SKRIPSI

EVALUASI DAN PEMETAAN KELOMPOK PASIEN PENDERITA STROKE DENGAN TEKNIK CLUSTERING DAN ASOSIASI (STUDI KASUS : RUMAH SAKIT UMUM DAERAH POLEWALI MANDAR)

Disusun dan diajukan oleh :

RINI FEBRIANTI RIAL
D121171301

PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023

LEMBAR PENGESAHAN SKRIPSI

EVALUASI DAN PEMETAAN KELOMPOK PASIEN PENDERITA STROKE DENGAN TEKNIK CLUSTERING DAN ASOSIASI (STUDI KASUS : RUMAH SAKIT UMUM DAERAH POLEWALI MANDAR)

Disusun dan diajukan oleh

RINI FEBRIANTI RIAL
D121171301

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 28 Juni 2023 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping,
Dr. Amil Ahmad Ilham, S.T., M.IT.	Anugrayani Bustamin, S.T., M.T.
Nip. 197310101998021001	Nip. 199012012018074001

Ketua Program Studi,

Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.
Nip. 19750716 200212 1 004

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;
Nama : Rini Febrianti Rial
NIM : D121171301
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Evaluasi Dan Pemetaan Kelompok Pasien Penderita Stroke Dengan Teknik Klustering Dan Asosiasi (Studi Kasus: Rumah Sakit Umum Daerah Polewali Mandar)

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

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

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.

Gowa, 3 Juli 2023
Yang Menyatakan

Rini Febrianti Rial

ABSTRAK

RINI FEBRIANTI RIAL . *Evaluasi Dan Pemetaan Kelompok Pasien Penderita Stroke Dengan Teknik Clustering Dan Asosiasi (Studi Kasus Rumah Sakit Umum Daerah Polewali Mandar)* (dibimbing oleh Amil Ahmad Ilham dan Anugrayani Bustamin)

Stroke merupakan salah satu penyebab kematian dan kecacatan neurologis yang utama di seluruh dunia. Stroke menyebabkan penderitanya kehilangan kemampuan dan keterampilan. Merujuk pada data rekam medik pasien Rumah Sakit Umum Daerah Polewali Mandar berjumlah 500 data maka penelitian ini bertujuan untuk mengevaluasi dan memetakan karakteristik gejala pasien penyakit stroke. Penelitian ini menggunakan 2 metode yaitu *clustering* yang dimana kita dapat melihat pengelompokkan pasien penyakit stroke dengan menggunakan algoritma *K-Means* dengan nilai k yang merupakan jumlah klaster ditentukan menggunakan metode *Elbow* dan SSE (70,940554) dan nilai *Silhoutte Score* (0,64) dan menghasilkan 7 klaster. Metode asosiasi digunakan untuk mencari keterkaitan antar variabel untuk melihat pola gejala pasien penyakit *Stroke*. Algoritma *apriori* sebagai metode *asosiasi* digunakan dengan 2 parameter yaitu *min support* 70% dan *minimum confidence 85%* yang menghasilkan 7 *rules* dengan keterkaitan antara variabel yang cukup tinggi.

Kata Kunci : *Stroke, Clustering, Asosiasi, K-Means, Apriori, Minimum Support, Minimum Confidence, Silhouette Score.*

ABSTRACT

RINI FEBRIANTI RIAL . *Evaluation and Mapping of Stroke Patient Groups Using Clustering and Association Techniques (Case Study of Polewali Mandar Regional General Hospital)* (supervised by Amil Ahmad Ilham dan Anugrayani Bustamin)

*Stroke is one of the leading causes of death and neurological disability worldwide. Stroke causes sufferers to lose abilities and skills. Referring to the medical record data of Polewali Mandar Regional General Hospital patients totaling 500 data, this study aims to evaluate and characterize the symptoms of stroke patients. This study uses 2 methods, namely clustering where we can see the grouping of stroke patients using the K-Means algorithm with a value of k which is the number of clusters determined using the Elbow and SSE methods (*70.940554*) and the value of the Silhouette Score (0.64) and produces 7 clusters. The association method is used to look for interrelationships between variables to see the pattern of symptoms of stroke patients. The a priori algorithm as an association method is used with 2 parameters, namely 70% min support and 85% minimum confidence which produces 7 rules with a fairly high interrelationship between variables.*

Keywords: Stroke, Clustering, Association, K-Means, Apriori, Minimum Support, Minimum Confidence, Silhouette Score.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
NHS	*Non Hemoragik Stroke*
HS	*Hemoragik Stroke*
KDD	*Knowledge Discovery from Database*
SSE	*Sum of Squared Error*
FS	*Feature Selection*
IS	*Instance Selection*
PCA	*Principal Component Analysis*
HT	*Hipertensi*
DM	*Diabetes Melitus*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah *subhanahu wa ta'ala,* yang telah memberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir dengan judul**"Evaluasi Dan Pemetaan Kelompok Pasien Penderita *Stroke* Dengan Teknik Clustering Dan Asosiasi (Studi Kasus Rumah sakit Umum Daerah Polewali Mandar)".**

Penulis menyadari bahwa penyusunan dan penulisan tugas akhir ini tidak dapat terselesaikan dengan baik tanpa adanya bantuan, bimbingan serta dukungan dari berbagai pihak. Oleh karena itu , penulis ingin menyampaikan ucapan terima kasih banyak kepada

1. Allah *subhanahu wa ta' ala,* atas segala rahmat bantuan serta karunianya diberikan kepada penulis hingga saat ini.
2. Kedua orang tua penulis, Bapak Rial Hasan dan Ibu Jasmurni yang selalu memberikan dukungan, doa, dan semangat yang tiada hentinya.
3. Bapak Dr. Amil Ahmad Ilham, S.T., M.IT. selaku pembimbing I dan ibu Anugrayani Bustamin, S.T., M.T. selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis menyelesaikan tugas akhir.
4. Bapak Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng. selaku ketua departemen teknik informatika Universitas Hasanuddin.
5. Segenap staf dan dosen Departemen Teknik Informatika, Fakultas Teknik Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir.
6. Adik-adik penulis Sri Wahyuni Rial , Muh. Afif Rial, Tarisha Ramadhani dan Auliah Syahrani yang telah memberikan dukungan kepada penulis untuk tetap semangat dalam menyusun tugas akhir.
7. Six Girl sahabat penulis Jum, Priska, Reka, Suci, dan Ilmi yang selalu membersamai di setiap proses dan moment selama masa perkuliahan di Teknik Informatika. Serta Fitri dan Aries yang juga selalu membantu penulis dalam pengerjaan tugas akhir ini.

8. Penghuni Lab UBICON yang senantiasa membantu dan mendengar keluhan penulis selama menyelesaikan tugas akhir.
9. Teman-teman RECOGNIZER atas dukungan, bantuan, dan semangat yang diberikan selama ini.
10. Serta berbagai pihak atas segala dukungan dan bantuannya yang tidak dapat penulis tuliskan satu persatu.

Akhir kata, penulis berharap semoga Tuhan Yang Maha Esa berkenan membalas segala kebaikan dari semua pihak yang telah banyak membantu. Penulis menyadari masih terdapat kekurangan dalam penyusunan Laporan Tugas Akhir ini baik isi maupun penyajian. Oleh karena itu penulis mengharapkan adanya saran serta masukan yang membangun demi kesempurnaan laporan ini. Penulis berharap semoga Laporan Tugas Akhir ini dapat memberi manfaat bagi para pembaca dan semua pihak. Amiin.

Makassar, 21 Maret 2023

Penulis,

Rini Febrianti Rial

BAB I
PENDAHULUAN

1.1 Latar Belakang

Stroke adalah suatu keadaan dimana ditemukan tanda-tanda klinis yang berkembang cepat berupa defisit neurologik fokal dan global, yang dapat memberat dan berlangsung lama selama 24 jam atau lebih dan atau dapat menyebabkan kematian tanpa adanya penyebab lain yang jelas selain vascular. stroke terjadi apabila pembuluh darah otak mengalami penyumbatan atau pecah. Akibatnya sebagian otak tidak mendapatkan pasokan darah yang membawa oksigen yang diperlukan sehingga mengalami kematian sel/jaringan. Selain itu, penyakit stroke merupakan penyebab kematian kedua dan penyebab disabilitas ketiga di dunia. Stroke dapat dialami oleh siapa saja. Ada beberapa penyakit yang dapat meningkatkan risiko stroke yaitu hipertensi, diabetes, kolesterol tinggi, obesitas, penyakit jantung, kelainan darah, dan mempunyai riwayat serangan jantung adapun gejala yang ditimbulkan pada penyakit stroke (Kementerian Kesehatan RI,2019).

Terdapat dua jenis stroke yaitu stroke perdarahan atau stroke hemoragik dan stroke non perdarahan disebut stroke iskemik. Insiden stroke karena sumbatan (iskemik) antara 70-80% dan stroke karena perdarahan (hemoragik) sebesar 15-30%. Stroke iskemik disebabkan antara lain karena trombosis otak (penebalan dinding arteri) dan emboli, sedangkan stroke hemoragik dapat disebabkan oleh aneurisma dan angioma (Saefuloh M, & Wayunah,2016).

Prevalensi stroke berdasarkan terdiagnosis nakes dan gejala tertinggi terdapat di Sulawesi Selatan (17,9%), DI Yogyakarta (16,9%), Sulawesi Tengah (16,6%), diikuti Jawa Timur sebesar 16 per mil, dan Sulawesi Barat (15,5%) (Riskesdas, 2013) Sulawesi Barat termasuk dalam 5 provinsi terbesar yang terdiagnosis memiliki penderita stroke. (Riskesdas, 2013). Profil kesehatan Indonesia menunjukkan stroke menjadi penyebab kematian

nomor satu di rumah sakit umum di Indonesia pada tahun 2005. Situasi derajat kesehatan di Polewali Mandar, angka kesakitan (morbidity) dengan penyakit stroke termasuk sepuluh penyakit terbesar dan berada pada urutan ke delapan rawat inap di RSUD Polewali (Dinkes, 2010).

Berdasarkan data rekam medik RSUD Polewali Mandar pada tahun 2017-2022 penyakit stroke berjumlah 500 pasien yang terdiri dari 406 pasien dengan Non Hemoragik Stroke (NHS) dan 96 pasien dengan Hemoragik Stroke (HS). Oleh sebab itu, dengan banyaknya kasus penyakit stroke di Indonesia terutama di Provinsi Sulawesi Barat Kabupaten Polewali Mandar, maka penelitian ini dilakukan untuk membantu dalam melakukan analisis dan pemetaan karakteristik terhadap penyakit stroke pada pasien menggunakan data *mining*, sehingga diharapkan dapat memberikan informasi sebagai dasar untuk melakukan tindakan yang diperlukan dalam pengendalian pencegahan kasus stroke.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang dijelaskan, maka rumusan masalah pada tugas akhir ini adalah:

a. Bagaimana mengidentifikasi kelompok pasien penderita stroke RSUD Polewali Mandar?
b. Bagaimana menentukan faktor-faktor terkait penyebab stroke pada setiap kelompok pasien penderita Stroke RSUD Polewali Mandar ?
c. Bagaimana memvisualisasikan hasil cluster pasien penderita stroke RSUD Polewali Mandar?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini antara lain:

a. Untuk mengidentifikasi kelompok pasien penderita stroke RSUD Polewali Mandar

b. Untuk menentukan faktor-faktor terkait penyebab stroke pada setiap kelompok pasien penderita Stroke RSUD Polewali Mandar
c. Untuk memvisualisasikan hasil cluster pasien penderita stroke RSUD Polewali Mandar

1.4 Manfaat Penelitian

Dengan dilakukannya penelitian ini, diharapkan manfaat yang didapatkan yaitu memberikan informasi terkait dengan karakteristik pasien stroke yang ditangani RSUD Polewali Mandar sehingga diharapkan dapat menjadi dasar dalam penetapan program pencegahan stroke yang tepat. Selain itu, penelitian ini juga dapat menjadi acuan dan bahan bacaan bagi pengembangan selanjutnya.

1.5 Ruang Lingkup

Untuk bagian ruang lingkup pada penelitian ini, penulis membagi menjadi beberapa poin:

a. Pengambilan data dilakukan dengan menggunakan data rekam medik pasien stroke Iskemik, dan stroke Hemoragik di Rumah Sakit Umum Daerah Polewali Mandar
b. Visualisasi hasil Clustering akan ditampilkan berbentuk pemetaan
c. Variabel-variabel penelitian terdiri dari :
    - Usia pasien
    - Tekanan Darah
    - Detak Jantung
    - Kolesterol
    - Keluhan Utama
    - Riwayat Penyakit
    - Jenis *Stroke*
    - Alamat
    - Gula
    - Jenis Kelamin

1.6 Sistematika Penulisan

Pada bagian ini memberikan gambaran singkat mengenai isi tulisan secara keseluruhan pada tugas akhir ini, maka akan diuraikan beberapa tahapan dari penulisan secara sistematis, yaitu:

BAB 1 PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, metode penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini membahas landasan teori yang digunakan untuk menganalisis masalah yang akan diteliti serta hal-hal lain yang berhubungan dengan variabel-variabel data yang digunakan, *Data Mining ,Clustering, Association Rules, Algoritma K-Means, Algoritma Apriori.*

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi mengenai, waktu dan lokasi penelitian, instrumen penelitian, pengumpulan data, penerapan metode penelitian, penerapan algoritma, teknik pengolahan data, serta informasi akhir berupa pengetahuan

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang pembahasan hasil dari implementasi algoritma dan sistem yang berhasil dibangun.

BAB V PENUTUP

Pada bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut.

BAB II
TINJAUAN PUSTAKA

2.1 *Stroke*

Stroke adalah suatu keadaan dimana ditemukan tanda-tanda klinis yang berkembang cepat berupa defisit neurologik fokal dan global, yang dapat memberat dan berlangsung lama selama 24 jam atau lebih dan atau dapat menyebabkan kematian tanpa adanya penyebab lain yang jelas selain vascular. Stroke terjadi apabila pembuluh darah otak mengalami penyumbatan atau pecah. Akibatnya sebagian otak tidak mendapatkan pasokan darah yang membawa oksigen yang diperlukan sehingga mengalami kematian sel/jaringan. Selain itu, penyakit stroke merupakan penyebab kematian kedua dan penyebab disabilitas ketiga di dunia stroke dapat dialami oleh siapa saja. Ada beberapa penyakit yang dapat meningkatkan risiko stroke yaitu hipertensi, diabetes, kolesterol tinggi, obesitas, penyakit jantung, kelainan darah, dan mempunyai riwayat serangan jantung adapun gejala yang ditimbulkan pada penyakit stroke (Kementerian Kesehatan RI,2019).

Stroke merupakan salah satu penyakit serebrovaskular dan penyebab utama kematian di indonesia, jumlah penderita stroke di bawah usia 45 tahun di seluruh dunia terus meningkat. Kematian fisik akibat Stroke diperkirakan akan meningkat dengan kematian akibat penyakit jantung dan kanker. Stroke adalah penyebab kematian ketiga paling umum di Amerika Serikat dan penyebab utama kecacatan permanen (Handayani & Dominica,2019).

Ada dua jenis stroke, yaitu stroke iskemik dan stroke hemoragik. Stroke iskemik merupakan komplikasi dari beberapa penyakit pembuluh darah, ditandai dengan penurunan tekanan darah secara tiba-tiba, takikardia, kulit pucat dan pernapasan tidak teratur, sedangkan stroke hemoragik biasanya disebabkan oleh pendarahan intrakranial, dantekanan darah sistoliknya meningkat (Nasution,2019)

Dari semua definisi stroke di atas dapat diambil kesimpulan bahwa stroke adalah suatu serangan mendadak yang terjadi di otak dan dapat mengakibatkan kerusakan pada sebagian atau secara keseluruhan dari otak yang disebabkan oleh

gangguan peredaran pada pembuluh darah yang mensuplai darah ke otak, biasanya berlangsung lebih dari 24 jam. Jadi, batasan stroke adalah segala sesuatu gangguan pada otak yang disebabkan oleh gangguan peredaran darah ke otak, bukan karena kecelakaan atau trauma di otak.

Stroke bisa dialami siapa saja dan umur berapa saja , kategori Menurut Depkes, umur seseorang dikategorikan ke beberapa tingkatan yang tentunya hal tersebut sudah diperhitungkan sebelumnya. Adapun kategori umur menurut WHO yaitu balita 0-5 tahun, kanak-kanak 5-11 tahun, remaja awal 12-16 tahun, remaja akhir 17-25 tahun, dewasa awal 26-35 tahun, dewasa akhir 36-45 tahun, lansia awal 46-55, lansia akhir 56-65, manula di atas 65 tahun.

2.2 Data Mining

Data *mining* adalah sebuah proses untuk menemukan informasi yang berguna dalam penyimpanan data yang besar. Teknik data *mining* dikerahkan untuk menjelajahi kumpulan data dalam skala yang besar untuk menemukan pola baru dan berguna yang tidak diketahui (Tan *et al.*, 2019).

Data *mining* telah ada sejak tahun 1930-an, ditemukan oleh seorang ahli matematika, ahli logika, dan seorang kriptografer, bernama Alan Turing. Menurut Jiawei Han (Jiawei Han *et al*, 2012), data *mining* yang merupakan bagian dari *machine learning* menyelidiki bagaimana komputer dapat belajar atau meningkatkan kinerjanya berdasarkan data.

Berikut adalah beberapa pendekatannya:

a. Supervised Learning

Supervised Learning adalah sebuah pendekatan untuk data yang memiliki informasi kelas/label. Label merupakan variabel yang menjadi identitas dari tiap data dalam kumpulan data, sehingga tujuan dari pendekatan ini ialah mengklasifikasi suatu data ke data yang sudah ada.

b. Unsupervised Learning

Salah satu penerapan *unsupervised learning* adalah klasterisasi. Proses belajar disebut sebagai *unsupervised* (tidak diawasi) sebab set data yang diberikan tidak memiliki label, sehingga keluaran dari unsupervised learning tidak dapat

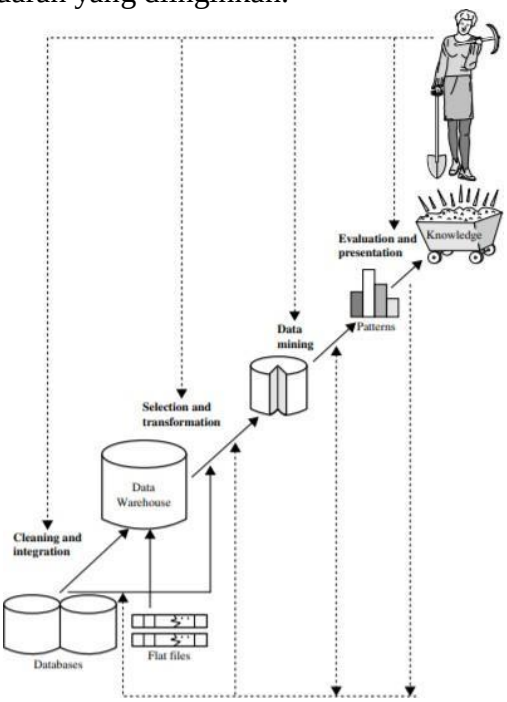
ditebak karena model tersebut belajar sendiri untuk menemukan informasi dari set data yang telah diberikan.

c. Semi-supervised learning

Semi-supervised learning adalah kombinasi antara supervised dan unsupervised learning. Data yang diberikan pada pendekatan ini memiliki kombinasi antara data yang berlabel dan tidak berlabel.

d. *Active learning*

Active learning adalah kasus yang khusus atau spesial pada *machine learning*, yang membuat mesin meminta peneliti atau pengguna secara interaktif memberi label data dengan keluaran yang diinginkan.



Gambar 1 Menampilkan tahapan *data mining*. Jiawei Han *et al* (2011)

Menyebutkan proses dari *Data Mining* atau *Knowledge Discovery from Database* (KDD) yang terbagi ke dalam 7 tahap, yaitu sebagai berikut:

a. Data *Cleaning* : tahap untuk membersihkan data yang hilang , *noise* dan tidak konsisten.
b. Data *Integration* : tahap dimana beberapa sumber data dapat digabungkan.
c. Data *Selection :* tahap untuk data yang relevan dengan analisis diambil dari *database.*

d. Data *Transformation :* tahap untuk data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk penambanngan dengan melakukan operasi *summary* atau *aggregation.*
e. Data *Mining :* tahap penting dimana metode cerdas diterapkan untuk mengekstrak pola data.
f. Pattern *evaluation :* tahap untuk mengidentifikasi pola yang benar-benar menarik dan mewakili pengetahuan berdasarkan ukuran keterkaitannya (*distance/interestingness measure).*
g. *Knowledge presentation :* tahap dimana teknik gambaran visualisasi dan pengetahuan digunakan untuk menyajikan pengetahuan kepada pengguna

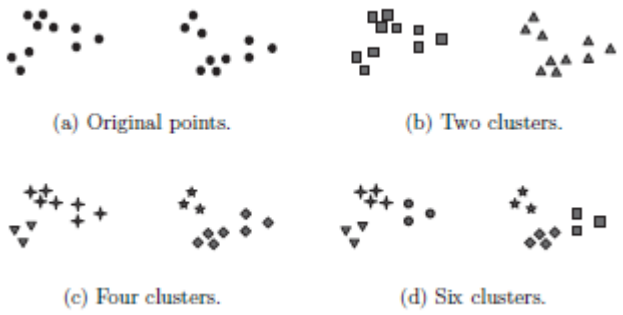
2.3 Clustering

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).

Daniel (2015) menyebutkan *clustering* mengacu pada pengelompokan data, observasi, atau kasus ke dalam kelas-kelas objek serupa. *Cluster* adalah kumpulan data yang mirip satu sama lain dan berbeda dengan data di *cluster* lain. *Clustering* berbeda dari klasifikasi dimana tidak ada variabel target *class* untuk pengelompokan. Tugas *clustering* tidak mencoba untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel target.

Sebagai gantinya, algoritma *clustering* berusaha untuk mengelompokkan seluruh kumpulan data ke dalam sub kelompok atau kelompok yang relatif

homogen, dimana kesamaan dalam *cluster* yang sama dimaksimalkan, dan kesamaan dalam *cluster* yang berbeda ini diminimalkan.



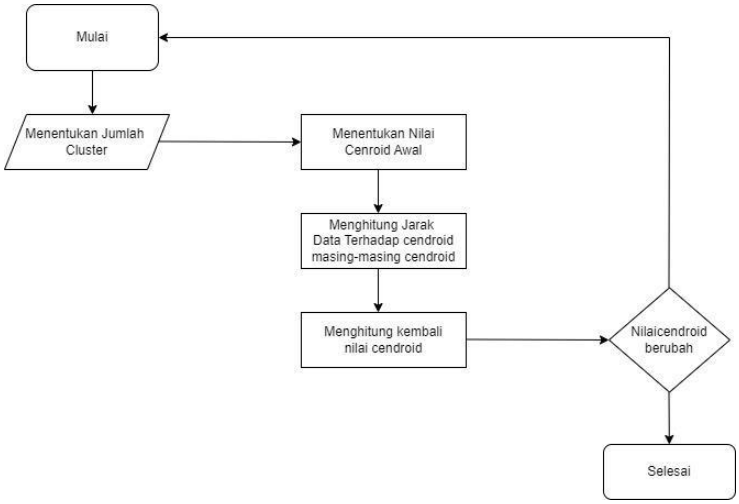
Gambar **2** Menunjukkan ilustrasi 3 cara berbeda untuk melakukan *clustering*

2.3.1 Algoritma K-Means

K-Means merupakan salah satu metode pengelompokkan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam kelompok yang lain. Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengelompokkan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok. (Prasetyo, 2012).

Ide dasar algoritma k-means sangatlah sederhana, yaitu meminimalkan Sum of Squared Error (SSE) antara objek-objek data dengan sejumlah k centroid. Algoritma k-means bekerja dengan empat langkah, yang diilustrasikan dalam pseudo code di bawah ini. Pertama, dari himpunan data yang akan di klasterisasi, dipilih sejumlah k objek secara acak sebagai centroid awal. Kedua, setiap objek yang bukan centroid dimasukkan ke klaster terdekat berdasarkan ukuran jarak tertentu. Ketiga, setiap centroid diperbarui berdasarkan rata-rata dari objek yang ada di dalam setiap klaster. Keempat, langkah kedua dan ketiga tersebut diulang

ulang (diiterasi) sampai semua centroid stabil atau konvergen, dalam arti centroid yang dihasilkan dalam iterasi saat ini sama dengan semua centroid yang dihasilkan pada iterasi sebelumnya. (Suryanto, 2019).



Gambar 3 Tahap Data Mining (Han et al., 2012)

Tahapan proses analisis dataset menggunakan algoritma *K-Means* dengan langkah-langkah pemodelan *k-means clustering* dari mulainya digunakan hingga menghasilkan pengelompokkan data pada akhir (pada saat iterasi ke-n saat tidak terjadi perubahan pusat (*cluster/centroid*) dengan taksiran bahwa tolak ukur terhadap input adalah jumlah *dataset* sebanyak n data dan jumlah *inisialisasi centroid* (pusat klaster) hingga pada saat *iterasi* ke-n saat tidak terjadi perubahan pusat *cluster/centroid* menggunakan persamaan *Euclidean Distance* Algoritma *K-Means Clustering* (R. A. Malik et al., 2018) Sebagai berikut :

1. Memasukkan data awal yang akan diklaster atau dikelompokan
2. Tentukan nilai K sebagai jumlah cluster yang akan dibentuk
3. Inisialisasi k dari data sebanyak jumlah klaster secara acak sebagai pusat klaster (*Centroid*).
4. Hitung jarak antar masing-masing data dengan pusat klaster (*Centroid*), dengan menggunakan persamaan Euclidean Distance

$$d(x+y) \quad = \sum_{k=0}^{n} \quad \sqrt{|xn - yn|^2|} \qquad (1)$$

Atau

$$d(x,y)\sum_{k=0}^{n} \quad \sqrt{\sqrt{(x1i-y1i)^2}+\ (x2i-y1i)^2+\cdots+\sqrt{(x1i-y1i)^2}+\ (x2n-y2i)^2} \tag{2}$$

Dimana:

d (x, y) = jarak data x ke pusat klaster j
xn = data ke n pada atribut klaster x
yn = titik pusat ke n pada atribut y
n = banyaknya objek

5. Kelompok setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya.
6. Tentukan posisi pusat klaster (*centroid*) baru (k)

Jika pusat *cluster* tidak berubah maka proses klaster telah selesai. Jika belum maka ulangi langkah ke-4 sampai pusat *cluster (centroid)* tidak berubah lagi.

2.3.2 Elbow Method

Metode Elbow merupakan metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik melihat persentase perbandingan antara jumlah cluster yang akan membentuk siku-siku sebuah titik. Metode ini memberikan ide/gagasan dengan memilih nilai cluster kemudian menjumlahkan nilai dari cluster yang akan digunakan sebagai model data dalam menentukan cluster terbaik. Dan selain itu persentase dari perhitungan yang dihasilkan adalah perbandingan antara jumlah cluster yang ditambahkan. Hasil persentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumbernya informasi. Jika nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilai yang mengalami penurunan terbesar maka nilai cluster adalah yang terbaik. Untuk mendapatkan perbandingannya adalah menghitung SSE (Sum of Square Error)

dari setiap nilai cluster. Karena semakin besar jumlah cluster K, nilai SSE akan semakin kecil (Nainggolan dkk. 2019)

2.3.3 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:

$$SSE = \sum_{i=1}^{n} (d_i - C_i)^2 \quad (3)$$

Keterangan:

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

n = jumlah data

di = nilai data ke-i

ci = nilai *centroid* cluster 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).

2.3.4 Silhouette Coefficient

Untuk mengukur kualitas suatu *clustering*, kita bisa menggunakan nilai koefisien siluet rata-rata dari semua objek dalam kumpulan data. Koefisien siluet dan ukuran intrinsik lainnya juga dapat digunakan dalam metode siku untuk secara heuristik menurunkan jumlah klaster dalam kumpulan data dengan mengganti jumlah varians dalam klaster. Nilai koefisien siluet adalah antara -1 dan 1. Ketika nilai koefisien siluet o mendekati 1, klaster yang mengandung o kompak dan o jauh dari klaster lain, yang merupakan kasus yang lebih disukai. Namun, ketika nilai koefisien siluet negatif (yaitu, b(o) < a(o)), ini berarti bahwa, dengan harapan, o lebih dekat ke objek di klaster lain daripada ke objek di klaster yang sama dengan o. Dalam banyak kasus, ini adalah situasi yang buruk dan harus dihindari (Pei dkk., 2012)

Menghitung nilai *Silhouette Coefficient* untuk setiap data ke-i (Kaufman 2009)

$$S(i) = \frac{b((i)-a(i)}{(bi,\ ai)} \qquad (4)$$

S(i) = Silhouette Coefficient

a(i)= rata-rata jarak antara titik *i* dengan seluruh objek yang berada pada cluster yang sama

b(*i)*= rata-rata jarak antara titik *i* dengan seluruh objek yang berada pada cluster yang berbeda

Nilai Silhouette Coefficient berada pada rentan (-1) hingga (1). Semakin tinggi nilainya maka semakin bagus pula kualitasnya.

Ukuran nilai *koefisien silhouette* menurut Kaufman dan Rousseuw (1990) adalah seperti yang ditunjukkan pada tabel berikut:

Tabel. 1 *Interpretasi nilai silhoutte score (Kauffan dan Rousseuw. 1990)*

Koefisien Silhoutte	Interpretasi
0.71-1	Terdapat ikatan yang sangat baik (*strong structure)* antara objek dan kelompok yang terbentuk
0.51-0.70	Terdapat ikatan yang cukup baik

	(*medium structure)* antara objek dan kelompok yang terbentuk
0.26-0.5	Terdapat ikatan yang lemah (*weak structure)* antara objek dan kelompok yang terbentuk
$\leq$ 0.25	Tidak terdapat ikatan antara objek dan kelompok yang terbentuk

2.4 Asosiasi

Ide dari aturan asosiasi adalah untuk memeriksa semua kemungkinan hubungan if-then antar item dan memilih hanya yang paling mungkin (most likely) sebagai indikator dari hubungan ketergantungan antar item. Biasanya digunakan istilah antecedent untuk mewakili bagian "jika" dan consequent untuk mewakili bagian "maka". Dalam analisis ini. Antecedent dan consequent adalah sekelompok item yang tidak punya hubungan secara bersama(Listriani et al., 2016).

Metodologi dasar aturan asosiasi dijelaskan sebagai berikut(Listriani et al., 2016):

1. Pembentukan Pola Frekuensi Tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam suatu *database*. Nilai *support* adalah nilai penunjang atau persentase kombinasi sebuah item bersamaan dalam suatu *database*. Semakin besar nilai *support* menandakan semakin banyak data pendukung yang ditemukan dalam *database*. Nilai *support* sebuah item diperoleh dari persamaan:

$$Support\ A = \frac{Jumlah\ kejadian\ mengandung\ A}{total\ kejadian} x\ 100\% \qquad (5)$$

Nilai *support* dari 2 item diperoleh dari persamaan:

$$Support\ (A, B) = \frac{Jumlah\ kejadian\ mengandung\ A\ dan\ B}{total\ kejadian} \text{x } 100\% \qquad (6)$$

Sementara nilai *support* untuk 3 itemset diperoleh dari persamaan:

Support (A,B, C) = $\frac{jumlah\ kejadian\ mengandung\ A\ ,B\ dan\ C}{total\ kejadian}$ x 100% (7)

2. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, maka akan dibentuk aturan asosiasi yang memenuhi syarat minimum untuk *confidence.* Nilai *confidence* adalah nilai keyakinan berupa kuatnya hubungan antar item yang didapatkan. Semakin besar nilai *confidence* menandakan semakin besar kemungkinan kombinasi item muncul secara bersamaan. Persamaan dari nilai *confidence* dengan menggunakan kondisi "jika A maka B" adalah sebagai berikut. Nilai *confidence* adalah nilai keyakinan berupa kuatnya hubungan antar item yang didapatkan. Semakin besar nilai *confidence* menandakan semakin besar kemungkinan kombinasi item muncul secara bersamaan. Persamaan dari nilai *confidence* dengan menggunakan kondisi "jika A maka B" adalah sebagai berikut.

Persamaan untuk menentukan Confidence :

$$Confidence\ (A \rightarrow B) = \frac{Jumlah\ kejadian\ mengandung\ A\ dan\ B}{Total\ kejadian\ mengandung\ A} x100\%$$

(8)

2.4.1 Algoritma Apriori

Algoritma apriori adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan *frequent itemsets* untuk aturan asosiasi *Boolean* (Defit, 2013). Algoritma *apriori* termasuk jenis aturan asosiasi pada *data mining*. Selain apriori, yang termasuk pada golongan ini adalah metode *Generalized Rule Induction* dan Algoritma *Hash Based*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market analysis*.

Algoritma *apriori* merupakan algoritma yang paling populer dikenal sebagai dengan paradigma *general and test*, yaitu pembuatan kandidat kombinasi

item yang mungkin berdasarkan aturan tertentu lalu diuji apakah kombinasi item tersebut memenuhi syarat support minimum (M. Malik et al., 2019)

Beberapa istilah yang sering digunakan dalam algoritma *apriori* antara lain (Fitriyanto, 2017):

1. *Support* (dukungan): probabilitas (kemungkinan) pelanggan membeli beberapa produk secara bersamaan dari seluruh transaksi. *Support* untuk aturan "X, Y" adalah probabilitas atribut atau kumpulan atribut X dan Y yang terjadi bersamaan.
2. *Confidence* (tingkat kepercayaan): probabilitas kejadian beberapa produk dibeli bersamaan dimana salah satu produk sudah pasti di beli.
3. *Minimum Support*: parameter yang digunakan sebagai batasan frekuensi kejadian atau *support count* yang harus dipenuhi suatu kelompok data untuk dapat dijadikan aturan.
4. *Minimum Confidence*: parameter yang mendefinisikan minimum *level* dari *confidence* yang harus dipenuhi oleh aturan yang berkualitas.
5. *Itemset*: kelompok produk
6. Kandidat *Itemset (C*k*): itemset-itemset* yang akan dihitung *support count*-nya.
7. *Frequent itemset*(Fk): *itemset* yang sering terjadi, atau *itemset-itemset* yang sudah melewati batas *minimum support* yang telah ditentukan.

Penggunaan Algoritma apriori membantu dalam proses pengambilan keputusan, dengan Algoritma ini dapat membantu dalam membentuk kandidat kombinasi item yang mungkin, kemudian dilakukan pengujian apakah kombinasi tersebut memenuhi parameter *support* dan *confidence minimum* yang merupakan nilai minimum yang diberikan oleh pengguna (Nurajizah, 2019).

Cara kerja Algoritma *apriori d*apat dijelaskan sebagai berikut (Haris, 2016):

1. Tentukan *minimum support*
2. Iterasi 1: hitung *item-item* dari *support* (transaksi yang memuat seluruh item) dengan memindai database untuk 1-*itemset*, setelah *itemset* didapatkan, apabila telah memenuhi *minimum support,* 1-*itemset* tersebut akan menjadi pola frekuensi tinggi

3. Iterasi 2: untuk mendapatkan 2-*itemset*, perlu dilakukan kombinasi dari k-*itemset* sebelumnya. Kemudian pindai database sekali lagi untuk menghitung *item-item* yang memuat *support*. *Itemset* yang memenuhi *minimum support* akan dipilih sebagai pola frekuensi tertinggi dari kandidat.
4. Tetapkan nilai k-*itemset* dari *support* yang telah memenuhi nilai *minimum support* dari k-*itemset*.
5. Lakukan proses untuk *iterasi* selanjutnya hingga tidak ada lagi k-itemset yang memenuhi nilai *minimum support.*

2.4.2 Lift Ratio

Lift ratio adalah suatu ukuran (parameter) untuk mengetahui kekuatan aturan asosiasi yang telah dibentuk dari nilai *support* dan *confidence*. Nilai *lift ratio* biasanya digunakan sebagai penentu apalah aturan asosiasi valid atau tidak valid (Simanjuntak & Windarto, 2020).

Lift ratio digunakan untuk mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift Ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. Lift ratio dihitung berdasarkan rumus pada persamaan 2.5 (Wicaksono, 2015) berikut:

$$\text{Lift Ratio} = \frac{Support\ (A\ U\ B)}{Support\ (A)\ x\ Support\ (B)} \tag{9}$$

Lift ratio biasanya digunakan sebagai penentu apakah aturan keakuratan suatu asosiasi, nilai *lift* yang baik yaitu jika nilai *lift* > 1 maka pada nilai ini proses transaksi dikatakan valid, karena keterkaitan item bergantung positif (Rahardjo et al., 2021).

2.4.3 Conviction

Conviction adalah perhitungan untuk menentukan nilai akurasi minimum pada metode association rules. Pada proses ini dihitung performansi yaitu akurasi

untuk rule yang dihasilkan oleh sistem. Mengukur akurasi dari metode yang digunakan dengan rumus :

$$\text{Conviction (A ->B)} = \frac{1-Support\ B}{1-Confidence\ (A \rightarrow B)} \quad (10)$$

Nilai range pada conviction ini berada pada 0,5,…,1,…,∞ dengan ketentuan conviction dianggap memiliki nilai tak hingga (infinite) apabila nilai dari confidence (A→B) sama dengan 1. Jika conviction menghasilkan nilai rules yang semakin menjauh dari 1 bahkan sampai tak hingga, maka akan dianggap semakin akurat

2.5 Preprocessing

Preprocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah yang diproses oleh sistem. (Abdan,2017). Data awal sangat rentan memiliki ketidak konsistenan data, kesalahan *input* data, kesalahan pengetikan dan lain lain. Data awal berukuran terlalu besar sehingga proses data *mining* yang dilakukan kepada data tersebut akan menjadi efektif. Selain itu seringkali data pada dunia nyata tidak cukup akurat untuk menggambarkan kondisi yang terwakilkan oleh data, karena itu dalam beberapa sisi perlu adanya sistem yang menimbulkan nilai-nilai tertentu untuk meningkatkan akurasi data (Hermawan et al., 2011).

Ada 7 (tujuh) tahapan proses data *mining*, dimana 4 (empat) tahapan pertama disebut juga dengan data *preprocessing* (yang terdiri dari data *cleaning*, data *integration*, data *selection*, dan data *transformation*), yang dalam implementasinya membutuhkan sekitar 60% dari keseluruhan proses (Junaedi et al., 2011).

Hal mendasar yang harus disiapkan dalam penyiapan data, dapat dijelaskan (García et al., 2015) sebagai berikut:

- Data *Cleaning* (Membersihkan data)

Data *cleaning* atau data *cleansing* termasuk operasi data yang buruk, menyaring data yang salah dari kumpulan data dan mengurangi detail data yang tidak diperlukan. Perlakuan terhadap *missing* data (data hilang) dan *noise* data (data yang menyimpang / kebisingan data) terdapat dalam proses ini. Tugas data *cleaning* lainnya adalah mendeteksi perbedaan dan data kotor (fragmen dari data asli yang tidak masuk akal).

- Data *Integration* (Menggabungkan dan menyesuaikan data)

  Dalam proses mengintegrasi data, ini merupakan penggabungan data dari beberapa penyimpanan data. Proses ini harus dilakukan dengan hati-hati agar menghindari redudansi dan inkonsistensi dalam kumpulan data yang dihasilkan. Penggabungan data terutama dilakukan pada data yang memiliki representasi berbeda karena berbagai hal pada sebuah data.

- Data *Transformation* (Memberikan data yang akurat)

  Pada tahap *preprocessing* ini data yang telah diintegrasikan, kemudian dinormalisasi dan dilakukan proses generalisasi data, proses ini memastikan tidak adanya data yang berlebihan. Sehingga data akan dihimpun dalam sebuah tempat penyimpanan, yang dependensinya harus masuk akal, data juga akan ditransformasikan dalam bentuk yang sesuai.

  Pada tahap ini pula data akan dikonversikan atau diubah bentuknya. Contohnya seperti mengubah bentuk data angka menjadi suatu bentuk kategori yang berbeda. Tahapan transformasi data ini berguna untuk mengurangi jumlah data. Cara-cara transformasi antara lain: *aggregate* data, *smoothing*, *attribute construction*, normalisasi data dan *discretization* data.

- *Missing Data Imputation* (Menangani data yang hilang)

  Proses ini berisi pembersihan data, dimana tujuannya adalah mengisi variabel yang berisi *missing values* dengan beberapa data intuitif. Dalam sebagian besar kasus, menambahkan perkiraan yang masuk akal dari nilai data yang sesuai lebih baik daripada membiarkannya kosong.

- *Noise Identification* (Mendeteksi dan mengelolah *noise* (kebisingan))

  Pada tahap ini termasuk dalam langkah data *cleaning* dan juga dikenal sebagai pemulusan dalam transformasi data. Tujuan utamanya adalah

mendeteksi kesalahan acak atau varians dalam variabel yang diukur. Pada proses ini akan dideteksi noise daripada penghilangan noise.

- Data *Reduction*

  Pada tahapan ini, data reduksi atau mereduksi data. karena jumlah data yang besar, maka tingkat akurasinya pun dikhawatirkan akan rendah atau bahkan tidak akurat. Sebab itulah diperlukannya reduksi data, reduksi merupakan suatu proses untuk mengurangi jumlah data namun dengan mempengaruhi proses analisis data. Dengan menggunakan pengurangan data, maka akan menjadikan proses penyimpanan menjadi lebih efisien. Dalam kasus reduksi data, data yang dihasilkan biasanya mempertahankan struktur dan integrasi penting dari data asli, tetapi jumlah data dirampingkan. Proses pada reduksi data dapat dijelaskan sebagai berikut, dimana *Feature Selection* (FS) adalah mengurangi dimensi data. *Instance Selection* (IS) adalah menghapus contoh yang berlebihan dan bertentangan. *Discretization* adalah bagaimana cara menyederhanakan domain dari sebuah atribut, lalu yang terakhir ialah *Feature Extraction* dan/atau *Instance Generation* dimana bagaimana cara mengisi kekosongan data.

2.6 Principal Component Analysis (PCA)

Principal component analysis (PCA) atau disebut juga transformasi Karhunen-Loeve adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasikan *linear* sehingga terbentuk sistem koordinat baru dengan variasi maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan. Metode ini mengubah dari sebagian besar variabel asli yang saling berkorelasi menjadi satu himpunan variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi lagi). Prinsip dasar dari algoritma *Principal Component Analysis* adalah mengurangi satu set data namun tetap mempertahankan sebanyak mungkin variasi dalam set data tersebut. Secara matematis PCA mentransformasikan sebuah variabel yang berkorelasi ke dalam bentuk yang bebas tidak berkorelasi (Firliana et al., 2015).

Menurut jolliffe (2002), Prosedur pengerjaan *Principal Component Analysis* bertujuan untuk menyederhanakan dan menghilang *factor* yang kurang dominan dan kurang relevan tanpa mengurangi maksud dan tujuan dari data asli dari variabel (Nasution et al., 2019) .

2.7 Pemetaan

Pemetaan adalah pengelompokkan suatu kumpulan wilayah yang berkaitan dengan beberapa letak geografis wilayah yang meliputi dataran tinggi, pegunungan, sumber daya dan potensi penduduk yang berpengaruh terhadap sosial kultural yang memiliki ciri khas khusus dalam penggunaan skala yang tepat (Munir, 2012).

Peta adalah penggambaran dua dimensi pada bidang datar keseluruhan atau sebagian dari permukaan bumi yang diproyeksikan dengan perbandingan atau skala tertentu (Nasution, 2016).

Jadi, dari dua definisi diatas dan disesuaikan dengan penelitian ini maka pemetaan merupakan proses pengumpulan data untuk dijadikan sebagai langkah awal dalam pembuatan peta, dengan menggambarkan penyebaran kondisi alamiah tertentu secara meruang, memindahkan keadaan sesungguhnya ke dalam peta dasar, yang dinyatakan dengan penggunaan skala peta.

2.8 Penelitian Terkait

Berikut ini merupakan beberapa penelitian terkait dengan penelitian yang dilakukan:

1. **The Global kernel k-means Clustering Algorithm for Cerebral Infarction Classification**

   Penelitian ini menggunakan metode Clustering dan menggunakan algoritma K-Means untuk mengklasifikasikan apakah ada infark pada otak seseorang atau tidak. Dimana hasil yang didapatkan dalam proses klasifikasi ini menggunakan metode Clustering Algoritma K-Means ialah mencapai akurasi 78,06%

2. **Analisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Dengan Menggunakan Metode DBSCAN Clustering**

   Penelitian ini menggunakan Metode Clustering dan menggunakan Algoritma DBSCAN untuk menganalisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan. Dimana hasil menghasilkan 4 cluster yang menyimpan data-data penyakit pasien dengan karakteristik yang berdekatan

3. **Analisa Keterkaitan *Risk Factor Stroke* dengan Jenis *Stroke* yang Diderita Menggunakan Algoritma ECLAT**

   Penelitian ini menggunakan metode Asosiasi dan menggunakan Algoritma ECLAT untuk menganalisis keterkaitan *Risk factor Stroke* dengan Jenis *Stroke* . Dimana hasil yang didapatkan yaitu dimana *representative rule* dengan diagnosa *Stroke iskemik* memiliki akurasi terendah 76.47% dan akurasi tertinggi 95.23% sedangkan *representative rule* dengan diagnosa *Stroke hemoragik* memiliki tingkat akurasi terendah 6.66% dan akurasi tertinggi 24%.

4. **Pemodelan Data Mining Algoritma Apriori Dalam Sistem Analisa Pola Keterjangkitan Penyakit di Puskesmas**

   Penelitian ini menggunakan metode Asosiasi dan menggunakan Algoritma Apriori untuk menganalisis pola keterjangkitan penyakit di puskesmas. Dimana hasil yang didapatkan yaitu 15% pasien yang terjangkit Batuk juga terjangkit Flu, dan terendah didapatkan hasil Diare juga terjangkit Migrain sebesar 13%.

5. **Analisis Visualisasi dan Pemetaan data tanaman padi di Indonesia menggunakan Microsoft Power BI**

   Penelitian ini menggunakan Microsoft Power BI untuk memvisualisasikan data tanaman padi di indonesia. Hasil dari penelitian ini memperoleh informasi mengenai data tanaman padi dengan tingkat produksi tertinggi di indonesia.