

BAB II

LANDASAN TEORI

2.1 DATA MINING

2.1.1 Pengertian Data

Menurut Edhy Sutanta dalam jurnal Hermansyah Sembiring dan Nurhayati (2012:2) menyatakan: “Data adalah sebagai bahan keterangan tentang kejadian nyata atau fakta-fakta yang dirumuskan dalam sekelompok lambang tertentu yang tidak acak yang menunjukkan jumlah, tindakan, atau hal”

Menurut Eka Iswandy (2015:4) menyatakan:

“Data adalah sesuatu yang belum mempunyai arti bagi penerimanya dan masih memerlukan adanya suatu pengolahan. Data bisa berwujud suatu keadaan, gambar, suara, huruf, angka, matematika, bahasa ataupun simbol-simbol lainnya yang bisa kita gunakan sebagai bahan untuk melihat lingkungan, obyek, kejadian ataupun suatu konsep”.

Menurut Sutarman (2012:3) mengatakan: “Data adalah fakta dari suatu pernyataan yang berasal dari kenyataan, dimana pernyataan tersebut merupakan hasil pengukuran atau pengamatan”.

Jadi dapat disimpulkan bahwa data adalah sekumpulan fakta-fakta hasil pengukuran dan pengamatan yang belum memiliki arti dan dapat diolah menjadi suatu informasi tertentu.

2.1.2 Pengertian Data Mining

Menurut Aline Embun Pramadhani (2014:833) dalam jurnalnya:

“*Data Mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya”.

Menurut Alfannisa Annurullah Fajrin dan Algifanri Maulana (2018:29) mengatakan:

“*Data Mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan *Data Mining* adalah kenyataan bahwa *Data Mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu”.

Berawal dari beberapa disiplin ilmu, *Data Mining* bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

- a. Jumlah data yang sangat besar
- b. Dimensi data yang tinggi
- c. Data yang heterogen dan berbeda sifat

Heni Sulastri dan Acep Irham Gufroni (2017:2) menyatakan: “ *Data Mining* merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data”.

Jadi dapat disimpulkan bahwa *Data Mining* adalah proses penggalian data secara mendalam untuk mengetahui hal yang berarti dan tidak diketahui keberadaanya.

2.1.3 Tahapan *Data Mining*

Secara sederhana *Data Mining* biasa dikatakan sebagai proses penyaring atau menambang pengetahuan dari sejumlah data yang besar. Istilah *Data Mining* dan *Knowlegde Discovering in Database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data

yang besar. Proses (KDD) secara garis besar dapat dijelaskan sebagai berikut (Menurut Fayyad dalam buku Kusriani dan Emha Taufiq Luthfi, 2009 : 6) :

a. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *Data Mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. *Pre-processing / Cleaning*

Sebelum proses *Data Mining* dapat dilaksanakan, dapat dilakukan proses *cleaning* data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*), juga dilakukan proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

c. *Transformation*

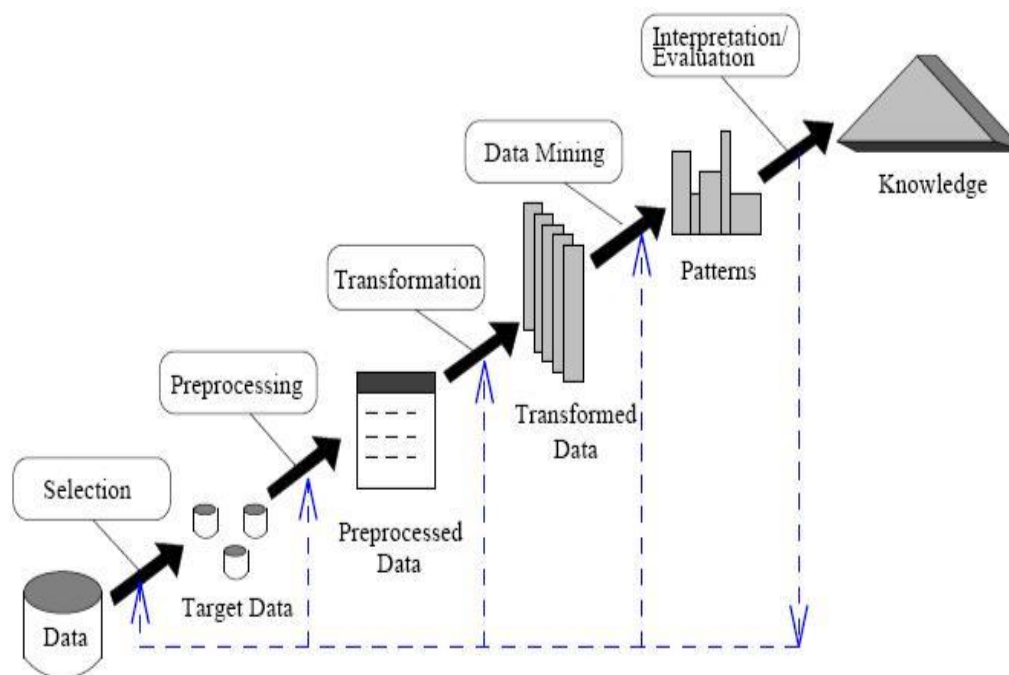
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d. Data Mining

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data Mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

e. Interpretation/ Evaluation

Pola yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



Gambar 2.1 Proses KDD (Fayyad, 2009)

Tahapan yang dilakukan pada proses *Data Mining* diatas diawali dari seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data, transformasi, *Data Mining* serta tahap interpretasi dan evaluasi yang menghasilkan *output* berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik.

2.1.4 Pengelompokan *Data Mining*

Menurut Larose dalam Kusri dan Emha Taufiq Luthfi (2009 : 10) *Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

a. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

b. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

c. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

d. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. Pengklasteran

Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklasteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu

kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

f. Asosiasi

Tugas asosiasi dalam *Data Mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.2 CLUSTERING

2.2.1 Pengertian Clustering

Menurut Heni Sulastri dan Acep Irham Gufroni (2017:2) mengatakan: “*Clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain”

Menurut Garcia dalam jurnal Asroni, Ronald Adrian (2015:2) menyatakan:

“*Clustering* adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial”.

Menurut Sri Rahayu dkk (2015:2) dalam jurnalnya menyatakan:

“*Clustering* merupakan salah satu metode *Data Mining* yang melakukan pemisahan/pemecahan/segmentasi data ke dalam sejumlah kelompok (*Cluster*) menurut karakteristik tertentu yang diinginkan, dalam pekerjaan pengelompokan, label dari setiap data belum diketahui dan dengan pengelompokan diharapkan dapat diketahui kelompok data untuk kemudian diberi label sesuai keinginan”.

Jadi dapat disimpulkan bahwa *Clustering* adalah metode untuk mengelompokkan data yang memiliki kemiripan dan kemudian di beri label sesuai keinginan.

Setelah menganalisis data-data penduduk miskin yang di peroleh dari Dinas Sosial maka dengan metode *Clustering* lebih cocok diolahnya, karna *Clustering* mengelompokkan data yang memiliki kemiripan.

2.2.2 Algoritma dalam *Clustering*

Adapun algoritma *Clustering* dalam *Data Mining* meliputi:

a. *Fuzzy C-Means*

Menurut Muhardi & Nisar (2015:5) dalam jurnalnya menyatakan: “FCM adalah suatu teknik pengklasteran data yang keberadaan tiap-tiap data dalam suatu *Cluster* ditentukan oleh nilai/derajat keanggotaan tertentu”

Menurut Qonita U. Safitri, dkk (2017:25) dalam jurnalnya: “fuzzy digunakan untuk menyatakan kelompok/ himpunan yang dapat dibedakan dengan kelompok lain berdasarkan derajat keanggotaan secara samar atau tidak tegas”

b. *K-Medoids*

Menurut Dyang Falila Pramesti, dkk (2017:256) menyatakan:

“*K-Medoids* atau *Partitioning Around Medoids* (PAM) adalah algoritma *Clustering* yang mirip dengan *K-Means*. Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* atau PAM menggunakan objek sebagai perwakilan (*medoid*) sebagai pusat *Cluster* untuk setiap *Cluster*”.

Menurut Yulison Herry Chrisnanto dan Gunawan Abdillah (2015:445)

dalam jurnalnya:

“*K-Medoids* adalah sebuah algoritma yang merepresentasikan *Cluster* yang dibentuk menggunakan *medoids*. Proses pembentukan *Cluster*

dimulai dengan menentukan k objek dari dataset secara acak sebagai *medoid*, selanjutnya hitung *cost* setiap objek non-medoid dengan k objek, *cost* terkecil setiap objek non-medoid terhadap medoid akan masuk dalam *Cluster* dimana *medoid* tersebut berada”.

c. *K-Means*

Menurut Aniq Noviciatie Ulfah dan Shofwatul ‘Uyun (2015:42)

menjelaskan:

“*K-Means Clustering* termasuk dalam *partitioning Clustering* yang disebut juga *exclusive Clustering* yang memisahkan data ke k daerah bagian yang terpisah dan setiap data harus termasuk ke dalam *Cluster* tertentu dan memungkinkan setiap data yang termasuk *Cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *Cluster* yang lain”.

Menurut Jaroji, dkk (2016:3) menyatakan:

“Algoritma *K-Means* adalah algoritma yang mempartisi data ke dalam *Cluster – Cluster* sehingga data yang memiliki kemiripan berada pada satu *Cluster* yang sama dan data yang memiliki ketidaksamaan berada pada *Cluster* yang lain”

d. *Particle Swarm Optimization (PSO)*

Menurut Ardiana Rosita (2012:1) dalam jurnalnya menyatakan: “Algoritma *Particle Swarm Optimization (PSO)* adalah teknik optimasi berdasarkan populasi yang terinspirasi oleh perilaku sosial dari pergerakan burung atau ikan (*bird flocking* atau *fish schooling*).

Menurut Bayu Wibisana, dkk (2016:115) dalam jurnalnya menyatakan:

“PSO terinspirasi oleh kemampuan kawanan burung, ikan dan hewan untuk beradaptasi pada lingkungan, menemukan sumber makanan dan menghindari predator dengan pendekatan “berbagi informasi”.

e. *Hierarchical Clustering*

Menurut Gregorius S. Budhi, dkk, (2008 : C-27) dalam artikelnya menyatakan :

“*Hierarchical Clustering* adalah salah satu algoritma *Clustering* yang dapat digunakan untuk meng-klaster dokumen (*document Clustering*).

Dari teknik *hierarchical Clustering*, dapat dihasilkan suatu kumpulan partisi yang berurutan, di mana dalam kumpulan tersebut terdapat:

- a. *Cluster – Cluster* yang mempunyai poin – poin individu. *Cluster – Cluster* ini berada di level yang paling bawah.
- b. Sebuah *Cluster* yang di dalamnya terdapat poin – poin yang dipunyai semua *Cluster* didalamnya. *Single Cluster* ini berada di level yang paling atas.

f. *Complete Linkage*

Menurut Gede Aditra Pradnyana dan Ngurah Agus Sanjaya ER, (2012 : 3) dalam artikelnya menyatakan bahwa :

“*Complete Linkage* adalah suatu metode yang menggunakan prinsip jarak minimum yang diawali dengan mencari jarak terjauh antar dua buah *Cluster* dan keduanya membentuk *Cluster* baru. Pada awalnya, dilakukan perhitungan jarak terpendek dalam $D = \{d_{ik}\}$ dan menggabungkan objek-objek yang bersesuaian misalnya, U dan V , untuk mendapatkan *Cluster*

(UV). Kemudian jarak-jarak antara (UV) dan *Cluster W* yang lain dihitung dengan cara :

$$d_{(uv)w} = \max\{d_{uw}, d_{vw}\}$$

g. *Single Linkage*

Menurut Manning, dkk dalam artikel Rendy Handoyo, dkk, (2014 : 75-76) menyatakan bahwa : “*Single Linkage Clustering* merupakan contoh dari algoritma *Agglomerative Hierarchical Clustering*.”

Langkah-langkah dari metode *Single Linkage Clustering* adalah sebagai berikut” :

1. Menentukan k sebagai jumlah *Cluster* yang ingin dibentuk.
2. Setiap data dianggap sebagai *Cluster*. Kalau n = jumlah data dan c = jumlah *Cluster*, berarti ada c = n.
3. Menghitung jarak / *similarity* / *dissimilarity* antar *Cluster*.
4. Cari dua *Cluster* yang mempunyai jarak antar *Cluster* yang minimal dan gabungkan (c = c1). Setelah semua jarak diketahui, selanjutnya dikelompokkan dokumen-dokumen yang memiliki jarak terdekat.
5. Jika c > 3, kembali ke langkah 3.

2.3 ALGORITMA *K-MEANS*

Taslim dan Fajrizal (2016:110) Algoritma *K-Means Clustering* adalah algoritma sederhana dan efektif untuk menemukan *Cluster* dalam data dengan algoritma sebagai berikut:

Langkah 1 : Tentukan jumlah *Cluster*

Langkah 2 : Tentukan nilai yang menjadi lokasi pusat *Cluster* awal.

Langkah 3 : Hitung pusat *Cluster* terdekat untuk setiap record.

Langkah 4 : Untuk setiap *Cluster* k , hitung centroid *Cluster* dan memperbarui lokasi setiap pusat *Cluster*

Langkah 5 : Ulangi langkah 3 sampai 5 sampai konvergensi atau penghentian.

Algoritma *K-Means* dikenal dan banyak digunakan untuk metode *partitional*, yaitu membagi himpunan objek data ke dalam sub himpunan *Cluster* yang tidak saling tumpang tindih, sehingga setiap objek data tepat berada dalam satu *Cluster*.

Secara matematis metode *K-Means* dapat menggunakan rumus sebagai berikut:

$$C_i = \frac{1}{M} \sum_{j=1}^M X_{ij}$$

Rumus 1. Sentroid fitur ke-i

Dimana

C_i : Sentroid fitur ke-i

M : Jumlah data dalam sebuah kelompok

i : Fitur ke-i dalam sebuah kelompok

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

Keterangan:

p = Dimensi data x_1 = Posisi

titik 1 x_2 = Posisi titik 2

BCV (*Between Cluster Variation*) =

WCV (*Within Cluster Variation*) =

M = Mean (rata-rata)

D = Data

Algoritma standar dari *K-Means* adalah sebagai berikut :

1. Langkah 1 : Tentukan jumlah *clustering* yang diinginkan (misalkan : k3).
2. Langkah 2 : Pilih centroid awal secara acak. Pada langkah ini secara acak akan dipilih 3 buah data sebagai *centroid*.
3. Langkah 3 : Hitung jarak dengan *centroid*.

Pada langkah ini setiap data akan ditentukan *centroid* terdekatnya, dan data tersebut akan ditetapkan sebagai anggota kelompok yang terdekat dengan *centroid*.

Untuk menghitung jarak ke *centroid* masing-masing *cluster*

Misalkan data (x,y) , *centroid* M1 : (a_1,b_1) , *centroid* M2 : (a_2,b_2) , *centroid* M3 : (a_3,b_3) .

$$DM1 = \sqrt{(x - a1)^2 + (y - b1)^2} = ? \dots\dots\dots (2.3)$$

$$DM2 = \sqrt{(x - a2)^2 + (y - b2)^2} = ? \dots\dots\dots (2.4)$$

$$DM3 = \sqrt{(x - a3)^2 + (y - b3)^2} = ? \dots\dots\dots (2.5)$$

Buat tabel hasil perhitungan jarak selengkapnya antara masing-masing data dengan *centroid*, maka di dapatkan keanggotaan dari masing-masing *cluster*

Pada langkah ini dihitung pula rasio antara BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*) :

Karena *centroid* M1 = (a1,b1), M2 = (a2,b2), M3 = (a3,b3).

$$d(m1,m2) = \sqrt{(a1 - a2)^2 + (b1 - b2)^2} = ? \dots\dots\dots (2.6)$$

$$d(m1,m3) = \sqrt{(a1 - a3)^2 + (b1 - b3)^2} = ? \dots\dots\dots (2.7)$$

$$d(m2,m3) = \sqrt{(a2 - a3)^2 + (b2 - b3)^2} = ? \dots\dots\dots (2.8)$$

$$BCV = d(m1,m2) + d(m1,m3) + d(m2,m3) = ?$$

Dalam hal ini $d(m_i, m_j)$ menyatakan jarak *euclidean* dari m_i ke m_j

Menghitung WCV yaitu dengan memilih jarak terkecil yang terdapat pada tabel keanggotaan.

$$WCV = c1^2 + c2^2 + c3^2 + N = ?$$

$$\text{Sehingga besar rasio} = BCV/WCV = ?$$

Karena langkah ini merupakan iterasi 1 maka lanjutkan ke langkah berikutnya.

4. Langkah 4 : pembaruan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster*.

Setelah menghitung rata-rata nilai pada masing-masing *cluster* didapatkan *centroid* baru yaitu : $M1 = (a1,b1)$, $M2 = (a2,b2)$, $M3 = (a3,b3)$.

5. Langkah 5 : (iterasi ke 2) kembali kelangkah ke 3, jika masih ada data yang berpindah *cluster* atau jika nilai *centroid* diatas ambang, atau jika pada fungsi obyektif yang digunakan masih diatas ambang. Jika tidak maka iterasi dihentikan.

Berdasarkan referensi dari penelitian-penelitian sebelumnya yang menggunakan algoritma *K-Means*, algoritma ini memiliki tingkat akurasi yang lebih tinggi. Penelitian yang dilakukan Gusti Ngurah Wisnu Paramartha, dkk (2017) mereka memperoleh tingkat akurasi algoritma *K-Means* sebesar 90,3%, sedangkan penelitian yang dilakukan oleh Andi Sri Irtawaty (2017) dia memperoleh tingkat keakuratan metode ini yaitu sebesar 90%.

2.4 PENDUDUK MISKIN

Menurut Hasya Yanto mengatakan: “kemiskinan adalah keadaan dimana seseorang yang tidak bisa memenuhi kebutuhan hidupnya”.

Menurut Almira Qatrunnada Qurratu’ain dan Vita Ratnasari (2016:3) menyatakan:

“Kemiskinan adalah keadaan dimana terjadi ketidakmampuan untuk memenuhi kebutuhan dasar seperti makanan, pakaian, tempat berlindung, pendidikan, dan kesehatan”

Menurut Risky Dwi Setiyawan, dkk (2016:1) menyatakan:

“Kemiskinan secara konseptual dibedakan menurut kemiskinan relatif dan kemiskinan absolut, dimana perbedaannya terletak pada standar penilaiannya. Standar penilaian kemiskinan relatif merupakan standar kehidupan yang ditentukan dan ditetapkan secara subyektif oleh masyarakat setempat dan bersifat lokal serta mereka yang berada dibawah standar penilaian tersebut dikategorikan sebagai miskin secara relatif. Sedangkan standar penilaian kemiskinan secara absolut merupakan standar kehidupan minimum yang dibutuhkan untuk memenuhi kebutuhan dasar yang diperlukan, baik makanan maupun non makanan. Standar kehidupan minimum untuk memenuhi kebutuhan dasar ini disebut sebagai garis kemiskinan”.

Jadi dapat disimpulkan bahwa kemiskinan dimana seseorang tidak dapat memenuhi kebutuhan hidupnya seperti kebutuhan makanan, pakaian, tempat tinggal, pendidikan dan kesehatan.

2.5 TOOLS DALAM DATA MINING

Ada banyak *tools* yang tersedia untuk *Data Mining*. Tujuan utama *Data Mining tools* adalah untuk menemukan data, mengekstrak data, menyaring data, mendistribusikan informasi dan memonetisasinya. Berikut beberapa penjelasan *Data Mining tools* antara lain :

2.5.1 Orange Ailab

Orange Ailab adalah perangkat lunak open source yang memungkinkan pengguna yang tidak memahami sedikitpun tentang pemrograman dapat melakukan visualisasi dan analisis data. Fitur-fitur yang dimiliki di antaranya *scatterplots, bar charts, trees, dendrograms, networks* dan *heatmaps*. (Subekti Mujiasih 2011 : 191).

2.5.2 Rapid Miner

Siska Haryati, Aji Sudarsono, Eko Suryana (2015:133) Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini

2.5.3 SPSS

Menurut Priyatno (2013: 13), dalam buku Haryadi Sarjono dan Winda Julianita(2011 : 113) mengatakan bahwa SPSS adalah program atau *software* yang digunakan untuk mengolah data statistik. Dari berbagai program olah data statistik lainnya, *SPSS* merupakan Program yang paling banyak digunakan. Umumnya, *SPSS* dulu digunakan untuk mengolah data statistik pada ilmu sosial saja. Namun demikian seiring berjalannya waktu, *SPSS* mengalami perkembangan dan penggunaanya semakin kompleks untuk berbagi ilmu pengetahuan. Selain ilmu sosial, *SPSS* juga sering digunakan untuk ilmu ekonomi, psikologi, pertanian, teknologi, industri, dan lain-lain.

2.5.4 Weka

Menurut Dewi Sartika dan Dana Indra Sensuse, (2017) dalam artikel Vinny Alvionita (2018 : 3) mengatakan .” *Waikato Environment for Knowledge Analysis* (WEKA) merupakan perangkat lunak pembelajaran mesin yang populer yang ditulis dalam bahasa pemrograman *java*. *WEKA* dikembangkan di

Universitas Waikato, Selandia Baru. *WEKA* berisikan kumpulan algoritma beserta visualisasinya untuk analisis data dan pemodelan prediktif. Algoritma-algoritma pembelajaran mesin pada *WEKA* dapat dimanfaatkan untuk pemecahan masalah dibidang *Data Mining*. *WEKA* versi asli awalnya dirancang untuk menganalisis data dari domain pertanian, tetapi *WEKA* versi lengkap berbasis *java* (versi 3), yang mulai dibangun pada tahun 1997, yang sekarang dapat digunakan untuk menganalisis data dari berbagai domain, khususnya untuk pendidikan dan penelitian. *WEKA* memiliki implementasi semua teknik pembelajaran untuk klasifikasi dan regresi , yaitu *decision trees*, *rules set*, pengklasifikasian *teorema bayes*, *Support Vector Machines (SVM)*, logistik dan linier, *multi layers perceptrons* dan metode *nearest neighbour*. “

2.6 PENELITIAN TERKAIT

Untuk memperoleh wawasan dan pengetahuan tentang penelitian sejenis, maka peneliti melakukan tinjauan penelitian sejenis, dimana ringkasan tercantum pada tabel 2.1.

Table 2.1 Penelitian Terkait

No	Penulis (tahun)	Masalah	Metode	Hasil
1	Femi Dwi Astuti (2016)	Pada kondisi saat ini, penentuan pemberian bantuan di Kecamatan Bantul dilakukan berdasarkan status kemiskinan dari total skor data yang diperoleh dari hasil pendataan penduduk miskin. Semakin tinggi skornya maka dianggap semakin miskin. Hasil yang diperoleh dengan cara seperti ini mengakibatkan bantuan tidak tersalurkan dengan tepat.	<i>Algoritma Hard C-Means</i>	Dari hasil proses pengClusteran dengan jumlah Cluster 3, Cluster pertama berisi kelompok-kelompok keluarga yang aspek pemenuhan kebutuhan pangan, papan, kesehatan, air bersih dan listrik tergolong paling rendah jika dibandingkan dengan kelompok 2 dan kelompok 3. Cluster kedua berisi kelompokkelompok keluarga yang aspek pemenuhan kebutuhan sandang dan kekayaan 1 masih rendah jika dibandingkan dengan kelompok kedua dan ketiga. Cluster ketiga berisi kelompok-kelompok keluarga yang aspek pemenuhan pendidikan, kekayaan 2 dan jumlah jiwa masih rendah.

No	Penulis (tahun)	Masalah	Metode	Hasil
2	Irtania Muthia Rizki (2017)	Peneliti ingin melakukan analisis klaster dalam mengelompokkan rumah tangga miskin yang terdapat di Kecamatan Tabir Barat berdasarkan keadaan infrastruktur. Dalam menggambarkan variabel kemiskinan berdasarkan infrastruktur pada rumah tangga menggunakan 9 indikator, yaitu status bangunan tempat tinggal, jenis lantai terluas, jenis dinding terluas, jenis atap terluas, sumber air minum, fasilitas buang air besar, tempat pembuangan air tinja, sumber penerangan utama dan bahan bakar untuk memasak.	<i>Latent Class Cluster Analysis</i>	Latent Class <i>Cluster Analysis</i> menghasilkan 5 klaster ke klaster 1 (keadaan infrastruktur kurang) klaster 2 (keadaan infrastruktur cukup) klaster 3 (keadaan infrastruktur sangat kurang) klaster 4 (keadaan infrastruktur sangat baik) klaster 5 (keadaan infrastruktur baik)
3	Aniq Noviciatie Ulfah , Shofwatul ‘Uyun2 (2015)	Sebagai upaya pengentasan kemiskinan pemerintahan Kabupaten Gunungkidul membuat suatu penanggulangan kemiskinan dengan membuat rumusan yaitu membuat indikator-indikator kemiskinan serta mengklasifikasikannya kedalam tiga kategori yaitu keluarga tidak miskin, keluarga miskin dan keluarga sangat miskin berdasarkan 15 indikator kemiskinan	<i>Algoritma Fuzzy C-Means dan K-Means</i>	Hasil dari pengujian dilakukan dengan membandingkan hasil <i>Clustering</i> kedua algoritma tersebut dengan pengelompokan berdasarkan aturan BAPPEDA 2008. Hasil <i>Clustering</i> menggunakan algoritma FCM memiliki tingkat akurasi hanya 50%, sedangkan untuk algoritma <i>K-Means</i> memiliki tingkat akurasi lebih baik yaitu 83.33%..

No	Penulis (tahun)	Masalah	Metode	Hasil
4	Almira Qatrunnada Qurratu'ain dan Vita Ratnasari (2016)	Jawa Timur merupakan salah satu provinsi di Indonesia yang terdiri dari 38 kabupaten/ kota. BPS melaporkan bahwa jumlah penduduk miskin terbanyak berada di Provinsi Jawa Timur, kemudian disusul Jawa Tengah sebanyak 4,58 juta jiwa dan Jawa Barat sebanyak 4,35 juta jiwa. Ketidaktepatan sasaran dalam pengentasan kemiskinan menyebabkan tujuan pengentasan angka kemiskinan jauh dari harapan	<i>Regresi Panel</i>	Hasil dari penelitian ini yaitu Rata-rata persentase penduduk miskin di Jawa Timur dari tahun 2005 hingga tahun 2014 semakin menurun. Selain itu, selisih antara nilai minimum dan maksimum pada masing-masing variabel berbeda jauh. Hal ini mengindikasikan bahwa terjadi ketidakmerataan yang ada di kabupaten/kota
5	Risky Dwi Setiyawan, Dwi Sunaryono, Rizky Januar Akbar(2016)	Salah satu faktor penting dalam penanganan kemiskinan ini ialah data yang akurat dan program pemerintah yang tepat sasaran. Saat ini, pendataan tingkat kemiskinan masyarakat, khususnya di Kabupaten Madiun, masih dilakukan secara konvensional (menggunakan kertas kemudian dimasukkan kedalam sistem) dan tidak menampilkan program pemerintah apa saja yang seharusnya didapatkan dari hasil pendataan tersebut.	<i>Algoritma K-Means</i>	Hasil dari penelitian ini berupa: Dengan menggunakan aplikasi ini, surveyor cukup datang ke rumah warga dan memasukkan data survei pada aplikasi kemudian data langsung terkirim ke server. Dimana hasil yang sudah terkirim, dapat dilakukan validasi secara langsung hingga keluar hasil akhir berupa tingkat kesejahteraan masyarakat

Berdasarkan tabel 2.1 maka dapat di simpulkan bahwa data kemiskinan penduduk dapat di klasterkan dengan beberapa metode yaitu: Algoritma *Hard C-Means*, *Latten Class Cluster Analysis*, Algoritma *Fuzzy C-Means*, *Regresi Panel*, dan Algoritma *K-Means* karena algoritma ini memiliki tingkat akurasi yang lebih tinggi yaitu sebesar 90,3% (Gusti Ngurah Wisnu Paramartha, dkk:2017). Dengan masalah data kemiskinan yang ada di Provinsi Jambi tidak akurat maka penulis melakukan penelitian dengan data tersebut menggunakan algoritma *K-Means*. Pengolahan data tersebut nantinya akan digunakan untuk menentukan keluarga yang berhak menerima bantuan, untuk menentukan kriteria tersebut maka penulis melakukan perhitungan dengan beberapa atribut seperti: tanggungan dalam keluarga, anggota keluarga yang cacat, lanjut usia, penguasaan bangunan, luas bangunan, jenis lantai, jenis atap, jenis dinding, fasilitas tempat buang air besar, tempat pembuangan akhir tinja, sumber air minum, sumber penerangan utama, apakah memiliki dapur, jenis bahan bakar untuk memasak, pekerjaan utama kepala keluarga, dan pendapatan perbulan.