



**PENERAPAN ALGORITMA “*K-MEANS CLUSTERING*” UNTUK  
MENGETAHUI PREFERENSI JENIS USAHA MIKRO WARGA  
KECAMATAN INDRAMAYU TAHUN 2020**

**SKRIPSI**

**MUHAMMAD RAFFIZA AZKA**

**1910511062**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2023**



**PENERAPAN ALGORITMA “*K-MEANS CLUSTERING*” UNTUK  
MENGETAHUI PREFERENSI JENIS USAHA MIKRO WARGA  
KECAMATAN INDRAMAYU TAHUN 2020**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Ilmu  
Komputer**

**MUHAMMAD RAFFIZA AZKA**

**1910511062**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2023**

# PERNYATAAN ORISINALITAS

## PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri, dan semua sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Muhammad Raffiza Azka

NIM : 1910511062

Tanggal : 12 Juli 2023

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 12 Juli 2023

Yang Menyatakan,

  
Muhammad Raffiza Azka

## PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

### PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Raffiza Azka  
NIM : 1910511062  
Fakultas : Ilmu Komputer  
Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya berjudul :

#### **PENERAPAN ALGORITMA “K-MEANS CLUSTERING” UNTUK MENGETAHUI PREFERENSI JENIS UMKM WARGA KECAMATAN INDRAMAYU TAHUN 2020**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta  
Pada tanggal : 12 Juli 2023  
Yang Menyatakan,



Muhammad Raffiza Azka

# LEMBAR PERSETUJUAN

## LEMBAR PERSETUJUAN


Yang bertanda tangan di bawah ini:

Nama : Muhammad Raffiza Azka  
NIM : 1910511062  
Program Studi : Informatika  
Judul Skripsi/TA : Penerapan Algoritma "K-Means Clustering" untuk Mengetahui Preferensi Jenis Usaha Mikro Warga Kecamatan Indramayu Tahun 2020

Dinyatakan telah memenuhi syarat dan menyetujui untuk mengikuti ujian sidang skripsi.

Jakarta, ... Mei 2023

Mengetahui,  
Ketua Program Studi,



Dr. Widya Cholil, M.L.T.

Menyetujui,  
Dosen Pembimbing,



Nurhafifah Matondang, S.Kom, M.M, M.T.I

# LEMBAR PENGESAHAN

## LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Muhammad Raffiza Azka  
NIM : 1910511062  
Program Studi : SI Informatika  
Judul Tugas Akhir : Penerapan Algoritma "K-Means Clustering" Untuk Mengetahui Preferensi Jenis Usaha Mikro Warga Kecamatan Indramayu Tahun 2020

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

  
Bayu Hananto, S.Kom., M.Kom.


Penguji I

  
Anita Muliawati, S.Kom., MTI.  
Penguji II

  
Nurhafifah Matondang, S.Kom, M.M,  
MTI

Pembimbing

  
Dr. Ermatita, M.Kom.  
Dekan

  
Dr. Widya Cholil, M.I.T.  
Kepala Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 10 Juli 2023



**PENERAPAN ALGORITMA “K-MEANS CLUSTERING” UNTUK  
MENGETAHUI PREFERENSI JENIS UMKM WARGA KECAMATAN  
INDRAMAYU TAHUN 2020**

**MUHAMMAD RAFFIZA AZKA**

**ABSTRAK**

UMKM atau Usaha Mikro Kecil dan Menengah sudah menjadi salah satu sumber utama penghasilan bagi banyak warga Indonesia. Untuk mengembangkan Usaha Mikro di suatu daerah, kita perlu mengetahui preferensi Usaha Mikro warga daerah tersebut. Diharapkan dari hasil penelitian ini kita bisa mengetahui mana jenis Usaha Mikro yang paling banyak diminati. Dalam penelitian ini, metode yang akan digunakan adalah metode *K-Means Clustering*. Akan digunakan metode Elbow untuk menentukan banyak *cluster*. Setelah proses *clustering*, *cluster* akan diuji menggunakan metode *silhouette coefficient* dan *Davies Bouldin Index*. Untuk penelitian ini, *clustering* data Usaha Mikro akan dibagi menjadi beberapa label berdasarkan jenis Usaha Mikro. Penelitian akan menggunakan data dari Kecamatan Indramayu. Total data mentah yang akan digunakan dalam penelitian ini adalah 24.000. Dikarenakan data masih dalam keadaan yang kurang layak, perlu dilakukan pra-proses data data terlebih dahulu karena kondisi data yang agak berantakan agar proses penelitian tidak ada kendala. Penelitian ini akan menggunakan *Python* dan *Excel* untuk pra-proses data dan penerapan *K-Means Clustering*.

**Kata Kunci** : *UM, K-Means, Clustering, Metode Elbow, Silhouette Coefficient, Davies Bouldin Index, Indramayu, Pra-proses, .*

**IMPLEMENTATION OF THE “K-MEANS CLUSTERING”  
ALGORITHM TO KNOW THE PREFERENCE OF TYPES OF MICRO  
ENTERPRISES IN INDRAMAYU DISTRICT IN 2020**

**MUHAMMAD RAFFIZA AZKA**

**ABSTRACT**

*MSMEs or Micro, Small and Medium Enterprises have become one of the main sources of income for many Indonesians. To develop MSMEs in an area, we need to know the MSME preferences of residents of that area. It is hoped that from the results of this study we can find out which types of Micro Enterprises are most in demand. In this study, the method to be used is the K-Means Clustering method. Elbow method will be used to determine the number of clusters. After the clustering process, clusters will be tested using the silhouette coefficient method and the Davies Bouldin Index. For this research, data clustering will be divided into several labels based on the type of Micro Enterprises. The research will use Micro Enterprises data from Indramayu District. The total raw data to be used in this study is 24,000. Because the data is still in an inadequate condition, it is necessary to pre-process the data first because the data conditions are a bit messy so that there are no obstacles in the research process. This research will use Python and Excel to pre-process the data and apply K-Means Clustering.*

**KeyWords** : *Micro Enterprises, K-Means, Clustering, Elbow Method, Silhouette Coefficient, Davies Bouldin Index, Indramayu, Preprocess, .*



## KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa, karena atas berkat rahmat-Nya, penulis dapat menyelesaikan penulisan skripsi yang disusun sebagai salah satu syarat kelulusan Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Penulisan skripsi ini tidak mungkin berjalan dengan lancar tanpa bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada :

1. Allah SWT, yang sudah memberikan rahmat, karunia, kesehatan, dan kemudahan kepada penulis dalam mengerjakan tugas akhir ini sampai dengan selesai.
2. Kedua orang tua yang sudah mendoakan dan memberi dukungan sehingga skripsi ini bisa selesai.
3. Dr. Ermatita, M.Kom., selaku dekan Fakultas Ilmu Komputer.
4. DR. Widya Cholil, M.IT., selaku Ketua Program Studi Sarjana Jurusan S1 Informatika.
5. Nurhafifah Matondang, S.Kom., M.M., M.SI., selaku dosen pembimbing.
6. Seluruh pihak yang terlibat dalam kelancaran pembuatan skripsi ini dan yang belum disebutkan di atas.

Penulis berharap skripsi ini dapat memberi manfaat dan ilmu yang berguna bagi pembacanya.

Jakarta, 12 Juli 2023

Muhammad Raffiza Azka

## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
LEMBAR PERSETUJUAN .....	v
LEMBAR PENGESAHAN .....	vi
ABSTRAK .....	vii
<i>ABSTRACT</i> .....	viii
KATA PENGANTAR .....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR .....	xiv
DAFTAR SIMBOL .....	xvi
DAFTAR LAMPIRAN.....	xvii
BAB 1 .....	1
1.1. Latar Belakang .....	1
1.3. Tujuan.....	2
1.4. Ruang Lingkup.....	2
1.5. Manfaat Penelitian.....	2
1.6. Sistematika Penulisan.....	3
BAB 2 .....	4
2.1. UMKM .....	4
2.1.1. Pengertian .....	4
2.1.2. Kriteria UMKM .....	4
2.2. Algoritma <i>K-Means Clustering</i> .....	5
2.2.1. Pengertian .....	5
2.2.2. Alur Kerja .....	6
2.2.3. Rumus .....	6
2.3. Elbow Method .....	7
2.3.1. Pengertian .....	7
2.3.2. Rumus .....	8
2.4. <i>Silhouette Coefficient</i> .....	8
2.4.1. Pengertian .....	8
2.4.2. Rumus .....	8

2.5.	<i>Davies Bouldin Index</i> .....	10
2.5.1.	Pengertian .....	10
2.5.2.	Rumus .....	10
2.6.	Pra-proses Data .....	11
2.6.1.	Pengertian .....	11
2.6.2.	Tahapan Pra-proses .....	11
2.7.	<i>Python</i> .....	12
2.7.1.	Pengertian .....	13
2.7.2.	Library.....	13
2.8.	Penelitian Terdahulu .....	14
BAB 3	.....	17
3.1.	Tahapan Penelitian .....	17
3.2.	Kegiatan Penelitian .....	18
3.2.1.	Identifikasi Masalah.....	18
3.2.2.	Persiapan Data .....	18
3.2.3.	Penerapan Algoritma .....	19
3.2.4.	Evaluasi <i>Cluster</i> .....	20
3.2.5.	Hasil Penelitian dan Kesimpulan.....	20
3.3.	Alat Bantu Penelitian .....	21
3.4.	Jadwal Penelitian.....	21
BAB 4	.....	23
4.1.	Persiapan Data.....	23
4.1.1.	Akuisisi Data.....	23
4.1.2.	Praproses Data .....	24
4.2.	Penerapan Algoritma.....	31
4.2.1.	Pembuatan Model .....	31
4.2.2.	Penerapan Elbow Method.....	33
4.2.3.	Penerapan K-Means .....	33
4.3.	Evaluasi <i>Cluster</i> .....	35
4.3.1.	Visualisasi <i>Cluster</i> .....	35
4.3.2.	Pengujian Kualitas <i>Cluster</i> .....	35
4.4.	Hasil Penelitian .....	38
BAB 5	.....	59
5.1.	Kesimpulan.....	59

5.2. Saran.....	59
DAFTAR PUSTAKA .....	61
DAFTAR RIWAYAT HIDUP.....	64
LAMPIRAN.....	65

## DAFTAR TABEL

Tabel 3.1. Tabel Jadwal Penelitian .....	21
Tabel 4.1. Daftar Kolom di Dataset .....	23
Tabel 4.2. Data Utama .....	24
Tabel 4.3. Data bidang usaha .....	24
Tabel 4.4. Jenis usaha mikro .....	25
Tabel 4.5. Data bidang usaha dengan jenisnya .....	26
Tabel 4.6. Data utama dengan jenis .....	26
Tabel 4.7. Data Utama tanpa duplikasi dan kolom kosong .....	27
Tabel 4.8. Data final.....	28
Tabel 4.9. Data dengan <i>cluster</i> .....	38
Tabel 4.10. Label <i>cluster</i> .....	46
Tabel 4.11. Data <i>cluster 2</i> .....	47
Tabel 4.12. Jenis usaha mikro yang diminati.....	48

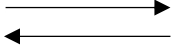



## DAFTAR GAMBAR

Gambar 3.1. Tahap Penelitian.....	17
Gambar 3.2. Tahap Algoritma K-Means .....	20
Gambar 4.1. Data dalam bentuk dataframe <i>python</i> .....	29
Gambar 4.2. Tipe data setiap kolom .....	29
Gambar 4.3. Deskripsi data kolom jumlah .....	30
Gambar 4.4. Dataframe dengan kolom jumlah setelah normalisasi .....	31
Gambar 4.5. Dataframe dengan kolom RASIO_DESA.....	32
Gambar 4.6. Visualisasi Elbow Plot .....	33
Gambar 4.7. Penerapan <i>K-Means Clustering</i> di <i>python</i> .....	34
Gambar 4.8. Dataframe dengan kolom label <i>cluster</i> .....	34
Gambar 4.9. Visualisasi Data <i>cluster</i> menggunakan scatter plot.....	35
Gambar 4.10. Skor Davies Bouldin Index .....	36
Gambar 4.11. Skor <i>Silhouette Coefficient</i> .....	36
Gambar 4.12. Visualisasi skor <i>Davies Bouldin Index</i> beberapa <i>cluster</i> .....	37
Gambar 4.13. Visualisasi skor <i>Silhouette Coefficient</i> beberapa <i>cluster</i> .....	37
Gambar 4.14. <i>Pie chart</i> perbandingan jumlah data di setiap <i>cluster</i> .....	47
Gambar 4.15. <i>Bar chart</i> perbandingan total jumlah jenis usaha.....	49
Gambar 4.16. Bar chart jenis usaha Bojongsari.....	50
Gambar 4.17. Bar chart jenis usaha Dukuh .....	50
Gambar 4.18. Bar chart jenis usaha Karanganyar.....	51
Gambar 4.19. Bar chart jenis usaha Karangmalang.....	51
Gambar 4.20. Bar chart jenis usaha Karangsong.....	52
Gambar 4.21. Bar chart jenis usaha Kepandean .....	52
Gambar 4.22. Bar chart jenis usaha Lemahabang.....	53
Gambar 4.23. Bar chart jenis usaha Lemahmekar .....	53
Gambar 4.24. Bar chart jenis usaha Margadadi .....	54
Gambar 4.25. Bar chart jenis usaha Pabean Udik.....	54
Gambar 4.26. Bar chart jenis usaha Paoman .....	55
Gambar 4.27. Bar chart jenis usaha Pekandangan .....	55
Gambar 4.28. Bar chart jenis usaha Pekandangan Jaya.....	56
Gambar 4.29. Bar chart jenis usaha Plumbon.....	56
Gambar 4.30. Bar chart jenis usaha Singajaya.....	57

Gambar 4.31. Bar chart jenis usaha Singaraja .....	57
Gambar 4.32. Bar chart jenis usaha Singaraja .....	58
Gambar 4.33. Bar chart jenis usaha Teluk Agung .....	58

## DAFTAR SIMBOL

### 1. Flow Chart

No	Nama	Simbol	Penjelasan
1.	<i>Arus Flow Direction</i>		Untuk menjelaskan arus prosedur dari atas kebawah maupun dari kanan kekiri dan berlaku sebaliknya yang sering juga disebut sebagai <i>connecting line</i> .
2.	<i>Processing</i>		Berfungsi untuk pemrosesan data oleh <i>user</i> maupun komputer (sistem).
4.	<i>Input-Output</i>		Berfungsi untuk mendeskripsikan data yang digunakan.
5.	<i>Terminator</i>		Berfungsi untuk memulai atau mengakhiri suatu kegiatan



## DAFTAR LAMPIRAN

Lampiran 1. Data UM Tahun 2020 Kecamatan Indramayu.....	65
Lampiran 2. Source Code.....	77
Lampiran 3. Hasil Turnitin .....	81