



**PREDIKSI PENYAKIT HEPATITIS MENGGUNAKAN
METODE *PRINCIPAL COMPONENT ANALYSIS* DAN
SUPPORT VECTOR MACHINE (PCA-SVM)**

Skripsi

Disusun sebagai salah satu syarat
untuk memperoleh gelar Sarjana Komputer
Program Studi Teknik Informatika

Oleh

Triyana Fadila

4611415007

**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG**

2020

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa skripsi saya yang berjudul “Prediksi Penyakit Hepatitis Menggunakan Metode *Principal Component Analysis* dan *Support Vector Machine* (PCA-SVM)” disusun berdasarkan penelitian saya dengan arahan dosen pembimbing. Dan saya menyatakan bahwa skripsi ini bebas plagiat, kecuali yang secara tertulis dirujuk dalam skripsi ini dan disebutkan dalam daftar pustaka. Apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Semarang, 28 Juli 2020



Triyana Fadila
4611415007

PERSETUJUAN PEMBIMBING

Nama : Triyana Fadila
NIM : 4611415007
Program Studi : S-1 Teknik Informatika
Judul Skripsi : Prediksi Penyakit Hepatitis Menggunakan Metode *Principal Component Analysis* dan *Support Vector Machine* (PCA-SVM).

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian skripsi Program Studi Teknik Informatika FMIPA UNNES.

Semarang, 28 Juli 2020

Pembimbing



Dr. Alamsyah, S.Si., M.Kom.
NIP. 197405172006041001

PENGESAHAN

Skripsi yang berjudul

Prediksi Penyakit Hepatitis Menggunakan Metode *Principal Component Analysis* Dan *Support Vector Machine* (PCA-SVM)

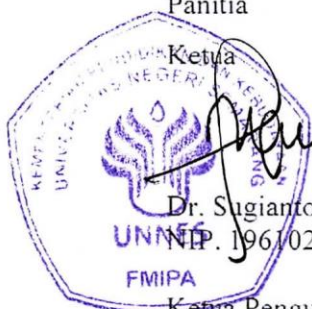
disusun oleh

Triyana Fadila

4611415007

Telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 28 Juli 2020.

Panitia



Ketua

Dr. Sugianto, M.Si.
NIP. 196702191993031001

Ketua Penguji

Endang Sugiharti, S.Si., M.Kom.
NIP. 197401071999032001

Pembimbing

Dr. Alamsyah, S.Si., M.Kom.
NIP. 197405172006041001

Sekretaris

Dr. Alamsyah, S.Si., M.Kom.
NIP. 197405172006041001

Anggota Penguji

Riza Arifudin, S.Pd., M.Cs.
NIP. 198005252005011001

MOTTO DAN PERSEMBAHAN

MOTTO

Hasbunallah wani'mal wakil ni'mal maula wani'man nasir.

There's no power nor strength except through Allah.

PERSEMBAHAN

Skripsi ini ku persembahkan kepada:

1. Orang tua tercinta, Bapak Fachruri dan Ibu Dairoh. Terima kasih atas do'a yang senantiasa mengalir dan kasih sayang yang tak pernah terputus.
2. Kakak dan adik tersayang, Ahmad Mustofa Kamal, Istik Maliya Muslimah dan Difa Rizkiana Fauziah. Terima kasih telah memberi naungan dan dukungan yang melimpah.
3. Sahabat-sahabat terdekat serta teman-teman Ilkom angkatan 2015.
4. Almamater Unnes.

PRAKATA

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan segala rahmat dan hidayah-Nya dalam penyusunan skripsi, sehingga penulis dapat menyelesaikan skripsi dengan judul “Prediksi Penyakit Hepatitis Menggunakan Metode *Principal Component Analysis* dan *Support Vector Machine* (PCA-SVM)”.

Skripsi ini dapat diselesaikan karena adanya kerja sama, bantuan dan motivasi dari berbagai pihak. Ucapan terima kasih ini penulis tujukan kepada yang terhormat:

1. Bapak Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang, yang telah memberikan kesempatan kepada penulis untuk menyelesaikan Studi Strata 1 di Jurusan Ilmu Komputer Program Studi Teknik Informatika FMIPA UNNES;
2. Bapak Dr. Sugianto, M.Si., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang yang telah memberikan ijin kepada penulis untuk menyusun skripsi;
3. Bapak Dr. Alamsyah, S.Si., M.Kom., selaku Ketua Jurusan Ilmu Komputer sekaligus Dosen Pembimbing yang telah meluangkan waktu, membantu, membimbing, mengarahkan dan memberikan saran dan motivasi sehingga penulis dapat menyelesaikan skripsi.
4. Ibu Endang Sugiharti, S.Si., M.Kom., selaku Dosen Penguji 1 yang telah memberikan kritik, saran serta masukan dalam proses penyelesaian skripsi ini.

5. Bapak Riza Arifudin, S.Pd., M.Cs., selaku Sekretaris Jurusan sekaligus Dosen Penguji 2 yang telah membantu, memberikan kritik, saran serta masukan dalam proses penyelesaian skripsi ini.
6. Ayah saya tercinta, Bapak Fachruri yang selalu mendoakan serta melalui keringatnya selalu memberi dukungan dalam proses penyelesaian skripsi ini.
7. Ibu saya tersayang, Ibu Dairoh yang harapan dan doanya selalu mengalir, selalu memberi dukungan serta motivasi dalam proses penyelesaian skripsi ini.
8. Kakak dan adik yang saya sayangi, Ahmad Mustofa Kamal, Istik Maliya Muslimah dan Difa Rizkiana Fauziah yang selalu mendukung saya dalam menyelesaikan skripsi ini.
9. Sahabat saya Alifa, Amanah, Anggita, Doni, Fachrizal, Puput, Revi, Rika, Selvi, Sinta, Umay, Vinca, teman-teman satu bimbingan dan seluruh teman-teman dari Ilkom angkatan 2015 yang telah memberikan banyak motivasi, inspirasi dan semangat.
10. Semua pihak yang tidak dapat disebutkan satu-persatu, terima kasih atas segala bentuk dukungan dan dorongannya.

Semoga seluruh pihak yang telah membantu penulis dalam penyelesaian skripsi ini mendapatkan imbalan kebaikan dari Allah SWT.

Semarang, 28 Juli 2020

Triyana Fadila

ABSTRAK

Triyana, Fadila. 2020. *Prediksi Penyakit Hepatitis Menggunakan Metode Principal Component Analysis dan Support Vector Machine (PCA-SVM)*. Skripsi. Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Dr. Alamsyah, S.Si., M.Kom.

Kata kunci: *Support Vector Machine, Principal Component Analysis, Hepatitis*

Seiring meningkatnya penggunaan teknologi, volume data yang dihasilkan dari penggunaan teknologi komputer juga terus meningkat tak terkecuali data yang dihasilkan dari bidang medis. Dengan *data mining*, data medis dapat dimanfaatkan untuk membantu dokter dalam mendiagnosa suatu penyakit termasuk hepatitis. Salah satu teknik *data mining* yang dapat digunakan untuk memprediksi suatu keputusan adalah klasifikasi. Dari banyaknya algoritma klasifikasi, *Support Vector Machine* (SVM) digunakan dalam penelitian ini karena dapat memberikan hasil yang baik dalam banyak penelitian. Namun, jika data yang diproses memiliki dimensi yang tinggi SVM akan membutuhkan waktu komputasi yang lama dan membutuhkan memori yang besar. Maka dari itu *Principal Component Analysis* (PCA) diterapkan untuk mengesktrasi fitur dan mereduksi jumlah fitur atau atribut. PCA dapat mereduksi dimensi data tanpa menghilangkan banyak informasi dari *dataset*. Data yang digunakan dalam penelitian ini adalah *dataset* hepatitis yang diambil dari *UCI Machine Learning Repository*. Atribut atau fitur *dataset* hepatitis berjumlah 19 atribut dan 1 atribut kelas. Jumlah atribut tersebut kemudian direduksi menggunakan metode PCA dari 19 menjadi 8 atribut. Data yang telah diekstrak dan direduksi kemudian digunakan untuk klasifikasi SVM. Hasil akurasi yang dicapai adalah sebesar 93,55%. Sedangkan hasil klasifikasi SVM pada *dataset* hepatitis tanpa metode PCA sebesar 87,1%. Dapat disimpulkan bahwa kinerja algoritma SVM dalam mengklasifikasi *dataset* hepatitis lebih baik setelah diterapkan metode PCA.

DAFTAR ISI

	halaman
PERNYATAAN KEASLIAN SKRIPSI.....	ii
PERSETUJUAN PEMBIMBING	iii
PENGESAHAN	iv
MOTTO DAN PERSEMBAHAN	v
PRAKATA.....	vi
ABSTRAK.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	6
BAB 2 TINJAUAN PUSTAKA	8
2.1 Tinjauan Pustaka	8
2.2 Landasan Teori.....	10
2.2.1 <i>Data Mining</i>	10
2.2.2 Klasifikasi.....	13
2.2.3 <i>Support Vector Machine</i>	14
2.2.4 <i>Principal Component Analysis</i>	17
2.2.5 Validasi dan Evaluasi.....	19
2.2.6 Hepatitis.....	21
BAB 3 METODE PENELITIAN	24
3.1 Studi Literatur	24
3.2 Pengambilan Data	24
3.3 Pengolahan Data.....	25
3.4 Klasifikasi Data.....	29

3.5 Metode yang Diusulkan	32
3.6 Perancangan Sistem.....	35
3.7 Penarikan Kesimpulan.....	36
BAB 4 HASIL DAN PEMBAHASAN	37
4.1 Hasil Penelitian	37
4.2 Pembahasan.....	58
BAB 5 PENUTUP	62
5.1 Kesimpulan	62
5.2 Saran.....	62
DAFTAR PUSTAKA	64
LAMPIRAN-LAMPIRAN	67
Lampiran 1. Hasil Imputasi <i>Missing Value</i> pada <i>Dataset Hepatitis</i>	68
Lampiran 2. Data Hasil Standardisasi.....	76
Lampiran 3. <i>Dataset</i> Hasil Proses <i>Principal Component Analysis (PCA)</i>	84

DAFTAR TABEL

Tabel	halaman
Tabel 2.1 <i>Confusion Matrix</i>	20
Tabel 2.2 Atribut Dataset Hepatitis.....	22
Tabel 3.1 Deskripsi Atribut <i>dataset</i> Hepatitis.....	25
Tabel 4.1 Sampel <i>Missing Value</i> pada Dataset Hepatitis.....	40
Tabel 4.2 Sampel Hasil Penanganan <i>Missing Value</i> dengan Modus	41
Tabel 4.3. Sampel Hasil Standardisasi.....	43
Tabel 4.4 Contoh Hasil Perhitungan Matrik Kovarian	44
Tabel 4.5 Nilai Eigen dan Vektor Eigen yang Telah Diurutkan.....	46
Tabel 4.6 <i>Feature Vector</i> Hasil Pemilihan Komponen.....	47
Tabel 4.7 Sampel <i>Dataset</i> PCA	48
Tabel 4.9 <i>Confussion Matrix</i> SVM.....	49
Tabel 4.8 <i>Confussion Matrix</i> SVM-PCA.....	50
Tabel 4.10 Penelitian Terkait	61

DAFTAR GAMBAR

Gambar	halaman
Gambar 2.1 Proses <i>Data Mining</i>	13
Gambar 2.1 Pemisahan Dua Kelas Data dengan Margin Maksimum.....	16
Gambar 3.1 Diagram Alir Metode <i>Principal Component Analysis</i> (PCA).....	29
Gambar 3.2 Diagram Alir Rancangan Algoritma SVM dengan PCA	33
Gambar 4.1 Sampel <i>Dataset</i> Hepatitis dengan Format <i>DATA file</i>	38
Gambar 4.2 Sampel <i>Dataset</i> Hepatitis dalam Format <i>CSV</i>	38
Gambar 4.3 Kode Program <i>Python</i> untuk Imputasi <i>Missing Value</i>	40
Gambar 4.4 Tampilan Halaman Awal Aplikasi.....	52
Gambar 4.7 Tampilan Utama Menu <i>Dahsboard</i>	54
Gambar 4.8 Tampilan Submenu <i>Dataset Asli</i>	54
Gambar 4.9 Tampilan Submenu <i>Dataset Diolah</i>	55
Gambar 4.10 Tampilan Submenu <i>Dataset PCA</i>	55
Gambar 4.11 Tampilan Submenu <i>Prediksi Hepatitis</i>	56
Gambar 4.12 Tampilan Submenu <i>Bantuan</i>	56
Gambar 4.13 Kode Program Proses <i>Mining</i> dengan Algoritma SVM.....	57
Gambar 4.14 Kode Program Algoritma <i>PCA</i> dan <i>SVM</i>	58
Gambar 4.15 Hasil Akurasi.....	60

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan pesatnya perkembangan teknologi, pemanfaatan sistem informasi berbasis komputer telah meluas di berbagai bidang tak terkecuali di bidang medis atau kesehatan. Industri kesehatan sejauh ini mampu menghasilkan sejumlah besar data yang akan terus meningkat. Peningkatan volume data ini membutuhkan cara otomatis agar data tersebut dapat diekstraksi saat diperlukan (Milovic & Milovic, 2012: 69). Banyaknya data rekam medis pasien dapat diolah dengan memanfaatkan teknik *data mining*. *Data mining* merupakan solusi yang mampu menemukan kandungan informasi yang tersembunyi berupa pola dan aturan dari sekumpulan data yang besar agar mudah dipahami (Handarko & Alamsyah, 2015: 158). *Data mining* berkaitan dengan bidang ilmu lain seperti sistem basis data, *data warehousing*, statistik, *machine learning*, pengambilan informasi, dan komputasi canggih. Selain itu, *data mining* dapat diterapkan pada disiplin ilmu seperti jaringan syaraf tiruan, pengenalan pola, analisis spasial data, basis data gambar dan pemrosesan sinyal (Sugiharti *et al.*, 2017: 903).

Data mining telah digunakan secara luas dalam bidang bioinformatika untuk menganalisa data biomedis (Hussien *et al.*, 2018: 1). Analisis tersebut banyak dilakukan untuk memprediksi atau mendiagnosa suatu penyakit, salah satunya adalah penyakit hepatitis. Besarnya jumlah data pasien hepatitis dapat digunakan para peneliti untuk mengekstraksi informasi yang berguna dengan memanfaatkan teknik

data mining seperti teknik klasifikasi. Klasifikasi merupakan teknik yang dapat digunakan untuk memprediksi data atau menggambarkan kelas data. Algoritma klasifikasi *data mining* tersebut dapat dimanfaatkan dan membantu ahli medis dalam mendiagnosa suatu penyakit (Mirqotussa'adah *et al.*, 2017: 136).

Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas atau konsep data (Han *et al.*, 2012: 18). Klasifikasi diaplikasikan pada banyak hal seperti deteksi penipuan, target pemasaran, prediksi kinerja, manufaktur, diagnosis medis dan sebagainya. Terdapat banyak algoritma yang dapat dipakai untuk klasifikasi salah satunya adalah *Support Vector Machine* (SVM). Menurut Purnami *et al.*, (2015: 477), SVM merupakan salah satu teknik *data mining* yang banyak diteliti dan menunjukkan keberhasilan dalam banyak aplikasi, algoritma ini dikenal dengan akurasi dan prediksi yang tinggi. Akurasi sendiri sering dijadikan sebagai ukuran dalam diagnosa medis, dimana kesalahan dalam mendeteksi kondisi medis yang fatal dapat mengakibatkan kematian pada pasien (Cai *et al.*, 2018: 146).

Penyakit hepatitis sendiri adalah penyakit peradangan pada hati (*liver*) yang dapat disebabkan oleh faktor genetik, infeksi virus, alkohol, dan obat-obatan. Berdasarkan laporan global *World Health Organization* (WHO), kematian yang disebabkan oleh virus hepatitis meningkat pada tahun 2017, setiap tahunnya terdapat 1,75 juta orang baru yang terinfeksi virus hepatitis khususnya hepatitis C. Menentukan diagnosa hepatitis bagi dokter termasuk tugas yang sulit (Polat & Güneş, 2006: 889), karena banyak faktor harus dipertimbangkan dalam prosedur

diagnosis penyakit (Nilashi *et al.*, 2019: 13). Seorang dokter umumnya menentukan keputusan

dengan mengevaluasi hasil tes seorang pasien atau merujuk pada keputusan sebelumnya yang ditentukan pada pasien lain dengan kondisi yang sama (Çalışır & Dogantekin, 2011: 10705). Maka dari itu penelitian ini bertujuan untuk memprediksi keputusan diagnosis hepatitis dengan menerapkan teknik klasifikasi menggunakan algoritma SVM. Pengembangan sistem diagnosis hepatitis yang akurat dapat membantu dalam mendeteksi hepatitis dan membantu pengambilan keputusan bagi dokter.

Data yang digunakan dalam penelitian ini adalah *dataset* hepatitis yang bersumber dari <https://archive.ics.uci.edu/ml/datasets/hepatitis> (*UCI Machine Learning Repository*). *Dataset* hepatitis mengandung 20 atribut/fitur termasuk atribut *class*. Untuk memperoleh hasil yang lebih baik, data perlu dianalisis terlebih dahulu dengan mengekstraksi data dan mereduksi jumlah atribut. Algoritma SVM membutuhkan waktu komputasi yang lama dan membutuhkan memori yang besar, jika data yang diproses memiliki dimensi yang tinggi atau memiliki atribut yang banyak (Kotu & Deshpande, 2015: 147). Oleh karena itu, metode *Principal Component Analysis* (PCA) dipilih untuk mengekstraksi fitur-fitur yang penting dan mereduksi jumlah fitur. Banyak algoritma *data mining* yang bekerja dengan baik apabila dimensi (jumlah atribut/fitur pada data) lebih rendah. Konsep PCA adalah mengurangi dimensi *dataset*, dengan mempertahankan sebanyak mungkin variabilitas yaitu informasi statistik sebanyak mungkin (Jolliffe & Cadima, 2016: 1).

Berdasarkan uraian permasalahan diatas, maka diusulkan penelitian dengan judul “*Prediksi Penyakit Hepatitis Menggunakan Metode Principal Component Analysis dan Support Vector Machine (PCA-SVM)*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah dalam penelitian ini adalah bagaimana penerapan algoritma *Support Vector Machine* dengan *Principal Component Analysis* untuk memprediksi penyakit hepatitis?

1.3 Batasan Masalah

Pada penelitian ini diperlukan batasan-batasan agar tujuan penelitian dapat tercapai. Adapun batasan masalah yang dibahas pada penelitian ini adalah sebagai berikut:

1. Algoritma klasifikasi yang digunakan adalah *Support Vector Machine* (SVM).
2. Data yang digunakan sebagai objek penelitian ini adalah dataset hepatitis yang bersumber dari *UCI Machine Learning Repository*.
3. Bahasa pemrograman yang digunakan adalah *Python*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah untuk menerapkan algoritma *Support Vector Machine* dengan *Principal Component Analysis* sebagai metode ekstraksi dan reduksi fitur dalam memprediksi hepatitis.

1.5 Manfaat Penelitian

Mengetahui perbandingan hasil klasifikasi dari algoritma *Support Vector Machine* sebelum dan sesudah diterapkan metode *Principal Component Analysis* dalam memprediksi hepatitis.

1.6 Sistematika Penulisan

1.6.1 Bagian Awal Skripsi

Bagian awal skripsi terdiri dari halaman judul, halaman pengesahan, halaman pernyataan, halaman motto dan persembahan, abstrak, kata pengantar, daftar isi, daftar gambar, daftar tabel dan daftar lampiran.

1.6.2 Bagian Isi Skripsi

Bagian isi skripsi terdiri dari lima bab, yaitu sebagai berikut:

1. BAB 1: PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan skripsi.

2. BAB 2: TINJAUAN PUSTAKA

Bab ini berisi penjelasan mengenai definisi maupun pemikiran-pemikiran yang dijadikan kerangka teoritis yang menyangkut dan mendasari pemecahan masalah dalam skripsi ini.

3. BAB 3: METODE PENELITIAN

Bab ini berisi penjelasan mengenai studi pendahuluan, tahap pengumpulan data, dan tahap pengembangan sistem.

4. BAB 4: HASIL DAN PEMBAHASAN

Bab ini berisi hasil penelitian beserta pembahasannya.

5. BAB 5: PENUTUP

Bab ini berisi simpulan dari penulisan skripsi dan saran yang diberikan penulis untuk mengembangkan skripsi ini.

1.6.3 Bagian Akhir Skripsi

Bagian akhir skripsi ini berisi daftar pustaka yang merupakan informasi mengenai buku-buku, sumber-sumber dan referensi yang digunakan penulis serta lampiran-lampiran yang mendukung dalam penulisan skripsi ini.

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Penelitian ini dikembangkan berdasarkan beberapa referensi yang mempunyai keterkaitan metode dan objek penelitian. Penggunaan referensi ini ditujukan untuk memberikan batasan-batasan terhadap metode dan sistem yang nantinya akan dikembangkan lebih lanjut. Berikut adalah hasil dari penelitian sebelumnya.

Penelitian mengenai diagnosis hepatitis telah banyak dilakukan, Kumar *et al.*, (2014) dalam penelitian mereka dengan judul “*Data Analysis and Prediction of Hepatitis Using Support Vector Machine (SVM)*”, menerapkan *Support Vector Machine* untuk memprediksi dan memantau tingkat hepatitis dalam pasien. Metode yang digunakan adalah metode *wrapper* pada WEKA untuk menghapus *record* yang tidak penting serta Rapid Miner untuk menganalisis data. Uji *Chi square* dilakukan sebagai tahap prapemrosesan data. Hasil penelitian ini menunjukkan bahwa menggunakan metode *wrapper* dan SVM dapat meningkatkan akurasi hingga 83%. Namun mereka tidak menggunakan waktu eksekusi sebagai pengukuran kinerja

Çalisir & Dogantekin (2011) melalui penelitian yang berjudul “*A New Intelligent Hepatitis Diagnosis System: PCA–LSSVM*”, menyajikan metode baru yang disebut PCA-LSSVM untuk diagnosa penyakit hepatitis. Metode yang diusulkan berdasarkan pada *Principal Component Analysis (PCA)* untuk ekstraksi

dan reduksi serta *Least Square Support Vector Machine* (LSSVM) untuk klasifikasi. Akurasi yang diperoleh pada penelitian tersebut adalah 96.12%.

Suwardika (2017) dalam jurnalnya yang berjudul “*Pengelompokan dan Klasifikasi pada Data Hepatitis dengan Menggunakan Support Vector Machine (SVM), Classification and Regression Tree (CART) dan Regresi Logistik Biner*” melakukan penelitian pada *dataset* hepatitis dengan menggunakan K-Means dan Kernel K-Means untuk klasterisasi data serta algoritma SVM, CART, dan Regresi Logistik Biner untuk mengklasifikasi data hasil klasterisasi. Akurasi terbaik yang diperoleh antara lain sebesar 83,2% dengan menggunakan algoritma CART pada *dataset* asli, 100% dengan menggunakan algoritma CART pada data hasil K-Means, serta sebesar 98,7% dengan menggunakan algoritma SVM dan CART pada data hasil Kernel K-Means.

Abdullah *et al.*, (2018) dalam penelitiannya yang berjudul “*High-Resolution Systems for Automated Diagnosis of Hepatitis*” menerapkan *Support Vector Machine* (SVM), *multilayer perceptron* (MLP) dan *K-nearest neighbor* (KNN) untuk meningkatkan akurasi dalam mendiagnosis hepatitis. Hasil akurasi tertinggi yang didapatkan dengan menggunakan algoritma SVM adalah 97.87% ketika sampel yang digunakan sejumlah 80 sampel dengan pembagian data sebesar 70% untuk data latih dan 30% untuk data uji. Untuk algoritma *multilayer perceptron* (MLP) diperoleh akurasi antara 97,8% dan 80% dengan pembagian data latih dan data uji yang berbeda. Sedangkan untuk algoritma KNN diperoleh akurasi tertinggi yaitu sebesar 100%.

Sedangkan Nilashi *et al.*, (2019) dalam penelitiannya yang berjudul “*A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique*”, menggunakan metode *hybrid* untuk mendiagnosa hepatitis. Metode yang digunakan yaitu *Non-linear Iterative Partial Least Squares* (NIPALS) untuk mereduksi dimensi, *Self-Organizing Map* (SOM) untuk mengkluster data, CART untuk seleksi fitur, dan *Neuro-Fuzzy Infer-ence System* (ANFIS) *ensemble* untuk proses prediksi hepatitis. Hasil yang diperoleh adalah akurasi sebesar 93,06%.

2.2 Landasan Teori

2.2.1 Data Mining

Data mining adalah kegiatan menemukan pola yang menarik dan pengetahuan dari sejumlah besar data. Menurut Witten *et al.*, (2011: 39), *data mining* adalah pemecahan masalah dengan menganalisa data yang sudah ada sebelumnya, dan didefinisikan sebagai proses dari penemuan pola pada suatu data. Diantara tujuan data mining adalah untuk melakukan klasifikasi, klusterisasi, menemukan pola asosiasi hingga melakukan peramalan (*predicting*) (Iskandar & Suprpto, 2015: 38).

Data mining dapat diterapkan pada semua jenis data, selama data tersebut berguna untuk target *system/aplikasi*. Secara umum, bentuk data untuk data mining adalah data dari *database*, data *warehouse* dan data transaksional. *Data mining* merupakan bagian dari proses *Knowledge Discovery in Databases* (KDD), yaitu proses transformasi data mentah menjadi informasi yang berguna. KDD merupakan proses ekstraksi informasi yang berpotensi berguna dan tidak diketahui sebelumnya dari sebuah kumpulan data (Bramer, 2007: 2). *Data mining* sebagai proses penemuan

pengetahuan (*knowledge discovery*) ditunjukkan pada Gambar 2.1 sebagai urutan iteratif dari langkah-langkah berikut (Han *et al.*, 2011: 6):

1. Pembersihan data (*Data cleaning*)

Merupakan proses untuk menangani nilai yang hilang, mengurangi *noise* ketika mengidentifikasi *outlier data* dan memperbaiki data yang tidak

konsisten. *Missing value*, *noise* dan data yang tidak konsisten akan membuat data menjadi tidak akurat.

2. Integrasi data (*Data intregation*)

Merupakan proses untuk menggabungkan data dari beberapa *database*. Integrasi data dapat meningkatkan akurasi dan kecepatan proses penambangan data. Namun, integrasi data harus dilakukan dengan hati-hati agar mengurangi dan menghindari data yang *redundant* dan data yang tidak konsisten.

3. Seleksi data (*Data selection*)

Merupakan proses pengambilan data yang relevan dari *database* untuk dianalisis.

4. Transformasi data (*Data transformation*)

Merupakan proses untuk mengubah data kedalam format yang dapat diproses oleh *data mining*. Sehingga proses penambangan data lebih efisien dan pola yang ditemukan mungkin lebih mudah dipahami.

5. Proses *mining*

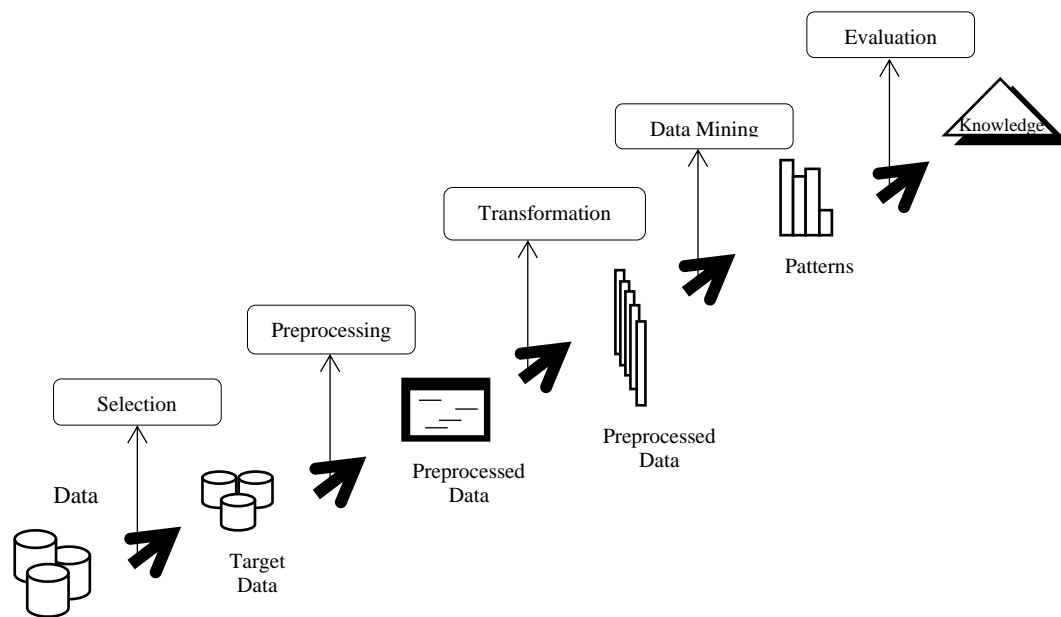
Merupakan proses utama dimana metode diterapkan untuk mendapatkan pola dan pengetahuan dari data.

6. Evaluasi pola (*Evaluation pattern*)

Untuk mengidentifikasi pola yang benar-benar berharga.

7. Presentasi pengetahuan (*Knowledge presentation*)

Untuk menyajikan pengetahuan yang didapat kepada pengguna.



Gambar 2.1 Proses *Data Mining*

2.2.2 Klasifikasi

Menurut Han *et al.*, (2011: 18), klasifikasi adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep. Klasifikasi merupakan teknik *data mining* yang dapat digunakan untuk memprediksi keanggotaan kelompok untuk data *instance* (Wijaya & Muslim, 2016: 22). Tiap teknik klasifikasi menggunakan suatu algoritma pembelajaran untuk mendapatkan suatu model yang paling memenuhi hubungan antara himpunan atribut dan label kelas dalam data masukan (Hermanto *et al.*, 2020: 213). Model diturunkan berdasarkan analisis data pelatihan yaitu pada objek data yang label kelasnya diketahui. Model dapat digunakan untuk memprediksi label kelas dari objek yang label kelasnya tidak diketahui atau dapat memprediksi kecenderungan data-data yang muncul di masa depan.

Ada 2 proses dalam klasifikasi, yaitu:

1. *Learning/training*

Pada tahap ini algoritma klasifikasi akan membangun model klasifikasi dengan menganalisis data latih. Atribut label kelas dan model atau pengelompokan yang dipelajari direpresentasikan dalam bentuk aturan.

2. *Classification* (Klasifikasi)

Tahap di mana model yang telah dibangun digunakan untuk memprediksi label kelas untuk data terkait. Data uji digunakan untuk memperkirakan akurasi dari aturan klasifikasi. Jika akurasi dianggap dapat diterima, aturan dapat diterapkan pada klasifikasi tupel data baru.

2.2.3 *Support Vector Machine*

Support vector machines (SVM) pertama kali diusulkan oleh Vladimir Vapnik dan dipresentasikan pertama kali pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. SVM adalah algoritma untuk klasifikasi data linear dan nonlinear (Han *et al.*, 2012: 437). SVM merupakan salah satu algoritma pembelajaran yang dapat menganalisis data dan mengenali pola, dan digunakan untuk melakukan klasifikasi maupun regresi (Kesavaraj & Sukumaran, 2013: 5). *Support vector machine* adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi.

Konsep dasar dari klasifikasi SVM adalah mencari *hyperplane* (garis batas) terbaik yang berfungsi sebagai pemisah dua buah kelas data. Pencarian bidang pemisah terbaik inilah yang menjadi inti dari *Support Vector Machine* (Munawarah

et al., 2016: 105). *Hyperplane* terbaik dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* dengan data terdekat dari masing-masing kelas. Data yang paling dekat dengan *hyperplane* pada masing-masing kelas tersebut disebut sebagai *support vector* (Yusuf & Priambadha, 2013: 14).

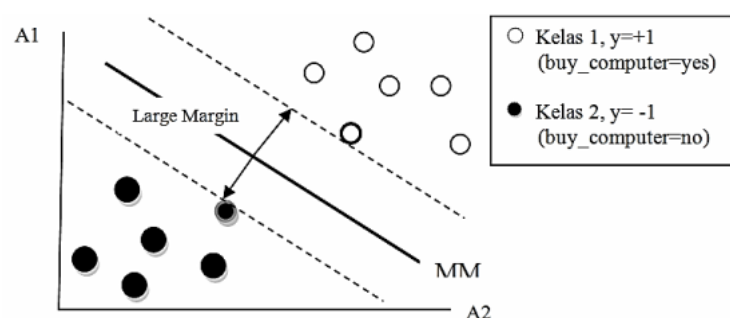
Menurut (Han *et al.*, 2011: 408), langkah awal suatu algoritma SVM adalah pendefinisian persamaan suatu *hyperplane* pemisah yang dituliskan dengan Persamaan 1.

$$W \cdot X + b = 0 \quad (1)$$

Di mana w merupakan suatu bobot vektor, yaitu $W = \{w_1, w_2, w_3, \dots, w_n\}$, n adalah jumlah atribut dan b merupakan suatu skalar yang disebut dengan bias. Jika berdasarkan pada atribut A1, A2 dengan permisalan tupel pelatihan $X = (x_1, x_2)$, dimana x_1 dan x_2 merupakan nilai dari atribut A1 dan A2, dan jika b dianggap sebagai suatu bobot tambahan w_0 , maka persamaan suatu *hyperplane* pemisah dapat ditulis seperti pada Persamaan 2.

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (2)$$

Setelah persamaan dapat didefinisikan, nilai x_1 dan x_2 dapat dimasukkan ke dalam persamaan untuk mencari bobot w_1 , w_2 , dan w_0 atau b . Grafik pemisahan dua kelas data dengan margin maksimum dapat dilihat pada Gambar 2.2.



Gambar 2.1 Pemisahan Dua Kelas Data dengan Margin Maksimum

SVM menemukan *hyperlane* pemisah maksimum, yaitu *hyperlane* yang mempunyai jarak maksimum antara tupel pelatihan terdekat. *Support vector* ditunjukkan dengan batasan tebal pada titik tupel. Dengan demikian, setiap titik yang terletak di atas *hyperplane* pemisah memenuhi Persamaan 3.

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (3)$$

Sedangkan, titik yang terletak di bawah *hyperlane* pemisah memenuhi rumus seperti pada Persamaan 4.

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (4)$$

Berdasarkan persamaan di atas, bobot dapat disesuaikan sehingga *hyperplanes* yang mendefinisikan sisi dari *margin* dapat ditulis dalam Persamaan 5 dan 6.

$$H_1 = w_0 + w_1x_1 + w_2x_2 \geq 0, \text{ untuk } y_i = +1 \quad (5)$$

$$H_1 = w_0 + w_1x_1 + w_2x_2 \leq 0, \text{ untuk } y_i = -1 \quad (6)$$

Artinya, setiap tupel yang berada di atas H_1 memiliki kelas +1, dan setiap *tupel* yang berada di bawah H_2 memiliki kelas -1. Untuk menemukan *Maksimum Margin Hyperplane* (MMH) pada SVM dapat dilakukan dengan menggunakan *Lagrangian formulation* dan kemudian solusi dapat dipecahkan dengan kondisi *Karush-Kuhn-Tucker* (KKT). Berdasarkan *Lagrangian formulation*, *Maksimum Margin Hyperplane* (MMH) dapat ditulis ulang sebagai suatu batas keputusan (*decision boundary*) seperti pada Persamaan 7:

$$d(X^T) = \sum_{i=1}^l y_i a_i X_i X^T + b_0 \quad (7)$$

Dimana y_i adalah label kelas *support vector* X_i dan X^T merupakan tupel tes. α_i dan b_0 merupakan parameter numerik yang ditentukan otomatis oleh optimalisasi algoritma SVM dan l merupakan jumlah dari *support vector*.

2.2.4 *Principal Component Analysis*

Principal Component Analysis (PCA) adalah sebuah teknik untuk membangun variabel-variabel baru yang merupakan kombinasi linear dari variabel-variabel asli. Metode ini ditemukan pertama kali oleh Karl Pearson pada tahun 1901, kemudian dikembangkan oleh Harold Hotelling pada 1933. Kemudian pada tahun 1947 teori ini dikenalkan kembali oleh Karhunen, dan dikembangkan oleh Loeve pada tahun 1963, sehingga teori ini juga dinamakan Karhunen-Loeve transform pada bidang ilmu telekomunikasi.

Menurut (Smith, 2002: 12), PCA dapat mengidentifikasi pola dalam data untuk menyoroti persamaan dan perbedaan di dalam data tersebut, dimana pola dalam data mungkin sulit ditemukan dalam data berdimensi tinggi (representasi grafis sulit dilakukan). Keuntungan utama lain dari PCA adalah bahwa setelah menemukan pola dalam data, data dapat dikompresi dengan mengurangi jumlah dimensi tanpa kehilangan banyak informasi.

Tujuan utama dari analisis komponen utama (PCA) adalah untuk mengurangi dimensi *dataset* yang terdiri dari sejumlah besar atribut yang mungkin saling berkorelasi, dengan mempertahankan sebanyak mungkin variansi *dataset*. Hal ini dicapai dengan mentransformasikan variabel asli yang mungkin berkorelasi menjadi variabel baru yang tidak berkorelasi. Variabel baru ini disebut sebagai *principal component* (PC). Jumlah maksimum variabel baru (PC) adalah sama dengan jumlah

variabel aslinya (Sharma, 1996: 58). PC tersebut kemudian diurutkan sehingga variabel-variabel pertama merupakan variabel yang mempertahankan sebagian besar variansi dari keseluruhan variabel aslinya (Jolliffe, 2002: 1).

Secara garis besar, langkah-langkah metode *Principal Component Analysis* (PCA) adalah sebagai berikut:

1. Standardisasi data

Standarisasi data biasanya diperlukan dalam berbagai algoritma *machine learning*. Sebuah *dataset* bisa saja memiliki perbedaan skala yang mencolok. Perbedaan skala ini dapat menyebabkan ketidaksimetrisan persebaran data (*variance*) sehingga *dataset* tersebut perlu distandarkan. PCA tidak dapat bekerja maksimal apabila variabel-variabelnya memiliki unit pengukuran yang berbeda dan cenderung mengabaikan fitur-fitur tertentu (Jolliffe & Cadima, 2016: 5). Untuk mengatasinya, biasanya perlu dilakukan standarisasi pada *dataset*. Standarisasi merupakan salah satu dari metode *scaling* data agar skala fitur berada di antara nilai minimum dan maksimum yang diberikan, umumnya antara 0 dan 1. *Scaling* data penting dilakukan karena banyak algoritma *machine learning* yang akan bekerja tidak maksimal apabila fitur-fiturnya memiliki skala yang berbeda. Demikian juga dengan algoritma *Support Vector Machine* (SVM) yang dapat menghasilkan performa tidak maksimal jika tidak dilakukan *scaling* terlebih dahulu pada datanya, hal ini dikarenakan SVM menggunakan fungsi jarak (*distance function*) untuk optimasinya.

2. Menghitung matrik kovarian (*covariance matrix*)

Kovarian adalah ukuran besarnya hubungan antara dua dimensi. Jika kovarian dihitung dari satu dimensi dengan dimensi itu sendiri, maka hasilnya adalah varians.

3. Menghitung nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*) dari matrik kovarian tersebut

Nilai eigen dan vektor eigen mengandung informasi yang berguna dari suatu *dataset* dan menentukan *principal component* (PC) yang terlibat.

4. Membentuk *feature vector*

Feature vector secara sederhana adalah matrik yang berupa vektor eigen yang ingin dipertahankan. Langkah ini merupakan awal dari konsep pengurangan dimensi. Vektor eigen yang dipilih adalah yang memiliki nilai eigen terbesar setelah diurutkan dari urutan terbesar ke terkecil.

5. Membentuk *dataset* baru

Langkah ini dapat dilakukan dengan mengalikan transpose dari *feature vector* dengan transpose dari *dataset* yang telah distandardisasi. Tujuannya adalah untuk mengubah orientasi data dari sumbu asli ke sumbu yang diwakili oleh *principal component*.

2.2.5 Validasi dan Evaluasi

Metode yang diusulkan diawali dengan membagi *dataset* menjadi data latih dan data uji. Dalam penelitian ini pembagian data dilakukan dengan menggunakan metode *splitter*. Metode ini membagi data menjadi dua sub himpunan dengan proporsi 80% untuk data latih dan 20% untuk data uji. Pembagian ini bersifat acak dengan konsistensi pengacakan (*random state*) tertentu, sehingga tiap eksekusi

nilainya tidak berubah. Data latih akan diproses dengan algoritma *Support Vector Machine* untuk menghasilkan model klasifikasi. Kemudian model klasifikasi tersebut diuji dengan data uji. Hasil dari proses tersebut digunakan untuk mengukur kinerja algoritma dari metode yang diusulkan.

Evaluasi model klasifikasi didasarkan pada pengujian untuk memprediksi objek yang benar dan salah, pengukuran kinerja klasifikasi dilakukan dengan *confusion matrix*. Menurut Han *et al.*, (2011: 365), *confusion matrix* adalah alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali tupel dari kelas yang berbeda. *Confusion matrix* dapat memberikan keputusan yang diperoleh dari proses *training* dan *testing* (Sutarti *et al.*, 2019: 70), dengan cara menghitung jumlah kelas terprediksi terhadap kelas aktual. Hasil tersebut dinyatakan dalam *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TP dan TN menyatakan bahwa *classifier* mengenali tupel dengan benar, artinya tupel positif dikenali sebagai positif dan tupel negatif dikenali sebagai negatif. Sedangkan FP dan FN menyatakan bahwa *classifier* salah dalam mengenali tupel, artinya tupel positif dikenali sebagai negatif dan tupel negatif dikenali sebagai positif. Tabel *confusion matrix* dapat ditunjukkan pada pada Tabel 2.1.

Tabel 2.1 *Confusion Matrix*

Klasifikasi	Kelas hasil prediksi			Jumlah
	Ya	Tidak		
Kelas aktual	Ya	TP	FN	P
	Tidak	FP	TN	N
	Jumlah	P'	N'	P+N

Metode *confussion matrix* digunakan untuk menghitung akurasi yang diperoleh dalam proses klasifikasi. Akurasi adalah persentase dari total data yang

diklasifikasikan secara benar. Pengukuran akurasi dapat dituliskan dengan Persamaan 15.

$$Accuracy = \frac{TP+TN}{P+N} \times 100\% \quad (15)$$

Selain akurasi, dalam klasifikasi juga dihitung tingkat kesalahan atau (*error rate*). Pengukuran tingkat kesalahan dapat dituliskan dengan Persamaan 16.

$$Error Rate = \frac{FP+FN}{P+N} \times 100\% \quad (16)$$

2.2.6 Hepatitis

Hepatitis adalah suatu penyakit peradangan hati yang umumnya disebabkan oleh virus (Handarko & Alamsyah, 2015: 158). Selain itu hepatitis juga dapat disebabkan alkohol dan autoimun. Hepatitis virus dapat terjadi melalui penggunaan jarum yang terkontaminasi virus (seperti melalui suntikan narkoba, suntikan obat, jarum transfuse, jarum tato, dan tindik tubuh melakukan hubungan seks dengan seseorang yang terinfeksi hepatitis, atau menjadi petugas kesehatan yang bekerja dengan pasien hepatitis juga beresiko tertular infeksi hepatitis. Terdapat 5 jenis virus hepatitis, yaitu A, B, C, D, dan E. Karakteristik dari masing-masing jenis ini berbeda, sehingga gejala dan pengobatannya juga beragam.

Hepatitis virus menyebar di seluruh dunia dan menjadi masalah utama kesehatan masyarakat global (Nilashi *et al.*, 2019: 13). Tidak semua kasus hepatitis menimbulkan gejala, Secara umum, gejala hepatitis antara lain demam, mual sampai muntah, letargi (kelelahan), mudah memar, jaundice (sakit kuning) dan sebagainya. Jika dibiarkan tanpa pengobatan, hepatitis dapat mengakibatkan sirosis (kerusakan hati permanen), dan pada akhirnya gagal hati. Cara terbaik untuk memeriksa hepatitis

adalah dengan tes darah, selain itu dapat juga melalui biopsi hati yaitu sepotong kecil jaringan hati diambil untuk pemeriksaan laboratorium. Selain itu, dokter dapat mendiagnosis hepatitis melalui pemeriksaan fisik untuk gejala hepatitis seperti kulit atau mata yang menguning. Pemeriksaan riwayat juga diperlukan untuk mengetahui dari mana pasien terkena virus hepatitis. Penyakit hepatitis dapat dicegah dengan cara menghindari faktor-faktor resiko penularan hepatitis dan melakukan vaksinasi atau imunisasi.

Data yang digunakan dalam penelitian ini adalah *dataset* hepatitis yang diperoleh dari <https://archive.ics.uci.edu/ml/datasets/hepatitis> (*UCI Machine Learning Repository*). Tujuan *dataset* adalah untuk memprediksi ada atau tidaknya penyakit hepatitis menggunakan hasil berbagai tes medis yang dilakukan pada pasien dan menentukan apakah pasien dengan penyakit hepatitis akan hidup atau mati (Sartakhti et al., 2012: 571). *Dataset* hepatitis memiliki 19 atribut (fitur) serta 1 atribut *class*. Terdapat 155 sampel yang terdistribusi ke dalam dua kelas yaitu kelas “Die” 32 (20.6%) sampel dan kelas “Live” 123 (79.4%) sampel. *Dataset* ini mengandung *missing value*. Atribut *dataset* hepatitis dapat dilihat pada Tabel 2.2.

Tabel 2.2 Atribut Dataset Hepatitis

No.	Nama atribut	Keterangan
1	<i>Class</i>	Die, Live
2	<i>Age</i>	7-78
3	<i>Sex</i>	Male, Female
4	<i>Steroid</i>	Yes, No
5	<i>Antivirals</i>	Yes, No
6	<i>Fatigue</i>	Yes, No
7	<i>Malaise</i>	Yes, No

8	<i>Anorexia</i>	Yes, No
9	<i>Liver big</i>	Yes, No
10	<i>Liver firm</i>	Yes, No
11	<i>Spleen palpable</i>	Yes, No
12	<i>Spiders</i>	Yes, No
13	<i>Ascites</i>	Yes, No
14	<i>Varices</i>	Yes, No
15	<i>Bilirubin</i>	0.3-8
16	<i>Alk phosphate</i>	26-295
17	<i>SGOT</i>	14-648
18	<i>Albumin</i>	2.1-5.3
19	<i>Protime</i>	0-100
20	<i>Histology</i>	Yes, No

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan terkait penerapan kombinasi algoritma *Support Vector Machine* (SVM) dan *Principal Component Analysis* (PCA) pada *dataset* hepatitis yang diperoleh dari *UCI Machine Learning Repository* dapat ditarik kesimpulan bahwa penerapan algoritma PCA pada klasifikasi *dataset* hepatitis dapat mereduksi dimensi atau fitur *dataset* melalui pembentukan *dataset* baru. Atribut yang semula berjumlah 19 atribut dapat direduksi menjadi 8 atribut dengan jumlah *instance* tetap yaitu 155 *instance*. Hasil klasifikasi dengan penerapan algoritma SVM menggunakan *dataset* PCA adalah akurasi sebesar 93,54%. Sedangkan hasil klasifikasi dengan menerapkan algoritma SVM tanpa proses *preprocessing* adalah akurasi sebesar 87,01%. Dengan demikian, dapat dikatakan bahwa penerapan PCA pada klasifikasi algoritma SVM dapat mereduksi data dan memprediksi hepatitis dengan akurasi yang baik yaitu sebesar 93,54 %.

5.2 Saran

Untuk pengembangan penelitian yang lebih lanjut, penulis memberikan saran sebagai berikut.

1. Melakukan uji coba menggunakan *dataset* yang memiliki fitur lebih banyak.
2. Menerapkan algoritma *resampling* untuk mengatasi ketidakseimbangan kelas data pada *dataset* hepatitis.

3. Menentukan kernel yang digunakan dalam menerapkan algoritma *Support Vector Machine*.

DAFTAR PUSTAKA

- Abdullah, H.N., Abd, B.H. & Muhi, S.H. 2018. High-Resolution Systems for Automated Diagnosis of Hepatitis. *2018 3rd Scientific Conference of Electrical Engineering, SCEE 2018*, 39–44.
- Bramer, M. 2007. *Principles of Data Mining*. London: Springer.
- Cai, T., He, H. & Zhang, W. 2018. Breast Cancer Diagnosis Using Imbalanced Learning and Ensemble Method. *Applied and Computational Mathematics*, 7(3): 146–154.
- Çalışır, D. & Dogantekin, E. 2011. A new intelligent hepatitis diagnosis system: PCA-LSSVM. *Expert Systems with Applications*, 38(8): 10705–10708.
- Delsen, M. S. N. V., Wattimena, A. Z., Saputri, S. D. 2017. Penggunaan Metode Analisis Komponen Utama untuk Mereduksi Faktor-Faktor Inflasi di Kota Ambon. *Jurnal Ilmu Matematika dan Terapan*, 11(2): 109-118.
- Han, J., Kamber, M. & Pei, J. 2011. *Data Mining: Concepts and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann Publishers Inc.
- Handarko, J.L. & Alamsyah. 2015. Implementasi Fuzzy Decision Tree Untuk Mendiagnosa Penyakit Hepatitis. *Unnes Journal of Mathematics*, 4(2): 157–164.
- Hermanto, Mustopa, A. & Kuntoro, A.Y. 2020. Algoritma Klasifikasi Naive Bayes dan Support Vector Machine dalam Layanan Komplain Mahasiswa. *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, 5(2): 211–220.
- Hussien, S.O., Elkhatem, S.S., Osman, N. & Ibrahim, A.O. 2018. A review of data mining techniques for diagnosing hepatitis. *Proceedings of: 2017 Sudan Conference on Computer Science and Information Technology*, 1–6.
- Iskandar, D. & Suprpto, Y.K. 2015. Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara Algoritma C 4.5 dan Naive Bayes. *Jurnal Ilmiah NERO*, 2(1): 37–43.
- Jolliffe, I.T. 2002. *Principal Component Analysis* (2nd ed.). New York: Springer.
- Jolliffe, I.T. & Cadima, J. 2016. Principal component analysis: a review and recent developments Subject Areas. *Phil.Trans.R.Soc.A 374*: 20150202.
- Kesavaraj, G. & Sukumaran, S. 2013. A study on classification techniques in data mining. *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*.
- Kotu, V. & Deshpande, B. 2015. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. San Francisco: Morgan Kaufmann Publishers

Inc.

- Kumar, C.B., Kumar, M.V., Gayathri, T. & Kumar, S.R. 2014. Data Analysis and Prediction of Hepatitis Using Support Vector Machine (SVM). *International Journal of Computer Science and Information Technologies*, 5(2): 2235–2237.
- Mäs, S., Henzen, D., Bernard, L. & Müller, M. 2018. Generic Schema Descriptions for Comma-Separated Values Files of Environmental Data. *21st International Conference on Geographic Information Science (AGILE 2018)*, 1–5.
- Milovic, B. & Milovic, M. 2012. Prediction and decision making in Health Care using Data Mining. *International Journal of Public Health Science (IJPHS)*, 1(2): 69–76.
- Mirqotussa'adah, M., Muslim, M.A., Sugiharti, E., Prasetyo, B. & Alimah, S. 2017. Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 8(2): 135–143.
- Muhtadi. 2017. Penerapan Principal Component Analysis (PCA) dalam Algoritma K-Means untuk Menentukan Centroid pada Clustering. *Jurnal Konstanta*, 1(1): 121–142.
- Munawarah, R., Soesanto, O. & Faisal, M.R. 2016. Penerapan Metode Support Vector Machine pada Diagnosa Hepatitis. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 4(1): 103–113.
- Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O. & Akbari, E. 2019. A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. *Journal of Infection and Public Health*, 12(1): 13–20.
- Polat, K. & Güneş, S. 2006. Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation. *Digital Signal Processing: A Review Journal*, 16(6): 889–901.
- Purnami, S.W., Andari, S. & Pertiwi, Y.D. 2015. High-Dimensional Data Classification Based on Smooth Support Vector Machines. *Procedia Computer Science*, 72: 477–484.
- Rahayu, G. & Mustakim. 2017. Principal Component Analysis untuk Dimensi Reduksi Data Clustering Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia. In: *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*. pp.201–208.
- Rea, A., & Rea, W. 2016. How Many Components should be Retained from a Multivariate Time Series PCA?. *arXiv preprint arXiv:1610.03588*, 1-49.

- Sartakhti, J.S., Zangoeei, M.H. & Mozafari, K. 2012. Hepatitis disease diagnosis using a novel hybrid method based on *Support Vector Machine* and simulated annealing (SVM-SA). *Computer Methods and Programs in Biomedicine*, 108(2): 570–579.
- Sharma, S. 1996. *Applied Multivariate Techniques*. New York: John Wiley & Sons, Inc. New.
- Smith, L.I. 2002. A tutorial on Principal Components Analysis (Computer Science Technical Report No. OUCS-2002-12).
- Sugiharti, E., Firmansyah, S. & Devi, F.R. 2017. Predictive evaluation of performance of computer science students of unnes using data mining based on naïve bayes classifier (NBC) algorithm. *Journal of Theoretical and Applied Information Technology*, 95(4): 902–911.
- Sutarti, S., Putra, A.T. & Sugiharti, E. 2019. Comparison of PCA and 2DPCA Accuracy with K- Nearest Neighbor Classification in Face Image Recognition. *Scientific Journal of Informatics*, 6(1): 64–72.
- Suwardika, G. 2017. Pengelompokan dan Klasifikasi pada Data Hepatitis dengan Menggunakan Support Vector Machine (SVM), Classification and Regression Tree (CART) dan Regresi Logistik Biner. *Journal of Education Research and Evaluation*, 1(3): 183–191.
- Wijaya, K.P. & Muslim, M.A. 2016. Peningkatan Akurasi pada Algoritma Support Vector Machine dengan Penerapan Information Gain untuk Mendiagnosa Chronic Kidney Disease. *Seminar Nasional Ilmu Komputer*, (Snik): 22–27.
- Witten, I.H., Frank, E. & Hall, M. a. 2011. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). San Francisco: Morgan Kaufmann Publishers Inc.
- Yusuf, A. & Priambadha, T. 2013. Support Vector Machines Yang Didukung K-Means Clustering Dalam Klasifikasi Dokumen. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 11(1): 15.