

PENERAPAN K-MEANS CLUSTERING PADA KLASIFIKASI RISIKO KESEHATAN IBU HAMIL

Ni Made Novia Nurtiani^{a1}, I Gede Santi Astawa, S.T., M.Cs.^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas
Udayana
Badung, Bali, Indonesia
¹madenovia2000@gmail.com
²santi.astawa@unud.ac.id

Abstract

The health condition of pregnant women greatly affects the growth and development of the fetus in the womb. There are many cases of maternal and infant deaths that occur in the world. Both caused by maternal health conditions during pregnancy and childbirth. Several factors affect the health condition of pregnant women, namely age, blood pressure, blood sugar levels in the body of pregnant women, body temperature of pregnant women. This study will apply the K-Means method to classify the health risks of pregnant women. The author will also use the Elbow method to find the right cluster to be classified. Based on the research conducted, it was found that 891 out of 1014 data were correctly labeled with an accuracy of 88% and the right number of clusters was 4.

Keywords: *K-Means Clustering, Klasifikasi, Maternal, Elbow Method, Cluster*

1. Pendahuluan

Kehamilan memiliki pengertian di mana terjadi pembuahan oleh sel sperma dengan sel telur yang nantinya berkembang menjadi janin pada dinding rahim. Setiap harinya, janin tersebut akan terus mengalami pertumbuhan dan perkembangan hingga waktunya dilahirkan ke dunia. Pertumbuhan dan perkembangan yang dialami oleh janin sangat dipengaruhi oleh kondisi fisik maupun mental ibu. Risiko mengalami gangguan kesehatan yang tertinggi apabila usia kehamilan pada saat hamil muda, hamil pertengahan, dan setelah persalinan. Menurut Kementerian Kesehatan (2021), kejadian ibu dan bayi yang meninggal di rumah sakit melebihi persentase 62%. Berdasarkan data yang dari Pusat Data dan Informasi (2012) kematian ibu saat persalinan disebabkan oleh pendarahan sebesar 28%, eklamsia dengan besar persentase 24%, dan terakhir infeksi sebesar 11%. Faktor lainnya yang secara tidak langsung menyebabkan kematian pada ibu adalah berkurangnya energi secara kronis saat masa hamil 57% serta mengalami kekurangan hemoglobin dalam darah selama masa hamil sebesar 40%.

K-Means merupakan sebuah metode klusterisasi yang bertujuan untuk mengelompokkan banyak data yang kedalam beberapa kelompok dengan karakteristik yang sama. Metode K-Means akan membagi data menjadi beberapa kelompok sehingga masing-masing data dengan karakteristik sama menjadi satu kelompok dan data dengan karakteristik berbeda dikelompokkan pada kelompok lainnya. K-Means clustering diawali dengan memilih jumlah kluster dan memilih centroid secara acak, yang kemudian akan dihitung nilai rerata jarak dari data terhadap kluster terdekat.

Penelitian sebelumnya melakukan penerapan metode K-Means pada pengklasifikasian data mahasiswa yang dilakukan oleh Totok Suprawoto (2016) dengan penelitian yang berjudul "Klasifikasi Data Mahasiswa Menggunakan Metode Kmeans Untuk Menunjang Pemilihan Strategi Pemasaran". Pada penelitian tersebut didapatkan hasil rerata IPK dengan pengelompokkan mahasiswa ke dalam 3 area. Pada penelitian yang dilakukan oleh Somantri dkk. (2016) dengan

judul “Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)”. Pada penelitian Somantri dkk. (2016), pengoptimasian terhadap hasil pengelompokan dilakukan dengan mengaplikasikan metode Support Vector Machine (SVM). Penelitian tersebut menghasilkan tingkat akurasi optimasi sebesar 86,21%. Berdasarkan penjelasan yang telah dijabarkan, penulis akan melakukan sebuah penelitian untuk melakukan pengklasifikasian terhadap risiko kesehatan ibu hamil dengan menerapkan metode K-Means Clustering, dengan judul penelitian, yaitu “Penerapan K-Means Clustering pada Klasifikasi Risiko Kesehatan Ibu Hamil”.

2. Metode Penelitian

Penelitian ini menerapkan metode K-Means untuk mengklasifikasi risiko kesehatan ibu hamil. Metode K-Means akan digunakan untuk mencari tingkat risiko dari kesehatan ibu hamil, kemudian dicari jumlah data yang memiliki tingkat yang sama dan dicari persentase keakurasian dari penggunaan model tersebut. Selain itu, metode Elbow akan digunakan untuk mencari jumlah kluster yang tepat untuk melakukan klasifikasi.

2.1 K-Means Clustering

Algoritma K-Means Clustering merupakan sebuah algoritma unsupervised learning yang bertujuan untuk mengelompokkan banyak data acak ke dalam beberapa kelompok yang di dalamnya berisi data yang memiliki karakteristik yang sama. K-Means Clustering adalah metode non-hierarchy. Pengelompokan dilakukan dengan melakukan teknik pengambilan sampel secara acak kemudian akan dihitung jarak antar data dengan sampel tersebut. Berikut merupakan konsep dasar dari metode K-Means Clustering.

- Menetapkan jumlah kluster.
- Memilih sembarang titik untuk dijadikan titik centroid dan mengalokasikan setiap data sesuai dengan cluster terdekat.
- Menghitung rata-rata jarak setiap kluster dengan data yang tergabung. Kemudian menggeser centroid ke means yang baru.
- Mengalokasikan kembali data dengan cluster terdekat.
- Mengulangi proses ke-3 hingga tidak ada perubahan kluster.

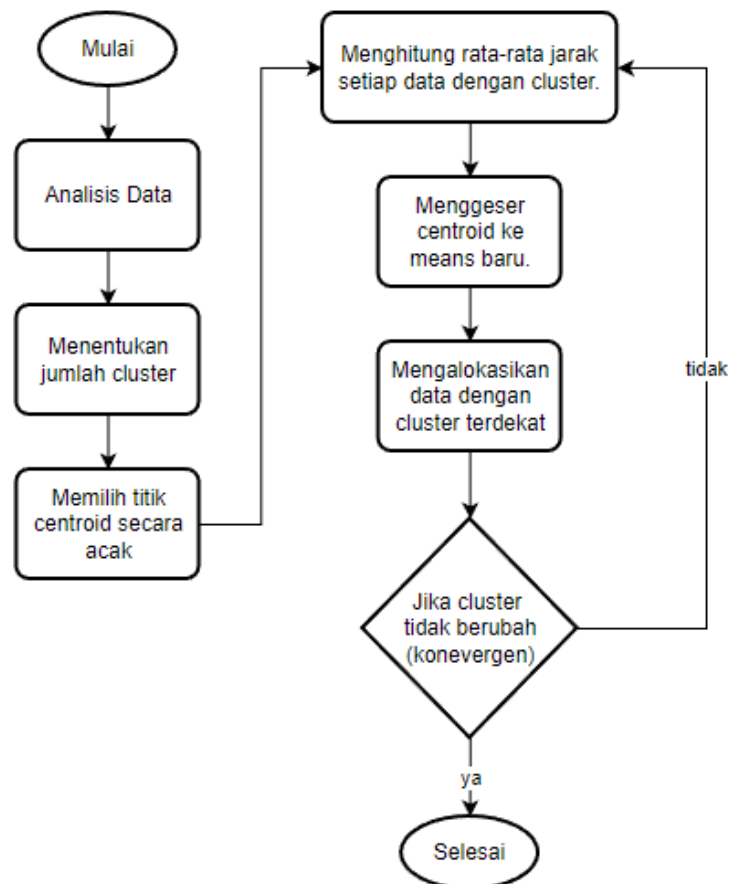
Jarak dari setiap data dengan centroid dapat dihitung dengan meaplikasikan rumus Manhattan Distance atau Euclidean Distance.

- Manhattan Distance:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

- Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$



Gambar 1. Flowchart

2.2 Pengumpulan Data

Pada penelitian ini, data yang digunakan merupakan dataset Maternal Health Risk yang didapatkan dari *open source kaggle*. Jumlah data yang tersedia dalam dataset tersebut sebanyak 1014 records dengan 7 buah atribut. Atribut pada dataset tersebut adalah *Age*, *SystolicBP*, *DiastolicBP*, *BS*, *BodyTemp*, *HeartRate*, dan *RiskLevel*.

Tabel 1. Dataset Maternal Health Risk

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
25	130	80	15.0	98.0	86	high risk
35	140	90	13.0	98.0	70	high risk
29	90	70	8.0	100.0	80	high risk
30	140	85	7.0	98.0	70	high risk
35	120	60	6.1	98.0	76	low risk
...
22	120	60	15.0	98.0	80	high risk
55	120	90	18.0	98.0	60	high risk
35	85	60	19.0	98.0	86	high risk
43	120	90	18.0	98.0	70	high risk
32	120	65	6.0	101.0	76	mid risk

2.3 Pembersihan Data

Analisis data yang dilakukan pada dataset tersebut yaitu dengan melakukan pembersih terhadap dataset. Pembersihan dilakukan dengan mencari atribut-atribut yang tidak dibutuhkan pada penelitian ini. Selain itu, atribut RiskLevel yang memiliki tipe nominal diubah menjadi data numerikal. Berikut adalah tampilan dataset Maternal Health Risk.

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15.0	98.0	86	high risk
1	35	140	90	13.0	98.0	70	high risk
2	29	90	70	8.0	100.0	80	high risk
3	30	140	85	7.0	98.0	70	high risk
4	35	120	60	6.1	98.0	76	low risk

Gambar 2. Tampilan dataset awal

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15.0	98.0	86	0
1	35	140	90	13.0	98.0	70	0
2	29	90	70	8.0	100.0	80	0
3	30	140	85	7.0	98.0	70	0
4	35	120	60	6.1	98.0	76	1

Gambar 3. Tampilan dataset setelah dilakukan pembersihan

3. Hasil dan Pembahasan

3.1. Penskalaan Data

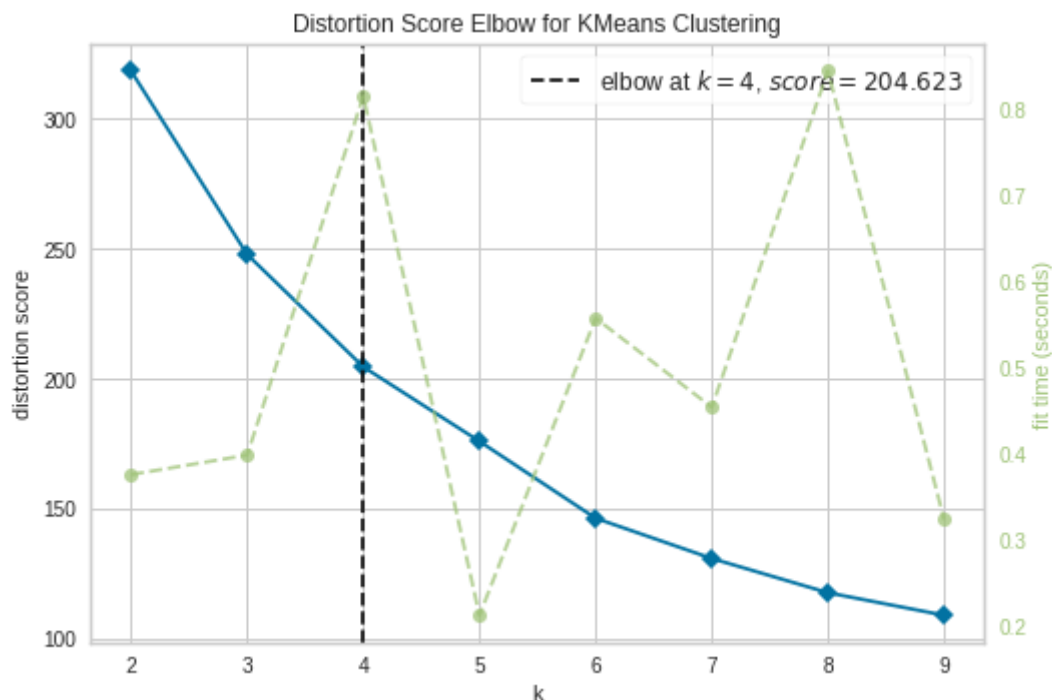
Penskalaan data dilakukan agar data sebelumnya yang telah diubah menjadi data numerik berada pada rentang nilai (skala) yang sama. Berikut adalah tampilan dari dataset setelah dilakukan penskalaan terhadap data tersebut.

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	0.250000	0.666667	0.607843	0.692308	0.0	0.951807	0.0
1	0.416667	0.777778	0.803922	0.538462	0.0	0.759036	0.0
2	0.316667	0.222222	0.411765	0.153846	0.4	0.879518	0.0
3	0.333333	0.777778	0.705882	0.076923	0.0	0.759036	0.0
4	0.416667	0.555556	0.215686	0.007692	0.0	0.831325	0.5

Gambar 4. Tampilan dataset setelah dilakukan pendataan

3.2. Mencari Jumlah Cluster

Metode Elbow adalah sebuah metode yang dapat menyarankan jumlah cluster yang tepat untuk model K-Means. Semakin sedikit jumlah cluster, data dalam satu cluster semakin seragam, begitupun sebaliknya, semakin banyak jumlah cluster, semakin beragam data dalam sebuah cluster. Pada penelitian ini didapatkan jumlah cluster yaitu 4, dengan gambar grafik sebagai berikut.



Gambar 5. Grafik Metode Elbow

3.3. K-Means dengan 4 cluster

Setelah mendapatkan jumlah kluster dengan metode Elbow, selanjutnya dilakukan pengaplikasian K-Means terhadap semua data yang ada dalam dataset Maternal Health Risk. Sehingga didapatkan hasil di mana sebanyak 891 dari 1014 sampel dapat dilabeli dengan benar. Kemudian dihitung akurasi dari penerapan model tersebut sebesar 88%.

4. Simpulan

Berdasarkan penelitian yang dilakukan, dapat dikatakan bahwa penerapan metode K-Means pada klasifikasi risiko kesehatan ibu hamil berhasil dilakukan dengan menggunakan metode Elbow untuk pencarian jumlah cluster yang tepat sehingga didapatkan hasil sebanyak 891 dari 1014 sampel data dilabeli dengan benar dan hasil akurasi sebesar 88%.

Referensi

- [1] Oman Somantri, Slamet Wiyono dan Dairoh "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)" *Scientific Journal of Informatics*, vol. 3, no. 1, 2016.
- [2] Totok Suprawoto, "Klasifikasi Data Mahasiswa Menggunakan Metode Kmeans Untuk Menunjang Pemilihan Strategi Pemasaran" *Jurnal Informatika dan Komputer (JIKO)*, vol. 1, no. 1, 2016.
- [3] Rokom, "sehatnegeriku.kemkes.go.id", 15 September 2021. [Daring]. Tersedia: <https://sehatnegeriku.kemkes.go.id/baca/umum/20210914/3738491/kemenkes-perkuat-upaya-penyelamatan-ibu-dan-bayi/> [Diakses : 30 September 2022]

- [4] Widya Sri Mulyaningsih, "medium.com", 6 Januari 2019. [Daring]. Tersedia: <https://medium.com/@16611021/klasifikasi-menggunakan-k-means-dan-naive-bayes-dalam-python-975b25e5c432> [Diakses: 30 September 2022]