

**ANALISIS KLASTER DENGAN METODE ENSEMBLE ROCK
UNTUK DATA CAMPURAN
(Studi Kasus *Stunting* di Provinsi Jawa Barat)**

SKRIPSI

Diajukan untuk memenuhi sebagian dari syarat Memperoleh
Gelar Sarjana Matematika



Oleh
Refki Ardy Prasetya
1601718

**DEPARTEMEN PENDIDIKAN MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2020**

**ANALISIS KLASTER DENGAN METODE *ENSEMBLE ROCK*
UNTUK DATA CAMPURAN
(Studi Kasus *Stunting* di Provinsi Jawa Barat)**

Oleh
Refki Ardy Prasetya

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Matematika pada Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Refki Ardy Prasetya 2020
Universitas Pendidikan Indonesia
Januari 2020

Hak Cipta dilindungi undang-undang.
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difotokopi, atau lainnya tanpa ijin dari peneliti

LEMBAR PENGESAHAN

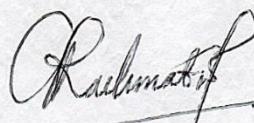
REFKI ARDY PRASETYA

**ANALISIS KLASTER DENGAN METODE ENSEMBLE ROCK UNTUK
DATA CAMPURAN**

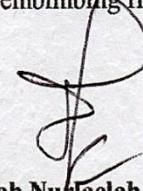
(Studi Kasus *Stunting* di Provinsi Jawa Barat)

disetujui dan disahkan oleh:

Pembimbing I,

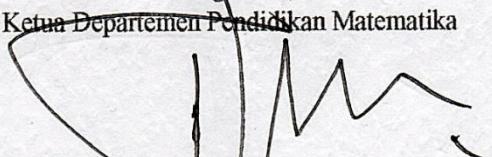

Hj. Dewi Rachmatin, S.Si., M.Si.
NIP. 196909291994122001

Pembimbing II,


Dr. Elah Nurlaelah, M.Si.
NIP. 196411231991032002

Mengetahui,

Ketua Departemen Pendidikan Matematika


Dr. H. Dadang Juandi, M.Si.
NIP. 196401171992021001

ANALISIS KLASTER DENGAN METODE ENSEMBLE ROCK
UNTUK DATA CAMPURAN
(Studi Kasus *Stunting* di Provinsi Jawa Barat)

ABSTRAK

Masalah *stunting* merupakan salah satu problematika gizi yang dihadapi di dunia, umumnya dialami oleh negara-negara berkembang dan miskin. Dalam pengukuran *stunting* memerlukan banyak variabel yang perlu dipertimbangkan untuk menentukan seorang balita mengalami *stunting* atau tidak. Oleh karena itu dapat dilakukan pengklasteran untuk data *stunting* tahun 2018 di Provinsi Jawa Barat dengan beberapa faktor yang dapat mempengaruhi *stunting*. Data *stunting* berupa data campuran (data kategorik dan data numerik). Salah satu permasalahan dalam analisis klaster adalah jika data yang akan dianalisis berupa data campuran. Pengklasteran untuk data campuran menggunakan metode *ensemble* ROCK (*Robust Clustering using linKs*) dengan cara menggabungkan *output* pengklasteran untuk data kategorik dan data numerik. Metode yang diterapkan untuk data kategorik adalah ROCK sedangkan untuk data numerik digunakan metode hirarki agglomeratif. Untuk menentukan hasil pengklasteran terbaik ditentukan berdasarkan kriteria rasio antara simpangan baku dalam (S_w) dan simpangan baku antar kelompok (S_b) yang terkecil. Berdasarkan 735 responden , metode *ensemble* ROCK dengan nilai $\theta = 0,10$ menghasilkan dua klaster dengan nilai rasio sebesar 0.0145 adalah hasil pengklasteran yang terbaik dengan karakteristik hasil pengklasteran metode *ensemble* ROCK yang diperoleh menjelaskan bahwa klaster satu lebih baik daripada klaster dua karena masih banyaknya anak yang mengalami obesitas dan simpulan PB/TB/U dengan kategori pendek yang banyak.

Kata Kunci : *Stunting*, Analisis klaster, Data Campuran, Data Kategorik, Data Numerik, Pengklasteran *Ensemble*, Rock.

**CLUSTER ANALYSIS WITH ENSEMBLE ROCK METHOD
FOR MIXED DATA**

(Case Study of Stunting In The Province of West Java)

ABSTRACT

Stunting is one of the nutritional problems faced in the world, generally experienced by developing and poor countries. In measuring stunting requires many variables that need to be considered to determine whether a toddler experiences stunting or not. Therefore, clustering can be done for 2018 stunting data in the Province of West Java with several factors that can affect stunting. Stunting data is in the form of mixed data (categorical data and numeric data). One of the problems in cluster analysis is if the data to be analyzed is mixed data. Clustering for mixed data uses the ROCK ensemble method (Robust Clustering using linKs) by combining clustering output for categorical data and nuclear data. The method applied for categorical data is ROCK while for numerical data the agglomerative hierarchy method is used. The best clustering results are determined based on the ratio criteria between the standard deviation in (S_w) and the smallest standard deviation between groups (S_b). Based on 735 respondents, the ROCK ensemble method with a value of $0, = 0.10$ produces two clusters with a ratio value of 0.0145 is the best clustering result with the characteristics of the ROCK ensemble clustering method obtained explains that cluster one is better than cluster two because there are still many children obese and PB / TB / U conclusions with many short categories.

Keywords: *Stunting, Cluster analysis, Mixed Data, Categorical Data, Numerical Data, Clustering Ensemble, Rock.*

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN KEASLIAN ISI SKRIPSI.....	ii
KATA PENGANTAR	iii
UCAPAN TERIMA KASIH.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.5.1 Manfaat Teoritis.....	5
1.5.2 Manfaat Praktis	5
BAB II KAJIAN TEORI.....	6
2.1 Stunting	6
2.2 Data Mining.....	7
2.3 Data Multivariat	9
2.4 Analisis Multivariat	10
2.5 Data Numerik dan Data Kategorik	10
2.6 Analisis Klaster	11
2.6.1 Ukuran Ketidakserupaan.....	12
2.6.2 Ukuran Keserupaan	13
BAB III METODOLOGI PENELITIAN.....	15
3.1 Sumber Data	15
3.2 Analisis Data	15

3.2.1	Pengklasteran Data Kategorik.....	15
3.2.2	Pengklasteran Data Numerik.....	17
3.2.3	Pengklasteran Data Campuran	20
3.2.4	Evaluasi Hasil Pengklasteran	21
3.3	Teknik Analisis Data	24
BAB IV HASIL DAN PEMBAHASAN		27
4.1	Data	27
4.1.1	Pemisahan Data.....	27
4.1.2	Transformasi Data.....	39
4.2	Pengklasteran untuk Data Numerik.....	39
4.3	Pengklasteran untuk Data Kategorik	42
4.4	Pengklasteran untuk Data Campuran	45
BAB V KESIMPULAN DAN SARAN.....		51
5.1	Kesimpulan.....	51
5.2	Saran	51
DAFTAR PUSTAKA		52

DAFTAR TABEL

Tabel 2.0.1. Data Multivariat	10
Tabel 4.1. Contoh Data Numerik	27
Tabel 4. 2. Contoh Data Kategorik	28
Tabel 4. 3. Statistik Deskriptif Untuk Peubah Numerik	29
Tabel 4. 4. Statistik Deskriptif Untuk Peubah Kategorik jenis kelamin	30
Tabel 4. 5. Statistik Deskriptif Untuk Peubah Kategorik Status Gizi.....	30
Tabel 4. 6. Statistik Deskriptif Untuk Peubah Kategorik Kelengkapan Imunisasi.....	31
Tabel 4. 7. Statistik Deskriptif Untuk Peubah Kategorik Jadwal Imunisasi.....	31
Tabel 4. 8. Statistik Deskriptif Untuk Peubah Kategorik asupan makanan	31
Tabel 4. 9. Statistik Deskriptif Untuk Peubah Kategorik Asupan	32
Tabel 4. 10. Statistik Deskriptif Untuk Peubah Kategorik Pernah dirawat	32
Tabel 4. 11. Statistik Deskriptif Untuk Peubah Kategorik Riwayat	33
Tabel 4. 12.Statistik Deskriptif Untuk Peubah Kategorik Penyulit Persalinan.....	33
Tabel 4. 13. Statistik Deskriptif Untuk Peubah Kategorik Pelayanan Persalinan	33
Tabel 4. 14. Statistik Deskriptif Untuk Peubah Kategorik Proses Persalinan	34
Tabel 4. 15. Statistik Deskriptif Untuk Peubah Kategorik Penyakit infeksi ibu.....	34
Tabel 4. 16.Statistik Deskriptif Untuk Peubah Kategorik Penyakit Kronis Ibu	34
Tabel 4. 17. Statistik Deskriptif Untuk Peubah Kategorik.....	35
Tabel 4. 18. Statistik Deskriptif Untuk Peubah Kategorik Pekerjaan Ibu	35
Tabel 4. 19. Statistik Deskriptif Untuk Peubah Kategorik Pekerjaan Suami.....	36
Tabel 4. 20.Statistik Deskriptif Untuk Peubah Kategorik Pendapatan	36
Tabel 4. 21.Statistik Deskriptif Untuk Peubah Kategorik Pengeluaran.....	37
Tabel 4. 22. Statistik Deskriptif Untuk Peubah Kategorik Jaminan Kesehatan.....	37
Tabel 4. 23. Statistik Deskriptif Untuk Peubah Kategorik Akses.....	37
Tabel 4. 24. Statistik Deskriptif Untuk Peubah Kategorik Pengasuh Balita.....	38
Tabel 4. 25. Statistik Deskriptif Untuk Peubah Kategorik Simpulan PB/TB/U	38
Tabel 4. 26.Contoh Hasil Standarisasi Data Numerik	39
Tabel 4. 27.Hasil Nilai Indeks Dunn Metode Hirarki Agglomeratif	40
Tabel 4. 28. Hasil Nilai Rasio Metode Hirarki Agglomeratif.....	41
Tabel 4. 29. Anggota Klaster Metode Complete Linkage	41

Tabel 4. 30. Nilai Rasio Hasil Pengklasteran Metode ROCK	44
Tabel 4. 31. Hasil Pengklasteran Metode ROCK dengan nilai $\theta = 0,15$	44
Tabel 4. 32.Nilai Rasio Hasil Pengklasteran Metode Ensemble ROCK.....	45
Tabel 4. 33. Anggota Pengklasteran Metode Ensemble ROCK	46
Tabel 4. 34.Krakteristik Peubah Numerik Metode Ensemble ROCK	46
Tabel 4. 35.Krakteristik Peubah Kategorik Metode Ensemble ROCK.....	47

DAFTAR GAMBAR

Gambar 2.1. Proses KDD (Hermawati, 2013)	8
Gambar 3.1. Langkah dari Algoritma CEBMDC	21

DAFTAR LAMPIRAN

Lampiran 1. Penjelasan Variabel.....	57
Lampiran 2. Data Kategorik	58
Lampiran 3. Data Numerik.....	71
Lampiran 4.Syntax Metode ROCK Untuk Data Kategorik.....	76
Lampiran 5. Hasil Output Metode ROCK Untuk Data Kategorik	79
Lampiran 6. Syntax Metode Hirarki Agglomeratif Untuk Data Numerik	86
Lampiran 7. Syntax Indeks Dunn Metode Hirarki Agglomeratif Untuk Data Numerik.....	102
Lampiran 8. Syntax Rasio Sw dan Sb Metode Hirarki Agglomeratif Untuk Data Numerik	103
Lampiran 9. Hasil Output Metode Complete Linkage (K=5) Untuk Data Numerik.....	106
Lampiran 10. Hasil Output Standarisasi Untuk Data Numerik	111
Lampiran 11. Syntax Metode Ensemble ROCK Untuk Data Campuran	118
Lampiran 12. Hasil Output Metode Ensemble ROCK Untuk Data Campuran.....	120

DAFTAR PUSTAKA

- Aderberg, M. (1973). *Cluster Analysis for Application*. New York: Academic Press.
- Agusta, Y. (2007). K-Means–Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, Vol. 3, 47-60.
- Alvionita. (2017). Metode ensemble ROCK dan SWFM untuk pengelompokan data campuran numerik dan kategori pada kasus aksesi jeruki [Thesis.]. Surabaya: Institut Teknologi Sepuluh November.
- Andeson, T. W., & Sclove, S. L. (1974). *Introductory Statistical Analysis*. Boston: Houghton Mifflin.
- BKKBN. (2018). *Analisis Situasi Potensi Program Ketahanan Keluarga Dalam Mencegah Stunting di Wilayah Provinsi Jawa Barat*. Bandung.
- Bolshakova, N., & Azuaje, F. (2001). Improving Expression Data Mining through Cluster Validity. *Departement of Computer Science. Ireland: TrinityCollege Dublin*.
- Davies, P. B. (2004). *Database System*. New York: Palgrave Macmillan.
- Dewangan, R. ,,, Sharma, L. K., & Akasapu, A. K. (2010). *Fuzzy clustering technique for numerical and categorical dataset*. International Journal on Computer Science and Engineering.
- Dewanti. (2013). Perbandingan Metode Cluster validity pada jenis data numerik dan kategori [Skripsi]. Bogor: Institut Pertanian Bogor.
- Dutta, M., Mahanta, A. K., & Arun, K. P. (2005). “QROCK : A Quick Version of the ROCK Algorithm for Clustering of Categorical Data”. *Proceedings of the 15 IEEE International Conference on Data Engineering*.
- Ghozali, I. (2005). *Aplikasi Analisis Multivariate dengan SPSS*. Semarang: Badan Penerbit UNDIP.
- Gudono. (2011). *Analisis Data Multivariat Edisi Pertama*. Yogyakarta: BPFE.
- Guha, S., Rastogi, R., & Shim, K. (2000). “ROCK: A Robust Clustering Algorithm for Categorical Attributes”. *Proceedings of the 15th International Conference on Data Engineering*.

- Hair, J. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. United State of America: Prentice-Hall Internatioan,Inc.
- Han, J., & Kamber, M. (2001). *Data Mining : Concepts and Techniques*. USA: Academic Press.
- Hariyadi, D., & Ekayanti, I. (2011). Analisis pengaruh perilaku keluarga sadar gizi terhadap stunting di Propinsi Kalimantan Barat. *Teknologi dan Kejuruan*, 34: 1,:71-80.
- Hee, Z., Xu, X. I., & Deng, S. (2005). Clustering mixed numeric and categorical data: A cluster ensemble approach. *China: Harbin Institute of technology*.
- Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi Offset.
- Hofmaan, D., Sawaya , A., Verreschi, I., Tucker , K., & Robert, S. (2000). Studies of metabolic rate and fat oxidation in shantytown children from São Paulo, Brazil. *Am J Clin Nutrition*, 72: 702-708.
- Indonesia, U. (2013). Ringkasan Kajian Gizi Ibu dan Anak. Akses www.unicef.org Tanggal 22 November 2019.
- Jackson, A., & Calder, P. ., (2004). Handbook Of Nutrition And Immunity (Severe Undernutrition and Immunity). *M. Eric Gershwin, M. E. Nestel, P. & Keen*, (p. 77). Humana Press.
- Johnson, R., & Whincern, D. (2007). *Applied multivariate statistical Analysis Sixth Edition*. New Jersey: Prentice Hall.
- Kader, D. G., & Perry, M. (2007). “Variability for Categorical Variables”. *Journal of Statistics Education*, Vol 15, No. 2.
- Light, R. J., & Margolin, B. H. (1971). “An Analysis of Variance for Categorical Data”. *Journal of American Statistical Association*, Vol. 66, No.335.
- Mann, J., & Truswell, A. S. (2002). *Essensial of Human Nutrition* (p. 65). Oxford University Press.
- Mann, J., & Truswell, A. S. (2009). Gizi Kesehatan Masyarakat, Gizi dan Perkebanginan Anak, Penerbit Buku Kedokteran EGC. *Terjemahan Public Health Nutrition*, Editor. *Gibney, M. J, Margetss, B.M.,Kearney, J.M. & Arab, L Blackwell* (p. 65). Publishing Ltd,Oxford.

- Metisen, B. M., & Sari, H. L. (2015). Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokkan Penjualan Produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110-118.
- Mucha, N. (2012). Implementing Nutrition-Sensitive Development. *Reaching Consensus.briefing paper*, Akses:www.bread.org/institute/papers/nutrition-sensitive-interventions.pdf tanggal 26 November 2019.
- Okada, T. (1999). "Sum of Square Decomposition for Categorical Data". *Kwansei Gakuin Studies in Computer Science, Vol 14*.
- Padmadas, S., Hutter, I., & Willekens, F. (2002). Weaning Initiation patterns and subsequents linear rowth progression among children aged 24 years in India. *International Journal of Epidemiology*, 31: 855-63.
- Picauly, I., & Magdalena, S. (2013). Analisis determinan dan Pengaruh Stunting Terhadap Prestasi Belajar Anak Sekolah di Kupang dan Sumba Timur , NTT. *Jurnal Gizi dan Pangan*, 8 (1) : 55-62.
- Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative Dalam Analisis Klaster Pada Data Tingkat Polusi Udara. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, Vol 3, No.2.,
- Rahayu, D. P. (2013). "Analisis Karakteristik Kelompok dengan Menggunakan Cluster Ensemble". *Jurnal Matematika, Sains, dan Teknologi*, Vol 14,No 1.
- Rencher, A. C. (2002). *Method of Multivariate Analysis. Second Edition*. New York: Jhon Wiley & Sons, Inc.
- Riskesdas. (2013, November 22). *Penyajian Pokok-pokok Hasil Riset Kesehatan Dasar 2013. Badan Penelitian dan Pengembangan Kesehatan*. Akses www.litbang.depkes.go.id tanggal 26 November 2019: Kementrian Kesehatan RI. Retrieved from Badan Penelitian dan Pengembangan Kesehatan: www.litbang.depkes.go.id
- Safri, M., & Aidi, M. N. (2011). Pengelompokan Kelas Kekuatan pada Beberapa Jenis Kayu Berdasarkan Sifat Mekanik dengan Menggunakan Analisis Gerombol. *Jurnal Paradigma Ekonomika*, 1(4), 24-36.

- Sarwono, J. (2013). *Statistik Multivariat Aplikasi untuk Riset Operasi*. Yogyakarta: C.V Andi Offset.
- Satato, B. D., Khotimah, B. K., & Muhammad, A. (2015). Pengelompokan Tingkat Kesehatan Masyarakat menggunakan Shelf Organizing MapsDengan Cluster Validation Idb dan I-Dunn. *Seminar Nasional AplikasiTeknologi Informasi*.
- Sharma, S. (1996). *Applied Multivariate Technique*. New York: John Wiley and Sons, Inc.
- Simamora. (2005). *Analisis Multivariat Pemasaran*. Jakarta: Gramedia Pustaka Utama.
- Sitepu, R., & Gultom, B. (2011). Clustering Analysis for Air Pollution Level on Industrial Sector in South Sumatera. *Jurnal Penelitian Sains*, 14(3), 11-17.
- 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.
- Unicef. (1990). *Strategy for improved nutrition of children and women in developing countries*. New York.
- Unicef. (2012). *Ringkasan Kajian Gizi*. Jakarta: Pusat Promosi Kesehatan - Kementerian Kesehatan RI.
- Unicef. (2013). *Improving Child Nutrition The achievable imperative for global progress*. Diakses:www.unicef.org/media/files/nutrition_report_2013.pdf tanggal 1 Desember 2019.