

LAPORAN TAHUN TERAKHIR
PENELITIAN DASAR UNGGULAN PERGURUAN TINGGI
(PDUPT)



SISTEM DIAGNOSA PENYAKIT ANAK DI BAWAH LIMA TAHUN
(BALITA) MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS
(PCA) DAN ALGORITMA APRIORI

TAHUN KE – 2 DARI RENCANA 2 TAHUN

Dr. RIMULJO HENDRADI, S.Si., M.Si.	0002117101
INDAH WERDININGSIH, S.Si., M.Kom.	0017028004
PURBANDINI, S.Si., M.Kom.	0012077103
Ir. ELLY ANA, M.Si.	0004126209

DIBIYAI OLEH:
DIREKTORAT RISET DAN PENGABDIAN MASYARAKAT
DIREKTORAT JENDERAL PENGUATAN RISET DAN PENGEMBANGAN
KEMENTERIAN RISET, TEKNOLOGI, DAN PENDIDIKAN TINGGI
SESUAI DENGAN PERJANJIAN PENDANAAN PENELITIAN DAN PENGABDIAN
KEPADA MASYARAKAT
NOMOR: 122/SP2H/PTNBH/DRPM/2018

UNIVERSITAS AIRLANGGA
NOPEMBER 2018

**LAPORAN TAHUN TERAKHIR
PENELITIAN DASAR UNGGULAN PERGURUAN TINGGI
(PDUPT)**



MILIK
PERPUSTAKAAN
UNIVERSITAS AIRLANGGA
SURABAYA

KKC
KK
LP.45/19
Hen
S

**SISTEM DIAGNOSA PENYAKIT ANAK DI BAWAH LIMA TAHUN
(BALITA) MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS
(PCA) DAN ALGORITMA APRIORI**

TAHUN KE - 2 DARI RENCANA 2 TAHUN

Dr. RIMULJO HENDRADI, S.Si., M.Si.	0002117101
INDAH WERDININGSIH, S.Si., M.Kom.	0017028004
PURBANDINI, S.Si., M.Kom.	0012077103
Ir. ELLY ANA, M.Si.	0004126209

**DIBIYAI OLEH:
DIREKTORAT RISET DAN PENGABDIAN MASYARAKAT
DIREKTORAT JENDERAL PENGUATAN RISET DAN PENGEMBANGAN
KEMENTERIAN RISET, TEKNOLOGI, DAN PENDIDIKAN TINGGI
SESUAI DENGAN PERJANJIAN PENDANAAN PENELITIAN DAN PENGABDIAN
KEPADA MASYARAKAT
NOMOR: 122/SP2H/PTNBH/DRPM/2018**

**UNIVERSITAS AIRLANGGA
NOPEMBER 2018**

HALAMAN PENGESAHAN

Judul : Sistem Diagnosa Penyakit Anak di Bawah Lima Tahun (Balita) Menggunakan Principal Componen Analysis (PCA) dan Algoritma Apriori

Peneliti/Pelaksana
 Nama Lengkap : Dr RIMULJO HENDRADI, S.Si, M.Si
 Perguruan Tinggi : Universitas Airlangga
 NIDN : 0002117101
 Jabatan Fungsional : Lektor
 Program Studi : Sistem Informasi
 Nomor HP : 0818500836
 Alamat surel (e-mail) : rimuljohendradi@fst.unair.ac.id

Anggota (1)
 Nama Lengkap : INDAH WERDINGSIII S.Si, M.Kom
 NIDN : 0017028004
 Perguruan Tinggi : Universitas Airlangga

Anggota (2)
 Nama Lengkap : PURBANDINI S.Si, M.Kom
 NIDN : 0012077103
 Perguruan Tinggi : Universitas Airlangga

Anggota (3)
 Nama Lengkap : Ir ELLY ANA M.Si
 NIDN : 6004126209
 Perguruan Tinggi : Universitas Airlangga

Institusi Mitra (jika ada)
 Nama Institusi Mitra : -
 Alamat : -
 Penanggung Jawab : -
 Tahun Pelaksanaan : Tahun ke 2 dari rencana 2 tahun
 Biaya Tahun Berjalan : Rp 93,900,000
 Biaya Keseluruhan : Rp 193,670,000

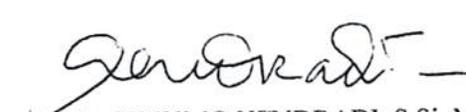


Mengetahui,
 Dekan FST



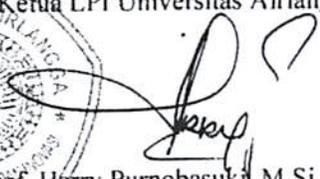
(Prof. Widi Darmanto, M.Si, Ph.D)
 NIP/NIK 196106161987011001

Kota Surabaya, 14 - 11 - 2018
 Ketua,



(Dr RIMULJO HENDRADI, S.Si, M.Si)
 NIP/NIK 197102111997021001

Menyetujui,
 Ketua LPI Universitas Airlangga



(Prof. Herry Purnobasuki, M.Si, Ph.D)
 NIP/NIK 196705071991021001

RINGKASAN

Tujuan utama dari penelitian ini membangun Sistem Diagnosa Penyakit Anak di Bawah Lima Tahun (Balita) Menggunakan *Principal Component Analysis* (PCA) dan Algoritma Apriori. *Principal Component Analysis* (PCA) digunakan untuk mereduksi fitur yang mempengaruhi penyakit anak di Bawah Lima Tahun (Balita) berdasarkan korelasi antar atribut. Klasifikasi penyakit anak di Bawah Lima Tahun dibagi menjadi 16 penyakit yaitu: Batuk, Pneumonia, Pneumonia Berat, Diare, Diare Dehidrasi Ringan, Diare Dehidrasi Berat, Diare Persisten, Diare Persisten Berat, Disentri, Demam, Demam dengan Tanda Bahaya Umum, Campak, Campak dengan Komplikasi Berat, Campak dengan Komplikasi, Demam mungkin Demam Berdarah Dengue (DBD), dan DBD.

Penelitian ini terdapat tiga tahapan yang harus dilakukan, yaitu reduksi variabel, identifikasi dan klasifikasi. Tahap identifikasi dilakukan pada tahun pertama dan klasifikasi dilakukan pada tahun kedua. Sedangkan reduksi variabel dilakukan di kedua tahun. Hasil dari reduksi digunakan untuk proses identifikasi dan klasifikasi.

Variabel yang digunakan adalah umur, jenis kelamin, berat badan, tinggi badan, dan 26 gejala. Data yang digunakan merupakan data rekam medis pasien dari balai kesehatan di Desa Soko, Kecamatan Glagah, Kabupaten Lamongan dan salah satu rumah sakit di Surabaya.

Reduksi variabel merupakan pengurangan variabel dengan mengeliminasi komponen yang lemah. Hasil dari reduksi variabel berupa variabel yang berpengaruh untuk proses identifikasi dan klasifikasi. Setelah mengetahui pola yang terjadi langkah selanjutnya adalah mengklasifikasikan penyakit tersebut. Klasifikasi digunakan untuk mendiagnosa penyakit balita. Klasifikasi menggunakan *K-Nearest Neighbor* (KNN) Classifier dan rule based system. Rule based system terdiri dari forward chaining dan CBR. CBR berfokus pada 3 perhitungan, yaitu Nearest Neighbor Similarity (NNS), Minkowski Distance Similarity (MDS) dan Euclidean Distance Similarity (EDS). Evaluasi system dilakukan dengan membandingkan nilai akurasi klasifikasi tanpa reduksi fitur dan menggunakan PCA dan KNN classifier, PCA dan Forward chaining, PCA dan Case Base Reasoning (CBR) classifier.

Berdasarkan hasil uji coba dengan menggunakan data sebanyak 600 data diperoleh akurasi yang optimal diperoleh sebesar 93% menggunakan CBR-MDS classifier menggunakan reduksi variabel. Pada penelitian ini diharapkan dapat dibangun system diagnosa penyakit balita yang dapat digunakan masyarakat dan membantu tenaga medis di balai kesehatan untuk mendiagnosa secara dini dari penyakit balita.

”

Kata kunci: Penyakit Balita, Classifikasi, Reduksi variabel, PCA, dan CBR

PRAKATA

Puji syukur terlimpah ke hadirat Tuhan Yang Maha Esa yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan Laporan Akhir dengan judul “Sistem Diagnosa Penyakit Anak di Bawah Lima Tahun (BALITA) Menggunakan Principal Component Analysis (PCA) dan Algoritma Apriori”.

Penulis mengucapkan terima kasih yang sebesar-besarnya, kepada Dirjen Pendidikan Tinggi (DIKTI) yang telah mempercayai dan membiayai Penelitian Dasar Unggulan Perguruan Tinggi (PUPT), penulis juga mengucapkan terima kasih kepada LPI Universitas Airlangga yang telah memfasilitasi penelitian ini sehingga dapat terlaksana dengan baik. Penulis juga mengucapkan terima kasih kepada Indah Werdiningsih.,S.Si., M.Kom, Purbandini.,S.Si.,M.Kom, dan Ir. Elly Anna.,M.Si sebagai tim penelitian PUPT, serta semua pihak yang telah bekerjasama dan membantu atas terselesainya penelitian ini.

Penulis berharap semoga isi dari Laporan Akhir ini dapat bermanfaat bagi kita semua. Kritik dan saran yang membangun sangat diharapkan bagi kemajuan yang lebih baik di masa yang akan datang.

Surabaya, Nopember 2018

Penulis



DAFTAR ISI

LEMBAR PENGESAHAN	i
RINGKASAN	ii
PRAKATA.....	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL.....	vii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Datamining	5
2.2 <i>Principal component analysis (PCA)</i>	6
2.3. K-Nearest Neighbor Classifier	7
2.4 Case Based Reasoning Cycle (CBR).....	8
2.5 Perancangan Sistem.....	12
2.7 Pengujian Sistem	13
BAB III TUJUAN DAN MANFAAT	15
BAB IV METODE PENELITIAN	16
4.1. Pengambilan Data.....	16
4.2. Pengelolaan Data	16
4.3 Perancangan Sistem.....	22
4.4 Pengujian Sistem	23
4.5. Implementasi Sistem.....	24
4.6 Evaluasi Sistem.....	24
BAB V HASIL DAN LUARAN YANG DICAPAI.....	25
5.1 Pengumpulan Data dan Informasi	25
5.2 Pengolahan dan Analisis Data	25
5.3 Uji Coba Sistem.....	38
5.4 Evaluasi Sistem.....	42
5.5 Luaran Yang Dicapai.....	43

BABVI RENCANA TAHAPAN BERIKUTNYA.....	44
BAB VII KESIMPULAN DAN SARAN.....	44
7.1 Kesimpulan.....	44
7.2 Saran.....	44
DAFTAR PUSTAKA.....	45
LAMPIRAN.....	47

DAFTAR GAMBAR

Gambar 2.1 Principal component	6
Gambar 2.2 CBR Cycle.....	9
Gambar 4.1 Alur <i>Case Based Reasoning</i>	20
Gambar 4.2 <i>Flowchart</i> Sistem.....	23
Gambar 5.1. Diagram flow untuk system klasifikasi menggunakan K-NN.....	31
Gambar 5.2. Diagram flow untuk system klasifikasi menggunakan PCA dan K-NN.....	32
Gambar 5.3 Halaman diagnosa dengan menggunakan KNN	39
Gambar 5.4 Halaman diagnosa dengan menggunakan Forward Chaining.....	40
Gambar 5.5 Halaman diagnosa dengan menggunakan CBR.....	40
Gambar 5.6 Halaman hasil diagnosa dengan menggunakan KNN.....	41
Gambar 5.7 Halaman hasil diagnosa dengan menggunakan Forward Chaining.....	41
Gambar 5.8 Halaman hasil diagnosa dengan menggunakan CBR.....	42
Gambar 5.9 Grafik akurasi system.....	42

DAFTAR TABEL

Tabel 2.1. Komponen <i>Flowchart</i>	13
Tabel 4.1. Klasifikasi penyakit Balita	19
Tabel 5.1. Hasil perhitungan nilai kovarian.....	26
Tabel 5.2. Nilai eigen dan vektor eigen.....	27
Tabel 5.3. Matrik korelasi.....	27
Tabel 5.4. Nilai Maximum dan Minimum dari data training.....	28
Tabel 5.5. Skor Weighted Voting.....	30
Tabel 5.6. Gejala.....	33
Tabel 5.7. Penyakit.....	34
Tabel 5.8. Rule yang dihasilkan.....	34
Tabel 5.9. Range Berat Badan.....	35
Tabel 5.10. Range Tinggi Badan.....	35
Tabel 5.11. Pemilihan Nilai Tertinggi.....	37
Tabel 5.12. Hasil Uji Coba	38

BAB I PENDAHULUAN

1.1 Latar Belakang

Anak Balita (bawah lima tahun) merupakan aset bangsa yang perlu diberikan pendidikan dan pengasuhan yang baik untuk peningkatan kualitas bangsa (Fauzi, 2016). Masa ini merupakan pertumbuhan dasar anak, selain itu juga terjadi perkembangan kemampuan berbahasa, berkeaktivitas, kesadaran sosial, emosional dan intelegensia berjalan sangat cepat yang merupakan landasan bagi perkembangan anak selanjutnya. Kesehatan anak menjadi bagian paling penting bagi pemerintah untuk dijaga melalui program-program peningkatan kesehatan anak. Peran orang tua juga penting dalam pengawasan terhadap anaknya, sehingga tumbuh kembang anak sesuai yang diharapkan. Anak yang sehat dalam tumbuh kembangnya merupakan dambaan bagi para orang tua.

Salah satu penyakit infeksi yang sering dialami balita adalah demam, diare, dan batuk (Yanto, 2017). Demam merupakan keluhan utama yang sering ditemui pada banyak penyakit, baik yang disebabkan oleh infeksi, dehidrasi, gangguan pusat pengaturan panas, dan keracunan. Diare mengacu pada kehilangan cairan dan elektrolit secara berlebihan yang terjadi dengan bagian feces tidak terbentuk. Diare adalah kondisi frekuensi defekasi yang lebih dari 3 kali sehari, serta konsistensi feses yang cair. Batuk merupakan salah satu penyakit yang lazim pada anak. Batuk memiliki ciri khas sehingga dapat dikenali. Satu hal yang perlu diingat bahwa batuk hanyalah sebuah gejala, bukan suatu penyakit. Batuk baru bisa ditentukan sebagai tanda suatu penyakit jika ada gejala lain yang menyertainya.

Desa Soko merupakan salah satu desa yang terdapat di Kecamatan Glagah, Kabupaten Lamongan. Di Desa Soko, terdapat Balai sakit. Balai sakit tersebut merupakan klinik yang memberikan pelayanan, pengobatan, dan perawatan pada anak sakit. Setiap pasien yang datang dicatat rekam medisnya. Surabaya merupakan salah satu kota terbesar di surabaya. Data rekam medis yang terkumpul di penyimpanan klinik dan salah satu rumah sakit di Surabaya sangatlah banyak dan mempunyai atribut pasien yang sangat banyak, misalnya umur, tinggi badan, berat badan, jenis kelamin, gejala penyakit. Hal tersebut menyebabkan tenaga medis di balai tersebut kesulitan dalam pengolahan tumpukan data pasien.

Oleh karena itu, untuk mengurangi jumlah atribut membutuhkan reduksi atribut (Rafayah et al, 2005). Principal Component Analysis (PCA) adalah salah satu teknik statistik yang terkenal yang bertujuan mereduksi dimensi data tanpa menghilangkan informasi penting



di dalam data tersebut. Pada dasarnya PCA mengubah dan mendekomposisikan sejumlah besar variable yang berkorelasi menjadi sejumlah kecil variable yang tidak berkorelasi dan dapat mengurangi dimensi dari data (Pang-Ning Tan et al, 2006). PCA memiliki beberapa keunggulan seperti mengurangi redundansi data, mengurangi kompleksitas, database kecil dan pengurangan noise. Serta dengan metode ini dapat diketahui korelasi antar atribut sebagai sarana untuk mengetahui apakah tiap-tiap atribut memang diperlukan.

Jabbar (2013) menerapkan KNN dengan Feature Subset Selection untuk menentukan variabel yang memberikan kontribusi lebih untuk prediksi penyakit. Metode ini secara tidak langsung dapat mengurangi jumlah tes yang akan diambil oleh pasien. Model prediksi ini bahkan membantu para dokter dalam proses pengambilan keputusan yang efisien dengan atribut yang lebih sedikit untuk mendiagnosa penyakit jantung.

Klasifikasi adalah teknik yang digunakan untuk menemukan kelas-kelas data yang tidak diketahui (Beniwal & Arora, 2012). Berbagai metode untuk klasifikasi antara lain bayesian, decision trees, rule based, dan neural networks. Selain itu, salah satu metode untuk klasifikasi adalah k-nearest neighbour classifier (KNN classifier). KNN classifier adalah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan k tetangga terdekatnya. KNN menjadi terkenal diantara teknik data mining karena paling sederhana dan kecepatan konvergensi yang relatif tinggi. KNN juga disebut lazy learning karena tidak melalui fase training, dan memory based classification karena sampel pelatihan harus berada di memori pada saat proses sedang berjalan (Jabbar, dkk., 2013). Principal Component Analysis (PCA) dan K-Nearest Neighbor Classifier (KNN) dapat diterapkan untuk klasifikasi Penyakit Balita (Dananjaya, 2017). Penelitian tersebut membandingkan akurasi PCA dan KNN dengan KNN saja. Nilai akurasi PCA dan KNN lebih bagus dibandingkan akurasi tanpa PCA. Namun akurasi yang dihasilkan kurang bagus. Hal ini dikarenakan nilai akurasi yang didapatkan dengan metode K-Nearest Neighbors (KNN) tidak melakukan base learning, KNN melakukan klasifikasi dengan database yang sudah ada, dan tidak dapat melakukan klasifikasi pada data yang tidak masuk kedalam database yang membuat nilai akurasi kurang baik (Ahn & Kim, 2009).

Rule based system adalah sebuah sistem yang dapat melakukan base learning yang tidak masuk ke database. Case Based Reasoning (CBR) dan system pakar merupakan bagian dari rule based system. Sistem pakar yang menggunakan metode forward chaining memberikan hasil diagnosa sesuai dengan fakta – fakta yang diinputkan user (subakti,2002). Sistem pakar dengan metode forward chaining dalam bidang diagnosa kesehatan telah digunakan dalam beberapa penelitian diantaranya Pembuatan Aplikasi Diagnosa Penyakit

Jantung Berdasarkan Faktor Resiko dan Gejala dengan Metode Forward Chaining dan Certainly Factor (Dewi, 2009). Implementasi Metode Forward Chaining untuk Pendeteksian Dini Penyakit Diabetes Melitus (Harahap, 2009), Aplikasi Sistem Pakar Diagnosa Penyakit Menular pada Balita dengan Metode Forward Chaining (Putra, 2012), dan forward chaining dapat digunakan untuk mendiagnosa penyakit anak (Yanto, 2017). Klasifikasi KNN merupakan bagian dari tahapan awal Case Based Reasoning (CBR), dimana CBR melakukan proses pembelajaran pada kasus baru dengan empat tahap yang membuat nilai akurasi lebih baik dibandingkan hanya menggunakan KNN (Ahn & Kim, 2009). CBR merupakan metode pemecahan masalah yang dapat digunakan kembali serupa kasus untuk menemukan solusi terhadap masalah baru dengan mengacu pada kasus basis (Ahn & Kim, 2009). Metode CBR sudah berkembang cukup luas di dunia medis. Dalam penalaran diagnosis dibidang medis dengan metode CBR adalah menggunakan tipe pencocokan pola, yang intinya proses penalaran berbasis kasus didasarkan pada pengalaman pasien-pasien sebelumnya (Mulyana & Hartati, 2009). Proses yang dilakukan dengan metode CBR dilakukan dengan siklus empat langkah, dimana proses tersebut membuat CBR mudah untuk memperbaharui basis pengetahuannya sehingga dapat mengatasi masalah yang kompleks dan tidak terstruktur (Ahn & Kim, 2009).

Pada penelitian CBR untuk Diagnosis Penyakit Jantung (Wahyudi & Hartati, 2017), digunakan tiga jenis metode klasifikasi untuk melakukan diagnosis terhadap 6 jenis penyakit jantung. Metode klasifikasi yang digunakan adalah Nearest Neighbor Similarity (NNS), Minkowski Distance Similarity (MDS) dan Euclidean Distance Similarity (EDS), dengan tingkat akurasi NNS. Akurasi yang paling bagus diperoleh dengan menggunakan MDS. CBR dapat meningkatkan nilai akurasi dalam mendiagnosa penyakit.

Berdasarkan uraian di atas maka pada penelitian ini dibangun system diagnose penyakit balita. Sistem diagnosa penyakit balita membutuhkan proses reduksi atribut, dan klasifikasi penyakit. Reduksi atribut menggunakan PCA dan klasifikasi penyakit menggunakan Rule based system. Rule based system yang dibangun adalah system pakar dan CBR. Hasil akurasi antara KNN dibandingkan dengan akurasi yang diperoleh dari rule based, kemudian akurasi antara system pakar dibandingkan dengan CBR. Sistem diagnosa yang dibangun menggunakan CBR diharapkan memiliki nilai akurasi yang paling optimal sehingga system yang dibangun diharapkan dapat digunakan tenaga medis untuk mendiagnosa penyakit balita secara dini berdasarkan data rekam medis.

1.2 Rumusan Masalah

Berdasarkan uraian diatas, permasalahan dalam penelitian ini sebagai berikut :

- 1. Bagaimana mereduksi atribut yang mempengaruhi penyakit anak di bawah lima tahun dengan PCA berdasarkan korlasi antar atribut.**
- 2. Bagaimana membuat sistem diagnose pola penyakit anak dibawah lima tahun dengan Rule Based System (CBR).**

BAB II

TINJAUAN PUSTAKA

Pada bagian ini dijelaskan beberapa tinjauan pustaka yang akan digunakan pada penelitian.

2.1 Datamining

Data mining merupakan proses pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan tehnik gambar dari statistik, mesin pembelajaran dan sistem manajemen database (Han, Jiawei et al, 2001) Data mining yang disebut juga dengan *Knowledge-Discovery in Database (KDD)* adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*) atau pengelompokan (*clustering*).

Secara sederhana, data mining dapat diartikan sebagai proses mengekstrak atau “menggali” pengetahuan yang ada pada sekumpulan data. Banyak orang yang setuju bahwa data mining adalah sinonim dari *Knowledge-Discovery in Database* atau yang biasa disebut KDD. Dari sudut pandang yang lain, data mining dianggap sebagai satu langkah yang penting didalam proses KDD. Menurut Han, J. and Kamber, M, 2001, proses KDD ini terdiri dari langkah-langkah sebagai berikut :

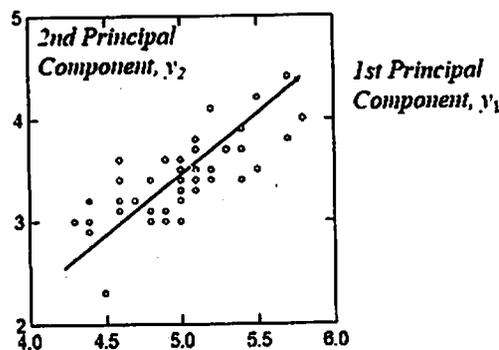
1. *Data Cleaning*, proses menghapus data yang tidak konsisten dan kotor
2. *Data Integration*, penggabungan beberapa sumber data
3. *Data Selection*, pengambilan data yang akan dipakai dari sumber data
4. *Data Transformation*, proses dimana data ditransformasikan menjadi bentuk yang sesuai untuk diproses dalam data mining
5. *Data Mining*, suatu proses yang penting dengan melibatkan metode untuk menghasilkan suatu pola data
6. *Pattern Evaluation*, proses untuk menguji kebenaran dari pola data yang mewakili knowledge yang ada didalam data itu sendiri
7. *Knowledge Presentation*, proses visualisasi dan teknik menyajikan knowledge digunakan untuk menampilkan knowledge hasil mining kepada user.

2.2 *Principal component analysis (PCA)*

Principal component analysis (PCA) adalah salah satu teknik statistik yang terkenal yang bertujuan mereduksi dimensi data tanpa menghilangkan informasi penting di dalam data tersebut. Pada dasarnya PCA mengubah dan mendekomposisikan sejumlah besar variable yang berkorelasi menjadi sejumlah kecil variable yang tidak berkorelasi dan dapat mengurangi dimensi dari data (Pang-Ning Tan et al, 2006).

Secara umum prosedur PCA adalah menemukan eigen vector dari matrik kovarian dan eigen vector untuk mendefinisikan atribut yang baru. Prosedure PCA sebagai berikut (Pang-Ning Tan et al, 2006).

1. Input data dinormalisasi terlebih dahulu, sehingga masing-masing atribut mempunyai dimensi yang sama.
2. Menghitung vektor orthogonal k yang menyediakan sebuah basis dari input data yang sudah dinormalisasi, yaitu vektor satuan yang masing-masing titik tegak lurus dengan lainnya yang disebut *Principal component*. Proses diatas diulangi k kali. Hal ini sama dengan mencari eigenvector dari matriks covariance dengan k eigenvalue terbesar. Gambar *Principal Component* dapat dilihat pada Gambar 2.1.
3. *Principal component* diurutkan secara signifikan. *Principal component* secara essensial sebagai satu set baru dari sumbu untuk data (memberikan informasi yang penting dari varians)
4. Sejak *Principal component* diurutkan secara signifikan, ukuran data dapat dikurangi dengan mengeliminasi component yang lemah (dimana data yang memiliki varians yang rendah)



Gambar 2.1 Principal component (Pang-Ning Tan et al, 2006)

2.3. K-Nearest Neighbor Classifier

KNN merupakan salah satu teknik data mining yang paling sederhana dan merupakan teknik data mining lazy learning karena tidak memiliki fase training, (Jabbar, dkk., 2013). KNN juga disebut sebagai klasifikasi berdasarkan memori, karena sampel pelatihan perlu berada dimemori pada saat dijalankan. KNN adalah contoh pembelajaran instance-based, dimana data training set disimpan, sehingga klasifikasi untuk unclassified record baru dapat ditemukan hanya dengan membandingkannya dengan record paling mirip pada training set (Larose, 2005).

Algoritma ini penentuan nilai k sangat berpengaruh dalam prose klasifikasi. Nilai k sendiri merupakan jumlah tetangga terdekat yang akan dijadikan estimasi dari record baru. Menetapkan nilai k yang terlalu kecil memungkinkan klasifikasi akan terpengaruhi oleh noise atau outliers. Akan tetapi, dengan memilih nilai k yang besar menyebabkan kemiripan tiap record semakin sedikit dan menyebabkan record baru mempunyai kemiripan dengan record dengan jarak yang jauh. Untuk mencegah hal tersebut pada proses klasifikasi akan diterapkan fungsi weighted voting. Weighted voting merupakan metode pembobotan KNN berdasarkan jarak data baru terhadap data lain. Dengan fungsi weighted voting penentuan kelas berdasarkan bobot dari masing-masing kelas yang ada, dimana tetangga dekat memiliki bobot lebih besar dalam keputusan klasifikasi daripada tetangga yang lebih jauh (Larose, 2005).

Algoritma KNN mempunyai beberapa tahapan, sebagai berikut:

1. Normalisasi dataset dengan menggunakan rumus min-max normalization untuk data numerik. Rumus min-max normalization:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (2.1)$$

Keterangan :

X' : data yang telah dinormalisasi

X : data sebelum dinormalisasi

Sedangkan untuk data kategorik dapat digunakan sebuah fungsi different from. Fungsi ini digunakan untuk membandingkan nilai variabel ke-i terhadap record lain, berikut ketentuannya:

$$\text{different}(x_i, y_i) = \begin{cases} 0 & \text{if } x_i = y_i \\ 1 & \text{otherwise} \end{cases} \quad (2.2)$$

Diasumsikan bahwa record x_i dan y_i mempunyai jarak minimum jika record mempunyai nilai yang sama. Dan sebaliknya jika record mempunyai nilai yang berbeda jaraknya maksimum.

2. Menghitung jarak record baru terhadap record lain untuk menemukan kedekatannya. Semakin kecil jaraknya maka semakin besar kemungkinan record tersebut berada dalam kelas yang sama. Untuk menghitung nilai jarak digunakan rumus Euclidean distance, sebagai berikut:

$$d_{euclidean}(x, y) = \sqrt{\sum_1^i (x_i - y_i)^2} \quad (2.3)$$

Dimana $x = x_1, x_2, \dots, x_i$ dan $y = y_1, y_2, \dots, y_i$ mewakili nilai variabel ke- i terhadap dua record.

3. Menentukan kelas dari record baru dengan fungsi weighted voting. Setelah ditemukan nilai jarak Euclidean kemudian dilakukan voting antar kelas. Kelas kelas tersebut nantinya akan menunjukkan nilai yang berbeda-beda. Kemudian record baru dapat diketahui kelasnya dengan melihat hasil voting tertinggi.

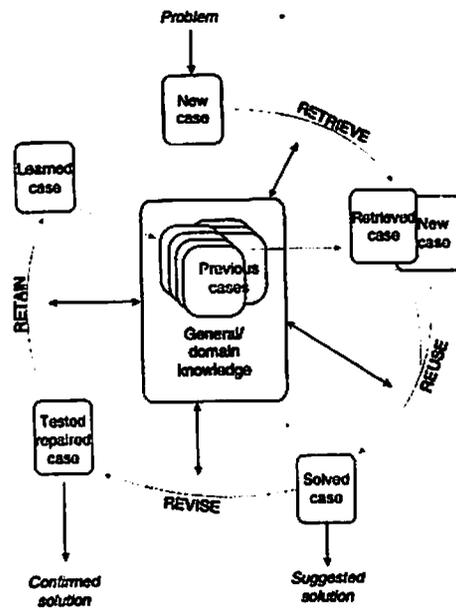
Berikut rumus weighted voting:

$$weighted\ voting = \sum_1^j \left(\frac{1}{d_j^2} \right) \quad (2.4)$$

Adapun $d = 1, d_2, \dots, d_j$ merupakan jarak Euclidean

2.4 Case Based Reasoning Cycle (CBR)

Case Based Reasoning Cycle (CBR Cycle) merupakan framework yang digunakan dalam CBR. CBR cycle adalah model dinamis yang mengidentifikasi sub proses utama dari siklus CBR, sedangkan yang kedua adalah tampilan berorientasi tugas dimana terjadi dekomposisi tugas dan masalah terkait. Untuk menghasilkan solusi suatu masalah, CBR harus melakukan beberapa tahap proses dimana CBR harus mencari kemiripan kasus baru dengan kasus yang tersimpan, atau ketika ada perubahan terhadap solusi suatu kasus. Tahapan proses yang terjadi dalam CBR dapat dilihat pada Gambar 2.2.



Gambar 2.2 CBR Cycle

1. Retrive, menemukan kembali kasus yang sama atau yang paling mirip dengan kasus baru
2. Reuse adalah menggunakan kembali informasi dan pengetahuan dari basis kasus untuk memecahkan masalah kasus baru (proses ini disebut “tansfer solusi”),
3. Revise adalah merevisi atau memperbaiki solusi yang diusulkan.
4. Retain adalah menyimpan pengalaman untuk memecahkan masalah yang akan datang kedalam basis kasus.

2.4.1 Langkah-Langkah Menggunakan CBR

Langkah-langkah menggunakan CBR adalah pemecahan masalah menggunakan informasi yang tersimpan pada kasus sebelumnya. Berdasarkan tahapan yang ada dalam suatu sistem CBR, diperlukan beberapa langkah dalam menentukan solusi, yaitu:

1. Basis Kasus

Membangun basis kasus, yang digunakan sebagai tempat penyimpanan. Pada langkah ini, setiap kasus yang disimpan dibagi menjadi beberapa bagian utama bagian, yaitu: identitas usia anak, gejala-gejala penyakit dan hasil diagnosa penyakit. Solusi pembagian ini dilakukan untuk memudahkan penyimpanan data kasus kedalam basis kasus, serta memudahkan dalam pengambilan data yang sesuai dengan kasus baru.

2. Pembobotan Oleh Ahli

Tujuan Pembobotan Parameter adalah untuk mengekspresikan seberapa besar pengaruh suatu parameter terhadap parameter lainnya. Terdapat banyak jenis metode untuk melakukan

pembobotan, salah satunya metode ranking, metode rating, metode perbandingan pasangan, dan metode analisis trade off. Proses pembobotan dilakukan oleh ahli yang memiliki ilmu dan pengalaman dibidang tersebut.

3. Similarity Lokal

Similarity lokal adalah proses melakukan kedekatan antar atribut secara lokal atau yang memiliki atribut yang sama. Similarity lokal dilakukan dengan mendefinisikan kedekatan antar nilai dalam atribut. Perhitungan similarity local dihitung berdasarkan tipe data fiturnya, untuk tipe data numerik ditunjukkan pada persamaan 2.5 dan 2.6.

1. Rumus similarity local untuk tipe data numerik:

$$f(s, t) = 1 - \frac{|s-t|}{R} \quad (2.5)$$

Keterangan:

s, t : Nilai atribut yang ingin dibandingkan

R : Range nilai atribut

2. Rumus similarity local untuk tipe data Boolean:

$$f(s, t) = \begin{cases} 1 & \text{jika } s=t \\ 0 & \text{lainnya} \end{cases} \quad (2.6)$$

Keterangan:

$s, t \in \{ \text{benar, salah} \}$

4. Tingkat Keyakinan

Tingkat keyakinan adalah pencerminan tingkat kepastian yang diinginkan oleh pengukur setelah memutuskan tidak akan melakukan pengukuran yang sangat banyak. Tingkat keyakinan dihitung dengan menghitung tingkat keyakinan kasus baru terhadap kasus lama yang diaplikasikan dengan persamaan 2.7.

$$\mu_{(T,S)} = \frac{J(S,T)}{J(T)} \quad (2.7)$$

Keterangan:

$\mu_{(T,S)}$: Tingkat keyakinan antar kasus T (target case) dan S (source case)

$J(S, T)$: Banyaknya fitur yang terdapat dalam target case yang muncul pada fitur source case.

$J(T)$: Banyaknya fitur yang terdapat dalam target case

5. Similarity Global

Similaritas adalah proses menunjukkan keserupaan antara atribut permasalahan terhadap atribut yang sama dari sebuah kasus. Similarity local dan global memiliki cara kerja yang sama dengan cakupan similarity yang berbeda, similarity local mencari keserupaan antar atribut pada satu variable sedangkan similarity global mencari keserupaan antar atribut pada seluruh varable yang ada.

Algoritma yang digunakan dalam proses similarity adalah Nearst Neighbor Similarity, minkowski distance similarity dan euclidean distance similarity. Rumus untuk melakukan perhitungan kedekatan similarity global ditambah dengan pembobotan dan telah dimodifikasi dengan menambah faktor tingkat keyakinan similarity adalah sebagai berikut:

1. Nearst Neighbour Similarity:

$$SimNN(T, S) = \frac{\sum_{i=1}^n (f_i(S_i, T_i) \cdot w_{i,p(s)})}{\sum_{i=1}^n (w_{i,p(s)})} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} \quad (2.8)$$

Keterangan:

SimNN(S,T) : Similaritas global antara kasus T (target case) dan S (source case)

N : Banyaknya fitur yang ada

$f_i(S_i, T_i)$: Kesamaan fitur ke-i dari source case dan target case/ fungsi similaritas local

S_i : Fitur ke-i yang ada dalam source case

T_i : Fitur ke-i yang ada dalam target case

$w_{i,p(s)}$: Nilai bobot fitur ke-i pada penyakit dari source case

$P(S)$: Persentase tingkat keyakinan pakar terhadap suatu kasus dalam source case

$J(S_i, T_i)$: Banyaknya fitur yang terdapat dalam target case yang muncul pada fitur source case

$J(T_i)$: Banyaknya fitur yang terdapat dalam target case

2. Minkowsky Distance Similarity:

$$SimMD(S, T) = \left[\frac{\sum_{i=1}^n (w_{i,p(s)})^3 * |f_i(S_i, T_i)|^3}{\sum_{i=1}^n (w_{i,p(s)})^3} \right]^{\frac{1}{3}} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} \quad (2.9)$$

Keterangan:

SimMD(S,T) : Similaritas global antara kasus T (target case) dan S (source case)

N : Banyaknya fitur yang ada

$w_{i,p(s)}$: Nilia bobot fitur ke-i pada penyakit dari source case

- R : Faktor minkowski (integer positif) (dalam hal ini $r=3$)
 $f_i(S_i, T_i)$: Kesamaan fitur ke-i dari source case dan target case / fungsi similaritas local
 S_i : Fitur ke-i yang ada dalam source case
 T_i : Fitur ke-i yang ada dalam target case
 $P(S)$: Persentase tingkat keyakinan pakar terhadap suatu kasus dalam source case
 $J(S_i, T_i)$: Banyaknya fitur yang terdapat dalam target case yang muncul pada futur source case
 $J(T_i)$: Banyaknya fitur yang terdapat dalam target case

3. Euclidian Distance Similarity:

$$SimMD(S, T) = \left[\frac{\sum_{i=1}^n (w_{i,p(s)})^2 \cdot |f_i(S_i, T_i)|^3}{\sum_{i=1}^n (w_{i,p(s)})^2} \right]^{\frac{1}{2}} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} \quad (2.10)$$

Keterangan:

- $SimED(S, T)$: Similaritas global antara kasus T (target case) dan S (source case)
 N : Banyaknya fitur yang ada
 $w_{i,p(s)}$: Nilia bobot fitur ke-i pada penyakit dari source case
 R : Faktor minkowski (integer positif) (dalam hal ini $r=3$)
 $f_i(S_i, T_i)$: Kesamaan fitur ke-i dari source case dan target case / fungsi similaritas local
 S_i : Fitur ke-i yang ada dalam source case
 T_i : Fitur ke-i yang ada dalam target case
 $P(S)$: Persentase tingkat keyakinan pakar terhadap suatu kasus dalam source case
 $J(S_i, T_i)$: Banyaknya fitur yang terdapat dalam target case yang muncul pada futur source case
 $J(T_i)$: Banyaknya fitur yang terdapat dalam target case

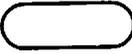
2.5 Perancangan Sistem

Perancangan sistem menentukan bagaimana suatu sistem akan menyelesaikan apa yang mesti diselesaikan. Tahapan ini menyangkut mengkonfigurasi dari komponen-komponen perangkat lunak dan perangkat keras dari suatu sistem, sehingga instalasi dari sistem akan benar-benar memuaskan rancangan bangun yang telah ditetapkan pada akhir tahap analisa sistem. Tujuan utama perancangan sistem adalah:

1. Untuk memenuhi kebutuhan para pemakai system
2. Untuk memberikan gambaran yang jelas dan rancang bangun yang lengkap kepada programmer dan user yang terlibat.

Salah satu cara penggambaran alur sistem dapat menggunakan *flowchart*. *Flowchart* adalah suatu diagram yang menampilkan aliran data dan rangkaian tahapan operasi dalam suatu system. Simbol-simbol pada *flowchart* terdapat pada Tabel 2.1.

Tabel 2.1. Komponen *Flowchart*

No	Nama	Bentuk	Keterangan
1	Terminator		Simbol yang menyatakan titik ujung dari sebuah flowchart. Titik awal, titik akhir ataupun titik interupsi
2	Proses		Simbol yang menggambarkan setiap fungsi pengolahan data
3	Keputusan		Menggambarkan keputusan untuk menentukan operasi mana yang harus dijalankan dari berbagai alternatif jalur operasi yang tersedia
4	Garis alir		Mengaitkan simbol satu dengan simbol yang lain . mengindikasi urutan informasi dan operasi yang harus dijalankan
5	Input/output		Membuat data tersedia untuk diproses dan mencatat informasi hasil suatu pemrosesan.

2.7 Pengujian Sistem

Pengujian Sistem adalah proses menjalankan suatu program dengan tujuan untuk menemukan kesalahan dari sistem yang telah dikembangkan. Pengujian yang baik adalah yang memiliki probabilitas tinggi untuk menemukan kesalahan yang belum ditemukan sebelumnya (Pressman, 2012). Sebelum menetapkan metode pengujian, seorang perancang sistem harus mengerti atau memahami prinsip dasar yang menuntun pengujian perangkat lunak. Pengujian sistem terdapat 2 pendekatan dalam teknik pengujian yaitu:

1. ***Black Box Testing***

Black box testing adalah menguji perangkat lunak dari segi spesifikasi fungsional tanpa menguji desain dan kode program. Pengujian ini bertujuan untuk menunjukkan fungsi perangkat lunak, apakah masukan dan keluaran dari perangkat lunak sesuai dengan spesifikasi yang dibutuhkan.

2. ***White Box Testing***

White box testing merupakan langkah pengujian untuk mendapatkan program yang benar, dilakukan dengan mendefinisikan seluruh jalur logika, mengembangkan test case untuk mengerjakan program dan kemudian mengevaluasi hasilnya. Secara garis besar dapat diartikan bahwa ***white box testing*** merupakan petunjuk untuk mendapatkan program yang benar / sesuai dengan keinginan pengguna.

BAB III

TUJUAN DAN MANFAAT

3.1 Tujuan

Berdasarkan uraian diatas maka tujuan dari penelitian ini sebagai berikut:

1. Mereduksi atribut yang mempengaruhi penyakit balita dengan PCA berdasarkan korelasi antar variabel.
2. Membangun sistem diagnosa penyakit balita dengan rule based system.

3.2 Manfaat

Manfaat yang diharapkan dari penelitian ini sebagai berikut:

1. Melatih kemampuan menganalisa suatu masalah terkait seleksi atribut
2. Membantu pihak kesehatan untuk mengetahui pola penyakit balita.
3. Memberikan kemudahan untuk mengetahui jenis penyakit yang diderita balita sakit sehingga dapat digunakan untuk membantu tenaga medis untuk mendiagnosa penyakit balita secara dini.

BAB IV

METODE PENELITIAN

Secara garis besar, penelitian dilakukan dalam dua tahap. Tahap pertama untuk mereduksi variabel yang mempengaruhi penyakit anak di bawah lima tahun dengan PCA berdasarkan korelasi antar variabel. Tahap kedua bertujuan untuk membuat sistem diagnosa penyakit balita dengan rule based system. Tahapan yang dilakukan dalam penelitian ini sebagai berikut :

4.1. Pengambilan Data

Data yang digunakan dalam penelitian ini diambil dari klinik Desa Soko, Kecamatan Glagah, Kabupaten Lamongan berupa data penyakit yang diderita anak dengan keluhan batuk, diare dan demam pada Desember 2016 sampai dengan Mei 2018. Pengumpulan data yang digunakan dalam penelitian ini adalah melalui studi literatur dan wawancara.

Studi literature dilakukan untuk mengetahui dan memahami penggunaan PCA dan rule based system dengan menggunakan bantuan buku, jurnal dan sumber lain yang terkait. Wawancara dilakukan dengan salah satu tenaga medis yang berpraktek di Desa Soko, Kecamatan Glagah, Kabupaten Lamongan dan salah satu rumah sakit di Surabaya. Wawancara untuk mengetahui jenis penyakit yang sering menyerang balita sehingga dapat dijadikan sistem terkomputerisasi untuk mengetahui pola penyakit tersebut. Data dan informasi yang diperoleh digunakan sebagai gambaran masalah-masalah dan analisis sistem yang diteliti.

4.2. Pengelolaan Data

Data yang diperoleh dianalisis untuk mengetahui input, proses, dan output sistem. Input pada sistem ini diidentifikasi sebagai data yang diolah pada tahap proses. Data yang diinputkan adalah tanggal berobat, umur, jenis kelamin, berat badan, tinggi badan, dan 26 gejala, yaitu batuk, demam, diare, anak tidak bisa minum atau menyusu, anak memuntahkan makanan yang dimakan, anak menderita kejang, anak tampak letargis atau tidak sadar, napas cepat, tarikan dinding dada ke dalam, stridor, berak cair atau lembek, mata cekung, cubitan kulit perut kembali lambat, rewel/mudah marah, haus, minum dengan lahap, diare 14 hari atau lebih, ada darah dalam tinja, kaku kuduk (anak tidak bisa menunduk hingga dagu mencapai dada), ruam kemerahan di kulit, pilek atau mata merah, kekeruhan pada kornea mata, luka di mulut, mata bernanah, demam 2 - 7 hari, demam mendadak tinggi dan terus menerus.

Tahap pertama adalah reduksi variabel, digunakan untuk mengetahui pengaruh variabel yang mempengaruhi penyakit balita. Hasil dari PCA adalah variabel yang telah direduksi

berdasarkan korelasi antar variabel. Setelah diperoleh variabelnya, maka system mengelola dengan menggunakan rule based system. Setelah itu, menghasilkan suatu hasil atau yang biasa disebut dengan output. Sedangkan pada bagian output yang dihasilkan oleh sistem adalah jenis penyakit. Terdapat 16 jenis penyakit, yaitu Batuk, Pneumonia, Pneumonia Berat, Diare, Diare Dehidrasi Ringan, Diare Dehidrasi Berat, Diare Persisten, Diare Persisten Berat, Disentri, Demam, Demam dengan Tanda Bahaya Umum, Campak, Campak Dengan Komplikasi Berat, Campak Dengan Komplikasi, Demam Mungkin DBD, DBD.

4.2.1. Reduksi Variabel

Reduksi variabel digunakan untuk mengetahui pengaruh variabel yang mempengaruhi penyakit balita. Hasil dari PCA adalah variabel yang telah direduksi berdasarkan korelasi antar variabel. Berikut merupakan langkah-langkah dari proses yang akan dijalankan dengan PCA, yaitu (Pang-Ning Tan et al, 2006):

- Memasukkan semua data yang terdiri dari atribut umur, berat badan, tinggi badan, dan suhu badan.
- Menghitung standart deviasi dari data tersebut. Standar deviasi dari himpunan data sampel (disini dibagi $(n-1)$ untuk sampel, dibagi n bila untuk seluruh populasi) :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}} \quad (4.1)$$

- Menghitung varian dan covarian dari data tersebut. Persamaan varian dan covarian sebagai berikut :

$$\text{var}(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)} \quad (4.2)$$

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)} \quad (4.3)$$

- Menghitung nilai eigen dan vektor eigen dari matrik covarian diatas.

Nilai eigen merupakan nilai karakteristik suatu matriks. Secara sederhana, nilai eigen merupakan nilai yang mempresentasikan suatu matriks dalam perkalian dengan suatu vektor, dapat ditulis sebagai:

$$Ax = \lambda x \quad (4.4)$$

di mana A suatu matriks, x merupakan vector eigen yang merupakan solusi dari matriks $(A-\lambda)$ untuk setiap nilai λ yang ada di mana $x \neq 0$, dan λ merupakan nilai eigen dari matriks A . Nilai eigen matriks A dapat dicari dengan

$$Ax - \lambda x = 0 \quad (4.5)$$

$$(A - \lambda)x = 0 \quad (4.6)$$

- e. Menghitung *Principal component*
- f. Setelah *Principal component* diketahui, maka akan dicari korelasi antara data dan *principal component*
- g. Mengurangi variabel pada data dengan mengeliminasi *component* yang rendah

4.2.2. Proses klasifikasi

Setelah dilakukan reduksi variabel, langkah selanjutnya adalah proses klasifikasi. Proses klasifikasi dilakukan dengan menggunakan KNN dan rule based system. Rule based system yang digunakan adalah system pakar dan CBR.

1. K-NN Classifier

K-NN classifier pada tahap ini digunakan untuk menentukan jenis penyakit balita. Variabel yang digunakan untuk proses klasifikasi pada sistem yaitu variabel yang dihasilkan oleh PCA. Terdapat 26 gejala pada data yang diolah, yaitu batuk, demam, anak tidak bisa minum atau menyusu, anak memuntahkan makanan yang dimakan, anak menderita kejang, anak tampak letargis atau tidak sadar, napas cepat, tarikan dinding dada ke dalam, stridor, berak cair atau lembek, mata cekung, cubitan kulit perut kembali lambat, rewel/mudah marah, haus, minum dengan lahap, diare 14 hari atau lebih, ada darah dalam tinja, kaku kuduk (anak tidak bisa menunduk hingga dagu mencapai dada), ruam kemerahan di kulit, pilek atau mata merah, kekeruhan pada kornea mata, luka di mulut, mata bernanah, demam 2 - 7 hari, demam mendadak tinggi dan terus menerus, tinja berwarna hitam. Klasifikasi penyakit balita dapat dilihat pada Tabel 4.1.

Dalam penentuan kelas dari data baru pada data balita sakit diterapkan fungsi weighted voting. Berikut merupakan langkah-langkah pengolahan data dengan KNN classifier:

- a. Data yang telah diolah dengan PCA, didapatkan variabel yang berpengaruh untuk klasifikasi. Variabel tersebut kemudian digabungkan dengan variabel jenis kelamin, dan gejala-gejala menjadi dataset baru.
- b. Dataset baru dinormalisasi dengan menggunakan rumus min-max normalization. Begitu pula dengan data untuk testing juga dinormalisasi.
- c. Menghitung jarak record testing terhadap record dari dataset dengan rumus Euclidean distance.
- d. Menentukan kelas dari record testing. Fungsi weighted voting digunakan untuk menghitung bobot setiap kelas. Kelas dengan nilai terbesar maka record testing termasuk di dalam kelasnya.

Tabel 4.1. Klasifikasi penyakit Balita

Keluhan	Klasifikasi Penyakit
Batuk	Batuk
	Pneumonia
	Pneumonia Berat
Diare	Diare
	Diare Dehidrasi Ringan
	Diare Dehidrasi Berat
	Diare Persisten
	Diare Persisten Berat
	Disentri
Demam	Demam
	Demam dengan Tanda Bahaya Umum
	Campak
	Campak Dengan Komplikasi Berat
	Campak Dengan Komplikasi
	Demam Mungkin DBD
	DBD

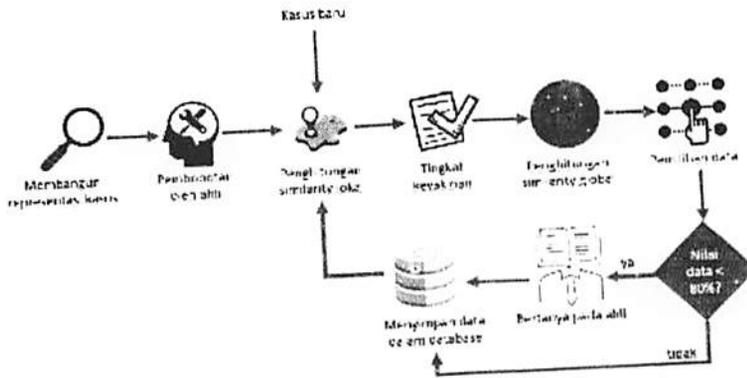
2. Sistem Pakar

Berikut penjelasan gambaran umum system dari system pakar, sebagai berikut :

- Input, merupakan pertanyaan-pertanyaan yang muncul dalam aplikasi sistem pakar.
- Knowledge base (domain pengetahuan), pengetahuan mengenai klasifikasi penyakit balita yang dijadikan sebagai rule-based.
- Working memory, fakta-fakta yang diinputkan oleh pengguna kedalam aplikasi sistem pakar.
- Inference Engine, proses pencocokan fakta-fakta yang ada pada working memory dengan domain pengetahuan, untuk menarik kesimpulan dari masalah yang dihadapi.
- Klasifikasi penyakit, hasil kesimpulan dari proses diagnosa sistem pakar
- Tindakan/pengobatan, tata cara penanganan yang bisa dilakukan oleh pengguna.

3. CBR

Proses CBR adalah memasukkan kasus baru dan mencari kesesuaian kasus baru dengan kasus yang terdapat didalam *database*. Dalam metode ini, langkah-langkah yang dilakukan untuk mengolah kasus baru digambarkan dalam Gambar 4.1.



Gambar 4.1 Alur Case Based Reasoning

a. **Membangun Representasi Kasus**

Basis kasus dibuat untuk menentukan kasus yang dijalankan oleh metode CBR. Proses representasi kasus adalah mengetahui variabel-variabel yang dibutuhkan untuk kepentingan penyimpanan retrieve. Proses representasi kasus dalam hal ini adalah jenis kelamin, usia, berat badan, tinggi badan, suhu badan dan 26 gejala.

b. **Pembobotan**

Pembobotan dibuat untuk mengetahui pengaruh suatu parameter terhadap parameter lainnya, dalam hal ini pembobotan dibuat untuk mengetahui pengaruh parameter sebuah variabel gejala terhadap penyakit. Pembobotan diperoleh dari melakukan analisa statistic dari data testing yang diperoleh dan kemudian didiskusikan dalam wawancara dengan pakar. Pemberian nilai bobot yang semakin tinggi memberi tanda bahwa variabel yang diberi bobot tersebut sangat mempengaruhi penyakit tertentu.

c. **Penghitungan Similarity Lokal**

Fungsi similarity lokal adalah untuk menghitung nilai similarity secara lokal antara variabel yang terdapat dalam database dan nilai data kasus baru. Tujuan similarity local adalah mengetahui keserupaan antar variabel gejala antara kasus baru yang didiagnosis dengan data yang terdapat dalam database. Proses penghitungan similarity lokal dihitung berdasarkan tipe data numerik dan boolean. Perhitungan secara numerik untuk variabel gejala dengan tipe umum seperti tinggi badan, berat badan dan suhu badan. Perhitungan secara boolean untuk variabel gejala dengan tipe khusus yang memiliki jawaban ya atau tidak seperti batuk, pilek, demam, dan lain lain. Hasil dari similarity lokal berupa nilai yang digunakan untuk proses perhitungan tingkat keyakinan.

d. Penghitungan Tingkat Keyakinan

Fungsi tingkat keyakinan adalah melakukan perhitungan kepastian. Proses yang dilakukan dalam tingkat keyakinan adalah membandingkan banyaknya variabel gejala yang terdapat dalam kasus baru yang muncul dalam kasus lama database dengan banyaknya variabel gejala yang terdapat dalam database. Hasil dari penghitungan tingkat keyakinan menentukan dilanjutkan atau tidaknya pada proses penghitungan berikutnya. Nilai yang memiliki tingkat keyakinan > 0.75 tidak dapat digolongkan kedalam salah satu penyakit, maka kasus baru yang memiliki tingkat keyakinan tertinggi kurang dari 0.75 disimpan untuk kemudian dilakukan revisi oleh pakar. Nilai tingkat keyakinan yang lebih dari 0.75 dapat melanjutkan penghitungan ke proses similarity global.

e. Similarity Global

Similarity global adalah proses yang dilakukan untuk mengetahui nilai kedekatan antara data kasus baru dengan seluruh data kasus lama dalam database. Proses penghitungan similarity global dilakukan dengan membandingkan hasil nilai kasus baru dan kasus lama dalam database yang didapat dari hasil similarity local yang kemudian dikalikan dengan bobot ahli yang sesuai dengan penyakit yang didiagnosis. Similarity global juga tingkat keyakinan Proses similarity global dalam penelitian ini dilakukan dengan 3 jenis metode similarity yaitu proses similarity dengan nearest neighbor similarity (NNS), Minkowsky distance similarity (MDS), dan euclidian distance similarity (EDS). Hasil dari masing-masing metode similarity kemudian diurutkan dan dipilih nilai tertinggi untuk dilakukan proses pemilihan nilai tertinggi.

f. Pemilihan Nilai Tertinggi

Pemilihan nilai tertinggi adalah proses melakukan pengurutan nilai dari seluruh hasil perhitungan similarity global dengan 3 jenis similarity yaitu NNS, EDS dan MDS. Dari hasil similarity yang ada, dipilih masing-masing satu nilai similarity global yang tertinggi. Nilai tertinggi yang didapat kemudian dilihat diagnosisnya dengan cara melihat detail data pemilihan nilai tertinggi dan mengambil hasil diagnosisnya. Kasus baru kemudian diberi hasil diagnosis sesuai dengan kasus lama yang memiliki nilai tertinggi.

g. Proses Retrieve

Proses retrieve adalah proses pilihan. Proses ini dilakukan ketika nilai yang tertinggi yang didapatkan dalam proses perhitungan tingkat keyakinan $< 0,75$. Artinya kecocokan antara kasus baru dengan data kasus lama yang terdapat dalam database tidak lebih dari 75% (threshold), yang menyebabkan walaupun kasus baru memiliki kesamaan variabel gejala dengan data kasus lama dalam database, kasus tersebut tidak dapat didiagnosis. Langkah yang

dilakukan bila terjadi kasus reverse adalah menyimpan permasalahan kasus baru tersebut dan menanyakannya pada pakar yang ahli dibidangnya.

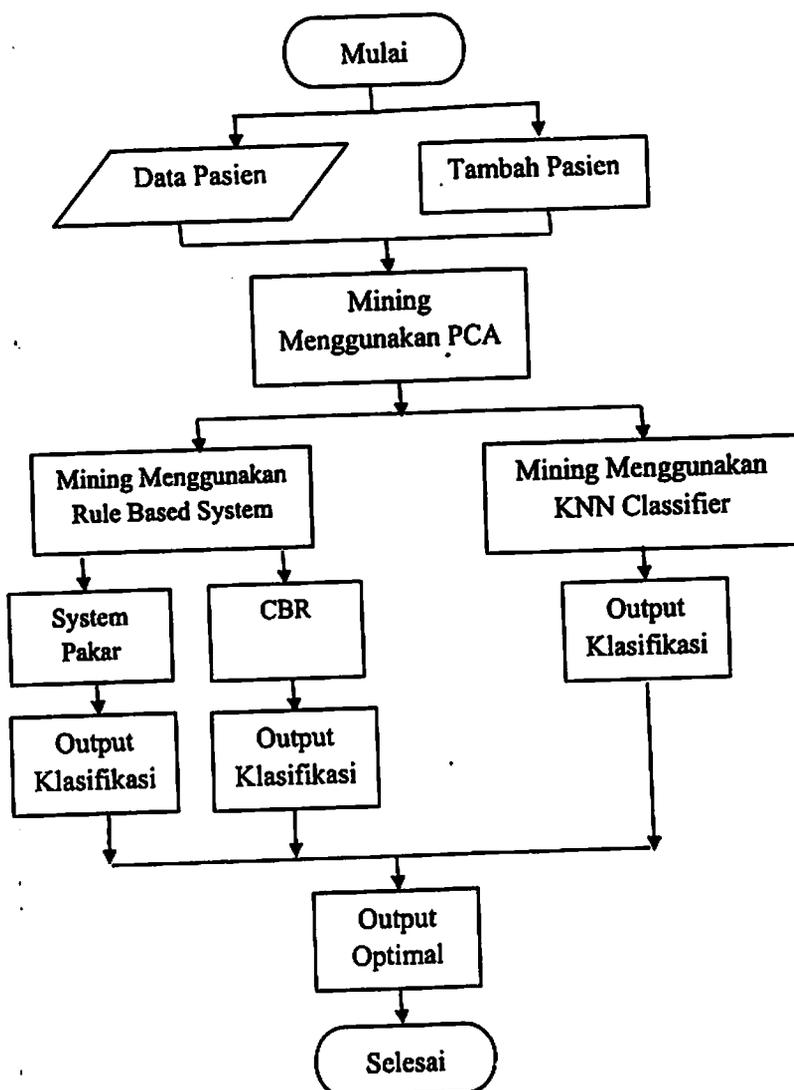
h. Menyimpan Data dalam Database

Proses menyimpan data dalam database adalah proses memasukkan data permasalahan kasus baru yang telah melalui proses CBR dan telah mendapatkan diagnosis kedalam database. Proses ini berfungsi sebagai referensi dan menambah pengetahuan sistem untuk melakukan diagnosis masalah pada kasus-kasus berikutnya. Semakin banyak database yang dimiliki sistem, maka semakin banyak referensi dan peluang sistem untuk melakukan diagnosis masalah secara tepat pada kasus-kasus berikutnya.

4.3 Perancangan Sistem

Sebelum pembuatan aplikasi, perlu dibuatnya rancangan sistem terlebih dahulu. Pembuatan rancangan ini diharapkan agar aplikasi tersebut dapat berfungsi sebagai gambaran umum yang diharapkan untuk membantu pengguna dalam mendiagnosa penyakit pada balita secara dini. Berikut adalah gambaran sistem yang digunakan sebagai acuan untuk pembuatan aplikasi. Gambaran sistem secara umum dijelaskan pada Gambar 4.2. Berikut penjelasan dari perancangan sistem yang dibangun, sebagai berikut:

- a. Input data pada sistem digunakan untuk menginputkan atau mengimport data yang sudah ada digunakan sebagai proses reduksi. Inputan untuk reduksi adalah umur, berat badan, tinggi badan, suhu badan, dan gejala.
- b. Pada tahap tambah pasien digunakan untuk menambah pasien yang baru melakukan registrasi yang sedang berobat atau menambahkan dari data rekam medis.
- c. Pada tahap selanjutnya, data diolah oleh PCA yang sudah dijelaskan di pengelolaan data.
- d. Tahap selanjutnya adalah klasifikasi dengan KNN Classifier dan Rule based System. Rule based system terdiri dari system pakar dan CBR.
- e. Setelah di kelola dengan proses klasifikasi, langkah selanjutnya adalah membandingkan akurasi yang diperoleh dari masing masing classifier. Akurasi yang terbaik merupakan akurasi yang optimal untuk diagnose penyakit anak.



Gambar 4.2 Flowchart Sistem

4.4 Pengujian Sistem

Pengujian sistem diperlukan untuk menguji seberapa jauh sebuah sistem dapat menjalankan fungsinya dengan baik. Adapun teknik pengujian yang digunakan untuk menguji adalah teknik *blackbox testing*. Uji coba sistem dilakukan dengan menggunakan data sampling berupa data Balita yang sakit dengan keluhan batuk, diare dan demam dari tempat klinik.

4.5. Implementasi Sistem

Implementasi sistem pada identifikasi pola penyakit pada balita menggunakan algoritma apriori dilakukan dengan aplikasi berbasis *desktop* dengan bahasa pemrograman Java dan *database* yang digunakan untuk membangun sistem adalah MySQL.

4.6 Evaluasi Sistem

Berdasarkan uji coba sistem, evaluasi sistem dilakukan untuk mengukur akurasi aplikasi yang dibangun dalam klasifikasi penyakit. Evaluasi sistem membandingkan output sistem klasifikasi penyakit balita dengan menerapkan PCA dan KNN classifier terhadap sistem klasifikasi dengan PCA dan Rule Base system classifier. Setiap output sistem dibandingkan terhadap data asli untuk memperoleh akurasi masing-masing. Setelah didapatkan akurasi masing-masing akan dibandingkan mana yang lebih bagus antara akurasi klasifikasi dengan PCA dan KNN classifier atau akurasi klasifikasi dengan PCA dan Rule Base classifier.



BAB V HASIL DAN LUARAN YANG DICAPAI

5.1 Pengumpulan Data dan Informasi

Pengumpulan data dan informasi pada penelitian ini mengenai penyakit Balita. Penyakit Balita dengan keluhan batuk, demam, dan diare. Pengumpulan data dan informasi yang digunakan adalah melalui studi literatur dan wawancara.

Wawancara dilakukan terhadap salah satu tenaga medis di balai kesehatan Desa Soko, Kecamatan Glagah, Kabupaten Lamongan dan di salah satu rumah sakit Surabaya. Hasil wawancara yang diperoleh, yaitu dapat diketahui cara pengklasifikasian penyakit balita berdasarkan gejala dan keluhan yang diderita pasien. Tujuan lain dari wawancara sekaligus untuk memperoleh data balita sakit yang berobat di klinik atau rumah sakit.

5.2 Pengolahan dan Analisis Data

Data dan informasi yang sudah didapatkan dari proses pengumpulan data dan informasi akan diolah agar siap digunakan dalam proses penerapan algoritma untuk identifikasi penyakit balita. Jumlah data yang telah dikumpulkan sebanyak 600 data balita. Data yang telah diperoleh data pasien pada Desember 2016 sampai dengan Mei 2018.

5.2.1 Reduksi variabel

Principal Component Analysis (PCA) digunakan untuk reduksi variabel. Hasil PCA berupa korelasi antar variabel. variabel yang memiliki korelasi paling rendah akan dieliminasi. variabel yang digunakan untuk PCA adalah umur, tinggi badan, berat badan, jenis kelamin, suhu badan, dan 26 gejala. Langkah-langkah pada PCA adalah:

a. Menentukan variabel yang akan dianalisa.

Variabel yang dapat dianalisa hanya umur, tinggi badan, berat badan, jenis kelamin, suhu badan, dan 26 gejala. Untuk mempermudah penulisan variabel tersebut dibuatkan kode penamaannya, yaitu umur = X1, berat badan = X2, tinggi badan = X3, suhu tubuh = X4, jenis kelamin = X5, batuk = X6, diare = X7, demam = X8, anak tidak dapat minum atau menyusu = X9, muntah = X10, kejang = X11, anak-anak tampak tidak sadar = X12, nafas cepat = X13, tarikan dinding dada di dalam = X14, Stridor = X15, berak cair atau lembek = X16, mata cekung = X17, Cubitan kulit perut kembali lambat = X18, rewel / iritasi = X19, haus = X20, minum dengan lembut = X21, diare 14 hari atau lebih = X22, ada darah di tinja = X23, anak tidak bisa menunduk hingga dagu mencapai dada = X24, ruam kemerahan di kulit = X25, batuk

pilek atau mata merah = X26 , Kekeruhan pada kornea mata = X27, luka di mulut = X28, mata bernanah = X29, demam 2 - 7 hari = X30, Demam mendadak tinggi dan terus menerus = X31

b. Membuat matrik kovarian.

Matrik kovarian dapat dibuat dengan menghitung nilai kovarian antar variabel data. Setiap variabel X dihitung nilai hubungan antar variabel itu sendiri serta terhadap variabel lain. Misalnya, untuk variabel X1 dapat dihitung nilai kovarian terhadap X ditulis cov(X1, X1), X1 terhadap X2 ditulis cov(X1, X2), X1 terhadap X3 ditulis cov(X1, X3), X1 terhadap X4 ditulis cov(X1, X4). Cara yang sama berlaku untuk X2, X3, dan X4. Hasil perhitungan kovarian dapat dilihat pada Tabel 5.1.

Contoh Perhitungan kovarian untuk Cov(X1,X2):

$$\begin{aligned}
 &= (20 - 23.41)(10 - 10.89) + (24 - 23.41)(12 - 10.89) + \dots + (12 - 23.41)(9 - 10.89) (150 - 1) \\
 &= (-3.41)(-0.89) + (0.55)(1.11) + \dots + (-11.41)(-1.89) 149 \\
 &= 3.03 + 0.61 + \dots + 21.56 149 \\
 &= 42.334
 \end{aligned}$$

Tabel 5.1 Hasil perhitungan nilai kovarian

Kovarian	Hasil
Cov (X1, X1)	244.714
Cov (X1, X2)	42.334
Cov (X1, X3)	233.043
Cov (X1, X4)	2.176
Cov (X2, X1)	42.334
Cov (X2, X2)	10.149
Cov (X2, X3)	44.124
Cov (X2, X4)	0.613
Cov (X3, X1)	233.043
Cov (X3, X2)	44.124
Cov (X3, X3)	302.037
Cov (X3, X4)	2.961
Cov (X4, X1)	2.176
Cov (X4, X2)	0.613
Cov (X4, X3)	2.961
Cov (X4, X4)	0.74

Dari hasil perhitungan diatas dapat dibuat matrik kovarian seperti dibawah ini:

$$\text{Matrix kovarian} = \begin{bmatrix} 244.714 & 42.334 & 233.043 & 2.176 \\ 42.334 & 10.149 & 44.124 & 0.613 \\ 233.043 & 44.124 & 302.037 & 2.961 \\ 2.176 & 0.613 & 2.961 & 0.74 \end{bmatrix}$$

c. **Menghitung nilai eigen dan vektor eigen.**

Setelah matrik kovarian dibentuk maka dapat dicari nilai dan vektor eigen. Nilai dan vektor eigen dapat diselesaikan dengan menggunakan aplikasi yang disediakan oleh website: www.comnuan.com. Hasil pencarian nilai dan vektor eigen dapat dilihat pada Tabel 5.2.

Tabel 5.2 Nilai eigen dan vektor eigen

Nilai Eigen	542.8898	35.0508	2.4901	0.7403
Vektor Eigen	-0.6696	-0.7318	-0.1265	0.0096
	-0.1187	-0.0631	0.9887	-0.0663
	-0.7331	0.6786	-0.0452	-0.0079
	-0.0073	0.0082	0.0665	0.9977

d. **Menentukan Principal Component**

Vektor eigen diatas digunakan untuk menentukan Principal Component. Dari perhitungan tersebut dapat dibuat empat Principal Component yang dibentuk dengan mengalikan vektor eigen dengan variabel X.

$$Y1 = (-0.6696).X1 - 0.1187.X2 - 0.7331.X3 - 0.0073.X4$$

$$Y2 = (-0.7318).X1 - 0.0631.X2 + 0.6786.X3 + 0.0082.X4$$

$$Y3 = (-0.1265).X1 + 0.9887.X2 - 0.0452.X3 + 0.0665.X4$$

$$Y4 = 0.0096.X1 - 0.0663.X2 - 0.0079.X3 + 0.9977.X4$$

e. **Menemukan korelasi antara varibel dan Principal Component.**

Setiap Principal Component yang dibentuk dihitung nilai korelasinya dengan seluruh variabel. Berikut contoh perhitungan korelasi untuk Y1 dan X1.

$$r_{Y1,X1} = \frac{(-0.6696) \times \sqrt{542.8898}}{244.714} = -0.06375$$

Penyelesaian dari persamaan diatas diubah kedalam bentuk matrik, sehingga didapatkan sebuah matrik korelasi. Dengan mengamati matrik yang dibentuk dapat diketahui varibel yang kurang berpengaruh. Matrik korelasi yang dihasilkan dapat dilihat pada Tabel 5.3.

Tabel 5.3. Matrik korelasi

	X1	X2	X3	X4
Y1	-0.06107	-0.01113	-0.06896	-0.00067
Y2	-0.45480	-0.04140	0.40942	0.00521
Y3	-0.00069	0.00527	-0.00024	0.00054
Y4	0.01639	-0.11175	-0.0073	1.11736

Tabel 5.3 menunjukkan variabel X1 mempunyai anggota dengan nilai terkecil. Hal ini menunjukkan bahwa variabel X1 kurang berpengaruh untuk identifikasi penyakit balita. Dengan demikian variabel Umur(X1) tidak digunakan. Variabel yang digunakan adalah Berat Badan(X2), TinggiBadan(X3), dan Suhu Badan(X4).

Berdasarkan hasil eksperimen dengan perhitungan yang sama di atas menggunakan data training, maka variabel yang dihasilkan adalah 18 variabel, yaitu berat badan, tinggi badan, suhu tubuh, jenis kelamin, batuk, diare, demam, anak tidak boleh minum atau menyusu, anak-anak memuntahkan makanan mereka, anak-anak menderita kejang, nafas cepat, stridor, berak cair atau lembut, mata cekung, Cubitan kulit perut kembali lambat, rewel/ iritasi, diare 14 hari atau lebih, kekeruhan di kornea. Variabel ini digunakan untuk variabel dalam proses klasifikasi.

5.2.2 Klasifikasi

Proses selanjutnya adalah klasifikasi penyakit balita dengan menggunakan KNN Classifier dan rule based system. Variabel yang digunakan adalah variabel yang diperoleh dari PCA, yaitu berat badan, tinggi badan, suhu tubuh, jenis kelamin, batuk, diare, demam, anak tidak boleh minum atau menyusu, anak-anak memuntahkan makanan mereka, anak-anak menderita kejang, nafas cepat, stridor, berak cair atau lembut, mata cekung, Cubitan kulit perut kembali lambat, rewel/ iritasi, diare 14 hari atau lebih, kekeruhan di kornea.

1. K-NN Classifier

Cara kerja classifier menghitung jarak antar data testing terhadap data training. Berikut ini merupakan langkah-langkah KNN Classifier, yaitu :

a. Normalisasi

Data Normalisasi data dilakukan terhadap data training dan data testing. Variabel Jenis Kelamin mempunyai dua nilai yaitu "Laki-laki" dan "Perempuan", dinormalisasi nilainya menjadi "1" dan "0". Nilai "1" untuk Laki-laki dan nilai "0" untuk Perempuan. Untuk variabel Gejala mempunyai nilai "Ada" dan "Tidak ada", dinormalisasi menjadi bernilai "1" dan "0". Nilai "1" untuk normalisasi nilai "Ada" dan "0" untuk nilai "Tidak ada". Untuk variabel Umur, Berat Badan, dan Suhu Badan digunakan persamaan min-max normalization. Persamaan tersebut menggunakan nilai tertinggi dan nilai terkecil dari data training untuk setiap variabel.

Tabel.5.4 menunjukkan nilai Maximum dan Minimum dari data training.

Tabel 5.4. Nilai Maximum dan Minimum dari data training.

	Berat Badan	Tinggi Badan	Suhu badan
Max	21	120	40
Min	4.4	21	36

Contoh perhitungan min-max normalization untuk data pasien "Y" yang mempunyai umur 20 bulan, berat badan 10 kg, tinggi badan 79 dan suhu badan 37.5°C.

$$X_2' = \frac{X_2 - \min(X_2)}{\max(X_2) - \min(X_2)} = \frac{10 - 4.4}{21 - 4.4} = 0.33735$$

$$X_3' = \frac{X_3 - \min(X_3)}{\max(X_3) - \min(X_3)} = \frac{79 - 21}{120 - 21} = 0.58586$$

$$X_4' = \frac{X_4 - \min(X_4)}{\max(X_4) - \min(X_4)} = \frac{37.5 - 36}{40 - 36} = 0.375$$

Setelah dinormalisasi dengan min-max normalization kini pasien "Y" mempunyai nilai Berat Badan = 0.33735, Tinggi Badan = 0.58586, dan Suhu Badan = 0.375.

b. Hitung jarak Euclidean

Ukuran kemiripan suatu data terhadap kelas ditentukan dengan seberapa besar jarak kemiripannya. Semakin kecil jaraknya maka semakin besar kemiripan sifatnya terhadap suatu kelas. Persamaan Euclidean Distance digunakan untuk mengukur jarak kemiripan tersebut. Untuk setiap data testing dihitung jaraknya terhadap setiap data dalam data training. Contoh perhitungan jarak Euclidean data pasien "Z" terhadap data pasien "Y"

$$(Z, Y) = \sqrt{(1 - 0)^2 + (0.2048 - 0.3373)^2 + \dots + (0 - 0)^2} = 1.76339$$

Pada perhitungan tersebut mempunyai hasil akhir 1.76339, itu menunjukkan nilai jarak untuk data balita "Z" dan "Y". Rumus yang sama digunakan untuk menghitung jarak pasien "Z" terhadap pasien yang lain.

c. Weighted Voting

Kesalahan dalam memprediksi kelas dapat terjadi meskipun telah dihitung jarak kedekatannya dengan persamaan Euclidean Distance. Hal itu disebabkan kerana adanya noise atau data yang menyimpang jauh dari kelas asalnya.

Weighted Voting diperlukan untuk menentukan letak kelas dari data baru. Voting tersebut dilakukan dengan melakukan pembobotan untuk setiap kelas kelas yang ada. Dalam kasus ini terdapat 16 kelas penyakit balita, yaitu: Batuk, Pneumonia, Pneumonia Berat, Diare, Diare Dehidrasi Ringan, Diare Dehidrasi Berat, Diare Persisten, Diare Persisten Berat, Disentri, Demam, Demam dengan Tanda Bahaya Umum, Campak, Campak Dengan Komplikasi Berat, Campak Dengan Komplikasi, Demam Mungkin DBD, DBD. Dengan

menghitung total dari nilai Euclidean Distance untuk setiap kelas dapat ditentukan hasil akhir dari proses klasifikasi.

Perhitungan dibawah ini merupakan contoh Weighted Voting untuk kelas "Batuk" dari data training terdapat 45 data yang mempunyai diagnosa "batuk".

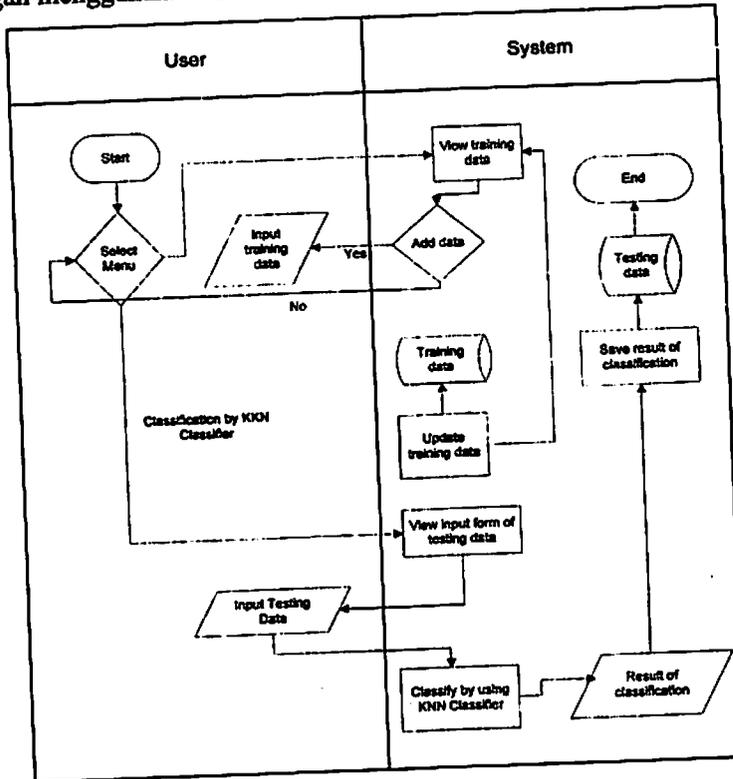
$$\text{voting (Cough)} = \left(\frac{1}{1.5413^2}\right) + \dots + \left(\frac{1}{1.05003^2}\right) = 146.8976$$

Untuk kelas penyakit balita yang lain juga diterapkan perhitungan yang sama dengan nilai yang berbeda-beda. Tabel 5.5 menunjukkan hasil dari Weighted Voting seluruh kelas penyakit balita. Kelas "Batuk" mempunyai skor "146.8976105", yang mana merupakan skor tertinggi dibandingkan kelas lain. Sehingga dapat ditarik kesimpulan Bahwa pasien "Z" menderita penyakit "Batuk".

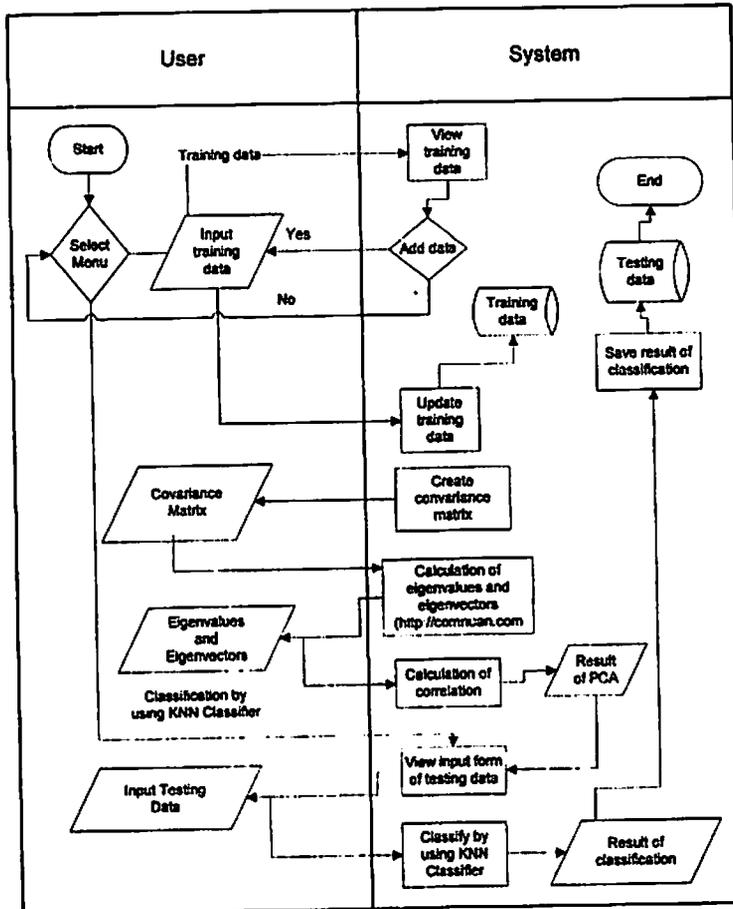
Tabel 5.5. Skor Weighted Voting

Penyakit balita	Bobot
Batuk	146.8976
Pneumonia	5.364010
Pneumonia Berat	3.265106
Diare	3.651260
Diare Dehidrasi Ringan	1.000566
Diare Dehidrasi Berat	0.848560
Diare Persisten	0.121155
Diare Persisten Berat	0.220797
Disentri	0.461117
Demam	5.183737
Demam dengan Tanda Bahaya Umum	4.284291
Campak	5.061522
Campak Dengan Komplikasi Berat	0.420538
Campak Dengan Komplikasi	4.064063
Demam Mungkin DBD	1.383750
DBD	0.661613

Perancangan sistem aplikasi PCA dan K-NN Classifier untuk klasifikasi balita digambarkan oleh Diagram flow yang dapat dilihat pada Gambar 5.1 menunjukkan klasifikasi untuk balita dengan menggunakan K-NN dan PCA dan Gambar 5.2 menunjukkan klasifikasi untuk balita dengan menggunakan K-NN



Gambar 5.1. Diagram Flow untuk system klasifikasi menggunakan K-NN



Gambar 5.2. Diagram Flow untuk system klasifikasi menggunakan PCA dan K-NN

2. Sistem Pakar

Sistem pakar yang dibangun menggunakan adalah *forward chaining*, apabila bagian premis (jika) terpenuhi maka bagian konklusi (maka) juga bernilai benar. Berikut langkah – langkah penelusuran dengan *forward chaining*, sebagai berikut:

- Langkah 1 : Mengajukan pertanyaan kepada user
- Langkah 2 : Menampung inputan dari *user* sebagai fakta yang diketahui pada *short term memory* yang disimpan pada tiap variable pertanyaan yang diajukan
- Langkah 3 : memeriksa rule berdasarkan fakta pada *short term memory* menggunakan metode *forward chaining*.
- Langkah 4 : jika rule ditemukan maka konklusi rule ditampung pada *short term memory*, apabila ada fakta baru maka langkah satu sampai dengan langkah empat diulang. Jika *rule* tidak ditemukan maka berikan *default output*.
- Langkah 5 : memberikan solusi.

Fakta yang sudah didapat, dibuat aturan (*rule*) berdasarkan penyakit. Gejala dapat dilihat pada Tabel 5.6. Keluhan terdiri dari K1 = batuk, K2= Diare, dan K3= Panas. Jenis penyakit dapat dilihat pada Tabel 5.7. Aturan dihasilkan berdasarkan gejala, keluhan, dan jenis penyakit. Aturan yang dihasilkan sebanyak 18 rule, yang dapat dilihat dalam Tabel 5.8.

Tabel 5.6. Gejala

Code	Gejala	Code	Gejala
G1	Anak tidak bisa minum atau menyusu	G14	Diare 14 hari atau lebih
G2	Anak memuntahkan makanan yang dimakan	G15	Ada darah dalam tinja
G3	Anak menderita kejang	G16	Demam
G4	Anak tampak letargis atau tidak sadar	G17	Kaku kuduk (anak tidak bisa menunduk hingga dagu mencapai dada)
G5	Napas cepat	G18	Ruam kemerahan di kulit
G6	Tarikan dinding dada ke dalam	G19	batuk pilek atau mata merah
G7	Stridor	G20	Kekeruhan pada kornea mata
G8	Berak cair atau lembek	G21	Luka di mulut
G9	Mata cekung	G22	Mata bernanah
G10	Cubitan kulit perut kembali lambat	G23	Demam 2 - 7 hari
G11	Gelisah, rewel/mudah marah	G24	Demam mendadak tinggi dan terus menerus
G12	Haus	G25	Tinja berwarna hitam / diare
G13	Anak tampak letargis atau tidak sadar	G26	Pilek /batuk

Tabel 5.7. Penyakit

Code	Jenis penyakit	Code	Jenis penyakit
P1	Tanda gejala umum	P10	Diare
P2	Batuk	P11	Demam
P3	Pneumonia	P12	Demam dengan Tanda Bahaya Umum
P4	Pneumonia Berat	P13	Campak
P5	Diare	P14	Campak Dengan Komplikasi Berat
P6	Diare Dehidrasi Ringan	P15	Campak Dengan Komplikasi
P7	Diare Dehidrasi Berat	P16	Demam Mungkin DBD
P8	Diare Persisten	P17	DBD
P9	Diare Persisten Berat	P18	Demam

Tabel 5.8. Rule yang dihasilkan

Rule	IF	THEN	Rule	IF	THEN
1	G1 OR G2 OR G3 OR G4	P1	10	P5 AND G15	P10
2	K1 AND G5	P2	11	K3 AND G16	P11
3	K1 AND G6	P3	12	P1 AND P11 OR G17	P12
4	K1 AND P1 OR G7	P4	13	P11 AND G18 AND G19 OR G21	P13
5	K2 AND G8	P5	14	P13 AND P1 AND G21 OR G20	P14
6	P5 AND G10 AND G11 OR G12 OR G13	P6	15	P13 AND G21 OR G22	P15
7	P5 AND G10 OR G12 OR G13	P7	16	P11 AND G23 AND G24	P16
8	P5 AND G14	P8	17	P11 AND G23 AND G24 OR G25 OR G11	P17
9	P8 AND P6 OR P7	P9	18	P11 AND G16 OR G26	P18

3. CBR

CBR terdiri dari 8 tahap, yaitu : membangun representasi kasus, pembobotan oleh pakar, penghitungan similarity local, penghitungan tingkat keyakinan, penghitungan similarity global, pemilihan nilai tertinggi, proses revise dan penyimpanan data di database.

a. Membangun Representasi Kasus

Variabel dibagi menjadi dua jenis variabel yaitu umum dan khusus. variabel umum ditujukan terhadap variabel yang memiliki sifat umum, seperti tinggi badan, berat badan, suhu badan, usia, dan jenis kelamin. Variabel khusus merupakan variabel yang hanya berlaku pada penyakit-penyakit tertentu, seperti 26 variabel gejala. Masing-masing variabel dikelompokkan sebagai berikut (DepKes, 2010) :

1. Jenis kelamin

Jenis kelamin memiliki 2 item-set, yaitu laki-laki dan perempuan.

2. Berat Badan

Berat badan memiliki lima item-set, yaitu yaitu B1, B2, B3, B4, dan B5. Pengelompokan berat badan dapat dilihat pada Tabel 5.9.

Tabel 5.9 Range Berat Badan

Atribut Berat Badan	Range
B1	3 – 9 Kg
B2	10 – 12 Kg
B3	13 – 14 Kg
B4	15 – 16 Kg
B5	17 – 19 Kg

3. Tinggi Badan

Tinggi badan memiliki lima item-set, yaitu T1, T2, T3, T4, dan T5. Pengelompokan tinggi badan dapat dilihat pada Tabel 5.10.

Tabel 5.10 Range Tinggi Badan

Atribut Tinggi Badan	Range
T1	49 – 76 Cm
T2	77 – 87 Cm
T3	88 – 96 Cm
T4	97 – 103 Cm
T5	104 – 110 Cm

b. Pembobotan

Pembobotan dilakukan oleh seorang pakar. Pembobotan adalah memberikan bobot untuk setiap variabel. Pembobotan dihitung dengan analisa statistik, diawali dengan melakukan perhitungan banyaknya variabel yang muncul dalam target kasus pada data training. Pembobotan dilakukan beberapa kali iterasi. Proses penghitungan pembobotan dengan analisa statistik memiliki dua syarat yang harus dipenuhi. Syarat pertama, nilai similarity global tertinggi seharusnya adalah nilai perbandingan antara 1 data testing dan data training yang sesuai dengan diagnosis data dokter. Apabila nilai similarity global tertinggi yang didapat tidak sesuai dengan diagnosis dokter, maka nilai bobot dari diagnosis sistem harus diubah agar menghasilkan nilai similarity yang lebih rendah dari nilai similarity global hasil diagnosis dokter. Proses perubahan bobot disesuaikan dengan kesamaan variabel dari data testing dan data training. Jika nilai kesamaan variabel adalah 1 dan nilai bobotnya tinggi, maka bobot tersebut harus dikurangi. Jika nilai kesamaan variabel gejala adalah 0 maka bobot tersebut harus ditambah. Syarat kedua adalah melakukan normalisasi pembobotan dengan cara menjaga

nilai total seluruh bobot penyakit memiliki jumlah yang sama, dalam hal ini jumlah bobot penyakit yang ditentukan berjumlah 290.

c. Penghitungan Similarity Lokal

Penghitungan similarity local adalah proses membandingkan nilai antara variabel gejala antara kasus baru dengan kasus lama. Similarity local untuk type data numeric digunakan untuk variabel usia, jenis kelamin, berat badan, dan tinggi badan. Similarity local untuk type data Boolean digunakan untuk variabel gejala.

d. Penghitungan Tingkat Keyakinan

Proses yang dilakukan dalam tingkat keyakinan adalah menghitung perbandingan antara banyaknya variabel yang terdapat dalam kasus baru yang muncul pada data kasus lama dengan banyaknya variabel yang terdapat dalam kasus lama dalam database. Hasil tingkat keyakinan digunakan untuk mengetahui apakah kasus baru dapat melanjutkan proses perhitungan ke tahap berikutnya atau harus menjalankan proses reverse. Nilai tingkat keyakinan dari keseluruhan data yang memiliki nilai < 0.75 (threshold) dimasukkan kedalam proses reverse.

e. Penghitungan Similarity Global

Algoritma yang digunakan dalam proses similarity global adalah Nearest Neighbor Similarity, minkowski distance similarity, dan euclidean distance similarity. Contoh perhitungan similarity global antara kasus baru dengan kasus lama dengan data 1 dan data 2 sebagai berikut :

a. Perhitungan similarity global antara kasus baru dengan kasus lama dengan data 1

1. Nearest Neighbor Similarity:

$$SimNN(T, S) = \frac{\sum_{i=1}^n (f_i(S_i, T_i) \cdot w_{L,p(s)})}{\sum_{i=1}^n (w_{L,p(s)})} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} = \frac{282.8}{290} * 50\% * 1 = 0.49$$

2. Euclidian Distance Similarity:

$$SimMD(S, T) = \left[\frac{\sum_{i=1}^n (w_{L,p(s)})^2 * |f_i(S_i, T_i)|^2}{\sum_{i=1}^n (w_{L,p(s)})^2} \right]^{\frac{1}{2}} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} = \frac{3728.03}{290} * 150\% * 1 = 0.1$$

3. Minkowsky Distance Similarity:

$$SimMD(S, T) = \left[\frac{\sum_{i=1}^n (w_{L,p(s)})^3 * |f_i(S_i, T_i)|^3}{\sum_{i=1}^n (w_{L,p(s)})^3} \right]^{\frac{1}{3}} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} = \frac{51367.09}{290} * 150\% * 1 = 0.06$$

b. Perhitungan similarity global antara kasus baru dengan kasus lama dengan data 2

1. Nearest neighbor Similarity:

$$SimNN(T, S) = \frac{\sum_{i=1}^n (f_i(S_i, T_i)(w_{i,p(s)}))}{\sum_{i=1}^n (w_{i,p(s)})} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} = \frac{249.15}{290} * 40\% * 0 = 0$$

2. Minkowsky Distance Similarity:

$$SimMD(S, T) = \left[\frac{\sum_{i=1}^n (w_{i,p(s)})^3 * |f_i(S_i, T_i)|^3}{\sum_{i=1}^n (w_{i,p(s)})^3} \right]^{\frac{1}{3}} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} = \frac{3113.3}{290} * 40\% * 0 = 0$$

3. Euclidian Distance Similarity:

$$SimMD(S, T) = \left[\frac{\sum_{i=1}^n (w_{i,p(s)})^2 * |f_i(S_i, T_i)|^2}{\sum_{i=1}^n (w_{i,p(s)})^2} \right]^{\frac{1}{2}} * P(S) * \frac{J(S_i, T_i)}{J(T_i)} = \frac{37068.7}{290} * 40\% * 0 = 0$$

f. Pemilihan Nilai Tertinggi

Nilai tertinggi dibandingkan antara penghitungan similarity global antar data testing. Nilai terbesar yang didapat menunjukan diagnose pada kasus baru. Hasil pemilihan nilai tertinggi dapat dilihat pada Tabel 5.11.

Tabel 5.11. Pemilihan Nilai Tertinggi

No	Uji	Data 1	Data 2	Hasil
1	NNS	0,49	0	Diare Dehidrasi Ringan
2	MDS	0.1	0	Diare Dehidrasi Ringan
3	EDS	0.06	0	Diare Dehidrasi Ringan

g. Proses Revise

Proses revise adalah proses yang hanya dilakukan apabila data yang dibandingkan pada saat proses penghitungan tingkat keyakinan memiliki nilai < 0.75. Proses ini dilakukan dengan menanyakan kasus yang tidak dapat didiagnosa oleh sistem kepada ahli. Penentuan sistem dapat atau tidak dapat menjawab diagnose adalah dari nilai tingkat keyakinan. Tingkat keyakinan yang memiliki nilai < 0.75 memiliki arti bahwa nilai tersebut tidak menyerupai seluruh data yang terdapat dalam database.

h. Menyimpan Data Dalam Database

Apabila kasus baru sudah dihitung dengan similarity global dan mendapatkan diagnose, maka data kasus baru akan dimasukkan kedalam database. Data tersebut selanjutnya dianggap