

SKRIPSI

**Clustering Tingkat Kesehatan Balita Menggunakan Metode Fuzzy
Subtractive Clustering Pada Kota Makassar**

**Disusun dan diajukan oleh
FABYOLA LARASATI MASYITA
D42116510**



**DEPARTEMEN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
MAKASSAR
2021**

LEMBAR PENGESAHAN SKRIPSI
CLUSTERING TINGKAT KESEHATAN BALITA MENGGUNAKAN
METODE FUZZY SUBTRACTIVE CLUSTERING PADA KOTA
MAKASSAR

Disusun dan diajukan oleh
FABYOLA LARASATI MASYITA
D42116510

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 22 November 2021 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

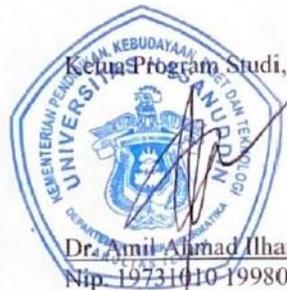
Pembimbing Pendamping,



Dr. Ir. Ingrid Nurtanio, MT
Nip. 196108131988112001



Elly Wami, ST., MT.
Nip. 198202162008122001

Ketua Program Studi,

Dr. Amil Ahmad Ilham, S.T., M.IT.
Nip. 197310101998021002

PERNYATAAN KEASLIAN

Yang bertanda tangan di bahwa ini :

Name : Fabyola Larasati Masyita
NIM : D42116510
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa hasil karya tulisan saya berjudul :

Clustering Tingkat Kesehatan Balita Menggunakan Metode Fuzzy Subtractive Clustering Pada Kota Makassar

Adalah karya tulisan saya sendiri. Bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi ini yang saya tulis ini benar benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain. Maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 19 November 2021

Yang menyatakan,



FABYOLA LARASATI MASYITA

ABSTRAK

Tingkat kesehatan balita mencerminkan tingkat kesehatan bangsa, sebab balita sebagai generasi penerus bangsa memiliki kemampuan yang dapat dikembangkan dalam meneruskan pembangunan bangsa. Salah satu teknik yang dapat digunakan untuk mengetahui tingkat kesehatan balita dengan menggunakan teknik *data mining*. Tujuan utama dalam penelitian ini adalah agar mengetahui pengelompokan tingkat kesehatan balita untuk setiap kelurahan pada Kota Makassar.

Pada penelitian ini, algoritma *fuzzy subtractive clustering* diterapkan untuk mengetahui pengelompokan tiap lingkungan (kelurahan) yang ada di Kota Makassar menjadi beberapa klaster. Klaster yang terbentuk ditinjau berdasarkan kemiripan karakteristik lingkungan dari indikator kesehatan balita, seperti angka kematian balita, presentase balita gizi kurang, presentase balita gizi kurus, presentase balita pendek, BBLR (berat badan lahir rendah), presentase balita menderita pneumonia, dan diare dengan menggunakan data yang berasal dari Dinas Kesehatan Kota Makassar tahun 2018.

Hasil penelitian menggunakan algoritma *fuzzy subtractive clustering* dengan $r = 0.5$ pada *dataset* balita di setiap kelurahan yang berasal dari Dinas Kesehatan Kota Makassar tahun 2018 baik dengan pengerjaan tanpa *library* maupun menggunakan *library* pada Matlab memberikan hasil klaster yang sama, yaitu 8 klaster. Sebanyak 16 kelurahan (34.8%) berada pada klaster 1 (persentase penderita gizi kurang tertinggi ketiga). Sebanyak 4 kelurahan (8.7%) berada pada klaster 2 (angka kematian tertinggi kedua). 5 kelurahan (10.9%) berada pada klaster 3 (persentase penderita pneumonia tertinggi kedua). 6 kelurahan (13%) berada pada klaster 4 (persentase penderita gizi kurang, kurus, dan pendek tertinggi). 4 kelurahan (8.7%) berada pada klaster 5 (persentase penderita pneumonia tertinggi). 3 kelurahan (6.5%) berada pada klaster 6 (persentase penderita BBLR tertinggi). 3 kelurahan (6.5%) berada pada klaster 7 (persentase penderita diare tertinggi). 5 kelurahan (10.9%) berada pada klaster 8 (angka kematian tertinggi). Adapun nilai *Silhouette Coefficient* yang didapatkan 0.2417, nilai *SSE* 7.6842, dan nilai *MAE* 0.5870. Dengan diketahuinya hasil klaster pada setiap kelurahan, maka dapat dijadikan oleh pemerintah sebagai bahan kebijakan agar mampu meningkatkan pelayanan kesehatan di setiap kelurahan berdasarkan masalah yang dihadapi, sehingga dapat menekan angka kematian dan penyakit serta status gizi balita dapat ditingkatkan.

Kata Kunci: *data mining, clustering, algoritma fuzzy subtractive clustering, tingkat kesehatan*

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji dan syukur senantiasa penulis panjatkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir ini yang berjudul “**Clustering Tingkat Kesehatan Balita Menggunakan Metode Fuzzy Subtractive Clustering Pada Kota Makassar**” sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Dalam proses pembuatan laporan akhir ini tidak lepas dari bantuan, bimbingan serta dukungan dari berbagai pihak, dari masa perkuliahan sampai penyusunan Tugas Akhir. Oleh karena itu dengan segala kerendahan hati, penulis mengucapkan rasa terima kasih sebesar-besarnya kepada:

1. Tuhan Yang Maha Esa yang melalui berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan tugas akhir ini.
2. Kedua orang tua penulis, Ibu Yuliana Indar dan Bapak Ichsyah Syahrir yang selalu memberikan motivasi, dukungan, doa, semangat dan kasih sayang serta selalu sabar dalam mendidik penulis sejak kecil.
3. Ibu Dr. Ir. Ingrid Nurtanio, M.T., selaku dosen pembimbing I dan Ibu Elly Warni, S.T., M.T., selaku dosen pembimbing II yang selalu menyediakan waktu, tenaga dan pikirannya yang luar biasa untuk mengarahkan penulis dalam penyusunan Tugas Akhir ini.
4. Bapak Dr. Amil Ahmad Ilham, S.T., M.I.T., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas bimbingannya selama masa perkuliahan penulis.
5. Segenap Dosen dan Staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak membantu penulis selama masa perkuliahan.

6. Keluarga besar penulis, Puang Indar, Puang Emma, Puang Imran, Jihan, Jira, dan Syifa yang selalu mendoakan serta menyemangati penulis, serta Moly yang selalu menemani penulis.
7. Sahabat-sahabat penulis semasa kuliah, Rya Dita Purnama, Nishrina Nurul Amirah, Afifah Ilham, Ismayanti, dan Patricia Vhiola Palada yang selalu menemani serta memberikan bantuan penulis semasa perkuliahan.
8. Sahabat NINUNINU, Enun, Rani, Haifa, Cia, Lulu, dan Agista yang selalu menjadi tempat berbagi cerita dan memberikan perhatian dan semangat kepada penulis.
9. Teman-teman IGNITER16 atas dukungan dan semangat yang telah diberikan.
10. Orang-orang berpengaruh lainnya yang tidak sempat penulis sebutkan satu persatu yang telah memberikan dukungan, semangat dan doanya selama penyusunan Tugas Akhir ini.

Akhirnya dengan segala kerendahan hati, penulis menyadari masih terdapat kekurangan dalam penyusunan laporan ini baik isi maupun cara penyajian. Oleh karena itu penyusun mengharapkan adanya saran dan kritik yang bersifat membangun demi kesempurnaan laporan ini.

Makasar, November 2020

Penulis

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan masalah.....	2
1.3. Tujuan Penelitian.....	2
1.4. Batasan Masalah.....	2
1.5. Manfaat.....	3
1.6. Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. <i>Data Mining</i>	5
2.2. <i>Supervised dan Unsupervised Learning</i>	7
2.3. <i>Clustering</i>	10
2.4. <i>Fuzzy Subtractive Clustering</i>	12
2.5. Performa Hasil <i>Clustering</i>	16
2.5.1. <i>Silhouette Coefficient</i>	17
2.5.2. <i>Sum of Squared Error (SSE)</i>	18
2.5.3. <i>Mean Absolute Error (MAE)</i>	19
2.6. Penelitian Terkait.....	19

BAB III METODOLOGI PENELITIAN.....	21
3.1. Kajian Pustaka.....	22
3.2. Pengumpulan Data.....	22
3.3. Perancangan Sistem.....	22
3.3.1. <i>Data Collection</i>	22
3.3.2. <i>Data Cleaning / Integration</i>	23
3.3.3. <i>Data Selection</i>	24
3.3.4. <i>Data Transformation</i>	25
3.3.5. <i>Data Mining</i>	25
3.3.6. <i>Pattern Evaluation</i>	25
3.3.7. <i>Knowledge Presentation</i>	25
3.4. Implementasi Sistem.....	26
3.5. Pengujian dan Evaluasi Sistem.....	28
3.6. Pembahasan.....	29
BAB IV HASIL DAN PEMBAHASAN.....	30
4.1. Konsep Penerapan Algoritma <i>Fuzzy Subtractive Clustering</i>	30
4.1.1. Contoh <i>Dataset</i>	30
4.1.2. Menentukan Nilai.....	31
4.1.3. Normalisasi Data.....	31
4.1.4. Menentukan potensi awal tiap-tiap data.....	32
4.1.5. Mencari titik dengan potensi tertinggi.....	33
4.1.6. Mengembalikan pusat kluster ke bentuk semula.....	36
4.1.7. Nilai Sigma Kluster.....	37

4.1.8.	Menghitung Derajat Keanggotaan	38
4.1.9.	Evaluasi Kluster	41
4.2.	Pengelompokan Tingkat Kesehatan Balita.....	43
4.2.1.	Penerapan Algoritma <i>Fuzzy Subtractive Clustering</i> dalam Mengelompokkan Tingkat Kesehatan Balita dengan Matlab	43
4.2.1.1.	Memasukkan Data	43
4.2.1.2.	Inisialisasi Parameter	44
4.2.1.3.	Normalisasi Data.....	44
4.2.1.4.	Penentuan Potensi Awal Tiap Data	45
4.2.1.5.	Pencarian Titik Potensi Tertinggi	46
4.2.1.6.	Penentuan Titik Pusat Kluster.....	46
4.2.1.7.	Mengembalikan Pusat Kluster Ke Bentuk Semula	48
4.2.1.8.	Menghitung Nilai Sigma Kluster	49
4.2.1.9.	Menghitung Derajat Keanggotaan.....	49
4.2.1.10.	Menghitung Nilai <i>Silhouette</i>	51
4.2.1.11.	Menghitung Nilai <i>SSE</i>	53
4.2.1.12.	Menghitung Nilai <i>MAE</i>	54
4.2.2.	Penerapan Algoritma <i>Fuzzy Subtractive Clustering</i> dalam Mengelompokkan Tingkat Kesehatan Balita dengan <i>Library</i> Matlab.....	56
4.2.2.1.	<i>Library</i>	56
4.2.2.2.	Penentuan Titik Pusat Kluster.....	56
4.2.3.	Pembahasan.....	57
4.2.3.1.	Statistik	57

4.2.3.2.	Klaster 1	60
4.2.3.3.	Klaster 2	61
4.2.3.4.	Klaster 3	61
4.2.3.5.	Klaster 4	62
4.2.3.6.	Klaster 5	63
4.2.3.7.	Klaster 6	64
4.2.3.8.	Klaster 7	64
4.2.3.9.	Klaster 8	65
4.2.3.10.	Ringkasan Hasil <i>Clustering</i>	65
4.3.	Pengembangan Aplikasi Menggunakan GUI Matlab	67
BAB V PENUTUP		69
5.1.	Kesimpulan	69
5.2.	Saran	69
DAFTAR PUSTAKA		71
LAMPIRAN		73

DAFTAR GAMBAR

Gambar 2.1. Tahapan <i>data mining</i> sebagai proses penemuan pengetahuan	5
Gambar 2.2. Model <i>Supervised Learning</i>	8
Gambar 2.3. Model <i>Unsupervised Learning</i>	10
Gambar 3.1. Diagram Tahapan Penelitian	21
Gambar 3.2. Sampel Data Status Gizi Balita	23
Gambar 3.3. Data Kesehatan Balita	24
Gambar 3.4. Data Kesehatan Balita	24
Gambar 3.5. <i>Flowchart</i> Algoritma <i>Fuzzy Subtractive Clustering</i>	27
Gambar 4.1. Kode Matlab dalam <i>input</i> data	43
Gambar 4.2. Kode Matlab dalam inisialisasi parameter	44
Gambar 4.3. Kode Matlab untuk normalisasi data.....	44
Gambar 4.4. Hasil normalisasi data	45
Gambar 4.5. Kode Matlab dalam menentukan potensi awal tiap data	45
Gambar 4.6. Kode Matlab dalam menentukan titik potensi tertinggi	46
Gambar 4.7. Kode Matlab untuk mencari pusat kluster	47
Gambar 4.8. Tampilan nilai rasio setiap iterasi pencarian titik pusat kluster	48
Gambar 4.9. Kode Matlab untuk mengembalikan nilai pusat kluster	48
Gambar 4.10. Hasil pengembalian nilai pusat kluster.....	49
Gambar 4.11. Kode Matlab untuk menghitung nilai sigma kluster	49
Gambar 4.12. Kode Matlab untuk menghitung nilai sigma kluster	49
Gambar 4.13. Kode Matlab untuk menghitung derajat keanggotaan setiap data.....	50
Gambar 4.14. Kode Matlab untuk menghitung derajat keanggotaan setiap data.....	50
Gambar 4.15. Kode Matlab untuk penentuan kluster dari setiap data	50
Gambar 4.16. Tampilan Matlab untuk penentuan kluster dari setiap data	51
Gambar 4.17. Kode Matlab untuk menghitung nilai <i>Silhouette</i> dari setiap data	51
Gambar 4.18. Tampilan dari nilai <i>Silhouette</i> dari setiap data	52
Gambar 4.19. Grafik Hasil perhitungan <i>Silhouette Coefficient</i>	53

Gambar 4.20. Kode Matlab untuk fungsi penghitung nilai <i>SSE</i>	53
Gambar 4.21. Kode Matlab untuk memanggil fungsi <i>SSE</i>	53
Gambar 4.22. Kode Matlab untuk fungsi penghitung nilai <i>MAE</i>	54
Gambar 4.23. Kode Matlab untuk memanggil fungsi <i>MAE</i>	54
Gambar 4.24. Kode Matlab dalam menggunakan fungsi <i>subclust</i> untuk mencari titik pusat kluster.....	56
Gambar 4.25. Tampilan <i>output</i> dalam setiap iterasi pencarian titik pusat klaster	57
Gambar 4.26. Kode Matlab untuk fungsi perhitungan nilai statistik klaster	58
Gambar 4.27. Kode Matlab untuk perhitungan nilai statistik klaster	58
Gambar 4.28. Nilai statistik klaster 1	58
Gambar 4.29. Persentase jumlah anggota setiap klaster	59
Gambar 4.30. Daftar kelurahan yang berada pada klaster 1	60
Gambar 4.31. Daftar kelurahan yang berada pada klaster 2	61
Gambar 4.32. Daftar kelurahan yang berada pada klaster 3	61
Gambar 4.33. Daftar kelurahan yang berada pada klaster 4	62
Gambar 4.34. Daftar kelurahan yang berada pada klaster 5	63
Gambar 4.35. Daftar kelurahan yang berada pada klaster 6	64
Gambar 4.36. Daftar kelurahan yang berada pada klaster 7	64
Gambar 4.37. Daftar kelurahan yang berada pada klaster 5	65
Gambar 4.38. Tampilan Awal Aplikasi	67
Gambar 4.39. Tampilan Hasil Klaster Aplikasi	67
Gambar 4.40. Tampilan <i>Output</i> dari 3 Variabel Visualisasi <i>Cluster</i>	68

DAFTAR TABEL

Tabel 4.1. Contoh <i>dataset</i> berisi data balita dari puskesmas pada setiap kelurahan...	30
Tabel 4.2. Hasil normalisasi data	32
Tabel 4.3. Hasil mengembalikan pusat kluster ke bentuk semula.....	37
Tabel 4.4. Hasil nilai sigma kluster	38
Tabel 4.5. Hasil perhitungan derajat keanggotaan	40
Tabel 4.6. Hasil anggota berdasarkan perhitungan derajat keanggotaan	40
Tabel 4.7. Hasil perhitungan <i>Silhouette Coefficient</i> , <i>MAE</i> , dan <i>SSE</i>	52
Tabel 4.8. Hasil perhitungan <i>SSE</i>	54
Tabel 4.9. Hasil perhitungan <i>MAE</i>	55
Tabel 4.10. <i>Silhouette Coefficient</i> , <i>MAE</i> , dan <i>SSE</i>	55
Tabel 4.11. Nilai statistik setiap kluster	59
Tabel 4.12. <i>Clustering</i> tingkat kesehatan balita untuk setiap kelurahan di Kota Makassar tahun 2018 berdasarkan angka kematian, persentase gizi kurang, gizi kurus, pendek, BBLR, penderita pneumonia dan diare	66

BAB I

PENDAHULUAN

1.1. Latar Belakang

Masalah kesehatan balita merupakan salah satu masalah utama dalam bidang kesehatan yang saat ini terjadi di negara Indonesia. Dalam indeks kesehatan global, Indonesia berada di posisi ke 97 dari 149 negara menurut laporan *The Legatum Prosperity Index 2020* (Legatum Institute, 2020). Derajat kesehatan balita mencerminkan derajat kesehatan bangsa, sebab balita sebagai generasi penerus bangsa memiliki kemampuan yang dapat dikembangkan dalam meneruskan pembangunan bangsa. Adapun dalam menentukan derajat kesehatan balita di Indonesia, terdapat beberapa indikator yang dapat digunakan, antara lain angka kematian balita, angka kesakitan serta status gizi (Hidayat, 2008). Dengan kondisi derajat kesehatan balita yang tinggi diharapkan mampu menciptakan sumber daya manusia yang produktif dan meningkatkan kualitas bangsa. Kualitas kesehatan balita dapat dijadikan sebagai acuan penting bagi Pemerintah, khususnya Pemerintah Daerah (Pemda) agar lebih terpacu untuk meningkatkan mutu pelayanan kesehatan bagi balita sehingga mampu menekan angka kematian dan kesakitan balita, serta dapat meningkatkan status gizi balita (Sari, 2015).

Dinas Kesehatan Kota Makassar secara intensif melakukan pendataan kesehatan penduduk termasuk balita sehingga diperoleh informasi yang berkaitan dengan data kesehatan untuk setiap kelurahan yang ada di Kota Makassar. Dalam rangka meningkatkan kualitas kesehatan Kota Makassar, perlu dilakukan suatu terobosan pengolahan data sehingga mampu memberikan gambaran umum dan menyeluruh tentang perkembangan kesehatan masyarakat khususnya balita untuk setiap kelurahannya (Pemerintah Kota Makassar Dinas Kesehatan, 2016).

Pada penelitian ini bertujuan mengimplementasikan metode *Subtractive Fuzzy Clustering* untuk pengelompokan tiap kelurahan yang ada di Kota Makassar menjadi beberapa kelompok berdasarkan kemiripan karakteristik daerah, yang ditinjau

berdasarkan indikator kesehatan. Hasil *output* pada penelitian ini berupa visualisasi informasi pengelompokan kesehatan balita untuk setiap kelurahan pada Kota Makassar yang diharapkan mampu untuk menjadi bahan pertimbangan agar penanganan dari masalah kesehatan tersebut menjadi lebih terfokus dan tepat sasaran. Dengan adanya metode ini tingkat kesehatan balita berdasarkan indikator kesehatan untuk setiap kelurahan menjadi lebih akurat. Sehingga diharapkan mampu membantu meningkatkan derajat kesehatan, khususnya balita di Indonesia.

1.2. Rumusan masalah

Berdasarkan latar belakang masalah di atas, maka dapat dirumuskan permasalahan penelitian, yaitu “Bagaimana mengetahui pengelompokan tingkat kesehatan balita untuk setiap kelurahan pada Kota Makassar?”

1.3. Tujuan Penelitian

Dalam melakukan penelitian ini, tujuan yang ingin dicapai yaitu agar mengetahui pengelompokan tingkat kesehatan balita untuk setiap kelurahan pada Kota Makassar.

1.4. Batasan Masalah

Untuk menghindari kesalahpahaman dan meluasnya pokok bahasan, maka penulis memberikan batasan masalah sebagai berikut:

1. Hanya difokuskan untuk menentukan status gizi pada anak usia kurang dari 5 tahun atau 0-60 bulan.
2. Data yang digunakan ialah *dataset* balita di setiap kelurahan yang berasal dari Dinas Kesehatan Kota Makassar tahun 2018.
3. Parameter yang digunakan adalah angka kematian balita, presentase balita gizi kurang, presentase balita gizi kurus, presentase balita pendek, BBLR (berat badan lahir rendah), presentase balita menderita pneumonia, dan presentase balita menderita diare.
4. Aplikasi pemrograman yang akan digunakan adalah Matlab.

1.5. Manfaat

Adapun manfaat dari penelitian ini ialah:

1. Dengan diketahuinya pengelompokan kesehatan balita kelurahan di Kota Makassar, maka dapat dijadikan oleh pemerintah sebagai bahan kebijakan untuk peningkatan pelayanan kesehatan di kelurahan yang mempunyai derajat kesehatan terendah sehingga dapat mengurangi angka kematian dan penyakit, serta meningkatkan status gizi balita.
2. Memberikan gambaran tentang tingkat kesehatan balita untuk setiap kelurahan di Kota Makassar.
3. Menjadi sumber informasi mengenai pengelompokan tingkat kesehatan balita dengan metode *Fuzzy Subtractive Clustering*.

1.6. Sistematika Penulisan

Untuk memberikan gambaran singkat mengenai isi tulisan secara keseluruhan, maka akan diuraikan beberapa tahapan dari penulisan secara sistematis, yaitu :

BAB I PENDAHULUAN

Bab ini menguraikan secara umum mengenai hal yang menyangkut latar belakang, perumusan masalah, tujuan, batasan penelitian, manfaat dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi teori-teori dasar dan penelitian-penelitian terkait dari berbagai literatur tentang hal hal yang berhubungan dengan *data mining*, serta metode dan algoritma yang diterapkan.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang perencanaan dan penerapan algoritma serta teknik pengolahan data, dimulai dari *preprocessing* hingga menemukan *clustering*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil pengolahan data serta pembahasan hasil penelitian yang telah dilakukan.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran untuk pengembangan sistem lebih lanjut.

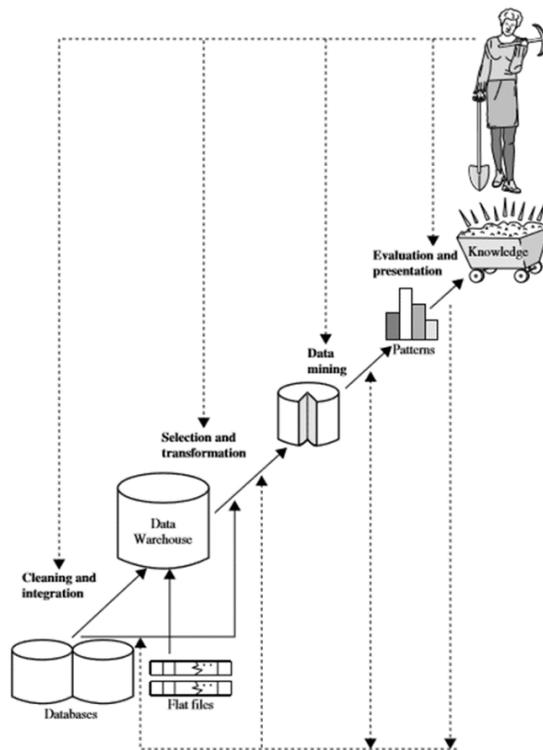
BAB II

TINJAUAN PUSTAKA

2.1. *Data Mining*

Data mining adalah suatu istilah yang digunakan untuk menggambarkan penemuan pengetahuan dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkandung dari berbagai *database* besar (Turban, Aronson & Liang, 2005).

Istilah *data mining* dan *knowledge discovery in databases* (KDD) sering digunakan secara bergantian untuk menggambarkan proses penggalian informasi tersembunyi dalam *database* besar. Faktanya, kedua istilah ini memiliki konsep yang berbeda, tetapi saling berkaitan. Salah satu tahapan dari keseluruhan proses KDD adalah *data mining* (Fayyad, 1996).



Gambar 2.1. Tahapan *data mining* sebagai proses penemuan pengetahuan

Sebagai rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap sebagai berikut (Han, Kamber & Pei, 2011):

1. *Data Collection* (Pengumpulan data)

Pengumpulan data merupakan kumpulan data yang relevan dalam *database* yang dikumpulkan dan diproses menggunakan teknik data mining. Pada tahap ini, didapatkan informasi seperti jumlah data yang digunakan dan variabel dari data.

2. *Data cleaning and Integration* (Pembersihan data dan integrasi data)

Pembersihan data dilakukan untuk menghilangkan *noise* data yang tidak konsisten atau tidak relevan. Integrasi data, merupakan penggabungan untuk beberapa sumber data. Dilakukan pada atribut-atribut yang mengidentifikasi entitas, seperti atribut nama, jenis produk, nomor pelanggan, dsb.

3. *Data selection* (Pemilihan data)

Pemilihan data yaitu mengambil data yang relevan dengan tugas analisis dari *database*.

4. *Data transformation* (transformasi data)

Transformasi data dilakukan dengan mengubah dan mengkonsolidasikan data ke dalam bentuk yang sesuai untuk proses pengolahan data.

5. *Data mining*

Data mining merupakan proses esensial menggunakan metode intelegen untuk mengekstrak pola data.

6. *Pattern evaluation* (evaluasi pola)

Mengidentifikasi pola yang benar-benar menarik yang mepresentasikan pengetahuan berdasarkan ukuran ketertarikan. Pada tahap ini, hasil teknologi *data mining* dievaluasi dalam bentuk model unik dan model prediktif untuk menilai apakah asumsi yang ada memang terpenuhi.

7. *Knowledge presentation* (presentasi pengetahuan)

penggunaan visualisasi dan teknologi representasi pengetahuan untuk menyajikan pengetahuan yang ditambang kepada pengguna. Presentasi pola

yang ditemukan untuk menghasilkan aksi. Pada tahap terakhir dari proses *data mining* ini adalah bagaimana mengambil keputusan atau tindakan berdasarkan hasil analisis yang diperoleh.

Adapun langkah 1 sampai 4 adalah bentuk *preprocessing* data yang berbeda, di mana data perlu disiapkan untuk digunakan dalam pertambangan. Langkah penambangan data dapat berinteraksi dengan pengguna atau basis pengetahuan. Pola yang menarik dapat disajikan kepada pengguna dan dapat disimpan sebagai pengetahuan baru di dasar pengetahuan.

Pada gambar 2.1 menunjukkan bahwa *data mining* merupakan salah satu langkah dalam proses pengetahuan, meskipun penting karena telah mengungkap pola tersembunyi untuk evaluasi. Namun, dalam lingkungan industri, media, dan penelitian, istilah *data mining* sering digunakan yang mengacu pada seluruh seluruh proses penemuan pengetahuan (Han, Kamber & Pei, 2011).

2.2. Supervised dan Unsupervised Learning

Secara garis besar, terdapat dua jenis pembelajaran dari *machine learning*, yaitu *supervised learning* dan *unsupervised learning*.

1. Supervised learning

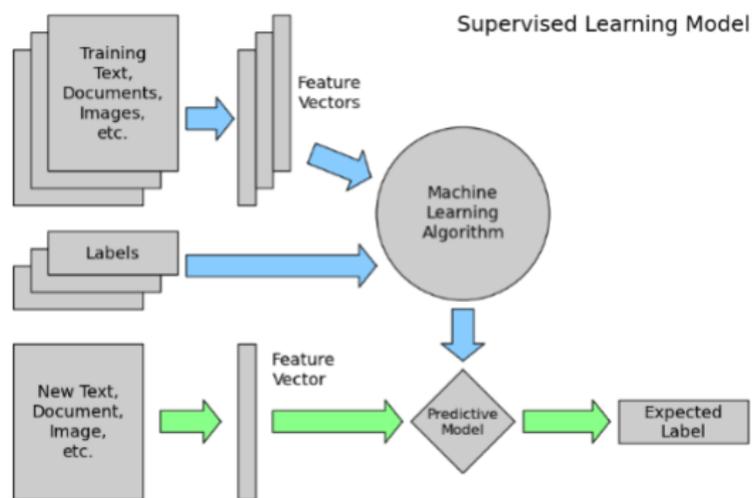
Jika diterjemahkan secara literal, *supervised learning* adalah pembelajaran terarah/terawasi. Metode *supervised learning* didasarkan pada kumpulan sampel data yang memiliki label. Kumpulan sampel digunakan untuk meringkas karakteristik distribusi ukuran perilaku dalam setiap jenis aplikasi sehingga membentuk model perilaku dari data (Wu, 2011).

Supervised learning dalam model pembelajaran mesin sederhana dibagi menjadi dua langkah, yaitu pelatihan dan pengujian (*training and testing*). Dalam proses pelatihan, sampel pada data latihan digunakan sebagai masukan (*input*) fitur mana yang dipelajari oleh algoritma pembelajaran atau pelajar dan membangun model pembelajaran. Dalam proses pengujian, model pembelajaran menggunakan mesin eksekusi untuk membuat prediksi untuk pengujian atau data

produksi. Data yang diberi tanda adalah keluaran dari pembelajaran model yang memberikan prediksi akhir atau klasifikasi data.

Secara umum, *supervised learning* memberikan probabilitas untuk masukan yang tidak terdefinisi, seperti masukan di mana keluaran yang diharapkan diketahui. Proses ini menyediakan set data terdiri dari fitur dan label. Tugas utamanya adalah membangun estimator yang mampu memprediksi label suatu objek yang diberikan oleh kumpulan fitur. Kemudian, algoritma pembelajaran menerima sekumpulan fitur sebagai masukan bersama dengan keluaran yang benar dan belajar dengan membandingkan keluaran aktualnya dengan keluaran yang dikoreksi untuk menemukan kesalahan. Ini kemudian memodifikasi model yang sesuai. Model yang dibuat tidak diperlukan selama masukan tersedia, tetapi jika nilai masukan hilang, tidak mungkin menyimpulkan apapun tentang keluaran.

Selain itu, pembelajaran ini digunakan dalam aplikasi di mana data historis memprediksi kemungkinan peristiwa fitur. Ada banyak contoh praktis dari pembelajaran ini, misalnya aplikasi untuk memprediksi spesies iris mata dengan serangkaian pengukuran bunganya. Adapun tugas *supervised learning* dibagi menjadi dua kategori, yaitu klasifikasi dan regresi. Dalam klasifikasi, label bersifat diskrit sedangkan dalam regresi, label kontinu.



Gambar 2.2. Model *Supervised Learning*

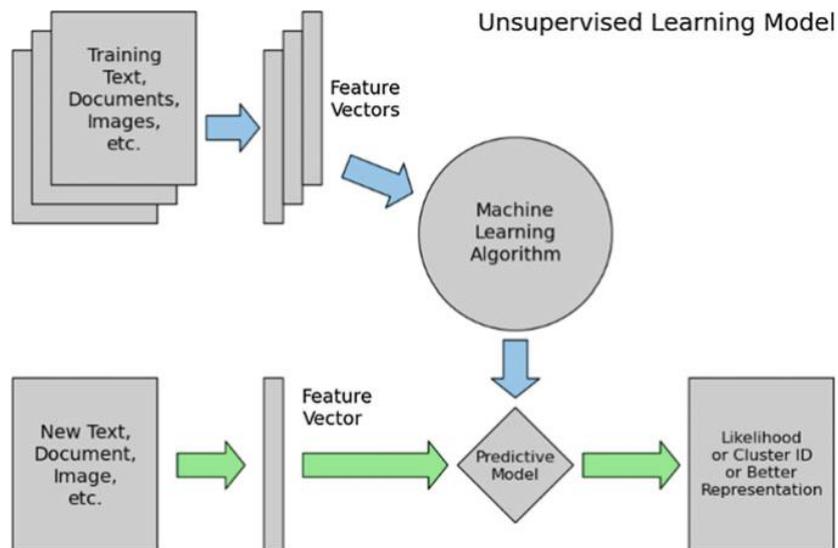
Seperti yang ditunjukkan pada Gambar 2.2, algoritma membuat perbedaan antara data yang diamati X yaitu, data latihan. Dalam banyak kasus, data terstruktur diberikan ke model selama proses pelatihan. Dalam proses ini, algoritma pembelajaran yang diawasi membangun model prediksi. Setelah pelatihannya, model yang dipasang akan mencoba memprediksi label yang paling mungkin untuk sekumpulan sampel baru x dalam set pengujian. Tergantung dari sifat target y , *supervised learning* dapat diklasifikasikan (Nasteski, 2017):

- Jika y memiliki nilai dalam satu set hasil kategorikal tetap bilangan bulat (*integers*), tugas untuk memprediksi y disebut klasifikasi.
- Jika y memiliki nilai *floating-point*, tugas untuk memprediksi y disebut regresi.

2. *Unsupervised learning*

Pada *unsupervised learning*, tidak ada variabel target yang diidentifikasi. Sebaliknya, algoritma *data mining* mencari pola dan struktur di antara semua variabel (Larose, 2005). *Unsupervised learning* merupakan pendekatan yang lebih rumit daripada *supervised learning*. Di sini tujuannya adalah mempelajari sesuatu dengan komputer dengan sendirinya. Ada dua pendekatan utama yang tersedia dalam jenis pembelajaran ini. Pendekatan pertama adalah mengajar dengan bantuan sistem penghargaan (*reward system*) yang merupakan indikator kesuksesan. Pendekatan ini paling cocok untuk kerangka masalah keputusan di mana tujuannya adalah membuat keputusan untuk memaksimalkan imbalan daripada menghasilkan klasifikasi. Jenis pendekatan kedua adalah *clustering* di mana tujuannya adalah menemukan pola yang serupa dalam set data pelatihan alih-alih memaksimalkan fungsi utilitas. Ada bermacam-macam teknik yang digunakan dalam *unsupervised learning* meliputi algoritma *K-Means clustering*, teknik pengurangan dimensi, dll. Beberapa area umum di mana jenis dari

metodologi pembelajaran yang paling cocok adalah menentukan fitur yang paling membedakan antar galaksi di mana pengamatan rinci dari detail galaksi hadir untuk masalah pemisahan sumber buta, dll.



Gambar 2.3. Model *Unsupervised Learning*

Pada gambar 2.3 menggambarkan perbedaan langkah-langkah yang terlibat dalam *unsupervised learning*. *Unsupervised learning* adalah salah satu cara paling efektif untuk menganalisis *big data* karena tidak diperlukan data set pelatihan di sini. Dalam domain data besar, analisis biasanya diperlukan pada kumpulan data dalam eksplorasi di mana kumpulan aturan yang telah ditentukan sebelumnya tidak tersedia. Jadi dalam situasi ini, *unsupervised learning* cukup efektif untuk menemukan pola yang berguna di atas dan di luar *noise* (Desarkar and Das, 2017).

2.3. Clustering

Clustering adalah metode *data mining* yang dapat digunakan untuk masalah-masalah yang tujuannya untuk mengelompokkan kumpulan contoh serupa ke dalam kelompok. Berlawanan dengan klasifikasi, *clustering* menggunakan *unsupervised learning*, yang artinya bahwa contoh set data masukan yang digunakan untuk pelatihan

tidak diberi label, yaitu tidak diketahui mereka milik kelompok mana. Klaster ditentukan dengan memeriksa struktur data dan objek pengelompokan itu serupa menurut beberapa metrik (Kulin, 2016).

Clustering mengacu pada pengelompokan catatan, observasi, atau kasus ke dalam kelas objek serupa. Klaster adalah sekumpulan catatan yang mirip satu sama lain, dan berbeda dengan catatan di klaster lain. *Clustering* berbeda dari *classification*, disebabkan karena tidak ada variabel target untuk *clustering*. Tugas *clustering* tidak mencoba untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel target. Sebaliknya, *clustering* mencari untuk memberi segmentasi kepada seluruh kumpulan data menjadi klaster yang relatif homogen, di mana kesamaan catatan dalam klaster dimaksimalkan dan kesamaan catatan di luar klaster diminimalkan (Larose, 2005).

Analisis klaster atau pengelompokan adalah proses membagi sekumpulan objek data (atau pengamatan) menjadi beberapa *subset*. Setiap *subset* adalah klaster, sehingga objek dalam klaster serupa satu sama lain. Himpunan klaster yang diperoleh dengan analisis klaster dapat disebut klaster. Dalam kasus ini, metode pengelompokan yang berbeda dapat menghasilkan pengelompokan yang berbeda dari kumpulan data yang sama. Partisi tidak dilakukan oleh manusia, tetapi dengan algoritma pengelompokan. Oleh karena itu, pengelompokan berguna karena dapat mengarah pada penemuan grup yang sebelumnya tidak dikenal dalam data.

Sebagai fungsi *data mining*, analisis klaster dapat digunakan sebagai alat independen untuk mendapatkan wawasan tentang distribusi data, mengamati karakteristik setiap klaster, dan fokus pada kumpulan klaster tertentu untuk analisis lebih lanjut. Analisis klaster juga dapat digunakan sebagai langkah pra-pemrosesan untuk algoritma lain, seperti karakterisasi, pemilihan *subset* atribut dan klasifikasi, dan kemudian operasi pada klaster yang terdeteksi dan atribut atau fitur yang dipilih.

Clustering disebut juga segmentasi data dalam beberapa aplikasi, karena pengelompokan membagi kumpulan data yang besar menjadi beberapa kelompok berdasarkan kesamaannya. *Clustering* juga dapat digunakan untuk deteksi *outlier*, di

mana *outlier* (nilai yang jauh dari klaster manapun) mungkin lebih menarik daripada kasus umum. Aplikasi pendeteksian *outlier* meliputi pendeteksian penipuan kartu kredit dan pengawasan kegiatan kriminal di perdagangan elektronik (Han, Kamber & Pei, 2011).

2.4. Fuzzy Subtractive Clustering

Subtractive Clustering merupakan metode pengelompokan yang diperkenalkan oleh Stephen L. Chiu pada tahun 1994. Metode ini merupakan penyempurnaan dari *Mountain Method* (MM) yang diperkenalkan oleh Yager dan Filev pada tahun 1992. *Fuzzy subtractive clustering* merupakan algoritma *unsupervised clustering* yang dapat membentuk jumlah dan pusat klaster sesuai dengan kondisi data. Dasar dari metode pengelompokan *fuzzy subtractive* adalah untuk mengukur densitas (potensi) titik-titik data dalam ruang (variabel). Konsep dasar pengelompokan subtraktif adalah menentukan daerah dengan kepadatan tinggi titik-titik sekitarnya dalam variabel (Kusuma, Furqon & Muflikhah, 2017). Titik dengan tetangga paling banyak akan dipilih sebagai pusat klaster (*centroid*). Kemudian densitas titik-titik yang telah dipilih sebagai pusat kelompok akan berkurang. Kemudian, algoritma akan memilih titik lain dengan tetangga terbesar untuk menjadi pusat klaster lain. Lakukan berulang kali untuk menguji semua poin (Kusumadewi & Purnomo, 2010).

Menurut (Chiu, 1994), penentuan dalam memilih nilai *accept ratio* dan *reject ratio* akan mempengaruhi hasil dari *clustering*. Nilai dari *accept ratio* akan mempengaruhi data yang akan diterima, semakin besar nilai *accept ratio* maka data yang diterima terlalu sedikit untuk menjadi pusat kelompok. Sedangkan jika nilai *reject ratio* terlalu kecil, maka akan terlalu banyak pusat kelompok yang diterima. Maka dari itu, spesifikasi yang diberikan oleh Chiu adalah *accept ratio* = 0,5 dan *reject ratio* = 0,15. Jari-jari berupa vektor yang akan menentukan tingkat pengaruh pusat klaster pada setiap variabel. Oleh karena itu, jika suatu titik data memiliki densitas yang besar maka dia akan memiliki banyak tetangga dekat, sehingga semakin sedikit klaster yang terbentuk.

Adapun algoritma dari *Fuzzy Subtractive Clustering* adalah sebagai berikut (Kusumadewi & Purnomo, 2010):

1. Memasukkan data yang akan dikelompokkan : X_{ij} dengan $i = 1, 2, \dots, n$; dan $j = 1, 2, \dots, m$.

2. Menetapkan nilai:

a. $r; j=1, 2, \dots, m$

r adalah jari-jari setiap atribut data, merupakan vektor yang akan menentukan seberapa besar pengaruh *centroid* pada tiap-tiap dimensi data.

b. q (*squash factor*)

Faktor pengali ke jari-jari yang akan menentukan kedekatan suatu *centroid* yang mana keberadaannya terhadap *centroid* yang lainnya akan dikurangi.

c. *accept ratio*

Merupakan bilangan pecahan yang menunjukkan potensi terhadap *centroid* pertama, jika potensi lebih besar dari *accept ratio*, maka keberadaan titik tersebut akan diterima sebagai *centroid* baru.

d. *reject ratio*

Merupakan bilangan pecahan yang menunjukkan potensi terhadap *centroid* pertama, jika potensi lebih kecil dari *reject ratio*, maka titik tersebut akan diabaikan untuk dipertimbangkan menjadi *centroid* baru selanjutnya.

e. X_{min} (minimum data yang diperbolehkan) .

f. X_{max} (maksimum data yang diperbolehkan).

3. Melakukan normalisasi

$$X_{minj} = \min[x_{ij} \mid j = 1, 2, \dots, m]; i = 1, 2, \dots, n$$

$$X_{maxj} = \max[x_{ij} \mid j = 1, 2, \dots, m]; i = 1, 2, \dots, n$$

Hitung :

$$X_{ijnorm} = \frac{X_{ij} - X_{minj}}{X_{maxj} - X_{minj}} \quad (2.1)$$

4. Menentukan potensi awal tiap-tiap titik data $i = 1$, kerjakan hingga $i = n$, dimana :

$T_j = X_{ijnorm}$

$$Dist_{ij}(x_k) = \frac{T_j - x_{ijnorm}}{r_j} \quad (2.2)$$

$j = 1, 2, \dots, m$; $k = 1, 2, \dots, n$

Potensi awal : jika $m = 1$ maka

$$D_k = \sum_{i=1}^n e^{-4(Dist_{kj}^2)} \quad (2.3)$$

jika $m > 1$, maka

$$D_k = \sum_{i=1}^n e^{-4(\sum_{i=1}^n Dist_{ij}(x_k)^2)} \quad (2.4)$$

hingga $i = i + 1$

5. Mencari titik dengan potensi tertinggi, dimana

$$M = \max [D_i \mid i = 1, 2, \dots, n] \quad (2.5)$$

dan $h = i$, sedemikian hingga $D_i = M$

6. Menentukan pusat kluster dan mengurangi potensinya terhadap titik-titik di sekitarnya

- a. *Center*, yang selanjutnya disebut dengan pusat kluster.
- b. $V_j = ; j = 1, 2, \dots, m$.
- c. $C =$ jumlah kluster
- d. $Z = M$

e. $Rasio = \frac{Z}{M}$ (2.6)

f. Jika $Rasio > accept\ ratio$, maka kondisi = 1; (ada calon pusat baru). Jika tidak : $rasio > reject\ ratio$, (calon baru akan diterima sebagai pusat jika keberadaannya akan memberikan keseimbangan terhadap data-data yang letaknya cukup jauh dengan pusat kluster yang telah ada), kemudian mengerjakan : $Md = -1$ dan $i = 1$ sampai $i = C$

$$G_{ij} = \frac{v_j - center_{ij}}{r} \quad j = 1, 2, \dots, m. \quad (2.7)$$

$$Sd = \sum_{j=1}^m (G_j)^2 \quad (2.8)$$

Jika $(Md < 0)$ atau $(Sd < Md)$, maka $Md = Sd$ dan $Smd = \sqrt{Md}$:

- 1) Jika $(Rasio + Smd) \geq 1$, maka kondisi = 1. Hal ini berarti data diterima sebagai pusat kluster.
 - 2) Jika $(Rasio + Smd) < 1$, maka kondisi = 2. Hal ini berarti data tidak akan dipertimbangkan kembali sebagai pusat kluster.
- g. Jika kondisi = 1 (calon pusat baru diterima sebagai pusat kluster), selanjutnya dikerjakan :

Jumlah kluster (C) = C + 1 $Center_c = V$

Kemudian mengurangi potensi dari titik-titik di dekat kluster:

$$S_{ij} = \frac{V_{ij} - X_{ij}}{r * q} \quad j = 1, 2, \dots, m ; i = 1, 2, \dots, n \quad (2.9)$$

$$D_{C_i} = M \times e^{-4 \left[\sum_{j=1}^m (S_{ij})^2 \right]}, \quad i = 1, 2, \dots, n \quad (2.10)$$

D_{C_i} merupakan nilai pengurang potensial setiap titik

$$D_i \text{ baru} = D_i \text{ lama} - D_{ci}$$

Jika $D_i \text{ baru} \leq 0$, maka $D_i \text{ baru} = 0$, $i = 1, 2, \dots, n$

$Z = \max [D_i \mid i = 1, 2, \dots, n]$. Pilih $h = i$, sedemikian hingga $D_i = Z$

- h. Jika kondisi = 2 (calon pusat baru tidak diterima sebagai pusat baru), maka

$$D_h = 0$$

$$Z = \max [D_i \mid i = 1, 2, \dots, n] \quad (2.11)$$

pilih $h = i$, sedemikian hingga $D_i = Z$.

7. Mengembalikan pusat kluster dari bentuk yang sudah dinormalisasi ke bentuk semula :

$$Center_{semula} = Center_{ternormalisasi} * (X_{max_j} - X_{min_j}) + X_{min_j} \quad (2.12)$$

8. Menghitung nilai sigma kluster (nilai parameter fungsi keanggotaan *Gauss*)

$$\sigma_j = \frac{r_j * (X_{max_j} - X_{min_j})}{\sqrt{8}} \quad (2.13)$$

9. Menghitung derajat keanggotaan setiap data pada setiap pusat kelompok yang terbentuk dengan fungsi *Gauss*.

$$\mu_{li} = e^{-\frac{\sum_{j=i}^m (x_{ij} - c_{ij})^2}{2\sigma_j^2}} \quad (2.14)$$

Dimana pusat kluster (C) dan sigma (σ) akan digunakan untuk menentukan nilai parameter fungsi keanggotaan *Gauss*.

2.5. Performa Hasil *Clustering*

Pada metode pengelompokan menggunakan konsep *fuzzy*, data dapat menjadi anggota dari semua kluster yang memiliki nilai keanggotaannya. Semakin tinggi nilai keanggotaan suatu kluster, maka semakin besar kecenderungan untuk menjadi anggota kluster.

2.5.1. *Silhouette Coefficient*

Silhouette Coefficient merupakan metode yang digunakan untuk melihat kualitas dan kekuatan klaster, dan seberapa baik suatu objek ditempatkan dalam suatu klaster. Metode *Silhouette Coefficient* merupakan metode kombinasi metode *cohesion* dan *separation*. Metode *Cohesion* adalah ukuran seberapa dekat hubungan antar objek dalam sebuah klaster. Sedangkan metode *separation* merupakan ukuran seberapa jauh atau terpisahnya sebuah klaster dengan klaster lainnya (Kusuma, Furqon & Muflikhah, 2017).

Adapun tahapan dari *Silhouette Coefficient* sebagai berikut (Handoyo, Rumani & Nasution, 2014):

1. Hitung rata-rata jarak dari suatu dokumen misalkan i dengan semua dokumen lain yang berdada dalam satu klaster

$$a(i) = \frac{1}{|A| - 1} \sum_j \in_{A, j \neq i} d(i, j) \quad (2.15)$$

dimana j adalah dokumen lain dalam suatu klaster A , $d(i, j)$ adalah jarak antara dokumen i dengan j , $a(i)$ merupakan rata-rata jarak, dan A merupakan jumlah anggota klaster dokumen i

2. Hitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di klaster lain, dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|} \sum_j \in C d(i, j) \quad (2.16)$$

Dengan $d(i, C)$ adalah jarak rata-rata dokumen i dengan semua objek pada klaster lain C dimana $A \neq C$, dan $d(i, j)$ adalah jarak dokumen i ke dokumen anggota klaster selain dokumen i .

3. Nilai *Silhouette Coefficient* nya adalah

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.17)$$

Dimana $s(i)$ merupakan nilai dari *Silhouette Coefficient*, $b(i)$ merupakan nilai minimum dari $d(i,C)$, dan $a(i)$ adalah jarak rata-rata i terhadap seluruh data pada klaster yang sama.

Nilai yang dihitung menggunakan metode *Silhouette Coefficient* terletak pada kisaran antara nilai -1 sampai 1. Jika $s(i) = 1$ berarti objek I berada pada klaster yang tepat. Jika nilai dari $s(i) = 0$ maka objek I berada diantara dua klaster, sehingga objek tersebut tidak jelas harus dimasukkan pada klaster A atau klaster B. Jika $s(i) = -1$ maka terjadi kesalahan dalam menentukan entitas selama proses *clustering*, sehingga objek I lebih tepat untuk dimasukkan dalam klaster lain (Kusuma, Furqon & Muflikhah, 2017).

2.5.2. *Sum of Squared Error (SSE)*

Sum of Squared Error (SSE) merupakan salah satu cara untuk mengukur *clustering* dengan menggunakan teknik statistik yang mampu mencari apakah objek cocok pada satu klaster. Adapun perhitungan dari SSE ialah sebagai berikut:

$$SSE = \sum_{i=1}^n (d_i - c_i)^2 \quad (2.18)$$

Keterangan:

SSE = nilai kuadrat selisih antara koordinat *centroid* ke setiap data

n = jumlah data

d_i = nilai data ke- i

c_i = nilai *centroid* klaster ke- i

Bila objek sangat cocok dengan klaster tersebut maka nilai SSE adalah nol atau berarti tidak ada *error* atau sangat cocok. Namun hal itu jarang terjadi, oleh karena itu, *clustering* yang baik adalah yang memiliki nilai SSE serendah mungkin. Semakin rendah nilai SSE maka semakin sama. SSE yang tinggi maka memiliki derajat perbedaan antara objek dan klaster yang dituju (Shofiani, 2017).

2.5.3. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) merupakan rata-rata perbedaan absolut antara peringkat aktual dan peringkat prediksi atau ukuran deviasi rekomendasi dari nilai sebenarnya yang telah ditentukan pengguna. Nilai MAE yang lebih rendah menunjukkan prediksi yang lebih baik (Fadilla, Andarsyah & Awangga, 2020).

$$MAE = \frac{1}{n} \sum_{i=1}^n |d_i - c_i| \quad (2.19)$$

Keterangan:

MAE = nilai rata-rata absolut selisih antara koordinat *centroid* ke setiap data

n = jumlah data

d_i = nilai data ke- i

c_i = nilai centroid klaster ke- i

2.6. Penelitian Terkait

Penelitian yang terkait merupakan salah satu bahan acuan bagi penulis untuk melakukan penelitian, sehingga penulis dapat memperkaya teori yang digunakan untuk mengkaji penelitian yang dilakukan. Penelitian terkait penggunaan teknologi *data mining* metode *clustering* dengan menggunakan algoritma *fuzzy subtractive clustering* telah banyak digunakan dalam penelitian.

Salah satu penelitian terkait pemanfaatan algoritma *fuzzy subtractive clustering* adalah penelitian yang berjudul “Implementasi Metode *Fuzzy Subtractive Clustering* Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan”. Pada penelitian ini menggunakan metode *fuzzy subtractive clustering* dengan mengelola data titik api yang menghasilkan pengklasterkan potensi kebakaran hutan yang dikelompokkan dalam potensi tinggi dan potensi sedang, dengan nilai rata-rata *Silhouette Coefficient* sebesar 0.45 dan hasil klaster yang terbentuk sebanyak dua *cluster*. (Kusuma, Furqon & Muflikhah, 2017).

Penelitian kedua yang berjudul “Penerapan Algoritma *K-Means* Untuk Menentukan Tingkat Kesehatan Bayi Dan Balita Pada Kabupaten Dan Kota Di Jawa Tengah”. Pada penelitian tersebut, dilakukan *clustering* menggunakan algoritma *k-means* untuk mengetahui tingkat kesehatan bayi dan balita dengan menggunakan atribut seperti angka kematian bayi, angka kematian balita, presentase balita gizi kurang, presentase balita gizi buruk, BBLR, presentase balita pneumonia, dan presentase balita diare. Adapun hasil penelitian berdasarkan analisa setiap indikator kesehatan dengan tiga kategori, yaitu tinggi, sedang, dan rendah. (Sari, 2015).