

**PENERAPAN DATA MINING UNTUK MENENTUKAN TINGKAT
KEMISKINAN MENURUT PROVINSI DI INDONESIA MENGGUNAKAN
K-MEANS**

SKRIPSI



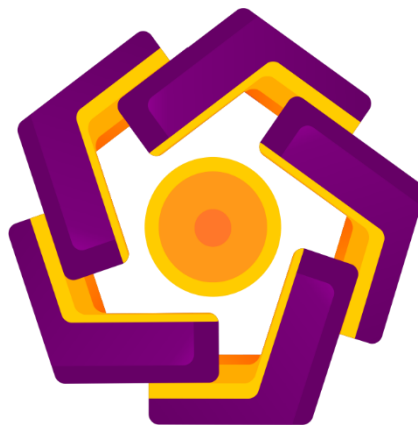
**disusun oleh
Agnes Dhea Hefra
20.21.1457**

**FALKUTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENERAPAN DATA MINING UNTUK MENENTUKAN TINGKAT
KEMISKINAN MENURUT PROVINSI DI INDONESIA MENGGUNAKAN
K-MEANS**

SKRIPSI

untuk memenuhi sebagian persyaratan mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Agnes Dhea Hefra
20.21.1457

FALKUTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022

PERSETUJUAN

SKRIPSI

**PENERAPAN DATA MINING UNTUK MENENTUKAN
TINGKAT KEMISKINAN MENURUT PROVINSI DI
INDONESIA MENGGUNAKAN K-MEANS**

yang dipersiapkan dan disusun oleh

Agnes Dhea Hefra
20.21.1457

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Agustus 2022

Dosen Pembimbing,

Ike Verawati, M.Kom
NIK. 190302237

PENGESAHAN

SKRIPSI

PENERAPAN DATA MINING UNTUK MENENTUKAN TINGKAT KEMISKINAN MENURUT PROVINSI DI INDONESIA MENGGUNAKAN K-MEANS

yang dipersiapkan dan disusun oleh

Agnes Dhea Hefra

20.21.1457

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, M.Kom

NIK. 190302185

Supriatin, M.Kom

NIK. 190302239

Ike Verawati, M.Kom

NIK. 190302237

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

Tanggal 26 Agustus 2022

DEKAN FALKUTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom.,M.Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Pangkalpinang, 20 Agustus 2022



Agnes Dhea Hefra

NIM. 20.21.1457

MOTTO

“Manusia yang berkualitas adalah manusia yang mampu mematahkan ketidakmampuan dirinya”

(Agnes Dhea Hefra)

“Sukses bukan siapa dirimu dan apa gelarmu tapi bagaimana kamu mau berusaha belajar dan menikmati prosesnya”

(Agnes Dhea Hefra)

“Jika hari ini kamu mengalami kegagalan, maka bangkitlah ada keberhasilan yang menunggumu di hari esok”

(Agnes Dhea Hefra)



PERSEMBAHAN

Puji Syukur saya panjatkan kepada Tuhan Yang Maha Esa, atas berkat dan rahmat-Nya saya telah berhasil menyelesaikan Tugas Akhir ini. Saya persembahkan Tugas Akhir ini untuk :

- Kedua orang tua yang selalu mendukung saya, baik dalam doa maupun perbuatan.
- Keluarga besar saya yang selalu mendukung saya selama kuliah.
- Dosen kami tercinta, Ibu Ike Verawati, M.Kom. selaku dosen pembimbing yang telah membantu penulis dalam pembuatan Skripsi ini.
- Paulus Elimson, Andre Atmaja Pandu Pratama telah memberikan *support* dalam memberikan masukan mengenai skripsi yang telah saya buat.
- Diri saya sendiri terimakasih sudah berjuang sejauh ini dan terimakasih sudah menjadikan kelulusan kado Ulang Tahun di tahun ini
- Seluruh Dosen Universitas Amikom Yogyakarta atas ilmu yang telah diberikan selama masa kuliah saya.
- Teman-teman saya di Transfer Informatika 2020 dan teman-teman di Universitas Amikom Yogyakarta yang mensupport saya segera menyelesaikan tugas akhir ini.

KATA PENGANTAR

Puji Bagi Allah sang pencipta Alam Semesta yang Maha Pengasih dan lagi Penyayang, Karena atas pertolongan-Nya kami dapat menyelesaikan Skripsi dengan judul “Penerapan Data Mining Untuk Menentukan Tingkat Kemiskinan Menurut Provinsi Di Indonesia Menggunakan K-Means ”, kemudian Shalawat dan salam Penulis haturkan kepada Baginda Rasul Muhammad Sallahualaihi wasallam, keluarga, para Sahabat serta seluruh Ummatnya.

Dengan selesainya Tugas Akhir. Penulis mengucapkan banyak terimakasih kepada:

1. Prof.Dr.M. Suyanto, MM Selaku Rektor Universitas Amikom Yogyakarta yang telah memberikan kesempatan kepada Penulis untuk menimba ilmu di kampus.
2. Hanif Al Fatta. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, serta selaku dosen Pembimbing yang telah membantu dan memberi dukungan kepada penulis di dalam proses penulisan Tugas Akhir ini.
3. Windha Mega Pradnya Duhita, M.Kom Selaku ketua Program Studi S1 Informatika.
4. Pihak Prodi S1 Informatika Universitas Amikom Yogyakarta, terimakasih banyak karena telah mengijinkan penulis untuk melakukan penelitian.
5. Dosen-dosen yang telah mendidik, mengajarkan ilmunya kepada Penulis selama proses belajar di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih jauh dari sempurna, keterbatasan kemampuan dan pengetahuan penulis merupakan faktor utama ketidak mampuan ini.

DAFTAR ISI

PENERAPAN DATA MINING UNTUK MENENTUKAN TINGKAT KEMISKINAN MENURUT PROVINSI DI INDONESIA MENGGUNAKAN K-MEANS.....	ii
PENERAPAN DATA MINING UNTUK MENENTUKAN TINGKAT KEMISKINAN MENURUT PROVINSI DI INDONESIA MENGGUNAKAN K-MEANS.....	ii
PERSETUJUAN.....	iii
PENGESAHAN.....	iv
PERNYATAAN.....	v
MOTTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
INTISARI.....	xvi
ABSTRACT	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian.....	3

1.6.1	Metode Pengumpulan Data	4
1.6.2	Metode Studi Literatur	4
1.6.3	Metode Analisis	4
1.6.4	Metode Perancangan	5
1.6.5	Implementasi Sistem	5
1.6.6	Pengujian Sistem	5
1.7	Sistematika Penulisan	5
BAB II LANDASAN TEORI		7
2.1	Kajian Pustaka.....	7
2.2	Data mining	9
2.2.1	Pengertian Data Mining	9
2.2.2	Fungsi Data Mining.....	11
2.2.3	Ilmu-Ilmu Yang Berkaitan Dengan Data Mining.....	13
2.2.4	Teknik Pembelajaran Data Mining.....	14
2.3	Clustering	15
2.4	Metode K-means	16
2.4.1	Pengertian K-means.....	16
2.4.2	Kelebihan dan Kekurangan K-means.....	17
2.4.3	Tujuan K-means.....	18
2.4.4	Tahap Clustering Algoritma K-means.....	18
2.5	Kemiskinan	19
2.6	Flowchart.....	20
BAB III METODOLOGI PENELITIAN		22
3.1	Alur Penelitian	22
3.1.1	Studi Literatur.....	23

3.1.2	Pengumpulan Data.....	23
3.1.3	Preprocessing Data	23
3.1.4	Transformasi Data	23
3.1.5	Modeling.....	24
3.1.6	Implementasi	24
3.1.7	Evaluasi Hasil	24
3.2	Analisis Kebutuhan Sistem.....	24
3.2.1	Kebutuhan non Fungsional	24
3.2.1.1	Kebutuhan Perangkat Keras (Hardware)	25
3.2.1.2	Kebutuhan Perangkat Lunak (Software)	25
3.2.2	Bahan.....	25
3.3	Sumber Data	25
3.3.1	Sampel Data.....	26
3.4	K-means Clustering.....	27
BAB IV	PEMBAHASAN DAN HASIL.....	29
4.1	Pengumpulan Data	29
4.2	Preprocessing Data.....	30
4.3	Transformasi Data.....	31
4.4	Modelling.....	31
4.5	Implementasi.....	32
4.6	Perhitungan Algoritma simple K-Means	44
4.7	Evaluasi	50
4.7.1	Perhitungan Silhouette Index.....	50
BAB V	PENUTUP.....	55
5.1	Kesimpulan	55

5.2	Saran.....	56
	DAFTAR PUSTAKA	57
	LAMPIRAN.....	59



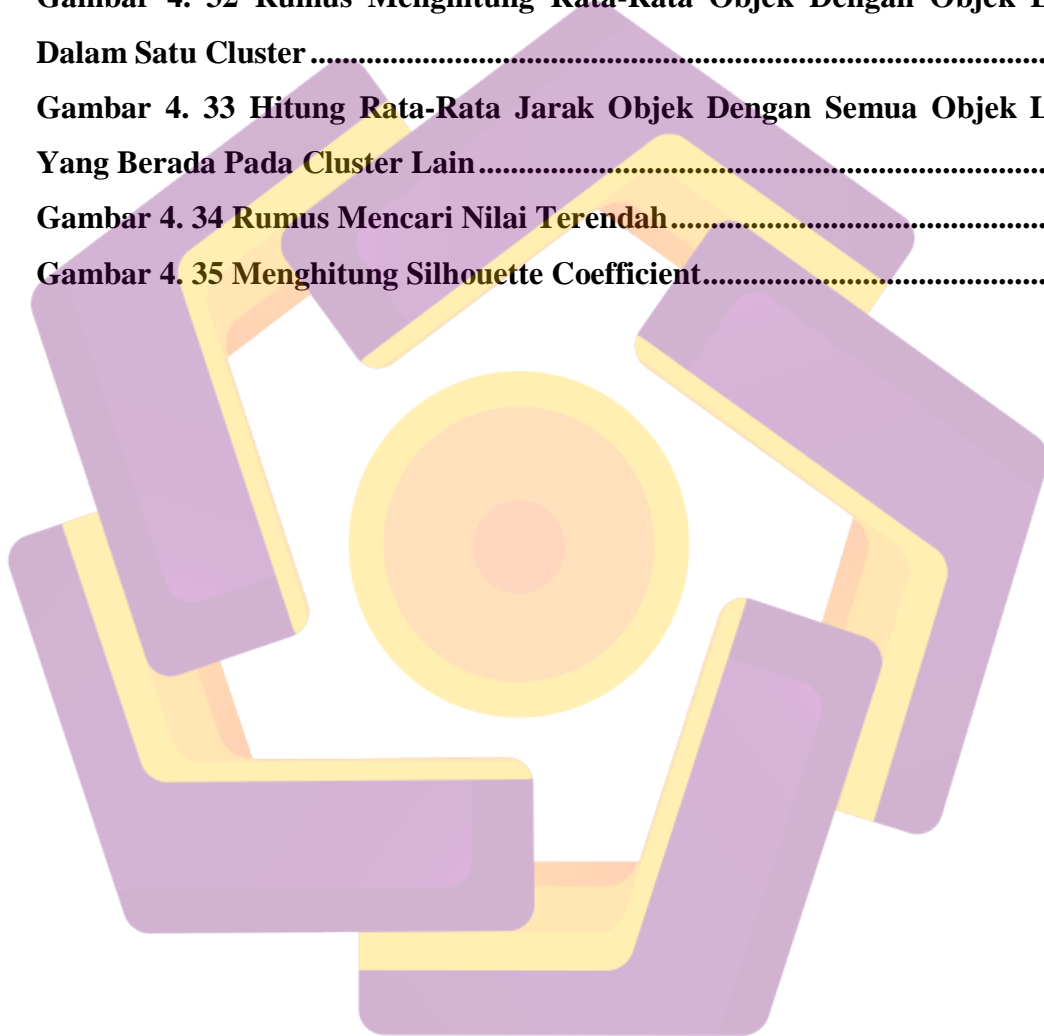
DAFTAR TABEL

Tabel 2. 1 Persamaan dan Perbedaan Penelitian.....	8
Tabel 2. 2 Daftar Simbol Flowchart Diagram.....	21
Tabel 3. 1 Sampel Data.....	26
Tabel 3. 2 Kriteria dan Bobot	28
Tabel 4. 1 Hasil Normalisasi Data.....	31
Tabel 4. 2 Sampel Data Badan Pusat Statistik 2018-2020.....	44
Tabel 4. 3 Parameter.....	45
Tabel 4. 4 Hasil Pemilihan Centroid Awal	46
Tabel 4. 5 Hasil Perhitungan Jarak Cluster Dengan Titik Centroid	46
Tabel 4. 6 Pengelompokan Cluster Hasil Perhitungan Centroid Pada Iterasi 1	47
Tabel 4. 7 Keanggotaan Cluster Hasil Perhitungan Centroid pada Iterasi 1..	47
Tabel 4. 8 Nilai Centroid dari Hasil Perhitungan Iterasi 1	47
Tabel 4. 9 Hasil Perhitungan Jarak Cluster dengan Titik Centroid Sebelumnya Untuk Menentukan Centroid pada Iterasi 2	48
Tabel 4. 10 Pengelompokan Cluster Hasil Perhitungan Centroid Pada Iterasi 2.....	Error! Bookmark not defined.
Tabel 4. 11 Keanggotaan Cluster Hasil Perhitungan Centroid pada Iterasi 2	49
Tabel 4. 12 Nilai Centroid yang Dihasilkan Pada Iterasi 2	49
Tabel 4. 13 Hasil Perhitungan Jarak Cluster Untuk Menentukan Centroid pada Iterasi 3	50
Tabel 4. 14 Nilai Centroid yang Dihasilkan Pada Iterasi 3	50
Tabel 4. 15 Hasil Perhitungan Silhouette Index.....	52
Tabel 4. 16 Hasil Silhouette Index Setiap Variabel.....	53
Tabel 4. 17 Hasil Pola Cluster	53

DAFTAR GAMBAR

Gambar 2. 1 Penemuan Ilmu Baru	10
Gambar 2. 2 Fungsi Data Mining	12
Gambar 2. 3 Tahap Algoritma K-means	18
Gambar 3. 1 Alur Penelitian.....	22
Gambar 4. 1 Data BPS.....	29
Gambar 4. 2 Nilai Rata-Rata Tahun 2018-2020.....	30
Gambar 4. 3 Mencari Missing Value.....	30
Gambar 4. 4 Grafik Cluster Optimal Dengan Metode Elbow	31
Gambar 4. 5 Dataset pada tampilan Microsoft Excel.....	32
Gambar 4. 6 Import Library.....	33
Gambar 4. 7 Bentuk Dataframe dari Dataset	33
Gambar 4. 8 Perintah Dataset.....	34
Gambar 4. 9 Tampilan Dataframe	34
Gambar 4. 10 Tampilan Info Dari Dataset.....	35
Gambar 4. 11 Perintah Frekuensi HLS	35
Gambar 4. 12 Tampilan Histogram HLS	36
Gambar 4. 13 Perintah Frekuensi PPK	36
Gambar 4. 14 Tampilan Histogram PPK	37
Gambar 4. 15 Perintah Frekuensi JPM	37
Gambar 4. 16 Tampilan Histogram JPM	38
Gambar 4. 17 Perintah Multivariate Analysis	38
Gambar 4. 18 Tampilan Pair Plot	39
Gambar 4. 19 Perintah Minmax Scaler	39
Gambar 4. 20 Hasil Minmax Scaler	40
Gambar 4. 21 Perintah Modeling	40
Gambar 4. 22 Metode Elbow.....	40
Gambar 4. 23 Perintah SSE	41
Gambar 4. 24 Nilai Optimum	41
Gambar 4. 25 Perintah K-means	42
Gambar 4. 26 Perintah Nilai Pusat.....	42

Gambar 4. 27 Centroid Awal	42
Gambar 4. 28 Cetak Cluster	42
Gambar 4. 29 K-means	43
Gambar 4. 30 Perintah Grafik Scatterplot 3d.....	43
Gambar 4. 31 Grafik Scatter Plot Hasil Akhir.....	44
Gambar 4. 32 Rumus Menghitung Rata-Rata Objek Dengan Objek Lain Dalam Satu Cluster	51
Gambar 4. 33 Hitung Rata-Rata Jarak Objek Dengan Semua Objek Lain Yang Berada Pada Cluster Lain.....	51
Gambar 4. 34 Rumus Mencari Nilai Terendah.....	52
Gambar 4. 35 Menghitung Silhouette Coefficient.....	52



INTISARI

Kemiskinan merupakan masalah yang cukup serius terutama untuk di Negara berkembang seperti Indonesia. Data kemiskinan di setiap provinsi akan berbeda. Hal ini dipengaruhi oleh banyak indikator pendukungnya. Dengan menentukan dan mengukur indikator-indikator kemiskinan, maka akan mempermudah mengenal tingkat kemiskinan suatu wilayah. Dengan dikelompokannya karakteristik suatu wilayah berdasarkan indikator kemiskinan, maka pemerintah dapat tepat dan cepat dalam mengambil kebijakan untuk menanggulangi kemiskinan di suatu wilayah. Teknik pemanfaatan data menjadi sebuah informasi baru disebut data mining.

Salah satu metode data mining yang cukup populer yaitu clustering dengan menggunakan algoritma k-means. K-means dapat mengelola data tanpa diberitahu lebih dahulu label kelasnya. Data penelitian ini menggunakan data yang diterbitkan oleh BPS (Badan Pusat Statistik) di tahun 2018-2020 yang terdiri dari 34 provinsi dengan 3 variabel. Variabel yang penulis gunakan adalah angka harapan lama sekolah menurut jenis kelamin (tahun) dan pengeluaran per kapita menurut jenis kelamin (ribu rupiah per tahun) dan jumlah penduduk miskin (ribu jiwa) menurut provinsi dan daerah. Cluster yang digunakan dalam penelitian ini adalah sebanyak 3 kelompok.

Hasil analisis menunjukkan bahwa sebanyak 4 provinsi termasuk pada cluster rata-rata tingkat kemiskinan tinggi, 12 provinsi termasuk pada cluster rata-rata tingkat sedang dan 18 provinsi lainnya termasuk pada cluster rata-rata tingkat kemiskinan rendah.

Kata Kunci : Data Mining, Kemiskinan, K-means

ABSTRACT

Poverty is a serious problem, especially in developing countries like Indonesia. Poverty data in each province will be different. This is influenced by many supporting indicators. By determining and measuring poverty indicators, it will be easier to identify the poverty level of a region. By grouping the characteristics of a region based on poverty indicators, the government can be precise and fast in making policies to reduce poverty in a region. The technique of utilizing data into new information is called data mining.

One of the most popular data mining methods is clustering using the k-means algorithm. K-means can manage data without notifying the class label in advance. This research data uses data published by BPS (Central Statistics Agency) in 2018-2020 which consists of 34 provinces with 3 variables. The variables that the writer uses are the expected years of schooling by gender (years) and expenditure per capita by gender (thousands of rupiah per year) and the number of poor people (thousands of people) by province and region. The clusters used in this study were 3 groups.

The results of the analysis show that as many as 4 provinces are included in the cluster of high average poverty rates, 12 provinces are included in the average cluster of medium levels and 18 other provinces are included in the cluster average low poverty levels.

Keywords: Data Mining, Poverty, K-means