

Visualisasi Algoritma sebagai Sarana Pembelajaran *K-Means Clustering*

Alethea Suryadibrata¹, Julio Christian Young²

^{1,2}Program Studi Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia
alethea@umn.ac.id
julio.christian@umn.ac.id

Diterima 30 Maret 2020

Disetujui 17 Juni 2020

Abstract—Algorithm Visualization (AV) is often used in computer science to represent how an algorithm works. Educators believe that visualization can help students to learn difficult algorithms. In this paper, we put our interest in visualizing one of Machine Learning (ML) algorithms. ML algorithms are used in various fields. Some of the algorithms are used to classify, predict, or cluster data. Unfortunately, many students find that ML algorithms are hard to learn since some of these algorithms include complicated mathematical equations. We hope this research can help computer science students to understand K-Means Clustering in an easier way.

Index Terms—Algorithm Visualization, K-Means Clustering, Machine Learning

I. PENDAHULUAN

Memberikan penjelasan yang dapat dimengerti pelajar merupakan tantangan dari seorang pengajar. Berbagai studi dilakukan untuk membantu pengajar menyelesaikan masalah tersebut salah satunya melalui proses visualisasi materi ajar. Proses visualisasi yang dimaksud adalah serangkaian kegiatan untuk merepresentasikan informasi dalam berbagai bentuk seperti animasi, teks, dan diagram statis [1]. Pada bidang Ilmu Komputer, proses visualisasi banyak digunakan untuk menjelaskan bidang ilmu yang melibatkan konsep pemrograman dan algoritma.

Pada penelitian terdahulu, Cetin dan Andrews-Larson telah mencoba memvisualisasikan tahapan-tahapan dalam suatu algoritma *sorting* [2]. Berdasarkan proses visualisasi, peneliti menyimpulkan bahwa materi algoritma yang tadinya bersifat *cryptic* menjadi lebih mudah dimengerti oleh para pelajar. Kemudian, pada penelitian lainnya, Nathasya dkk. telah mencoba mengintegrasikan proses visualisasi untuk proses pembelajaran materi struktur data [3]. Melalui penelitian yang dilakukan, proses visualisasi bahkan terbukti meningkatkan motivasi, performa dan capaian pembelajaran para pelajar (para pelajar mendapatkan nilai ujian yang lebih baik dan lebih bersemangat dalam mengerjakan tugas). Selain digunakan untuk materi-materi pembelajaran yang sederhana, visualisasi juga dapat digunakan untuk membantu proses pembelajaran algoritma yang bersifat lebih kompleks. Hal ini dibuktikan melalui

penelitian yang dilakukan oleh Sriadhi dkk., yang memvisualisasikan proses dari salah satu algoritma pada bidang kriptografi, RC4 [4].

Selain bidang kriptografi, pembelajaran mesin juga merupakan salah satu bidang yang bersifat kompleks dan dirasa sulit oleh para pelajar. Hal ini dikarenakan algoritma pembelajaran mesin banyak melibatkan rumus-rumus matematika yang bersifat abstrak dan *cryptic*. Padahal, algoritma pembelajaran mesin cukup banyak digunakan untuk menyelesaikan berbagai permasalahan di dunia nyata seperti memprediksi pergerakan data runtun waktu [5], pengenalan karakter dalam dokumen cetak [6], ataupun pada berbagai bidang lainnya.

Berdasarkan alasan di atas, penelitian ini mencoba untuk memvisualisasikan salah satu algoritma pembelajaran mesin, yaitu algoritma *K-Means Clustering*. Algoritma ini merupakan salah satu algoritma pembelajaran mesin yang sering digunakan untuk mengelompokkan sekumpulan data berdasarkan informasi yang dimilikinya.

II. TELAAH LITERATUR

A. Visualisasi Algoritma

Visualisasi algoritma merupakan salah satu alternatif untuk mempelajari suatu algoritma dengan menggunakan sekumpulan elemen grafis yang berisikan informasi-informasi penting dalam suatu algoritma. Elemen grafis digunakan untuk visualisasi seperti titik, garis, dan sebagainya [7].

Terdapat 2 variasi pada visualisasi algoritma yaitu:

1. Visualisasi algoritma statis, di mana penjelasan algoritma melibatkan elemen-elemen grafis yang tidak bergerak.
2. Visualisasi algoritma dinamis (sering juga disebut animasi algoritma), di mana proses visualisasi melibatkan elemen-elemen grafis yang saling berinteraksi. Teknik visualisasi algoritma dinamis cenderung lebih canggih dan sulit untuk diimplementasikan.

B. Pembelajaran Mesin

Pemelajaran mesin (*machine learning*) adalah pemrograman komputer untuk mengoptimalkan performa menggunakan sampel data atau pengetahuan sebelumnya [8]. Pemelajaran mesin merupakan bagian dari kecerdasan buatan (*Artificial Intelligence / AI*). Untuk menjadi cerdas, sebuah sistem di dalam lingkungan yang berubah-ubah harus memiliki kemampuan untuk belajar. Jika sistem dapat belajar untuk beradaptasi dalam perubahan yang ada, maka desainer sistem tidak perlu meramalkan serta menyediakan solusi untuk semua situasi yang dapat terjadi.

Pemelajaran mesin juga membantu kita untuk mencari solusi dari masalah-masalah dari data visual, pengenalan suara, dan robotik [8]. Pemelajaran mesin dapat menghasilkan prediksi untuk masa depan, atau deskriptif untuk mendapatkan pengetahuan dari data, atau bahkan keduanya. Pemelajaran mesin menggunakan teori statistik dalam membangun suatu model matematika. Peran ilmu komputer pada pemelajaran mesin dapat dibagi menjadi dua. Pertama, pada pelatihan, kita membutuhkan algoritma yang efisien untuk menyelesaikan masalah optimasi serta menyimpan dan memroses data yang kita miliki. Kedua, setelah model dipelajari, representasi serta solusi yang diberikan juga harus efisien.

C. K-Means Clustering

Pada bidang pemelajaran mesin, istilah *clustering* merujuk pada sebuah tugas untuk memisahkan/mengelompokkan sekumpulan data ke dalam sejumlah cluster berdasarkan hasil pengukuran suatu metrik yang menyatakan kesamaan/ jarak dari data [9]. *K-Means* merupakan salah satu algoritma untuk proses *clustering* yang bersifat sederhana secara matematis dan efektif, sehingga dalam lima dekade terakhir, algoritma ini masih sangat sering digunakan untuk menyelesaikan berbagai macam permasalahan.

Langkah-langkah algoritma *K-Means Clustering*:

1. Tentukan suatu nilai k yang menyatakan jumlah kelompok (*cluster*).
2. Inisialisasi posisi sejumlah k vektor secara acak, di mana setiap vektor merepresentasikan pusat (*center*) dari sebuah kelompok.
3. Berdasarkan setiap vektor pusat yang ada, hitung jarak dari setiap vektor pusat ke setiap data dalam representasi vektor. Berdasarkan perhitungan jarak yang dilakukan, sebuah titik akan dijadikan anggota dari sebuah kelompok berdasarkan jarak titik pusat kelompok terdekat dengannya.
4. Hitung kembali vektor pusat dari setiap kelompok dengan mencari rata-rata dari semua vektor data dalam kelompok.

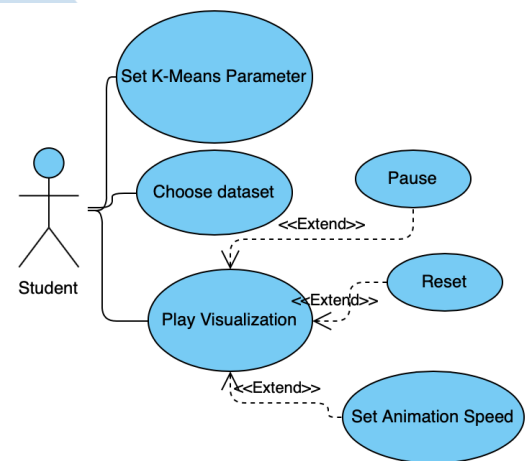
5. Jika nilai vektor pusat tidak mengalami perubahan, maka setiap titik pusat telah merepresentasikan kelompok dengan benar dan proses pengelompokan selesai. Jika nilai vektor pusat masih berubah, kembali ke tahapan nomor 3.

III. METODE PENELITIAN

Bagian ini menjelaskan bagaimana perancangan pembuatan aplikasi dilakukan serta fitur-fitur yang dianggap penting untuk meningkatkan interaktifitas aplikasi dengan pengguna. Fokus pada penelitian ini adalah untuk membuat visualisasi yang dapat digunakan oleh pelajar yang ingin memahami cara kerja algoritma *K-Means Clustering*.

A. Use Case Diagram

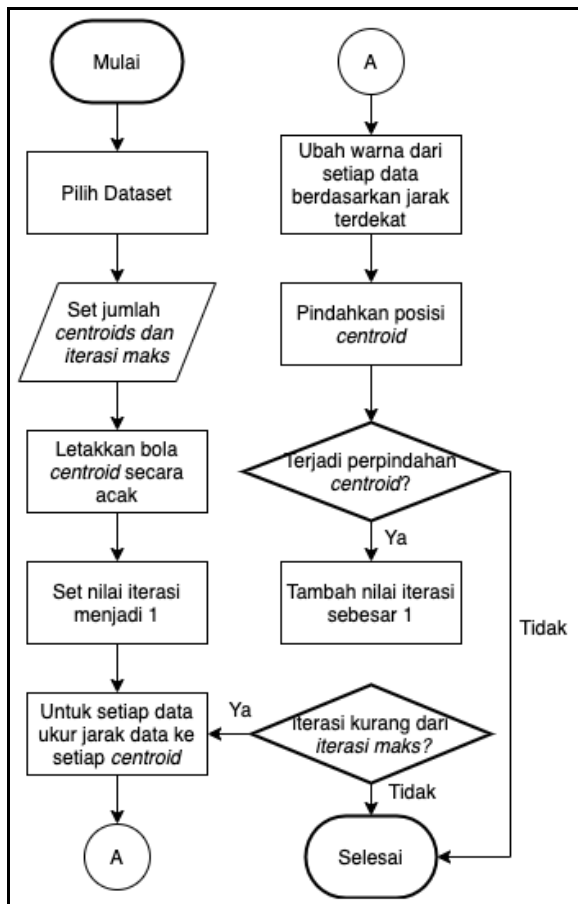
Pada perancangan aplikasi, dibuatlah sebuah *Use Case Diagram* untuk merepresentasikan interaksi *user* dengan aplikasi. Berdasarkan Gambar 1, pengguna dapat memasukan jumlah cluster dan jumlah iterasi ke dalam aplikasi sebelum visualisasi dimulai.



Gambar 1. *Use Case Diagram* visualisasi algoritma *K-Means Clustering*

B. Flowchart Algoritma K-Means Clustering

Pada penelitian ini, proses visualisasi dilakukan dengan merepresentasikan data sebagai sekumpulan bola-bola yang tersebar dalam sebuah ruang 3D. Di dalam aplikasi, setiap kelompok data akan direpresentasikan sebagai sebuah warna. Untuk setiap bola yang merepresentasikan sebuah data, warna dari bola akan berubah-ubah menyesuaikan kelompok di mana ia berada. Gambar 2 menunjukkan *flowchart* algoritma *K-Means Clustering* dan bagaimana proses visualisasi dilakukan dengan menggunakan algoritma tersebut.



Gambar 2. Flowchart visualisasi algoritma *K-Means Clustering*

C. Fitur-fitur Interaktif Aplikasi

Dalam sebuah proses visualisasi, fitur-fitur interaktif digunakan agar aplikasi menjadi lebih menarik sehingga dapat meningkatkan motivasi belajar. Selain itu, fitur-fitur interaktif juga dapat digunakan untuk mengikuti kecepatan belajar masing-masing pelajar. Dalam aplikasi yang dikembangkan, salah satu fitur interaktif yang diimplementasikan adalah animasi dari proses perhitungan jarak antara setiap vektor data dan setiap vektor pusat yang ada menggunakan sebuah garis pembantu. Dalam proses visualisasi, garis pembantu berperan untuk menunjukkan proses pencarian jarak terdekat dari setiap vektor data.

Selain itu, fitur interaktif kedua yang terdapat dalam aplikasi adalah animasi perubahan warna dari setiap vektor data berdasarkan warna dari titik pusat *cluster* yang paling dekat dengannya. Proses ini digunakan untuk menunjukkan kepada para peserta ajar bagaimana titik pusat dari setiap *cluster* dapat terbentuk dan berubah-ubah pada setiap iterasi dalam algoritma. Kemudian, untuk mempermudah pengguna dalam proses belajar, aplikasi juga menyediakan fitur untuk mengatur kecepatan animasi dan fitur jeda (*pause*) yang dapat digunakan penggunaannya sesuai dengan kebutuhan.

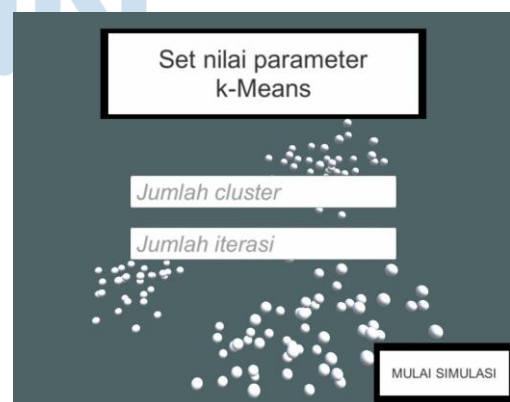
IV. IMPLEMENTASI

Aplikasi yang digunakan untuk membuat visualisasi algoritma *K-Means Clustering* dibuat menggunakan sebuah *game engine* bernama *Unity* dan merupakan aplikasi 3D. Sebelum menjalankan visualisasi, pengguna tentunya harus menentukan dataset yang ingin digunakan; dan beberapa parameter seperti nilai k , *tolerance*, jumlah iterasi maksimum dari algoritma. Gambar 3 dan Gambar 4 menunjukkan proses pemilihan dataset dan pengaturan parameter dalam aplikasi.

Melalui Gambar 3, dapat terlihat bahwa pengguna dapat memilih dataset yang ingin digunakan menggunakan tanda panah yang berada di kiri dan kanan layar dan memulai simulasi dengan menekan tombol "Mulai Simulasi". Kemudian, setelah memilih suatu dataset, sebelum simulasi dimulai, pengguna akan diminta untuk mengisi nilai parameter *K-Means* yang ingin digunakan seperti pada Gambar 4.

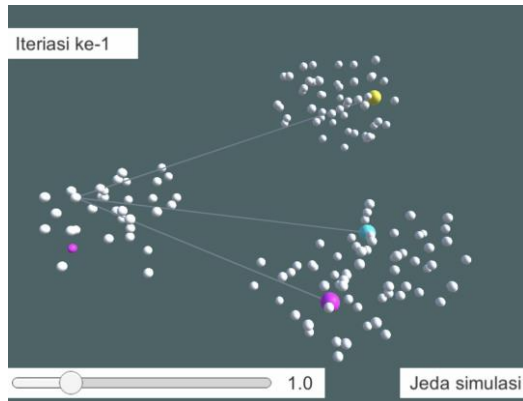


Gambar 3. Proses pemilihan dataset dalam aplikasi



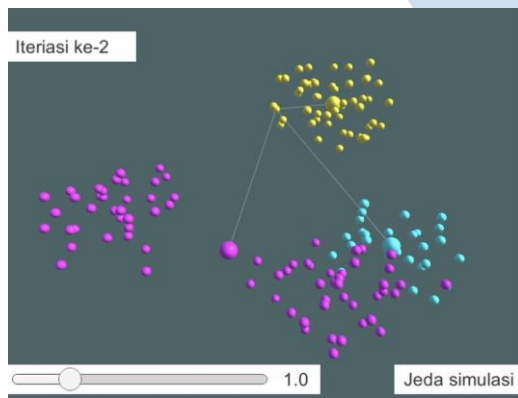
Gambar 4. Proses parameterisasi algoritma *K-Means*

Saat visualisasi dijalankan, mula-mula *centroid* akan diletakkan secara acak dan direpresentasikan sebagai suatu warna. Kemudian, animasi yang menunjukkan perbandingan jarak untuk setiap data ke setiap *centroid* dilakukan (proses ini ditunjukkan pada Gambar 5).



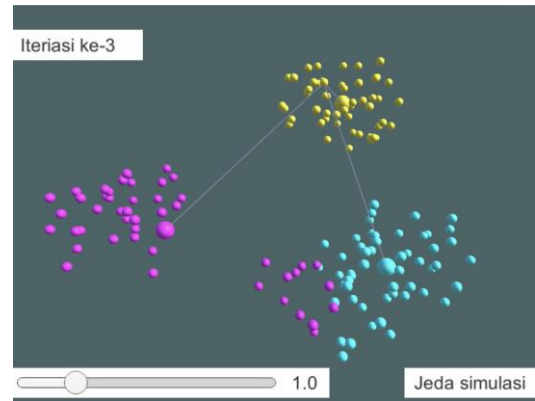
Gambar 5. Proses perbandingan setiap *centroid*

Pada Gambar 5, tiga garis yang mengarah dari setiap bola yang berwarna ke sebuah bola putih yang berada di bagian kiri layar menunjukkan proses perbandingan jarak dari setiap *centroid* ke sebuah data. Berdasarkan proses ini warna dari bola putih akan berubah berdasarkan jarak (garis) terpendek yang terhubung ke dirinya. Animasi dari proses ini akan dilakukan untuk setiap data dan kemudian dilanjutkan dengan proses update lokasi *centroid* sebelum algoritma berjalan ke iterasi selanjutnya. Gambar 6 menunjukkan lokasi *centroid* yang telah di-update berdasarkan iterasi pertama pada iterasi ke-2.

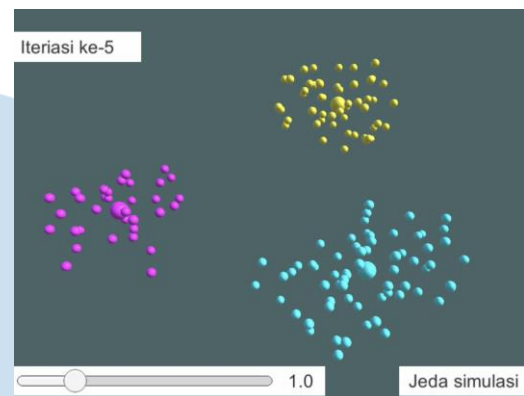


Gambar 6. Proses perpindahan lokasi *centroid* dan perhitungan jarak kembali pada iterasi ke-2

Proses simulasi akan terus dijalankan mengikuti alur dari algoritma *K-Means Clustering* yang telah dijelaskan pada subbab sebelumnya. Gambar 7 dan Gambar 8 menunjukkan proses lanjutan yang terjadi di dalam aplikasi sampai simulasi berakhir.



Gambar 7. Perpindahan *centroid* berdasarkan perhitungan pada iterasi ke-2



Gambar 8. *Centroid* telah mencapai tidak lagi mengalami perpindahan

V. EVALUASI

Berdasarkan aplikasi visualisasi algoritma *K-Means Clustering* yang telah dikembangkan dilakukan *Focus Group Discussion* (FGD) untuk menggali ketepatan, kelebihan serta kekurangan dari aplikasi visualisasi yang telah dikembangkan. FGD dilakukan bersama 3 orang pengajar di salah satu perguruan tinggi swasta yang mengampu mata kuliah pembelajaran mesin.

Berdasarkan FGD yang dilakukan para pengajar merasa aplikasi dapat membantu mempermudah proses pengajaran ataupun proses pembelajaran algoritma secara mandiri. Melalui elemen-elemen visual di dalam aplikasi konsep-konsep abstrak seperti pencarian jarak terdekat, perpindahan *centroid*, perhitungan ulang vektor *centroid* yang bersifat *cryptic* jika disajikan sebagai teks dirasa akan mempermudah proses pemahaman para peserta ajar. Selain itu, berdasarkan diskusi yang dilakukan, dengan menyediakan beberapa variasi dataset dan mengizinkan pengguna melakukan parameterisasi dari algoritma, aplikasi dirasa dapat membantu para pengajar dalam menjelaskan kelebihan dan kekurangan dari algoritma (seperti bagaimana

algoritma *K-Means Clustering* dalam menghadapi kondisi *density-connected region*).

Akan tetapi, terlepas dari kelebihan di atas, aplikasi dirasa masih memiliki beberapa kelemahan berkaitan dengan hal-hal definitif yang berkaitan dengan algoritma *K-Means* seperti bagaimana proses pengukuran jarak dilakukan di dalam aplikasi dan proses perpindahan *centroid*.

VI. SIMPULAN

Visualisasi algoritma *K-Means Clustering* telah dibuat dalam bentuk 3D menggunakan *Unity* dan bahasa pemrograman *C#*. Visualisasi ini diharapkan dapat membantu pelajar dalam memahami algoritma *K-Means Clustering*. Evaluasi telah dilakukan dalam bentuk FGD terhadap beberapa pengajar mata kuliah Pemelajaran Mesin. Untuk penelitian selanjutnya, dapat dilakukan evaluasi terhadap para pelajar berupa kuesioner yang mengacu pada model evaluasi tertentu. Aplikasi ini juga dapat dikembangkan lebih lanjut dengan menambahkan algoritma pemelajaran mesin lainnya.

DAFTAR PUSTAKA

- [1] C. D. Hundhausen, S.A. Douglas, dan J.T. Stasko, "A Meta-Study of Algorithm Visualization Effectiveness," *Journal of Visual Languages & Computing*, vol. 13(3), Juni 2002, hal. 259-290.
- [2] I. Cetin dan C. Andrews-Larson, "Learning Sorting Algorithms through Visualization Construction", *Journal Computer Science Education*, vol. 26, 2016, hal 27-43.
- [3] R.A. Nathasya, O. Karnalim, dan M. Ayub, "Integrating Program and Algorithm Visualization for Learning Data Structure Implementation", *Egyptian Informatics Journal*, Vol. 20(3), November 2019, hal. 193-204.
- [4] S. Sriadhi, R. Rahim, dan A.S. Ahmar, "RC4 Algorithm Visualization for Cryptography Education", *Journal of Physics*, [2nd International Conference on Statistics, Mathematics, Teaching, and Research, 2017].
- [5] K. Johan, J.C. Young, dan S. Hansun, "LSTM-RNN Automotive Stock Price Prediction", *International Journal of Scientific & Technology Research*, vol. 9 (8), 2019, hal. 173-176.
- [6] D.P. Chandra dan A. Suryadibrata, "Implementasi Jaringan Saraf Tiruan Backpropagation untuk Pengenalan Karakter pada Dokumen Tercetak", *Ultima Computing: Jurnal Sistem Komputer*, vol. 11, 2019, hal 46-51.
- [7] A. Levitin, "Introduction to the Design & Analysis of Algorithms", Pearson-Addison Wesley, 2007.
- [8] E. Alpaydin. "Introduction to Machine Learning" London : MIT Press, 2009.
- [9] J. Qi dkk., "An Effective and Efficient Hierarchical K-Means Clustering Algorithm", *International Journal of Distributed Sensor Networks*. Vol.13(8), 2017.



UMN