

LAPORAN SKRIPSI

ALGORITMA K-MEANS UNTUK PENGELOMPOKAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA

Oleh :

MUHAMMAD SHOFIYUDDIN

2011-51-182



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MURIA KUDUS

2016



UNIVERSITAS MURIA KUDUS

PENGESAHAN STATUS SKRIPSI

JUDUL : ALGORITMA K-MEANS UNTUK PENGELOMPOKAN
KABUPATEN/KOTA DI INDONESIA BERDASARKAN
INDIKATOR INDEKS PEMBANGUNAN MANUSIA

SAYA : MUHAMMAD SHOFIYUDDIN

Mengijinkan Skripsi Teknik Informatika ini disimpan di Perpustakaan Program Studi Teknik Informatika Universitas Muria Kudus dengan syarat-syarat kegunaan sebagai berikut:

1. Skripsi adalah hal milik Program Studi Teknik Informatika UMK Kudus
2. Perpustakaan Teknik Informatika UMK dibenarkan membuat salinan untuk tujuan referensi saja
3. Perpustakaan juga dibenarkan membuat salinan Skripsi ini sebagai bahan pertukaran antar institusi pendidikan tinggi
4. Berikan tanda \ sesuai dengan kategori Skripsi

Sangat Rahasia (Mengandung isi tentang keselamatan/kepentingan Negara Republik Indonesia)

Rahasia (Mengandung isi tentang kerahasiaan dari suatu organisasi/badan tempat penelitian Skripsi ini dikerjakan)

Biasa

Disahkan Oleh :

Muhammad Shofiyuddin

Alamat : Besito 03/06 Gebog Kudus
Tanggal : 21 Juni 2016

Endang Supriyati, M.Kom

NIDN.0629077402
Tanggal : 21 Juni 2016



UNIVERSITAS MURIA KUDUS
PERNYATAAN PENULIS

JUDUL : ALGORITMA K-MEANS UNTUK PENGELOMPOKAN
KABUPATEN/KOTA DI INDONESIA BERDASARKAN
INDIKATOR INDEKS PEMBANGUNAN MANUSIA

NAMA : MUHAMMAD SHOFIYUDDIN

NIM : 2011-51-182

“Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.

Kudus, 19 Mei 2016

Muhammad Shofiyuddin
Penulis



UNIVERSITAS MURIA KUDUS
PERSETUJUAN SKRIPSI

JUDUL : ALGORITMA K-MEANS UNTUK PENGELOMPOKAN
KABUPATEN/KOTA DI INDONESIA BERDASARKAN
INDIKATOR INDEKS PEMBANGUNAN MANUSIA

NAMA : MUHAMMAD SHOFIYUDDIN

NIM : 2011-51-182

Skripsi ini telah diperiksa dan disetujui.

Kudus, 19 Mei 2016

Pembimbing Utama

Pembimbing Pembantu

Endang Supriyati, M.Kom
NIDN.0629077402

Tutik Khotimah, M.Kom
NIDN. 0608068502

Komite Skripsi

Muhammad Imam Ghozali, M.Kom
NIDN.



UNIVERSITAS MURIA KUDUS

PENGESAHAN SKRIPSI

JUDUL : ALGORITMA K-MEANS UNTUK PENGELOMPOKAN
KABUPATEN/KOTA DI INDONESIA BERDASARKAN
INDIKATOR INDEKS PEMBANGUNAN MANUSIA

NAMA : MUHAMMAD SHOFIYUDDIN

NIM : 2011-51-182

Skripsi ini telah diujikan dan dipertahankan di hadapan Dewan Pengaji pada Sidang Skripsi tanggal **09 Juni 2016**. Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar **Sarjana Komputer (S.Kom)**.

Kudus, 21 Juni 2015

Ketua Pengaji

Pengaji 1

Arief Susanto, S.T, M.Kom
NIDN. 0603047104

Tri Listyorini, M.Kom
NIDN. 06160888502

Mengetahui
Dekan Fakultas Teknik

Ka Prodi Teknik Informatika

Muhammad Dahlan, S.T, M.T
NIDN. 0601076901

Ahmad Jazuli, M.Kom
NIDN. 0406107004

ABSTRACT

Human development is one of government obligations for increasing social welfare. In human development, the government need to make priority of any undeveloped area so the well-being can wide-spread well. The problem in Indonesia is that the human development has not spread evenly between one and another. Human Development Index (HDI) is measurement indicator to determine wheater the county/city/even country has good human development or not. HDI consists of 4 indicators, life expectancy index, literacy index, average years of schooling and puchasing power parity . The goal of this research is to group the 497 cities /counties (2012) in Indonesia with K-Means algorithm. The used feature extraction is Principal Component Analysis. PCA generates two principal component with total of variance 98.68%. The clustering result with Davies Bouldin Index value 0.6164 and membership pattern of each object 317, 161, 19. 317 objects are categorized as high prosperity, 161 objects are belonged in middle between two clusters and 19 objects have low prosperity.

Keywords : K-Means, Clustering, Principal Component Analysis, Human Development Index.



ABSTRAK

Pembangunan manusia merupakan hal yang wajib dilakukan oleh pemerintah dalam upaya peningkatan kesejahteraan rakyat. Saat ini masih banyak ketimpangan pembangunan manusia antar wilayah di Indonesia. Salah satu indikator untuk mengukur tingkat keberhasilan pembangunan manusia pada suatu negara adalah Indeks Pembangunan Manusia (IPM). IPM terdiri dari 4 indikator, yaitu angka harapan hidup, angka melek huruf, lama sekolah dan pengeluaran perkapita. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan karakteristik capaian pembangunan yang dimilikinya menggunakan algoritma K-Means. Data yang digunakan yaitu data IPM tahun 2012, dengan jumlah kabupaten dan kota sebanyak 497. Adapun ekstraksi ciri menggunakan Principal Component Analysis. PCA menghasilkan dua variabel sebagai komponen utama dengan total nilai keragaman 98.68%. Hasil pengklasteran yang diperoleh berupa dengan nilai evaluasi Davies Bouldin Index 0.6164 dan pola keanggotaan 317, 161, 19. Cluster dengan jumlah 317 memiliki kesejahteraan yang tinggi, sedangkan yang berjumlah 161 merupakan pertengahan antara kedua cluster. Adapun cluster yang berjumlah 19 masuk ke dalam kategori tidak sejahtera.

Kata kunci : K-Means, Clustering, Principal Component Analysis, Indeks Pembangunan Manusia.



KATA PENGANTAR

Puji syukur ke hadirat Allah SWT karena atas rahmat dan hidayah-Nya penulis mampu menyelesaikan penyusunan skripsi ini dengan judul “Algoritma K-Means untuk Pengelompokan Kabupaten/Kota di Indonesia berdasarkan Indikator Indeks Pembangunan Manusia”.

Skripsi ini disusun guna melengkapi salah satu persyaratan untuk memperoleh gelar kesarjanaan Progam Studi Teknik Informatika Fakultas Teknik Universitas Muria Kudus. Kiranya dalam penyusunan skripsi ini tidak akan terselesaikan dengan baik tanpa bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya, penghargaan yang setinggi-tingginya dan permohonan maaf atas segala kesalahan yang pernah penulis lakukan kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini, terutama kepada :

1. Bapak Dr. Suparnyo, S.H, M.S, selaku Rektor Universitas Muria Kudus.
2. Bapak Muhammad Dahlan, ST, MT, selaku Dekan Fakultas Teknik Universitas Muria Kudus.
3. Bapak Ahmad Jazuli, M.Kom, selaku Ketua Program Studi Teknik Informatika Universitas Muria Kudus.
4. Ibu Endang Supriyati, M.Kom, selaku pembimbing skripsi penulis.
5. Ibu Tutik Khotimah, M.Kom, selaku pembimbing skripsi penulis.
6. Semua pihak yang telah membantu penyusunan skripsi ini yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari sepenuhnya bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan, untuk itu penulis mengharap kritik dan saran dari berbagai pihak untuk sempurnanya sebuah karya tulis. Selain itu penulis juga berharap semoga karya tulis ini dapat memberikan manfaat bagi semua.

Kudus, 19 Mei 2015

Penulis

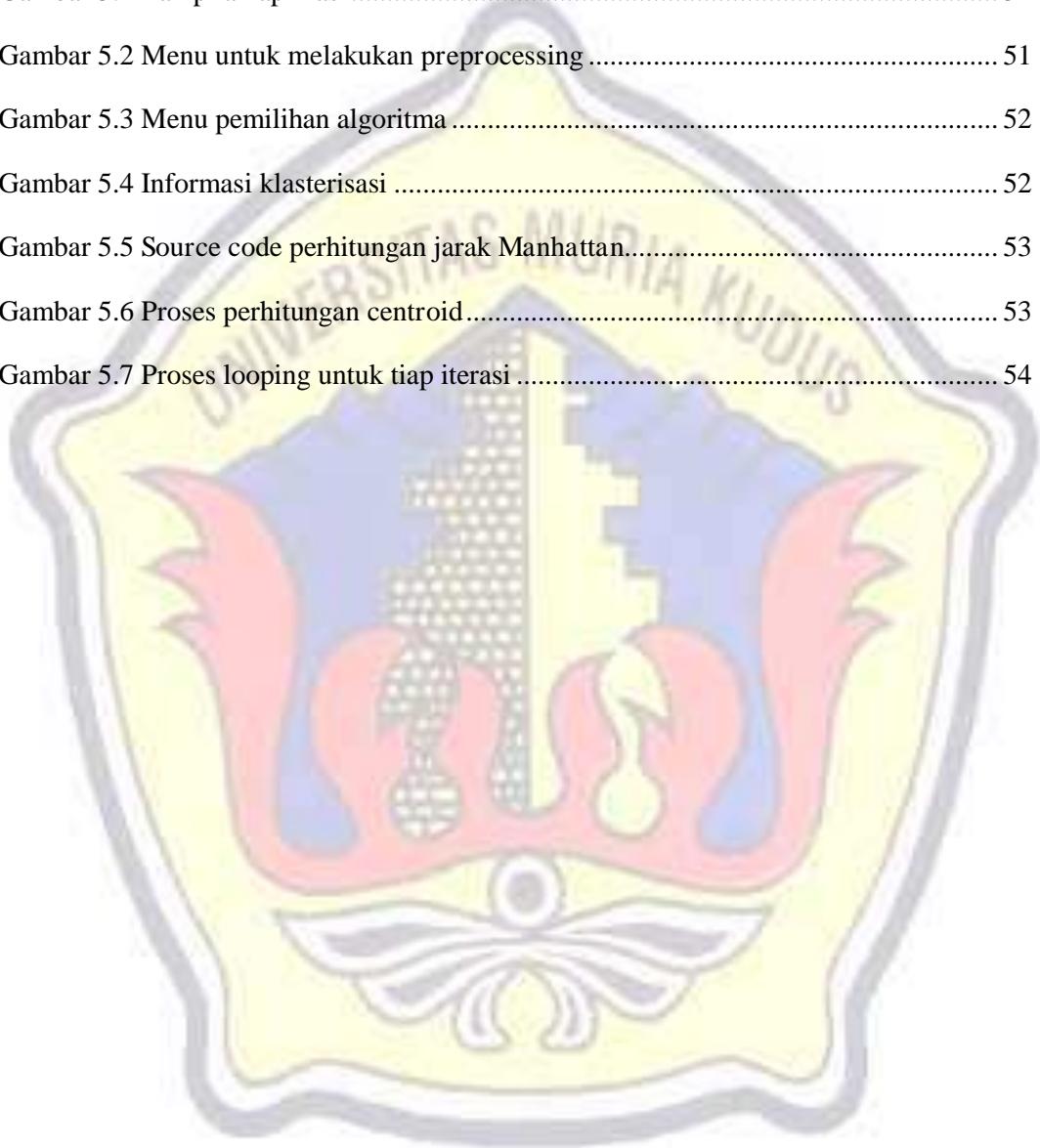
DAFTAR ISI

| | |
|---|-----|
| LAPORAN SKRIPSI..... | i |
| PENGESAHAN STATUS SKRIPSI | ii |
| PERNYATAAN PENULIS | iii |
| PERSETUJUAN SKRIPSI | iv |
| PENGESAHAN SKRIPSI | v |
| ABSTRACT | vi |
| ABSTRAK | vi |
| DAFTAR ISI..... | ix |
| DAFTAR GAMBAR | x |
| DAFTAR TABEL..... | xii |
| DAFTAR LAMPIRAN..... | xv |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Batasan Masalah | 3 |
| 1.3 Rumusan Masalah..... | 3 |
| 1.4 Tujuan Penelitian | 3 |
| 1.5 Manfaat Penelitian | 4 |
| 1.6 Sistematika Penulisan | 4 |
| BAB II TINJAUAN PUSTAKA..... | 5 |
| 2.1 Penelitian Terkait | 5 |
| 2.2 Landasan Teori..... | 6 |
| 2.2.1 Pengertian Indikator..... | 6 |
| 2.2.2 Indeks Pembangunan Manusia..... | 7 |
| 2.2.3 Principal Component Analysis..... | 8 |
| 2.2.4 Clustering..... | 8 |
| 2.2.5 Jarak Manhattan..... | 8 |
| 2.2.6 Davies Bouldin Index (DBI) | 9 |
| 2.2.7 Algoritma K-Means | 10 |
| 2.3 Kerangka Pikir | 11 |
| BAB III METODE PENELITIAN | 12 |
| 3.1 Sumber Data..... | 12 |
| 3.2 Alur Analisis | 12 |
| 3.3 Pengumpulan Data | 13 |

| | | |
|----------------|--|----|
| 3.4 | Studi Pustaka..... | 13 |
| 3.5 | Preprocessing Data..... | 13 |
| 3.6 | Ekstraksi Ciri | 13 |
| 3.7 | Penerapan Algoritma K-Means..... | 14 |
| 3.8 | Evaluasi Cluster | 14 |
| 3.9 | Analisis Hasil..... | 14 |
| BAB IV | ANALISIS DAN PEMBAHASAN..... | 15 |
| 4.1 | Deskripsi Data..... | 15 |
| 4.2 | Preprocessing | 15 |
| 4.3 | Ekstraksi Ciri dengan Principal Component Analysis | 16 |
| 4.4 | Pengukuran Jarak Manhattan..... | 40 |
| 4.4.1 | Clustering Menggunakan Ekstraksi Ciri..... | 40 |
| 4.4.2 | Clustering Tanpa Ekstraksi Ciri..... | 46 |
| BAB V. | IMPLEMENTASI..... | 51 |
| 5.1 | Antarmuka Aplikasi | 51 |
| 5.2 | Implementasi Algoritma | 53 |
| 5.3 | Hasil Pengujian | 54 |
| 5.3.1 | Evaluasi untuk Clustering dengan Ekstraksi Ciri | 54 |
| 5.3.2 | Evaluasi untuk Clustering tanpa Ekstraksi Ciri | 56 |
| 5.3.3 | Clustering dengan Ekstraksi Ciri | 57 |
| 5.3.4 | Clustering tanpa Ekstraksi Ciri | 58 |
| 5.3.5 | Karakteristik Masing-masing Cluster | 58 |
| BAB VI | PENUTUP | |
| 6.1 | Kesimpulan | 60 |
| 6.2 | Saran | 60 |
| DAFTAR PUSTAKA | | 61 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Kerangka pikir..... | 11 |
| Gambar 2.1 Alur analisis penelitian..... | 12 |
| Gambar 5.1 Tampilan aplikasi | 51 |
| Gambar 5.2 Menu untuk melakukan preprocessing | 51 |
| Gambar 5.3 Menu pemilihan algoritma | 52 |
| Gambar 5.4 Informasi klasterisasi | 52 |
| Gambar 5.5 Source code perhitungan jarak Manhattan..... | 53 |
| Gambar 5.6 Proses perhitungan centroid..... | 53 |
| Gambar 5.7 Proses looping untuk tiap iterasi | 54 |



DAFTAR TABEL

| | |
|---|----|
| Tabel 4.1 Data sebelum dinormalisasi | 15 |
| Tabel 4.2 Data sebelum dinormalisasi | 16 |
| Tabel 4.3 Data setelah dinormalisasi | 16 |
| Tabel 4.4 Matriks covariance 4 x 4 | 17 |
| Tabel 4.5 Perhitungan covariance X_1 dan X_1 | 17 |
| Tabel 4.6 Perhitungan covariance X_1 dan X_2 | 18 |
| Tabel 4.7 Perhitungan covariance X_1 dan X_3 | 18 |
| Tabel 4.8 Perhitungan covariance X_1 dan X_4 | 19 |
| Tabel 4.9 Perhitungan covariance X_2 dan X_1 | 20 |
| Tabel 4.10 Perhitungan covariance X_2 dan X_2 | 20 |
| Tabel 4.11 Perhitungan covariance X_2 dan X_3 | 21 |
| Tabel 4.12 Perhitungan covariance X_2 dan X_4 | 21 |
| Tabel 4.13 Perhitungan covariance X_3 dan X_1 | 22 |
| Tabel 4.14 Perhitungan covariance X_3 dan X_2 | 22 |
| Tabel 4.15 Perhitungan covariance X_3 dan X_3 | 23 |
| Tabel 4.16 Perhitungan covariance X_3 dan X_4 | 23 |
| Tabel 4.17 Perhitungan covariance X_4 dan X_1 | 24 |
| Tabel 4.18 Perhitungan covariance X_4 dan X_2 | 24 |
| Tabel 4.19 Perhitungan covariance X_4 dan X_3 | 25 |
| Tabel 4.20 Perhitungan covariance X_4 dan X_4 | 25 |
| Tabel 4.21 Hasil seluruh covariance..... | 26 |
| Tabel 4.22 Hasil eigenvalue..... | 36 |
| Tabel 4.23 Eigenvector untuk eigenvalue 1 dan eigenvalue 2..... | 38 |
| Tabel 4.24 Perkalian antara variabel normalisasi dengan eigenvector | 38 |

| | |
|---|----|
| Tabel 4.25 Perkalian antara variabel normalisasi dengan eigenvector | 39 |
| Tabel 4.26 Hasil penjumlahan keempat variabel | 39 |
| Tabel 4.27 Inisialisai titik centroid | 40 |
| Tabel 4.28 Hasil perhitungan jarak pada iterasi pertama..... | 40 |
| Tabel 4.29 Hasil perhitungan centroid baru pada iterasi pertama..... | 41 |
| Tabel 4.30 Hasil perhitungan jarak pada iterasi kedua | 41 |
| Tabel 4.31 Hasil perhitungan centroid baru pada iterasi kedua..... | 42 |
| Tabel 4.32 Hasil perhitungan jarak pada iterasi ketiga | 42 |
| Tabel 4.33 Hasil perhitungan centroid baru pada iterasi ketiga..... | 43 |
| Tabel 4.34 Hasil perhitungan jarak pada iterasi keempat | 43 |
| Tabel 4.35 Hasil perhitungan centroid baru pada iterasi keempat | 43 |
| Tabel 4.36 Hasil perhitungan jarak pada iterasi kelima..... | 44 |
| Tabel 4.37 Hasil perhitungan centroid baru pada iterasi kelima..... | 44 |
| Tabel 4.38 Hasil perhitungan jarak pada iterasi keenam | 45 |
| Tabel 4.39 Hasil perhitungan centroid baru pada iterasi keenam..... | 45 |
| Tabel 4.40 Hasil perhitungan jarak pada iterasi ketujuh..... | 45 |
| Tabel 4.41 Hasil perhitungan centroid baru pada iterasi ketujuh | 46 |
| Tabel 4.42 Titik inisialisasi centroid..... | 46 |
| Tabel 4.43 Hasil perhitungan jarak pada iterasi pertama..... | 47 |
| Tabel 4.44 Hasil perhitungan centroid baru pada iterasi pertama..... | 47 |
| Tabel 4.45 Hasil perhitungan jarak pada iterasi kedua | 48 |
| Tabel 4.46 Hasil perhitungan centroid baru pada iterasi kedua..... | 48 |
| Tabel 4.47 Hasil perhitungan jarak pada iterasi ketiga | 49 |
| Tabel 4.48 Hasil perhitungan centroid baru pada iterasi ketiga..... | 49 |
| Tabel 4.49 Hasil perhitungan jarak pada iterasi keempat | 50 |
| Tabel 4.50 Hasil perhitungan centroid baru pada iterasi keempat..... | 50 |

| | |
|--|----|
| Tabel 5.1 Hasil rata-rata jarak Manhattan untuk setiap cluster | 54 |
| Tabel 5.2 Hasil perhitungan SSB | 55 |
| Tabel 5.3 Hasil perhitungan DBI..... | 55 |
| Tabel 5.4 Hasil rata-rata jarak Manhattan untuk tiap cluster | 56 |
| Tabel 5.5 Hasil perhitungan SSB | 56 |
| Tabel 5.6 Hasil perhitungan R | 57 |
| Tabel 5.7 Nilai centroid akhir..... | 57 |
| Tabel 5.8 Nilai centroid akhir..... | 58 |



DAFTAR LAMPIRAN

- Lampiran 1 : Buku bimbingan skripsi
- Lampiran 2 : Lembar revisi sidang
- Lampiran 3 : Data asli dan setelah dinormalisasi
- Lampiran 4 : Covariance
- Lampiran 5 : Komponen 1 dan komponen 2
- Lampiran 6 : Iterasi 1, 2, dan 3 (dengan ekstraksi ciri)
- Lampiran 7 : Iterasi 4, 5, 6 dan 7 (dengan ekstraksi ciri)
- Lampiran 8 : Iterasi 1, 2, 3 dan 4 (tanpa ekstraksi ciri)

