

IDENTIFIKASI PENYAKIT RETINOPATI DIABETIKA MENGGUNAKAN ALGORITMA KERNEL K-MEANS

Edi Prayitno

Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
e-mail: ediprayitno@mhs.unesa.ac.id

Abstrak

Retinopati diabetika merupakan suatu gejala komplikasi dari penyakit *Diabetes Mellitus* yang menyebabkan retina tidak dapat mengirimkan gambar penglihatan ke otak secara normal akibat peningkatan glukosa pada darah. Identifikasi penyakit retinopati diabetika masih sering dibahas oleh banyak peneliti. Dalam skripsi ini algoritma *Kernel K-Means* akan digunakan untuk mengidentifikasi penyakit retinopati diabetika. Algoritma *Kernel K-Means* merupakan algoritma pengembangan dari algoritma *K-Means* yang direalisasikan melalui pernyataan jarak dalam bentuk fungsi *Kernel*. Dataset yang akan digunakan dalam skripsi ini yaitu dataset *Diabetic Retinopathy Debrecen* diambil dari UCI Repository Machine Learning sebanyak 1151 data yang terdiri dari 19 atribut dan 1 atribut menunjukkan kelas yang akan dibentuk. Identifikasi penyakit retinopati diabetika dimulai dengan memisahkan data input dan target output. Data hasil input dinormalisasi kemudian dilakukan proses identifikasi menggunakan algoritma *Kernel K-Means* dengan menentukan banyaknya iterasi dan sub pusat *cluster* (nilai *M*). Performa hasil identifikasi menggunakan algoritma *Kernel K-Means* memberikan hasil akurasi tertinggi sebesar 82.88% dengan waktu eksperimen 12.22 detik pada iterasi 10 dan nilai *M*=300.

Kata kunci : identifikasi, retinopati diabetika, *Kernel K-Means*

Abstract

Diabetic Retinopathy is a phenomenon of a complication of Diabetes Mellitus disease that causes retina can not transmit visual images to the brain normally due to increased glucose in the blood. The identification of diabetic retinopathy disease is still frequently discussed by many researchers. In this thesis *Kernel K-Means* algorithm will be used to identify diabetic retinopathy. *Kernel K-Means* algorithm is the modified algorithm of *K-Means* algorithm realized through distance statement in *Kernel* function form. The dataset that used in this thesis is the *Diabetic Retinopathy Debrecen* dataset taken from UCI Repository Machine Learning totaling 1151 data consisting of 19 attributes and 1 attribute showing the class that will be formed. The identification of diabetic retinopathy disease begins by separating the input data and target output. The data of the normalized input then performed the identification process using *Kernel K-Means* algorithm by determining the number of iterations and sub center cluster (*M* value). The performance of identification using *Kernel K-Means* algorithm gives the highest accuracy result until 82.88% with experimental time 12.22 seconds on iteration 10th and value *M* = 300.

Keywords : identification, diabetic retinopathy, kernel k-means

PENDAHULUAN

Diabetes Mellitus (DM) adalah penyakit gangguan metabolik akibat pankreas tidak cukup memproduksi insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif (INFODATIN, 2014). Menurut *International Diabetes Federation* (IDF), pada tahun 2015 sebesar 415 juta orang di dunia terkena diabetes. Diperkirakan pada tahun 2040 penderita diabetes akan meningkat menjadi 642 juta orang (Kemenkes RI. 2014:1).

World Health Organization (WHO) melaporkan bahwa posisi Indonesia berada pada urutan ke tujuh dunia untuk prevalensi penderita diabetes tertinggi di dunia bersama dengan negara India, China, Rusia, Amerika Serikat, Brazil dan Meksiko dengan jumlah estimasi orang terjangkit diabetes sebesar 10 juta (IDF Atlas 2015). Kecenderungan peningkatan prevalensi orang yang memiliki penyakit DM di Indonesia dapat dilihat dari kenaikan presentase 5,7% tahun 2007 menjadi 6,9% tahun 2016. Berdasarkan *The DiabCare Asia 2008 Study* yang melibatkan 1785 penderita DM di

Indonesia, sebanyak 42% penderita DM di Indonesia mengalami komplikasi retinopati diabetika (Soewondo, P., dkk. 2010:239).

Retinopati diabetika merupakan suatu komplikasi dari penyakit DM yang menyebabkan retina tidak dapat mengirimkan gambar penglihatan ke otak dengan normal akibat peningkatan glukosa pada darah. Efek visual dari kebocoran digambarkan dengan adanya perubahan pada retina seperti perubahan venous, *cotton wool spots*, mikroaneurisma, *hard exudate* dan pendarahan (Faust, O., et al. 2013:3). Kendala yang saat ini ada yaitu keterbatasan dokter spesialis mata untuk mendeteksi dini penyakit retinopati diabetika sehingga dapat segera diobati. Pada penelitian ini akan mengidentifikasi apakah seseorang tergolong penderita retinopati diabetika atau normal dari sebuah data. Data yang digunakan pada penelitian ini berasal dari UCI (University California Irvine) Repository Machine Learning.

UCI Repository Machine Learning diperoleh dari Universitas Debrecen, Hongaria. Dalam penelitian ini data yang diperoleh dari UCI akan disimulasikan dengan algoritma *Kernel K-Means*. *Kernel K-Means* adalah pengembangan dari Algoritma *K-Means* yang menggunakan metode *Kernel* untuk memetakan data yang berdimensi tinggi pada space baru sehingga dapat dipisahkan secara *linear*. Hal ini dilakukan agar dapat meningkatkan hasil akurasi. *Clustering* dengan Algoritma *Kernel K-Means* akan mengidentifikasi ke dalam retinopati diabetika tidak ada (kelas 0), ataupun ada retinopati diabetika (kelas 1).

KAJIAN PUSTAKA

A. *Diabetes Mellitus*

Diabetes Mellitus (DM) merupakan suatu penyakit metabolik dengan karakteristik hiperglikemia yang terjadi akibat kelainan kerja insulin, sekresi insulin, atau kedua-duanya (Purnamasari, 2009). Menurut *World Health Organization* (WHO), DM didefinisikan sebagai gangguan metabolisme atau suatu penyakit kronis yang ditandai dengan kenaikan kadar gula darah disertai

dengan lipid, gangguan metabolisme karbohidrat, dan protein sebagai akibat dari insufisiensi fungsi insulin. Empat klasifikasi *Diabetes Mellitus*: *Diabetes Mellitus Gestasional* (*Diabetes* kehamilan), *Diabetes Mellitus* tipe 1, *Diabetes Mellitus* tipe 2, dan *Diabetes Mellitus* tipe khusus lain (ADA, 2009).

B. Retinopati Diabetika

Retinopati diabetika adalah salah satu komplikasi mikrovaskular pada *Diabetes Mellitus* (DM) tipe 1 dan tipe 2 yang terjadi akibat proses hiperglikemia dalam jangka waktu yang relatif lama. Retinopati Diabetika adalah penyakit mata yang diakibatkan oleh DM. Retinopati Diabetika pada awalnya menyebabkan pandangan menjadi kabur dan dapat berkembang menjadi kebutaan jika tidak segera diobati.



Gambar 1. Kondisi Retina Mata Terdeteksi Retinopati Diabetika

C. Algoritma *K-Means*

Algoritma *K-Means* diperkenalkan pada tahun 1976 oleh J.B. MacQueen, yaitu salah satu Algoritma *clustering* yang sederhana dan sangat umum yang mengelompokkan data sesuai dengan karakteristik ciri-ciri yang serupa. Pada Algoritma *K-Means* setiap obyek akan masuk ke kelompok tertentu dalam suatu proses tertentu sehingga data yang memiliki karakteristik berbeda dikelompokkan ke dalam cluster yang lain dan data yang mempunyai karakteristik sama dikelompokkan ke dalam kelompok yang sama. Algoritma *K-Means* ditunjukkan sebagai berikut :

Langkah 1 Menentukan banyak *K-cluster* yang ingin dibentuk.

Langkah 2 inialisasi data ke dalam *K-cluster* secara acak.

Langkah 3 Menghitung jarak data input terhadap masing-masing centroid menggunakan rumus jarak *Euclidian* (*Euclidian Distance*).

$$d(x_i, c_j) = \sqrt{(x_i - c_j)^2}$$

Langkah 4 Mengklasifikasikan setiap data dengan jarak terdekat.

Langkah 5 Mengupdate nilai centroid.

$$c_s(t + 1) = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$$

Langkah 6 Melakukan perulangan dari langkah 2 hingga langkah 5 hingga anggota tiap *cluster* tidak ada yang berubah.

D. Algoritma Kernel K-Means

K Means secara umum hanya dapat memisahkan data secara linier. Namun, untuk data yang non-linier *K Means* harus dimodifikasi agar dapat mengakomodasi data yang tidak linear. Metode yang tidak sama yang digunakan untuk melaksanakan algoritma ini dalam ruang *kernel* adalah *Kernel K-Means*. *Kernel K-Means* pada prinsipnya mirip dengan *K-Means*, Girolami (2002) pengembangan dari *K Means* ke *Kernel K-Means* direalisasikan melalui pernyataan jarak dalam bentuk fungsi *kernel*.

Metode yang digunakan dalam penelitian ini mengadopsi *Kernel Trick* dalam proses *Clustering*. *Kernel Trick* merupakan suatu metode yang digunakan untuk menghitung kesamaan dalam ruang yang ditransformasikan dengan menggunakan sekelompok atribut standar. Untuk menerapkan *Kernel Trick* dilakukan transformasi dimensi pada data *x* dengan didefinisikan sebagai $x = \phi(x)$. Dapat diketahui bahwa pada **Gambar 2.** dibawah merupakan diagram alur untuk mengidentifikasi penyakit retinopati diabetika. Dengan menggunakan jarak *Euclid* pada *K-Means* dan fungsi

pemetaan, Maka Algoritma *Kernel K-Means* ditunjukkan sebagai berikut :

Langkah 1 Tentukan jumlah *cluster k* dan banyaknya iterasi.

Langkah 2 Inialisasi *k* pusat *cluster* secara acak

Langkah 3 Alokasikan semua data ke cluster terdekat berdasarkan jarak yang dihitung menggunakan rumus fungsi pemetaan ϕ dan fungsi *kernel*

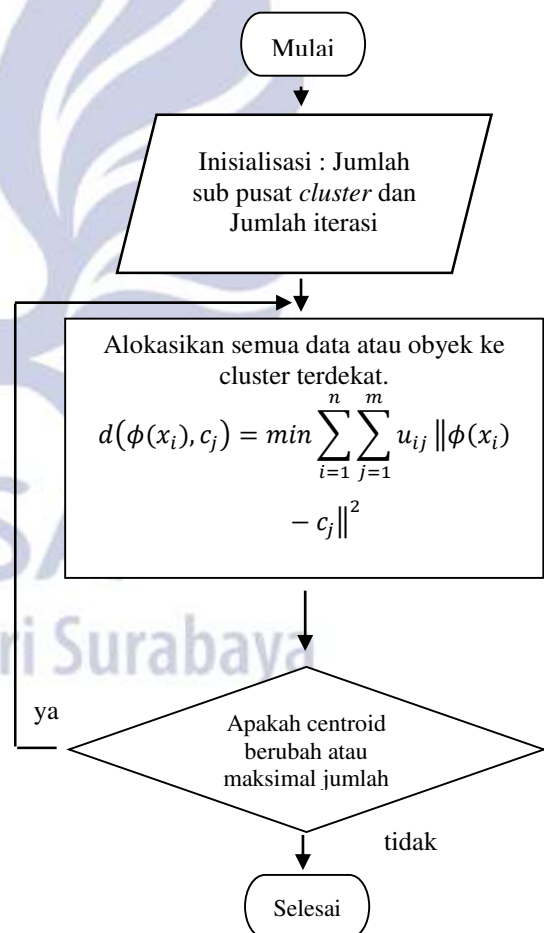
$$d(\phi(x_i), c_j) =$$

$$\min \sum_{i=1}^n \sum_{j=1}^m u_{ij} \|\phi(x_i) - c_j\|^2,$$

Dimana :

$$c_j = \frac{1}{n_j} \sum_{i=1}^n u_{ij} \phi(x_i)$$

Langkah 4 Melakukan perulangan ke langkah 3 sampai pusat cluster tidak berubah.



Gambar 2. Alur Algoritma Kernel K-Means

METODE PENELITIAN

A. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah jenis penelitian eksperimen. Eksperimen adalah penelitian dengan melakukan manipulasi variable bebas pada suatu keadaan yang terkendali (variable kontrol) yang bertujuan untuk mengidentifikasi hubungan sebab akibat dari satu atau lebih variable terikat.

B. Metode Pengumpulan Data

Data yang akan digunakan didapat dari UCI Repository Machine Learning database, yaitu *diabetic retinopathy debrecen dataset*. Data *diabetic retinopathy debrecen* diperoleh dari Universitas Debrecen, Hongaria. Dataset tersebut pada **Tabel 1.** berjumlah 1151 data yang mempunyai 20 atribut dan 1 atribut yang terakhir menunjukkan kelas yang akan dibentuk dengan type data yang berbeda-beda.

Tabel 1. Fitur Data

No.	Fitur	Type Data
1.	Hasil biner pada kualitas penilaian	Biner
2.	Hasil biner pada pra-skrining	Biner
3.	Hasil deteksi MA.	Integer
4.	Hasil deteksi MA	Integer
5.	Hasil deteksi MA	Integer
6.	Hasil deteksi MA	Integer
7.	Hasil deteksi MA	Integer
8.	Hasil deteksi MA	integer
9.	Hasil normalisasi pada eksudat	Double
10.	Hasil normalisasi pada eksudat	Double
11.	Hasil normalisasi pada eksudat	Double
12.	Hasil normalisasi pada eksudat	Double
13.	Hasil normalisasi pada eksudat	Double
14.	Hasil normalisasi pada eksudat	Double
15.	Hasil normalisasi pada eksudat	Double
16.	Hasil normalisasi pada eksudat	Double
17.	Jarak <i>Euclidean</i> pada pusat	Double

	macula dan pusat cakram optic	
18.	Diameter cakram optik	Double
19.	Hasil biner pada klasifikasi berbasis AM/FM	Biner
20.	Kelas	Biner

PEMBAHASAN

A. Pra-pemrosesan Data

Identifikasi penyakit retinopati diabetika diawali dengan menentukan sub pusat *cluster* dan jumlah iterasi. Kemudian membagi *diabetic retinopathy debrecen dataset* yang berukuran 20x1151 kedalam 2 bagian yaitu data input dan target output, Sehingga diperoleh data input berukuran 19x1151 data dan kelas output berukuran 1x1151 data. Pada **Tabel 2.** dibawah merupakan distribusi data yang terdiri dari kelas 1 sebesar 540 data dan kelas 2 sebesar 611 data. Kemudian data input di normalisasi kedalam range [0,1] untuk mereduksi perhitungan komputasi yang terlalu besar. Dengan menentukan sub pusat *cluster* (nilai M) dan banyaknya iterasi (perulangan) data akan dikenali dengan algoritma *Kernel K-Means* dan diperoleh hasil akurasi.

Tabel 2. Distribusi Data

Jumlah Data		Jumlah Fitur
Kelas 1	Kelas 2	
540	611	19

B. Pembahasan

Data *diabetic retinopathy Debrecen* kemudian diimplementasikan menggunakan Algoritma *Kernel K-Means*, dengan bantuan software Matlab R2017a yang kemudian diklusterkan dalam kelas yang bersesuaian. Pada Algoritma *Kernel K-Means* metode yang digunakan adalah *Kernel Eksponensial*. Kemudian dilakukan percobaan dengan menentukan masing-masing iterasi sebesar 10, 20 dan 30. Dengan nilai sub pusat *cluster* (nilai M) sebesar 100, 150, 200, 250 dan 300 yang disajikan pada Tabel 3, Tabel 4, Dan Tabel 5 berikut ini :

Tabel 3. Hasil Implementasi Algoritma Iterasi 10

Nilai M	Akurasi	Waktu (detik)
100	70.63%	8.70
150	76.63%	12.12
200	78.19%	11.66
250	80.10%	9.96
300	82.88%	12.22
Rata-rata Akurasi	77.69%	

Tabel 4. Hasil Implementasi Algoritma Iterasi 20

Nilai M	Akurasi	Waktu (detik)
100	72.46%	12.48
150	75.33%	9.84
200	78.71%	15.96
250	80.19%	9.68
300	81.58%	14.18
Rata-rata Akurasi	77.65%	

Tabel 5. Hasil Implementasi Algoritma Iterasi 30

Nilai M	Akurasi	Waktu (detik)
100	72.63%	9.21
150	76.89%	10.48
200	78.97%	12.51
250	80.63%	10.17
300	82.02%	13.45
Rata-rata Akurasi	78.23%	

Dari Tabel 3, 4, dan 5 dengan masing-masing iterasi sebesar 10, 20, dan 30 terlihat bahwa metode *Kernel Eksponensial* memberikan hasil akurasi terendah sebesar 70.63% dengan waktu 8.70 detik pada iterasi 10 dan nilai $M=100$, Sedangkan hasil akurasi tertinggi yaitu pada iterasi 10 dan nilai $M=300$ sebesar 82.88% dengan waktu 12.22 detik. Pada kasus ini, pemilihan iterasi dan sub pusat *cluster* atau nilai M yang semakin besar belum tentu menunjukkan bahwa akurasi semakin tinggi. Hal ini juga mungkin diakibatkan oleh pola dataset *diabetic retinopathy Debrecen* dan tingkat kecenderungan akibatnya pada masing-masing atribut berbeda.

KESIMPULAN

Identifikasi penyakit retinopati diabetika menggunakan algoritma *Kernel K-Means* dilakukan dengan cara menentukan jumlah iterasi dan jumlah sub

pusat *cluster*. Hal ini dilakukan untuk mengidentifikasi ada retinopati diabetika atau tidak ada tanda-tanda retinopati diabetika. Algoritma *Kernel K-Means* memperoleh hasil akurasi tertinggi sebesar 82.88% dengan waktu 12.22 detik pada iterasi 10 dan pusat nilai $M=300$.

DAFTAR PUSTAKA

- American Academy of Ophthalmology Retina Panel. 2016. *Preferred Practice Pattern Guidelines Diabetic retinopathy*. Di akses dalam: www.aaopt.org/ppp.
- American Diabetes Association. 2016. *Standards of medical care in diabetes, Retinopati Diabetic*. America: American Academy of Ophthalmology Retina Panel.
- Faust, O., Acharya, R. U., Ng, E. Y. K., Ng, K. H., Suri, J.S. 2010. "Algorithms for the Automated Detection of Diabetic Retinopathy Using Digital Fundus Image: A Review". *J Med Syst*. DOI 10.1007/s10916-101-9454-7.
- G. Tzortzis and A. Likas. 2009. "The global kernel - means clustering algorithm," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1181–1194.
- Inderjit S Dhillon, Yuqiang Guan, Brian Kulis (2004). *Kernel K-Means: Spectral Clustering and Normalized Cuts*. 551-556.
- J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations," *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297.
- Kementerian Kesehatan. 2014. *Infodatin Situasi dan Analisis Diabetes*. Jakarta: Pusat Data dan Informasi Kementerian Kesehatan RI.
- M Girolami (2002). *Mercer kernel-based clustering in feature space*. *Neural Networks, IEEE*

- Transactions on, Vol. 13, No. 3., pp. 780-784.
- Santoso, H. B., Albertus Joko Santoso, Eduard Rusdianto. (2013). Hybrid Clustering Method for Stock Price and Commodity Price. *International Journal of Science and Advanced Technology*. Vol. 3, No. 7, July, 21- 29.
- Shi, W., Weihua Zeng. (2013). Genetic K-Means Clustering Approach for Mapping Human Vulnerability to Chemical Hazards in the Industrialized City: A Case Study of Shanghai, China. *International Journal of Environmental Research and Public Health*. Vol. 10, 2578 – 2595.
- Soewondo P., Soegondo S., Suastika K., Pranoto A., Soeatmadji D.W., TjokroprawiroA. 2010. “The DiabCare Asia 2008 study – Outcomes on control and complications of type 2 diabetic patients in Indonesia”. *Med J Indones*. Vol. 19 (4): pp 235-244.
- V, V. V., & Ravikumar, A. 2014. *Study of Data mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus*. *International Journal of Computer Application*, 95(17), 12-16.
- Venesia, Harry, Laya. 2014. *Kecenderungan Penderita Retinopati Daibetik*. *Jurnal e-CliniC (eCl)*, volume 2 nomor 2.
- Wijayakusuma, Hembing. 2004. *Bebas Diabetes Mellitus Ala Hembing*. Jakarta: Puspa Swara.
- Wild S, Roglic G, Green A, Sicree R, King H. 2004. “Global prevalence of diabetes: estimates for the year 2000 and projections for 2030”. *Diabetes Care*. Vol. 27 (5): pp 1047-1053.
- Jiang, Ling-Li., Cao, Yu-Xiang., Yin, Hua-Kui., Den, Kong-Shu. 2013. “An Improved Kernel K-Mean Cluster Method and Its Application in Fault Diagnosis of Roller Bearing”. *Scientific Research*.