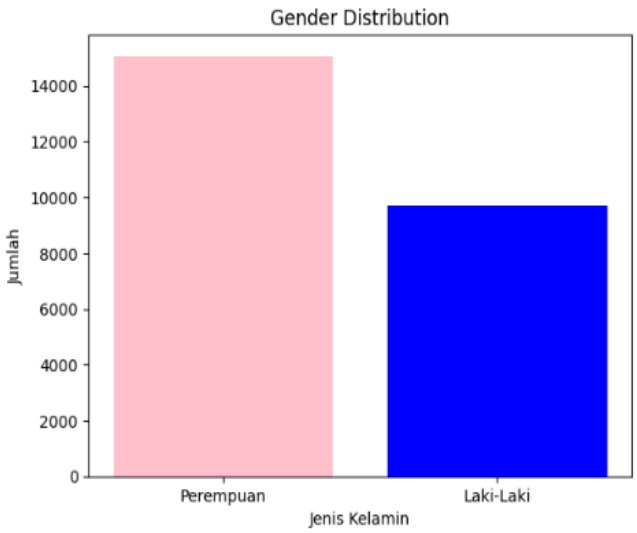
Gambar 4. 14 visualisasi *outlier* variabel

4.3.3 *EDA(Exploratory Data Analysis)*

Exploratory Data Analysis (EDA) merupakan tahap awal yang penting dalam proses analisis data. EDA bertujuan untuk memahami struktur, karakteristik, serta pola-pola yang terdapat dalam dataset. Melalui EDA, peneliti dapat mengidentifikasi distribusi data, nilai yang hilang, pencilan (*outlier*), serta hubungan antar variabel. Analisis ini menjadi landasan dalam pengambilan keputusan terhadap proses pra-pemrosesan data selanjutnya, serta memberikan gambaran awal terhadap potensi informasi yang dapat diperoleh dari data. Pada tahap ini, panulis menggunakan berbagai teknik visualisasi dan statistik deskriptif guna mempermudah interpretasi data.

1. **Distribusi data jenis kelamin**

   Distribusi data merupakan gambaran penyebaran nilai-nilai dalam suatu variabel. Melalui analisis distribusi, peneliti dapat mengetahui bagaimana data tersebar, apakah simetris, condong ke kiri (*left-skewed*), atau condong ke kanan (*right-skewed*). Pemahaman terhadap distribusi sangat penting untuk menentukan jenis transformasi data yang tepat serta untuk memilih metode analisis yang sesuai. Selain itu, distribusi data juga membantu dalam mengidentifikasi adanya *outlier* atau nilai ekstrem yang dapat memengaruhi hasil analisis. Pada gambar 4.15 merupakan hasil visualisasi distribusi data kolom jenis kelamin.

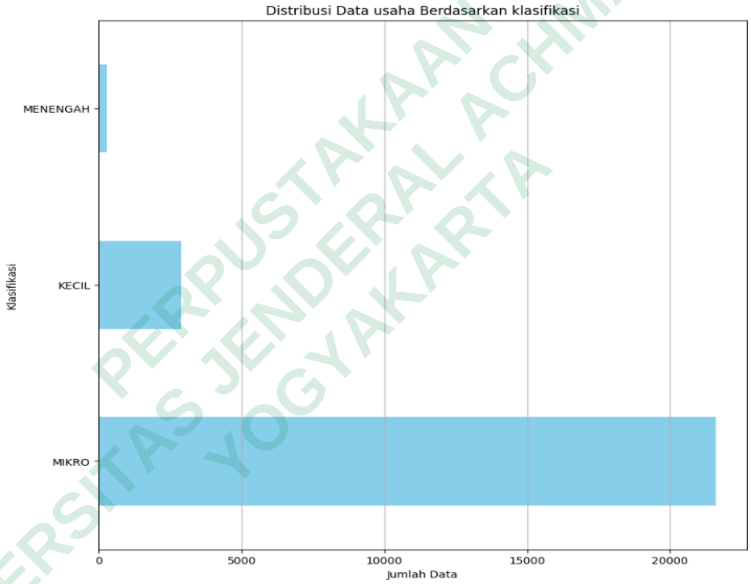


Gambar 4. 15 Distribusi data kolom jenis kelamin

Pada gambar di atas bisa terlihat bahwa jumlah pemilik UMKM dengan jenis kelamin Perempuan lebih banyak dibandingkan Laki-laki. Dimana sebanyak 15.072 data dengan jenis kelamin perempuan dan 9.718 jumlah data laki-laki.

2. Distribusi data klasifikasi usaha

Penulis melakukan analisis distribusi data terhadap klasifikasi usaha untuk mengetahui sebaran jumlah UMKM berdasarkan jenis klasifikasinya. Langkah ini bertujuan untuk memahami proporsi masing-masing kategori usaha dalam dataset, serta melihat apakah terdapat dominasi pada jenis usaha tertentu. Pada gambar 4.16 merupakan visualisasi distribusi Klasifikasi usaha.



Gambar 4. 16 Visualisasi distribusi data Klasifikasi Usaha

Dari hasil visualisasi tersebut terlihat bahwa klasifikasi usaha Mikro mendominasi dimana sebanyak 21.608 data dan menyusul klasifikasi Kecil dengan 2.882 data dan yang terakhir menengah sebanyak 300 data.

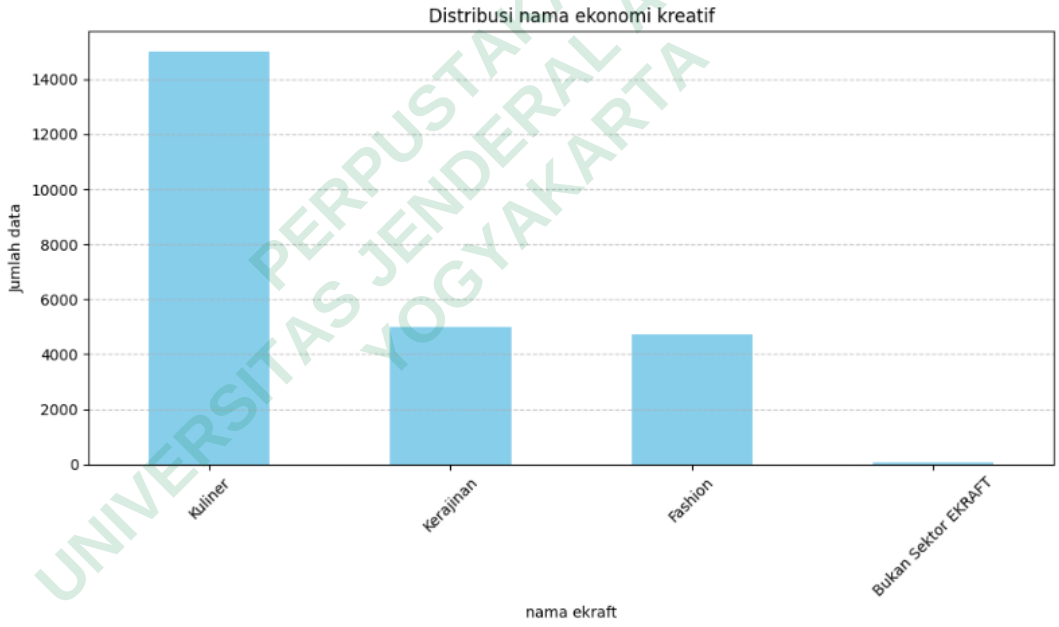
3. Distribusi data ekraf

Penulis melakukan analisis distribusi data ekonomi kreatif (ekraf) untuk mengetahui penyebaran jenis usaha kreatif yang terdapat dalam dataset UMKM. Analisis ini membantu mengidentifikasi sektor-sektor ekraf yang

paling dominan serta memberikan gambaran awal mengenai potensi ekonomi kreatif yang ada di wilayah tersebut menggunakan kode sebagai berikut.

```
ekraft_count= umkm['nama ekraft'].value_counts()

# Plot bar chart
plt.figure(figsize=(10, 6))
ekraft_count.plot(kind='bar', color='skyblue')
plt.title('Distribusi nama ekonomi kreatif')
plt.xlabel('nama ekraft')
plt.ylabel('Jumlah data')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Pada gambar 4.17 merupakan visualisasi distribusi data ekraf.



Gambar 4. 17 Visualisasi Distribusi data ekraf

Dari hasil visualisasi tersebut dapat diketahui bahwa ekonomi kreatif kuliner mendominasi data UMKM dimana sebanyak 14.989 data kemudian di susul dengan ekraf kerajinan sebanyak 5.012 data dan juga *fashion* sebanyak 4.705 data dan yang terakhir bukan sektor ekraf sebanyak 84 data.

4.3.4 *Data Reduction*

Penulis melakukan *data reduction* sebagai upaya untuk menyederhanakan jumlah fitur dalam dataset tanpa menghilangkan informasi yang relevan. Tahapan ini bertujuan untuk mengurangi kompleksitas data, meningkatkan efisiensi proses analisis, serta meminimalisir risiko *overfitting* saat membangun model. *Data reduction* menjadi penting terutama ketika terdapat fitur yang tidak terlalu berpengaruh terhadap variabel target atau memiliki korelasi tinggi dengan fitur lainnya. Pada tahap ini penulis melakukan *feature selection* dimana memilih fitur yang nantinya akan berguna ketika melakukan pengelompokkan, pada gambar 4.18 merupakan hasil variable yang sudah dilakukan *feature selection*.

	jenis_kelamin_encoded	klasifikasi_encoded	ekraft_encoded	nama desa
0	1	0	3	BAUSASRAN
1	1	0	2	BAUSASRAN
2	0	0	1	BAUSASRAN
3	1	0	2	BAUSASRAN
4	1	1	2	BAUSASRAN
...	...	...	...	...
25060	1	0	3	TRIMURTI
25061	1	0	3	TRIMURTI
25062	1	0	3	TRIMURTI
25063	1	1	3	TRIMURTI
25064	1	0	3	TRIRENGGO

24790 rows × 4 columns

Gambar 4. 18 Hasil Data Selection

Pada gambar di atas terlihat jelas bahwa variable yang ada pada data UMKM tersisa hanya beberapa variable diantaranya, jenis_kelamin_encoded, klasifikasi_encoded, ekraft_encoded dan juga yang tarakhir nama desa. Setelah melakukan data reduction Langkah selanjutnya yaitu Tahap modeling.

penelitian ini menghasilkan data yang lebih bersih, konsisten, dan siap digunakan dibandingkan dengan penelitian sebelumnya. Seluruh data telah melalui

tahapan pembersihan terhadap nilai duplikat, penanganan *missing values*, serta deteksi dan penyesuaian terhadap nilai *outlier* menggunakan metode seperti IQR. Tidak hanya itu, transformasi data juga dilakukan melalui *label encoding* pada kolom kategorikal agar sesuai dengan kebutuhan algoritma. Berbeda dengan beberapa studi terdahulu yang hanya melakukan pembersihan dasar dengan pengecekan *missing value*

4.4 MODELING

Penulis melakukan tahap modeling untuk mengelompokkan data kategori UMKM berdasarkan variabel-variabel yang telah melalui proses preprocessing, yaitu jenis kelamin, klasifikasi usaha, subsektor ekonomi kreatif (ekraf), dan nama desa. Pada tahap ini digunakan dua algoritma *unsupervised learning*, yaitu *K-Means Clustering* dan *Agglomerative clustering* . Keduanya merupakan metode yang umum digunakan dalam pengelompokan data tanpa label, dengan pendekatan yang berbeda dalam membentuk klaster.

Pemilihan kedua algoritma tersebut dilakukan untuk membandingkan hasil pengelompokan berdasarkan karakteristik data kategori UMKM. *K-Means* bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan kedekatan terhadap *centroid* tertentu, sedangkan *Agglomerative clustering* menggunakan pendekatan hierarki dengan menggabungkan data secara bertahap berdasarkan kemiripan antar data. Perbandingan ini bertujuan untuk mengevaluasi metode mana yang lebih efektif dalam memberikan representasi klaster yang sesuai dengan kondisi dan distribusi kategori UMKM yang dianalisis.

4.4.1 *K-Means* Clustering

Pada penelitian ini, algoritma *K-Means Clustering* digunakan sebagai salah satu metode pengelompokan data. *K-Means* merupakan algoritma *unsupervised learning* yang bertujuan untuk membagi sekumpulan data ke dalam sejumlah kelompok (cluster) berdasarkan kemiripan karakteristik. Setiap data akan ditempatkan ke dalam cluster dengan jarak terdekat terhadap pusat cluster (*centroid*) yang dihitung secara iteratif. Proses awal algoritma dimulai dengan

menentukan jumlah cluster (k) dan memilih *centroid* awal secara acak dengan menggunakan rumus persamaan **(1)**.

Selanjutnya, data akan dikelompokkan ke dalam cluster berdasarkan jarak terdekat ke *centroid*, kemudian posisi *centroid* akan diperbarui berdasarkan rata-rata dari data dalam masing-masing *cluster*. Proses ini akan terus berulang hingga tidak ada perubahan signifikan pada posisi *centroid* atau jumlah iterasi maksimum tercapai, kemudian penulis melakukan pengecekan *elbow* sebagai dasar penentuan *centroid*(k) yang nantinya digunakan untuk clustering dengan kode sebagai berikut:

```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

# Buat ulang crosstab awal (bersih dari kolom tambahan)
ekraf_per_desa_elbow = pd.crosstab(umkm['nama desa'],
umkm['nama ekraft']).reset_index()
X_elbow = ekraf_per_desa_elbow.drop(columns=['nama desa'])

# Cek Elbow Method
inertia = []
K = range(1, 11)

for k in K:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(X_elbow)
    inertia.append(kmeans.inertia_)

# Plot Elbow
plt.figure(figsize=(8, 5))
plt.plot(K, inertia, 'bo-', label='Inertia (WCSS)')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Inertia (WCSS)')
plt.title('Elbow Method untuk Menentukan K Optimal')

# Tambahkan garis merah vertikal di k=3 (atau ganti sesuai
elbow sebenarnya)
plt.axvline(x=3, color='red', linestyle='--', label='k = 3')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```


Untuk hasil pengecekan *elbow* dapat dilihat pada gambar 4.19

Gambar 4. 19 *Elbow* metode *K-Means* Clustering

Pada gambar 4.19 menunjukkan hasil penerapan *Elbow Method* untuk menentukan jumlah cluster (k) yang optimal dalam algoritma *K-Means*. Terlihat bahwa nilai inertia atau *Within-Cluster Sum of Squares* (WCSS) mengalami penurunan tajam dari k = 1 hingga k = 3, yang mengindikasikan peningkatan kualitas klasterisasi. Namun, setelah k = 3, penurunan nilai mulai melambat dan tidak signifikan. Titik ini disebut sebagai "*elbow*" atau siku, yang menandakan jumlah cluster optimal karena setelah titik ini, menambah jumlah *cluster* tidak memberikan peningkatan yang berarti. Oleh karena itu, berdasarkan grafik, jumlah cluster yang paling ideal untuk digunakan adalah tiga (k = 3) yang akan digunakan untuk melakukan clustering terhadap ekonomi kreatif.

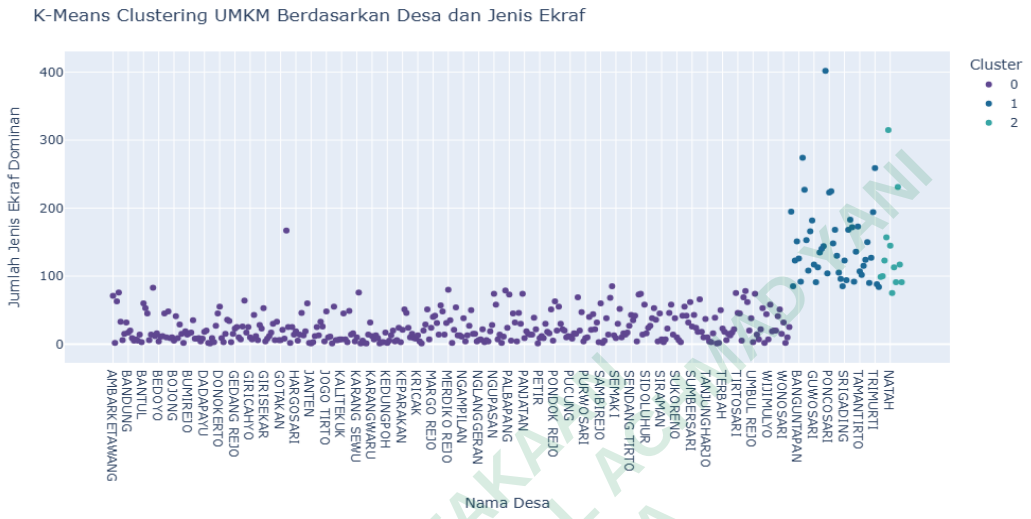
K-Means Clustering Ekonomi Kreatif

K-Means Clustering digunakan dalam analisis data ekonomi kreatif untuk mengelompokkan desa-desa berdasarkan karakteristik pelaku usaha kreatif yang ada di wilayah tersebut. Metode ini memungkinkan identifikasi pola dan pengelompokan desa dengan potensi ekonomi kreatif yang serupa, berdasarkan atribut numerik yang telah dikodekan sebelumnya. Dengan membagi data ke dalam beberapa cluster, hasil analisis ini diharapkan dapat memberikan gambaran wilayah

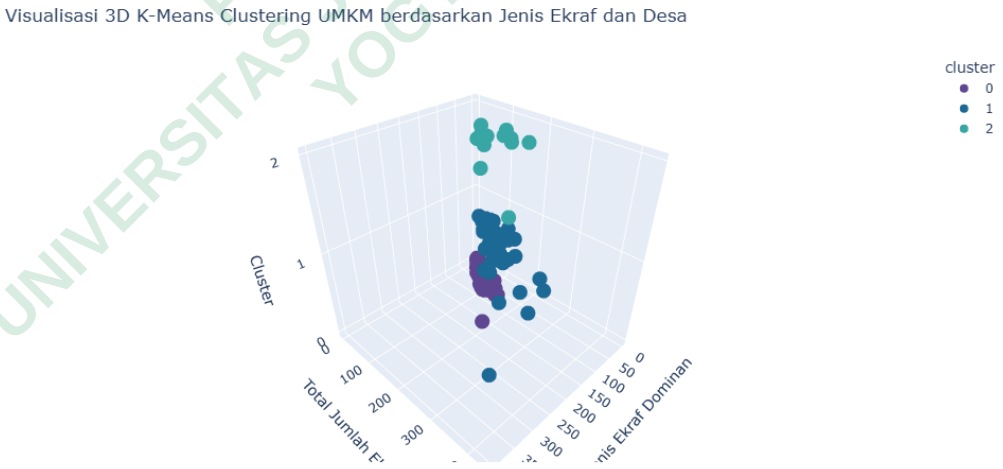
dengan konsentrasi jenis usaha kreatif tertentu, yang nantinya dapat digunakan sebagai dasar dalam membantu pemangku kepentingan dan juga wisatawan dalam melihat kategori UMKM yang sesuai dengan minat mereka. Dalam pengelompokan desa berdasarkan ekonomi kreatif penulis menentukan jumlah K(*Cendroid*)= 3 sesuai dengan hasil metode *elbow*. Untuk melakukan penerapan algoritma *K-Means* dapat menggunakan kode sebagai berikut:

```
import pandas as pd
from sklearn.cluster import KMeans
import plotly.express as px

# Buat crosstab jumlah ekraf per desa
ekraf_per_desa = pd.crosstab(umkm['nama desa'], umkm['nama
ekraft']).reset_index()

# Ganti nama kolom agar konsisten
ekraf_per_desa.columns.name = None  # hapus nama kolom
X = ekraf_per_desa.drop(columns=['nama desa'])

# Clustering K-Means
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
ekraf_per_desa['cluster'] = kmeans.fit_predict(X).astype(str)

# Hitung kategori dominan dan nilai tertingginya
ekraf_per_desa['Dominan Ekraf'] = X.idxmax(axis=1)
ekraf_per_desa['Nilai Dominan'] = X.max(axis=1)

# Visualisasi berdasarkan cluster
fig = px.scatter(ekraf_per_desa,
                 x='nama desa', y='Nilai Dominan',
                 color='cluster',
                 hover_data=['Dominan Ekraf'],
                 color_discrete_sequence=px.colors.qualitative
.Prism)

fig.update_layout(title='K-Means Clustering UMKM Berdasarkan
Desa dan Jenis Ekraf',
                  xaxis_title='Nama Desa', yaxis_title='Jumlah
Jenis Ekraf Dominan',
                  legend_title='Cluster', width=1000)
fig.show()
```

dari kode tersebut akan menghasilkan pemetaan *clustering*, pada gambar 4.20 merupakan hasil *clustering* data ekonomi kreatif berdasarkan nama desa, dan gambar 4.21 dengan tampilan 3D



Gambar 4. 20 Hasil Pemetaan Clustering Algoritma *K-Means*

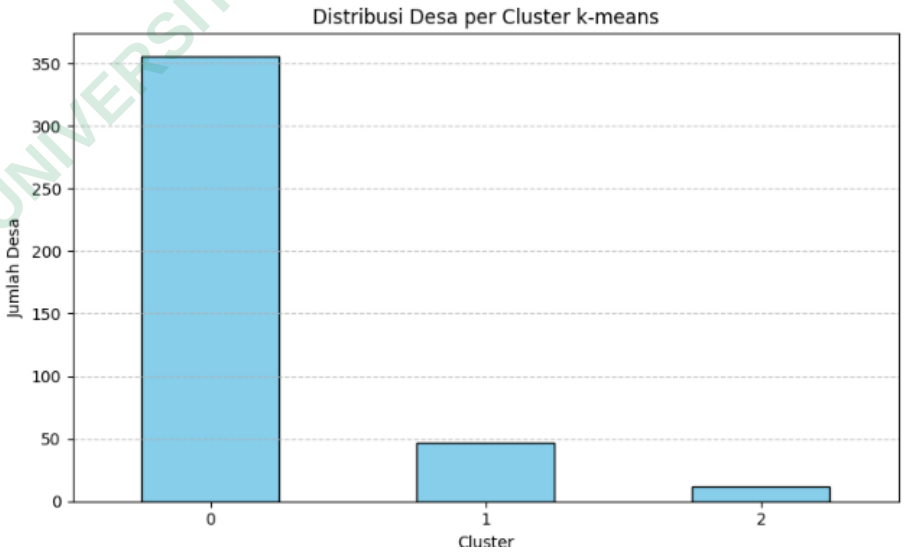


Gambar 4. 21 Tampilan 3D Clustering Algoritma *K-Means*

Dari hasil clustering menggunakan algoritma *K-Means* terdapat 3 cluster dengan keterangan sebagai berikut:

1) Cluster 0 = Kuliner, Cluster ini didominasi oleh desa-desa dengan jumlah pelaku usaha terbesar di sektor kuliner. Karakteristik umum dari cluster ini adalah tingginya kontribusi sektor kuliner dibandingkan sektor lainnya seperti fashion dan kerajinan. Meskipun terdapat beberapa nilai pada sektor lain, sektor kuliner menjadi pendorong utama dominasi usaha ekonomi kreatif pada desa-desa dalam cluster ini.
2) Cluster 1 = Fashion, Cluster ini menunjukkan dominasi usaha kreatif pada sektor fashion. Desa-desa yang tergabung dalam cluster ini umumnya memiliki pelaku usaha fashion yang lebih banyak dibanding sektor lainnya. Karakteristik khasnya adalah proporsi usaha fashion yang signifikan, sementara sektor kuliner dan kerajinan cenderung lebih rendah.
3) Cluster 2 = Kerajinan, Cluster ini menampilkan desa-desa dengan dominasi usaha di sektor kerajinan. Karakteristiknya adalah tingginya jumlah pelaku ekonomi kreatif yang bergerak dalam bidang kerajinan tangan atau produk berbasis keterampilan. Desa-desa pada cluster ini memiliki konsentrasi tinggi dalam kegiatan produksi kerajinan dibanding sektor kuliner maupun fashion.

Kemudian penulis juga melakukan pengecekan distribusi hasil clustering untuk melihat jumlah data tiap cluster, pada gambar 4.22 merupakan distribusi hasil cluster algoritma *K-means.*



Gambar 4. 22 Distribusi hasil Cluster *K-Means*

Berdasarkan hasil klasterisasi menggunakan algoritma K-means, diperoleh tiga kelompok utama dengan distribusi jumlah desa yang berbeda-beda. Cluster 0 dengan total 356 desa, cluster 1 terdiri dari 47 desa, dan cluster 2 merupakan kelompok terkecil dengan hanya 12 desa. Pada tabel 4.3 merupakan cara perhitungan manual untuk menentukan *cendroid* dari algoritma *K-Means* dengan rumus persamaan **(1)** menggunakan 6 sampel data.

Tabel 4. 3 Hitung manual sampel data Algoritma *K-Means*

Nama Desa	Rata Rata	C1	C2	C3	Jarak Terdekat	Keterangan
Bausasran	2,167	0,067	0,233	0,333	0,067	Cluster 0
Suryatmajan	2,429	0,329	0,029	0,071	0,029	Cluster 1
Tegal Panggung	2,571	0,471	0,171	0,071	0,071	Cluster 2
Pringgokumunan	2,286	0,186	0,114	0,214	0,114	Cluster 1
Sosromenduran	2,143	0,043	0,257	0,357	0,043	Cluster 0
Baciro	2,714	0,614	0,314	0,214	0,214	Cluster 2
UPDATE CENDROID						
Nama Desa	Rata Rata	C1	C2	C3	Jarak Terdekat	Keterangan
Bausasran	2,167	0,012	0,190	0,476	0,012	Cluster 0
Suryatmajan	2,429	0,274	0,071	0,214	0,071	Cluster 1
Tegal Panggung	2,571	0,417	0,214	0,071	0,071	Cluster 2
Pringgokumunan	2,286	0,131	0,071	0,357	0,071	Cluster 1
Sosromenduran	2,143	0,012	0,214	0,500	0,012	Cluster 0
Baciro	2,714	0,560	0,357	0,071	0,071	Cluster 2

Nilai yang dilabeli dengan warna kuning merupakan titik *condroid* secara random dalam menentukan cluster dimana K=3, K1= 2,143 dan di bulatkan menjadi 2.1 begitu juga dengan K2= 2,429 dan dibulatkan menjadi 2.4 serta K3= 2,571 dan dibulatkan menjadi 2.5 yang bisa di lihat pada tabel 4.4

Tabel 4. 4 Titik Centroid *K-Means*

K(Sebelum Update)		K(Setelah Update)	
K=3		K=3	
K1	2,1	k1	2,154761905
K2	2,4	k2	2,357142857
K3	2,5	k3	2,642857143

Pada tabel 4.4 di kolom k(setelah *update*) merupakan hasil penjumlahan rata rata cluster jumlah data, misal cluster 0 dimana memiliki rata rata $2{,}167 + 2{,}143 \div 2$ yang menghasilkan 2.154, Perhitungan berhenti dikarenakan algoritma telah mencapai titik stabil (*konvergen*) dan juga berhenti jika telah mencapai jumlah maksimum iterasi yang ditentukan.

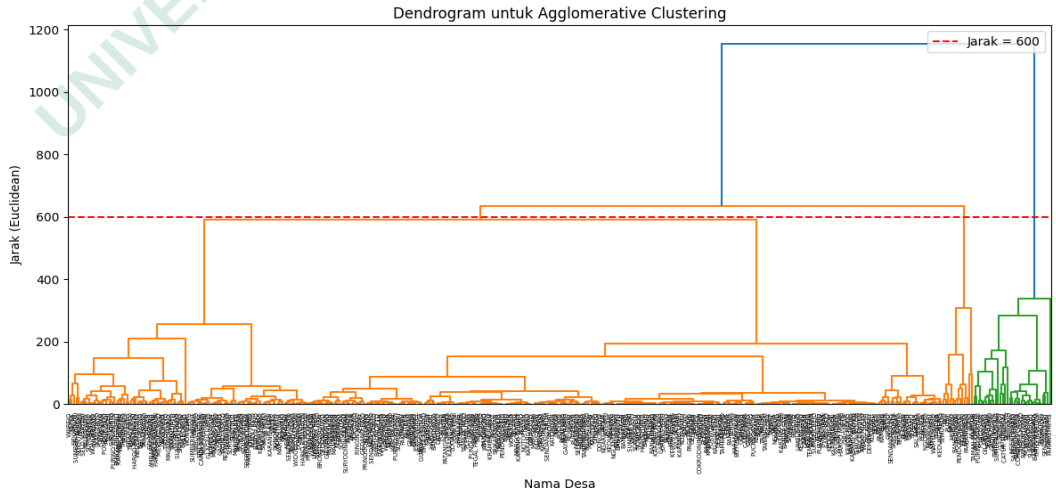
4.4.2 *Agglomerative clustering*

Agglomerative clustering merupakan salah satu metode dalam algoritma *hierarchical clustering* yang bekerja dengan pendekatan *bottom-up*, di mana setiap data awalnya dianggap sebagai satu cluster tersendiri dan secara bertahap digabungkan berdasarkan kedekatan hingga membentuk satu hierarki besar. Dalam penelitian ini, *Agglomerative clustering* digunakan untuk mengelompokkan data kategori UMKM berdasarkan kesamaan karakteristik tertentu, seperti sektor ekonomi kreatif, sehingga dapat membantu dalam mengidentifikasi pola kemiripan antar desa serta mendukung perumusan kebijakan yang lebih terarah. Untuk mengimplementasikan *agglomerative clustering* bisa menggunakan *Single Linkage* dengan rumus sama dengan persamaan **(3).** Sebelum melakukan Clustering penulis terlebih dahulu melakukan pengecekan dendogram untuk melihat jumlah *centroid* optimal untuk *agglomerative clustering* dengan kode sebagai berikut:

```
import pandas as pd
import matplotlib.pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage
```

```
# Buat ulang crosstab (pastikan sama seperti sebelumnya)
ekraf_per_desa_dendro = pd.crosstab(umkm['nama desa'],
umkm['nama ekraft']).reset_index()
X_dendro = ekraf_per_desa_dendro.drop(columns=['nama desa'])

# Buat linkage matrix menggunakan metode ward
linked = linkage(X_dendro, method='ward')
# Plot dendrogram
plt.figure(figsize=(12, 6))
dendrogram(
    linked,
    labels=ekraf_per_desa_dendro['nama desa'].values,
    orientation='top',
    distance_sort='descending',
    show_leaf_counts=False
)
# Tambahkan garis horizontal pada jarak 600
plt.axhline(y=600, color='red', linestyle='--', label='Jarak =
600')

plt.title('Dendrogram untuk Agglomerative clustering ')
plt.xlabel('Nama Desa')
plt.ylabel('Jarak (Euclidean)')
plt.legend()
plt.tight_layout()
plt.show()
```


Pada gambar 4.23 merupakan hasil dendogram *agglomerative clustering*



Gambar 4. 23 Dendogram algoritma *agglomerative clustering*

Berdasarkan dendrogram yang dihasilkan dari proses *Agglomerative clustering* , terlihat bahwa pemotongan pada ketinggian jarak sekitar 600 menghasilkan tiga klaster utama sebagai pembagian yang paling optimal. Ketiga klaster ini terbentuk dari penggabungan bertahap antar data desa berdasarkan kemiripan (jarak *Euclidean*) hingga membentuk struktur pohon. Pemilihan jumlah tiga klaster menunjukkan titik potong yang mencerminkan keseimbangan antara homogenitas di dalam klaster dan heterogenitas antar klaster, sehingga hasil pengelompokan ini dianggap paling representatif dalam menggambarkan pola atau struktur hubungan antar desa berdasarkan variabel yang dianalisis.

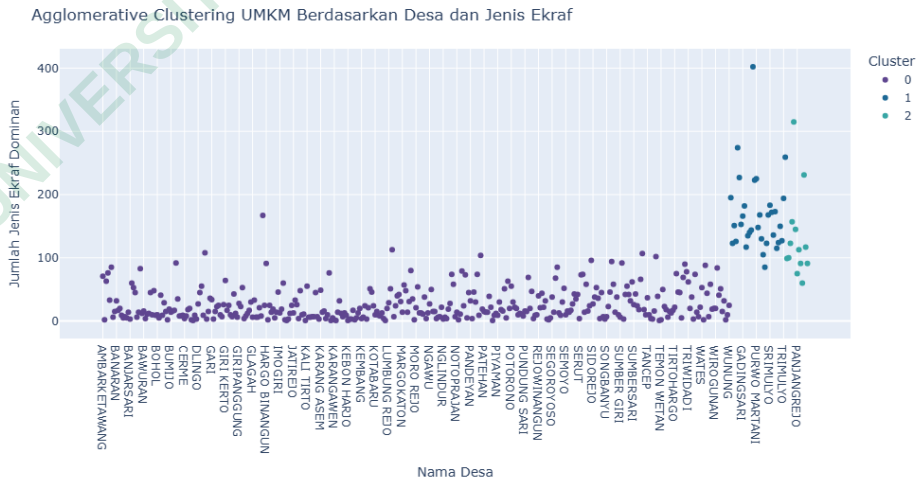
Agglomerative Clustering Ekonomi Kreatif

Agglomerative clustering Ekonomi Kreatif merupakan proses pengelompokan data pelaku usaha ekonomi kreatif berdasarkan kesamaan karakteristik pada masing-masing desa. Metode ini menggabungkan data secara bertahap mulai dari unit terkecil (setiap desa sebagai satu *cluster*) hingga terbentuk sejumlah cluster akhir yang telah ditentukan, dimana penulis menentukan jumlah k(*cendroid*)=3. Hasil dari pengelompokan ini memberikan gambaran pola dominasi subsektor ekonomi kreatif seperti kuliner, *fashion*, atau kerajinan di tiap wilayah. Dengan analisis ini, dapat diketahui konsentrasi jenis usaha kreatif tertentu pada desa-desa tertentu, yang berguna untuk membantu pemangku kepentingan dan juga wisatawan dalam melihat kategori UMKM yang sesuai dengan minat mereka. Untuk melakukan clustering menggunakan algoritma *agglomerative clustering* dapat menggunakan kode sebagai berikut:

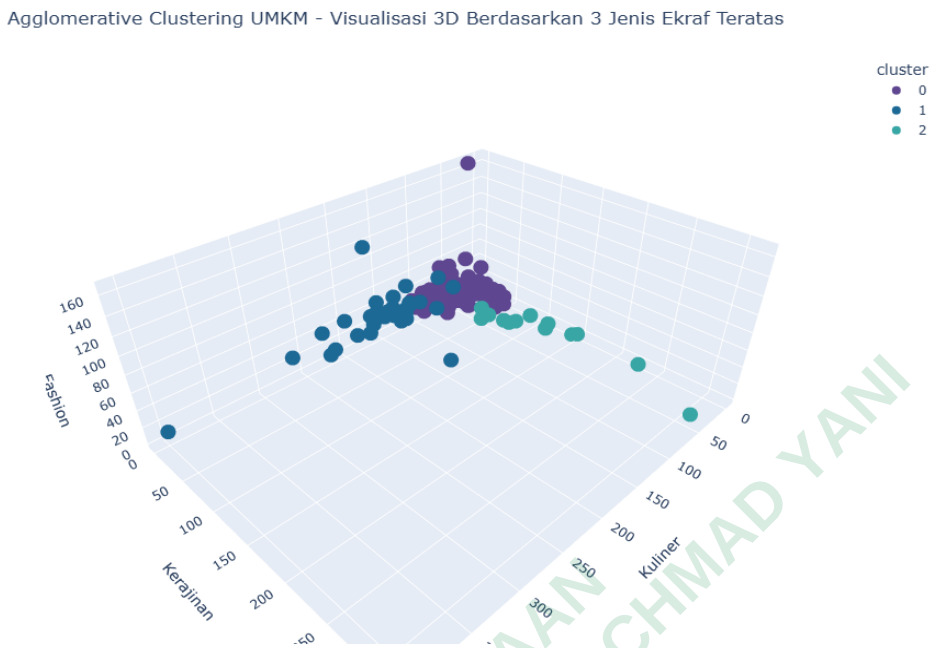
```
import pandas as pd
from sklearn.cluster import AgglomerativeClustering
import plotly.express as px

# Buat crosstab jumlah ekraf per desa
ekraf_per_desa_agglo = pd.crosstab(umkm['nama desa'],
umkm['nama ekraft']).reset_index()

# Ganti nama kolom agar konsisten
ekraf_per_desa_agglo.columns.name = None  # hapus nama kolom
X = ekraf_per_desa_agglo.drop(columns=['nama desa'])
```

```
# Clustering Agglomerative
agglo = AgglomerativeClustering(n_clusters=3)
ekraf_per_desa_agglo['cluster'] =
agglo.fit_predict(X).astype(str)

# Hitung kategori dominan dan nilai tertingginya
ekraf_per_desa_agglo['Dominan Ekraf'] = X.idxmax(axis=1)
ekraf_per_desa_agglo['Nilai Dominan'] = X.max(axis=1)

# Visualisasi berdasarkan cluster
fig = px.scatter(ekraf_per_desa_agglo,
                 x='nama desa', y='Nilai Dominan',
                 color='cluster',
                 hover_data=['Dominan Ekraf'],
                 color_discrete_sequence=px.colors.qualitative
.Prism)
fig.update_layout(title='Agglomerative clustering  UMKM
Berdasarkan Desa dan Jenis Ekraf',
                  xaxis_title='Nama Desa', yaxis_title='Jumlah
Jenis Ekraf Dominan',
                  legend_title='Cluster', width=1000)
fig.show()
```


Pada gambar 4.24 merupakan hasil clustering ekonomi kreatif berdasarkan nama desa, dan gambar 4.25 hasil clustering dengan tampilan 3D.



Gambar 4. 24 Hasil clustering ekonomi kreatif *agglomerative clustering*



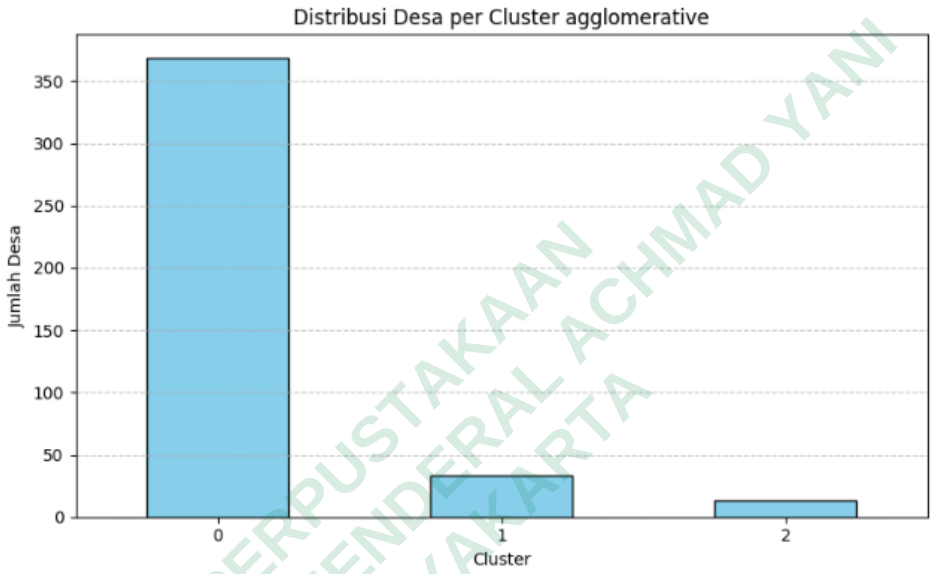
Gambar 4. 25 Hasil *Clustering* Tampilan 3D

Dari hasil clustering ekonomi kreatif berdasarkan nama desa dengan menggunakan algoritma *agglomerative clustering* terdapat 3 cluster sesuai dengan *centroid* yang penulis tentukan di awal dimana penjelasan clusternya sebagai berikut:

1) Cluster 0 = Kuliner, Cluster ini didominasi oleh desa-desa dengan jumlah pelaku usaha terbesar di sektor kuliner. Karakteristik umum dari cluster ini adalah tingginya kontribusi sektor kuliner dibandingkan sektor lainnya seperti fashion dan kerajinan. Meskipun terdapat beberapa nilai pada sektor lain, sektor kuliner menjadi pendorong utama dominasi usaha ekonomi kreatif pada desa-desa dalam cluster ini.
2) Cluster 1 = Fashion, Cluster ini menunjukkan dominasi usaha kreatif pada sektor fashion. Desa-desa yang tergabung dalam cluster ini umumnya memiliki pelaku usaha fashion yang lebih banyak dibanding sektor lainnya. Karakteristik khasnya adalah proporsi usaha fashion yang signifikan, sementara sektor kuliner dan kerajinan cenderung lebih rendah.
3) Cluster 2 =Kerajinan, Cluster ini menampilkan desa-desa dengan dominasi usaha di sektor kerajinan. Karakteristiknya adalah tingginya jumlah pelaku

ekonomi kreatif yang bergerak dalam bidang kerajinan tangan atau produk berbasis keterampilan. Desa-desa pada cluster ini memiliki konsentrasi tinggi dalam kegiatan produksi kerajinan dibanding sektor kuliner maupun fashion.

Selanjutnya penulis juga melakukan pengechekan distribusi hasil clustering untuk melihat jumlah data tiap cluster algoritma *Agglomerative clustering*, pada gambar 4.26 merupakan distribusi hasil cluster algoritma *Agglomerative Clustering.*



Gambar 4. 26 Hasil distribusi jumlah cluster *agglomerative*

Berdasarkan hasil klasterisasi menggunakan algoritma Agglomerative Clustering, diperoleh tiga kelompok utama dengan distribusi jumlah desa yang berbeda-beda. Cluster 0 dengan total 369 desa, cluster 1 terdiri dari 33 desa, dan cluster 2 merupakan kelompok terkecil dengan hanya 13 desa. Pada tabel 4.5 merupakan cara penghitungan manual untuk algoritma *Agglomerative Clustering* menggunakan persamaan **(3)** dengan sampel data sebanyak 6 data.

Tabel 4. 5 Hitung manual *agglomerative clustering*

Nama Desa	Rata Rata	
Bausasran	2,167	TARGET CLUSTER= 3
Suryatmajan	2,429	
Tegal Panggung	2,571	

Pringgokumunabn	2,286	
Sosromendur an	2,143	
Baciro	2,714	

SETIAP DATA ADALAH CLUSTER

D	Bausasran	Suryatmajan	Tegal Panggung	Pringgokumunan	Sosromenduran	Baciro
Bausasran	0,000	0,262	0,405	0,119	0,024	0,548
Suryatmajan	0,262	0,000	0,143	0,143	0,286	0,286
Tegal Panggung	0,405	0,143	0,000	0,286	0,429	0,143
Pringgokumunan	0,119	0,143	0,286	0,000	0,143	0,429
Sosromendur an	0,024	0,286	0,429	0,143	0,000	0,571
Baciro	0,548	0,286	0,143	0,429	0,571	0,000

MUNCUL CLUSTER BARU DARI JARAK TERDEKAT ITERASI 1

D	Bau/sosro	Suryatmajan	Tegal Panggung	Pringgokumunan	baciro	klaster
Bau/sosro	0	0,262	0,405	0,119	0,548	1
Suryatmajan	0,262	0	0,143	0,143	0,286	2
Tegal Panggung	0,405	0,143	0	0,286	0,143	3
Pringgokumunan	0,119	0,143	0,286	0	0,429	4
Baciro	0,548	0,286	0,143	0,429	0	5

MUNCUL CLUSTER BARU DARI JARAK TERDEKAT ITERASI 2

D	Bau/sosro/pringgo	Suryatmajan	Tegal Panggung	baciro	klaster
Bau/sosro/pringgo	0	0,143	0,286	0,429	1
Suryatmajan	0,143	0	0,143	0,286	2
Tegal Panggung	0,286	0,143	0	0,143	3
Baciro	0,429	0,286	0,143	0	4

MUNCUL CLUSTER BARU DARI JARAK TERDEKAT ITERASI 3				
D	Bau/sosro/pringgo	Surya/tegal	baciro	klaster
Bau/sosro/pringgo	0	0,143	0,429	1
surya/tegal	0,143	0	0,143	2
baciro	0,429	0,143	0	3

Perhitungan berhenti secara otomatis ketika jumlah cluster yang terbentuk telah mencapai jumlah centroid atau target cluster yang telah ditentukan. Hal tersebut menunjukkan bahwa proses pengelompokan telah selesai karena struktur hierarki telah menyatu pada tingkat yang diinginkan. Namun, jika jumlah *centroid* diubah, misalnya ditambah atau dikurangi, maka proses penggabungan akan dilanjutkan kembali hingga mencapai jumlah *cluster* yang baru tersebut.

4.5 EVALUASI

Evaluasi algoritma clustering merupakan tahap penting dalam memastikan bahwa hasil pengelompokan yang dilakukan benar-benar mencerminkan pola alami dalam data. Setelah proses clustering selesai, diperlukan pengukuran kuantitatif terhadap kualitas dan validitas dari *cluster* yang terbentuk. Hal ini dilakukan agar hasil tidak hanya terbentuk berdasarkan proses matematis semata, namun juga dapat diterima secara logis. Tiga metrik umum yang digunakan dalam evaluasi *clustering* adalah *Silhouette Score, Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Ketiganya memberikan gambaran tentang seberapa baik objek-objek dalam satu cluster memiliki kemiripan dan seberapa jauh mereka dari cluster lain.

Melalui tahapan evaluasi, peneliti dapat menilai sejauh mana metode clustering yang digunakan seperti *K-Means* atau *Agglomerative clustering* mampu mengelompokkan data secara efektif. Nilai evaluasi yang baik menunjukkan bahwa cluster yang terbentuk memiliki konsistensi internal yang tinggi dan pemisahan antar cluster yang jelas. Sebaliknya, hasil evaluasi yang rendah dapat menjadi indikator perlunya pemilihan algoritma lain, penyesuaian parameter, atau bahkan transformasi data sebelum proses clustering dilakukan ulang. Tahapan ini menjadi

kunci dalam menjamin keandalan dan akurasi dari keseluruhan proses analisis data.algoritma terdabaik dalam evaluasi nya nanti akan digunakan sebagai bahan dasar pada visualisasi dashboard kategori UMKM.

4.5.1 Evaluasi *K-Means*

Penulis melakukan evaluasi algoritma *K-Means* untuk mengukur sejauh mana algoritma mampu membentuk kelompok yang efektif terhadap data yang dianalisis. Proses evaluasi ini biasanya dilakukan dengan menggunakan beberapa metrik seperti, *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Masing-masing metrik memberikan perspektif berbeda dalam menilai kualitas cluster, mulai dari seberapa padat dan terpisahnya antar-cluster, hingga sejauh mana anggota dalam satu cluster memiliki kemiripan yang tinggi. Dengan melakukan evaluasi ini, peneliti dapat menilai apakah hasil pengelompokan sudah optimal atau perlu dilakukan penyesuaian, seperti perubahan jumlah cluster atau pemilihan algoritma yang lebih sesuai. Untuk melakukan evaluasi model dapat menggunakan kode sebagai berikut:

```
from sklearn.metrics import silhouette_score,
davies_bouldin_score, calinski_harabasz_score

def evaluate_kmeans(X, labels, nama_evaluasi):
    sil_score = silhouette_score(X, labels)
    db_score = davies_bouldin_score(X, labels)
    ch_score = calinski_harabasz_score(X, labels)

    print(f"=== Evaluasi Clustering: {nama_evaluasi} ===")
    print(f"Silhouette Score         : {sil_score:.4f}
(semakin tinggi, semakin baik)")
    print(f"Davies-Bouldin Index     : {db_score:.4f} (semakin
rendah, semakin baik)")
    print(f"Calinski-Harabasz Index  : {ch_score:.4f} (semakin
tinggi, semakin baik)")
    print("")

# ==== Evaluasi Clustering Ekraf ====
X_ekraf = ekraf_per_desa.drop(columns=['nama desa', 'cluster',
'Dominan Ekraf', 'Nilai Dominan'])
labels_ekraf = ekraf_per_desa['cluster'].astype(int)
evaluate_kmeans(X_ekraf, labels_ekraf, "Ekraf per Desa")
```

Pada gambar 4.27 merupakan hasil *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*.

```
=== Evaluasi Clustering: Ekraf per Desa ===
Silhouette Score         : 0.7016 (semakin tinggi, semakin baik)
Davies-Bouldin Index     : 0.8582 (semakin rendah, semakin baik)
Calinski-Harabasz Index  : 395.2165 (semakin tinggi, semakin baik)
```

Gambar 4. 27 Hasil Evaluasi *K-Means Clustering*

Berdasarkan hasil evaluasi clustering yang ditampilkan pada gambar 4.27, terdapat dua hasil evaluasi untuk dua jenis pengelompokan: ekonomi kreatif (ekraf) per desa dan klasifikasi usaha per desa. Penilaian dilakukan dengan menggunakan tiga metrik utama, yaitu *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*.

Untuk ekraf per desa, nilai *Silhouette Score* sebesar 0.7016 menunjukkan bahwa pengelompokan cukup baik karena mendekati nilai maksimum 1. *Davies-Bouldin Index* berada di angka 0.8582, yang relatif rendah dan mengindikasikan pemisahan antar cluster yang cukup baik. Sementara itu, *Calinski-Harabasz Index* sebesar 395.2165 menunjukkan struktur cluster yang cukup baik karena semakin tinggi nilainya maka semakin baik kualitas cluster.

4.5.2 Evaluasi *Agglomerative clustering*

Penulis melakukan evaluasi terhadap algoritma *Agglomerative clustering* untuk menilai seberapa baik hasil pengelompokan yang dihasilkan dalam membagi data ke dalam cluster yang bermakna. Dengan menggunakan metrik evaluasi seperti *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*, performa dari pengelompokan dapat dianalisis secara objektif. *Silhouette Score* memberikan gambaran mengenai kesesuaian data dalam cluster-nya masing-masing, sementara *Davies-Bouldin Index* menilai tingkat kemiripan antar cluster semakin rendah nilainya, semakin baik. Di sisi lain, *Calinski-Harabasz Index* mengukur seberapa terpisah dan kompak cluster yang terbentuk. Hasil evaluasi ini menjadi dasar untuk

menyimpulkan efektivitas metode *Agglomerative clustering* dalam mengelompokkan data desa berdasarkan ekonomi kreatif maupun klasifikasi usaha, serta menjadi acuan dalam perbaikan atau pengambilan keputusan lebih lanjut. Untuk melakukan evaluasi model dapat dilakukan dengan code sebagai berikut:

```
import pandas as pd
from sklearn.metrics import silhouette_score,
calinski_harabasz_score, davies_bouldin_score

# Fungsi evaluasi clustering
def evaluate_clustering(X, labels, nama_evaluasi):
    sil_score = silhouette_score(X, labels)
    db_score = davies_bouldin_score(X, labels)
    ch_score = calinski_harabasz_score(X, labels)

    print(f"=== Evaluasi Clustering: {nama_evaluasi} ===")
    print(f"Silhouette Score        : {sil_score:.4f}
(semakin tinggi, semakin baik)")
    print(f"Davies-Bouldin Index    : {db_score:.4f} (semakin
rendah, semakin baik)")
    print(f"Calinski-Harabasz Index  : {ch_score:.4f} (semakin
tinggi, semakin baik)")
    print("")

# ==== Evaluasi untuk CLUSTER BERDASARKAN EKRAF ====
columns_to_drop_ekraf = ['nama desa', 'cluster', 'Dominan
Ekraf', 'Nilai Dominan']
available_columns_ekraf = [col for col in
columns_to_drop_ekraf if col in ekraf_per_desa_agglo.columns]
X_ekraf =
ekraf_per_desa_agglo.drop(columns=available_columns_ekraf)
labels_ekraf = ekraf_per_desa_agglo['cluster'].astype(int)
evaluate_clustering(X_ekraf, labels_ekraf, "Ekraf per Desa")
```

Pada gambar 4.28 merupakan hasil evaluasi algoritma *Agglomerative clustering* .

```
=== Evaluasi Clustering: Ekraf per Desa ===
Silhouette Score        : 0.7028 (semakin tinggi, semakin baik)
Davies-Bouldin Index    : 0.6779 (semakin rendah, semakin baik)
Calinski-Harabasz Index  : 332.5230 (semakin tinggi, semakin baik)
```

Gambar 4. 28 Evaluasi Algoritma *Agglomerative clustering*

Berdasarkan hasil evaluasi clustering *Agglomerative clustering* yang ditampilkan pada gambar 4.28, terdapat dua hasil evaluasi untuk dua jenis pengelompokan: ekonomi kreatif (ekraf) per desa dan klasifikasi usaha per desa. Penilaian dilakukan dengan menggunakan tiga metrik utama, yaitu *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*.

Untuk ekraf per desa, nilai *Silhouette Score* sebesar 0.7028 menunjukkan bahwa pengelompokan cukup baik karena mendekati nilai maksimum 1, yang berarti objek dalam cluster cukup mirip satu sama lain dan berbeda dengan objek di cluster lain. Nilai *Davies-Bouldin Index* sebesar 0.6779 termasuk rendah, yang menunjukkan bahwa antar cluster memiliki pemisahan yang cukup baik dan kompak. Sedangkan nilai *Calinski-Harabasz* Index sebesar 332.5230 juga menunjukkan kualitas struktur cluster yang cukup baik, karena semakin tinggi nilainya, semakin baik pemisahan dan kekompakan antar cluster.

Dari hasil evaluasi clustering menggunakan dua algoritma, yaitu *K-Means* dan *Agglomerative clustering* , yang dapat di lihat jelas perbedaannya pada tabel 4.6

Tabel 4. 6 Hasil Evaluasi

	K-Means	*Agglomerative clustering*
Silhouette Score	0.701	0.702
Davies-Bouldin Index	0.858	0.677
Calinski-Harabasz Index	395.216	332.523

Dari tabel 4.6 dapat ditarik beberapa kesimpulan terkait performa masing-masing metode dalam mengelompokkan data Ekraf per Desa dan Klasifikasi Usaha per Desa. Untuk data Ekraf per Desa, baik *K-Means* maupun *Agglomerative clustering* menghasilkan nilai *Silhouette Score* yang hampir sama, yaitu sekitar 0.70, yang menunjukkan bahwa kedua metode mampu membentuk cluster dengan pemisahan yang cukup baik dan anggota cluster yang konsisten. Namun, nilai *Davies-Bouldin Index* pada *Agglomerative clustering* (0.6779) lebih rendah dibandingkan dengan

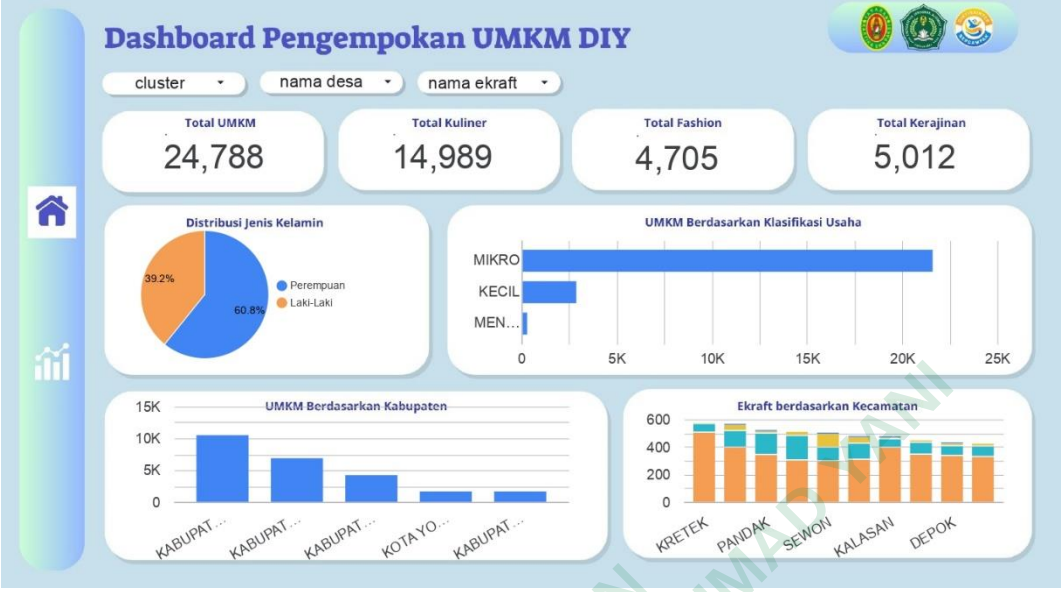
K-Means (0.8582), menandakan cluster yang terbentuk dengan *Agglomerative* lebih berbeda dan tidak tumpang tindih. Meskipun *K-Means* memiliki nilai *Calinski-Harabasz Index* yang lebih tinggi (395.2165 banding 332.5230), perbedaan ini tidak cukup signifikan untuk menandakan superioritas yang jelas, karena indeks ini hanya menunjukkan tingkat kepadatan dan pemisahan cluster secara global.

Dengan mempertimbangkan seluruh metrik evaluasi, *Agglomerative clustering* dapat dikatakan lebih stabil dan konsisten dalam menjaga pemisahan cluster yang baik secara keseluruhan, sedangkan *K-Means* unggul dalam membentuk cluster yang lebih padat dan terstruktur terutama pada data Klasifikasi Usaha. Oleh karena itu pada tahap Visualisasi data penulis memilih untuk memetakan hasil cluster dari algoritma *Agglomerative clustering* .

4.6 VISUALISASI

Visualisasi dashboard merupakan tahap akhir dalam proses analisis data yang berfungsi untuk menyajikan hasil olahan data secara ringkas, interaktif, dan mudah dipahami. Setelah data dikumpulkan, dibersihkan, dan dianalisis, langkah selanjutnya adalah menyusun visualisasi dalam bentuk dashboard agar informasi yang diperoleh dapat diakses dan dimanfaatkan oleh pengguna dengan lebih efisien. Dashboard biasanya menampilkan berbagai jenis grafik seperti diagram batang, garis, pie chart, hingga peta interaktif, yang dapat digunakan untuk mengamati tren, membandingkan data, dan mendeteksi pola penting dalam waktu singkat.

Tahap ini sangat penting karena menjadi media utama dalam menyampaikan insight kepada pihak yang berkepentingan, seperti manajer, pemangku kebijakan, atau publik secara umum. Dengan tampilan yang intuitif dan informatif, dashboard membantu pengambilan keputusan berbasis data tanpa harus memeriksa laporan yang panjang atau rumit. Oleh karena itu, perancangan dashboard harus memperhatikan tujuan analisis, kebutuhan pengguna, serta kejelasan visualisasi agar informasi yang ditampilkan benar-benar relevan dan mendukung proses pengambilan keputusan secara efektif. Pada gambar 4.29 merupakan visualisasi dashboard menggunakan *software looker studio.*

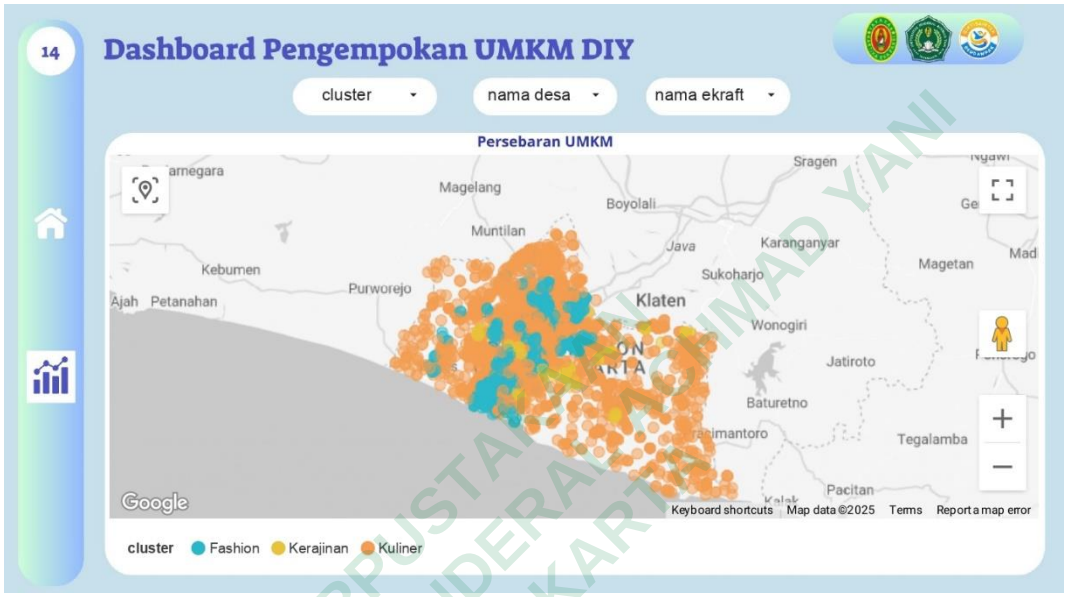


Gambar 4. 29 Visualisasi Data

Pada gambar 4.29 merupakan dashboard pengelompokan kategori UMKM DIY yang menyajikan visualisasi data mengenai Usaha Mikro, Kecil, dan Menengah yang tersebar di wilayah Daerah Istimewa Yogyakarta. Dashboard tersebut dilengkapi dengan fitur filter interaktif yang memungkinkan pengguna untuk menyaring data berdasarkan cluster, nama desa, dan jenis ekonomi kreatif (ekraf). Dengan adanya filter ini, pengguna dapat mengeksplorasi data secara lebih mendalam sesuai dengan kebutuhan analisis. Bagian atas dashboard menampilkan statistik utama seperti total UMKM sebanyak 24.788 unit, dengan rincian 14.989 bergerak di sektor kuliner, 4.705 di sektor fashion, dan 5.012 di sektor kerajinan. Di samping itu, pie chart menunjukkan distribusi pelaku UMKM berdasarkan jenis kelamin, dengan persentase perempuan sebesar 60,8 persen dan laki-laki sebesar 39,2 persen. Grafik batang horizontal menampilkan klasifikasi usaha, di mana mayoritas UMKM berada pada kategori mikro, disusul oleh kecil, dan hanya sedikit yang termasuk menengah.

Distribusi geografis kategori UMKM juga divisualisasikan dalam bentuk grafik batang vertikal berdasarkan kabupaten dan kota. Terlihat bahwa Kabupaten Sleman dan Bantul menjadi wilayah dengan jumlah UMKM tertinggi, diikuti oleh Kota Yogyakarta dan wilayah lainnya. Selain itu, grafik lainnya menunjukkan

sebaran usaha ekonomi kreatif berdasarkan kecamatan, dengan lima kecamatan teratas yaitu Kretek, Pandak, Sewon, Kalasan, dan Depok. Dashboard ini bertujuan memberikan gambaran umum mengenai kondisi UMKM DIY serta mendukung pengambilan keputusan berbasis data. Kemudian untuk persebaran data kategori UMKM dapat di lihat pada gambar 4.30



Gambar 4. 30 Peta Persebaran UMKM DIY

Pada gambar 4.30 menampilkan persebaran kategori UMKM di wilayah Daerah Istimewa Yogyakarta berdasarkan sektor ekonomi kreatif, yaitu kuliner, fashion, dan kerajinan. Titik-titik pada peta menunjukkan lokasi UMKM, dengan warna yang merepresentasikan clusternya: oranye untuk kuliner, biru untuk fashion, dan kuning untuk kerajinan. Peta ini memberikan gambaran sebaran spasial UMKM di DIY, di mana terlihat bahwa kategori UMKM kuliner mendominasi di hampir seluruh wilayah, sedangkan fashion dan kerajinan tersebar lebih terbatas. Visualisasi ini berguna untuk memahami konsentrasi dan distribusi kategori UMKM secara geografis yang dapat membantu wisatawan dalam melihat kategori UMKM.

4.7 PEMBAHASAN

Penelitian ini membandingkan dua algoritma *unsupervised learning*, yaitu *K-Means* dan *Agglomerative clustering* , untuk mengelompokkan data kategori

UMKM di Provinsi Daerah Istimewa Yogyakarta berdasarkan variabel nama desa dan sektor ekonomi kreatif. Tujuan dari perbandingan ini adalah untuk menilai efektivitas masing-masing algoritma dalam mengidentifikasi pola kelompok kategori UMKM yang relevan secara spasial dan sektoral.

Berdasarkan hasil implementasi dan evaluasi, *Agglomerative clustering* terbukti lebih unggul dibandingkan *K-Means* dalam konteks pengelompokan kategori UMKM di wilayah ini. Dibuktikan dengan analisis hasil *cluster* yang terdapat pada lampiran 7 dimana terdapat perbedaan hasil clustering di beberapa desa yang dapat di lihat pada tabel 4.7.

Tabel 4. 7 Hasil Perbedaan clustering algoritma

K-means		*Agglomerative Clustering*	
Clustering Benar	Clustering Salah	Clustering Benar	Clustering Salah
309	106	324	91

K-means mengelompokkan kategori UMKM dengan benar sejumlah 309 UMKM dari 415 data yang ada, sementara *Agglomerative Clustering* mengelompokkan kategori UMKM dengan benar sebanyak 324 UMKM dari total 415 data, Keunggulan *Agglomerative clustering* dibuktikan juga dengan hasil evaluasi *Silhouette Score Agglomerative* mencapai 0,702, sedikit lebih tinggi dibandingkan *K-Means* yang memperoleh *0,701*. Sementara itu, nilai *Davies-Bouldin Index Agglomerative* sebesar 0,677, lebih rendah dari *K-Means* sebesar *0,858*. Hal ini disebabkan oleh pendekatan hierarki bottom-up yang digunakan, yang mempertimbangkan jarak antar data dalam membentuk struktur klaster bertingkat. Sebagai hasilnya, klaster yang dihasilkan oleh *Agglomerative clustering* lebih stabil dan konsisten, terutama ketika variabel lokasi desa memainkan peran penting dalam penyebaran data.

Hasil pengelompokan dari algoritma *K-Means* maupun *Agglomerative clustering* dapat dimanfaatkan tidak hanya oleh pemerintah daerah atau pelaku UMKM, tetapi juga oleh wisatawan yang ingin menjelajahi potensi ekonomi kreatif di Provinsi DIY. Dengan adanya klasterisasi kategori UMKM berdasarkan lokasi

desa dan jenis usaha, wisatawan dapat lebih mudah dalam melihat kategori UMKM yang sesuai dengan minat mereka, seperti kuliner khas daerah, kerajinan lokal, maupun pertunjukan seni budaya.

Sebagai contoh, jika hasil klaster menunjukkan bahwa suatu desa memiliki konsentrasi tinggi UMKM di sektor kerajinan, maka desa tersebut dapat dipromosikan sebagai destinasi wisata belanja produk lokal. Sebaliknya, klaster desa dengan dominasi UMKM kuliner dapat diarahkan sebagai rute wisata kuliner