# IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK PEMBAGIAN KELAS SISWA

**SKRIPSI**

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Program Studi Teknik Informatika

OLEH :

**NISIA YUANITA**
NPM : 12.1.03.02.0017

FAKULTAS TEKNIK (FT)

UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA

**UN PGRI KEDIRI**

2016

Skripsi Oleh :

**NISIA YUANITA**
NPM : 12.1.03.02.0017

Judul :

**IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK PEMBAGIAN KELAS SISWA**

Telah disetujui untuk diajukan Kepada

Panitia Ujian/Sidang Skripsi Program Studi Teknik Informatika

FT UN PGRI Kediri

Tanggal : 3 Agustus 2016

| Pembimbing I | Pembimbing II |
|---|---|
| Ahmad Bagus Setiawan, S.T., M.Kom., M.M | Risky Aswi Ramadhani, M.Kom |
| NIDN.0703018704 | NIDN.0708049001 |

ii

Skripsi Oleh :

**NISIA YUANITA**

NPM : 12.1.03.02.0017

Judul :

**IMPLEMENTASI *K-MEANS CLUSTERING* UNTUK PEMBAGIAN KELAS SISWA**

Telah dipertahankan di depan panitia Ujian/Sidang Skripsi

Jurusan Teknik Informatika FT UN PGRI Kediri

Pada Tanggal : 12 Agustus 2016

**Dan Dinyatakan telah Memenuhi Persyaratan**

Panitia Penguji :

1. Ketua Penguji : Ahmad Bagus Setiawan, S.T., M.Kom., M.M
2. Penguji I : Ir. Juli Sulaksono M.M., M.Kom
3. Penguji II : Risky Aswi Ramadhani, M.Kom

Mengetahui
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.pd
NIP.19640202 199103 1 002

iii

# IMPLEMENTASI K-MEANS CLUSTERING UNTUK PEMBAGIAN KELAS SISWA

Nisia Yuanita
12.1.03.02.0017
Teknik – Teknik Informatika
nesya.yuan0129
Ahmad Bagus Setiawan, S.T., M.Kom., M.M dan Risky Aswi Ramadhani, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

## ABSTRAK

Dalam membagi kelas siswa sering kali dijumpai sistem yang digunakan masih manual sehingga pembagian kelas menjadi tidak merata. Hal tersebut memungkinkan terjadi persaingan yang tidak sehat baik antar siswa dan antar kelas sehingga dinilai kurang efektif dan efisien.

Permasalahan penelitian ini adalah (1) Bagaimana cara membangun aplikasi guna membantu pengambilan keputusan untuk pengelompokan dan pembagian kelas siswa? (2) Bagaimana mengimplementasikan *K-Means* untuk pengelompokan dan pembagian kelas siswa secara merata ?

Penelitian ini menggunakan nilai UTS siswa kelas VII semester gasal MTs Sunan Ampel yang terdiri dari nilai bidang studi matematika, bahasa indonesia, bahasa inggris, dan ilmu pengetahuan alam.

Kesimpulan penelitian ini adalah (1) Dihasilkan sebuah aplikasi untuk pengelompokan kelas siswa menggunakan metode *k-means*. (2) Dari sampel 150 siswa, 51 siswa masuk di kelas A, 50 siswa masuk di kelas B, dan siswa dengan jumlah 49 masuk di kelas C.

Berdasarkan simpulan hasil penelitian ini direkomendasikan : (1) Jumlah sampel dan penambahan variabel lebih di perbanyak. (2) Menggunakan perangkat lunak lain dan algoritma lain seperti *Fuzzy Clustering* dan *K-Means*++.

**Kata Kunci :** *Data Mining*, *K-Means Clustering*, Pembagian Kelas.

## I. LATAR BELAKANG

Ilmu pengetahuan dan teknologi mempunyai peranan yang sangat penting dalam pertumbuhan era globalisasi. Hal tersebut mendorong upaya-upaya pemanfaatan teknologi dalam berbagai bidang kehidupan, salah satunya adalah membangun suatu sistem yang dapat dimanfaatkan untuk pembagian kelas siswa.

Pembagian kelas di MTs Sunan Ampel saat ini masih menggunakan sistem manual. Agar pembagian kelas dalam tiap-tiap tingkat lebih merata, maka dibutuhkan sebuah alat bantu untuk memudahkan pekerjaan dalam pembagian kelas siswa. Data yang dibutuhkan untuk Pembagian kelas siswa berupa nilai akademik yang meliputi bidang studi Matematika, Bahasa Indonesia, Bahasa Inggris, dan Ilmu Pengetahuan Alam. Nilai-nilai tersebut diambilkan dari nilai UTS tingkat VII pada semester gasal.

Tujuan membangun sebuah sistem untuk pembagian kelas agar dapat dijadikan sebagai alat bantu untuk mempermudah admin dalam pembagian kelas siswa. Agar sistem yang akan dibangun berhasil diterapkan dengan baik, maka digunakan *K-Means* untuk pengelompokkan secara merata berdasarkan kriteria nilai yang sudah ditentukan.

## II. METODE

### A. *Data Mining*

*Data mining* diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan (Tan, 2006).

### B. *K-Means*

*K-Means* merupakan salah satu metode data *clustering* non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. (Agusta, 2007).

Pengelompokan data dengan metode *K-Means* secara umum dilakukan dengan algoritma sebagai berikut :

1. Pilih jumlah klaster *k.*
2. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.

3. Tempatkan setiap data/ obyek ke cluster terdekat. Kedekatan dua obyek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling dekat antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana.
4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/objek dalam cluster tertentu. Jika dikehendaki bisa juga memakai median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek dengan memakai pusat cluster yang baru. Jika pusat cluster sudah tidak berubah lagi, maka proses peng*cluster*an selesai. Atau, kembali lagi ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.(Santosa, 2007).

**CONTOH KASUS :**

pengelompokan kelas siswa menggunakan *K-Means Clustering*, maka dibutuhkan beberapa data siswa kelas VII semester gasal dengan parameter-parameter sebagai berikut :

Jumlah Cluster : 3

Jumlah data : 12

Jumlah atribut : 4 meliputi bidang studi Matematika, Bahasa Indonesia, Bahasa Inggris, dan IPA

Tabel 1. nilai siswa

| Nama Siswa | Mat | Bind | Bing | Ipa |
|---|---|---|---|---|
| Abil Oktaviano | 72 | 84 | 83 | 86 |
| Afif Nur Rifai | 80 | 72 | 82 | 80 |
| Alfredo Sundy | 75 | 80 | 80 | 84 |
| Ahmat Nur | 82 | 82 | 73 | 80 |
| Akhmad Alan | 80 | 70 | 82 | 83 |
| Aminiya Ulfa | 82 | 80 | 87 | 88 |
| Anwar Azis | 78 | 81 | 83 | 82 |
| Apriliani Fira | 80 | 85 | 80 | 82 |
| Detik Kusuma | 82 | 80 | 85 | 85 |
| Devon Fahreza | 72 | 82 | 83 | 85 |
| Difa Galih | 80 | 80 | 80 | 82 |
| Difan Ade | 84 | 82 | 85 | 86 |

Iterasi ke-1

1. Penentuan pusat awal *cluster*

   Penentuan awal pusat *cluster* menggunakan nilai dari beberapa data siswa secara *random* / acak

Tabel 2. Pusat *Cluster*

| | | | | |
|---|---|---|---|---|
| C1 | 80 | 80 | 81 | 81 |
| C2 | 82 | 81 | 80 | 82 |
| C3 | 76 | 81 | 80 | 87 |

2. Perhitungan jarak pusat cluster Untuk pengukuran jarak antara data dengan pusat cluster digunakan Euclidean Distance dengan rumus :

$$d_{ik} = \sqrt{\sum_{j}^{m} (Cij - Ckj)^2}$$

$Cij$ : Pusat *Cluster*

$Ckj$ : Data

Berikut adalah perhitungan menggunakan *euclidean distance* dari data ke-1 terhadap pusat *cluster* :

C1 =

$$\sqrt{\begin{array}{l}(72-80)^2+(84-80)^2\\ +(83-81)^2\\ +(86-81)^2\end{array}}$$

$$= \sqrt{8^2+4^2+2^2+5^2}$$

$$= \sqrt{64+16+4+25}$$

$$= \sqrt{109}$$

yang kemudian akan didapatkan matriks jarak sebagai berikut :

D1 =

Tabel 3. Hasil Perhitungan Jarak

| Nama Siswa | C1 | C2 | C3 |
|---|---|---|---|
| Abil Oktaviano | 10,44 | 11,57 | 5,91 |
| Afif Nur Rifai | 8,12 | 9,64 | 12,24 |
| Alfredo Sundy | 6,16 | 7,34 | 3,31 |
| Ahmat Nur | 8,54 | 7,34 | 11,61 |
| Akhmad Alan | 81,63 | 11,40 | 12,52 |
| Aminiya Ulfa | 81,24 | 9,27 | 9,32 |
| Anwar Azis | 3,16 | 5 | 6,16 |
| Apriliani Fira | 4,24 | 4,47 | 7,549 |
| Detik Kusuma | 6 | 5,91 | 8,12 |
| Devon Fahreza | 9,38 | 10,90 | 5,47 |
| Difa Galih | 1,41 | 2,236 | 6,48 |
| Difan Ade | 7,81 | 6,78 | 9,53 |

Setiap kolom pada mariks menunjukkan nilai jarak data terhadap pusat *cluster*. Baris pertama pada matriks menunjukan nilai jarak data terhadap titik pusat *cluster* pertama, baris kedua pada matriks menunjukkan nilai jarak data terhadap titik pusat *cluster* kedua dan seterusnya.

3. Pengelompokan data

Jarak hasil dari perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Berikut ini akan ditampilkan data matriks pengelompokan *group*.

GI =

Tabel 4. Jarak Terdekat Dengan Pusat *Cluster*

| C1 | | | | C2 | | | | C3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | 72 | 84 | 83 | 86 |
| 80 | 72 | 82 | 80 | | | | | | | | |
| | | | | | | | | 75 | 80 | 80 | 84 |
| | | | | 82 | 82 | 73 | 80 | | | | |
| | | | | 80 | 70 | 82 | 83 | | | | |
| | | | | 82 | 80 | 87 | 88 | | | | |
| 78 | 81 | 83 | 82 | | | | | | | | |
| 80 | 85 | 80 | 82 | | | | | | | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 82 | 80 | 85 | 85 | | | | |
| | | | | | | | | 72 | 82 | 83 | 85 |
| 80 | 80 | 80 | 82 | | | | | | | | |
| | | | | 84 | 82 | 85 | 86 | | | | |

Keterangan :

a. Jika nilai D1 lebih dekat jaraknya dengan nilai hasil perhitungan data *cluster* ke-1 (C1) atau data *cluster* ke-2 (C2) atau data *cluster* ke-3 (C3) maka GI bernilai 0 dan tidak termasuk *group* / kelompok *cluster* baru. Jika nilai D1 lebih jauh jaraknya dari nilai hasil perhitungan data *cluster* ke-1 (C1) atau data *cluster* ke-2 (C2) atau data *cluster* ke-3 (C3) dan bernilai selain 0 maka termasuk kelompok *cluster* baru.

b. Jika DI lebih jauh jaraknya dari nilai hasil perhitungan data *cluster* ke-1 (C1) atau data *cluster* ke-2 (C2) atau data *cluster* ke-3 (C3) maka GI bernilai selain 0 dan termasuk *group* / kelompok *cluster* baru.

4. Penentuan pusat *cluster* baru

Setelah diketahui anggota tiap-tiap *cluster*, kemudian pusat *cluster* baru dihitung berdasarkan data anggota tiap-tiap *cluster* sesuai dengan rumus pusat anggota *cluster* sehingga didapatkan perhitungan:

Tabel 5. Pusat *Cluster* Baru

| | | | | |
|---|---|---|---|---|
| C1 | 79,5 | 80 | 81,25 | 81,5 |
| C2 | 82 | 79 | 82,4 | 84,4 |
| C3 | 73 | 82 | 82 | 85 |

Iterasi ke-2

Ulangi langkah pertama pada iterasi ke-2 hingga iterasi ke... N hingga posisi data tidak mengalami perubahan,, sehingga diperoleh hasil *Clustering* sebagai berikut :

Tabel 6. Hasil *Clustering*

| Nama Siswa | Cluster |
|---|---|
| Abil Oktaviano | C3 |
| Afif Nur Rifai | C1 |
| Alfredo Sundy | C1 |
| Ahmat Nur | C1 |
| Akhmad Alan | C2 |
| Aminiya Ulfa | C2 |
| Anwar Azis | C1 |
| Apriliani Fira | C1 |
| Detik Kusuma | C2 |
| Devon Fahreza | C3 |
| Difa Galih | C1 |
| Difan Ade | C2 |

## I. HASIL DAN KESIMPULAN

Berdasarkan penelitian yang dilakukan di MTs Sunan Ampel, maka dihasilkan sebuah aplikasi untuk pengelompokan kelas siswa menggunakan teknik *data mining* dengan metode *k-means*. Variabel yang digunakan adalah nilai UTS

pada semester gasal meliputi bidang studi Matematika, Bahasa Indonesia, Bahasa Inggris, dan Ilmu Pengetahuan Alam. Data siswa yang digunakan untuk pengelompokan kelas yaitu data siswa tingkat VII yang berjumlah 150 data. Siswa dengan jumlah 51 menempati kelas A, siswa dengan jumlah 50 menempati kelas B, dan siswa dengan jumlah 49 menempati kelas C.

## II. DAFTAR PUSTAKA

Agusta, Y. 2007. *K-Means – Penerapan, Permasalahan, dan Metode Terkait*.

Arikunto, S. 2013. *Prosedur Penelitian-Suatu Pendekatan Praktik*. Jakarta : Rineka Cipta.

Eka ,Fenty, et. al. 2015. *Implementasi algoritma K-Means untuk menentukan kelompok pengayaan materi mata pelajaran ujian nasional*. Universitas Islam Negeri Syarif Hidayatullah.

Febriani, Lusia. et, al. 2012. *Pengelompokan mahasiwa sistem informasi berdasarkan tingkat kompetensi akademik dengan Fuzzy K-Means*.

Jogiyanto. 2010. *Analisis dan Desain Sistem Informasi*. Yogyakarta : Andi.

Mardiani. 2014. *Perbandingan algoritma K-Means dan EM untuk clusterisasi nilai mahasiswa berdasarkan asal sekolah*. STMIK GI MDP

Muzakir, Ari. 2014. *Analisa dan pemanfaatan algoritma K-Means Clustering pada data nilai siswa sebagai penentu penerima beasiswa*. Universitas Bina Darma Palembang.

Nugroho, B. 2005. *Database Relational dengan MySQL*. Yogyakarta : Andi

Parno, 2012. *Data Flow Diagram.*

Prasetyo, E. 2012. *Data Mining-Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta : Andi.

Santosa, B. 2007. *Data Mining-Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu

Suprihatin. 2011. *Klastering K-Means untuk penentuan nilai ujian*. Yogyakarta : Universitas Ahmad Dahlan.