

**PERBANDINGAN METODE *K-MEANS* DAN METODE
DBSCAN PADA PENGELOMPOKAN RUMAH KOST
MAHASISWA DI KELURAHAN TEMBALANG SEMARANG**



SKRIPSI

Disusun Oleh :

SISCA AGUSTIN DIANI BUDIMAN

NIM. 24010212130032

DEPARTEMEN STATISTIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

SEMARANG

2016

**PERBANDINGAN METODE *K-MEANS* DAN METODE
DBSCAN PADA PENGELOMPOKAN RUMAH KOST
MAHASISWA DI KELURAHAN TEMBALANG SEMARANG**

Disusun Oleh:

SISCA AGUSTIN DIANI BUDIMAN

24010212130032

Tugas Akhir sebagai salah satu syarat untuk memperoleh
gelar Sarjana Sains pada Jurusan Statistika

**DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2016

HALAMAN PENGESAHAN I

Judul : Perbandingan Metode *K-Means* dan Metode DBSCAN pada
Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang
Semarang

Nama : Sisca Agustin Diani Budiman

NIM : 24010212130032

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir dan dinyatakan lulus pada tanggal 27
September 2016

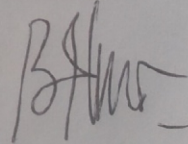
Semarang, 30 September 2016

Mengetahui,
a.n. Ketua Departemen Statistika
Sekretaris Departemen Statistika
Fakultas Sains dan Matematika



Drs. Agus Rusgiyono, M.Si
NIP. 196408131990011001

Panitia Penguji Ujian Tugas Akhir
Ketua,



Dr. Rukun Santoso, M.Si
NIP. 196502251992011001

HALAMAN PENGESAHAN II

Judul : Perbandingan Metode *K-Means* dan Metode DBSCAN pada
Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang
Semarang

Nama : Sisca Agustin Diani Budiman

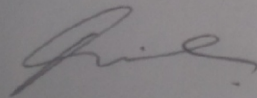
NIM : 24010212130032

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 27 September 2016

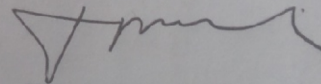
Semarang, 30 September 2016

Pembimbing I



Diah Safitri, S.Si, M.Si
NIP. 197510082003122001

Pembimbing II



Dra. Dwi Ispriyanti, M.Si
NIP. 195709141986032001

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan rahmat dan karunia Nya, sehingga Tugas Akhir yang berjudul ”Perbandingan Metode *K-Means* dan Metode DBSCAN pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang“ ini dapat diselesaikan. Banyak pihak yang telah membantu dalam penyelesaian Tugas Akhir ini. Oleh karena itu, rasa hormat dan terimakasih penulis ingin sampaikan kepada :

1. Ibu Dra. Dwi Ispriyanti, M.Si selaku Ketua Departemen Statistika Universitas Diponegoro Semarang dan selaku dosen pembimbing II.
2. Ibu Diah Safitri, S.Si, M.Si selaku dosen pembimbing I yang telah memberikan bimbingan dan pengarahan dalam penyusunan Tugas Akhir ini.
3. Semua Bapak dan Ibu Dosen Departemen Statistika Universitas Diponegoro yang telah memberikan ilmunya yang sangat berguna.
4. Pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu yang telah membantu penulisan Tugas Akhir ini.

Penulis menyadari bahwa dalam penulisan Tugas Akhir ini masih banyak kekurangannya. Untuk itu penulis mengharapkan kritik dan saran demi kesempurnaan laporan tugas akhir ini.

Semarang, September 2016

Penulis

ABSTRAK

Mahasiswa sama halnya dengan masyarakat atau rumah tangga, juga melakukan aktivitas ekonomi sehari-hari termasuk konsumsi. Kebutuhan mahasiswa untuk memilih tempat tinggal juga merupakan salah satu bentuk dari kegiatan konsumsi. Terdapat banyak faktor yang mempengaruhi preferensi mahasiswa dalam pemilihan rumah kost, di antaranya harga, fasilitas, lokasi, pendapatan, gaya hidup, dan lain-lain. Harga sewa kost dan fasilitas yang ditawarkan berpengaruh signifikan positif terhadap preferensi mahasiswa dalam memilih rumah kost. Dilakukan pengelompokan berdasarkan harga sewa dan fasilitas yang ditawarkan agar diketahui kondisi rumah kost mahasiswa di Kelurahan Tembalang. Pengelompokan merupakan salah satu tugas utama dalam data mining dan telah banyak diaplikasikan di berbagai bidang. Metode yang digunakan untuk mengelompokkan adalah *K-Means* dan DBSCAN dengan jumlah kelompok sebanyak tiga. Selanjutnya hasil dari kedua metode dibandingkan menggunakan nilai Indeks *Silhouette* untuk menentukan metode mana yang lebih baik dalam mengelompokkan rumah kost mahasiswa. Berdasarkan hasil penelitian yang telah dilakukan ditemukan bahwa metode *K-Means* bekerja lebih baik dibanding DBSCAN dalam mengelompokkan rumah kost mahasiswa yang dibuktikan dengan nilai Indeks *Silhouette* pada *K-Means* sebesar 0,463 lebih tinggi dari nilai Indeks *Silhouette* pada DBSCAN yaitu sebesar 0,281.

Kata Kunci: rumah kost mahasiswa, data mining, pengelompokan, *K-Means*, DBSCAN

ABSTRACT

Students as well as community or household, as well as economic activities daily, including consumption. The student needs to choose a place to stay is also one form of consumption activities. There are many factors that affect student preferences in the selection of boarding houses, including price, amenities, location, income, lifestyle, and others. The rental price boarding and facilities offered significant positive effect on student preferences in choosing a boarding house. Do groupings based on rental rates and facilities offered in order to know the condition of the student boarding house in the Village Tembalang. Grouping is one of the main tasks in data mining and have been widely applied in various fields. The method used to classify is *K-Means* and DBSCAN with a number of groups of three. Furthermore, the results of both methods were compared using the *Silhouette* index values to determine which method is better to classify the student boarding house. Based on the research that has been conducted found that the *K-Means* method works better than DBSCAN to classify the student boarding house as evidenced by the value of the *Silhouette* index on *K-Means* of 0.463 is higher than the value at DBSCAN *Silhouette* index is equal to 0.281.

Keywords: student boarding houses, data mining, clustering, *K-Means*, DBSCAN

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN I	Error! Bookmark not defined.
HALAMAN PENGESAHAN II	Error! Bookmark not defined.
KATA PENGANTAR	Error! Bookmark not defined.
ABSTRAK	Error! Bookmark not defined.
ABSTRACT	Error! Bookmark not defined.
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
BAB I PENDAHULUAN	Error! Bookmark not defined.
1.1. Latar Belakang	Error! Bookmark not defined.
1.2. Rumusan Masalah	Error! Bookmark not defined.
1.3. Batasan Masalah.....	Error! Bookmark not defined.
1.4. Tujuan.....	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
2.1. Data Mining.....	Error! Bookmark not defined.
2.2. Pemrosesan Data Awal (<i>Preprocessing Data</i>).....	Error! Bookmark not defined.
2.3. Analisis Kelompok	Error! Bookmark not defined.
2.3.1. Konsep Pengelompokan.....	Error! Bookmark not defined.
2.3.2. Tujuan Pengelompokan.....	Error! Bookmark not defined.
2.3.3. Jenis Pengelompokan	Error! Bookmark not defined.
2.4. <i>K-MEANS</i>	Error! Bookmark not defined.
2.3.1 Algoritma <i>K-Means</i>	Error! Bookmark not defined.
2.3.2 Karakteristik <i>K-Means</i>	Error! Bookmark not defined.
2.5. DBSCAN (<i>Density-Based Spatial Clustering Algorithm with Noise</i>)	Error! Bookmark not defined.
2.5.1. Konsep DBSCAN	Error! Bookmark not defined.
2.5.2. Algoritma DBSCAN	Error! Bookmark not defined.

2.6.	Sampel Acak Berlapis	Error! Bookmark not defined.
2.6.1.	Kerangka Sampel	Error! Bookmark not defined.
2.6.2.	Penarikan Sampel Acak Stratifikasi Proporsional (<i>Proportionate Stratified Random Sampling</i>).....	Error! Bookmark not defined.
2.7.	Rumah Kost	Error! Bookmark not defined.
2.8.	Harga	Error! Bookmark not defined.
2.9.	Fasilitas.....	Error! Bookmark not defined.
2.10.	Validasi Kelompok	Error! Bookmark not defined.
BAB III METODOLOGI PENELITIAN		Error! Bookmark not defined.
3.1.	Jenis dan Sumber Data	Error! Bookmark not defined.
3.2.	Teknik Pengolahan Data	Error! Bookmark not defined.
3.3.	Diagram Alir.....	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN		Error! Bookmark not defined.
4.1.	Analisis Deskriptif.....	Error! Bookmark not defined.
4.2.	<i>Preprocessing</i> (Pemrosesan Awal Data)	Error! Bookmark not defined.
4.3.	Proses Pengelompokan dengan Metode <i>K-Means</i>	Error! Bookmark not defined.
4.4.	Proses Pengelompokan dengan Metode DBSCAN.....	Error! Bookmark not defined.
4.5.	Perhitungan nilai Indeks <i>Silhouette</i>	Error! Bookmark not defined.
4.5.1.	Perhitungan nilai Indeks <i>Silhouette</i> pada <i>K-Means</i>	Error! Bookmark not defined.
4.5.2.	Perhitungan nilai Indeks <i>Silhouette</i> pada DBSCAN.....	Error! Bookmark not defined.
4.6.	Interpretasi Hasil Pengelompokan yang Paling Optimal	Error! Bookmark not defined.
BAB V PENUTUP.....		Error! Bookmark not defined.
5.1	Kesimpulan.....	Error! Bookmark not defined.
5.2	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA		Error! Bookmark not defined.
LAMPIRAN.....		Error! Bookmark not defined.

DAFTAR GAMBAR

Halaman

Gambar 1. Macam – macam pendefinisian kelompok pada titik yang sama. **Error!**

Bookmark not defined.

Gambar 2. Kategori titik pada DBSCAN.....**Error! Bookmark not defined.**

Gambar 3. Diagram Alir**Error! Bookmark not defined.**

Gambar 4. Plot data Hasil Normalisasi**Error! Bookmark not defined.**

Gambar 5. Pengelompokan menggunakan *K-Means* sebanyak 3 kelompok. **Error!**

Bookmark not defined.

Gambar 6. Pengelompokan menggunakan DBSCAN sebanyak 3 kelompok **Error!**

Bookmark not defined.

DAFTAR TABEL

	Halaman
Tabel 1. Populasi dan Sampel Rumah Kost.....	Error! Bookmark not defined.
Tabel 2. Statistik Deskriptif	Error! Bookmark not defined.
Tabel 3. Nilai Sentroid Awal	Error! Bookmark not defined.
Tabel 4. Perubahan Posisi Kelompok Pada Setiap Iterasi....	Error! Bookmark not defined.
Tabel 5. Hasil Perhitungan Iterasi Pertama.....	Error! Bookmark not defined.
Tabel 6. Indeks <i>Silhouette</i> pada <i>K-Means</i> dan DBSCAN....	Error! Bookmark not defined.
Tabel 7. Interpretasi Variabel untuk Masing – Masing Kelompok.....	Error! Bookmark not defined.
Tabel 8. Harga dan Fasilitas Masing – Masing Kelompok..	Error! Bookmark not defined.

DAFTAR LAMPIRAN

Halaman

- Lampiran 1. Kuesioner Pengelompokan Rumah Kost Mahasiswa **Error! Bookmark not defined.**
- Lampiran 2. Data Rumah Kost Mahasiswa.....**Error! Bookmark not defined.**
- Lampiran 3. Data normalisasi dengan metode Mix Max..... **Error! Bookmark not defined.**
- Lampiran 4. Clustering Kmeans Data Rumah Kost (k=3)... **Error! Bookmark not defined.**
- Lampiran 5. Clustering DBSCAN Data Rumah Kost (eps=0.1,minPts=2) ... **Error! Bookmark not defined.**
- Lampiran 6. Syntax R Clustering Kmeans.....**Error! Bookmark not defined.**
- Lampiran 7. Syntax R Clustering DBSCAN**Error! Bookmark not defined.**

BAB I

PENDAHULUAN

1.1. Latar Belakang

Tembalang merupakan salah satu wilayah di Semarang. Wilayah ini merupakan salah satu kawasan ramai di Ibukota Provinsi Jawa Tengah (Semarang). Salah satu faktor ramainya wilayah ini dikarenakan di kawasan Tembalang terdapat beberapa universitas di antaranya Universitas Diponegoro, Politeknik Negeri Semarang, dan Universitas Pandanaran. Menurut Kamus Besar Bahasa Indonesia (KBBI), mahasiswa adalah mereka yang sedang belajar di perguruan tinggi. Mahasiswa sama halnya dengan masyarakat atau rumah tangga, juga melakukan aktivitas ekonomi sehari-hari termasuk konsumsi. Kebutuhan mahasiswa untuk memilih tempat tinggal juga merupakan salah satu bentuk dari kegiatan konsumsi. Bagi setiap individu tidak terkecuali mahasiswa kebutuhan yang sangat penting untuk segera terpenuhi adalah tempat tinggal atau rumah.

Usaha rumah kost umumnya banyak ditemukan di daerah yang berdekatan dengan pusat kegiatan rutinitas, salah satunya di Kelurahan Tembalang. Terdapat banyak faktor yang mempengaruhi preferensi mahasiswa dalam pemilihan rumah kost, di antaranya harga, fasilitas, lokasi, pendapatan, gaya hidup, dan lain-lain. Menurut Rachmawati (2013) harga sewa kost dan fasilitas yang ditawarkan berpengaruh signifikan positif terhadap preferensi mahasiswa dalam memilih rumah kost.

Analisis kelompok (*cluster analysis*) adalah mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang

menggambarkan objek tersebut dan hubungan antar data (Tan *et al.*, 2006). Tujuan analisis kelompok adalah agar objek–objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok yang lain (Prasetyo, 2012).

Terdapat banyak metode yang dapat digunakan untuk mengelompokkan data, di antaranya ada *K-Means*, pengelompokkan hierarki, *Fuzzy C-Means*, *self-organizing map*, dan DBSCAN. *K-Means* adalah pengelompokan data nonhierarki (sekatan) yang mempartisi data ke dalam bentuk dua atau lebih kelompok, sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama. *Density-Based Spatial Clustering Algorithm with noise* (DBSCAN) adalah algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data. Konsep kepadatan dalam DBSCAN menghasilkan tiga macam status dari setiap data, yaitu inti (*core*), batas (*border*), dan *noise*. Jumlah kelompok yang terbentuk bergantung pada dua nilai yaitu Epsilon (Eps) yang merupakan radius (jarak maksimal) data anggota kelompok dari data inti, dan *Minimal Points* (MinPts) yang merupakan banyaknya minimal data kelompok dalam radius Eps. *K-Means* dan DBSCAN masuk kedalam pengelompokan sekatan (*partitioning*) dan eksklusif, yang artinya *K-Means* dan DBSCAN merupakan metode pengelompokan yang membagi set data ke dalam sejumlah kelompok yang tidak tumpang–tindih antara satu kelompok dengan kelompok yang lain. DBSCAN dapat mendeteksi data yang mempunyai jarak yang jauh dari suatu kelompok (*noise*) yaitu data yang tidak ikut dalam kelompok mayoritas (Tan *et al.*, 2006).

Dalam melakukan pengelompokan metode *K-Means* sering dibandingkan dengan metode yang lain. Berdasarkan penelitian yang telah ada, Flowrensia (2010) melakukan perbandingan pengelompokan *K-Means* dan K Medoid pada data yang mengandung pencilan. Merliana (2015) melakukan Perbandingan Metode *K-Means* dengan *Fuzzy C-Means* untuk analisa karakteristik mahasiswa berdasarkan kunjungan ke perpustakaan. Pada penelitian yang telah ada, salah satu contoh perbandingan pengelompokan dengan metode DBSCAN yaitu Ayu (2015) melakukan perbandingan kinerja *Fuzzy C-Means* dan DBSCAN dalam segmentasi citra USG kepala janin.

Berdasarkan penelusuran pustaka belum ditemukan penggunaan metode *K-Means* dan DBSCAN untuk pengelompokan rumah kost mahasiswa. Data hasil penelitian merupakan data *partitioning* dan dapat dipastikan hanya menjadi satu anggota dalam kelompok maka cocok dilakukan pengelompokan dengan metode *K-Means* dan metode DBSCAN. Selain itu dengan banyaknya data yang dihasilkan memungkinkan adanya data noise sehingga cocok dilakukan pengelompokan menggunakan DBSCAN. Hasil yang diperoleh dari penelitian ini adalah kelompok mana yang paling optimal diantara kedua metode tersebut. Berdasarkan kelompok yang paling optimal kemudian dilakukan interpretasi adalah beberapa kelompok dengan interval harga tertentu dengan jumlah fasilitas yang didapatkan dengan harga tersebut. Berdasarkan uraian tersebut, maka peneliti mengambil judul untuk Tugas Akhir ini adalah “Perbandingan Metode *K-Means* dan Metode DBSCAN dalam Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang”.

1.2. Rumusan Masalah

Rumusan masalah yang diambil pada Tugas Akhir ini adalah:

1. Menentukan cara pengelompokan menggunakan metode *K-Means* dan metode DBSCAN pada rumah kost mahasiswa di Kelurahan Tembalang Semarang.
2. Menentukan cara perbandingan pengelompokan menggunakan metode *K-Means* dan metode DBSCAN pada rumah kost mahasiswa di Kelurahan Tembalang Semarang.
3. Bagaimana interpretasi dari kelompok yang paling optimal yang dihasilkan dari perbandingan metode *K-Means* dan metode DBSCAN pada rumah kost mahasiswa di Kelurahan Tembalang Semarang.

1.3. Batasan Masalah

Batasan masalah pada Tugas Akhir ini adalah :

1. Lokasi Penelitian bertempat di Kelurahan Tembalang Semarang.
2. Data yang digunakan adalah data primer dari hasil penelitian menggunakan kuesioner pada mahasiswa yang menempati rumah kost di Kelurahan Tembalang Semarang.
3. Waktu pengambilan data bulan Maret 2016 – Agustus 2016.

1.4. Tujuan

Tujuan dari Tugas Akhir ini adalah :

1. Mengelompokkan rumah kost mahasiswa di Kelurahan Tembalang Semarang menggunakan metode *K-Means* dan metode DBSCAN.

2. Melakukan perbandingan untuk menentukan kelompok yang paling optimal dalam pengelompokan menggunakan metode *K-Means* dan metode DBSCAN pada rumah kost di Kelurahan Tembalang Semarang.
3. Melakukan interpretasi dari kelompok yang paling optimal yang dihasilkan dari perbandingan metode *K-Means* dan metode DBSCAN pada rumah kost mahasiswa di Kelurahan Tembalang Semarang.