PENGELOMPOKAN SEKOLAH MENENGAH ATAS (SMA) DI KABUPATEN KEDIRI DENGAN METODE *ENSEMBEL ROBUST CLUSTERING USING LINKS* (ROCK)

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

ANI KHOLIFAH

B2A016010

PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2021

HALAMAN PENGESAHAN

Skripsi dengan Judul "**Pengelompokan Sekolah Menengah Atas (SMA) di Kabupaten Kediri dengan *Metode Ensembel Robust Clustering Using Links* (ROCK)**" yang disusun oleh :

Nama	: Ani Kholifah
NIM	: B2A016010
Program Studi	: S-1 STATISTIKA

telah disetujui oleh dosen pembimbing pada tanggal : 2 Mei 2021

Pembimbing Utama	Pembimbing Pendamping
Indah Manfaati Nur, S.Si., M.Si	Prizka Rismawati Arum, M.Stat
NIK. 28.6.1026.221	NIP.CP.1026.071

Mengetahui,
Ketua Program Studi Statistika

Indah Manfaati Nur, S.Si., M.Si
NIK. 28.6.1026.221

SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH

Yang bertandatangan di bawah ini, saya :

Nama	: Ani Kholifah
NIM	: B2A016010
Fakultas/Jurusan	: Matematika dan Ilmu Pengetahuan Alam/Statistika
Jenis Penelitian	: Skripsi
Judul	: Pengelompokan Sekolah Menengah Atas (SMA) di Kabupaten Kediri dengan *Metode Ensembel Robust Clustering Using Links* (ROCK)"
Email	: akholifah13@gmail.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada Perpustakaan Unimus atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan.
2. Memberikan hak menyimpan, mengalih mediakan/ mengalih formatkan, mengelola dalam bentuk pangkalan data (*database*), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada Perpustakaan Unimus, tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta.
3. Bersedia dan menjamin untuk menganggung secara pribadi tanpa melibatkan pihak perpustakaan Unimus, dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 2 Mei 2021
Yang Menyatakan,

(Ani Kholifah)
NIM. B2A.016.010

Pengelompokan Sekolah Menengah Atas (SMA) di Kabupaten Kediri dengan Metode Ensembel Robust Clustering Using Links (ROCK)

Oleh: Ani Kholifah
Univeristas Muhammadiyah Semarang

Article history
Submission :
Revised :
Accepted :

Keyword:
Senior High School, Cluster, Ensembel ROCK

Abstract

The quality of senior high school education must always be improved so as to be able to create quality graduates because at this stage students have begun to form their individual abilities according to their goals. In 2020, it was noted that SMA in Kediri is 47 schools. Each school has different characteristics. Quite a lot of data and the different characteristics of each school make high school in Semarang need to be classified. The purpose of this study is to group high schools in the city of Semarang based on indicators of high school infrastructure and human resources. Development of clustering or clustering methods is generally used to process one type of numeric or categorical data type, but the data in this study are of mixed type (numeric and categorical) where the numerical variables are student/group ratio, number of permanent teachers, number of non-permanent teachers, number of non-permanent teachers, land area, and electric power and categorical variables including high school accreditation, high school status, and time of implementation. The method used is the Ensuring Robust Clustering Using Links (ROCK) where numeric data grouping uses non Hierarchical Algorithm method and Grouping categorical data using the ROCK method. The ROCK method will be used again in determining the final cluster.

PENDAHULUAN

Pendidikan yaitu sebuah proses pembelajaran bagi setiap individu untuk mencapai pengetahuan dan pemahaman yang lebih tinggi mengenai suatu objek dan spesifikasi tertentu (KBBI). Kementerian Pendidikan dan Kebudayaan (Kemendikbud) mewajibkan pendidikan di Indonesia selama 12 tahun dari SD, SMP, dan SMA. Mutu pendidikan SMA harus selalu di tingkatkan sehingga mampu menciptakan lulusan yang berkualitas karena pada tahapan inilah peserta didik sudah mulai di bentuk kemampuan individunya sesuai tujuan .

Pada tahun 2020, tercatat bahwa SMA di Kabupaten Kediri baik negeri maupun swasta berjumlah SMA 47 sekolah (Cabang Dinas Pendidikan Wilayah Kediri). Setiap sekolah memiliki karakteristik yang berbeda. Data yang cukup banyak serta adanya perbedaan karakteristik tiap sekolah menjadikan sekolah SMA di Kabupaten Kediri perlu untuk di kelompokan. Pengelompokan atau sering disebut clustering merpakan sebuah metode di dalam data mining yang bertujuan untuk mengelompokkan data dengan karakteristik yang sama kedalam satu kelompok dan data dengan karakteristik berbeda ke dalam kelompok lainnya.

Penelitian pengelompokan data campuran sudah beberapakali di lakukan, Ichsan (2018) melakukan penelitian dengan judul pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan pembangunan kualitas sumberdaya manusia dan pembangunn ekonomi dengan *menggunakan* metode ensembel ROCK.

Prakoso (2017) melakukan pengelompokan SMA di Sidoarjo menggunakan similarity weight and filter method (SWFM). Alvionita (2017) melakukan penelitian metode ensembel ROCK dan SWFM untuk mengelompokan data campuran numerik dan kategorik pada kasus aksesi jeruk dan hasil penelitian menujukan bahwa metode ensembel ROCK memberikan kinerja pengelompokan lebih baik dari pada metode ensembel SWFM.

Secara umum metode dalam analisis klaster dibagi dua yakni Hierarchical Clustering (metode hirarki) dan Non-Hierarchical Clustering (metode tak hirarki) .Pada penelian ini akan menggunakan metode Non-Hierarchical Clustering pada pengelompokan data numerik . metode ini merupakan metode yang jumlah clusternya di tentukan terlebih dahulu, sehingga objek- objek akan di kelompokan pada k kelompok yang telah di tentukan oleh peneliti. Metode Non-Hierarchical Clustering yang akan di gunakan adalah algoritma K-Mean dan K-Medoid.

Pengelompokan data kategorik akan menggunakan metode RObust Clustering using linKs (ROCK). Metode ROCK menggunakan konsep link sebagai ukuran kemiripan untuk membentuk cluster-nya. Metode ROCK dapat menangani outlier dengan cukup efektif. Tahapan selanjutnya metode ROCK akan di gunakan kembali pada final cluster untuk pengabunggan cluster-cluster yang telah di hasilkan melalui kedua algoritma tersebut dengan menganggapnya sebagai data baru bertipe kategorik.

LANDASAN TEORI

Analisis Cluster

Menurut Tan (2006), analisis cluster adalah sebuah proses untuk mengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Ukuran kemiripan diukur dengan menggunakan ukuran jarak. Hasil dari anlisis cluster dapat di pengaruhi oleh objek yang di kelompokan, ukuran kemiripan/ketidak miripan, skala ukuran, dan metode clustering yang digunakan.

Ukuran kemiripan

Kemiripan antar pasangan objek x dan y di notasikan sebagai si (x,y).Dalam menentukan mirip tidaknya suatu objek, dapat di lihat dari besarnya nilai si (x,y). Pada objek yang memiliki kemiripan maka nilai si (x,y) akan besar dan sebaliknya apabila nilai si (x,y) kecil makan objek merupakan pasangan yang tidak mirip. Untuk setiap pasangan objek x dan y berlaku 3 kondisi berikut (Kandardzic, 2011):

1. $0 \leq si\ \ (x,y) \leq 1$, kemiripan bernilai 0 dan 1.
2. $si\ (x,y) = 1$,, setiap objek mirip dengan dirinya sendiri.
3. $si\ \ (x,y) = si\ (y,x)$, kemiripan besifat simetri

Cluster Ensemble

Strehl dan Gosh (2002) memperkenalkan sebuah metode yang digunakan untuk mengombinasikan sekumpulan solusi gerombol yang disebut Cluster Ensemble. Metode Cluster Ensemble memiliki keunggulan dibanding metode penggerombolan lain. Penelitian yang dilakukan oleh Strehl dan Gosh (2002) menunjukkan bahwa metode Cluster Ensemble mampu meningkatkan kualitas dan kekekaran solusi gerombol. Tantangan untuk mendapatkan solusi gerombol dengan kualitas yang baik dan adanya keragaman solusi gerombol yang dihasilkan dari metode yang berbeda merupakan motivasi dikembangkannya Cluster Ensemble.

Penggerombolan pada Cluster Ensemble dilakukan dengan mengombinasikan berbagai solusi dari berbagai metode penggerombolan hingga diperoleh satu penggerombolan akhir yang lebih baik. Input yang dibutuhkan adalah solusi penggerombolan yang telah diperoleh dengan menggunakan berbagai hasil penggerombolan tanpa melihat karakteristik data awal. Secara umum, penggerombolan objek dengan metode cluster ensemble dilakukan dalam dua tahap menurut Iam-on dan Garret (2010), yaitu:

1. Membentuk anggota ensemble yang anggotanya adalah solusi dari berbagai metode penggerombolan yang berbeda.

2. Mengombinasikan seluruh anggota ensemble untuk memperoleh satu solusi akhir yang dinamakan fungsi Consensus.

Pengelompokan Data Numerik

Penelitian ini menggunakan metode non hirarki sebagai metode pengelompokan data numerik. Metode ini dimulai dengan proses penentuan jumlah cluster terlebih dahulu. Terdapat beberapa metode non hirarki yang dapat d gunakan, salah satunya k-mean dan k-medoid.

- Algoritma K-Means

Algoritma K-Means merupakan algoritma yang relatif sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar obyek dengan atribut terentu ke dalam kelompok-kelompok sebanyak K. Metode ini menjadi salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih kelompok. Algoritma K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Pada algoritma K-Means jumlah cluster K telah ditentukan terlebih dahulu. Cara kerja algoritma K-Means:

1. Tentukan K sebagai jumlah cluster yang ingin dibentuk,
2. Bangkitkan K centroid (titik pusat cluster) awal secara random
3. Dalam menentukan buah pusat cluster awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data input. Pusat awal cluster didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan merandom pusat awal dari data.
4. Hitung jarak setiap data ke masing-masing centroids, untuk mengukur jarak antara data dengan pusat cluster digunakan Euclidian Distance.

- Algoritma K-Medoid

Dalam metode K-medoid setiap cluster dipresentasikan dari sebuah objek di dalam cluster yang disebut dengan medoid. Tujuannya adalah menemukan kelompok K-cluster (jumlah cluster) diantara semua objek data di dalam sebuah kelompok data. Clusternya dibangun dari hasil mencocokkan setiap objek data yang paling dekat dengan cluster yang dianggap sebagai medoid sementara. Langkah-langkah menghitung medoids, yaitu:

1. Pilih point k sebagai inisial centroid/nilai tengah (medoids) sebanyak k cluster.
2. Cari semua point yang paling dekat dengan medoid, dengan cara menghitung jarak vector antar dokumen. (menggunakan Euclidian distance)
3. Secara random, pilih point yang bukan medoid.
4. Hitung total distance
5. If TD baru < TD awal, tukar posisi medoid dengan medoids baru, jadilah medoid yang baru.
6. Ulangi langkah 2 - 5 sampai medoid tidak berubah.

Pengelompokan Data Kategorik

Metode *clustering* yang digunakan untuk tipe data kategorik adalah algoritma *ROCK*. *ROCK* pertama kali diperkenalkan oleh Guha, Rastogi, & Shim pada tahun 1999. Metode *ROCK* menggunakan konsep *link* sebagai ukuran kemiripan untuk membentuk *cluster*-nya. Metode *ROCK* dapat menangani *outlier* dengan cukup efektif. Pemangkasan *outlier* memungkinkan untuk membuang yang tidak ada tetangga, sehingga titik tersebut tidak berpartisipasi dalam pengelompokan. Namun dalam beberapa situasi, *outlier* dapat hadir sebagai *cluster-cluster* yang kecil (Guha, Rastogi, & Shim, 1999).

Clustering untuk data kategorik dengan algoritma *ROCK* dilakukan dengan tiga langkah. Adapun langkahnya yaitu sebagai berikut:

1. Menghitung *similaritas* menggunakan rumus *Jaccard coefficient* (Rahayu, 2009). Ukuran kemiripan antara objek ke$-i$dan objek ke $-j$ di hitung dengan rumusan

   $$si\left(X_i, X_j\right) = \frac{|X_i \cap X_j|}{|X_i \cup X_j|}, X_i \neq X_j$$

   Dimana :

   $si\left(X_i, X_j\right)$ = Ukuran kemiripan antara objek ke$-i$dan objek ke $-j$

   $i = 1,2,3, \dots, n\ d \quad j = 1,2,3, \dots, n$

   X_i = himpunan pengamatan ke $-i$ dengan $X_i = \{ X_{1i}, X_{2i}, X_{3i,\dots,} X_{mk} \quad _i$

   X_j = himpunan pengamatan ke $-j$ dengan $X_j = \{ X_{1j}, X_{2j}, X_{3j,\dots,} X_{mk} \quad _j$

   |X| = bilangan kardinal atau jumlah anggota dari himpunan.

2. Langkah kedua adalah menentukan tetangga. Pengamatan dinyatakan sebagai tetangga

jika nilai $si\ (X_i, X_j) \geq \Theta$.

3. Langkah terakhir adalah menghitung *link* antar objek pengamatan. Besarnya *link* dipengaruhi oleh nilai *threshold* (Θ) yang merupakan parameter yang ditentukan oleh pengguna yang dapat digunakan untuk mengontrol seberapa dekat hubungan antara objek . besarnya nilai Θ yang di inputkan adalah 0 < Θ < 1.

Metode ROCK menggunakaan informasi tentang *link* sebagai ukuran kemiripan antar objek. Jika terdapat objek pengamatan X_i, X_j dan X_k dimana X_i tetangga dari X_j dan X_j tetangga dari X_k maka diikatakan X_i memiliki *link* dengan X_k walaupun X_i bukan tetangga dari X_k. Cara untuk menghitung *link* untuk semua kemungkinan pasangan dari *n* objek dapat menggunakan matriks A . matriks A merupakan matriks berukuran *nxn* yang bernilai 1 jika X_i dan X_j dinyatakan mirip (tetangga) dan bernilai 0 jika X_i dan X_j tidak mirip(bukan tetangga). Jumlah *link* antar pasangan X_i dan X_j di peroleh dari hasil kali antara baris ke $X_i i$ dan kolom ke X_j dari matriks A. Jika *link* antara X_i dan X_j semakin besar maka semakin besar pula kemungkinan X_i dan X_j berada dalam satu kelompok yang sama.

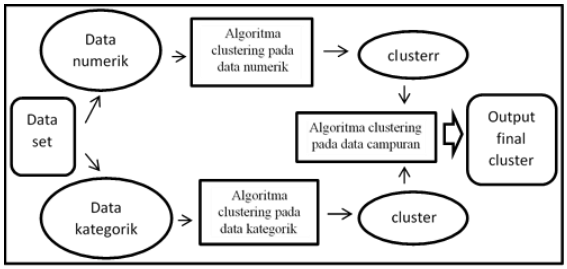
Algoritma *ROCK* yang didasarkan atas ukuran kebaikan (*goodness measure*) antar kelompok dengan rumusan pada persamaan *Goodness measure* adalah persamaan yang digunakan untuk menghitung jumlah *link* dibagi dengan kemungkinan *link* yang terbentuk berdasarkan ukuran kelompoknya (Tyagi & Sharma, 2012).

$$g\left(C_i, C_j\right) = \frac{li\ [C_i, C_j]}{(n_i, n_j)^{1+2f(\Theta)} - {n_i}^{1+2f(\Theta)} - {n_j}^{1+2f(\Theta)}}$$

Dengan $li\left[C_i, C_j\right] = \sum_{X_i \in C_i, X_j \in C_j} li\ \ (X_i, X_j)$ yang menyatakan jumlah *link* dari semua kemungkinan pasangan objek yang ada dalam C_i dan C_j serta n_i dengan n_j masing-masing menyatakan jumlah anggota dalam kelompok ke$-i$ dan j, sedangkan $f(\theta) = \frac{1-\theta}{1+\theta}$.

Pengelompokan Data Campuran

Salah satu medode yang bisa di gunakan untuk menyelesaikan masalah yang berkaitan dengan clustering data dengan tipe campuran (kategorik dan numerik) adalah metode ROCK . Pertama, data asli yang bertipe campuran dipisah menjadi dua yaitu data dengan tipe kategorik dan data dengan tipe numerik. Selanjutnya, kedua data tersebut diproses secara terpisah dengan menggunakan algoritma clustering yang sesuai dengan tipe masing-masing data. Terakhir, cluster-cluster yang dihasilkan oleh kedua algoritma digabungkan dan dipandang sebagai data baru dengan tipe kategorik, kemudian diproses dengan menggunakan algoritma clustering data kategorik untuk mendapatkan hasil akhir.Langkah dari pengelompokan data campuran ditunjukkan oleh Gambar 2.2 berikut:



Gambar 2.1 Langkah dari pengelompokan data campuran

Adapun langkah ensembel ROCK sebagai berikut :

1. Memisahkan data menjadi data kategorik dan data numerik
2. Melakukan clustering data numerik dengan menggunakan algoritma clustering data numerik dengan metode non hirarki.
3. Melakukan clustering data kategorik dengan menggunakan algoritma clustering data kategorik metode ROCK.
4. Menggabungkan output dari kedua algoritma tersebut menjadi data kategorik
5. Menggunakan kembali metode ROCK untuk melakukan clustering terhadap data hasil gabungan.

Kinerja Hasil *Clustering*

Kinerja Hasil *Clustering* Pengukuran kinerja hasil clustering merupakan langkah untuk mengetahui validitas suatu cluster. Cluster yang baik akan memiliki kehomogenan yang tinggi antar anggota dalam kelompok dan

keheterogenan yang tinggi antar kelompok (Hair, Black, Babin, & Anderson, 2010). Terdapat dua uji validitas yaitu validasi ukuran dan validasi metode.

- Validasi ukuran

Validasi ukuran yang digunakan dalam pemilihan jumlah cluster optimum pada variabel data berskala numerik adalah ukuran dunn index dan davie bouldin index. Dimana Dunn index adalah salah satu pengukuran validitas cluster yang diajukan oleh J.C.Dunn. Menurut Satato, Khotimah, & Muhammad (2015), validitas cluster berlandaskan pada fakta bahwa cluster yang terpisah pada umumnya memiliki jarak antar cluster yang besar dan jarak dalam cluster yang kecil. Dunn index tidak memiliki rentang nilai, nilai terbesar yang dihasilkan merupakan hasil ukuran terbaik terbaik (Dewanti, 2013).

$$D = min_{\,j=i+1..n_c} \ \ (min_{\,j=i+1..n_c} \frac{d\,[C_i,C_j]}{max_{\,k=1..n_c}\,(C_k)})$$

Sedangkan Davies bouldin index adalah salah satu metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Pengelompokan, kohesi dapat diartikan sebagai jumlah dari kedekatan data terhadap centroid dari cluster yang diikuti. Sedangkan separasi didasarkan pada jarak antar centroid dari clusternya. Semakin kecil nilai davies bouldin index maka semakin optimum jumlah cluster tersebut. Nilai Davies Bouldin Index (DBI) :

$$DBI = \frac{1}{K}\sum_{j=i}^{K} max \ \ (R_{i,j})$$

Dimana :

K = Jumlah cluster yang digunakan

$R_{i,j}$ = Jarak antara n cluster i dengan cluster j

- Validasi metode

Validasi metode yang di gunakan untuk menentukan metode pengelompokan data numerik terbaik dapat diketahui dari rasio nilai S_w dan S_B. Nilai perbandingan rasio S_w dan S_B ini juga di gunakan dalam memilih nilai *thresold* terbaik dari pengelompokan data kategorik dan data campuran.

Menurut Alvionita (2017), ukuran keragaman untuk data kategorik dikembangkan oleh Light dan Nargolin (1971), Okada (1999) serta Kader dan Perry (2007). Jika terdapat sebanyak n pengamatan dengan n_k merupakan jumlah pengamatan dengan kategori ke-k dimana k = 1,2,3,...K dan $\sum_{k=1}^{K} n_k = n$. Selanjutnya, n_k merupakan jumlah pengamatan dengan kategori ke-k dan kelompok ke-c , dimana c = 1,2,3,...C dengan C adalah jumlah yang terbentuk, sehingga $n_c = \sum_{k=1}^{K} n_k$ merupakan jumlah pengamatan pada kelompok ke-c $n_k = \sum_{c-1}^{K} n_k$ merupakan jumlah pengamatan pada kategori ke-k. Total jumlah pengamatan dapat di tuliskan menjadi

$$n = \sum_{c-1}^{K} n_c = \sum_{k=1}^{K} n_k = \sum_{c-1}^{K}\sum_{k=1}^{K} .$$

Jumlah kuadrat total atau SST pada sebuah peubah data kategorik dirumuskan sebagai berikut :

$$\text{SST} = \frac{n}{2} - \frac{1}{2n}\sum_{K=1}^{K} {n_k}^2$$

Rumus total jumlah kuadrat dalam kelompok atau SSW :

$$\text{SSW} = \sum_{c=1}^{C}\left(\frac{n_c}{2} - \frac{1}{2n_c}\sum_{k=1}^{K} {n_k}^2\right) = \frac{n}{2} - \frac{1}{2}\sum_{c=1}^{c}\frac{1}{n_c}\sum_{k}^{K} {n_k}^2$$

Rumus Jumlah kuadrat antar kelompok atau SSB :

$$\text{SSB} = \frac{1}{2}\sum_{c-1}^{c}\frac{1}{n_c}\sum_{k}^{K} {n_k}^2 \ \frac{1}{2n_c}\sum_{k}^{K} {n_k}^2$$

,

Rumus Mean of square (MST), mean of square (MSW), dan mean of square beetwen (MSB) sec ara bertutut-turut sebagai berikut:

$$\text{MST} = \frac{SST}{n-1}$$

$$\text{MSW} = \frac{SSW}{\text{n} - \text{C}}$$

$$\text{MSB} = \frac{SSB}{\text{C} - 1}$$

Rumus simpangan bakudalam kelompok(S_w) dan rumus simpangan baku antar kelompok (S_B) paada data kategorik adalah :

$$S_w = [MSW]^{1/2}$$

$$S_B = [MSB]^{1/2}$$

Sama halnya pada data numerik, untuk melihat kinerja suatu metode pengelompokan dapat di lihat dari nilai rasio antara (S_w) dan (S_B). Semakin kecil nilai rasio maka semakin baik juga kinerja suatu metode. kinerja metode yang baik antinya terdapat homogenitas maksimum dalam cluster dan heterogenitas yang maksimum antar *cluster.*

METODE PENELITIAN

Jenis Penelitian dan Sumber Data

Jenis penelitian yang akan digunakan adalah penelitian kuantitatif dengan menggunakan metode cluster ensemble ROCK. Data yang digunakan adalah data indikator sarana prasarana dan SDM SMA tahun 2020. Data tersebut merupakan data sekunder yang diperoleh dari Dinas Pendidikan Kabupaten Kediri.

Variabel Penelitian

Tabel 3.1. Variabel Numerik.

Variabel	Keterangan
X1	Rasio siswa/rombel
X2	Jumlah guru tetap
X3	Jumlah guru non- tetap
X4	Luas lahan
X5	Daya listrik

Tabel 3.2 Variabel Penelitian Berskala Kategori

Variabel	Keterangan
X6	Akreditasi SMA
X7	Status SMA
X8	Waktu penyelenggaraan

Langkah Penelitian

Langkah-langkah dalam penelitian ini yaitu :

1. Mengumpulkan data yang sesuai dengan penelitian.
2. Menyiapkan data yang di mulai dari membersihkan data, mengurangi data, dan memisahkan untuk data yang tidak sesuai.
3. Memisahkan data yang berskala numerik dan berskala kategorik.
4. melakukan pengelompokan pada variabel berskala numerik menggunakan metode non hirarki dengan jarak ecludien melalu K-Means dan K-Medoids.
5. Menghitung nilai dunn index dan davies bouldin index.
6. Melakukan langkah 4 dengan mengganti k = 3,4,dan 5.
7. Menentukan cluster terbaik berdasarkan nilai terbesar dunn index dan nilai terbesar davies bouldin index.
8. Menghitung nilai rasio dari S_w dan S_B untuk K-means dan K-Medoids dari cluster terpilih.
9. Membandingkan hasil rasio dari S_w dan S_B dan menentukam kelompok dengan melihat nilai rasio terkecil.
10. Melakukan pengelompokan pada variabel berskala kategorik menggunakan metode hirarki ROCK ***(****robust clustering using links).*

HASIL PENELITIAN dan PEMBAHASAN

Statistika Deskriptif

SMA di Kabupaten Kediri memiliki Rata-rata rasio siswa per rombel (x1) sebesar 27 siswa per rombel. Variabel jumlah guru tetap (x2) memiliki rata-rata 24 guru sedangkan rata-rata variabel jumlah guru non-tetap (x3) sebesar 5 guru. Luas lahan (x5) memiliki rata-rata 11.111 m^2 dan rata-rata daya listrik sebesar 38777 watt. Mayoritas SMA di kabupaten Kediri memiliki kerakteristik berakreditasi A (57%), berstatus negri (55%), dan waktu penyelenggaraan pada selama 6 hari dalam seminggu (18 sekolah), 5 hari dalam seminggu(25 sekolah) ataupun di lakukan pada siang hari selama 6 hari dalam seminngu (1 sekolah).

Pengelompokan Data Numerik dan Validasinya

Sebelum masuk tahab pengelompokan, data numerik memiliki rentang nilai yang sangat berbeda pada masing- masing atributnya atau dengan kata lain satuan setiap atribut berbeda sehingga perlu dilakukan standarisasi

Tabel 4.2. Standarisasi Data

Rasio siswa/rombel	Jumlah guru Tetap	Jumlah guru Non tetap	Luas Lahan	Daya Listrik
-154.992	-.67991	-.76116	143.988	-.71552
-119.089	-128.064	-.96171	-.81211	-.71552
.72822	-.98028	-.96171	-.53264	-.71552
-.30142	-.92020	.84320	-.90404	-.57114
-119.089	-128.064	-.76116	-.74051	-.71552

Pada pengelompokan data numerik di gunakan metode non hierarki dengan ukuran jarak euclidean. Metode non hierarki tersebut adalah K-mean dan K-medoid. Pengelompokan dilakukan dengan membentuk k=2, k=3, k=4 dan k=5. Selanjutnya hasil dari pengelompokan tersebut di bandingkan untuk mendapatkan kelompok yang paling optimum. Hasil kelompok optimum dapat di tentukan dengan melihat nilai terbesar dari dunn index dan nilai terkecil dari davies bouldin index. Berikut ini hasil nilai dunn index dan davies bouldin index pada metode K-mean.

Tabel 4.3 Hasil Nilai Dunn Index Dan Davies Bouldin Index Pada *K-Mean*

Jumlah Kelompok	Dunn Index	Davies Bouldin Index
2	0.3435909	1.039.839
3	0.3507582	1.006.778
4	0.4335883	0.8274516
5	0.456778	0.7902123

Dari Tabel 4.3 dapat di lihat bahwa kelompok paling optimum pada metode *K-mean* di hasilkan oleh jumlah kelompok k=5 dimana memiliki nilai dunn index paling besar yaitu 0.456778 dan davies bouldin index paling kecil yaitu 0.7902123.

Untuk hasil nilai dunn index dan bouldin index pada metode *K-medoid* dapat di lihat pada tabel dibawah ini.

Tabel 4.4 Hasil Nilai Dunn Index Dan Davies Bouldin Index Pada *K-Medoid*

Jumlah Kelompok	Dunn Index	Davies Bouldin Index
2	0.282104	1.126.107
3	0.3507582	1.167.939
4	**0.4335883**	**0.9020573**
5	0.4217366	0.926616

Tabel di atas menujukan bahwa dari dunn index kelompok paling optimum dengan metode *K-medoid* di hasilkan oleh jumlah kelompok k=4 dimana memiliki nilai dunn index paling besar yaitu 0.4335883 begitupula pada davies bouldin index kelompok paling optimum di hasilkan oleh k=4 dengan nilai davies bouldin index paling kecil yaitu 0.9020573.

Tabel 4.5 Hasil Perhitungan Nilai Rasio s_W /s_B

	Nilai s_W	Nilai s_B	Nilai rasio s_W /s_B
K-mean	81.400	239.351	0.3400873
K-medoid	88.977	239.435	0.3716138

Dari hasil nilai rasio s_W /s_B di peroleh metode terbaik yaitu metode K-mean dengan nilai s_W sebesar 81.400 dan nilai s_B sebesar 239.351 sehingga di peroleh nilai rasio s_W /s_B sebesar 0.3400873. Maka dapat disimpulkan bahwa pengelompokan data berskala numerik metode K-mean dengan 5 cluster merupakan clustering yang tepat untuk metode non hierarki. Berikut ini adalah tabel anggota cluster untuk metode K-mean.

Tabel 4.6. Hasil Pengelompokan Data Berskala Numerik

	Anggota Kelompok
Kelompok 1	SMAN 1 Grogol, SMAN 1 Kandat,SMAN 1 Mojo, SMAN 1 Papar, SMAN 1 Pare, SMAN 7 Kediri
Kelompok 2	SMA Bastren Darul Fatihin, SMAS Dharma Wanita Pare, SMAS Islam Al Wahid Kepung, SMAS Islam Gurah, SMAS Mardi Utomo, SMAS Diponegoro, SMAS K Petra Kediri, SMAS Kartanegara Kediri
Kelompok 3	SMAS Queen Al Falah, SMAN 1 Kediri
Kelompok 4	SMA Islam Khoirul Iman Kepung, SMA Islam Plus Hidayatut Thullab, SMAS Gajah Mada Puncu, SMAS Kadiri Kras, SMAS Kalijogo, SMAS Muhammadiyah 1 Pare, SMAS Psm Plemahan, SMA Katolik St. Augustinus, SMAS Al

	Anwar, SMAS Ar Risalah, SMAS Islam Ypa, SMAS Pawyatan Daha Kediri, SMAS Sultan Agung, SMAS Wahidiyah, SMAN 1 Plemahan, SMAN 1 Plosoklaten, SMAN 1 Puncu, SMAN 1 Purwoasri, SMAN 1 Wates, SMAN 2 Pare, SMAN 1 Kandangan, SMAN 1 Ngadiluwih, SMAN 8 Kediri
Kelompok 5	SMAN 2 Kediri, SMAN 3 Kediri, SMAN 4 Kediri, SMAN 6 Kediri

Pengelompokan Data Kategorik dan Validasinya

Data kategorik yang di gunakan pada penelitian ini meliputi 3 variabel yaitu variabel akreditasi, status SMA, dan waktu penyelenggara. Metode yang di gunakan untuk pengelompokannya adalah metode ROCK dengan nilai *threshlod* sebesar 0.05, 0.1, 0.12, 0.15, 0.17, 0.2, 0.22, 0.25, 0.27, dan 0.3. penggunaan nilai *threshlod* untuk mengatasi apabila Hasil yang di peroleh dalam pengelompokan menujukan semua objek pengamata berada dalam satu kelompok. Pemilihan kelompok optimum pada metode ROCK dilakukan dengan membandingkan nilai rasio s_W /s_B yang dihasilkan. Kelompok yang optimum adalah kelompok dengan kriteria nilai rasio s_W /s_B terkecil. Berikut merupakan hasil pengelompokan data kategori dengan k=2.

Tabel 4.7 Hasil Kinerja Pengelompokan Data Kategori

Threshold	Rasio s_W /s_B
0.05	**6.39×10^{-16}**
0.07	6.39×10^{-16}
0.10	6.39×10^{-16}
0.12	6.39×10^{-16}
0.15	6.39×10^{-16}
0.17	6.39×10^{-16}
0.20	6.39×10^{-16}
0.22	5.01×10^{-2}
0.25	5.01×10^{-2}
0.27	5.01×10^{-2}
0.30	5.01×10^{-2}

Berdasarkan tabel di atas terlihat bahwa kelompok yang di hasilkan nilai *threshlod* 0.05, 0.1, 0.12, 0.15, 0.17, dan 0.20 sama yaitu 6.39×10^{-16}. Begitupula pada nilai *threshlod* 0.22, 0.25, 0.27, dan 0.3 yang menujukan hasil pengelompokannya sama yaitu 5.01×10^{-2}. Karena rasio s_W /s_B terkecil terdapat pada nilai *threshlod* 0.05, 0.1, 0.12, 0.15, 0.17, dan 0.20 dengan hasil pengelompokan sama, maka nilai *threshlod* yang di gunakan dapat di pilih salah satu dari nilai *threshlod* di atas. Dari hasil pengelompokan di dapatkan kelompok 1 memiliki 23 anggota dan kelompok 2 memiliki 20 anggota. Anggota pada masing-masing kelompok dapat di lihat pada tabel di bawah ini.

Tabel 4.8. Anggota Hasil Pengelompokan Data Berskala Kategorik

Kelompok	Anggota Kelompok
Kelompok 1	SMA Bastren Darul Fatihin, SMA Islam Khoirul Iman Kepung, SMA Islam Plus Hidayatut Thullab, SMAS Dharma Wanita Pare, SMAS Gajah Mada Puncu, SMAS Islam Al Wahid Kepung, SMAS Islam Gurah, SMAS Kadiri Kras, SMAS Kalijogo, SMAS Mardi Utomo, SMAS Muhammadiyah 1 Pare, SMAS Psm Plemahan, SMAS Queen Al Falah, SMA Katolik St. Augustinus, SMAS Al Anwar, SMAS Ar Risalah, SMAS Diponegoro, SMAS Slam Ypa, SMAS K Petra Kediri, SMAS Kartanegara Kediri, SMAS Pawyatan Daha Kediri, SMAS Sultan Agung, SMAS Wahidiyah
Kelompok 2	SMAN 1 Grogol, SMAN 1 Kandangan, SMAN 1 Kandat, SMAN 1 Mojo, SMAN 1 Ngadiluwih, SMAN 1 Papar, SMAN 1 Pare, SMAN 1 Plemahan, SMAN 1 Plosoklaten, SMAN 1 Puncu, SMAN 1 Purwoasri, SMAN 1 Wates, SMAN 2 Pare, SMAN 1 Kediri, SMAN 2 Kediri, SMAN 3 Kediri, SMAN 4 Kediri, SMAN 6 Kediri, SMAN 7 Kediri, SMAN 8 Kediri

Pengelompokan Data Campuran dan Validasinya

Setelah mendapatkan hasil pegelompokan dari data berskala numerik dan kategorik maka selanjutnya hasil dari masing – masing pengelompokan tersebut di gabungkan dan dianggap sebagai data baru berskala

kategorik (tahab *ensemble*). Pengelompokan data gabungan ini menggunakan kembali metode ROCK yang sebelumnya di pakai untuk mengelompokan data kategorik.

Nilai *threshlod* yang di gunakan sebesar θ= 0,01, θ = 0.05, θ = 0,10, θ = 0,25, θ = 0,5, θ = 0,75 θ = 0,80 dan θ = 0,95. Pemilihan kelompok optimum pada metode ROCK dilakukan dengan membandingkan nilai rasio s_W /s_B yang dihasilkan. Kelompok yang optimum adalah kelompok dengan kriteria nilai rasio s_W /s_B terkecil. Berikut merupakan hasil pengelompokan data campuran dengan metode ROCK.

Tabel 4.9 Hasil Pengelompokan Ensembel ROCK

Threshold	rasio s_W /s_B
0.01	7.27 x 10^{-11}
0.05	7.27 x 10^{-11}
0.10	7.27 x 10^{-11}
0.25	7.27 x 10^{-11}
0.50	1.46 x 10^{-10}
0.75	1.46 x 10^{-10}
0.80	1.46 x 10^{-10}
0.95	1.46 x 10^{-10}

Sama halnya pada pengujian data berskala kategorik bahwa kelompok yang di hasilkan oleh nilai *threshlod* θ= 0.01, θ = 0.05, θ = 0.10, dan θ = 0.25 sama yaitu 7.27 x 10^{-11}. Begitupula pada nilai *threshlod* θ = 0.5, θ = 0.75 θ = 0.80 dan θ = 0.95 yang menujukan hasil pengelompokannya sama yaitu 1.46 x 10^{-10}. Karena rasio s_W /s_B terkecil terdapat pada nilai *threshlod* θ= 0.01, θ = 0.05, θ = 0.10, dan θ = 0.25 dengan hasil pengelompokan sama, maka nilai *threshlod* yang di gunakan dapat di pilih salah satu dari nilai *threshlod* di atas.

Tabel 4.10 Anggota Kelompok dengan Metode Ensemble ROCK Dengan Nilai = 0.01

Kelompok	Anggota Kelompok
Kelompok 1	SMA Bastren Darul Fatihin, SMA Islam Khoirul Iman Kepung, SMA Islam Plus Hidayatut Thullab, SMAS Dharma Wanita Pare,SMAS Gajah Mada Puncu, SMAS Islam Al Wahid Kepung, SMAS Islam Gurah, SMAS Kadiri Kras, SMAS Kalijogo, SMAS Mardi Utomo, SMAS Muhammadiyah 1 Pare, SMAS Psm Plemahan, SMAS Queen Al Falah, SMA Katolik St. Augustinus, SMAS Al Anwar, SMAS Ar Risalah, SMAS Diponegoro, SMAS Slam Ypa, SMAS K Petra Kediri, SMAS Kartanegara Kediri, SMAS Pawyatan Daha Kediri, SMAS Sultan Agung, SMAS Wahidiyah
Kelompok 2	l, SMAN 1 Kandangan, SMAN 1 Kandat, SMAN 1 Mojo, SMAN 1 Ngadiluwih, SMAN 1 Papar, SMAN 1 Pare, SMAN 1 Plemahan, SMAN 1 Plosoklaten, SMAN 1 Puncu, SMAN 1 Purwoasri, SMAN 1 Wates, SMAN 2 Pare, SMAN 1 Kediri, SMAN 2 Kediri, SMAN 3 Kediri, SMAN 4 Kediri, SMAN 6 Kediri, SMAN 7 Kediri, SMAN 8 Kediri

SIMPULAN dan SARAN

Simpulan

SMA di Kabupaten Kediri memiliki Rata-rata rasio siswa per rombel (x1) sebesar 27 siswa per rombel. Variabel jumlah guru tetap (x2) memiliki rata-rata 24 guru sedangkan rata-rata variabel jumlah guru non-tetap (x3) sebesar 5 guru. Luas lahan (x5) memiliki rata-rata 11.111 m^2 dan rata-rata daya listrik sebesar 38777 watt. Mayoritas SMA di kabupaten Kediri memiliki kerakteristik berakreditasi A (57%), berstatus negri (55%), dan waktu penyelenggaraan pada selama 6 hari dalam seminggu (18 sekolah), 5 hari dalam seminggu(25 sekolah) ataupun di lakukan pada siang hari selama 6 hari dalam seminngu (1 sekolah).

Hasil dari pengelompokan SMA di Kabupaten Kediri dengan metode ensembel Robust clustering using links (ROCK) di dapatkan jumlah kelompok sebanyak 2 kelompok. kelompok pertama beranggotakan 23 sekolah dan kelompok kedua beranggotakan 20 sekolah.

Pengelompokan SMA di Kabupaten Kediri menggunakan ensembel ROCK menghasilkan pengelompokan terbaik dengan rasio s_W dan s_B sebesar 7.27 x 10^{-11} pada *threshlod* 0.01.

Saran

Saran yang diberikan oleh peneliti untuk penelitian selanjutnya tentang pengelompokan ensembel ROCK adalah Perlu adanya penambahan variabel jumlah lulusan per kelas dalam variabel numerik agar hasil pengelompokan lebih maksimal dan Pengelompokan data numerik pada penelitian ini adalah dengan metode non hirarki menggunakan jarak euclidean dan metode yang digunakan yaitu K-mean dan K-medoid, sehingga masih terdapat beberapa metode pengelompokan data numerik dan ukuran jarak lain yang bisa di gunakan.

Daftar Pustaka

Rahayu, D. P., (2013), "Analisis Karakteristik Kelompok dengan Menggunakan Cluster Ensemble", Jurnal Matematika, Sains, dan Teknologi, Vol 14,No 1.

Sharma, S., (1996), Applied Multivariate Technique, John Wiley and Sons, Inc, New York.

Alvionita. (2017). Metode Ensembel ROCK dan SWFM untuk Pengelompokkan Data Campuran Numerik dan Kategorik pada Kasus Aksesi Jeruk. Institut Teknologi Sepuluh Nopember. Surabaya

Johnson RA, Wichern DW. 2002. Applied Multivariate Statistical Analysis. New Jersey: Prentice Hall.

Iam-on N, Garrett S. 2010. LinkCluE: A MATLAB Package for Link-Based Cluster Ensemble. Journal of Statistical Software. 36(9):1-36.

Strehl A, Gosh J. 2002. A Knowledge Reuse Framework for Combining Partitionings. The Journal of Machine Learning Research. 3(1):583-617.

Sumertajaya IM, Mattjik AA. 2011. Sidik Peubah Ganda dengan Menggunakan SAS. Bogor (ID): Departemen Statistika IPB.

He, Z., Xu, X., dan Deng, S., (2005b), "Clustering Mixed Numeric and Categorical Data: A Cluster Ensemble Approach". Department of Computer Science and Engineering, Harbin Institute of Technology

He, Z., Xu, X., dan Deng, S., (2005a), "A Cluster Ensemble Method For Clustering Categorical Data", Information Fusion, 6, hal 143-151.

Kader, D. G. dan Perry, M., (2007), "Variability for Categorical Variables", Journal of Statistics Education, Vol 15, No. 2.

Mulyono, M. S., (2006), Pengelompokan Data Kategorik dengan Algoritma ROCK,BUniversitas Airlangga, Surabaya.

Light, R. J., dan Margolin, B. H., (1971), "An Analysis of Variance for Categorical Data", Journal of American Statistical Association, Vol. 66, No.335.

Tyagi, A. dan Sharma, S., (2012), "Implementation of ROCK Clustering Algorithm for the Optimazation of Query Searching Time", International Journal on Computer Science and Engineering, Vol 4, No 05.

Guha, S., Rastogi, R., dan Shim, K., (2000), "ROCK: A Robust Clustering Algorithm for Categorical Attributes", Proceedings of the 15th International Conference on Data Engineering.

Hair, JR.J.F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis. United State of America: Prentice-Hall International,Inc.

Tyagi, A., & Sharma, S. (2012). Implementation of ROCK clustering algorithm for the optimazation of query searching time. International Journal on Computer Science and Engineering , Vol 4, No 05.

Tan, P., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. USA: Pearson Education,Inc .

Satato, B. D., Khotimah, B. K., & Muhammad, A. (2015). Pengelompokan Tingkat Kesehatan Masyarakat menggunakan Shelf Organizing Maps Dengan Cluster Validation Idb dan I-Dunn. Seminar Nasional Aplikasi Teknologi Informasi.

Han, J., & Kamber, M. (2001). Data Mining : Concepts and Techniques. USA:

Academic Press.